

Getting to Know CO:

Compute Ontario's Role in the DRI Ecosystem

Scholars Portal Days
May 5, 2023



Jeff Moon
Director, Data Strategy & Services
Compute Ontario
moonj@computeontario.ca

Hi, and welcome to this discussion introduction to Compute Ontario and its expanded role in the DRI ecosystem. I appreciate the opportunity to speak with you today. Bienvenue, et merci pour l'opportunit  de vous parler aujourd'hui. Let me start by acknowledging that I am speaking to you from Kingston Ontario, which is situated on traditional Anishinaabe (**Ah-nish-in-ah-bay**) and Haudenosaunee (**Ho-den-o-show-nee**) territory. We are grateful to be able to live, learn and play on these lands."

In keeping with the theme of this year's Scholars Portal Days, "**Refresh/Reframe/Renew**" and more specifically, with the title of this panel, "**New Partnerships, New Perspectives: Rethinking Approaches to Long-Standing Problems**", I thought I'd take a few moments to help solve the problem that many people don't know **what Compute Ontario is**, and **what it does**. My apologies to some in the Ontario Data Community who may have seen a version of this presentation recently – but I think reinforcement can't hurt.



Since June of 2022, I have had the privilege and pleasure of working with Compute Ontario, which is led by CEO Ranil Sonnadara and supported by the rest of the CO team shown here.

Compute Ontario's Role
in Ontario's DRI Ecosystem

Researchers



Starting with the basics, everything Compute Ontario does is aimed to help researchers conduct research successfully.

Compute Ontario's Role
in Ontario's DRI Ecosystem

DRI: ARC, RDM, & RS



Researchers

Compute Ontario does this by supporting DIGITAL RESEARCH INFRASTRUCTURE, including ADVANCED RESEARCH COMPUTING, RESEARCH DATA MANAGEMENT, AND RESEARCH SOFTWARE...

Compute Ontario's Role
in Ontario's DRI Ecosystem

DRI: ARC, RDM, & RS

Updated mandate explicitly acknowledging
Research Data Management
as an integral component of DRI



Researchers

RDM became a part of this picture through a recently updated mandate from the provincial government, explicitly acknowledging **Research Data Management (RDM)** as an integral component of DRI.

Compute Ontario's Role
in Ontario's DRI Ecosystem

Funded by the Ministry of Colleges and Universities (MCU) to bring together partners and stakeholders to coordinate provincial and federal investments and strategy for Ontario's DRI

Trusted source of unbiased strategic advice and leadership in the Ontario DRI space



DRI: ARC, RDM, & RS

Researchers

Compute Ontario is funded by the Ministry of Colleges and Universities (MCU) to bring together partners and stakeholders to coordinate provincial and federal investments and strategy for Ontario's DRI, AND serves as a **Trusted source of unbiased strategic advice and leadership in the Ontario DRI space**

Compute Ontario's Role
in Ontario's DRI Ecosystem

PARTNERS & STAKEHOLDERS

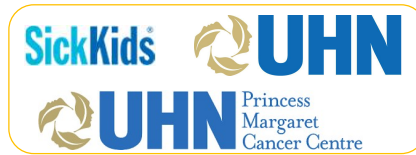


DRI: ARC, RDM, & RS

Researchers

Not surprisingly, there are a lot of **partners and stakeholders** in the Ontario DRI ecosystem... I'll provide a **whirlwind overview** of some of these here.

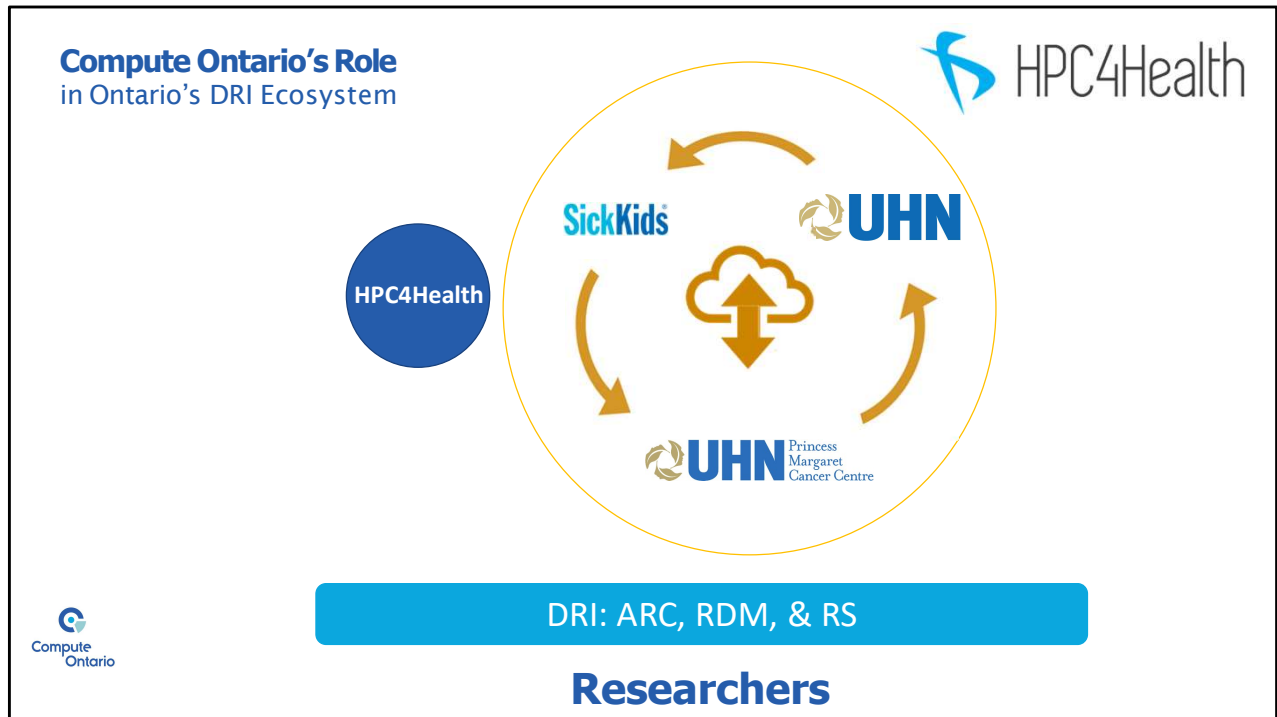
Compute Ontario's Role
in Ontario's DRI Ecosystem



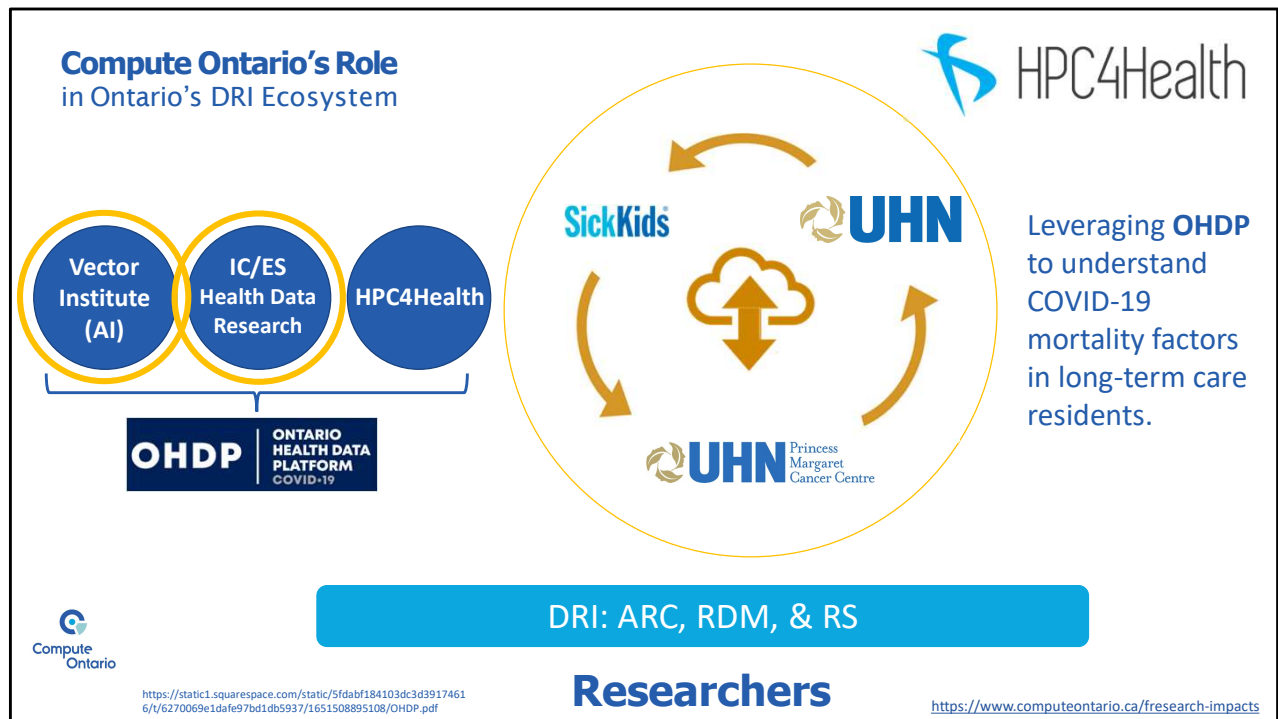
DRI: ARC, RDM, & RS

Researchers

Let's start with **HPC4Health** which is a consortium of the **Hospital for Sick Children + Princess Margaret Cancer Centre + the University Health Network or UHN** that together share the HPC resources provided by HPC4Health.



