Towards Research Platform as a Service (RPaaS) initiatives

Submitted to NDRIO in response to the call for Canadian DRI needs assessment white papers
https://engagedri.ca/ndrio-call-for-white-papers-on-canadas-future-dri-ecosystem/

Authors:

Carol Gauthier, Software Architect, Developer and Project Manager, Centre de calcul scientifique, Université de Sherbrooke

Pierre-Étienne Jacques, Associate Professor, Département de biologie, Université de Sherbrooke

Alain Veilleux, Director, Centre de calcul scientifique, Université de Sherbrooke

Guillaume Bourque, Professor, Department of Human Genetics, McGill University

\+ *List of supporters at the end of the document*

Summary

Research Platforms have become a significant part of the Canadian Digital Research Infrastructure (DRI), and their future will be essential for the Canadian research competitiveness at the international level. As part of its Research Software mandate, NDRIO is now well suited to pursue this work and bring it to the next level.

In this document, we propose a framework to help NDRIO meet this objective, which should include a stable and reliable ARC underlying infrastructures, mid-level access tools to HPC systems, a central team dedicated to support Research Platforms, increase funding to support the development and maintenance of Research Platforms, and a long term commitment to major Research Platform initiatives to bring them to the level of an official Service (RPaaS).

From HPC to Research Platforms and RPaaS

High Performance Computing (HPC) has evolved from a few dozen of very expensive supercomputers world wide in the 70s and 80s, to nowadays 100s of giant clusters deserving millions of researchers. Traditionally, researchers using HPC had a very high level of computer skills, and were coming from Natural Sciences and Engineering. In the last two decades, researchers from other fields (especially Life Sciences and Humanities) have been steadily increasing and could represent the majority of users in the near future. As the democratisation of HPC relentlessly advances, the average level of computer skills of the newcomers tends to decrease, as researchers want to concentrate on doing their actual science rather than developing the tools needed for their work. This trend is accelerating and it is imperative that NDRIO looks at this rapidly evolving landscape in order to provide the necessary tools to these large numbers of Canadian researchers to stay competitive.

Following that trend, researchers, including those from traditional HPC fields, have been more and more relying on existing softwares and high level tools to do their work. The advent of fast Internet, virtualization and cloud technologies has led to the development of high level pre-packaged set of tools that can be used via Web portals, often without the need to directly connect to HPC resources using traditional low level access (such as SSH). Over time, the term Research Platform has been coined to identify these sets of tools and modes of access.

Research Platforms are often the result of collaborations between research groups who want to share data, results, and technology with other Canadian and international groups. Support for the development of such platforms is often limited, as it is difficult to obtain funding for software if your main field of research is not Software Engineering or Computer Sciences. Even more difficult is finding funding to maintain these platforms on the long run to bring them at the level of a supported and reliable service. Hence we used the term RPaaS (Research Platform as a Service) in the title of our White Paper to deliberately highlight the concept of "Service" by these platforms. This addition is important because it implies stability, reliability, accessibility, security, and most importantly sustainability of the platforms.

As the Research Platforms become a significant part of the work of a growing number of researchers, their long term reliability and availability will be critical for the growth and quality of research in Canada. We clearly envision RPaaS as being a major part of HPC access and therefore an essential ingredient of the future Canadian DRI. In this regard, we believe that NDRIO can best tackle this challenge as part of its Research Software mandate.

Current status of Research Platforms in the Canadian DRI

Over the years, the physical infrastructure of the Canadian DRI, including HPC and other components globally referred to as Advanced Research Computing (ARC), has benefitted central funding from the Federal Government since the mid 2000s, with the clear objectives of providing Canadian researchers with World class, stable, reliable, and sustainable ARC resources. This gave birth to Compute Canada and lays the grounds for a true national infrastructure which provides HPC services with a high level of stability, reliability and accessibility. This was an essential foundation for the future of the Canadian DRI. Based on the announced directions, we are hopeful that NDRIO will continue to advance on this long journey.

In the meantime, research software development was still at an earlier stage of consolidation in Canada up until 2007, when CANARIE started to provide funding for software development through their Research Software Program. This program helped promote good software engineering practices, build a network of Research Software developers in Canada, and notably encourages software reusability. Also, with their direct support of various Research Platforms over the years, Canarie clearly filled a void and helped thousands of researchers and students concentrate on their research work rather than developing software.

As an example, we list below some of the many Research Platforms that CANARIE have funded and supported and that are still up and running at this moment, providing various levels of services to researchers of various domains, decision makers and even the general public:

- CANFAR (Astronomy) : https://www.canfar.net/
- CBRAIN (Medecine) : http://www.cbrain.ca/
- GenAP (Genomics): https://genap.ca/
- HEP (Particle-physics) : http://www.hepnetcanada.ca/
- iReceptor (Immunology) : http://ireceptor.org/
- Motus (Ecology) : *https://motus.org/*

Among the dozens of Canadian based Research Platforms, some have most of the ingredients to be qualified as a RPaaS. Yet, despite their advancements, recognition and usefulness, many have an unclear future due to long term funding uncertainties.

From Research Platform initiatives to RPaaS

1) Birth of Research Platforms:

Most, if not all, of the Research Platforms arise from direct needs from the researchers teams who initiate their developments. We believe this will remain the case in the future, because as the saying goes "Necessity is the mother of invention". Innovative platforms will necessarily come from the researcher teams themselves while tackling and generalizing their own problems. At first the research teams explore the platform's features and feasibility by experimenting, testing, building proofs of concept, and then even deploying on a small scale.

