



Scholars Portal activities within the Canadian RDM landscape

Meghan Goodchild (Scholars Portal / Queen's University)
Steve Marks (University of Toronto)

National Data Services Framework Summit
February 5, 2020



Agenda

- Overview of Scholars Portal and RDM
- Updates on projects funded by CANARIE's RDM grant program:
 - DuraCloud Canada
 - Dataverse for the Canadian Research Community
- Where do these services fit into a NDSF?

Scholars Portal and RDM

What is Scholars Portal?

- Service arm of the Ontario Council of University Libraries (OCUL)
- Provides shared technology and collections for Ontario university libraries and beyond
- Leadership in providing infrastructure and services around RDM
- 2 projects directly funded by CANARIE RDM grant program



Scholars Portal & RDM

Scholars Portal **Dataverse**

Dataverse is a research data repository to store, share, publish and discover research data
(dataverse.scholarsportal.info)



Ontario Library Research Cloud is a high capacity, geographically distributed storage and compute network
(cloud.scholarsportal.info)



Permafrost is a hosted digital preservation service providing tools, training and resources for processing digital objects for long-term preservation
(permafrost.scholarsportal.info)

DuraCloud Canada: Linking Data Repositories to Preservation Storage



So this DuraCloud thing.....

what is it



Wait, first let's talk about preservation

just for a minute, I promise



Digital Preservation

“Digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary.”*

This.



Not this.



*: <https://dpconline.org/handbook>



So this DuraCloud thing.....



