

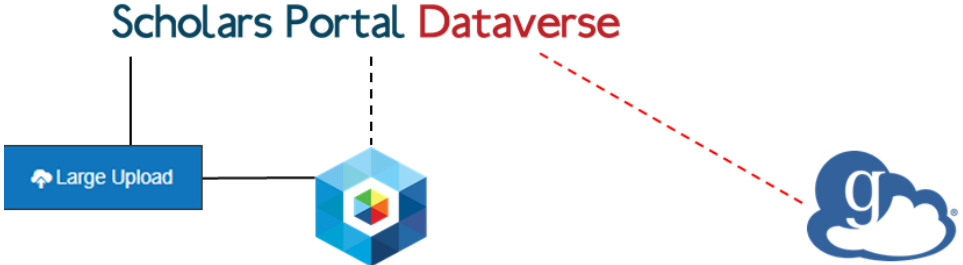

CANARIE project update: Dataverse for the Canadian Research Community

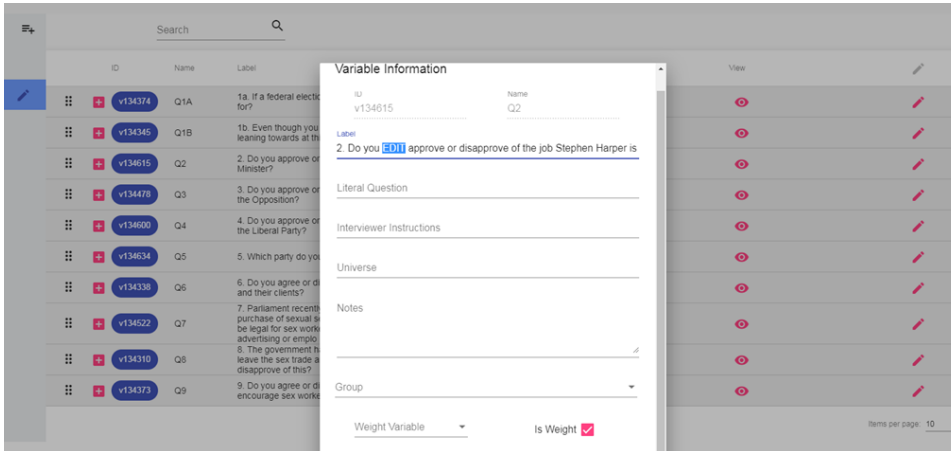
Background

Welcome to our Scholars Portal Dataverse blog, where we will be sharing news and updates about the Dataverse platform and service, including development work. Our first blog post provides an update about the development project "**Dataverse for the Canadian Research Community**". This project is funded by [CANARIE's RDM grant program](#) and led by Scholars Portal and University of Toronto Libraries, with support from CARL and Portage.

We're currently about half way through our 18 months of development work (October 2018-March 2020).

The aim of the grant is to enhance Dataverse to address the needs of a broad range of researchers in Canada through improved scalability, improved integrations with Canadian cloud storage and authentication providers, and better support for data curation workflows. These three areas of development are described further below and will be discussed in more detail in future blog posts.

<p>Scalability</p>	<p>The goals of the first leg of the project include:</p> <ul style="list-style-type: none"> • Optimize system architecture for scalable use • Connect to existing Canadian cloud data storage environments • Support large files in upload/download contexts <p>Planned deliverables:</p> <ul style="list-style-type: none"> • Develop and test connections to SWIFT object storage, such as the Ontario Library Research Cloud (OLRC) • Support Globus endpoints with file access mediated outside of Dataverse application • Develop large-file upload utility to support deposit of larger file sizes (2GB+) into Dataverse 	
<p>Authentication</p>	<p>The goals of the second leg of the project include:</p> <ul style="list-style-type: none"> • Integrate with Canadian authentication infrastructure • Streamline login workflows <p>Planned deliverables:</p> <ul style="list-style-type: none"> • Integrate Dataverse with CAF Shibboleth Login for single-sign on • Investigate further integration with ORCID to support linking research outputs 	

Data Curation	<p>The goals of the third leg of the project include:</p> <ul style="list-style-type: none"> Enhance multi-disciplinary support for data curation Enable users to adopt metadata standards and best practices <p>Planned deliverables:</p> <ul style="list-style-type: none"> Data Curation Tool, a modular application integrated within Dataverse that would allow users to create and edit variable-level metadata of tabular data files to aid in data re-usability 	 <p>The screenshot shows the Dataverse web interface. On the left, a table lists variables with columns for ID, Name, and Label. A modal window titled 'Variable Information' is open, showing details for variable v134615, including its name (Q2), label, and a description. The modal also has sections for 'Literal Question', 'Interviewer Instructions', 'Universe', 'Notes', and 'Group'.</p>
---------------	--	---

Status update

Our Project Timeline & Deliverables roadmap is included below. We have completed our first two deliverables and are currently working on the third.

For our first deliverable to **connect Dataverse with Swift as the primary storage service**, we stood up a test instance of Dataverse connected to the OLRC. The SP team tested upload and download functionality, as well as the integrity of files stored, with a variety of file types and sizes, along with other functionalities core to Dataverse. The idea behind this type of configuration would allow us to more easily scale the system, add storage resources, and run the platform more optimally.

We have also successfully **configured Dataverse to work with Shibboleth** for single sign-on using the University of Toronto as the test case. We are now initiating a pilot project with interested institutions to test out new sign-up and login workflows. More details to come in another blog post.

Currently, we are working on completing our third deliverable - developing the **Data Curation Tool**. We presented the DCT prototype at NADDI ([link to slides](#)) and at the Dataverse Community Meeting ([link to slides](#)). Feel free to test out the [Data Curation Tool - Prototype](#) and stay tuned for a future blog post describing the development of this tool.

In the fall, we will start to focus on the large-file support and storage connection pieces of the project.

We will be sharing more details about these deliverables and details about the development work in upcoming blog posts! If you have any comments or suggestions, please feel free to contact us at dataverse@scholarsportal.info. If you would like to see all the updates and have a Spotdocs account, click the "Watch this blog" button on the top right corner of the page to receive notifications.

Project Timeline & Deliverables

Q1-Q2 2019	Q3-Q4 2019	Q1-Q2 2020	Q3-Q4 2020
<p>Connect DV with Swift as primary storage</p> <p>Configure DV Shibboleth with CAF</p> <p>Data Curation Tool writes edits to DDI records</p>	<p>Develop large-file utility</p> <p>Connection with Globus endpoints</p> <p>Update UI and APIs with new connections</p>	<p>Migration from test environment to production</p> <p>Merge code into DV main code base</p>	<p>Launch Dataverse as national service</p>