What NDRIO can do: At this stage, the role of NDRIO is to provide a set of secure and reliable mid-level access tools (APIs, SSO, etc..) that will enable easy interconnections and communications between the Platforms and the underlying HPC infrastructures, as well as central team of HQP dedicated to the support of all the Research Platforms relying on the DRI.

2) Development of Research Platforms:

The next phase, consisting of the development of Research Platforms offering services to a larger users base, implies a more serious and planned development, as well as a more substantial funding than what the field-oriented Grant programs are usually willing to offer. CANARIE played an essential role by providing such funding since the late 2000s.

What NDRIO can do: By taking over the Research Software mandate from CANARIE, it is very important that NDRIO continues this work in supporting the development of new and existing Research Platforms. There are certainly improvements to be made on this way forward, and we would like to take the opportunity to make some suggestions on how NDRIO could deliver on this goal:

1. Organize calls to fund new Research Platforms led by researchers for development cycles of one to two years, followed by a maintenance period that may include further improvements in a similar way as CANARIE did. We believe that this model has proved to be productive and effective over the years.

2. Organize calls to fund major improvements of Research Platforms already adopted by a community of users to expand their users base, again following similar calls by CANARIE.
3. Allocate more money to fund these initiatives than what was previously available. This of course comes as no surprise, but it is clear that the need is there and we are convinced that there are many good quality projects just waiting to have the necessary funds to kick-off.
4. Make sure the development teams include scientists of the fields, software professionals and HPC specialists (most likely from existing local data centers) with a strong knowledge of the Canadian DRI. At the same time encouraging more synergy between the existing HPC staff in data centers and the research teams driving these initiatives.
5. Create a Research Software central support team to oversee the evolution of the various platforms with the objectives of promoting good practices, encouraging collaborations and software components reuse, and ultimately facilitating potential joint ventures between development teams.

3) Evolution of mature Research Platforms towards RPaaS:

As Research Platforms mature and deliver services to a larger number of researchers, and become one of the main gateways for them to perform their research, those platforms are naturally expected to deliver a higher level of services. As noted before, these expectations imply the following underlying qualities: 1) Stability 2) Reliability; 3) Accessibility; 4) Security; 5) Sustainability. Also, Research Platforms and research softwares in general can be very complex and do require a lot more HQP per user than what is needed for the underlying HPC infrastructure. That is where we believe NDRIO can best fill the void that CANARIE left following the transfer of the Research Software programs to NDRIO and make sure to provide sufficient long term funding to selected major Research Platforms so they reach the level of RPaaS.

What NDRIO can do: Here we envision NDRIO playing an essential role in making RPaaS a reality for the Canadian researchers community, which will require a long term commitment for each supported RPaaS. So a careful and thorough selection process will need to be established along with a good monitoring system of the RPaaS to make sure they continue offering the expected service level. Here are some propositions on how NDRIO could support this transition towards RPaaS:

1. Organize calls for 5 year long support among existing Research Platforms to acquire the status of RPaaS.
2. Criteria for choosing RPaaS will need to be carefully established. For example: Number and category of active users; Usage of the Platform (sessions, jobs, storage, …); How well the Platform is appreciated by the community; Are there other tools/platforms doing the equivalent.
3. RPaaS would be evaluated every year, with the possibility of defunding the project through a mid-term assessment if things go wrong or if the Platform does not meet the RPaaS criteria anymore.

4. The RPaaS funding should only cover the cost to maintain the required level of services, including minor updates and small continuous improvements. Major developments should go through the Research Platform development calls.
5. All RPaaS teams should include field specialists to directly assist the users.
6. We think that 2-3 FTE per RPaaS for the maintenance would be a strict minimum to ensure staff redundancy and a decent quality of service (some RPaaS could definitely need more). Some of the employees could be working part time on other components of the Canadian DRI, with dedicated time to the RPaaS projects.

Conclusion

CANARIE has been visionary and absolutely essential in the development of the Research Platforms that are now an integral part of the Canadian DRI. NDRIO, as part of its Research Software mandate, should pursue and take the lead to bring this success to the next level. We believe this can be accomplished by assuring a stable and reliable ARC underlying infrastructure and provide mid-level access tools for Platforms, by setting up a NDRIO central support team dedicated to Research Platforms, by supporting the development and improvement of Research Platforms, and finally by offering a strong commitment to major Platforms so they can raise their level to the RPaaS (Research Platform as a Service) status.

Supporters

Felix Breden, Scientific Director - iReceptor, Biological Sciences, Simon Fraser University

Alan Evans, PhD, FRSC, Professor of Neurology & Neurosurgery, Psychiatry, McGill University

Scientific Director - CBRAIN, Canadian Open Neuroscience Platform

Denis Lepage, Senior Director, Data Science and Technology, Birds Canada | Oiseaux Canada (Motus)

Stuart Mackenzie, Director, Migration Ecology, Birds Canada | Oiseaux Canada (Motus)

Xavier Roucou, Professeur, directeur de département, chaire de recherche du Canada, Département de Biochimie et de génomique fonctionnelle, Université de Sherbrooke

Sébastien Rodrigue, Professeur, Département de biologie, Université de Sherbrooke

P. T. Jayachandran, Professor, Physics Department, University of New Brunswick

Brian Corrie, Technical Director - iReceptor, Biological Sciences, Simon Fraser University

Joey Bernard, Computational Research Consultant, Physics Department, University of New Brunswick