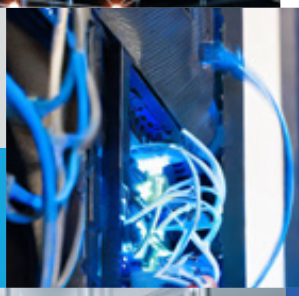
Advancing Research and Innovation

IN ONTARIO



2023 ANNUAL REPORT

Compute Ontario

Our Vision

Realize the full potential of digital research infrastructure in advancing research, innovation and creation in Ontario to support economic prosperity and improve the lives of Ontarians

Our Mission

Ensure that researchers in Ontario have access to the necessary digital research infrastructure equipment, services, skills and support to advance research, discovery and creation

Our Values

Trust, transparency and accountability

Collaboration, inclusivity and partnership

Innovation, agility and adaptability

Continuous improvement

Contents

Photo courtesy Seneca

Chair & CEO Message

Ontario's shared digital research infrastructure (DRI) is a sophisticated ecosystem of high-performance hardware, software and expertise. It exists to enable Ontario researchers to conduct cutting-edge research that leverages advanced computer simulation, modelling, data visualization, artificial intelligence and machine learning, and to support the training of the next generation of Ontario's skilled workers. The DRI ecosystem has revolutionized research, making it possible to advance drug design, genetics, materials design and manufacturing, air quality tracking and modelling, transportation, life sciences and health care at speeds we've never seen before. Thousands of Ontario researchers — working in hospitals, academic institutions and with industry partners — rely on this shared ecosystem, which has been jointly funded by the provincial and federal governments for more than 20 years. The ecosystem provides more efficient, cost-effective support than commercially available alternatives, and plays a key role in driving innovation and economic growth.

RANIL SONNADARA
President, CRO and CEO

WARREN HELLAND
Board Chair

In the past year alone, Ontario's DRI has supported more than 6,000 users, provided skill development to nearly 4,000 trainees and supported more than 800 collaborations with industry. Ontario's DRI ecosystem has also enabled the launch of 253 new companies and supported 338 patent applications over the last five years. With the emergence of machine learning technologies, Ontario's DRI ecosystem is vital to Ontario's vision of being a global hub of technology development, commercialization, and biomanufacturing.

This year, our ecosystem reached a critical juncture. The two largest shared computational systems have surpassed their expected productive lifetime, meaning the hardware is far less capable and less energy-efficient than newer systems. More importantly, despite the herculean efforts of staff, key elements are starting to fail, and both systems have increased vulnerability to cybersecurity attacks and ever-increasing maintenance costs. Investment decisions made by the province in the coming weeks will significantly impact whether Ontario's research and innovation community continues to have access to the computational systems and expertise necessary for Ontario to remain at the leading edge of innovation. Federal funding is available to replace these systems and support continued operations, but that funding is contingent upon match funding being provided by the province.

Compute Ontario's mission is to ensure that Ontario researchers continue to have access to this essential digital infrastructure. Our priority has been advocating for the necessary provincial matching funding to support the people and systems on which Ontario's researchers rely, and for a new, sustainable approach to funding for Ontario's DRI ecosystem in the longer term. A comprehensive funding proposal on behalf of the ecosystem was developed with valuable input from stakeholders, our Technology Leadership Advisory Council, and the Ministry of Colleges and Universities. It is currently under review. We have mobilized the ecosystem community — our Board members, host sites and consortia, research institutes, academic institutions, industry partners, the Ontario Council of University Researchers, the Health Research and Innovation Committee of the Ontario Hospital Association, and the Heads of Applied Research of Ontario colleges and many other allies. This network of passionate and informed individuals has collectively pursued thousands of conversations with decision-makers and influencers, to make the case for funding. However, as we write this message, the Ontario government has not yet committed to providing the required matching funds, and the future of Ontario's shared DRI platform is uncertain.

In the pages of this Annual Report, we share highlights of vital work that Ontario's DRI platform has enabled in the last year. This includes several examples of DRI-enabled research that are driving scientific breakthroughs, real-world innovation and industry collaboration in genetics, drug design and environmental sustainability. This is the kind of work that would become, if not impossible, more difficult and much more expensive in the absence of a shared DRI platform in Ontario.

We continue to advocate intensively for these resources in the coming year on behalf of Ontario's research community, and strongly encourage you to make sure that everybody is aware of their importance if you would like them to remain available.

WARREN HELLAND
Board Chair

RANIL SONNADARA
President, CRO and CEO

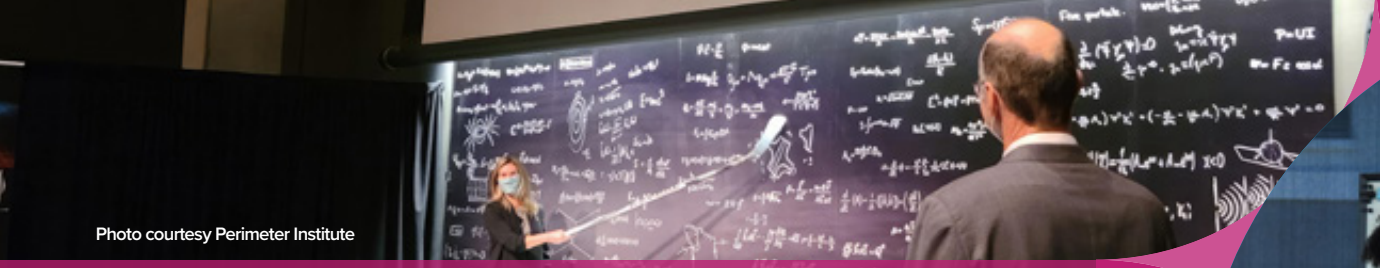


Photo courtesy Perimeter Institute

Strategic Leadership

Working toward a robust, sustainable DRI ecosystem to meet future research needs

Strategic leadership is a key component of Compute Ontario's mandate. In 2022 – 23, similar to last year, much of our effort was focused on advocating for a sustainable approach to funding and governance at the federal and provincial levels. Ontario's 22 years of funding for shared DRI has been instrumental in creating a strong DRI expert and user community. As such, Ontario has been well positioned to exploit breakthroughs in the application of AI and advanced analytics, and to address big data challenges in healthcare, genomics, pharmaceuticals and climate change. But the unpredictability of funding from year to year has held Ontario back from fully realizing this potential. Further, the most recent provincial budget contained no provision for DRI funding — at a time when Ontario's shared DRI platform is aging and in decline. Working with representatives from the Ontario Council of University Researchers, College Heads of Academic Research and the Ontario Hospital Association Health Research and Innovation Committee, Compute Ontario prepared and submitted a detailed funding proposal to the province, while advocating intensively for provincial DRI funding throughout 2022 – 23. As we publish this report, advocacy efforts continue as funding issues remain unresolved.

Planning is imperative not only with regard to funding but also from a technological standpoint. Ontario needs a shared DRI platform that can meet new and growing needs related to AI, quantum computing, cybersecurity and the expansion of advanced computing to nearly all sectors. What is the most effective and cost-efficient technological approach to achieve this goal? Compute Ontario's Technology Leadership Advisory Council (TLAC) is working to answer that question. As a priority this year, TLAC developed and launched a cloud pilot project to identify ideal ways to make use of commercial cloud resources within the overall portfolio of DRI. An open call resulted in five approved projects, which will receive free credits for cloud services from two major commercial cloud providers. The projects will run until March 31, 2024 and outcomes from these projects will inform the province's DRI plan.

"Collaboration and activation of talent and resources across industry and academic research institutions... is necessary to accelerate science and discovery, drive innovation and economic development... good R&D strategy is good fiscal policy."

— IBM Canada

Technical Leadership Advisory Council (TLAC)

Danny Gruner, SciNet

Francis Jeanson, OBI

John Morton, SHARCNET

Lincoln Stein, OICR

Amaz Taufique, Scholars Portal

In the spring of 2023, the Ontario Ministry of Health (MOH) invited Compute Ontario and six other organizations to form a Partners' Table tasked with developing a vision for data-driven health research and the next phase of evolution of the Ontario Health Data Platform (OHDP). This builds on the research use case completed last year by Compute Ontario, which made a series of recommendations regarding health data.

In 2022 – 23, Compute Ontario also continued its focused effort to encourage the integration of good data management into research processes. Making research data FAIR (Findable, Accessible, Interoperable, and Reusable) — an internationally recognized goal of RDM — benefits both researchers and society in myriad ways, including ensuring scientific integrity, fulfilling funder requirements, broadening impact of data beyond the original research team, and avoiding duplication of effort and investments. Training is key in this regard and there is great interest in training opportunities for both researchers and data professionals alike. This year's Compute Ontario Summer School included six RDM modules, as compared to one last year. Compute Ontario also participated in several conferences, panel discussions and working groups. These were valuable opportunities to initiate collaborative conversations about key RDM challenges: handling sensitive/proprietary data, managing access permissions, generating standards-compliant metadata, respecting Indigenous data sovereignty, and creating Data Management Plans in diverse disciplines. To help address these challenges, Compute Ontario conducted a third funding call for the development of DRI Training materials and formed an RDM Advisory Committee (RDMAC).