Using HPC4Health, each of these health care institutions can access their own, fully private, cloud-based HPC through a centralized "brain" that monitors each institution's usage and expands and shrinks capacity based on demand. This shared cloud model allows institutions to take advantage of each others' excess resource capacity, saving researchers from having to request additional resources themselves.



Continuing our focus on health-related partners: I-C-E-S is a not-for-profit research institute that supports cutting-edge studies on health care delivery and population outcomes using a range of demographic and health-related data. **[CLICK]**

The Vector Institute for Artificial Intelligence is a health-focused, independent, not-for-profit dedicated to research in the field of artificial intelligence (AI), machine, and deep learning to generate more accurate health insights and predictions. **[CLICK]**

Together, HPC4Health, ICES, and the Vector Institute support the Ontario Health Data Platform, or OHDP, a secure, regulatory-compliant research computing environment for applying AI and machine learning techniques. **[CLICK]** Researchers are leveraging the OHDP to address pressing COVID-19 and pandemic challenges such as those affecting long-term care residents.

Compute Ontario's Role in Ontario's DRI Ecosystem

Vector
Institute
(AI)

IC/ES
Health Data
Research

HPC4Health

Centre for
Advanced
Computing
(CAC)



DRI: ARC, RDM, & RS

Researchers

<https://static1.squarespace.com/static/5fdabf184103dc3d39174616/t/6243a9e2da48593da4b950ef/1648601570442/Nathan+Brinklow+CAC+Research+Impact+Story.pdf>

<https://www.computeontario.ca/fresearch-impacts>

The **Centre for Advanced Computing (CAC)** at **Queen's University** helps **fulfill** the HPC needs hundreds of Canadian research groups, comprising thousands of researchers working in a variety of fields.

Compute Ontario's Role
in Ontario's DRI Ecosystem



Vector
Institute
(AI)

IC/ES
Health Data
Research

HPC4Health

Centre for
Advanced
Computing
(CAC)

SHARCNET
19 colleges &
universities



GRAHAM
@
U Waterloo



https://static1.squarespace.com/static/5fdabf184103dc3d99174616/t/63d13483c125c142c3d26bda/1674654851736/CO-AnnualReport-2022_Trant.pdf

DRI: ARC, RDM, & RS

Researchers

<https://www.computeontario.ca/fresearch-impacts>

SHARCNET is a consortium of **19 universities and colleges** and runs one of the largest high-performance clusters in Canada: **Graham**, located at the University of Waterloo.

Compute Ontario's Role
in Ontario's DRI Ecosystem

SciNet

Vector Institute (AI) | IC/ES Health Data Research | HPC4Health | Centre for Advanced Computing (CAC) | SHARCNET 19 colleges & universities | SCINET UofT

NIAGARA @ U Toronto

Compute Ontario

DRI: ARC, RDM, & RS

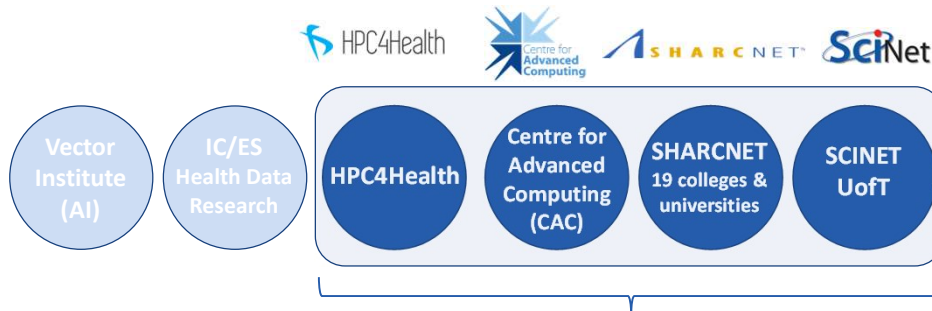
Researchers

https://static1.squarespace.com/static/5fdabf184103dc3d39174616/t/63e3c4e83bba8905c52caa90/1675871465242/CO-AnnualReport-2022_Diamond.pdf

<https://www.computeontario.ca/fresearch-impacts>

SciNet is a consortium-led supercomputer centre at the University of Toronto, hosting **Niagara**, one of the fastest supercomputers in Canada, plus a range of other systems. It provides Canadian researchers with computational resources and expertise necessary to perform their research on scales not previously possible in Canada on a wide range of fields - biomedical sciences, aerospace engineering, astrophysics and climate science.

Compute Ontario's Role in Ontario's DRI Ecosystem



Compute Ontario HPC Consortia



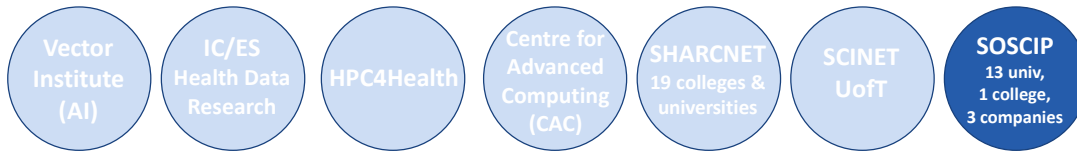
DRI: ARC, RDM, & RS

Researchers

<https://www.computeontario.ca/fresearch-impacts>

Together, **HPC4Health, CAC, SharcNet, and SciNet**, make up the **Compute Ontario HPC Consortia**, which we'll look at more closely later in this presentation.

