

SOSCIP Consortium on the Importance of Driving Industry Innovation with Canada's New Digital Research Infrastructure

As a special member of the DRI ecosystem, the SOSCIP Consortium is pleased to provide a submission on Canada's Future DRI Ecosystem. Much of the content of this white paper is directly related to SOSCIP's unique mission, the space the Consortium has established in the ecosystem, and the many successful outcomes and impacts it has generated. This context for application of Advanced Research Computing is needed in order to derive meaning from the current state of DRI, its challenges, and its future opportunities.

Launched in 2012, SOSCIP is Canada's leading R&D consortium harnessing Advanced Research Computing to drive industry innovation. SOSCIP's mission remains unchanged – to bring together industrial partners and academic researchers and provide them with sophisticated advanced computing technologies and expertise to solve social, technical and business challenges across sectors and to drive economic growth. Developed as a collaboration between Ontario post-secondary institutions and many small- and medium-sized enterprises (SMEs) as well as large companies across the province, SOSCIP has catalyzed the creation of a collaborative ecosystem that has accelerated economic impact, job creation, skills retention and the growth of Ontario-based firms. The SOSCIP ecosystem has a proven track record for success in delivering innovative outcomes that will contribute to a skilled and innovative Canadian economy.

SOSCIP directly supports Canadian companies as they tap into the data science, AI and broad computational expertise at our world-leading universities and colleges. As a result of SOSCIP's work, more than 750 HQP have gained valuable data science skills solving real-world industrial challenges. Over 60% of students engaged in SOSCIP projects who have completed their studies have been hired into high-tech Canadian companies. SOSCIP has engaged more than 250 industry personnel across more than 125 companies and has fostered the creation of or maintained over 800 positions in the past 8 years.

Put another way, SOSCIP has been Canada's only research consortium bringing advanced computing directly to industry in order to build partnerships to foster industry adoption of transformative technology. We build and facilitate partnerships between Canadian industry and academia to solve industry challenges and align the tools and funding needed to ensure success. Building a strong future for Canadians, the SOSCIP collaborative ecosystem is key to the fabric of economic development in Ontario. In short, SOSCIP accelerates the R&D required for commercialization, allowing innovative made-in-Canada technologies to move to market quickly.

Current Issues / Current State

Recent investments in new compute infrastructure by the federal and provincial governments have provided a timely investment in the computing resources necessary for Canada's academic community. Importantly, this infrastructure is currently functioning near capacity and is dedicated to fundamental research and is not generally available to support industrial research partnerships and applied research projects.

For this reason, alternative sources of support have been sought to enable SOSCIP to provide the essential dedicated advanced compute infrastructure for industry-academic partnerships. The collaborative research done by SOSCIP research project teams are as wide and varied as the researchers

themselves who can come from any academic or post secondary member institution. Our users employ a wide variety of software, mostly open-sourced, some commercial, on both HPC type machines.

Accessing DRI Tools, Services and Resources

Access to DRI tools, services and resources needed for the industry-academic research we support typically depends upon the area of research. For example, large companies who have project work in aerodynamics or materials science usually need access to an HPC-type machine but do not have large enough capacity for their growing demand. On the other hand, SMEs who are in an AI field generally do not have access to high-end GPUs. Commercial AI R&D software/services like Amazon Sagemaker and Google Datalab are not available on the current infrastructure and would be too expensive for SMEs to use intensively on commercial cloud. This appears to be a growing need that cannot be currently met.

Challenges to accessing and using DRI

There are a number of significant challenges that currently exist using the current DRI tools. One challenge is that of out-of-service situations typically due to network, power and cooling that can occur and sometimes can take a few days to get resolved. For this reason, the current infrastructure is not reliable enough for a company to run any 24/7 client-facing service such as web portal, public database, real time request or analysis. Technical support is generally not available other than working hours which is usually business hours on weekdays, and depending upon the client time zone, support may not be available.

