

# National Repository Updates

Brad Covey, Meghan Goodchild, Kaitlin Newson, Reyna Jenkyns, Lee Wilson

**Canadian Data Curation Forum**  
**October 18, 2019**

# Overview

- Federated Research Data Repository (FRDR)
  - Lee Wilson
- Scholars Portal Dataverse
  - Meghan Goodchild and Kaitlin Newson
- Canadian Integrated Ocean Observing System (CIOOS)
  - Brad Covey and Reyna Jenkyns

# Federated Research Data Repository

## Transition Y1

Lee Wilson

Service Manager | Portage/ACENET

613.482.9344 ext. 108 | [lee.wilson@ace-net.ca](mailto:lee.wilson@ace-net.ca) | @lee\_wilson001



# FRDR: A Collaborative Effort

**A collaboration among libraries, institutions, and federal organizations, involving many stakeholders across the RDM & ARC spheres**

- Partnership between Compute Canada Federation (CCF) and the Canadian Association of Research Libraries' Portage Network
- Hosted on CCF national hosting site infrastructure, with CCF providing development and technical support
- Service operated by Portage, including curation and data management support, with steering and input from CARL, the Network of Experts, and individual institutions
- Funding for the transitional year one provided by CANARIE

# ***FRDR = Federated Research Data Repository (DFDR en français)***

<https://www.frdr.ca/>

*A scalable, federated platform for digital research data management and the discovery of Canadian research data*

**Discovery**

**Deposit**

**Preservation**

# ***FRDR = Federated Research Data Repository (DFDR en français)***

<https://www.frdr.ca/>

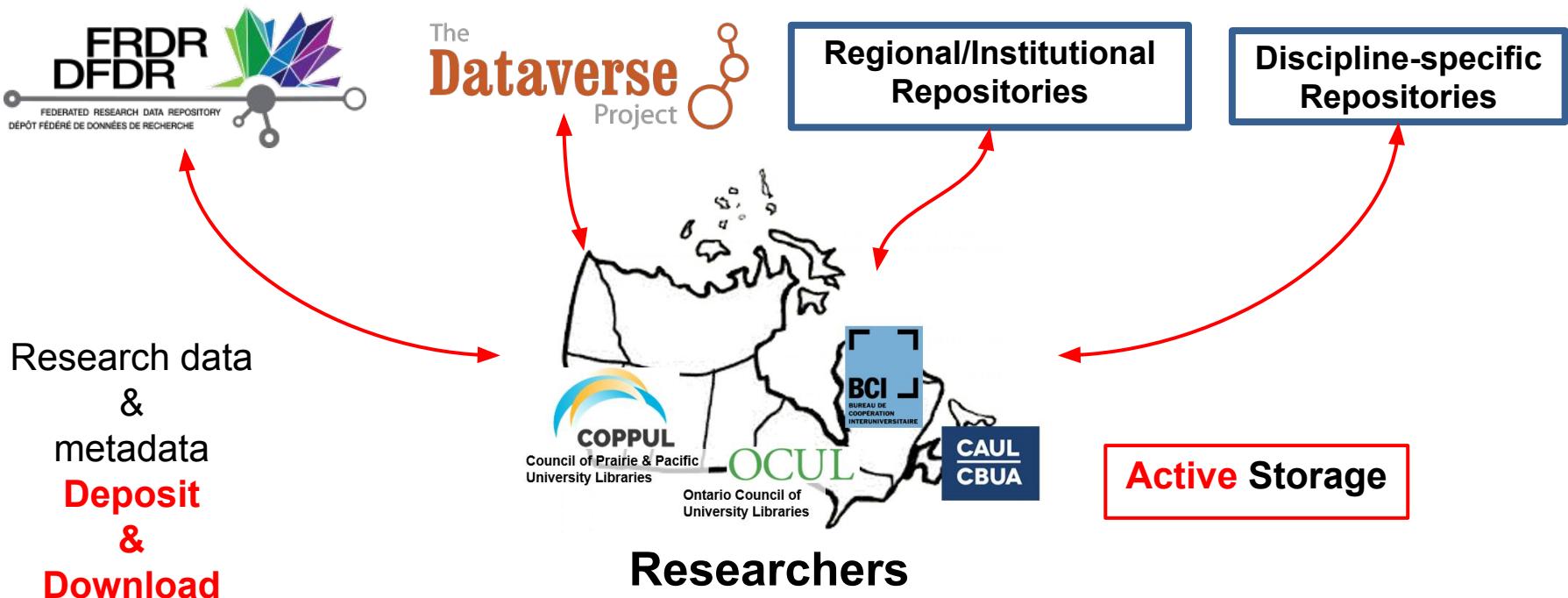
*A scalable, federated platform for digital research data management and the discovery of Canadian research data*