Compute Ontario's Role
in Ontario's DRI Ecosystem



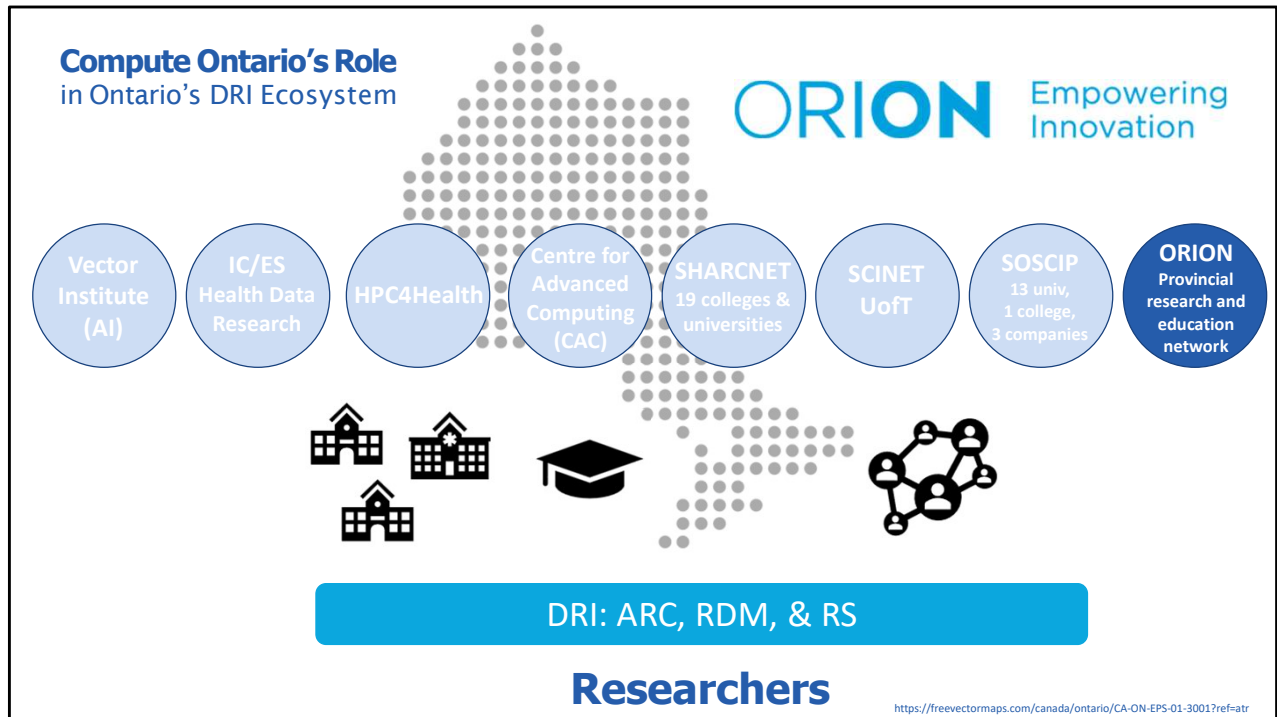
DRI: ARC, RDM, & RS

Researchers

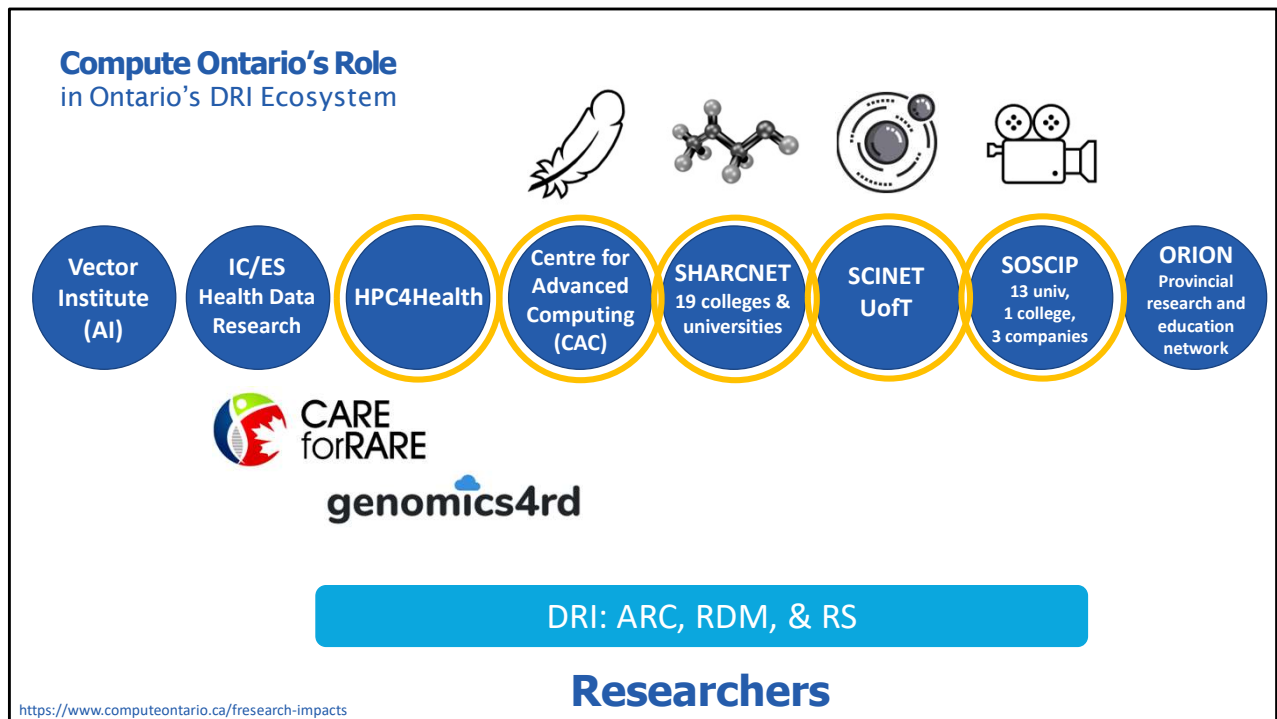
https://static1.squarespace.com/static/5fdabf184103dc3d39174616/t/63d91b92ce6221792759981b/1675172755128/CO-AnnualReport-2022_Movahedi.pdf

<https://www.computeontario.ca/fresearch-impacts>

SOSCIP is an academic-industry consortium founded in 2021 as a **collaboration between Ontario's research-intensive post-secondary institutions and small and medium-sized enterprises (SMEs) across the province.** SOSCIP is driving the uptake of AI and data science solutions and enabling the development of a knowledge-based and innovative economy in Ontario by supporting technical skill development and delivering high-quality outcomes.



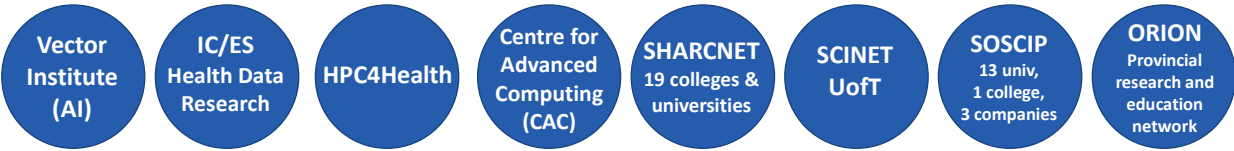
And finally, **ORION** is Ontario's provincial research and education network, spanning 6,000 kilometres and connecting over 1.7 million people in the research and education sector.



Taking a very brief look at impacts:

- **Health care initiatives** using **HPC4Health** include **Care4Rare Canada** a consortium of 20 Canadian clinical and research sites dedicated to helping children with rare genetic conditions and the related **[CLICK] Genomics4RD** a pan-Canadian rare disease data repository and analysis platform designed to centralize and provide harmonized access to data from thousands of “rare-disease” families to identify and connect rare diseases.
- **HPC resources at CAC** are helping to revitalize Indigenous languages
- **The University of Windsor is using SHARCNET resources to conduct Computational Chemistry research, to design unnatural molecules to cure disease.**
- **SCINET** resources support the **SUPER CRYOGENIC DARK MATTER SEARCH** or **SUPER CDMS**, a part of a collaborative international project to find the dark matter particle.
- **SOSCIP** is being used by Seneca College to develop a scalable machine learning algorithm to categorize videos

Compute Ontario's Role
in Ontario's DRI Ecosystem

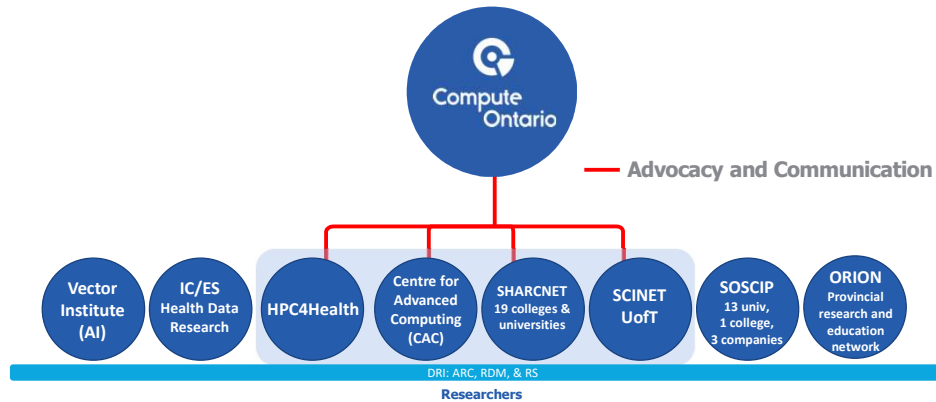


DRI: ARC, RDM, & RS

Researchers

Let's take a look at how Compute Ontario fits into this diverse DRI landscape.