Another significant challenge for SME clients is that they often cannot afford highly-skilled development personnel to smoothly migrate their work from a research-oriented infrastructure to a publicly available infrastructure without proper training and hands-on support. Obtaining sufficient allocation on existing infrastructure is another challenge for academic-company collaborations for industry-driven/commercially driven projects. Recently, companies have begun more often seeking and specifically asking for secured cloud-based infrastructure. Research-oriented HPC-type machines appear to not be secure enough for them.

Optimal Future State of DRI

Based upon SOSICIP's experience, our vision for a cohesive Canadian DRI ecosystem that would fulfill industry-academic research needs would qualitatively have the following features: be large capacity to ensure all kinds of projects can be supported regardless of discipline, high reliability so that users can be assured of steady consistent operations at any time, on any day of the week throughout the year, high security to gain confidence of users with sensitive and confidential personal or company data, and finally to allow for smooth migration to commercial platforms as needed by industry partners.

In the future, we envision using computing power beyond CPU/GPU/FPGA including dedicated AI accelerator ASICs like TPU, NPU, WSE, etc, and, of course, quantum computing that is just beginning to become more widely available.

One of the greatest challenges in the ARC ecosystem remains the significant costs associated with purchase, operation and maintenance of computing platforms. Similarly, companies that need access to significant compute infrastructure to advance their ideas and products face a significant barrier-to-entry when first exploring the use of new tools and technologies like data science and AI. As SOSICIP's mandate

is to drive the adoption of transformative technologies, it is essential that access to the compute infrastructure and tools are not cost-prohibitive to companies. Due to the costs associated with the computational resources, it is anticipated that periodic government support will be required to refresh SOSCIP's infrastructure.

Bridging the Gap with Tools, Services and Resources

In order to bridge the gap between the current state of DRI and the future improved state, services and resources must be planned specifically dedicated to industry-academic partnerships in mind. It is critical that these remain separate from those dedicated to pure discovery-based research. While the physical equipment may well be set up as an integrated contributed system arrangement in order to maximize functional benefits and optimize operating costs, nonetheless, a separate dedicated technical operations and project development teams must be maintained for industry-led industry-academic project work. This stems from the fact that priorities of companies are apart from those of the research ecosystem. A dedicated approach to resources committed exclusively to companies gives them the priority they need to tap in to project collaborations with researchers at the speed of business.

Going forward, in order to ensure the continued growth of industry-academic partnerships and the substantial business, social and job growth benefits, it is recommended that DRI investments must also include support for industry-led Advanced Research Computing infrastructure.

Furthermore, dedicated platforms to support industry with high-level security are increasingly being demanded. Next, software and service must be provided to help migrate to commercial platforms after projects are concluded otherwise project may never get underway since the results are stranded on the system and would not be able to commercially benefit the company. Finally, higher reliability and redundant cooling systems and power sourcing are essential in order to minimize system downtime.

Role of NDRIO

NDRIO has a unique and special role to play nationally to ensure that the benefits of DRI are realized for all of Canada to ensure that the maximum benefits and impacts are derived not only through fundamental research but also through industry-led project research given the limited funding that the federal government has provided.

With SOSCIP's years of proven experience in growing successes and generating impacts, the time has never been better for NDRIO to support engagement and connection with industry across Canada. Now that the SOSCIP formula for success has been well proven over years of increasing engagement with industry, it is an opportune time for NDRIO to recognize that the same beneficial impacts would also be realized across Canada if it were to support a similar national initiative. The benefits to the nation would be substantial to academic researchers, students, graduates and companies. The same opportunity for company access to Advanced Research Computing technologies for industry-academic projects should be opened to all parts of Canada. It is the next logical step in connecting the benefits of the DRI to industry, business and to society in general.

SOSCIP is uniquely positioned to directly support Canadian companies as they tap into the data science, AI and broad computational expertise at our world-leading universities and colleges. Building on the SOSCIP platform, we can deploy a highly efficient partnership model for private-sector innovation that is unique in Canada and the world. This innovative model can address Canadian firms' data science needs

and ensure the long-term growth and sustainability of programs that are developing skilled talent, scaling firms, supporting emerging industries and fuelling economic development. Industry-innovation on ARC systems can work in concert with academic research and is fundamental to talent development and skills training for Canadian firms.