**Discovery**

**Deposit**

**Preservation**

# Canadian Repository Landscape



# Canadian Repository Landscape

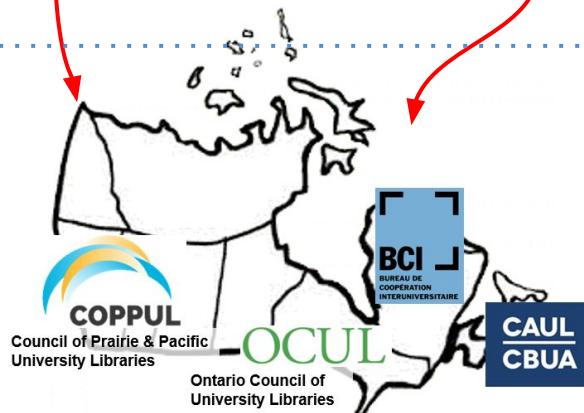
Repository Storage



Regional/Institutional  
Repositories

Discipline-specific  
Repositories

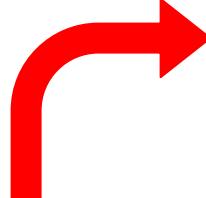
**Metadata**  
harvested from  
**45+** repositories  
indexing  
over **100,000**  
datasets



Researchers

Active Storage

# National Discovery Service



Repository Storage

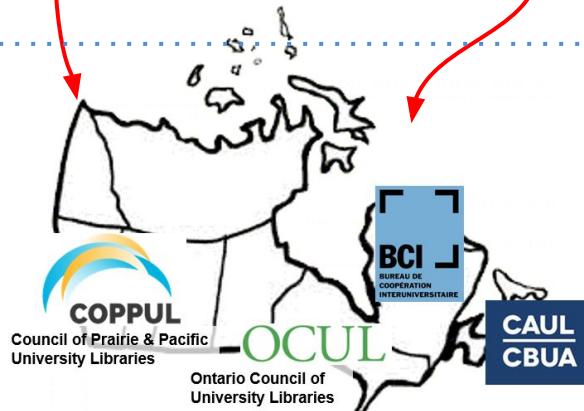


The  
**Dataverse**  
Project

Regional/Institutional  
Repositories

Discipline-specific  
Repositories

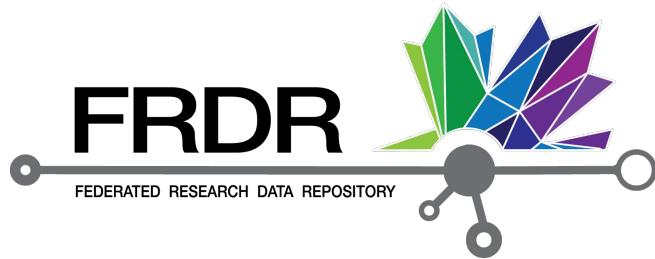
**Metadata**  
harvested from  
**45+** repositories  
indexing  
over **100,000**  
datasets



