2022 ANNUAL REPORT

ADVANCING RESEARCH & INNOVATION IN ONTARIO

Compute Ontario

TABLE OF CONTENTS

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 DRI 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

DRI BY THE NUMBERS

MCU FUNDING FOR ONTARIO DRI STAFF, APRIL 1, 2022 – MARCH 31, 2024:
$12 MILLION

FUNDING FOR ONTARIO DRI, APRIL 1, 2021 – MARCH 31, 2022:

FEDERAL
$6,802,664

PROVINCIAL INSTITUTIONAL/OTHER
$10,139,421

VALUE OF RAC AWARDS MADE TO ONTARIO PIS:
$15.5M (est.)

PERCENTAGE OF RAC AWARDS MADE TO ONTARIO PIS:
37%

TOTAL USAGE OF ALLIANCE-FUNDED SYSTEMS BY ONTARIO-BASED PIS IN 2021 – 22:
110,000 CORES & 900 GPUS

Number of institutions who received Research Allocation Competition (RAC) awards:

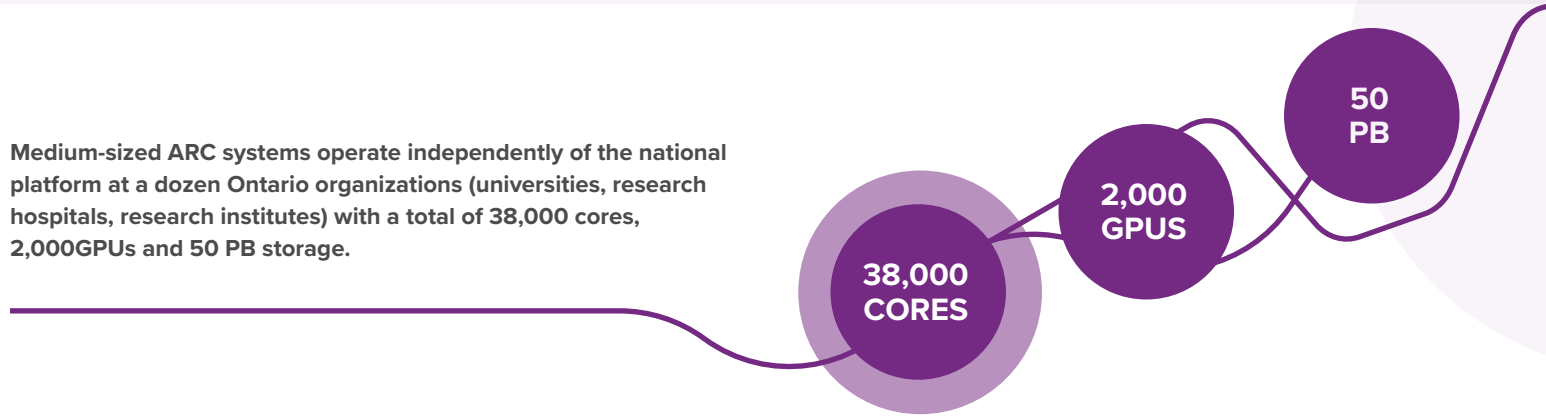
17 UNIVERSITIES

1 COLLEGE

4 RESEARCH INSTITUTES

4 RESEARCH HOSPITALS

Medium-sized ARC systems operate independently of the national platform at a dozen Ontario organizations (universities, research hospitals, research institutes) with a total of 38,000 cores, 2,000GPUs and 50 PB storage.



CHAIR & CEO MESSAGE

MARK DALEY
Board Chair

RANIL SONNADARA
President, CRO and CEO

This past year was one of continued evolution for the Digital Resource Infrastructure (DRI) ecosystem in Canada: Calcul Quebec launched a new high-performance computing cluster (Narval). WestGrid, the regional DRI coordinating body for the western provinces and BC, disbanded. Subsequently, PrairiesDRI emerged to coordinate DRI activities in Alberta and Saskatchewan, while coordination for BC was taken on by Research Universities of British Columbia (RUBC). National coordination of funding and services formally transitioned from Compute Canada to the Digital Research Alliance of Canada on March 31, 2022. Closer to home, Compute Ontario received an updated mandate from the provincial government, explicitly acknowledging Research Data Management (RDM) as an integral component of DRI and increasing our accountability for financial oversight and budget approval.

It has been a busy time!

By far the largest part of our attention and effort at Compute Ontario has been focused on working with the Alliance to troubleshoot various transition issues and attempt to shape funding and governance models that meet the expectations of both federal and provincial funders. An early and urgent issue centred on the availability of funding for Ontario DRI staff salaries to cover the period between the end of the Compute Canada funding and the start of the new Alliance funding year. Compute Ontario worked with the Ministry of Colleges and Universities (MCU) to secure $12 million in funding over two years to ensure that staff could be retained.

Finance continued to be a key theme in our work with the Alliance. Compute Ontario staff contributed significant amounts of time toward building business cases and financial models in support of a multi-year funding proposal (MYFP), submitted to ISED for review on Sept. 30, 2022. Through this process, we have advocated strongly for funding to refresh Ontario's Graham and Niagara systems, as they reach the end of their lifespan. We have worked hard to keep key stakeholders up to date, coordinate input and ensure that Ontario is truly speaking with one voice which represents the needs of our researchers. We have worked closely with our regional partners throughout the Alliance's planning processes, with the result that our cross-Canada relationships have never been stronger. As we enter Fiscal 2022 – 23, we are working with stakeholders and MCU to analyse funding requirements and advocate for co-investments.

We have continued to expand and deepen our relationships with the DRI community through our new Training Advisory Committee, engagement with Ontario colleges and research hospitals, and new connections and activities in the world of RDM. We have continued to support Ontario's ecosystem through project funding and coordination support for training and initiatives such as the Pan-Canadian AI Expansion project. We have advanced our knowledge of the ecosystem through a needs assessment for Ontario's colleges, an inventory of Ontario's DRI resources that are separate from the national DRI platform, a health research data use case study, and a cloud computing pilot project that is currently underway. These latter activities are particularly important at this time, as they will inform both Compute Ontario's strategic plan and DRI investment plans in the coming year.

MARK DALEY
Board Chair

RANIL SONNADARA
President, CRO and CEO

Photos courtesy of University of Windsor

Left to right: Dr. Sarfraz Ahmad, Dr. Abd Al-Aziz Abu-Saleh, Daniel Meister, Dr. John F.Trant, Dr. Farsheed Shahbazi, Dr. Muhammad Usman Mirza, Dr. Purshotam Sharma, Saeed Velayati

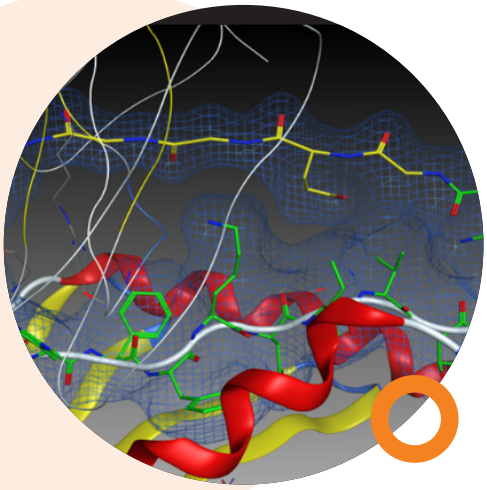
RESEARCH IMPACT

DESIGNING UNNATURAL MOLECULES TO CURE DISEASE

Dr. John Trant, an associate professor at the University of Windsor, works on the cutting edge of medicinal chemistry by designing, making, and testing new drugs. Integral to all of his team's work is computational chemistry, a discipline that, by definition, depends on access to Digital Research Infrastructure (DRI). The Trant Team uses computer simulations to design and test molecules with the goal of solving a diverse array of problems in biomedicine, from investigating the potential therapeutic properties of cannabinoids to attempting to cure celiac disease.