RDMAC membership

Nasser Abukhdeir, University of Waterloo

Brendan Behan, Ontario Brain Institute

Alexandra Cooper, Queen's University

Kate Davis, University of Toronto and Ontario Council of University Libraries Scholars Portal

Dylanne Dearborn, University of Toronto

Rose-Marie Dolinar, Humber College

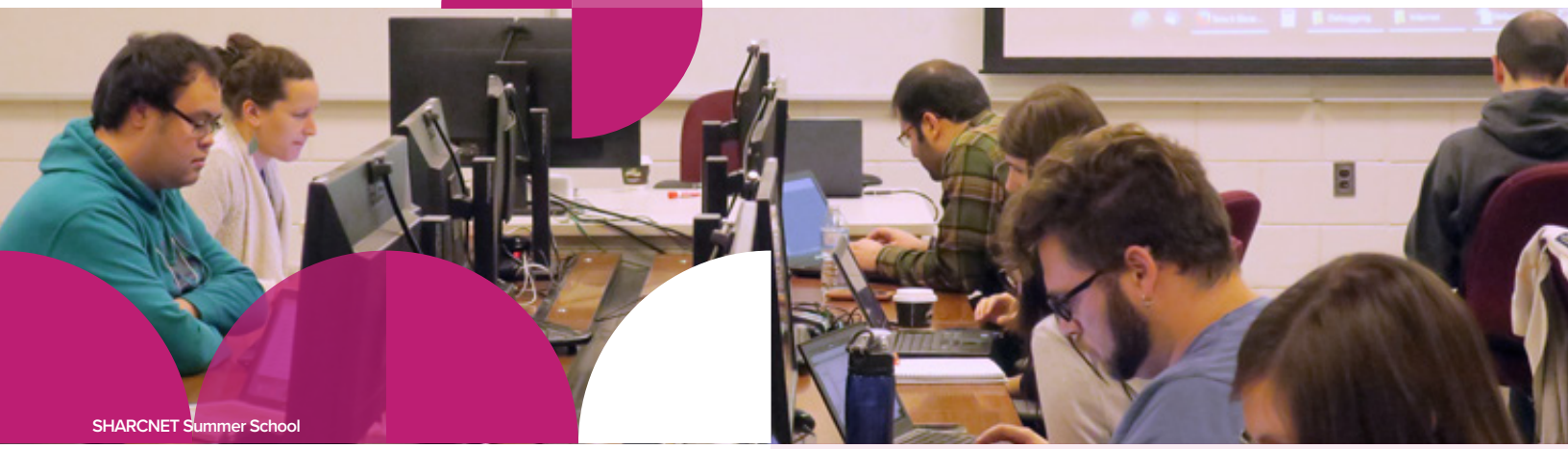
Oliver Goodison-Powell, Conestoga College

Jean-Pierre Hickey, University of Waterloo

Ping Liang, Brock University

Isaac Pratt, McMaster University

Amal Khalil, Queen's University/CAC



SHARCNET Summer School

Spotlight on Training:
Helping to build a high-tech talent pipeline

Ontario's DRI services are staffed by experts who, in addition to operating and maintaining high-performance computer systems, provide approximately 30,000 hours of training every year through the Compute Ontario Summer School and weekly colloquia series. This training is essential to academic researchers — whether undergraduate, graduate, post-doctoral, or faculty — who wish to use Advanced Research Computing (ARC) techniques including modelling, simulation, visualization, machine learning and artificial intelligence. Researchers who combine deep discipline-specific knowledge with advanced computing skills are in high demand in both academia and industry. Many combine their academic work with entrepreneurship to start innovative businesses. A key aspect of Compute Ontario's mandate is to help coordinate and increase accessibility of DRI training in Ontario. In 2022 – 23, Compute Ontario achieved several significant milestones in its efforts to advance training and build a collaborative foundation for developing a provincial DRI training strategy.

Training Advisory Committee

With diverse ARC, RS and RDM representation from across the ecosystem, the Training Advisory Committee (TAC) was created in December 2022 to bring greater coordination and collaboration to existing training initiatives, find opportunities to align RDM, RS and ARC training, and create a longer-term vision for DRI training in Ontario. As a first step, TAC is conducting selective inventory of existing training opportunities and resources to help identify strengths and gaps in the ecosystem.

TAC membership

Sergey Mashchenko, SHARCNET

Andrei Turinsky, Sick Kids

Ramses van Zon, SciNet

Elise Degen, CAC

Christina DeRoche, Canadore College

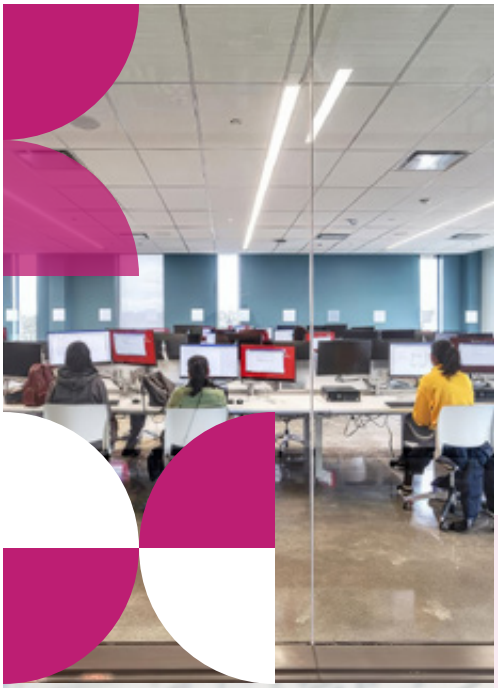
Michelle Edwards, University of Guelph

Meghan Goodchild, Queen's University

Berenica Vejvoda, University of Windsor

Danica Evering, McMaster University

Jane Fry, Carleton University



Partnership with eCampus Ontario

In August, Compute Ontario signed an MOU with eCampus Ontario, a nonprofit organization that aims to strengthen Ontario's post secondary education system by increasing access to online learning. Through this partnership, Compute Ontario intends to embed RDM elements into existing training resources, create additional RDM training resources, and make DRI training more widely available.

DRI Training Materials Funding

In September, Compute Ontario put out a funding call to support development of asynchronous DRI training materials for eventual sharing through the e-Campus Ontario platform. Compute Ontario also provided a $10,000 grant to support translation of a new Open Education Resource textbook, Research Data Management in the Canadian Context: A Guide for Practitioners and Learners.

Compute Ontario Summer School

This year's Compute Ontario Summer School ran from June 12 to 29, 2023 and offered an expanded program covering skills and technologies required for compute- and data-intensive research. In total, staff from SHARCNET, CAC, SciNet, Compute Ontario and others delivered 30 courses and 9,333 student-hours of training, for nearly 500 participants. The concentrated format of six hours per day, five days a week, over 3 weeks allowed for immersive learning at beginner, intermediate and advanced levels. Topics included high-performance computing tools, programming languages, machine learning, visualization, programming GPUs, bioinformatics, research data management, and more.

Joint Colloquia Series

In January, Compute Ontario supported SHARCNET, CAC and SciNet as they launched a new, jointly offered weekly colloquium series covering DRI-related topics targeted to a broad audience.

Innovation & Impact

The value of Ontario's digital research infrastructure

Digital research infrastructure (DRI) = Supercomputers that power artificial intelligence, machine learning and advanced data analytics + Highly skilled staff who support and train researchers to use advanced computing techniques + Data management practices that ensure privacy, security and discoverability

Compute Ontario

Ontario's Shared DRI Platform by the Numbers 2022 – 2023

Ontario-hosted supercomputing clusters that are shared nationally:

2

Ontario DRI experts who support the systems and their users:

70

Ontario-based researchers using shared DRI systems:

6,000+

Total usage of shared systems by Ontario-based researchers:

122,000 CPU Core-years, 900 GPUs

Researchers' cost to use the shared DRI platform:

$0

if similar services purchased from commercial providers:

$50M

Peer-reviewed publications of Ontario DRI-enabled research:

1100+

Students and postdoctoral fellows trained in high-tech skills by Ontario DRI experts:

approx. 4,000

*Ontario DRI-enabled innovation:

262
tech start-ups

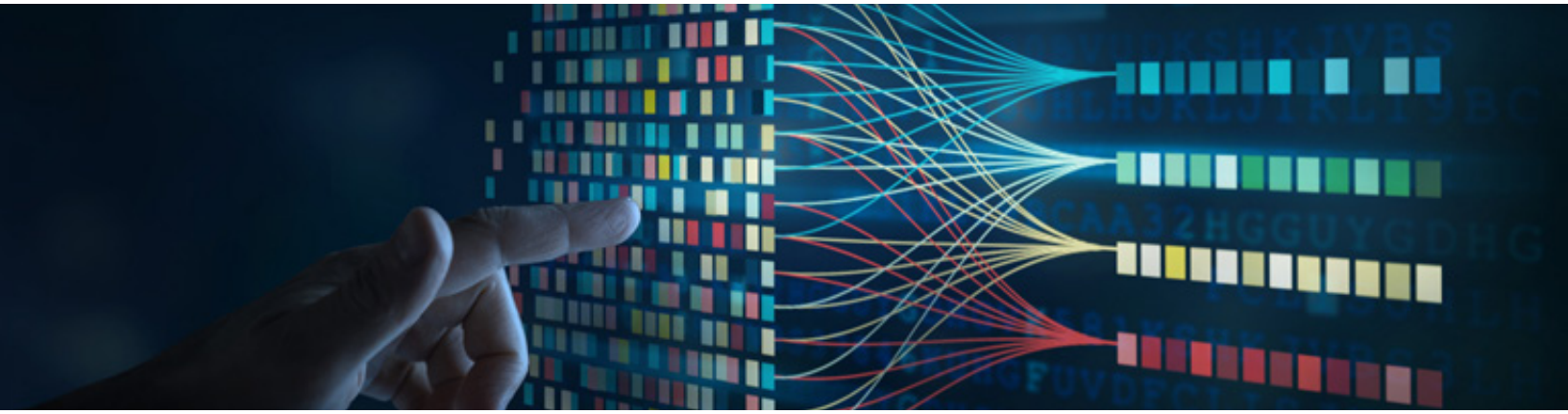
849
academic-industry collaborations

338
patent applications

*2017 – 22

Ontario's shared DRI platform is part of a publicly funded national network used by researchers at hospitals, research institutes, universities and colleges. It consists of supercomputing clusters run by two academic consortia, SHARCNET and SciNet, and several other domain-specific advanced research computing systems, including HPC4Health and Centre for Advanced Computing (CAC). More than 70 highly skilled staff are distributed across 13 institutions. Researchers throughout the province can access the infrastructure free of charge, enabling them to make use of sophisticated artificial intelligence, data analytics, and predictive simulations that cannot be performed on standard computers.

Ontario's DRI resources strategically deliver additional value by supporting Ontario's world-class research facilities, such as SNOLAB in Sudbury and the Centre for Applied Genomics led by SickKids. The work being done by researchers using these facilities, which receives significant public funding, supports a range of high-profile, international research collaborations.



Artificial intelligence (AI) has the potential to transform virtually all business and social sectors. However, the computational and storage demands of AI are extreme. For example, the largest machine learning models today need 10 billion times more compute power to train them than the most sophisticated models needed just ten years ago. Very few researchers can afford to purchase this amount of computing power commercially. Instead, they rely on Ontario's shared DRI platform.

Training opportunities enabled by Ontario's DRI ecosystem are essential. Rapid technological advancement within the advanced computing sector requires that skills be regularly refreshed, if researchers are to take full advantage of new capabilities and features. The shared DRI platform serves as a training ground where researchers develop and refresh these skills in critical areas such as machine learning, big data analytics, and advanced simulation techniques.

Ontario's shared DRI platform also helps to support large and ambitious provincial initiatives such as the Life Sciences Strategy and the Driving Prosperity plan. Many researchers rely on Ontario's DRI platform to develop new vaccines and therapeutics, advanced materials for transportation, new battery technologies, and the sensors and neural networks needed for autonomous vehicles.

To summarize, Ontario researchers use the province's shared DRI resources to help solve a wide variety of research problems that have direct economic and societal benefits. The stories you'll read on the next few pages are just a handful of examples of critically important research that would be impossible to perform without advanced computing — research that could not have been imagined 20 years ago.

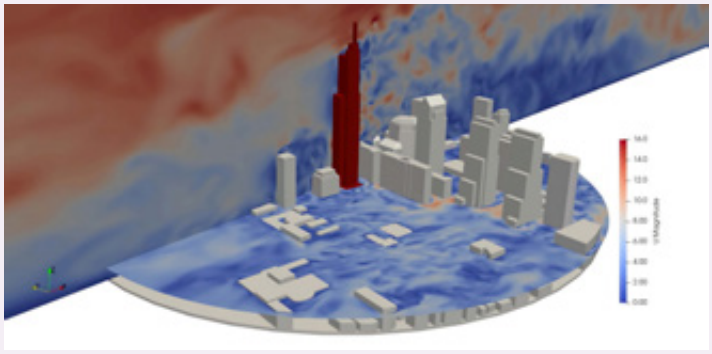
"The pace of technological advancement in artificial intelligence demands that we foster innovation by ensuring researchers have access to state-of-the-art computing power, data storage, and networking capabilities."

— integrate.ai

The impact of DRI-enabled research

Designing climate-resilient, sustainable buildings

Images courtesy of WindEEE Research Institute, Dr. Girma Bitsuamlak



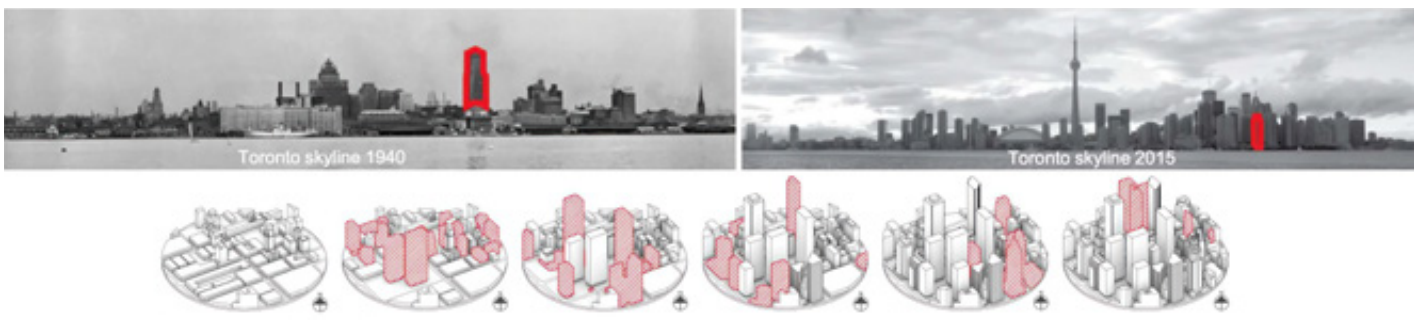
Data visualization of urban wind generate aerodynamic forces on a supertall building. Data visualizations are a valuable tool in data analysis, teaching, consulting and public education.

"Advances in computer technology allow us to predict the exact consequence of every design decision."

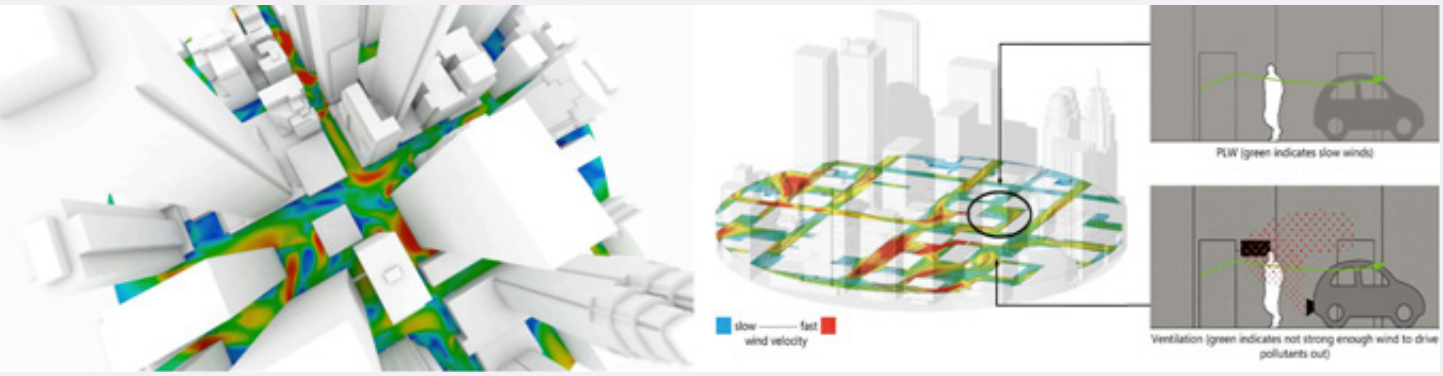
Wind tunnel experiments date back to the late 1800s, when people first began to build machines that fly. Wind tunnels still play an important role in measuring the effects of air pressure, velocity and turbulence, and evaluating the design of everything from golf balls to bike helmets. Unique wind tunnels also exist to test tall buildings for wind design. But when it comes to designing sustainable and resilient buildings, turbulence is not the only important element — and that's where DRI-enabled computational research becomes crucial.