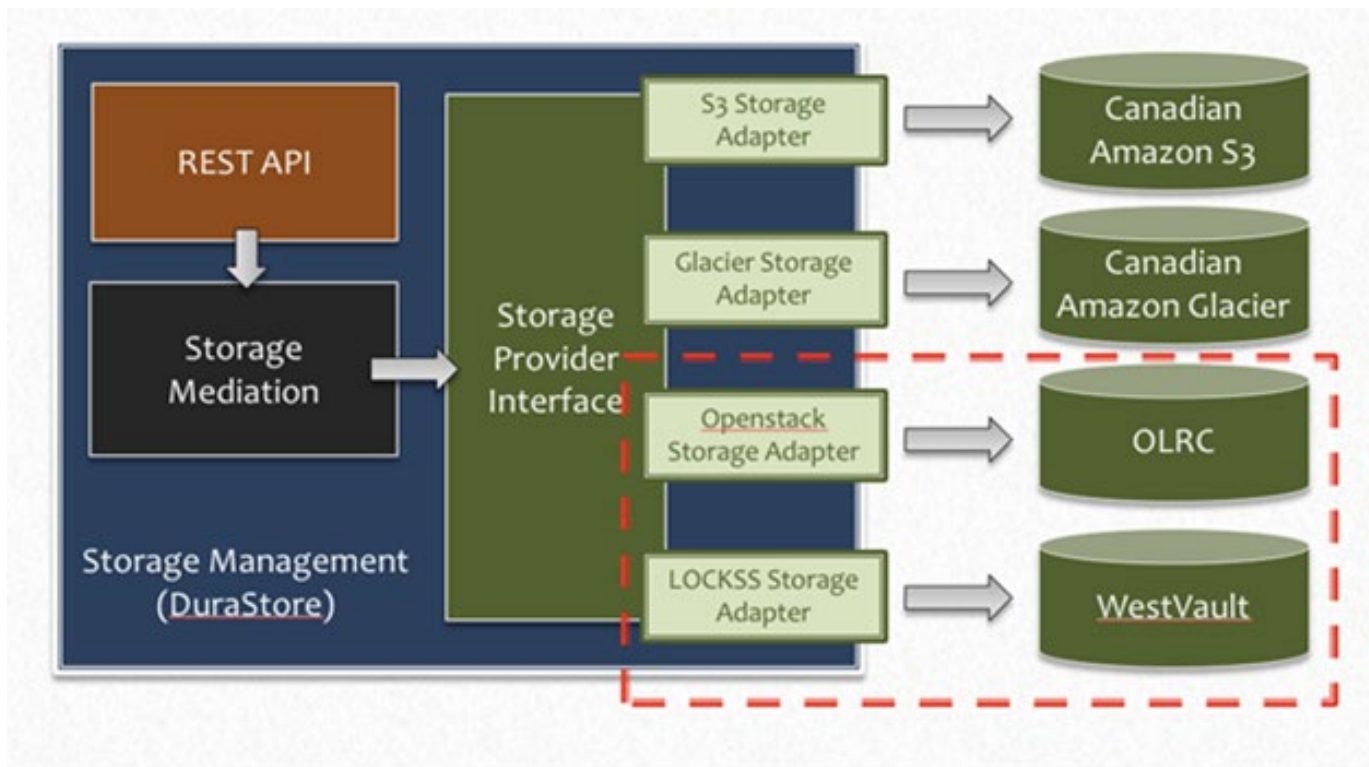
X





RDM-0063 - DuraCloud Canada

“The purpose of the proposed project is to connect these preservation storage services through a common deposit layer based on software from the DuraSpace foundation called DuraCloud.”





Spaces



- Spaces + Add Space
-
- kanata
 - myfirstspace
 - stevedata
 - testspace2

- Content Items Refresh + Upload
- Showing 1 - 7 of 7
- Documents/Five_Parsecs_From_Home.pdf
 - Documents/jonn mumford - chakra & kundalini workbook [ocr].pdf
 - STLs/Battleandroid1.stl
 - STLs/battleandroid2.stl
 - STLs/cavespider.stl
 - hi.txt
 - lifecycle clean.png

Provider: Swift S3

Space Detail

stevedata

- Delete Audit Log Manifest Make publicly readable

Permissions

Edit Add

User/Group	Read	Write
testuser	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Details

Items 7 Recount

Created 2020-02-03T21:35:57

History


There is no history available for this space.

DURACLOUD SYNC TOOL

Status Configuration

Overview

Stop Start

 Waiting

Sync Started	Tue Feb 04 11:09:34 EST 2020
Queue Size	0
Error Count	0

Active Syncs (0)

Download History

There are no active uploads at this time.

Recent Activity

File (hover for full path)	Action	Size	Duration	Complete	
hi.txt	Added	3 B	22 secs	02/04/11:11 A	^
hi.txt	Added	3 B	22 secs	02/04/11:11 A	v

< >

Queued for Synchronization (0) Errors (0)

There are no files currently queued for synchronization.



FAIR Intersections

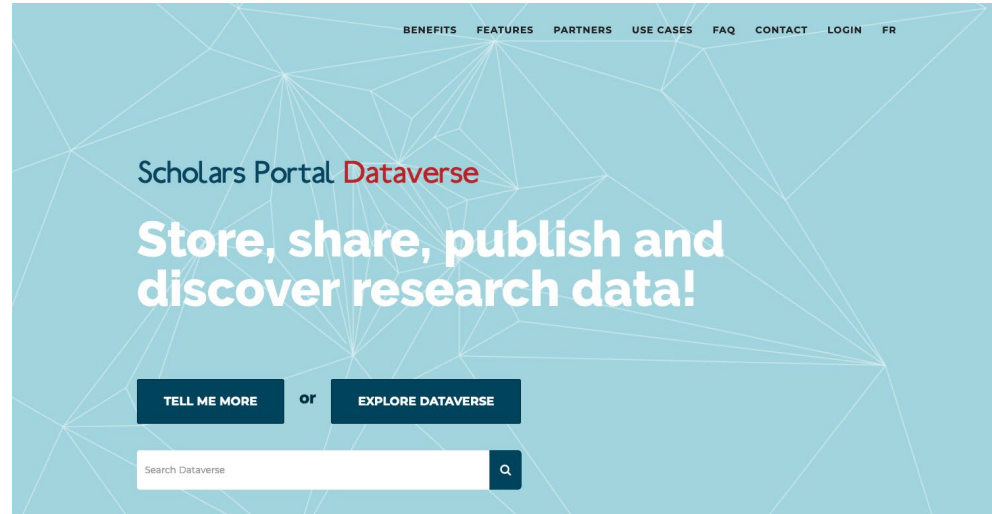
- Accessible
- Reusable
- But really....

Dataverse for the Canadian Research Community




What is Dataverse?

