Borealis: The Canadian Dataverse Repository

Annual Report 2022

Report to representatives of Canadian academic library consortia and institutional members of the Borealis service.

In 2022, after 10 years as Scholars Portal Dataverse, the service was officially renamed **Borealis**, the Canadian Dataverse Repository (https://borealisdata.ca). Borealis is a shared, bilingual research data repository service provided in partnership by the four Canadian regional academic library consortia: Bureau de Coopération Interuniversitaire (BCI), Council of Atlantic Academic Libraries (CAAL), Council of Prairie and Pacific University Libraries (COPPUL), and Ontario Council of University Libraries (OCUL), with funding from the Digital Research Alliance of Canada. The University of Toronto Libraries is the Borealis service provider.

A Year in Review
- Introducing Borealis
- CoreTrustSeal Certification Project
- Growth of service: expansion of Borealis beyond universities
- Community activities, training, and outreach

Looking Forward
- Sensitive Data in Borealis
- ODESI migration
- Large files support

A Year in Review

Introducing Borealis

In June 2022, Scholars Portal Dataverse was officially renamed Borealis, the Canadian Dataverse Repository. The new name was the culmination of the efforts of the Renaming Working Group, which included representation from the four regional academic library consortia and the Digital Research Alliance of Canada. The working group conducted an effective community consultation process including focus groups and meetings with key stakeholders.

In addition to the new name, the Borealis team also launched a new logo and website domain (borealisdata.ca) with a new look and feel. Redirects from the old website will be in place until at least January 2023, all DOIs continue to resolve, and repository features remain the same. The Borealis website now includes more information about repository use cases and research data profiles and a new About page that provides more details about the service, including a mission statement, current governance and organization, and our designated user community.

New policies were developed and published over the past year to support the launch of the new name and service, and to provide more information to institutions and researchers. These include a new Terms of Use, Technology Infrastructure and Security Information, Preservation Plan, and Privacy Statement. In September 2022, we also published an updated User Guide, which was developed to improve end-user guidance and support improved documentation for the CoreTrustSeal Certification Project (described below).

Borealis's evolution into a national repository collaboration provides academic libraries and research institutions in Canada an opportunity to invest in secure, scalable, reliable, open-source repository infrastructure, and to foster an open community for both RDM practitioners and researchers.

Borealis by the numbers (as of September 2022)

64 participating institutions	**5837** users	**10 million** downloads in 2022
4,549 datasets published in 2022	**15,433** published datasets	**1,336** datasets with direct links to publications and other scholarly outputs

CoreTrustSeal Certification Project

Over the past year, the Borealis team has also participated in and supported the national CoreTrustSeal (CTS) [Certification Cohort Project](#) led by the Digital Research Alliance of Canada (the Alliance). Through a cohort-within-a-cohort approach, participating Borealis institutions were involved in capacity building, training, and local policy development in pursuit of CTS certification. Within the larger Alliance project, the Borealis cohort included over ten participating institutions and experts from across the community, Dataverse North, and the Preservation Expert Group, including a dedicated Policy Working Group, who met regularly to discuss areas of interest.

Resources for those interested in pursuing or learning more about CTS certification are openly available on the [Borealis CTS Project website](#). These include:

- Background research, including successful Dataverse applications;
- Institutional and repository policy guidance;
- Institutional Policy Templates (developed by the Dataverse North Policy Working Group); and
- The [Borealis CTS Application Template](#), which assists institutions with CTS's 2020-2023 application requirements. Template updates are planned to support requirements in 2023 and moving forward.

Growth of service: expansion of Borealis beyond universities

The Borealis team continues to work on growing the community, with a focus on developing new models that allow a more diverse set of subscribers, including colleges and research organisations, to join the service. In 2022, an agreement was finalised between the College Libraries Ontario (CLO), Ontario Colleges Library Service (OCLS), and the University of Toronto Libraries (as Borealis Service Provider) for Ontario colleges. There are also ongoing preparations underway for the Quebec CEGEP consortium to join the service.

As a result of this work, the Borealis team welcomed several new subscribers this year from across Canada, including our first three colleges:

Colleges

- Aurora College, Northwest Territories
- Durham College, Ontario
- Fanshawe College, Ontario (official start date of January 2023)

Universities

- Concordia University of Edmonton, Alberta
- Thompson Rivers University, British Columbia

Community activities, training, and outreach

The Borealis team supports the Borealis Dataverse Community, a community of institutional collection administrators of Borealis and of other Dataverse repositories in Canada. The Community includes over 100 RDM support staff in libraries from the 64 partnering institutions who provide research data repository services to their researchers.