"Climate-resilient design goes beyond the turbulence modelling that is done efficiently in wind tunnels," says Dr. Girma Bitsuamlak, director of Western's WindEEE Research Institute. "It's a multi-physics problem." Designing a building to withstand extreme weather, for example, means looking at the combined effects of wind, heat or cold, rain or snow, airborne debris and multi-scale features of the terrain. This type of problem can only be investigated through computer simulations requiring vast amounts of data storage and processing power. While there are upper and lower size limits to experiments in a wind tunnel, in a computer simulation, there are none. Further, a computer simulation can model how micro-climate stressors affect not just one building, but whole neighbourhoods or cities. In one study, for example, the team investigated the effects of city growth on micro-climate conditions at street level, finding that, in some areas of downtown Toronto, winds can be accelerated by the built terrain to speeds that are unsafe for pedestrians and cyclists. The same study is used to assess how the wind helps clean up air pollutants.

Digital WindEEE: Microclimate variations due to urban topology changes



Advanced Research Computing



WindEEE Research Institute investigates complex interactions among micro-climate stressors, terrain and built infrastructure.

The purpose that guides the WindEEE Research Institute is the goal of enhancing performance of new and existing buildings and infrastructure in response to wind and other climate stressors. “The point is to help communities prepare for and adapt to extreme weather and to develop sustainable approaches to building design,” Dr. Bitsuamlak says. “How can we improve design for new buildings or retrofit older buildings to be safe, resilient and energy-efficient?”

The team meshes experimental work in Western’s WindEEE Dome — the most advanced wind facility in Canada — with Computational Fluid Dynamics, applying these methods to specific real-world problems, often brought to them by municipalities and architectural or engineering consultants. In one recent study, for example, the WindEEE Research Institute used computer modelling to simulate four styles of architectural detail typically found in Toronto’s downtown office and condominium towers. They investigated which type of detail is most effective in insulating the building from excessive heat loss or heat collection. (The winner? The “egg crate” design.) In another project, the team created a digital twin of a hockey arena, modeling the micro-physics of the building to investigate and solve a serious condensation issue. Dr. Bitsuamlak’s team also uses high-resolution computer simulations to determine how windows in a high-rise building can be configured to optimize lighting and minimize heat loss.

The advantage of computer simulation, based on observed data, is that you can easily ask “What if?” Various parameters can be changed and the results of these variations measured, relatively quickly and inexpensively. “Advances in computer technology have allowed us to deploy high-resolution models and nail down the exact consequence of every design decision,” Dr. Bitsuamlak says. “Then we can start implementing these design principles to improve tornado-resistance, heat efficiency and so on.”

Dr. Bitsuamlak notes that this work could not be done without extensive DRI resources. “It’s not only hardware, it’s expertise,” he says, adding that he could not have adopted these computational methods without SHARCNET’s expert staff and the training and software support they provide. After many years, Bitsuamlak and his team of students and post-doctoral fellows are now computational experts themselves, but they still rely on SHARCNET’s assistance when they hit a tough problem. ”I can dream big, because I know they are there to assist when I get stuck,” Dr. Bitsuamlak says. The next big dream? Adding artificial intelligence to the WindEEE tool box. “It takes massive data sets to support machine learning and AI,” he says, “and we’re getting close.”

Photo by Alex Dekker | Left to right: Oliver Trottier, Ethan Lee, Eugene Klyshko, Sarah Rauscher, Liam Haas-Neil, Justin Sung-Ho Kim, Jack Gwozdecky, Mark Atwood

Discovering the blueprint of disordered proteins to pave a path to new therapeutic strategies

Image courtesy Dr. Sarah Rauscher

Changes in proteins are at the root of many diseases. Viruses and bacteria, for example, target specific proteins and bind with them, while protein mutations are involved in diseases like cancer. Treatment design, therefore, often starts with getting a clear understanding of the structure of the proteins involved in the disease. Dr. Sarah Rauscher and her team at University of Toronto are using advanced computing methods to better understand disordered proteins — proteins whose structure changes depending on the surrounding conditions.

"We look at proteins as a machine," Dr. Rauscher says. "We are aiming to describe in precise detail how this machine works, to develop a blueprint." Rauscher points to the protein associated with childhood leukemia as an example. "In some cases, there are mutations in this protein that can make the disease more aggressive or difficult to treat," she says. "If we can understand this mutation at the molecular level — which we don't, yet — we can design more effective treatments."

"Fifteen years ago it took me 8 months to do all the simulations for my first Ph.D. paper. Now I could do them all in one day."

Computational methods are perfectly suited to studying the molecular structures of disordered proteins, which cannot be fully investigated through experiment. “By simulating these proteins, in collaboration with people who are working on them using experimental approaches, we can develop a complete picture of what they look like and how they move,” Dr. Rauscher says.

A key area of study for Dr. Rauscher’s team is the role of water in information transfer within proteins. Information transfer occurs, for example, when a molecular interaction in one location on the protein molecule triggers a change in another location. Proteins can only function in water, so water must play some role in protein interactions. “But what exactly is the water doing?” says Dr. Rauscher. “How is the water important to the protein’s movement?” Dr. Rauscher and her team recently published a method for computationally analyzing and describing this movement, such that the movement of individual water molecules can be tracked on a time scale of less than a millionth of a second. This allows researchers to observe the interaction between the water molecule and the protein molecule at a tremendously granular level.

These simulations are computationally challenging and some can take on the order of 1500 CPU cores for, in some cases, a full year. “Fifteen years ago when I was doing my Ph.D., it took me 8 months to do all the simulations for my first paper. Now I could do them all in one day,” Dr. Rauscher says, thanks to the resources available through Ontario’s shared DRI platform.

“Computational methods are perfectly suited to studying the molecular structures of disordered proteins, which cannot be fully investigated through experiment.”

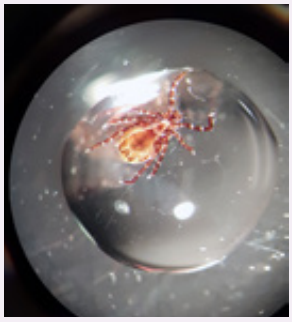
Using cutting-edge technology to understand tick-borne disease

Dr. Robert Colautti and his team at Queen's University investigate how genetic variation allows species to persist and thrive in challenging environments. From SARS-CoV-2 and Lyme disease to tick and plant invasions, the group is using a combination of large-scale experiments and cutting-edge genomics, analyzed with computational tools.

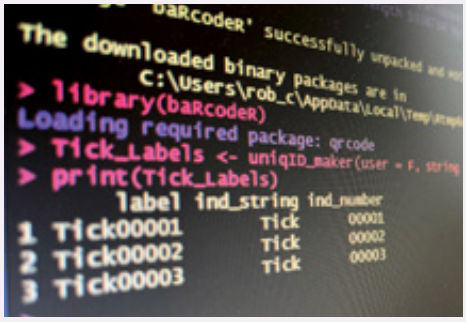
With the spread of deer ticks in Canada, and the likelihood that new species of ticks will be moving north from the U.S., it's important to have a complete picture of the pathogens they carry. Instead of focusing on a specific microbe, like the bacteria that cause Lyme disease, Dr. Colautti and his team are taking a broader approach. They use high-throughput gene sequencing to compare DNA of ticks and the microbes they carry. Using these tools, the team can identify these known pathogens as well as new variants and new pathogens.

All images courtesy of Colautti Lab

Dr. Robert Colautti



Close-up of code for the baRcodeR package.



Underside view of male deer tick in extraction buffer

"With new genomic technology, we can sequence the ticks and all the microbes they carry within a day – something that would have taken decades in the 1990s," Dr. Colautti says. This increase in speed has enabled researchers to use genetic sequencing to track COVID-19 variants and other diseases. The Colautti Lab is working on a proof of principle to show that a similar method can be used cost effectively to identify and track tick-borne pathogens. This involves development of customized bioinformatics "pipelines" that string together different algorithms, a process that is "heavily dependent on computational infrastructure," says Dr. Colautti. A typical project in his team involves the stores and analysis of several terabytes of genetic data on high-performance computers at the Centre for Advancing Computing (CAC).

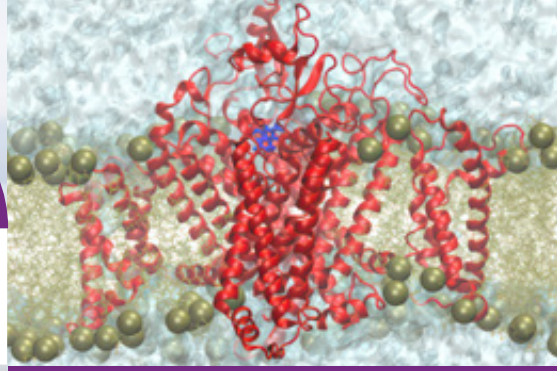
If the proof of principle is successful, there are potential applications in clinical medicine, epidemiology and public health. One possible direction could be the creation of a standards-of-practice for analyzing tick samples sent in by health professionals. Similar genomic data and bioinformatics tools are also helpful in other aspects of Dr. Colautti's work investigating how species evolve and spread into to new environments.

Like many scientists who have developed advanced computational expertise "on the job", Dr. Colautti and team are also making significant technical contributions to the field. They helped to analyze some of the first SARS-CoV-2 sequences in Canada, and they have developed an open-source tool called baRcodeR, which assists biology labs in generating unique identifier strings with printable barcodes. The use of bar codes to label biological samples reduces human error and facilitates the collection, tracking and curation of large research datasets.