In celiac and other auto-immune diseases, the immune system targets and destroys one of the body's own cells as if it were a virus or bacteria. At the molecular level, this involves an immune receptor binding to a protein. For example, in rheumatoid arthritis, immune receptors bind to a part of the collagen protein, resulting in damage to the fluid that protects the joints. At present, the only way to prevent this is to shut down immune system activity altogether. While this helps to reduce damage to organs and joints, it also greatly increases the patient's risk of serious infection.

Clearly, a targeted approach would be preferable. Scientists are looking for a way to shut down the immune receptors involved in the disease without impacting other receptors in the body. They seek to build a protein-like molecule that would accomplish this goal and effectively cure the disease. The foremost challenge to this strategy is the complex structure of immune receptors which have multiple binding sites of various shapes. This complex structure enables our very few (8 – 12) types of immune receptors to defend us against the thousands of different viruses, bacteria and fungi we might encounter in our day to day lives. Designing a new molecule that can fully bind all sites on an immune receptor is a tall order. To complicate the situation further, some auto-immune diseases can involve as many as 4 or 5 different types of immune receptors. Given the time and expense required to build and test even one new molecule in a lab, it's understandable that progress has been slow.



By contrast, the computational approach is fast and inexpensive. The Trant Team has taken off-the-shelf software and customized it to enable computerized molecule design. Using computer simulations, the team can quickly test hundreds of thousands of molecule designs to weed out the ones that will not work and identify the promising candidates. The team can then make and test a small number of molecules in vitro and in humanized mice (mice whose immune system has been removed and replaced with a human analog), with a greater likelihood of success.

"This would have been impossible even as recently as ten years ago," says Dr. Trant. "Computing speed was too slow, algorithms weren't optimized." Even now, he says, "you could not do this research with a couple of computing clusters on campus." The Trant Team annually uses nearly 4,000 CPU-years of processing power and 1.4 petabytes of storage provided largely by SHARCNET, a consortium of 19 universities and colleges. SHARCNET runs one of the largest high-performance clusters in Canada: Graham, located at the University of Waterloo.

ACTIVATING CULTURAL ARCHIVES FOR MARGINALIZED COMMUNITIES

Photos courtesy of Centre for Advanced Computing, Queen's University

It might seem surprising that the Vulnerable Media Lab (VML) at Queen's University is one of the most intensive users of large-scale storage at the Centre for Advanced Computing (CAC). However, VML digitizes and restores film and video, and when you consider that a film shot in 35-millimetre can run up to 72 frames per second, and each frame becomes a separate .tiff file when the film is digitized, that's a lot of data. To be precise, VML's storage runs to half a billion gigabytes.

But Dr. Susan Lord, director of the VML and a professor at Queen's, says her team relies on the CAC as much for expertise and advice as it does for storage services. "We have learned a lot from working together," she says. Much of that learning comes from pooling knowledge to solve challenges that are unique to the digitization of art and media, especially in the space VML inhabits, where knowledge sovereignty is crucial. VML is the digitization and restoration node of the Archive/Counter Archive project, a collaboration among 14 Canadian universities that is dedicated to "activating and remediating audiovisual archives created by Indigenous Peoples (First Nations, Métis, Inuit), Black communities and People of Colour, women, LGBT2Q+ and immigrant communities."*

As part of Archive/Counter-Archive, VML is working with ARNAIT Video Productions, a women's video collective in Igloolik that for 30 years has been producing documentary and feature films. Queen's holds ARNAIT's archive of 311 videos, of which about one third have been digitized to date. Dr. Lord's team has worked closely with ARNAIT over several years to ensure that the artists' collective and their community maintain control over these valuable cultural materials. This includes control over how the videos are accessed. Some films are widely available on the ARNAIT web site, while others are only accessible to scholars through a research portal set up by Queen's. It also includes control over how the materials are categorized, described and indexed.

"We asked ARNAIT to tell us how they want the works described," Dr. Lord says. Through discussion they arrived at the following process: An affiliate of ARNAIT and a translator sit down with elders in the community and watch the digitized videos, which are sent from Queen's on an iPad with an external hard drive. The elders provide a wealth of descriptive information provoked by the content of the video, all of which is sent back to Queen's. "The data we get from this approach go way beyond things like geographic location," Dr. Lord says. "We receive things like hand-written notes, stories and even drawings." Data librarians and others at Queen's now need to find a way to keep the integrity of this knowledge and use it as metadata. This is ground-breaking work.

Dr. Susan Lord

Another new challenge is the question of how to preserve "digital-born" artwork. This could be artwork that originates in digital media, such as computer animation, or artwork that combines live and digital elements, such as installations or performance pieces that incorporate augmented or virtual reality. Figuring out how to approach these problems will take ongoing collaboration with computational experts at the CAC, data management experts, artists and scholars. "This is where we're going next and it's really unknown territory," says Dr. Lord. "But I think we've built the collaborative, cross-disciplinary relationships and knowledge-sharing that will be needed to figure this out."

**from the Archive/Counter-Archive web site*

All photos by: Alex Dekker

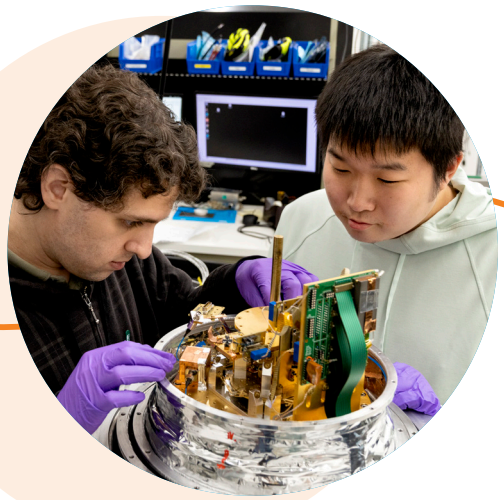
Left to right: Dr. Tyler Reynolds, Elspeth Cudmore (PhD student), Dr. Miriam Diamond (PI) Lucas Perna (PhD student), Zehua Sun (Undergrad student), Dr. Ziqing Hong (PI)

SOLVING THE MYSTERY OF DARK MATTER

Evidence for dark matter was first discovered when astronomers observed unexpected gravitational forces acting on distant cosmic objects. These gravitational effects implied the existence of matter that could not be seen because it does not emit radiation in any form, including light. Several decades later, this "dark" matter is still proving difficult to understand.