Compute Ontario's Role in Ontario's DRI Ecosystem

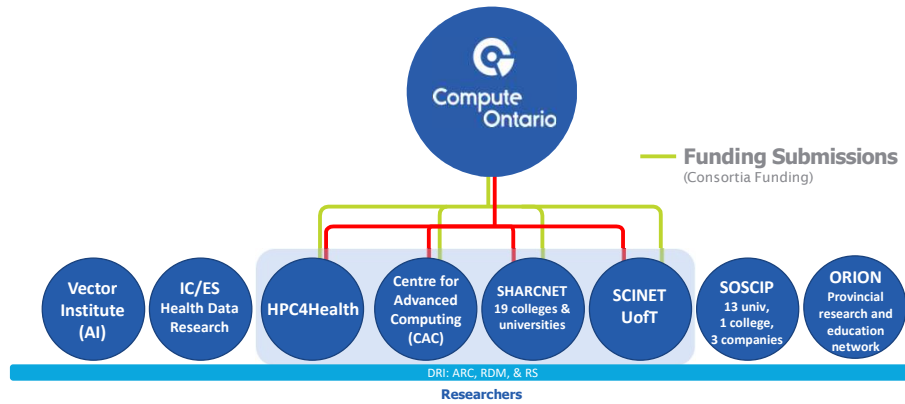


Updated February 2023

As already mentioned, Compute Ontario serves as a trusted source of unbiased strategic advice and leadership in the Ontario DRI space – which includes a lot of **ADVOCACY & COMMUNICATION** among Ontario's HPC consortium members.

Compute Ontario's Role in Ontario's DRI Ecosystem

— Advocacy and Communication

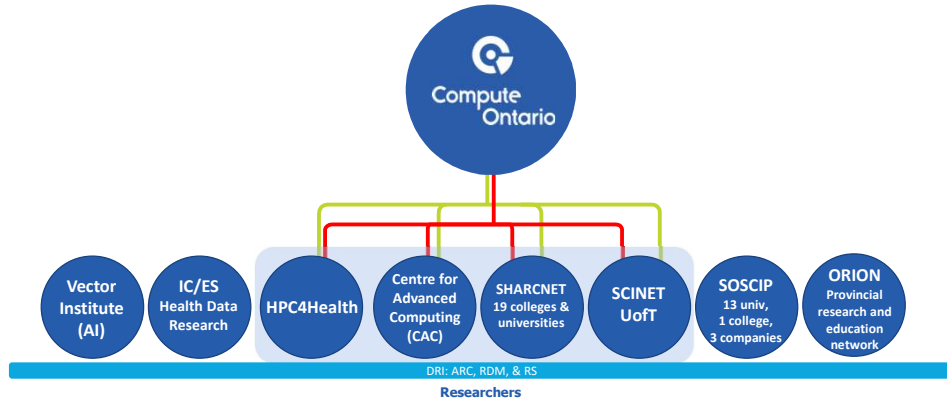


Updated February 2023

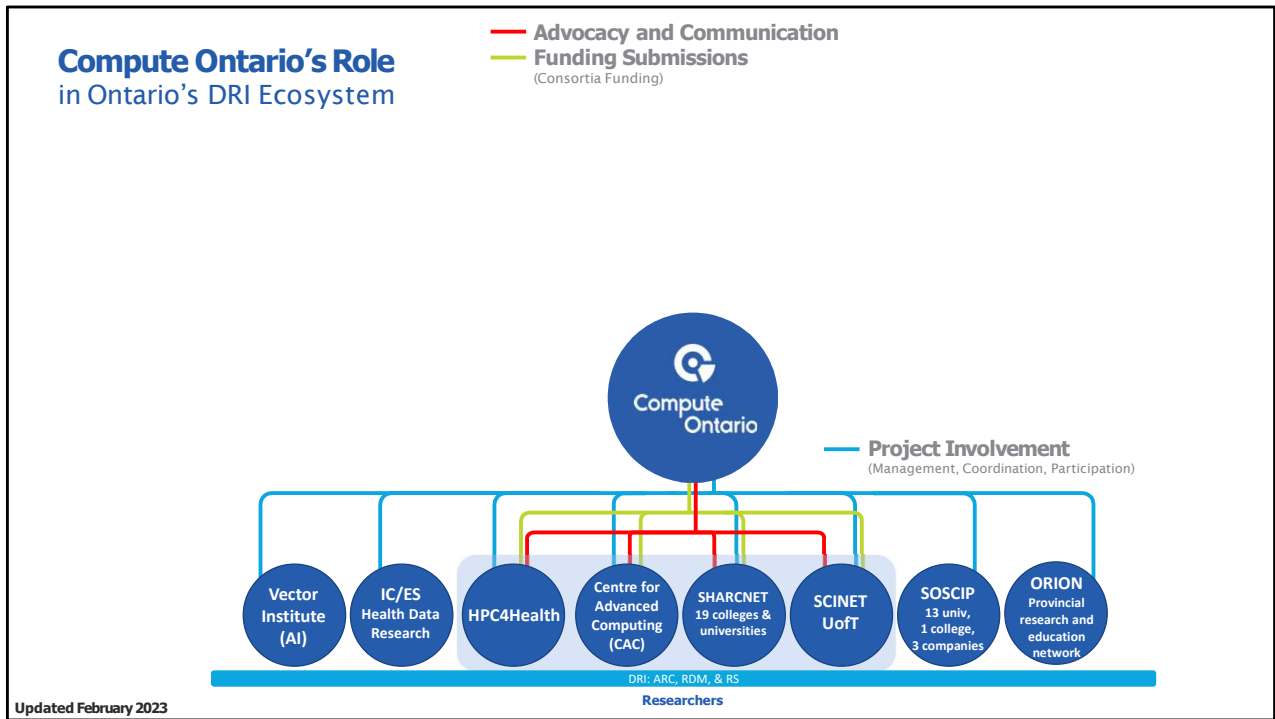
Compute Ontario also coordinates FUNDING SUBMISSIONS on behalf of the Consortia.

Compute Ontario's Role in Ontario's DRI Ecosystem

— Advocacy and Communication
— Funding Submissions
(Consortia Funding)



Updated February 2023



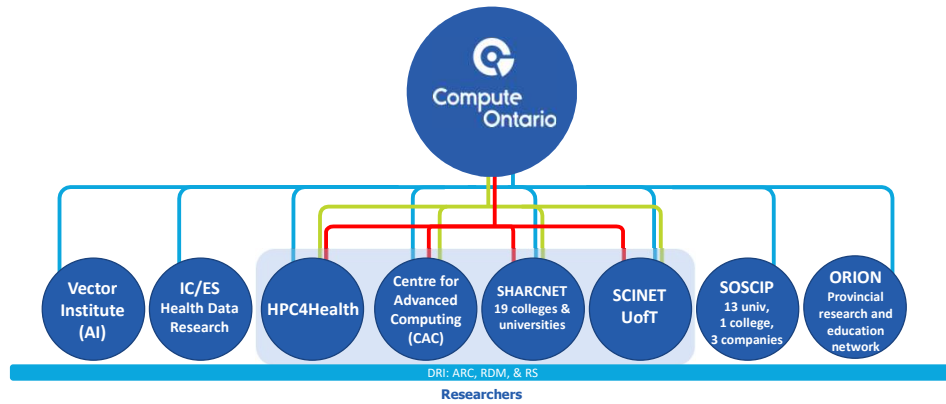
And has **Project Involvement with all of the partners & stakeholders.**

Compute Ontario:

- Coordinates Special Projects
- Brokers Knowledge and Expertise
- And promotes Best Practices in DRI

Compute Ontario's Role in Ontario's DRI Ecosystem

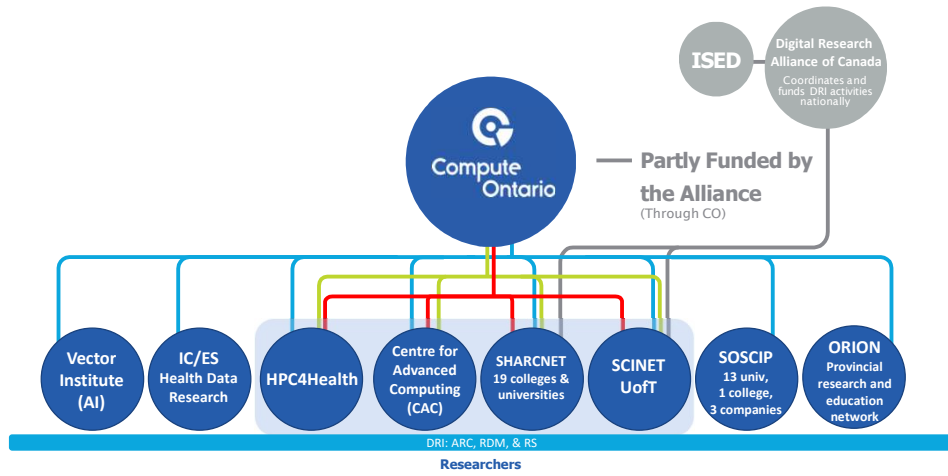
- **Advocacy and Communication**
- **Funding Submissions**
(Consortia Funding)
- **Project Involvement**
(Management, Coordination, Participation)