Dr. Colautti has released coding textbook called "R Crash Course for Biologists", and he is working on another. "So much is written for computer scientists and engineers, I wanted something that is approachable for biologists who often find math and coding intimidating at first," Dr. Colautti says.

Dr. Christopher Rowley

Sheyla Montero



Researchers in Rowley Group simulate the channels that allow sodium ions to cross the cell membrane when a neuron fires

Machine learning new models for drug-protein interactions

Dr. Christopher Rowley and his team at Carleton University are advancing the potential of neural networks to predict how molecules will interact with each other. “Neural networks allow us to leverage vast computational resources to build models that can predict the interactions between a drug and its target more accurately. This has the potential to finally make computer aided drug discovery a reliable tool,” Dr. Rowley says. Neural networks can predict the stability of a molecular structure in a fraction of the time needed by existing methods. They’re useful in situations where the calculations would be prohibitively complex or expensive. Until recently, neural networks have been effectively used to predict interactions within molecules. Dr. Rowley and others are developing neural networks that can also accurately predict interactions between molecules.

These ‘long-range’ interactions occur, for example, when a drug binds to its protein target or when gases are absorbed by a carbon-capture material. Neural networks can be used to simulate these interactions and predict things like how well different molecules will bond with each other or which version of a molecule’s structure will be more stable. With support from SHARCNET, in collaboration with a research team at Dalhousie University, Rowley Group has developed a neural network that can accurately predict weak, long-range interactions.

The team’s focus is enhancing this neural network to be applicable to a broader range of molecules and stronger types of interactions. The work the team is doing is also relevant to drug discovery through machine learning, in which computers search huge catalogues of drugs, looking for molecular structures that are likely to bond with the specific protein associated with a particular disease. Rowley Group is teaching their neural network to recognize drugs whose molecular structures lend themselves to covalent bonding, which can increase both drug toxicity and effectiveness. At the same time, Rowley Group works with other researchers and industry partners to embed neural networks into other simulations that, for example, calculate the binding affinity between a drug and a protein involved in the disease the drug is designed to treat.

Recently, Dr. Rowley’s team has begun a collaboration with WEX Pharmaceuticals Inc., a Canadian company which is now conducting late-stage clinical trials on a novel non-opioid pain medication derived from puffer fish toxin. In minuscule doses, tetrodotoxin binds to the sodium channels on peripheral neurons and blocks the transmission of pain signals to the brain. Rowley Group is working with WEX to use simulations to better understand variations in how people respond to the drug.

Sheyla Montero, a PhD candidate in the Rowley group is leading this project. She is using molecular dynamics simulations of sodium channels embedded on the surface membrane of a model neuron. By simulating how tetrodotoxin blocks some variations of these channels, she is to provide insight into how this molecule can block signals that cause chronic pain. particular disease. Rowley Group is teaching their neural network to recognize drugs whose molecular structures lend themselves to covalent bonding, which can increase both drug toxicity and effectiveness. At the same time, Rowley Group works with other researchers and industry partners to embed neural networks into other simulations that, for example, calculate the binding affinity between a drug and a protein involved in the disease the drug is designed to treat.

Recently, Dr. Rowley’s team has begun a collaboration with Wex Pharmaceuticals, a Canadian company which is now conducting clinical trials on a non-opioid pain medication derived from puffer fish toxin. In miniscule doses, tetrodoxin binds to the sodium channels on neurons and block the transmission of pain signals. Rowley Group is working with Wex to use simulations to better understand variations in how people respond to the drug.

Using big data to support genetic research in Canada

Images courtesy of CGEn

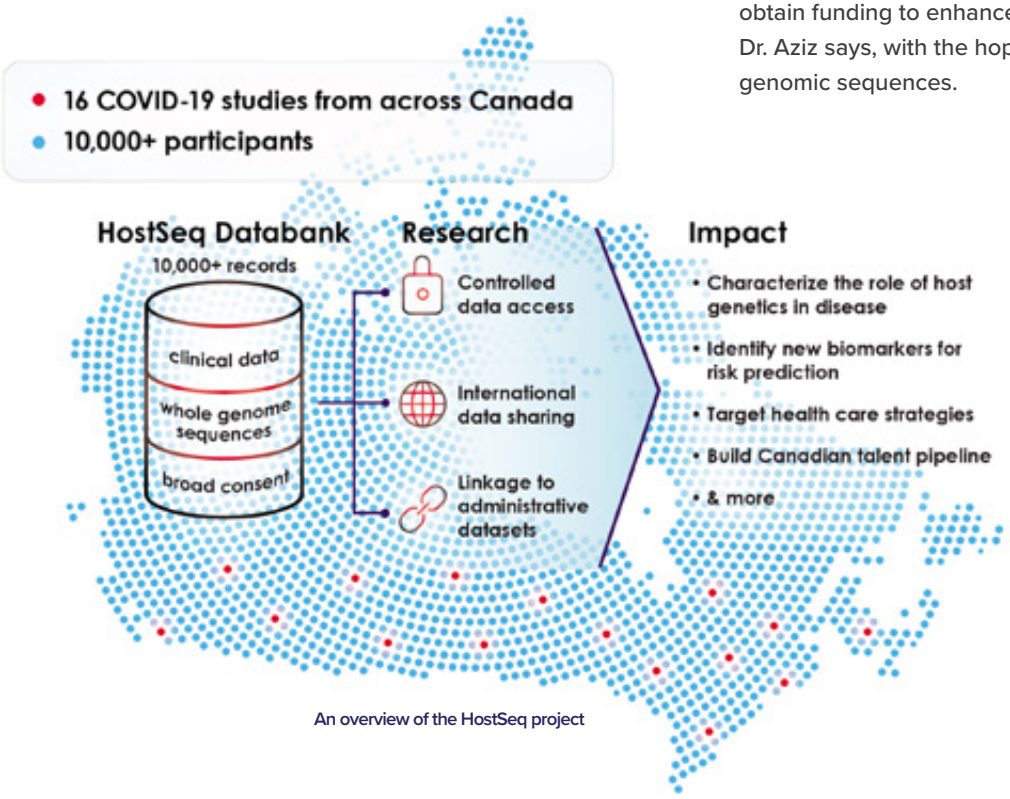
Dr. Naveed Aziz

With the completion of the HostSeq initiative in late 2021, researchers can now access the whole genome sequences of over 10,000 Canadians, all of whom have consented to their data being accessible for any scientific or medical research. This is a first in Canada; other genetic databanks are accessible exclusively for specific research purposes, such as treating rare genetic disease. For privacy and security reasons, researchers who wish to access HostSeq data must be approved by an arm's length Data Access Compliance Office.

"This kind of research requires enormous amounts of computing power and sophisticated software tools," says Dr. Naveed Aziz, CEO of CGEn. CGEn is the federally funded national platform for genome sequencing and analysis that led the HostSeq initiative. But he notes it is the challenge of data management and stewardship that "keeps me up at night." HostSeq's genomic sequences and associated clinical records amount to more than 1 petabyte of data. That's 1 million gigabytes or the equivalent of 500 billion pages of standard printed text.

A data set of this size and complexity would be useless without proper cataloging and annotating to ensure it is FAIR (findable, accessible, interoperable and reusable). Good data management is essential to enable researchers not only to find, download and analyze the data they need, but to have confidence in its accuracy and therefore in the results of their analysis. "If you are concerned that, for example, an element of your metadata might be innacurately recorded or stored, you lose confidence in the whole dataset," he says. He notes that clear, reproducible policies and processes also ensure that the dataset retains its integrity as new data is added.

Host Seq, which took a large team nearly three years to build, relies on Ontario's HPC4H consortium among other platforms for data storage and computing power. The databank has so far supported 26 different research projects. "We are now working to obtain funding to enhance the data set for Canadian studies," Dr. Aziz says, with the hope of eventually acquiring 100,000 genomic sequences.



An overview of the HostSeq project

DRI: Supporting Innovation and Industry

Ontario's shared DRI platform is used by researchers developing innovative solutions for real-world problems. The following is a limited sampling of recent and ongoing projects, and their impacts, that have been reported by researchers who make use of Ontario's shared DRI platform.