Researchers

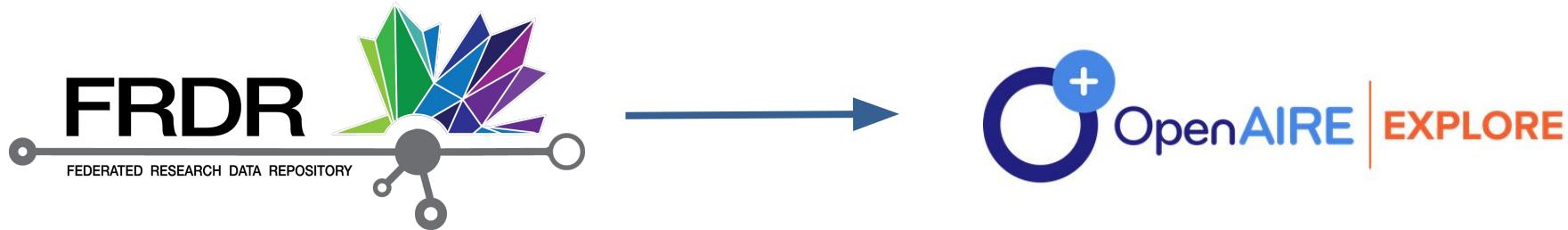
Active Storage

# National Discovery Service



- Improve discovery of research data/metadata across Canada
- Break down repository siloes
- Drive traffic to existing repository sites

# National Discovery Service



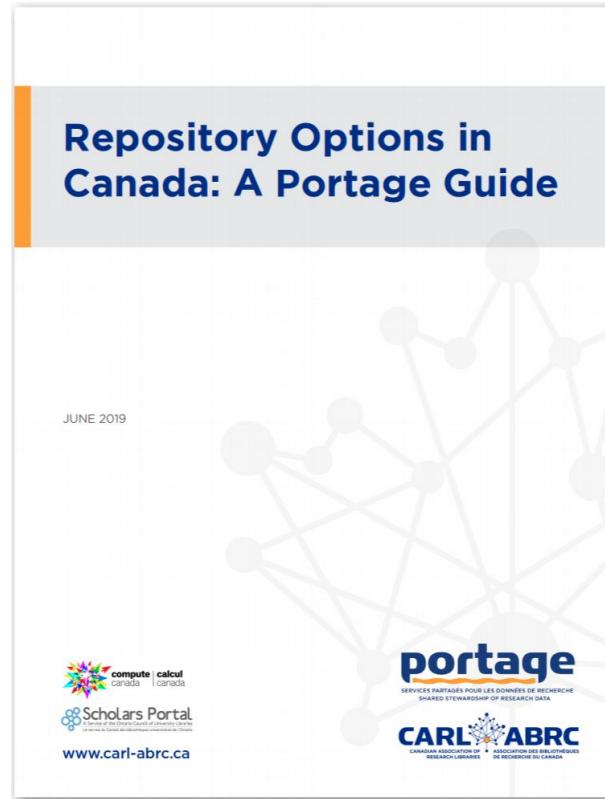
- Improve the discoverability of Canadian research data internationally
- Create interoperability between national and international platforms

<https://www.frdr.ca>

# Repository Options in Canada: A Portage Guide

- Developed collaboratively by Portage and OCUL Scholars Portal
- Provides an overview of the Canadian repository landscape
- Detailed information about FRDR and Scholars Portal Dataverse, including a breakdown of features and use-cases for when to recommend which platform
- Future updates will offer more information about domain-specific repositories

<https://portagenetwork.ca/portage-training-resources/>



# ***FRDR = Federated Research Data Repository (DFDR en français)***

<https://www.frdr.ca/>

*A scalable, federated platform for digital research data management and the discovery of Canadian research data*

**Discovery**

**Deposit**

**Preservation**

# Repository Overview

FRDR is a full-featured repository for research data publication:

- A place for Canadian researchers to deposit large datasets
- Designed for scalability (TB+ scale), leveraging Globus File Transfer
- Storage and compute geographically distributed at multiple hosting sites