Updated February 2023

Compute Ontario's Role in Ontario's DRI Ecosystem

- **Advocacy and Communication**
- **Funding Submissions**
(Consortia Funding)
- **Project Involvement**
(Management, Coordination, Participation)

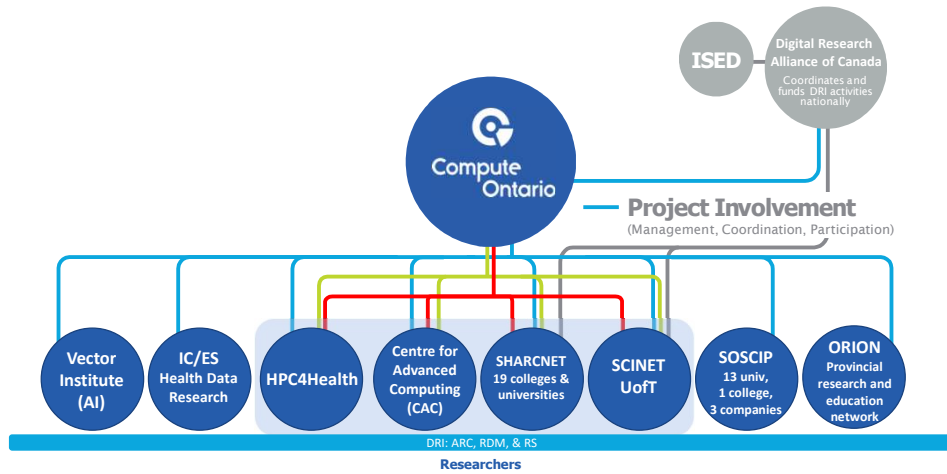


Updated February 2023

Alliance funding for two Compute Ontario's HPC consortia members, SHARCNET & SCINET, is brokered through Compute Ontario.

Compute Ontario's Role in Ontario's DRI Ecosystem

- **Advocacy and Communication**
- **Funding Submissions**
(Consortia Funding)
- **Project Involvement**
(Management, Coordination, Participation)
- **Partly Funded by the Alliance**
(Through CO)

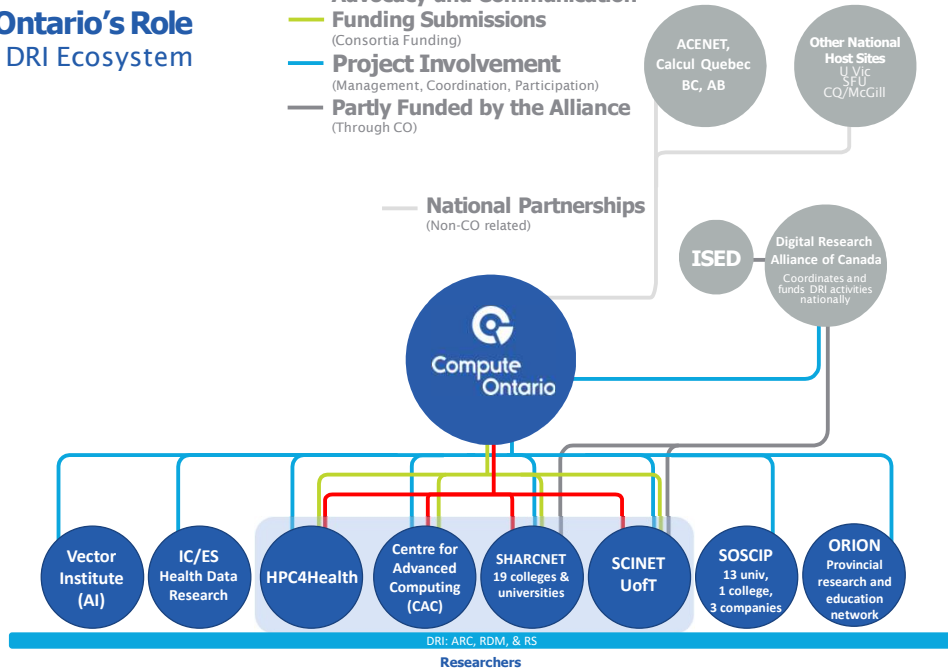


Updated February 2023

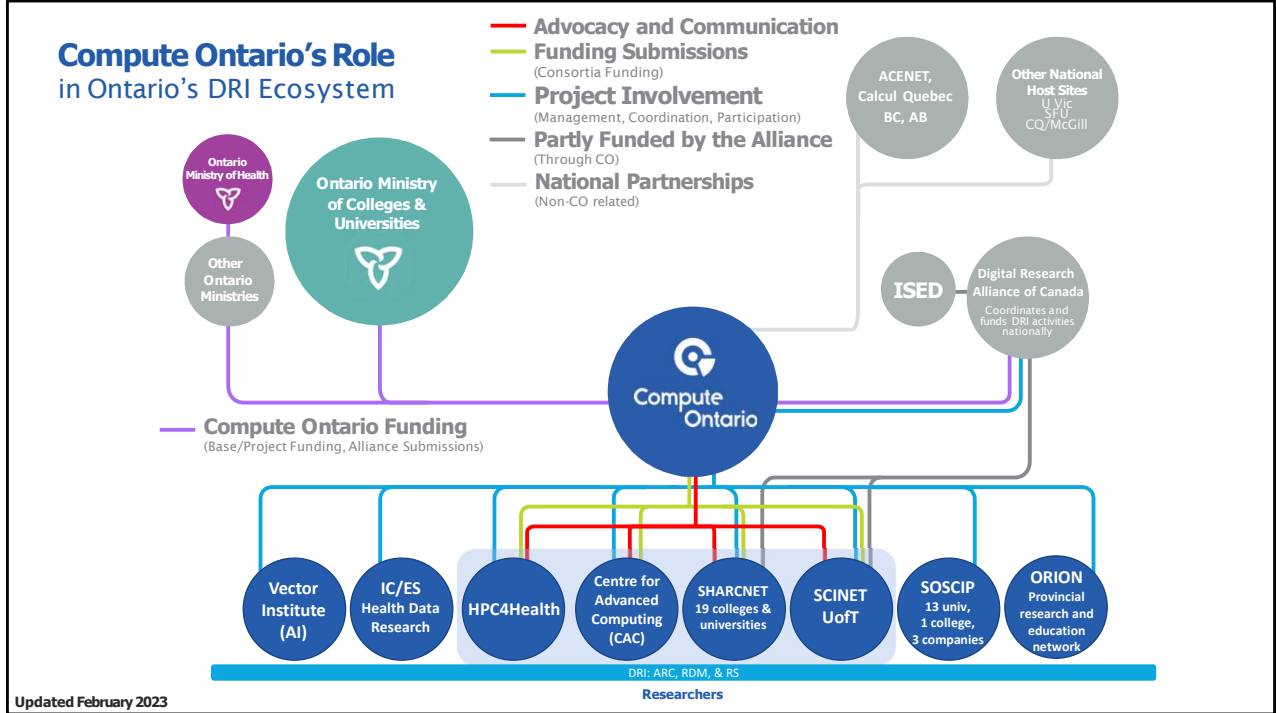
Compute Ontario's Role in Ontario's DRI Ecosystem

- **Advocacy and Communication**
- **Funding Submissions**
(Consortia Funding)
- **Project Involvement**
(Management, Coordination, Participation)
- **Partly Funded by the Alliance**
(Through CO)

— **National Partnerships**
(Non-CO related)