Improving patient care and reducing health-care costs

- Developing new pain therapeutics [Mike Hildebrand, Carleton University]
- Developing a process to perform thousands of diagnostic tests in a single hour from a drop of blood [Hannes Rost, University of Toronto]
- Developing new stem cell and anti-cancer therapies [Theodore Perkins, Ottawa Hospital Research Institute]
- Developing deep learning models to advance therapy for acute stroke [Maged Goubran, Sunnybrook Health Sciences Centre]

Ensuring public health and safety

- Improving the acoustic efficiency of commercial aircraft in order to reduce the noise signature around airports and surrounding residential areas [Phillipe Lavoie, University of Toronto]
- Modeling the human body and protective equipment in automotive crash and sports impact scenarios to improve safety systems and reduce injuries [Duane Cronin, University of Waterloo]

Advancing Ontario's agro-economy

- Making a grapevine genetic test service available in Canada for the first time with much better affordability and accuracy [Liang Ping, Brock University]
- Genomic evaluation of breeding values of dairy cattle, swine, and Canadian major crops [Zeny Feng, University of Guelph]
- Developing stress-tolerant crop plants [Owen Rowland, Carleton University]

Boosting high-tech industries and commercialization

- Designing next-generation integrated circuits [Piero Triverio, University of Toronto]
- Fabricating and testing of quantum computing devices [Rafael Kleiman, McMaster University]

Reducing greenhouse gas emissions

- Developing photobioreactors for the capture of carbon dioxide [Christopher DeGroot, University of Western Ontario]
- Enabling new technologies which reduce the environmental impact of aviation [Jeffrey Defoe, University of Windsor]
- Designing lower polluting engines and fuels, and innovating fossil fuel alternatives [Seth Dworkin, Toronto Metropolitan University]

Investigating financial systems and markets

- Developing machine learning models to more accurately forecast important financial risk measures [Fulei Lu, University of Guelph]
- Optimize large scale neural-networks to support applications in inventory and pricing problems [Joseph Milner, University of Toronto]

Ecosystem Support

In 2023 – 23, Compute Ontario contributed directly to the health of the ecosystem and the enhancement of DRI skills, knowledge and resources through financial services and advice, project coordination, training support and funding for small RDM and training projects.

Financial Services

This year, we supported the consortia by coordinating funding agreements and reporting, approving budgets and disbursing funds. In addition to these annual activities, we developed a comprehensive funding proposal requesting $56 million in provincial investment over fiscal 2024 and 2025, to renew Ontario's aging DRI infrastructure, provide salaries for consortia staff, and cover incremental cyber-security expenses. This provincial co-investment is required to access federal funding under ISED's Digital Research Infrastructure contribution program.

Funding RDM and Training Advances

At the end the fiscal year, Compute Ontario awarded nearly $380,000 to 7 collaborative teams to fund RDM projects to be completed by March 31, 2023. As detailed in the chart below, these teams generated a variety of valuable outcomes. They produced training materials and events, advanced the understanding of specific DRM issues and challenges, migrated an important data set into a national research data repository, and developed an interactive data visualization tool to improve discoverability and reusability of metadata in the Polar Data Catalogue.

Outcomes of 2022 – 2023 RDM and Training Projects

Research Team	Project Outcomes
Dr. Michelle Edwards (University of Guelph), Dr. Lucas Alcantara (University of Guelph), Dr. Carly Huitema (University of Guelph)	This project offered a series of four workshops to train research data users in: metadata creation to ensure new data and their associated dashboards are Findable; understanding the importance of data reusability and how to grow a FAIR data culture; Basic R programming skills for data tidying; and advanced research computing skills (R Shiny) needed to develop an interactive data explorer dashboard focused on improving data reusability and discovery.
Dr. Helen Chen (University of Waterloo), Dr. Alan Forster (Ottawa Hospital), Dr. Catherine Burns (University of Waterloo), Dr. Zahid Butt (University of Waterloo), Dr. Plinio Morita (University of Waterloo), Dr. William W.L. Wong (University of Waterloo)	This project used chimeric antigen receptor T cells (CAR-T) therapy data as a case study to explore and discuss the use of synthetic health data to advance research. Project outcomes included: a rapid scoping review of current use of synthetic data; a round table discussion on FAIR principles for synthetic health data; establishment of collaborative relationships with key stakeholders; exploration of data needs in CAR-T therapy and creation of synthetic health data; securing funding for further collaborative work on synthetic data governance framework and demonstration. The project advanced understanding of how to address and manage synthetic data and metadata needs and issues related to data ingestion, transformation, and preservation in the RDM workflow and process; and what policy changes are needed for data governance of sharing FAIR health synthetic data in Pan-Canadian networks.
Kate Davis (University of Toronto), Nana Boateng (University of Toronto), Alicia Urquidi Diaz (University of Toronto), Amber Leahey (University of Toronto), Bart Kawula (University of Toronto), Victoria Lubitch (University of Toronto), Guinsly Mondesir (University of Toronto), Hafsah Hujaleh (University of Toronto)	CO funding contributed to the successful migration of ODESI, a collection of statistical data maintained by the Ontario Council of University Libraries (OCUL), into Borealis, a national research data repository, creating a single site for shared, open repository and curation infrastructure, reducing duplication, and improving workflows for both library-created and researcher-deposited data collections.
Dr. Christina DeRoche (Canadore College), Oliver Goodison-Powell (Conestoga College), Dr. Timothy Larocque (Confederation College), Dr. Brett Goodwin (Fleming College), Donna Sevenpifer (Fanshawe College), Dr. Vicki Mowat (Sheridan College)	This project advanced overall learning and RDM development within the College sector by linking College Research Administrators and staff within Research offices to College Librarians and resources in support of the development of institutional strategies and capacity building. Further, this project developed two college-specific modules: Data Management Planning for the College Sector, and Data Deposit from a College Perspective.

Dr. Barbara Fallon (University of Toronto), Dr. Dale Turner (University of Toronto), Dylanne Dearborn (University of Toronto)	This project resulted in successful engagement with the University of Toronto research community to improve access to resources, support and training for RDM reflective of the diverse scholarship at the University of Toronto and in particular a commitment to respecting Indigenous data sovereignty, recognition of Indigenous knowledge systems and the need for co-development of research partnerships, processes and tools. Outcomes include identification of the need for (1) A single point of RDM tools, supports and expectations for researchers (2) Flexible RDM processes and tools responding to a continuum of needs including disciplinary and FNMI research, and (3) Coordinated efforts to develop and integrate a wide range of RDM supports and systems. Further, the team partnered with the Indigenous Research Network (IRN), an interconnected and collaborative community of researchers at U of T that respects and honours Indigenous cultures, to successfully gain insights on: (1) How Indigenous research is conceptualized; (2) How relationships with FNMI communities are nurtured and celebrated; and (3) How the data/findings from the research are defined, co-created, shared, and preserved.
Art Rhyno (University of Windsor), Dr. Berenica Vejvoda (University of Windsor), Dr. Paul Preney (University of Windsor)	This project successfully developed and deployed three JupyterHub training workshops and an associated Codefest event to help the University of Windsor build capacity for using Jupyter notebooks in support of advanced computational work and to assist the university in furthering its institutional RDM strategy goals by surfacing RDM services and RDM-supportive infrastructure in a JupyterHub environment.
Dr. Gregory Vey (University of Waterloo), Waleed Ashfaq (University of Waterloo)	This project successfully developed an interactive data visualization tool (https://www.polardata.ca/explore) to improve discoverability and reusability of polar metadata and data at the Polar Data Catalogue. Accompanying project documentation and source code are available at the same location, to serve as a reusable and adaptable model for new projects with related goals and/or functionality. This project yielded three specific deliverables: 1) the deployed data visualization, 2) the accompanying documentation (i.e. user guide and cookbook), and 3) the visualization source code.

Building Community

Communication, collaboration and partnership are key to achieving the shared goal of a robust, efficient and accessible DRI ecosystem. In 2022 – 23, we continued to build connections across the DRI community, expanding not only our networks but also our approaches to engaging with stakeholders.

Community Participation

In 2022 – 23 Compute Ontario served on an increased number of committees, conference panels and working groups to contribute insight and help to catalyze greater awareness and coordination, particularly in the area of Research Data Management (RDM). These included several Alliance committees, such as the Executive Management Committee; the Curation Event, Cloud Survey and Data Architecture Working Groups; the National Training Expert Group; and the RDM Network of Experts. Compute Ontario is also actively engaged with the Ontario Data Community, which includes RDM representatives from across Ontario.

Photo courtesy of Seneca

Compute Ontario Committees

Compute Ontario relies on the strategic input and hands-on assistance of three standing committees: Technology Leadership Advisory Council (TLAC), Research Data Management Advisory Committee (RDMAC) and Training Advisory Committee (TAC). These committees involve a wide range of stakeholders including (at time of writing) eight universities, three colleges, two research institutes, and a hospital. In addition, ad-hoc committees are formed as the need arises. The Advocacy Committee, as an example, was brought together for several months in 2023 to provide input and advice related to provincial/national funding and governance issues. In July 2023, a group of stakeholder representatives was brought together for a half-day working session to provide input into Compute Ontario's strategic priorities.

University and College Outreach

CO has intensified its engagement with universities through increasing representation on its Board of Directors, regular updates to the Ontario Council on University Research (OCUR), and ongoing communication with a core group of Vice Presidents, Research. CO is establishing positive relationships with Ontario colleges to gain an understanding of the DRI needs of college researchers. With the knowledge gleaned from these discussions, CO has begun meeting with college heads of research and representatives from Colleges Ontario.