# Repository Overview

- Direct deposit & download of large datasets through Globus
- Direct upload & download of small datasets through web browser
- Digital Object Identifiers (DOIs) through DataCite
- Embargo support (delayed publication)
- Support for custom metadata schemas and multiple data licenses
- ORCID integration (<https://orcid.org/>)
- Supports ‘Special Collections’

# Special Collection Page



UNIVERSITY OF  
**WATERLOO**



## Water Institute

Ranked among the top 10 water research institutes in the world, the [Water Institute](#) is a leader in water research and education. Distinguished by its commitment to 1) facilitating interdisciplinary collaboration, 2) inspiring innovation and 3) building international partnerships, the [Water Institute](#) is tackling some of the most complex water challenges of our time.

		Author(s)
Date ↑	Title ↑	
16-Apr-2019	<a href="#">Input and Output Variables for the biogeochemical modeling study of the Lakes Response to Climate Change, Internal Loads and Restoration Measures</a>	Markelov, Igor; Van Cappellen, Philippe; Couture, Raoul-Marie
3-Apr-2019	<a href="#">Reactive Silicon Data in Hamilton Harbour, Lake Ontario</a>	Ridenour, Christine; Van Cappellen, Philippe; Parsons, Chris; Marvin, Chris; Fitzpatrick, Mark
19-Mar-2019	<a href="#">Effects of Damming on River Nitrogen Fluxes: A Global Analysis</a>	Akbarzadeh, Zahra; Maavara, Taylor; Slowinski, Stephanie; Philippe, Van Cappellen
4-Feb-2019	<a href="#">Benthic nitrite exchanges in the Seine River (France)</a>	Akbarzadeh, Zahra; Laverman, Anniet; Rezanezhad, Fereidoun; Raimonet, Mélanie; Viollier, Eric; Shafei, Babak; Van Cappellen, Philippe

# FRDR in Limited Production

**Portage and CCF launched Limited Production Sep 21, 2017**

**Limited Production means:**

- Anyone can use FRDR to:
  - Search across harvested Canadian repositories
  - Download datasets stored in FRDR
- Data deposit is restricted to a select number of research groups
  - Portage commits to making their data discoverable and accessible into and beyond the full service launch
  - Groups working with us to refine our service ahead of production release

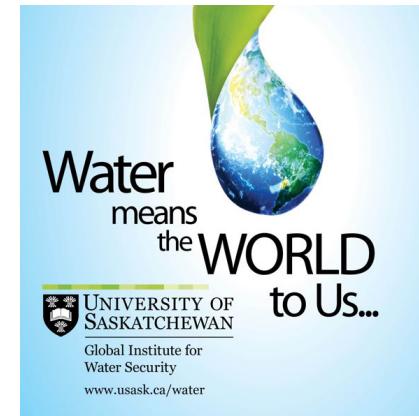
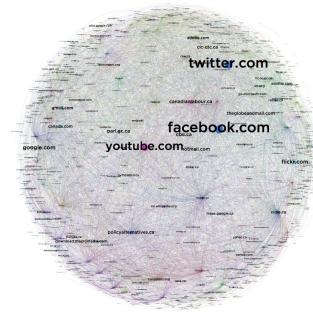
# FRDR Partners