The Borealis Dataverse Community Facilitation Team (comprised of representatives from the Borealis team, the Curation Officer and other members from the Digital Research Alliance of Canada, and the chair of the Dataverse North Expert Group) coordinate to deliver virtual monthly meetings that include Borealis updates on new features and platform improvements, as well as community updates about activities and events related to Dataverse use in Canada. Some highlights from the past year's [Monthly Community Meetings](#):

- 10 virtual monthly meetings were held for the community between September 2021-September 2022
- Updates covered Dataverse software upgrades, CoreTrustSeal certification, Borealis rebranding, policies, large file support, metrics, and community events
- Meetings provided opportunities for discussion that contributed to engagement across the community

A brief survey of community meeting participants in June 2022 revealed overall a high level of satisfaction with the community meeting, and in particular, an appreciation for the bilingual format, topical coverage, and focus on Borealis updates. Results of the survey were presented to the community in September 2022.

The Borealis team hosted several training sessions and events over the past year, including:
- September 2022: Borealis New Joiner Session
  - This session targeted new and recently joined members or administrators, providing an overview of the platform, administrative features, and helpful tips for getting started, followed by a platform demo. There were 50 community participants (36 in English, 14 in French)
- May 2022: Training Session for Ontario College Libraries
- November-December 2021: Alliance RDM three part webinar series "[Dataverse in Canada: Leveraging an open data repository for research data stewardship](#)"
  - Three sessions held over a month-long period, with over 100 participants from across Canada in attendance at each of the three sessions

As a complement to the new [User Guide](#), a new [Administrators Guide](#) was released in draft in September 2022, with full translation coming soon. New training events and videos are being planned with the Alliance's CEG & Dataverse North for 2023.

Looking Forward

Sensitive Data in Borealis

Researchers are looking for solutions to deposit and share de-identified sensitive data with increasing frequency. Dataveverse currently provides a number of features that allow researchers to restrict access to data and/or set limits on its use, including file level restrictions, granular permissions, request access options, and custom data licensing support.

In order to better support end-users and administrators working with existing functionality for restricted data deposit and sharing, the Borealis team is developing new documentation that will include an overview of current policies, clear options for restricted data deposit and sharing, and common tools and workflows. This work is being done in collaboration with the University of Toronto Libraries and will be shared with the community for feedback and comment on varying institutional contexts later this year.

The Borealis team is also working closely with the Alliance RDM team through participation in the Alliance's Sensitive Data Repository Project, a two-year funded project that aims to bring together a diverse range of stakeholders to address sensitive data infrastructure and policy-framework needs in Canada. Areas of Borealis planned participation include roadmap development, working group membership, and infrastructure development. More information will be shared with the community over the next year.

Looking ahead to future releases of the Dataverse software, there are a number of integrations under development to better support sensitive data, including several from the wider Global Dataverse Consortium Community such as [OpenDP](#), [DataTags](#), and [Remote Trusted Storage Agent.](#)

ODESI migration

Since 2009, OCUL's ODESI service (https://odesi.ca), has been a valued academic library data service supporting access to social science survey data and public opinion polls for researchers, students, staff, and instructors at over 40 academic institutions across Canada. With over 6000 datasets in both English and French, ODESI offers broad discovery, access, and online exploration and analysis of current and historical Canadian social science survey data. ODESI uses the Data Documentation Initiative (DDI) metadata standard and a custom search catalogue maintained by Scholars Portal and the [OCUL Markit! Program](#), to curate, publish, preserve, and share rich survey data for researchers and end-users. Underlying the current ODESI search catalogue is a licensed data repository software called Nesstar Server, which is reaching end-of-life and is no longer being supported by its vendor. In order

to support the future sustainability of the ODESI service, OCUL Scholars Portal will be migrating the data collections from ODESI's backend Nesstar repository to Borealis.

Moving forward, OCUL will be working with a variety of stakeholders and partners to develop a sustainable model for the ODESI service in the spirit of making the data more FAIR (Findable, Accessible, Interoperable, Reusable). More information about the migration, including an open webinar and FAQ for all ODESI users will be announced in the next few months. In the meantime, please visit the ODESI wiki space for more details and the latest information.

Large file support

The Borealis team continues to improve large file and volume support in the repository. Support for the open-source DVUploader Tool for batch file upload is now available. This tool can be used to batch upload files into the repository via command line (tested up to 50 GB with a maximum individual file size of 3 GB).

Additionally, this year Harvard University and the Institute for Quantitative Social Science (IQSS) partnered with Borealis to develop the proof-of-concept integration between Dataverse and Globus via the Dataverse - Globus Tool. Built primarily for Borealis use, the tool will provide a seamless, bilingual user interface for large file upload and download with Globus. It also provides improved user notifications and transfer selection tools. In October 2022, an open experimental release of this Globus integration was included in the Dataverse software (v. 5.12). The Borealis team is currently testing this release and developing configuration options for the Borealis context, such as increasing the maximum file size to 10 GB. More information will be shared with the Borealis Dataverse Community at upcoming monthly meetings. The team plans to continue to do testing with pilot institutional participants in early 2023 and aims to launch the Globus integration in spring 2023.