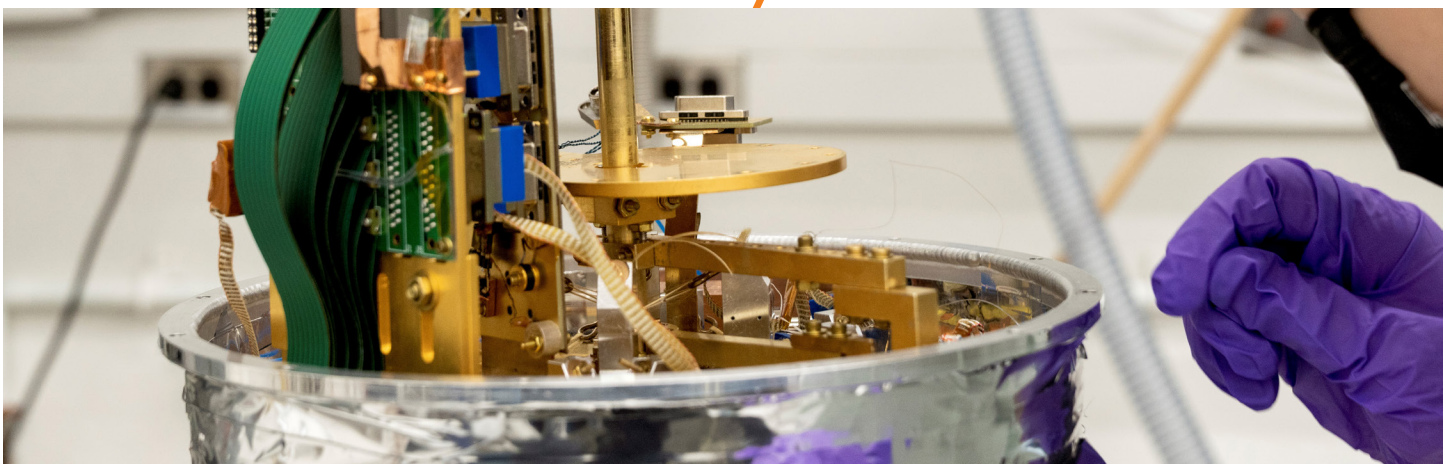
"We think dark matter — which makes up 25% of the matter in the universe — is a new subatomic particle that is not in the Standard Model of physics," says Dr. Miriam Diamond, an assistant professor at the University of Toronto. Dr. Diamond works on the Super Cryogenic Dark Matter Search (SuperCDMS) team, a collaborative international project that aims to find the dark matter particle. The Canadian SuperCDMS team and their partners in the U.S., India, France, Germany, Spain and the United Arab Emirates have banded together to undertake what Dr. Diamond calls "the greatest treasure hunt in human history."

"We think dark matter — **which makes up 25% of the matter in the universe** — is a new subatomic particle that's not in the Standard Model of physics,"



To find the dark matter particle, the project is building "the biggest and most sensitive detector we can afford," says Dr. Ziqing Hong, assistant professor at the University of Toronto. The detector, made of pure silicon and germanium crystals, is housed in a cryogenic chamber at a temperature close to absolute zero. The cryogenic chamber is located 2 kilometres below the earth's surface in the SNOLAB facility, which was built in a nickel mine near Sudbury for the purpose of neutrino detection.

In addition to operating the detection facility, Canada's role in SuperCDMS is focused on data acquisition. In this regard, a critical task for Dr. Diamond and her team is to filter out irrelevant signals ("noise") that could mask the signal produced by dark matter particles. Working deep underground acts to physically block background events such as cosmic rays or radioactive decay. But there are other kinds of noise, including noise caused by the experimental equipment, that must be identified and removed by precise, sophisticated computer algorithms. Creating, testing and refining these algorithms has been a major focus for Dr. Diamond and her team, which has quickly grown to comprise 7 graduate students, 4 postdoctoral fellows and 3 faculty members. In addition to filtering out noise, algorithms are used to convert raw signals into human-readable data that scientists can analyze.



Data storage and processing for this project requires hundreds of terabytes of disk space and hundreds of core-years of processing capacity. To avoid the risk that computer outages might disrupt the project schedule or result in data loss, storage and processing is being shared among Canada and three major centres in the U.S. About half the Canadian contribution to processing and storage has been supplied by SciNet, as part of a Digital Research Alliance of Canada allocation. SciNet is an Ontario DRI consortium led by the University of Toronto. We have been impressed with how stable the SciNet platform is and have really appreciated their flexibility when we have had to change our schedule because of things like COVID-19 outbreaks," Dr. Diamond says.

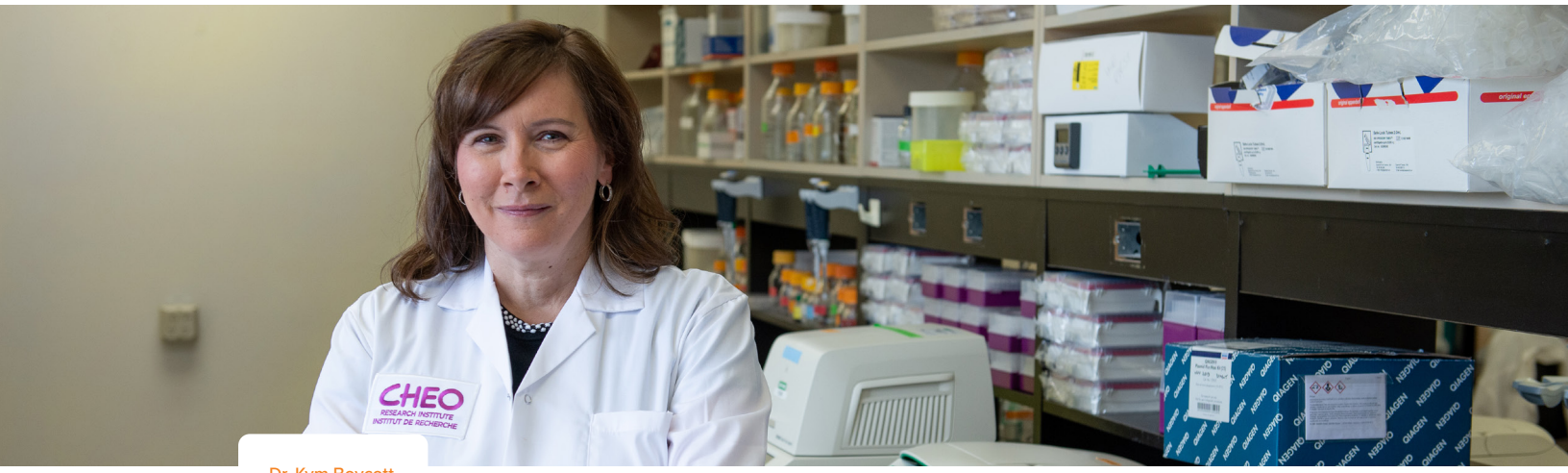
While SuperCDMS is about a year away from becoming fully operational, the work of building and testing the detectors and running simulations has already led to new learning about particle detection and other aspects of physics, Dr. Hong notes. The Canadian SuperCDMS team is supported by the McDonald Institute, NSERC and the Canadian Foundation for Innovation.

ADVANCING THE DIAGNOSIS OF RARE GENETIC DISEASES

Care4Rare Canada is a consortium of 20 Canadian clinical and research sites, led by Dr. Kym Boycott, a geneticist at the Children's Hospital of Eastern Ontario (CHEO) and senior scientist at CHEO Research Institute. Care4Rare is a 10-year research program which helps people — mainly children — who have rare genetic conditions, by identifying the precise DNA change or genetic variation responsible for the individual's symptoms. With this information, clinicians can make an accurate diagnosis that can help families understand their condition and access the resources they need.

While each rare disease is by itself uncommon, and therefore not well understood, more than 7,000 such diseases have been identified, together affecting at least 1 million Canadians. People with a rare disease spend on average 7 years consulting multiple specialists and undergoing many tests before their condition is diagnosed. However, many will remain undiagnosed following clinical investigations. Care4Rare grew out of a need for a personalized approach for these individuals. Over the past 10 years, Care4Rare has offered families with rare disease emerging genetic technologies which has led to the discovery of over 125 novel disease-causing genes.

To make these breakthroughs, the consortium had to overcome a major obstacle — siloing of data. "If you sequence any individual's genome, you might find several thousand variants," says Dr. Boycott.