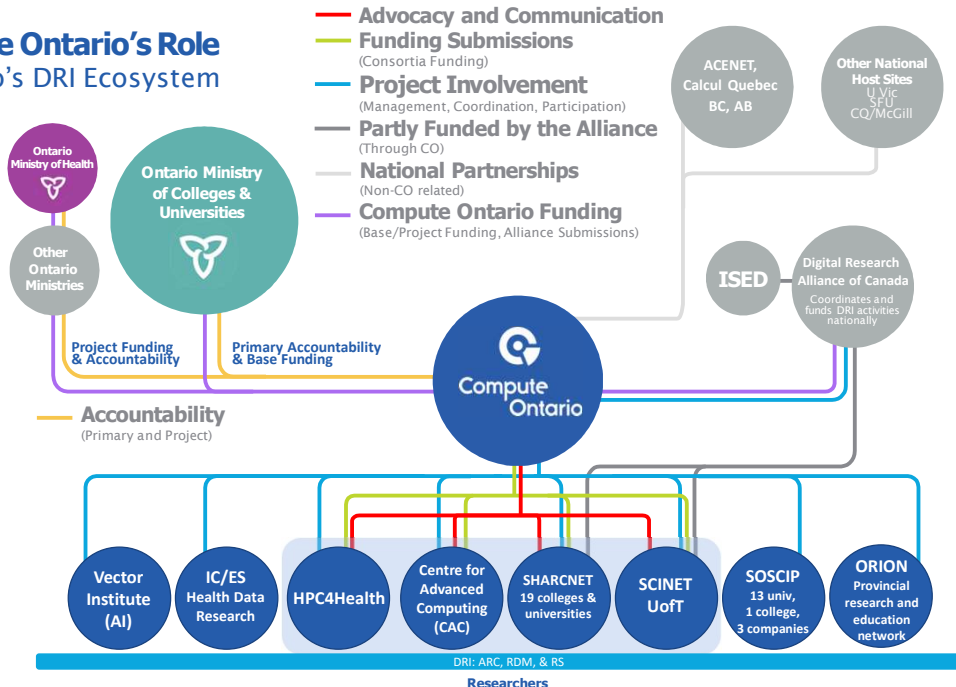


Updated February 2023



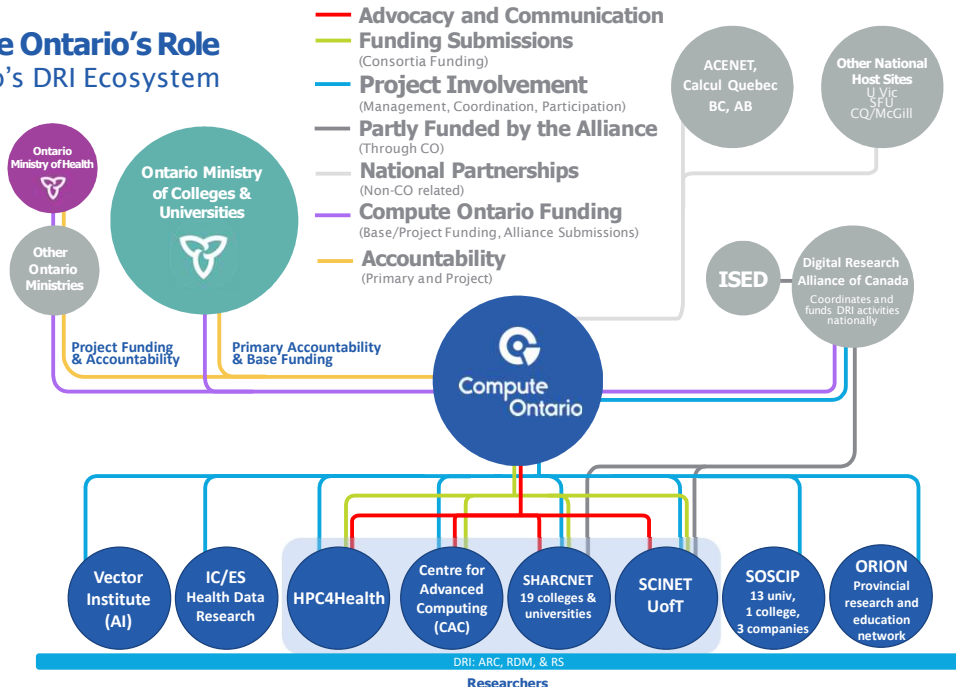
Compute Ontario is funded by the Ontario Ministry of Colleges and Universities, the Ministry of Health, and other Ministries.

Compute Ontario's Role in Ontario's DRI Ecosystem



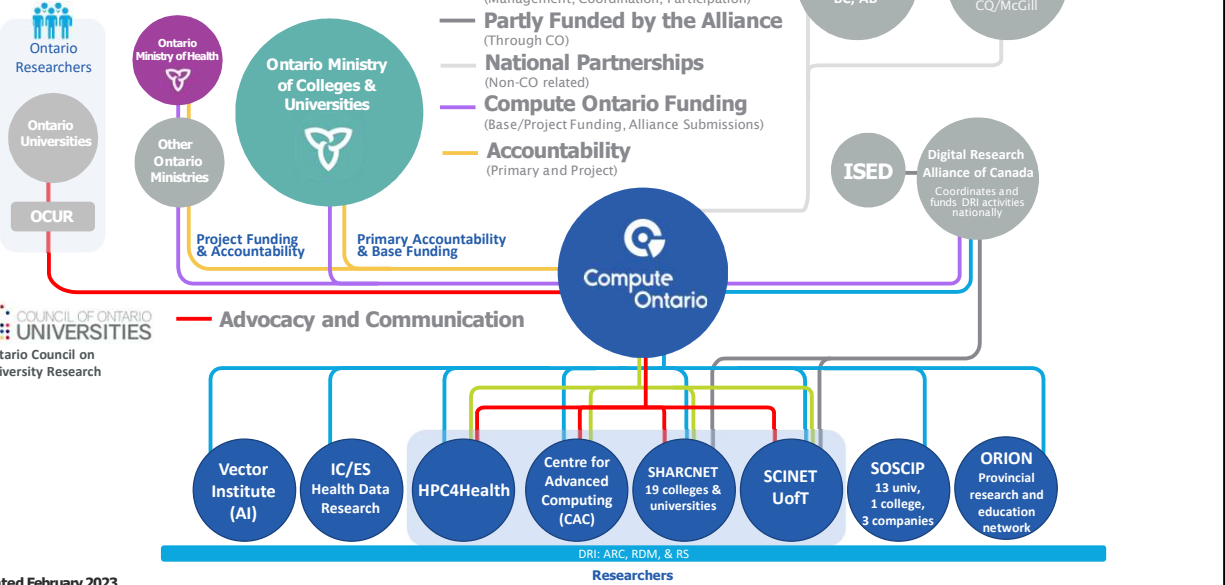
Updated February 2023

Compute Ontario's Role in Ontario's DRI Ecosystem



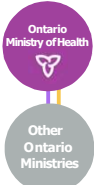
Updated February 2023

Compute Ontario's Role in Ontario's DRI Ecosystem

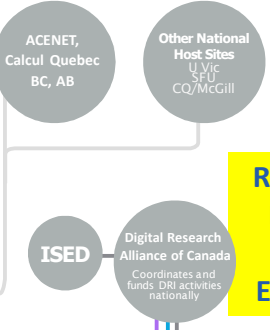


Updated February 2023

Compute Ontario's Role in Ontario's DRI Ecosystem



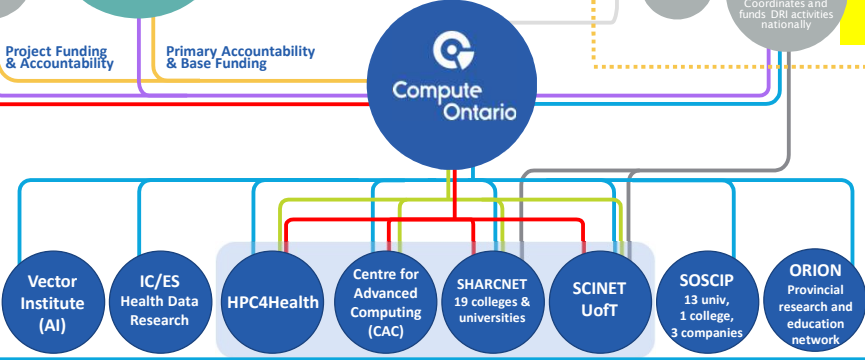
- **Advocacy and Communication**
- **Funding Submissions**
(Consortia Funding)
- **Project Involvement**
(Management, Coordination, Participation)
- **Partly Funded by the Alliance**
(Through CO)
- **National Partnerships**
(Non-CO related)
- **Compute Ontario Funding**
(Base/Project Funding, Alliance Submissions)
- **Accountability**
(Primary and Project)



Regional & National DRI Ecosystem

COUNCIL OF ONTARIO UNIVERSITIES
Ontario Council on University Research

Ontario DRI Ecosystem



DRI: ARC, RDM, & RS
Researchers

Updated April 2023

Director, Data Strategy & Services

Provides leadership for all activities related to Compute Ontario's data strategy portfolio, including, but not limited to RDM activities

Priority #1

Leadership & Vision: Work and plan strategically with other RDM and RS stakeholders to maintain, expand, and advance data management capacity in Ontario and across the country

Priority #3

Training & Services: Coordinate and facilitate access to data training, services, and infrastructure for researchers in Ontario

Priority #2

Relationship-building: Build and maintain strong, sustainable, trust-based partnerships with federal, national, regional, and local bodies, including Universities and Colleges

Priority #4

Representation: Serve as Compute Ontario's '*RDM voice*' on national and regional working groups, at technical conferences, events, trade shows, and other meetings, and on various committees, fostering a spirit of collaboration across the sector.