*“Big data solutions  
for the marine  
acoustic community”*

*Dalhousie University*

Web Archives for  
Longitudinal  
Knowledge



The Water Institute  
*U. Of Waterloo*

*“Breaking boundaries in water  
research”*



*Canadian Array for Realtime  
Investigations of Magnetic Activity*

*U. Alberta*



**S.D.S.**  
STONE TOOLS, DIET, AND SOCIALITY

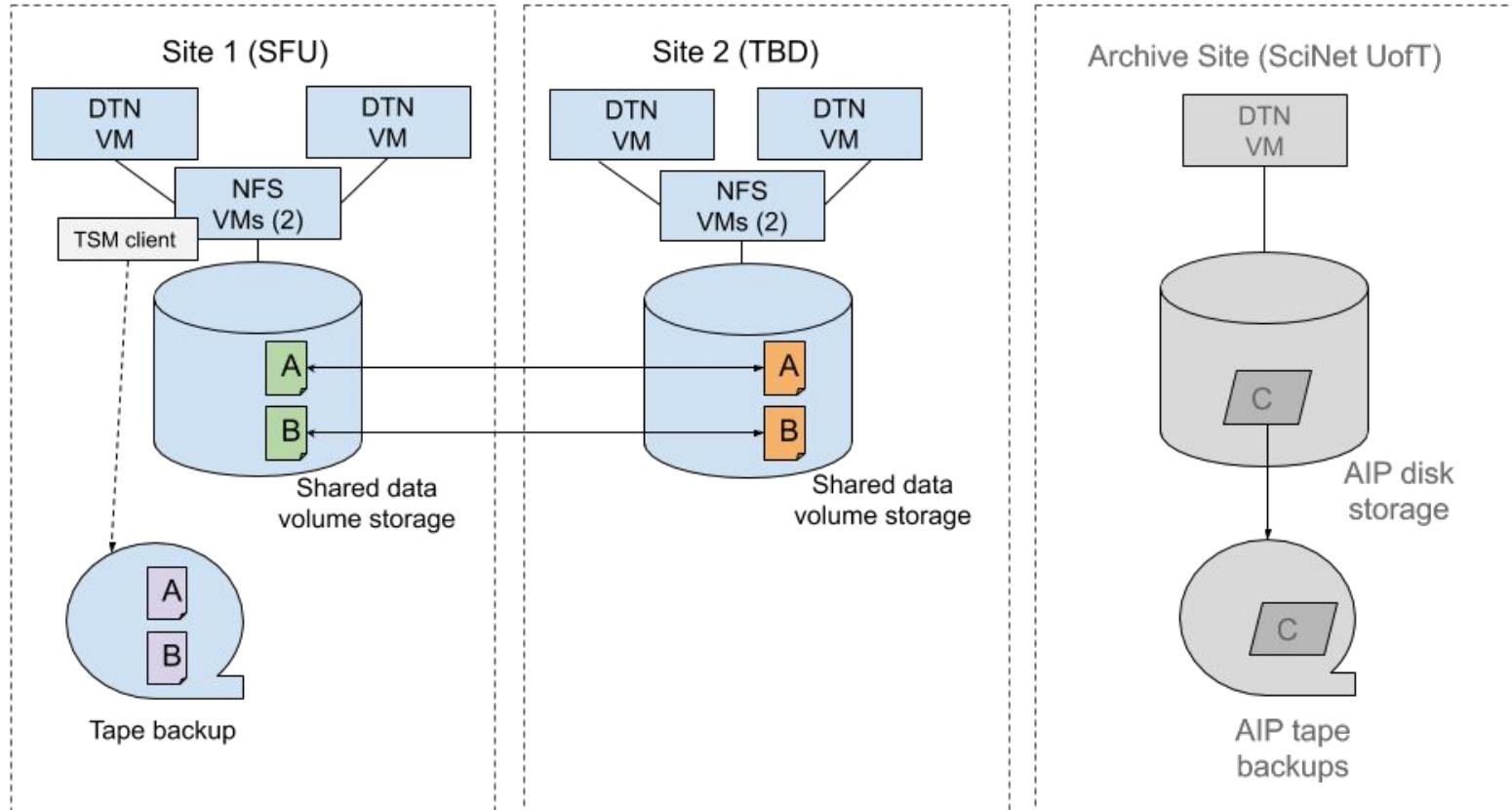
*U. Calgary*

*“Capturing Change in  
Canada’s Mountains”*  
*U. Victoria*

# FRDR Transitioning to Full Production

- \$2M of Directed Funding awarded to CARL Portage by CANARIE to strengthen national RDM activities ahead of the New DRI Organization
- Aim is to transition FRDR from limited to full production by Sept 30, 2020
- Funding will support storage hardware and system administrators at two national CCF hosting sites
- Hiring a new developer position and Technical Operations Manager