Dr. Kym Boycott

Photos courtesy of Children's Hospital of Eastern Ontario

"How do you know which of these variants is the key to the disease? One of the tools at our disposal is to compare one family's data with that of others to see if they have similar symptoms and genetic profiles. But with only one or two cases of any given rare disease, this means going outside of your institution to find additional families."

Enter HPC4Health, a partnership between SickKids and the Princess Margaret Cancer Centre at University Health Network that provides clinical researchers with secure computing services, while

"While each rare disease is by itself uncommon, and therefore not well understood, more than 7,000 such diseases have been identified, together affecting at least 1 million Canadians."

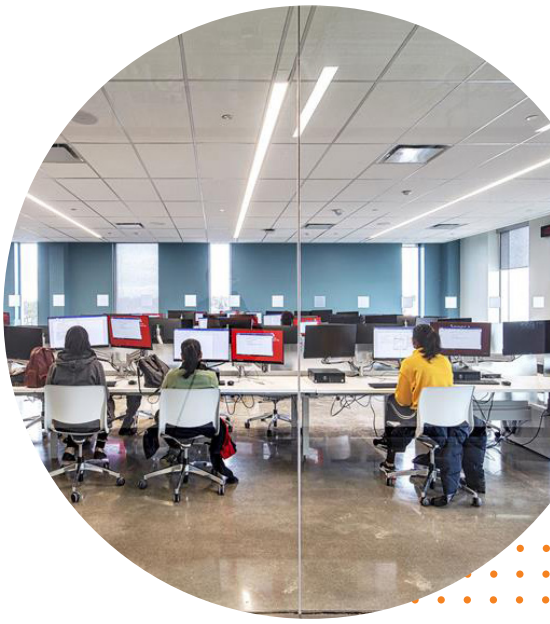
satisfying personal health information privacy requirements. HPC4Health supplies Care4Rare with storage for its extensive research data (anywhere from 54 to 425 GB per family), made accessible to consortium members through a platform called Genomics4RD, developed in collaboration with the Centre for Computational Medicine at SickKids.

Genomics4RD, launched in 2019, is the first pan-Canadian rare disease data repository and analysis platform. It is hosted on the advanced research computing cluster at HPC4Health and provides centralized access to harmonized data from thousands of rare disease families. Storage, authorization and access procedures were developed in collaboration with policy experts and stakeholders to ensure data stewardship.

Using Genomics4RD, approved researchers can compare cases to facilitate diagnoses. For example, Care4Rare recently identified a novel disease gene and linked it to a rare form of spastic paraplegia. The team was suspicious about the gene in a French-Canadian family and queried Genomics4RD. They discovered a second family with overlapping symptoms and variants in the same gene. This resulted in a diagnosis for both families and allowed the team to validate the connection between the gene and the disease as well as identify several additional families with this disease in other countries (Lemire et al., 2022; PMID: 34587489).

“Data-sharing is critical in rare disease research,” Dr. Boycott says. “A clinician might see only one person in their lifetime with a specific rare disease. Unless we share our knowledge with other researchers, we simply don’t have enough data to provide answers for these families.” Genomics4RD is working towards identifying additional ways to link this Canadian resource to international rare disease databases, continuing their global effort to solve all rare diseases.

AUTOMATING VIDEO CATEGORIZATION TO FIGHT MISINFORMATION



Photos courtesy of Seneca

Seneca Innovation has partnered with a privately-owned video technology company to develop a scalable machine learning algorithm to categorize videos.

Vubble, an Ontario-based company, serves its media and educational clients by collecting informational videos from online channels, curating them around topics the client has identified, then distributing them through a Vubble app on the client's web site. As part of the curation process, journalists assess the quality of the video and append a rich array of metadata, including a credibility rating assigned by the editorial team. Various AI tools are used to support this process.

In 2018, Vubble launched a project to automate video categorization in partnership with Dr. Vida Movahedi, a professor at Seneca College's School of Software Design and Data Science. "Vubble wanted to use AI to eliminate some of the easier, repetitive tasks so that the editorial team could focus on areas that require human creativity and judgment," Dr. Movahedi says.

Dr. Vida Movahedi

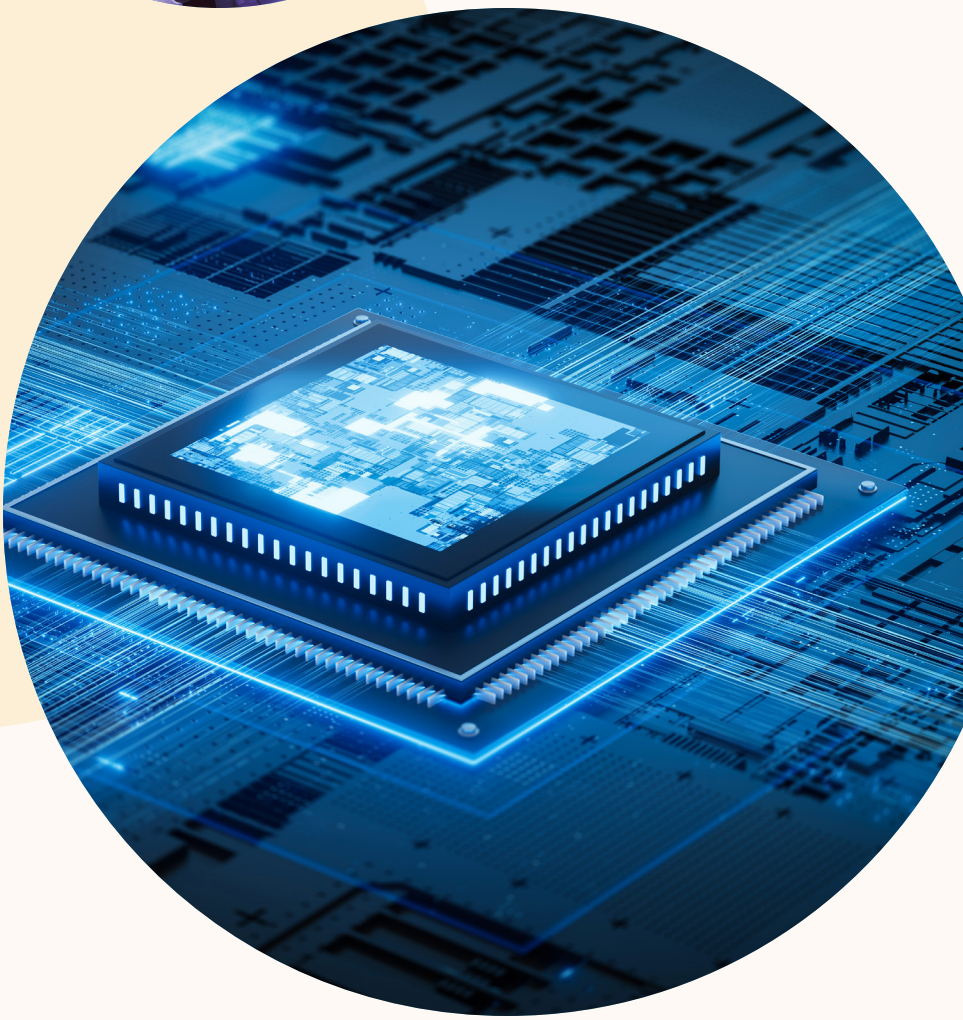
Dr. Movahedi and her students completed three projects between 2018 and 2021, using high-performance compute capacity provided by SOSCIP, a Canadian academic-industry consortium. The Seneca team designed an algorithm to predict one or more categories (up to 16 categories), such as Science, News, etc. for each video. Their initial data set included 80 sampled frames for each of 20,000 videos that had already been categorized and tagged by Vubble. Allowing any video to receive more than one category label increased the complexity of the problem.