- Open-source research data repository software
- Store, share, publish and discover research data
- Hosted by Scholars Portal on behalf of 51 universities across Canada



dataverse.scholarsportal.info

Dataverse features that enable FAIR



Findable	<ul style="list-style-type: none">• Automatic data citation with DOI• Rich metadata description possible• OAI-PMH harvesting protocol; API access
Accessible	<ul style="list-style-type: none">• Descriptive and citation metadata based on DDI standard (compliant with Dublin Core)• Optional discipline-specific metadata blocks
Interoperable	<ul style="list-style-type: none">• Create links to other scholarly outputs• Automatic file format identification, file verification (checksums), tabular data transformation
Reusable	<ul style="list-style-type: none">• Apply data reuse licence (CC0 default) + templates• Dataset and file versioning• Provenance information



Our project

- Aim to make enhancements to the Dataverse platform to address the needs of a broad range of users across Canada
- 3 key focus areas:
 - Authentication
 - Data Curation
 - Scalability



canarie



Our team

Kate Davis - PI
Amaz Taufique- Technical Lead
Amber Leahey- co-PI (*on leave*)
Meghan Goodchild - Project Manager
Kaitlin Newson

Developers

Jayanthi Chengan
Sunil Manikonda
Victoria Lubitch

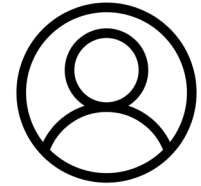
Systems

Bikram Singh
Sohaib Anwar
Carlos McGregor

Lee Wilson (Portage)

1) Authentication

- “As a user from an affiliated university, I want to log into Dataverse using the same login as other services I access.”
- Integrate Dataverse with Shibboleth for single-sign on
 - Canadian Access Federation (CAF)’s Research and Scholarship (R&S) entity profile
- Enables ease of collaboration between Dataverse (as service provider) and institutions (as identity providers)
- Ensures secure and trustworthy exchange of identity information



Shibboleth®

Scholars Portal Dataverse

➔ Log In

Log in or sign up with your institutional account — [learn more](#). Leaving your institution? Please contact [Dataverse Support](#) for assistance.

Your Institution

- ✓ Please select...
- Carleton University
- McGill University
- University of British Columbia
- University of Guelph
- University of Toronto
- University of Victoria
- University of Western Ontario

Continue

[Institution](#)

Other options

Username/Email

Sign In



2) Data Curation

- “As a user, I want to add details about the variables in my dataset after I’ve uploaded it to Dataverse.”
- Developed the Data Curation Tool
 - External web application that connects to Dataverse to create and edit metadata at the variable level using DDI standard
- Aim to improve data curation workflows and promote adoption of standards and best practices
- Works with files uploaded through tabular ingest process (XLSX, CSV, SAV etc.)
- Launched as part of our upgrade in October

Labour Force Survey (Curated)

LFS2016-01_PUMF_EN.tab

Tester, Curation, 2019, "Labour Force Survey (Curated)", <https://doi.org/10.5072/FK2/YLJJAY>, Scholars Portal Dataverse, V1, UNF:6:EerkQFr2ySzwCu4oV5JExA== [fileUNF]

< Hide Groups


Download Save




















Add Group +

Search

Items per page: 25 1 - 25 of 75 < >

All Variables

Number of hours 

<input type="checkbox"/>	ID	Name	Weight	View	
<input type="checkbox"/>	v14110	REC_N			
<input type="checkbox"/>	v14130	SURVY			
<input type="checkbox"/>	v14106	SURVM			
<input type="checkbox"/>	v14127	LFSSTA			
<input type="checkbox"/>	v14165	PROV			
<input type="checkbox"/>	v14168	CMA			
<input type="checkbox"/>	v14155	AGE_12			
<input type="checkbox"/>	v14136	AGE_6			
<input type="checkbox"/>	v14158	SEX			

Variable Information

ID: v14158 Name: SEX

Label: Sex of respondent

Literal Question

Interviewer Instructions

Post Question

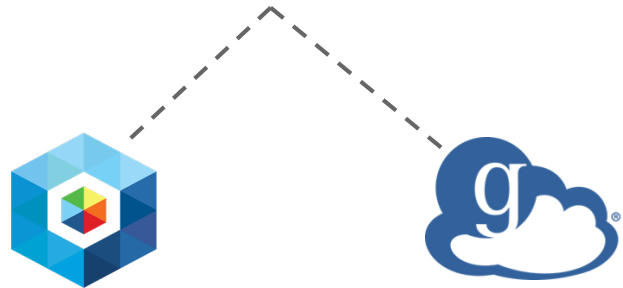
Universe

Notes

3) Scalability


- “As a user, I want to upload files larger than 2.5 GB to Dataverse.”
- Optimize system architecture for scalable use
- Connect to existing Canadian data storage environments
- Support large files in upload/download contexts

Scholars Portal **Dataverse**



Dataverse - Globus option on “add data” page


Files

Upload with HTTP via your browser 

Select files or drag and drop into the upload widget. File upload limit is 12.0 GB per file.