# FRDR Production Architecture



# FRDR as a National Service

- Canadian researchers (PIs) or their designates from eligible Canadian institutions (institutions able to apply for grant funding) can deposit
- Proposed storage allocation model (**for open data publication**):
  - 1-3 TB available to PIs by default
  - 3-10 TB available upon request, requires some justification, subject to curatorial review and resource availability
  - Requests for >10 TB will need to undergo an adjudication process and will be subject to curatorial review and resource availability
  - Managed repository storage may also be purchased by a large organization or institution

# Ramping up to Service Launch

- Informational campaign:
  - Who is eligible to use FRDR?
  - How do storage allocations work for FRDR?
  - How to access and use the platform?
- Work with Network of Experts (TEG/FRDR UX WG) to develop training materials, webinars, and an online module
- Interim governance model & FRDR policy approval
- Work with host site partners to bring new storage online and define service level expectations

# Thanks!

Lee Wilson

[lee.wilson@ace-net.ca](mailto:lee.wilson@ace-net.ca)

<https://www.frdr.ca/>



# **Scholars Portal Dataverse**

**Meghan Goodchild**

Research Data Management Systems Librarian  
Scholars Portal/Queen's University

**Kaitlin Newson**

Digital Projects Librarian  
Scholars Portal

# What is Dataverse?

- Open-source repository platform developed by the Institute of Quantitative Social Science (IQSS) at Harvard University
- 48 installations around the world
- Active development community
- Developed with quantitative social science data in mind, but not limited to any domain



# Scholars Portal Dataverse

The image shows the Scholars Portal Dataverse homepage. The background is a light blue color with a white network graph overlay, representing interconnected data. At the top, there is a navigation bar with links: BENEFITS, FEATURES, PARTNERS, USE CASES, FAQ, CONTACT, LOGIN, and FR. Below the navigation, the text "Scholars Portal Dataverse" is displayed in dark blue and red. The main headline "Store, share, publish and discover research data!" is prominently shown in large white text. At the bottom left, there are two dark blue buttons: "TELL ME MORE" and "EXPLORE DATAVERSE", separated by the word "or". Below these buttons is a search bar with the placeholder "Search Dataverse" and a magnifying glass icon.