My role at Compute Ontario was created to address FOUR MAIN PRIORITIES:

- Leadership & Vision
- Relationship Building
- Training & Services, and
- Representation

Support

2022 RDM Funding Call

University of Guelph – RDM Workshop Series

- Metadata creation
- Growing a FAIR data culture;
- Basic R programming skills for data tidying;
- R Shiny dashboard for data discovery.

University of Waterloo

- Curating FAIR synthetic health data while preserving privacy;
- Managing synthetic data and metadata
- Policy changes needed for governance of synthetic data.

University of Toronto

Project to integrate ODESI statistical data into the national Borealis research data repository.

University of Toronto

Augmenting resources, support & training for RDM with a focus on respecting Indigenous data sovereignty.

Conestoga College, Confederation College, Fleming College, Fanshawe College, Sheridan College

Workshops on RDM development and capacity building for Colleges.


University of Windsor – JupyterHub Workshop Series

Workshops & codefest on using Jupyter notebooks in support of RDM.

University of Waterloo

Interactive data visualization for discoverability and reusability of polar metadata and data at the Polar Data Catalogue.

One of the new initiatives related to the addition of RDM to the Compute Ontario mandate was the creation of RDM-focused funding calls. Successful applicants to the second of these calls, in 2022, are shown here. **[CLICK]** It is interesting, but perhaps not surprising, to note that 4 out of 7 of these are directly related to TRAINING INITIATIVES.



Training

RDM & ARC: Building Bridges to Support Researchers


Creating a Data Management Plan (DMP)

So you want to publish your data, now what? BOREALIS (Dataverse)!




Alicia Cappello

Felicity Tayler & Fatou Bah



Continuing with the Training theme, the long-standing Compute Ontario Summer School added THREE RDM-FOCUSED sessions in 2022,



**2023 CO
Summer School
June 12 - 29**

RDM Sessions on June 12 - 13

Research Data Management and HPC: Moving toward shared best practices
Introduction to Borealis, Odesi, and Geoportal
Safe Sharing of Data

Introduction to Alliance RDM Services: DMP Assistant, Federated Research Data Repository, & Lunaris
Research data about people - new policy and new opportunities

Research Data Management: Good Practices to Support Long-term Data Access



The 2023 school will have two parallel streams covering a wide range of HPC, Research Software, and RDM – including, I am happy to report, the **SIX RDM sessions shown here**. **I would encourage you all to check out the rich array of non-RDM course offerings, which are open to all and brought to you by dedicated specialists at CAC, SCINET, SHARCNET.** The Summer School runs from Monday June 12, to Wednesday, June 28th, but we've scheduled the RDM sessions right at the beginning, so mark your calendars for June 12 & 13.



COSS 2023 RDM Speakers	
Amber Leahey	Scholars Portal
Meghan Goodchild	Queen's University/Scholars Portal
Kristi Thompson	Western University
Victoria Smith	Alliance
Mark Goodwin	Alliance
Erin Clary	Alliance
Shiloh Williams	Alliance
Lina Harper	Alliance
Tamanna Moharana	Alliance
Jeff Moon	Compute Ontario

RDM Sessions on June 12 - 13

Research Data Management and HPC: Moving toward shared best practices
Introduction to Borealis, Odesi, and Geoportal
Safe Sharing of Data



Introduction to Alliance RDM Services: DMP Assistant, Federated Research Data Repository, & Lunaris
Research data about people - new policy and new opportunities

Thanks go to the many RDM specialists **from Ontario and across Canada** who have agreed to teach at the Summer School this year!

Outreach

Training Advisory Committee

Elise Degen, CAC - **Chair**

Michelle Edwards, UofGuelph – **co-Chair**

Sergey Mashchenko, SHARCNET

Andrei Turinsky, Sick Kids

Ramses van Zon, SciNet

Christina DeRoche, Canadore College

Berenica Vejvoda, University of Windsor

Danica Evering, McMaster University

Jane Fry, Carleton University

Amal Khalil, Queen's University/CAC



Turning to OUTREACH, Compute Ontario has struck two new Advisory Committees. The TRAINING ADVISORY COMMITTEE is dedicated to supporting and coordinating training efforts across the DRI ECOSYSTEM.

Outreach

Training Advisory Committee

RDM Advisory Committee

Rose-Marie Dolinar, Humber College

Jean-Pierre Hickey, U. Waterloo

Ping Liang, Brock U.

Nasser Abukhdeir, U. Waterloo

Oliver Goodison-Powell, Conestoga College

Dylan Dearborn, U. of Toronto

Alexandra Cooper, Queen's U.

Meghan Goodchild, Scholars Portal & Queen's U.

Isaac Pratt, McMaster U.

Kate Davis, Scholars Portal

Brendan Behan, Ontario Brain Institute

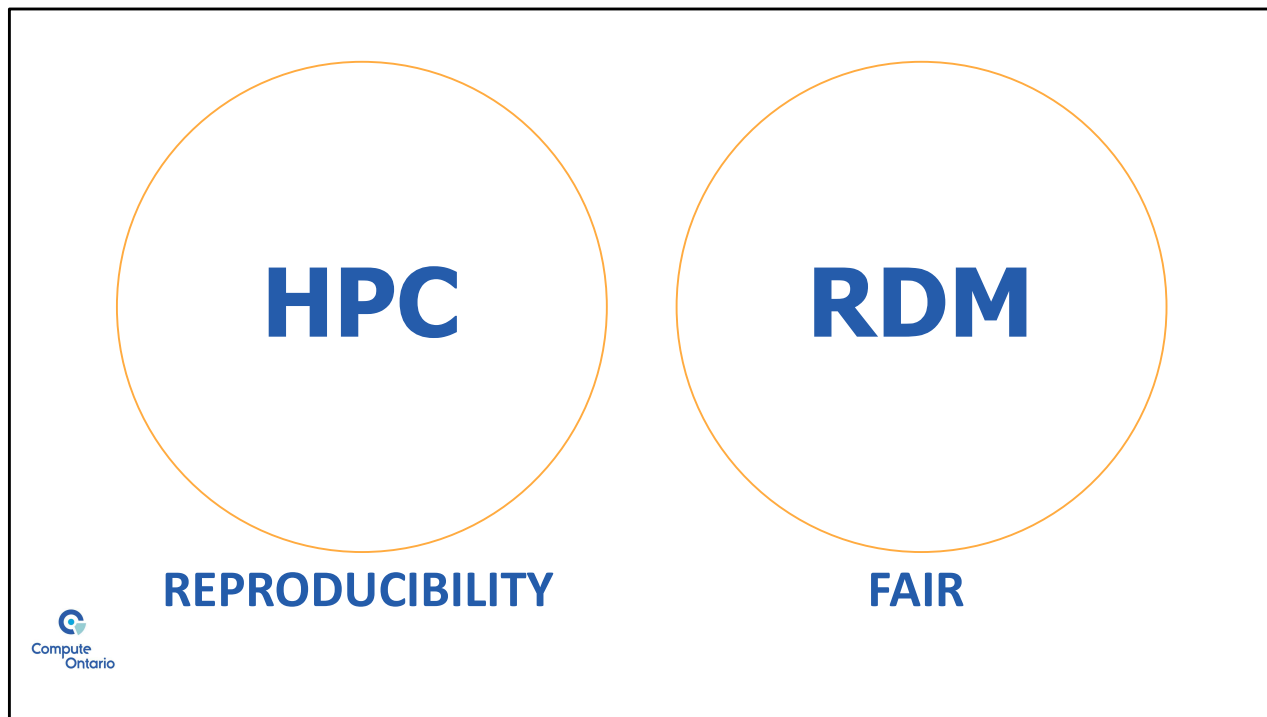


The RDM Advisory Committee will guide Compute Ontario **more specifically** regarding support and coordination of RDM efforts for Ontario Researchers.