[+ Select Files to Add](#)

Drag and drop files here.

Upload with Globus 

BEFORE YOU START: You will need to set up a free account with [Globus](#) and have [Globus Connect Personal](#) running on your computer to transfer files to and from the service.

[Upload with Globus](#)

Once Globus transfer has finished, you will get an email notification. Please come back here and press the following button:

[Globus Transfer has finished](#)

Click here to view the dataset page: [Globus Test 2020-01-16](#) .

Globus Connect Personal

File Manager

Collection: Meghan Goodchild's Laptop

Path: /~/OneDrive - Queen's University/Test Data/

Dataverse GCS test collection

/10.5072/FK2/KIDLKX/

select all up one folder refresh list columns view

dataverse_files (10).zip	12/11/2019 10:49am	74.85 MB
dataverse_files (300).zip	03/27/2019 10:38am	1.41 MB
eTheses	07/24/2018 02:45pm	-
Images	10/10/2019 09:55am	-
Images.zip	10/10/2019 09:58am	718.05 KB
LargeTest.zip	11/27/2017 02:55pm	5.26 GB
LargeTestFolder.tar.gz	11/28/2017 03:42pm	5.23 GB
MANIFEST.TXT	04/03/2019 08:08am	718 B
Queen's Journals	07/24/2018 02:44pm	-
Test 100 MB (2).zip	06/22/2018 07:46am	102.69 MB
Test AV 1GB.zip	05/25/2018 08:01am	1.05 GB
Test AV 2GB.zip	05/25/2018 08:01am	2.10 GB
Test AV 4 GB (2).zip	06/21/2018 09:13am	3.49 GB

- Share
- Transfer or Sync to...
- New Folder
- Rename
- Delete Selected
- Download
- Open
- Upload
- Get Link
- Show Hidden Items
- Manage Activation

cached	01/20/2020 09:04pm	1.01 KB
_json.cached	01/20/2020 09:04pm	4.46 KB
cached	01/20/2020 09:04pm	894 B
d	01/20/2020 09:04pm	2.81 KB
ied	01/20/2020 09:04pm	8.65 KB
ite.cached	01/20/2020 09:04pm	1.83 KB
ched	01/20/2020 09:04pm	627 B
ached	01/20/2020 09:04pm	2.78 KB
.cached	01/20/2020 09:04pm	4.80 KB
rg.cached	01/20/2020 09:04pm	1.94 KB
p	01/20/2020 06:42pm	102.69 MB
	01/20/2020 06:41pm	1.05 GB
	01/20/2020 06:38pm	2.10 GB

Start

Transfer & Sync Options

Start

Dataset page post upload



Globus Test 2020-01-16

Version 2.0

Goodchild, Meghan, 2020, "Globus Test 2020-01-16", <https://doi.org/10.5072/FK2/KIDLKX>, Scholars Portal Dataverse, V2

Cite Dataset

[Learn about Data Citation Standards.](#)

Dataset Metrics

0 Downloads

Description

test

Subject

Arts and Humanities

Files

Metadata

Terms

Versions

Search this dataset...

Find

Upload Files

Filter by

File Type: All Access: All

Sort

1 to 8 of 8 Files

Edit Files

Download

AcademicKnowledgeProcess.PNG
application/vnd.dataverse.file-globus - 1.0 MB - Jan 23, 2020 - 0 Downloads
SHA-1: 7c4e62ed6bd22ace8aa7996679a983b9

Download through Globus Download

CurationVsSelfDepositDCN.PNG
application/vnd.dataverse.file-globus - 262.9 KB - Jan 23, 2020 - 0 Downloads
SHA-1: e8f0ee2bf6767bcb3548031088b30caf

Download through Globus Download

CurationVsSelfDepositDCNGraph.PNG
application/vnd.dataverse.file-globus - 147.7 KB - Jan 23, 2020 - 0 Downloads
SHA-1: 0726acd43dba00739e2761a5942065eb

Download through Globus Download

Test 100 MB (2).zip
application/vnd.dataverse.file-globus - 97.9 MB - Jan 20, 2020 - 0 Downloads
SHA-1: 1fc35b98e03c69c90826acd875df3cf2