Ontario's DRI Community

Research Institutes

CLINICAL TRIALS ONTARIO, which works collaboratively with industry, research institutes, patients and others to improve the clinical trials environment and attract investment to the province, while supporting the highest ethical and quality standards.

FACIT, which bridges cancer research and innovation to real world opportunities that benefit investors, patients and our economy. ICES, a not-for-profit research institute encompassing a community of research, data and clinical experts, and a secure and accessible array of Ontario's health-related data.

ONTARIO BRAIN INSTITUTE, which works to establish Ontario as a world leader in brain research, commercialization and care. ONTARIO GENOMICS, a not-for-profit organization that manages cutting-edge genomics research projects and platforms.

ONTARIO INSTITUTE FOR CANCER RESEARCH (OICR), which collaborates with partners in Ontario and around the world to accelerate new cancer research.

PERIMETER INSTITUTE, an independent research centre in foundational theoretical physics.

ROTMAN RESEARCH INSTITUTE, BAYCREST, a premier international centre for the study of human brain function.

VECTOR INSTITUTE, an independent, not-for-profit corporation dedicated to research in the field of artificial intelligence (AI), excelling in machine and deep learning.

Academic and Medical Research Communities

Ontario universities

OCUR, the Ontario Council of University Researchers

OHA, the Ontario Hospital Association: Health Research and Innovation Committee

The Academic Consortia

CAC, the Centre for Advanced Computing, is based at Queen's University. CAC specializes in secure advanced computing resources for highly confidential data, and support for academic and medical researchers.

HPC4Health is a partnership between SickKids and the Princess Margaret Cancer Centre at University Health Network, providing clinical researchers with secure cloud-computing services, while satisfying personal health information privacy requirements.

SciNet is led by University of Toronto and hosts the Niagara system. SciNet provides Canadian researchers with computational resources and expertise necessary to perform their research on scales not previously possible in Canada.

SHARCNET is a consortium of 19 universities and colleges, covering a geographical span of about 1800 km from Windsor to Peterborough, and St. Catharines to Thunder Bay, making it the largest HPC consortium in Canada. SHARCNET is responsible for running the Graham system which is located at the University of Waterloo.

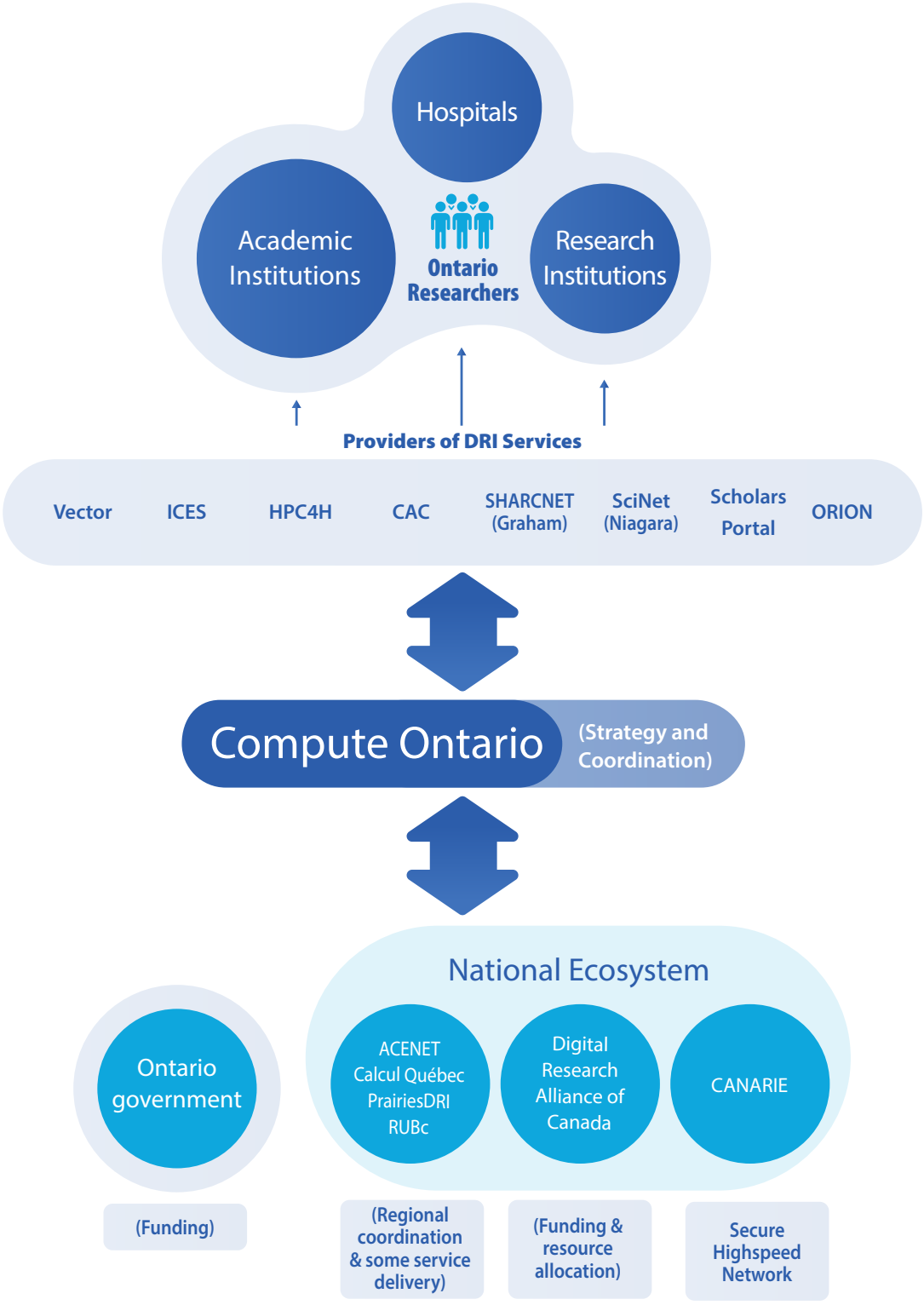
DRI Partners

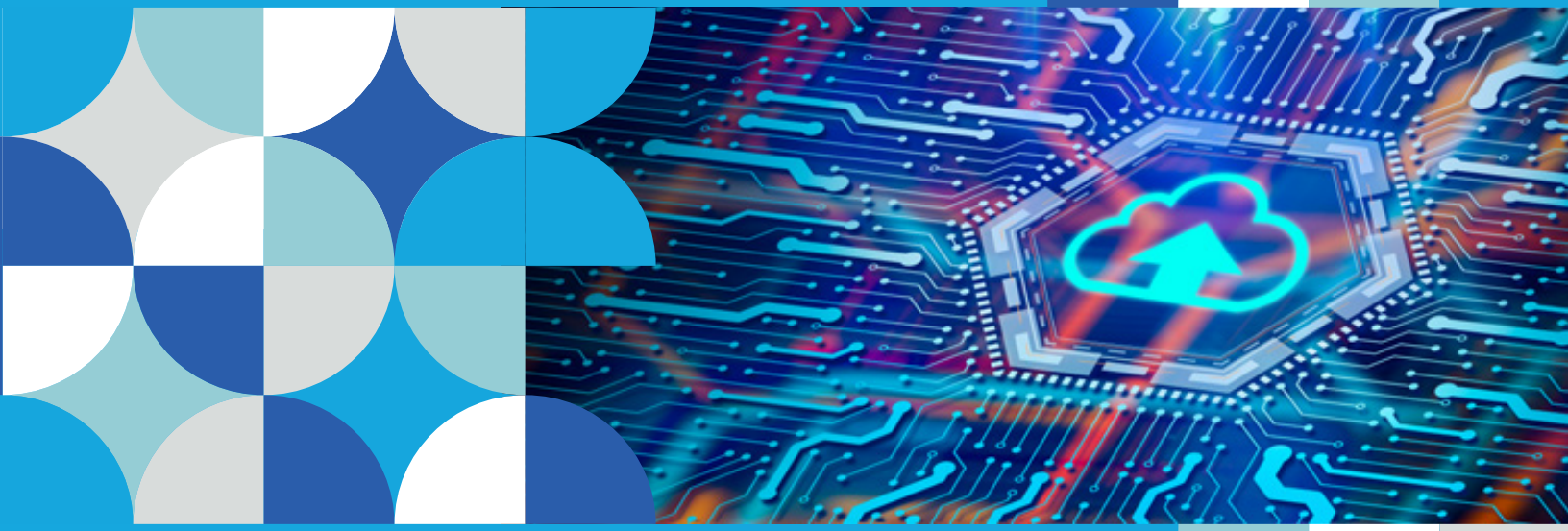
OCUL, the Ontario Council of University Libraries, and the OCUL Data Community

ORION, the only high-speed, fibre-optic network specifically committed to supporting research, education and innovation in Ontario.

The Digital Research Alliance of Canada, which coordinates and funds DRI activities in collaboration with Compute Ontario and other stakeholders.