Automatic video categorization is difficult. Changing backgrounds, viewpoints and levels of resolution interfere with accurate object recognition. Another key factor is context. Identifying the objects seen in a video is not sufficient for detecting the semantics. Consider an image of a person wearing a surgical mask. Is it an ad for a medical clinic, a public health announcement or an anti-masking message? Without context, it's difficult to say.

Dr. Movahedi and her team solved this problem by adding audio transcripts to the sample images in their dataset. Keywords in the audio file provided important context that increased the algorithm's accuracy.

In subsequent stages of the project, Dr. Movahedi's team expanded the number of video and audio cues and experimented with techniques such as stream learning, that allowed for continuous training of the categorizer as well as introduction of new tags, such as the COVID-19 pandemic. At each stage, the algorithm's category assignments were compared against the category assignments provided by Vubble's human journalists. Analyzing the mismatches allowed the team to fine-tune the algorithm. The Seneca team's work eventually became a key tool for Vubble — the Vubble Video Categorizer.

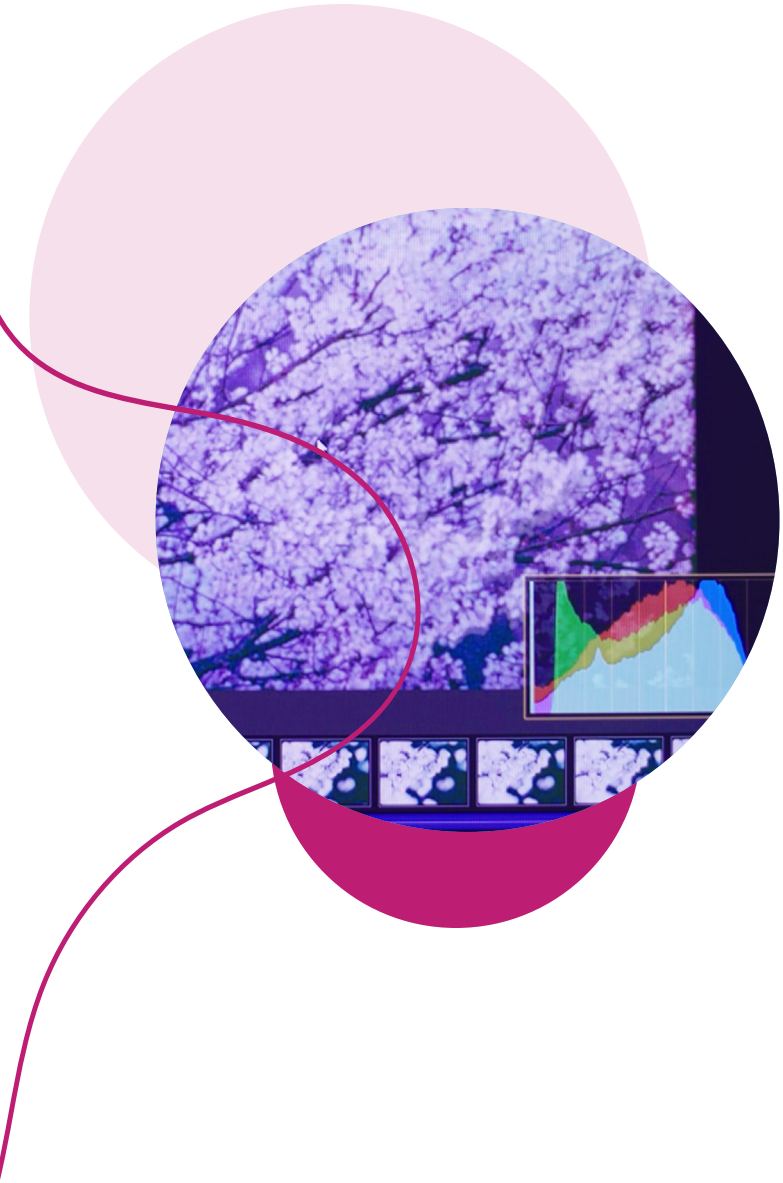
The video categorizer project is a wonderful example of the value of industry-academic collaboration. "Our students had a great learning experience," Dr. Movahedi says, and two went on to careers at Vubble. This project experience also helped to secure funding for an AI Research Centre at Seneca. "I'm delighted that we now have two years of funding to pursue more projects like this," Dr. Movahedi says.



STRATEGIC LEADERSHIP

PLANNING FOR THE FUTURE OF DIGITAL RESEARCH INFRASTRUCTURE

Strategic leadership is a key component of Compute Ontario's mandate. In 2021 – 22, much of our effort was focused on helping to shape new national funding and governance models that meet the expectations of both federal and provincial funders and the needs of Ontario's researchers. This work is ongoing. We also contributed significant time and staff expertise toward the Digital Research Alliance of Canada's multi-year funding proposal (MYFP) for 2023 – 25, and its supporting businesses cases.



With medical/life sciences researchers among the largest users of DRI resources, we were pleased to have the opportunity to contribute to the conversation about access to health data for research purposes. At the request of the Ministry of Health, we developed a research use case to answer the question, "What does the future state of data-driven research in an integrated health data ecosystem in Ontario look like?" This project took a researcher-centric view of how the ecosystem can best facilitate researcher access to data, while maintaining privacy and security for individuals' personal health information. Interviews with 36 health and research leaders, and a jurisdictional review, helped us to gain an understanding of barriers, facilitators and best practices related to health research data. The full report with our recommendations has been submitted to the Ministry of Health and will be released publicly next year.

At the same time, Compute Ontario's Technical Leadership Advisory Council (TLAC) has launched several important initiatives that will inform the Province's strategic investment in DRI resources. TLAC is conducting research to build out a more complete picture of DRI resources in Ontario, by documenting small- and medium-sized resources at the local/regional level that are not funded through the Alliance. TLAC is also working to quantify DRI needs for researchers at Ontario universities and colleges in terms of data processing, storage, cybersecurity, commercial cloud, energy consumption and HQP.

Through a new cloud pilot project, TLAC is exploring the role cloud services might play in a DRI strategy for Ontario. The pilot, launched in the fall of 2022, will help illuminate the true costs and benefits of using the cloud, while building participants' technical expertise. Compute Ontario has secured $110,000 worth of cloud credits for the pilot from several major vendors, who are also providing orientation and training for pilot participants.

TECHNICAL LEADERSHIP ADVISORY COUNCIL (TLAC)

- Danny Gruner, SciNet
- Francis Jeanson, Ontario Brain Institute (OBI)
- Fei Mao, SOSCIP
- John Morton, SHARCNET
- Lincoln Stein, Ontario Institute for Cancer Research (OICR)
- Amaz Taufique, Scholars Portal

SPOTLIGHT ON RESEARCH DATA MANAGEMENT (RDM)

THE RESEARCH DATA LIFE CYCLE

Good data management practices are essential to ensure the dependability, accuracy, and validity of the research data that underpin research results and published conclusions. Further, good data management helps make data FAIR: Findable, Accessible, Interoperable, and Reusable, an internationally recognized set of principles that increase the value of research data for both current and future research. Faced with an ever-growing volume of research data, applying practices that support FAIR principles helps researchers overcome challenges and maximize opportunities associated with creating, using, and re-using research data.

