

Indigenous language technologies and online resources:
Algonquian Dictionaries Project and Algonquian Linguistic Atlas
(www.atlas-ling.ca)(dictionaries.atlas-ling.ca)

For NDRIO, Dec 2020

PI: Dr. Marie-Odile Junker
Technical Director: Delasie Torkornoo
Carleton University

Our project is unique and representative of Digital Humanities with a focus on Indigenous language technologies and services to the Canadian population. Our large team, working in collaboration with Indigenous communities, organisations and universities across provinces, documents languages, creates dictionaries and other language learning and teaching tools. Most importantly, we offer online Indigenous language resources that would otherwise not be accessible to the Indigenous communities themselves and to the larger public. That aspect of our work is unusual as it fills a national gap in providing Indigenous language online resources that are not provided by any other institution.

We have been using Compute Canada (West Grid) since 2014. Here is our perspective on DRI tools, services and support; our suggestions for future state of DRI in Canada; and how NDRIO could achieve such a state. We are using the questions provided.

Current issues

- What are the main DRI tools, services and/or resources you currently use in your research?

At the core of our research infrastructure is customised open-source data management software. This software runs in a Linux Apache MySQL PHP web server environment. It has been crucial for us to access a cloud based virtual machine that we have been able to customize to run a persistent web server. On this stack we are running multiple applications built with a CakePHP framework, and customized WordPress sites. This is the primary way we disseminate the data we have collected with the general public.

- Do you have access to all the DRI tools, services and/or resources you need for your research? What are they? What is missing?

Through Compute Canada we have access to a reliable cloud VPS to create the web servers we need. However, we need storage for long term backup and archiving.

- What are your biggest challenges accessing and using the DRI tools, services and/or resources that do exist and are available to you?

Resources currently available are mostly geared towards the traditional sciences with an after thought to social sciences and humanities. There is tremendous computing power and storage for temporary use, however there is no easy access to persistent storage and computing power for slow growing projects. Projects like ours must project our computing needs in the future and ask for resources in anticipation of our project growth. This leads to situations where we either over-project or under-project our resource needs, thus leading to waste in the allocation process.

Another challenge is that we must renew access every year. Our principal investigator has to do that with the cumbersome and time-consuming Canadian common CV, otherwise we run the risk of our resources not being accessible, and we fear our work and data getting lost.

At the beginning, documentation was very sparse, and it is still insufficient.

As far as we know, there is no dedicated Linux system support personnel for those running web servers to help with a range issues with setting up, security, SSL and best practices.

There is no diversity in the number of cloud VPS images available.

Future DRI State

- What is your vision for a cohesive Canadian DRI ecosystem that would fulfill your research needs?
- What are the types of DRI tools, services and/or resources you would like to use, or envision using, in the future?
- What challenges do you foresee while using integrated DRI tools, services and/or resources?

Digital Humanities computing needs are going to skyrocket in the next decade as more multimedia data is being collected and curated. Storage, bandwidth and processing power needs will begin to grow. Dynamic scalable persistent resources will be required. There needs to be processes, scripts and documentations for robust migration, and all these should be made available to research teams as they outgrow their allocations. The availability of virtual machine images with pre-set web server stacks and technical support would be a game changer. Targeted training sessions and documentation on best practices for setup & maintenance of web servers would be great. Given that most projects would like to share their data with the

general public, an ecosystem that provides help with establishing an optimal online presence/portal is crucial. Currently there is also a need for long term storage for collections of digital artifacts/data that various research group have accumulated. The expertise of an archivist to help prepare these data collections for long term storage and possible future rediscovery is paramount. The expertise of competent system administrators to help with cloud setup and maintenance would be a valuable contribution to most research groups.

How to Bridge the Gap

- What are the tools, services and/or resources NDRIO should leverage to achieve your desired future state?

Keep improving documentation on accessing the tools, services and resources.

The repository of scripts for initializing cloud servers should be broadened. Some images could be prefabbed with some popular web server stacks pre-set LAMP (CentOS & Ubuntu), MEAN, etc. This would reduce setup cost for most research projects.

It would be a great advantage to have a Linux system support personnel for those running web servers, to help with a range issues with setting up, security, SSL and best practices. Given that most research groups only go through the process of setting up a webserver once or twice, the knowledge and experience required to do it right on the first try is virtually non-existent. Having staff who deals with this on a regular basis, even available for a 30 minutes consultation, would make world of difference. This will not only reduce the time and money on the part of the research team, but also help with the entire security of the infrastructure.

- How do you see NDRIO's role in addressing current gaps in the national DRI ecosystem?

Provide education on best practices for researchers and create a network so we can talk to each other and compare notes.

- What other suggestions do you have?

Have a clear policy on what happens to computing resources when the primary investigator disappears and on what happens to the data of those researchers that are no longer active.

Any decision regarding choices of technologies and systems should consider their energy demand and thus their carbon footprint. Wherever possible, the least energy-intensive options should be prioritized. Opportunities to put "waste heat" to use for another purpose should be sought. As an example, a switch to 5G (or any such future technology) should seriously consider the system-wide

energy demands this will impose. NDRIO should be contributing towards Canada's commitment to achieving a zero-carbon future.