Ontario's DRI Ecosystem





Financial Statements

Schedule of Expenditures of

COMPUTE ONTARIO / CALCUL ONTARIO

MINISTRY OF COLLEGES AND UNIVERSITIES FUNDING

And Independent Auditor’s Report thereon

Year ended September 30, 2023

KPMG LLP
140 Fullarton Street, Suite 1400
London, ON N6A 5P2
Canada
Telephone 519 672 4880
Fax 519 672 5684

INDEPENDENT AUDITOR'S REPORT

To the Management of Compute Ontario / Calcul Ontario and the Minister of Colleges and Universities (the "Ministry")

Report on Audit of the Schedule

Opinion

We have audited the accompanying schedule of expenditures of Compute Ontario / Calcul Ontario for the Ministry of Colleges and Universities (formerly the Ministry of Research and Innovation) (the "Ministry") funding for the year ended September 30, 2023 and notes to the schedule, including a summary of significant accounting policies (hereinafter referred to as the "schedule").

In our opinion, the accompanying schedule is prepared, in all material respects, in accordance with the financial reporting provisions in A7.0 of the two Ontario Transfer Payment Agreements dated October 22, 2021, and April 8, 2022, between Compute Ontario / Calcul Ontario (the "Organization") and the Ministry of Colleges and Universities.

Basis for Opinion

We conducted our audit in accordance with Canadian generally accepted auditing standards. Our responsibilities under those standards are further described in the ***"Auditor's Responsibilities for the Audit of the Schedule"*** section of our auditor's report.

We are independent of the Organization in accordance with the ethical requirements that are relevant to our audit of the schedule in Canada and we have fulfilled our other ethical responsibilities in accordance with these requirements.

We believe that the audit evidence we have obtained is sufficient and appropriate to provide a basis for our opinion.

Emphasis of Matter – Financial Reporting Framework

We draw attention to Note 2 to the schedule, which describes the applicable financial reporting framework.

The schedule is prepared to assist the Organization to meet the requirements of the funding agreement between Compute Ontario / Calcul Ontario and the Ministries as described in Note 1 to the schedule.

As a result, the schedule may not be suitable for another purpose.

Our opinion is not modified in respect of this matter.

Other Matter – Restriction on Use

Our report is intended solely for the Ministries and the Organization and should not be used by other parties.

Responsibilities of Management and Those Charged with Governance for the Schedule

Management is responsible for the preparation of the schedule in accordance with the financial reporting provisions in the three Agreements between the Organization and the Ministries dated October 22, 2021, April 8, 2022, and for such internal control as management determines is necessary to enable the preparation of the schedule that is free from material misstatement, whether due to fraud or error.

Those charged with governance are responsible for overseeing the Organization's financial reporting process.

Auditor's Responsibilities for the Audit of the Schedule

Our objectives are to obtain reasonable assurance about whether the schedule as a whole is free from material misstatement, whether due to fraud or error, and to issue an auditor's report that includes our opinion.

Reasonable assurance is a high level of assurance, but is not a guarantee that an audit conducted in accordance with Canadian generally accepted auditing standards will always detect a material misstatement when it exists.

Misstatements can arise from fraud or error and are considered material if, individually or in the aggregate, they could reasonably be expected to influence the economic decisions of users taken on the basis of the schedule.

As part of an audit in accordance with Canadian generally accepted auditing standards, we exercise professional judgment and maintain professional skepticism throughout the audit.

We also:

- Identify and assess the risks of material misstatement of the schedule, whether due to fraud or error, design and perform audit procedures responsive to those risks, and obtain audit evidence that is sufficient and appropriate to provide a basis for our opinion.

The risk of not detecting a material misstatement resulting from fraud is higher than for one resulting from error, as fraud may involve collusion, forgery, intentional omissions, misrepresentations, or the override of internal control.

- Obtain an understanding of internal control relevant to the audit in order to design audit procedures that are appropriate in the circumstances, but not for the purpose of expressing an opinion on the effectiveness of the Organization's internal control.
- Evaluate the appropriateness of accounting policies used and the reasonableness of accounting estimates and related disclosures made by management.
- Communicate with those charged with governance regarding, among other matters, the planned scope and timing of the audit and significant audit findings, including any significant deficiencies in internal control that we identify during our audit.

KPMG LLP

Chartered Professional Accountants, Licensed Public Accountants

London, Canada

December 15, 2023

COMPUTE ONTARIO / CALCUL ONTARIO

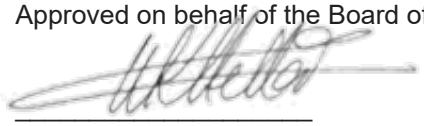
MINISTRY OF COLLEGES AND UNIVERSITIES FUNDING
Schedule of Expenditures

Year ended September 30, 2023, with comparative information for 2022

	Total 2023	Total 2022
Expenditures:		
Salaries, wages and benefits	$ 6,151,204	$ 1,852,604
Legal, HR, professional services	248,770	481,056
Special projects	569,960	297,542
IT, media and materials	35,680	54,222
Audit	16,923	24,374
Insurance	21,022	19,198
Conference, meetings and events	14,092	5,685
Miscellaneous, operating and administration	2,468	334
Training and development	1,360	165
Total expenditures	$ 7,061,479	$ 2,735,180

See accompanying notes to the schedule.

Approved on behalf of the Board of Directors:



Warren Helland

Chair, Board of Directors

COMPUTE ONTARIO / CALCUL ONTARIO

MINISTRY OF COLLEGES AND UNIVERSITIES FUNDING
Notes to Schedule of Expenditures

Year ended September 30, 2023

1. Project description:

Project Funding 1:

Compute Ontario / Calcul Ontario (the "Organization") signed an agreement dated October 22, 2021, with effective date October 1, 2021, with the Ministry of Colleges and Universities (formerly the Ministry of Research and Innovation) to fund the operating costs of the Organization (the "Project").

Under the agreement, The Ministry of Colleges and Universities (formerly the Ministry of Research and Innovation) (the "Ministry") will provide a maximum of $4,740,000 cash contribution to the Project for the project period, which ends on September 30, 2024. The objective of the Project is to create a cohesive DRI ecosystem for the province.

The schedule of expenditures of Compute Ontario / Calcul Ontario for the Ministry of Colleges and Universities funding presents the Organization's portion of eligible expenditures incurred for the Project during the reporting period.

Project Funding 2:

Compute Ontario / Calcul Ontario (the "Organization") signed an agreement dated April 8, 2022 with effective date April 1, 2022, with the Ministry of Colleges and Universities (formerly the Ministry of Research and Innovation) to fund the operating costs of the Organization (the "Project").

Under the agreement, The Ministry of Colleges and Universities (formerly the Ministry of Research and Innovation) (the "Ministry") will provide a maximum of $12,000,000 cash contribution to the Project for the project period, which ends on March 31, 2024. The objective of the Project is to leverage the national investment provided through The Alliance which supports costs for DRI staffing, operations and maintenance of the Ontario-based Advanced Research Computing (ARC) facilities.

The schedule of expenditures of Compute Ontario / Calcul Ontario for the Ministry of Colleges and Universities funding presents the Organization's portion of eligible expenditures incurred for the Project during the reporting period.

2. Significant accounting policies:

(a) Basis of accounting:

The schedule is prepared in accordance with the basis of accounting prescribed in the Reports Schedule of the Agreements, between Compute Ontario / Calcul Ontario and the Ministry.

The schedule has not been prepared in accordance with the presentation principles or the presentation of all the financial statements and related note disclosures required by Canadian accounting standards for not-for-profit organizations.

2. Significant accounting policies (continued):

(b) Expenditures:

Expenditures are eligible under the Project if they were incurred in the period of the Project and are directly related to the completion of the Project.

COMPUTE ONTARIO / CALCUL ONTARIO

MINISTRY OF COLLEGES AND UNIVERSITIES FUNDING
Schedule D

Year ended September 30, 2023

	CO Central	ARC Consortia	Total
Expenditures:			
Salaries, wages and benefits	$ 794,433	$ 5,356,771	$ 6,151,204
Legal, HR, professional services	248,770	–	248,770
Special projects	569,960	–	569,960
IT, media and materials	35,680	–	35,680
Audit	16,923	–	16,923
Insurance	21,022	–	21,022
Conference, meetings and events	14,092	–	14,092
Miscellaneous, operating and administration	2,468	–	2,468
Training and development	1,360	–	1,360
Total expenditures	$ 1,704,708	$ 5,356,771	$ 7,061,479

Compute Ontario Board of Directors

We extend our sincere gratitude to the members of our Board, who generously share their time, expertise and strategic guidance.

We would like to acknowledge and thank Board Director Shannon MacDonald, whose term ended this year.

WARREN HELLAND
Chair

ERIC BROAD
Director

SYLVAIN CHARBONNEAU
Secretary

LEAH COWEN
Director

MARK DALEY
Director

CHARMAINE DEAN
Director

SHANNON MACDONALD
Chair Audit & Resources Committee

ATEFEH MASHATAN
Director

CAROLYN MCGREGOR
Chair Industry Engagement Committee

SUSAN URSEL
Director

Compute Ontario

computeontario.ca

info@computeontario.ca