THE NATIONAL RDM LANDSCAPE

Recognizing the increasing need for thoughtful and rigorous RDM, the Tri-Agencies (NSERC, SSHRC, and CIHR) have launched an RDM Policy that requires all post-secondary institutions and hospitals eligible for tri-agency funds to have an institutional RDM strategy in place by March 2023. Some funding opportunities will also require data management plans to be included with grant applications. Starting in 2024, grant recipients will be required to deposit all digital data, metadata and code that supports their research conclusions into a recognized data repository. A number of Digital Research Alliance-supported initiatives are underway to help both institutions and researchers at each stage of the data life cycle:

Institutional RDM Strategy Development Template
Designed to guide institutions in the development of tailored institutional RDM strategies in compliance with Tri-Agency RDM requirements

Data Management Planning (DMP) Assistant
an open, national, bilingual data management planning tool that supports researchers in the creation of high-quality data management plans

Two national, bilingual, multidisciplinary repository options:
The Federated Research Data Repository (FRDR) — a scalable, federated repository with big data capabilities

Borealis, the Canadian Dataverse Repository — a flexible, easy-to-use repository, widely adopted by academic libraries and research institutions across Canada

Lunaris
A national data discovery portal that opens a wide range of Canadian research data to both domestic and international audiences

The Digital Research Alliance of Canada Data Champions Pilot
18 teams and individuals across Canada, funded by the Alliance to develop activities at the local, regional and/or national level to advance awareness, understanding, development and adoption of RDM tools, best practices and resources

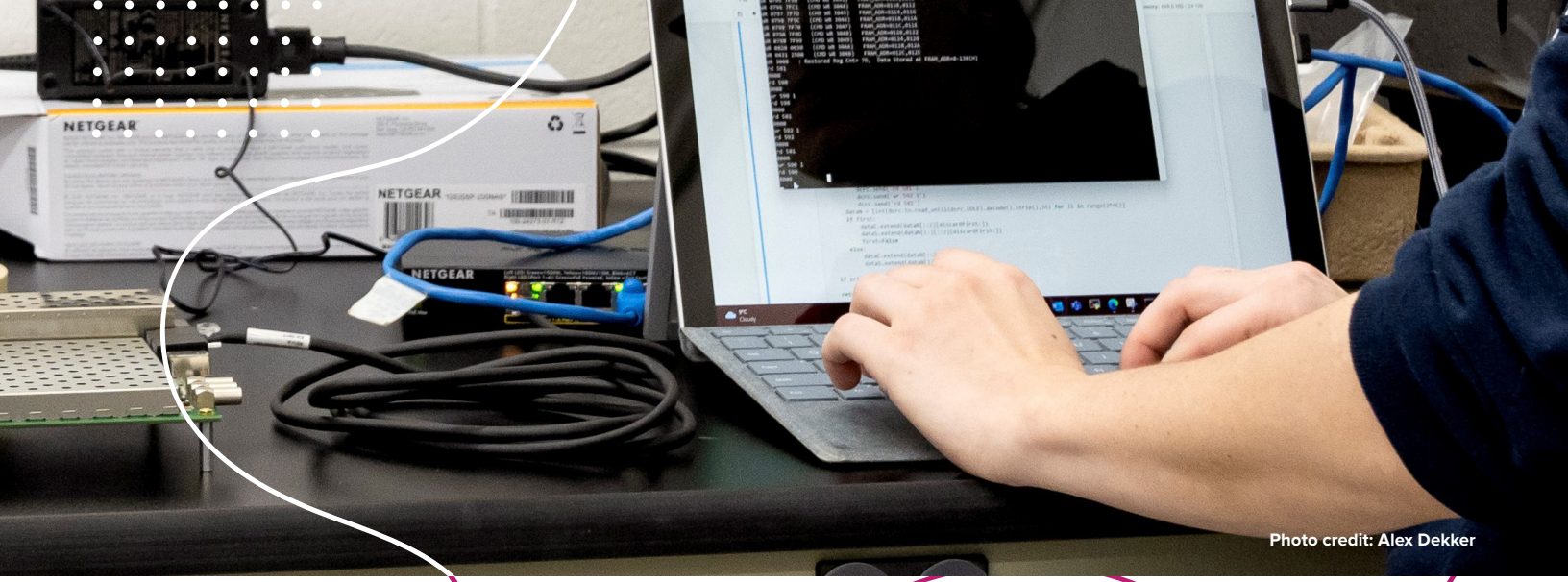


Photo credit: Alex Dekker

COMPUTE ONTARIO'S RDM MANDATE

In the spring of 2022, the Ministry of Colleges and Universities added RDM to Compute Ontario's mandate. In June we created a new role to bring focused attention and expertise to this area. As our new Director, Data Strategy and Services, Jeff Moon is focused on maintaining, expanding and advancing data management awareness, capacity, and adoption in Ontario and across Canada. With Jeff's leadership, Compute Ontario has begun to establish a network of relationships and collaborations with key RDM stakeholders and move this objective forward. Key activities have included:

- Providing an introduction to RDM at the 2022 Compute Ontario Summer School
- Facilitating funding for translation of an open education textbook on RDM
- Contributing expertise to help build RDM-related business cases for the Digital Research Alliance of Canada's multi-year funding proposal
- Connecting with academic institutions to discuss how Compute Ontario can help advance local RDM initiatives
- Maintaining and growing connections with Alliance Expert Groups, including being a member of the National Training Expert Group
- Forging connections and cooperation with Ontario Colleges through work with the Heads of Applied Research (HAR) group
- Maintaining and growing connections with other allied organizations such as CUCCIO, the Canadian Association of Research Libraries (CARL), the Ontario Data Community (ODC), and CANARIE
- Connecting and forging relationships with Ontario HPC sites and national ARC regions
- Facilitating a Compute Ontario RDM Funding call designed to advance innovative and sustainable approaches to solving RDM challenges with the goal of making research data FAIR (Findable, Accessible, Interoperable, and Reusable). *See pages 19 – 20 for project details.*

ECOSYSTEM SUPPORT

SUPPORTING A HEALTHY DRI ECOSYSTEM

In 2021 – 22, 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 and support for training.

FINANCIAL SERVICES

This year we supported the consortia by coordinating funding agreements, approving budgets, flowing funds and supporting the Major Science Initiatives (MSI) audit. In addition to these annual activities, we contributed significant time to ongoing work to develop a new national funding model, budget for 2023 – 25 and multi-year funding proposal (MYFP). We also worked with the Ministry of Colleges and Universities to secure $12 million in additional funding to secure the jobs of Ontario ARC staff over the next two years.

ECOSYSTEM COORDINATION

Working with our Technical Leadership Advisory Council (TLAC), Compute Ontario provided valuable coordination and support for the Digital Research Alliance of Canada's Pan-Canadian AI Project, which provided $65 million in funding for new AI sites in Ontario, Quebec and Alberta. TLAC developed and ran the site selection process for Ontario, which resulted in SciNet being selected. TLAC also developed and ran a process for allocating remaining (MSI) funds.

BUILDING KNOWLEDGE THROUGH PROJECTS

TLAC developed a cloud pilot project which will inform the province's cloud strategy while building technical expertise in cloud services. The pilot launched in September and will provide orientation and training to participants, as well as $110,000 worth of cloud services. Data will be gathered to determine the true costs, benefits and risks of using cloud services as compared with operating physical computing clusters.

Compute Ontario also issued its second Research Data Management project call, which award approximately $380,000 to 7 projects. The projects will advance innovative and sustainable approaches to solving RDM challenges — particularly those associated with ARC and RS — with the goal of making research data FAIR (Findable, Accessible, Interoperable, and Reusable). Projects are to be completed by March 31, 2023.