Download through Globus Download

Challenges and lessons learned

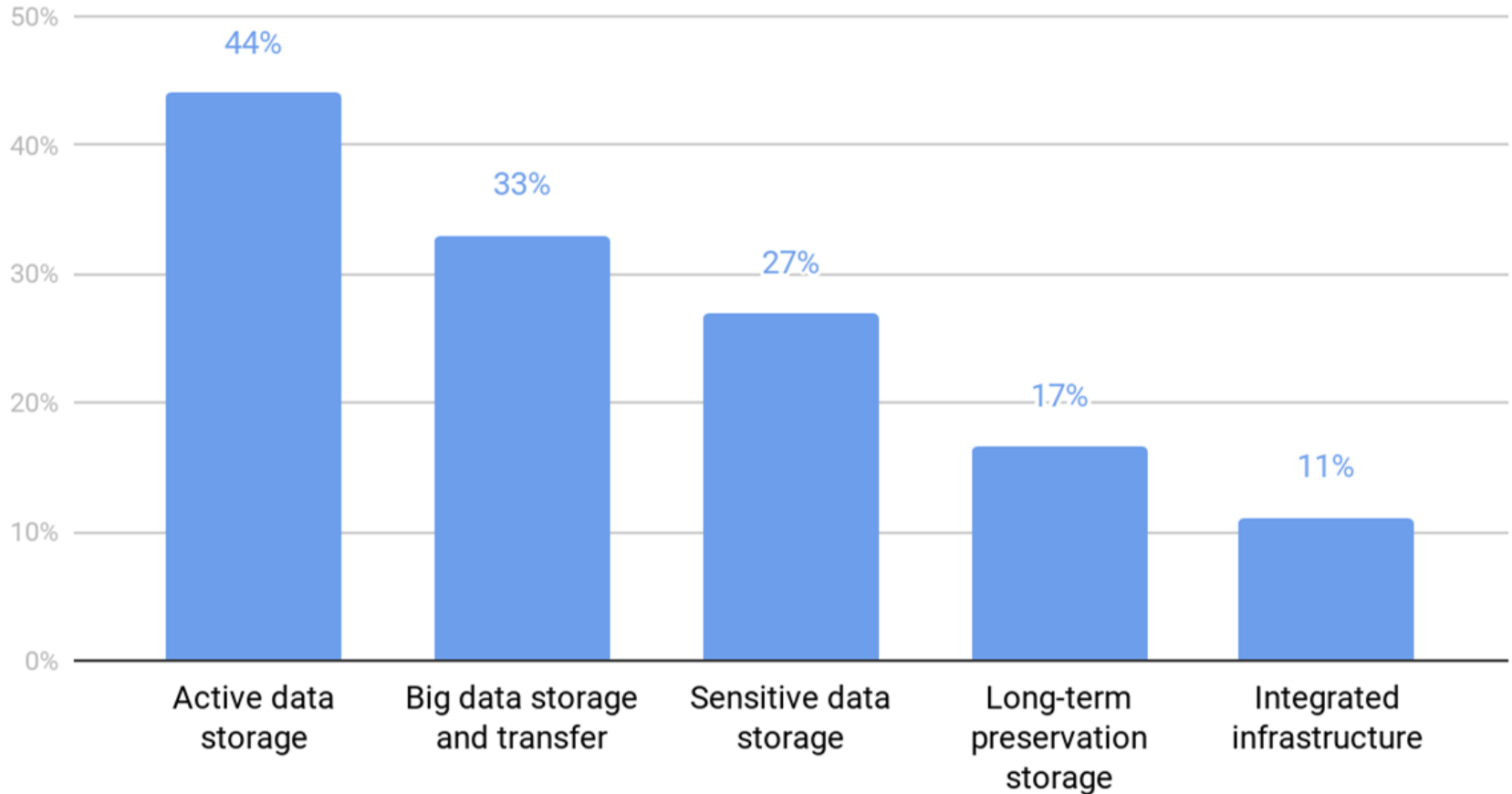


- Challenge working with open-source community for developments to be incorporated into core code
- Need for usability testing during development workflow
- Internationalization considerations
- Promotion of developments to ensure uptake by researchers and institutions



