Canadian Research Data Repository Solutions and Innovations

Jeffrey Moon & Lee Wilson, CARL Portage

The research landscape in Canada is changing rapidly, with growing awareness of the importance of research data and research data management (RDM) in supporting reproducibility, avoiding duplication, and facilitating further research. Recent Canadian funding announcements, granting agency policies, and journal requirements surrounding RDM have only accelerated this process. Against this backdrop, Portage, a Canadian library-based RDM network of expertise has partnered with other stakeholders to provide practical solutions for Canadian researchers in support of RDM best practices, Open Data, and upholding FAIR data principles. Specifically, Portage has collaborated with Compute Canada, our national high-performance computing organization, and the Ontario Council of University Libraries (OCUL) Scholars Portal team to provide researchers with two robust and complementary repository options.

The first option, the Federated Research Data Repository (or FRDR) developed in collaboration with Compute Canada, has three main features of practical use to researchers. First, it aims to provide a robust repository option for all Canadian researchers, with the ability to handle large files, maintain file hierarchies, and provide scalable storage capacity at low or no direct cost to researchers. Second, FRDR provides a national data discovery service by harvesting metadata from a growing number of Canadian data repositories to make their data discoverable from a single search portal. Third, FRDR provides preservation processing, using Archivematica, to convert deposited data into preservation-friendly 'archival information packages' or AIPs. The vision is to archive data in a network of preservation service providers.

The second repository option is a national instance of Dataverse, a Harvard-based open-source software used widely in Canada and around the world. The goal of the national instance is to move away from having several siloed institutional and/or regional instances of Dataverse to a single hosted instance that would be better positioned to leverage economies of scale and existing infrastructure and expertise. Under this model, institutions would accrue the benefits of having a hosted repository platform while still being able to operate the service locally. This initiative, led by OCUL Scholars Portal in collaboration with Portage and other regional Library consortia, was successful in attracting federal funding to support the internationalization of Dataverse (i.e. making it bilingual in Canada's two official languages: English and French) and launching a pilot phase involving a large number of Canadian universities. Additional funding was obtained independently to support the development of innovative geospatial search capabilities for both FRDR and Dataverse to leverage rich geo-coded metadata associated with Canadian research data.

These national-scale initiatives will have a profound impact on the RDM landscape in Canada, both for researchers and institutions, by providing robust and flexible repository options needed to fulfill emerging data-deposit requirements from funders and journal publishers, and to meet broader, internationally-accepted 'open science' ideals.