2022 RDM AND TRAINING PROJECTS

Dr. Michelle Edwards (University of Guelph), Dr. Lucas Alcantara (University of Guelph), Dr. Carly Huitema (University of Guelph)	This project will offer 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 will use chimeric antigen receptor T cells (CAR-T) therapy data as a case study with the goal of tackling the following three research questions: How to generate and curate useful FAIR synthetic CAR-T data at cohort level to represent a complex patient population, while preserving privacy in synthetic data and training federated learning; 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)	This project will integrate ODESI, a collection of statistical data maintained by the Ontario Council of University Libraries (OCUL), with Borealis, the national research data repository. To create 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 aims to enhance overall learning and RDM development within the Colleges by linking College Research Administrators and staff within Research offices to College Librarians and resources that would support the development of institutional strategies and capacity building. Further, this project supports building college specific resources that could be easily accessed by all researchers within the college system and further sustained through the Heads of Applied Research Data Subcommittee in Ontario.
Dr. Barbara Fallon (University of Toronto), Dr. Dale Turner (University of Toronto), Dylanne Dearborn (University of Toronto)	This project will embark on a series of engagements to provide the University of Toronto research community with access to resources, support and training for RDM that reflects the diverse scholarship at the University of Toronto. The project will reflect their commitment to respecting Indigenous data sovereignty, recognition of Indigenous knowledge systems and the need for co-development of research partnerships, processes and tools. The project will also raise awareness of existing resources, identify gaps and the development of processes and tools that are consistent with community-driven principles (such as Ownership, Control, Access and Possession (OCAP) and Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE)) is central to this project.

Art Rhyno (University of Windsor), Dr. Berenica Vejvoda (University of Windsor), Dr. Paul Preney (University of Windsor)	This project proposes 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 involves the development of an interactive data visualization designed to improve discoverability and reusability of polar metadata and data at the Polar Data Catalogue2. It will produce a direct increase in the utilization of polar metadata and data, thereby enhancing the activities of researchers and policy makers, as well as many other stakeholders and interested parties; and it will provide reusable deliverables intended to serve as adaptable examples for achieving data harvest and visualization capabilities required to support research computing and software efforts.

This initiative builds on the successful slate of projects completed in 2021, which meaningfully contributed to the dissemination of data management best practices as well as improving access to specific datasets.

OUTCOMES OF 2021 RDM AND TRAINING PROJECTS

Dr. Kulamakan Kulasegaram (University of Toronto / UHN); Dr. Lawrence Grierson (McMaster University)	Created three free training modules on data-sharing best practices health care/medicine which were officially launched in January 2022.
Drs. Kelly Cobey & David Moher (Ottawa Hospital Research Institute); Kevin Holmes (Algonquin College)	Created four free training modules on good data handling practices for researchers, distributed via journalologytraining.ca. Learners who score 80% or higher on all four modules receive a micro-credential from Ottawa Hospital and Compute Ontario. As of September 30, 2022 more than 60 certificates had been granted.
Dr. Michael Brudno (University of Toronto); Dr. Ken Evans (Indoc Research); Dr. Charles Victor (ICES); Dr. Lisa Strug (University of Toronto)	Developed and trialed encrypted Global Unique Identifiers to allow unique, auditable, and secure linking of data elements.
Drs. Philippe Van Cappellen & Rodney Smith (University of Waterloo); Mary Kruk (DataStream); Kelly Stathis & Erin Clary (Portage Network); Nancy Goucher (University of Waterloo); Dr. Jacob de Boer (Free University of Amsterdam)	Developed an open-source metadata template to standardize environmental microplastics Research Data Management, advancing the comparative analysis of datasets collected in different environments across the world.
Alan Darnell (University of Toronto); Kate Davis (Scholars Portal)	Developed a methodology for georeferencing the 1:50,000 series of Canadian topographic maps and making these available to researchers on the Scholars GeoPortal and the Dataverse Canada platform as a unified collection with support for searching, online viewing, and bulk downloading.
Dr. Mark Campbell (University of Toronto); University of Toronto's Flourish Research Initiative: Community Engaged Arts as a Method For Social Wellness; Ajah, the Critical Digital Methods Institute; UTSC's Digital Scholarship Unit	Created a data repository to share research on the impact of arts participation on health and wellness, with special emphasis on community-based activities.

BUILDING SKILLS THROUGH TRAINING COORDINATION

This year's Compute Ontario Summer School ran from May 30 to August 5, 2022. Twenty-one free courses totalling 94 hours of instruction time were delivered by advanced research computing and RDM specialists. A total of 525 unique participants attended the summer school to learn the skills and technologies required for compute and data-intensive research.

This year the summer school was a truly integrated experience for participants. The consortia developed one combined schedule of workshops, and registration and delivery were coordinated through a single Moodle site hosted on a shared server. With support from Compute Ontario, the consortia also adopted a standard post-course survey for all courses. Demographic and other data gathered through the survey will help to inform a coordinated training strategy, which will be developed by the Compute Ontario Training Advisory Committee launched in the fall of 2022.

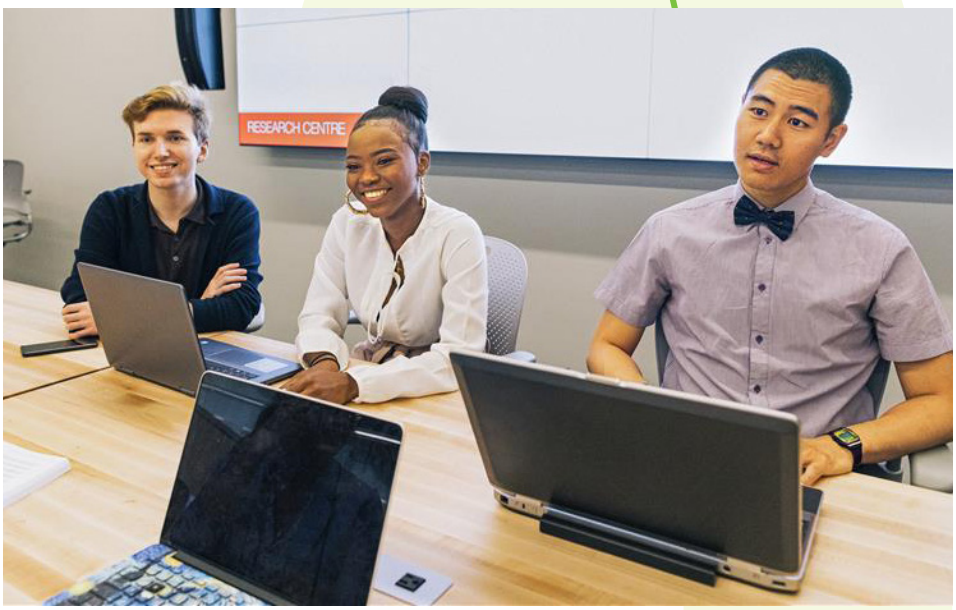


Photo courtesy of Seneca

TRAINING ADVISORY COMMITTEE

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

Amal Khalil
Queen's University/CAC

COMMUNITY BUILDING

Photo courtesy of Seneca

BUILDING COMMUNITY ACROSS THE ECOSYSTEM

Building connection and community is a key aspect of Compute Ontario's mandate, both for its own sake and because it is the foundation for achieving other goals. Common sense tells us that coordination makes us more efficient and more effective, whether in the realm of training or service delivery. It's common knowledge that collaboration is essential to generate innovation. Common ground enables us to speak with a unified voice to ensure that the ecosystem meets the needs of researchers.

In 2021 – 22, we intensified our efforts to strengthen connections across the DRI community, guided by a thoughtful stakeholder engagement strategy, and supported by a new software tool that helps us track activities and progress.