Vision

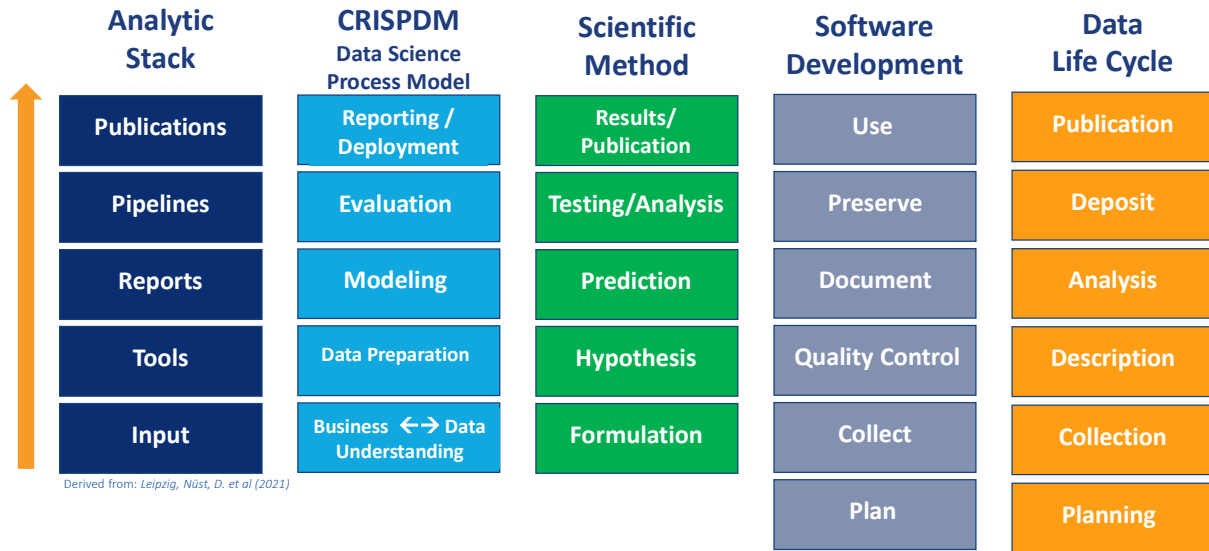


I started with Compute Ontario with the goal of learning about HPC and how it intersects with RDM, and reflecting this back to the community in practical ways. I think I have made some progress in this regard, thanks in no small part to all the patient people who have helped me in my journey thus far.



My understanding is still growing and evolving. One theme I've noticed that might help us move things forward is the **strong HPC focus on REPRODUCIBILITY**. [\[CLICK\]](#). **And while reproducibility is certainly** important, I believe RDM best practices take a broader view, seeking to make [\[CLICK\]](#) **data, metadata, and code FAIR [Findable, Accessible, Interoperable, and Reusable]** → which if done well, supports the goal of reproducibility.

Research Workflows



Derived from: Leipzig, Nüst, D. et al (2021)

With modifications, from Leipzig, Nüst, D., Hoyt, C. T., Ram, K., & Greenberg, J. (2021). The role of metadata in reproducible computational research. *Patterns* (New York, N.Y.), 2(9), 100322–100322. <https://doi.org/10.1016/j.patter.2021.100322>

41

And certainly, there are a range of **RESEARCH WORKFLOWS** beyond HPC that would benefit from RDM Best Practices...



Research Workflows

	Analytic Stack	CRISPDPM Data Science Process Model	Scientific Method	Software Development	Data Life Cycle	Text & Data Mining ¹	Other?
↑	Publications	Reporting / Deployment	Results/ Publication	Use	Publication	Publish/ Reproduce	?
	Pipelines	Evaluation	Testing/Analysis	Preserve	Deposit	Validate	?
	Reports	Modeling	Prediction	Document	Analysis	Measure	?
	Tools	Data Preparation	Hypothesis	Quality Control	Description	Data	?
	Input	Business ↔ Data Understanding	Formulation	Collect	Collection	Construct	?
			Plan	Planning	Theory	?	

Derived from: Leipzig, Nüst, D. et al (2021)

With modifications, from Leipzig, Nüst, D., Hoyt, C. T., Ram, K., & Greenberg, J. (2021). The role of metadata in reproducible computational research. *Patterns* (New York, N.Y.), 2(9), 100322–100322. <https://doi.org/10.1016/j.patter.2021.100322>

1. Prof. Andrew Piper, McGill U. Presentation at Text and Data Mining Colloquium, University of Toronto, May 2-3, 2023. txtlab.org

42

And I attended a TEXT AND DATA MINING Colloquium that added yet another workflow... and I'm sure there are other discipline-specific workflows that could be added here as well.

Analytic Stack	CRISFDM Data Science Process Model	Scientific Method	Software Development	Data Life Cycle	Text & Data Mining ¹	Other?
Publications	Reporting / Deployment	Results/ Publication	Use	Publication	Publish/ Reproduce	?
Pipelines	Evaluation	Interpretation	Provision	Deposit	Validate	?
Reports	Modeling	Prediction	Document	Analysis	Measure	?
Tools	Data Preparation	Hypothesis	Quality Control	Integration	Data	?
Input	Provenance Interoperability	Formulation	Collect	Collection	Construct	?
			Plan	Planning	Theory	?

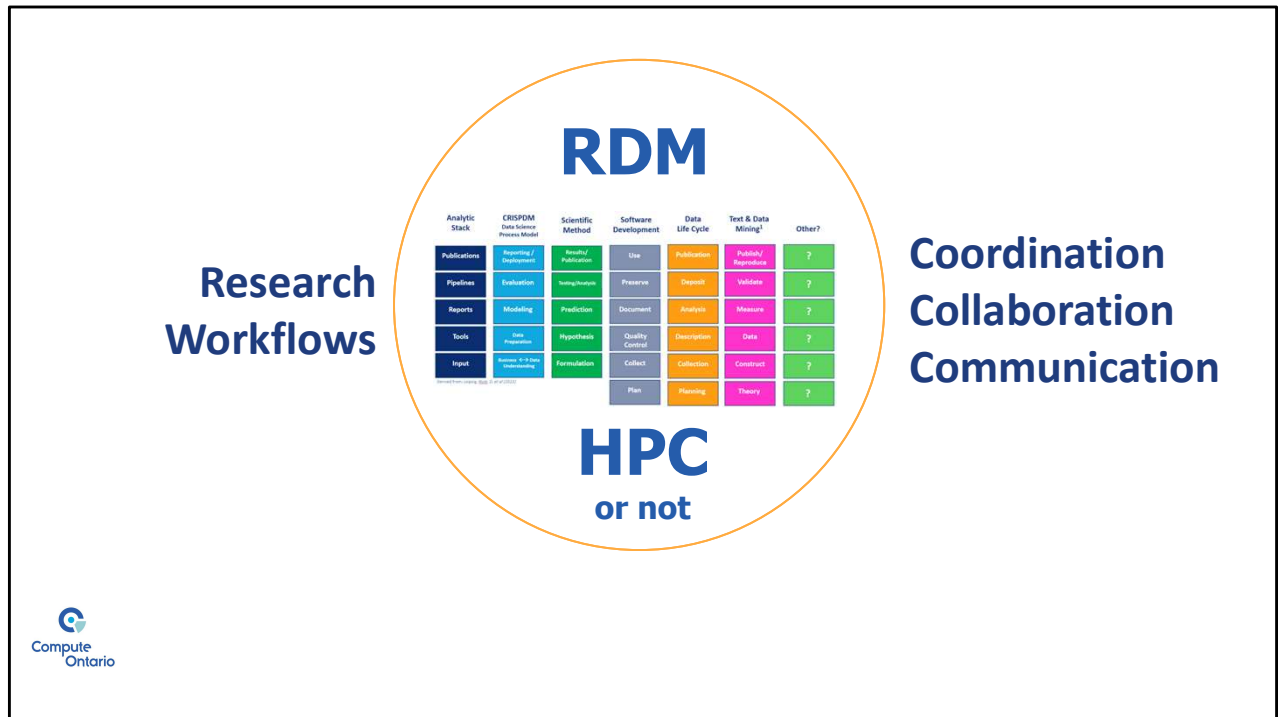
HPC
or not

RDM

Research Workflows

Compute Ontario

The point here is that **regardless of the Workflow, or whether or not High Performance Computing is needed to complete the associated research, good RDM practices are key.**



Our goal is to help articulate and socialize **RDM-friendly approaches** to conducting research, whether **HPC is involved or not**, and **Compute Ontario's role is to COORDINATE, COLLABORATE, AND COMMUNICATE** with partners & stakeholders to achieve this goal.

Questions?