CANARIE project completion: Dataverse for the Canadian Research Community

The Scholars Portal Dataverse team is pleased to announce the completion of the development project "Dataverse for the Canadian Research Community," funded by CANARIE's program for Research Data Management software tools.

The 18-month project involved innovative development of the Dataverse platform to better support Canadian researchers around data deposit and sharing with national and international collaborations. The development work addressed key themes around data curation, authentication, scalability, and large-file support (described below).

Data Curation

The Data Curation Tool (DCT) allows users to create and edit variable-level metadata for tabular files (e.g., SPSS, R, Excel, CSV), as a modular application launched from within Dataverse. The aims of the DCT are to improve data curation workflows within Dataverse, to improve the ability for data reuse, and to support the application of standards and best practices using the Data Documentation Initiative (DDI) metadata standard.

Other features include the ability to view summary statistics and charts about the data. User edits are saved back to Dataverse and can be exported outside the platform.

Special thanks to the University of Ottawa for the French translations.

Deliverables

- The DCT was launched in Scholars Portal Dataverse on October 31, 2019.
- Code openly available in GitHub
- Hosted webinar (see recording)
- Published blog post on features and functionality

Authentication

Scholars Portal configured Dataverse to work with Shibboleth for institutional single sign-on through the Canadian Access Federation (CAF), an identity management service for Canadian research institutions run by CANARIE. Dataverse requires each user's email, first name, last name, and affiliation, which are released under Research and Scholarship (R&S) entity profile.

The benefits of using the CAF's R&S include ease of collaboration between Dataverse as the service provider and institutions as the identity providers. CAF's vetting process ensures secure and trustworthy exchange of identity information.

For Dataverse users, this integration results in a simpler log-in process with one less username and password to manage.

Deliverables

- Launched in Scholars Portal Dataverse on October 31, 2019
- 14 institutions now participating
- Planning upcoming webinar with Portage and CANARIE on the benefits of joining CAF's R&S profile
- Documentation in Scholars Portal Dataverse Guide coming soon

Scalability

Scholars Portal connected Dataverse to cloud storage by hosting files in a test cluster of the Ontario Library Research Cloud (OLRC).

The aim is to optimize system architecture for scalable use and to leverage an existing, distributed Canadian data storage network.

Deliverables

- Developed in test environment with an innovative design
- Plan to upgrade to cloud storage for Scholars Portal Dataverse (end of 2020)

Large-file support

Scholars Portal developed proof-of-concept integration with Globus as a large-file transfer mechanism. Dataverse users would run Globus Connect Personal (free software) and have a Globus account to upload/download files to/from Dataverse.

Our testing demonstrated robust transfers up to 100 GB in size and up to 38,000 files. We are continuing to collaborate and consult with the Harvard's IQSS Dataverse team to bring this proof-of-concept development work into the core code.

Deliverables

- Developed in test environment
- Code available in GitHub
- Plan to launch in Scholars Portal Dataverse (early 2021)
- Blog post with demo of Globus deposit coming soon

Acknowledgements

The development work was a collaborative effort by the Scholars Portal Team:

- Direction/Organization
    - Kate Davis - PI
    - Amaz Taufique - Technical Lead
    - Amber Leahey - co-PI (on leave)
    - Meghan Goodchild - Project Manager
    - Kaitlin Newson
- Developers
    - Jayanthy Chengan
    - Sunil Manikonda
    - Victoria Lubitch
- Systems Support
    - Bikram Singh
    - Sohaib Anwar
    - Carlos McGregor
    - Dawas Zaidi

Special thanks for input and feedback:

- Lee Wilson (Portage)
- Danny Brooke, Gustavo Durand, and Tania Schlatter (IQSS Harvard)
- Jim Meyers
- Felicity Tayler and Pierre Leblanc (University of Ottawa)