BROADENING AND DEEPENING OUR ACADEMIC RELATIONSHIPS

Ontario's universities are key partners in the ecosystem, both as DRI users and as hosts and providers of DRI resources. As such, they have a very real stake in the new funding model being developed by the Digital Research Alliance of Canada. Compute Ontario is taking a systematic approach to keeping our university stakeholders apprised as Alliance plans develop. We present an update to the Ontario Council of University Researchers every quarter, along with email updates as needed between presentations. Frequent briefings are provided to a core group of VPRs whose institutions are most affected by DRI funding issues, and we have collaborated with this group to jointly advocate on behalf of researcher needs. We have also solicited delegates from each university to form a Research Advisory Council for Compute Ontario as an ongoing forum to ensure the researcher voice is heard in everything we do.

Recognizing that many Ontario colleges are increasing their focus on areas such as AI, data science and machine learning, we undertook an informal assessment this year to begin to understand the DRI needs of college researchers and their students. We intend to build on this initial outreach by setting up a College Advisory Council in the coming year.

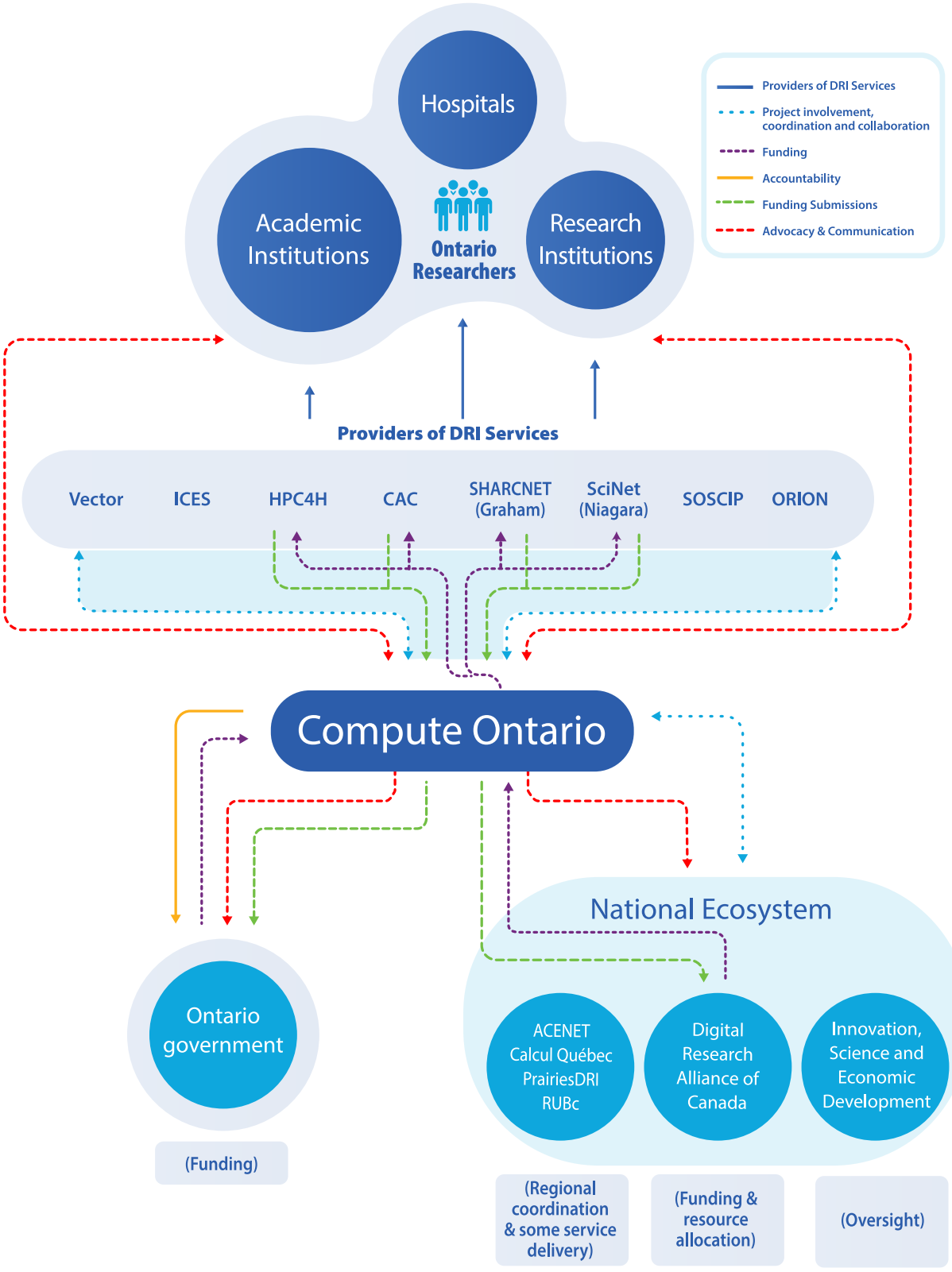
CONNECTING AND COLLABORATING MORE BROADLY

In 2021 – 22, Compute Ontario began to build connections that will help us to integrate Research Data Management into the DRI ecosystem. Initial meetings with data librarians and RDM experts are being held to establish common ground and begin to identify opportunities for collaboration and coordination. By the end of the fiscal year, we began providing expertise to universities and colleges to support development of their RDM strategies.

We collaborated intensively with our regional partners this year, supporting each other in navigating the transition from Compute Canada to the Alliance. We worked together to represent researcher interests throughout the Alliance's financial and service delivery planning process. We also found time for smaller projects, such as Powering Innovation in Mining with Advanced Computing, an industry workshop that that we sponsored jointly with SOSCIP, PINQ2, ACENET and Calcul Quebec.

Finally, we took the first step toward solidifying relationships with research institutes and hospitals, through a presentation to the Ontario Hospital Association's Health Research and Innovation Committee. In the coming year, we will be continuing this conversation.

COMMUNITY BUILDING



THE RESEARCH 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 and colleges

Ontario Council of University Researchers (OCUR)

Research hospitals

OHA Health Research and Innovation Committee

Individual researchers and students

THE DRI COMMUNITY

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, which collaborates to enhance research supports and create rich learning environments for Ontario's diverse and growing university population.

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

SOSCIP, a Canadian academic-industry consortium dedicated to high-performance computing.

Nationally, the **Digital Research Alliance of Canada** was launched in 2020 to coordinate and fund activities in advanced research computing (ARC), research data management (RDM) and research software (RS), working collaboratively with stakeholders across the country, including Compute Ontario; its regional counterparts **ACENET** (New Brunswick, Nova Scotia, PEI and Newfoundland and Labrador), **Calcul Québec, PrairiesDRI and Research Universities BC.**

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 especially thank outgoing Board Chair Mark Daley for his dedicated and graceful leadership, and we warmly congratulate Warren Helland, who was elected Chair on September 30, 2022.

We would like to acknowledge and thank Board Directors Selim Teja and Dereck Whitmell, whose terms ended this year, and welcome on board two new Directors: Eric Broda and Leah Cowen.

MARK DALEY
Board Chair

WARREN HELLAND
Chair Governance & Nominations Committee

SYLVAIN CHARBONNEAU
Secretary

CHARMAINE DEAN
Director

SHANNON MACDONALD
Chair Audit & Resources Committee

CAROLYN MCGREGOR
Chair Industry Engagement Committee

ATEFEH MASHATAN
Director

SELIM TEJA
Director

SUSAN URSEL
Director

DERECK WHITMELL
Director

Compute Ontario

computeontario.ca
info@computeontario.ca