NDSF connections

RDM barriers from OCUL RDM needs assessment



Building a national Dataverse service

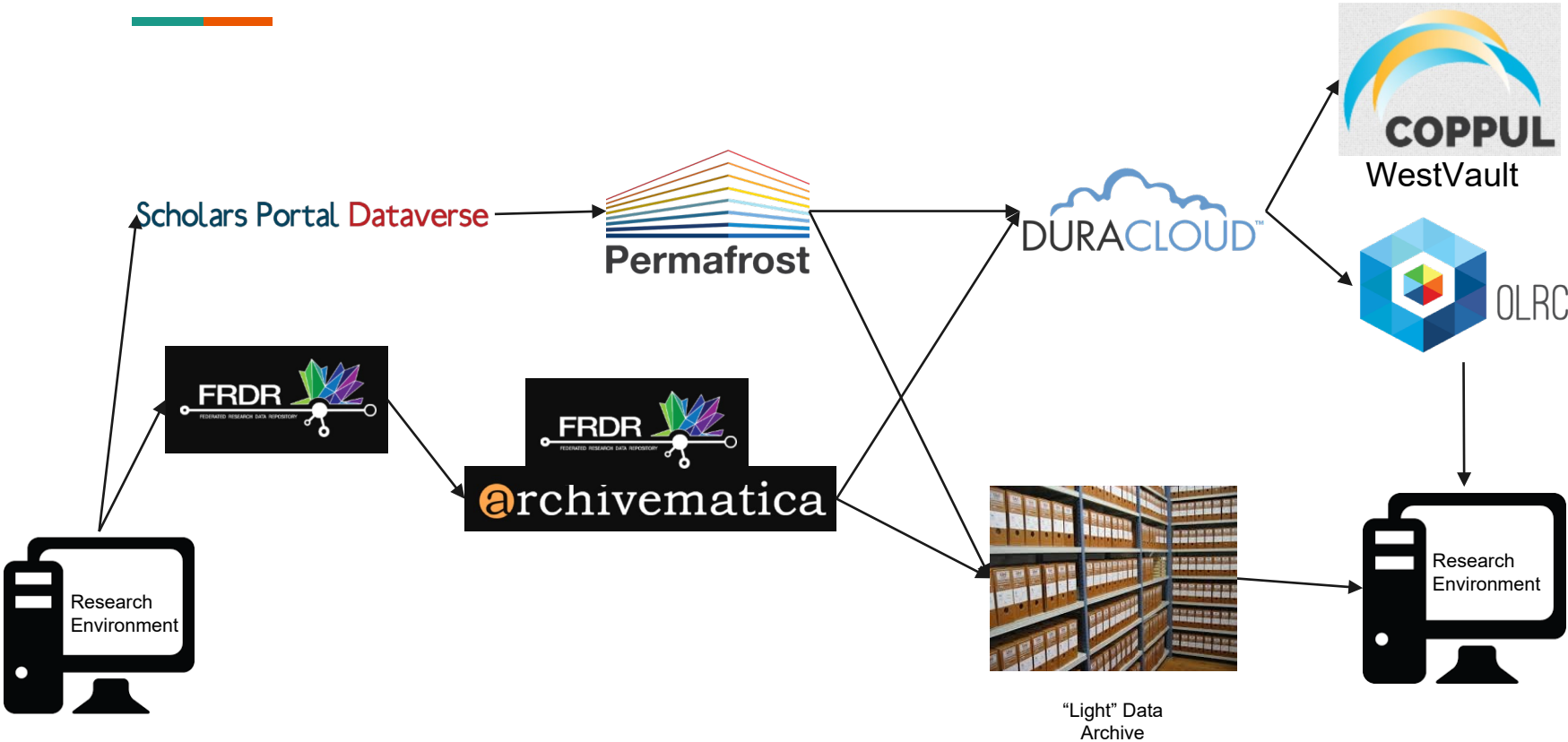
National multi-stakeholder engagement

- Engaging with Portage's Dataverse North Working Group
- Partnerships with other regional library consortia
- Defining national Dataverse governance structure
- Seeking national funding opportunities
 - CANARIE RDM grant program
 - Part of \$2M CANARIE Directed Funding program

Local connections and touchpoints

- Scholars Portal Dataverse Admin Contacts listserv and monthly meetings
- Training and outreach to help build local capacity
- Lowering barriers to access for researchers
- Continually evaluate and respond to researcher needs

Knitting together a national framework





Thank you!

meghan.goodchild@queensu.ca

steve.marks@utoronto.ca