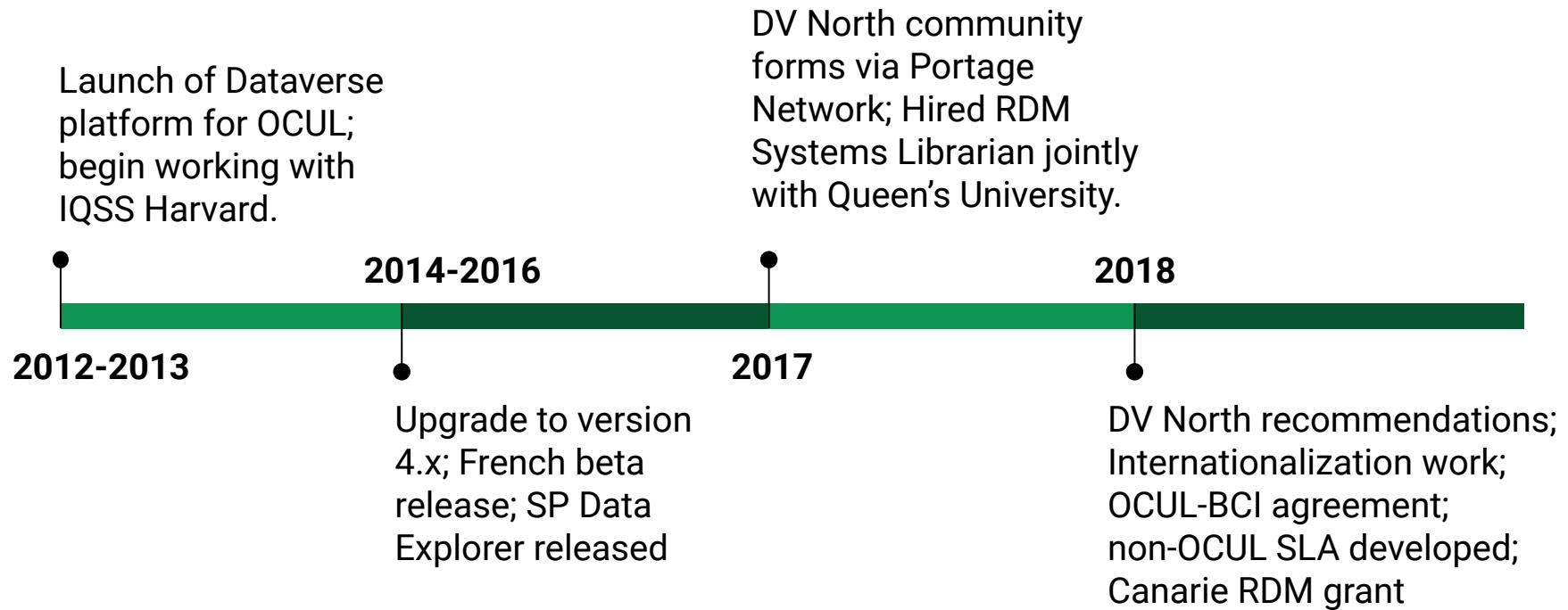
Scholars Portal Dataverse

**Store, share, publish and  
discover research data!**

TELL ME MORE    or    EXPLORE DATAVERSE

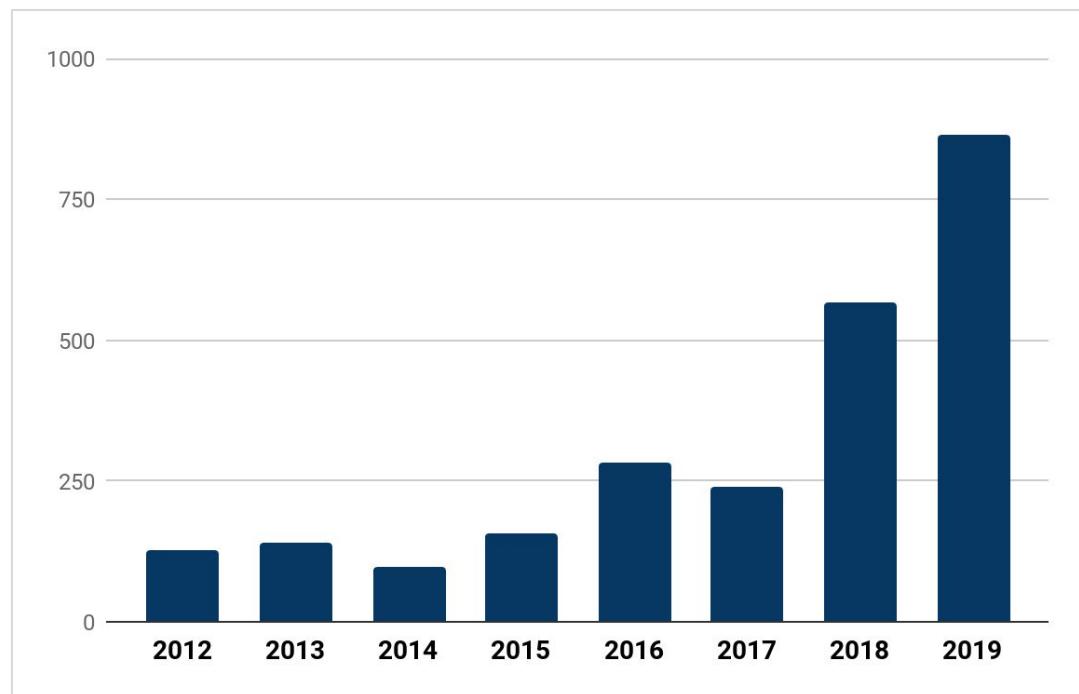
Search Dataverse

# Background - SP Dataverse Milestones



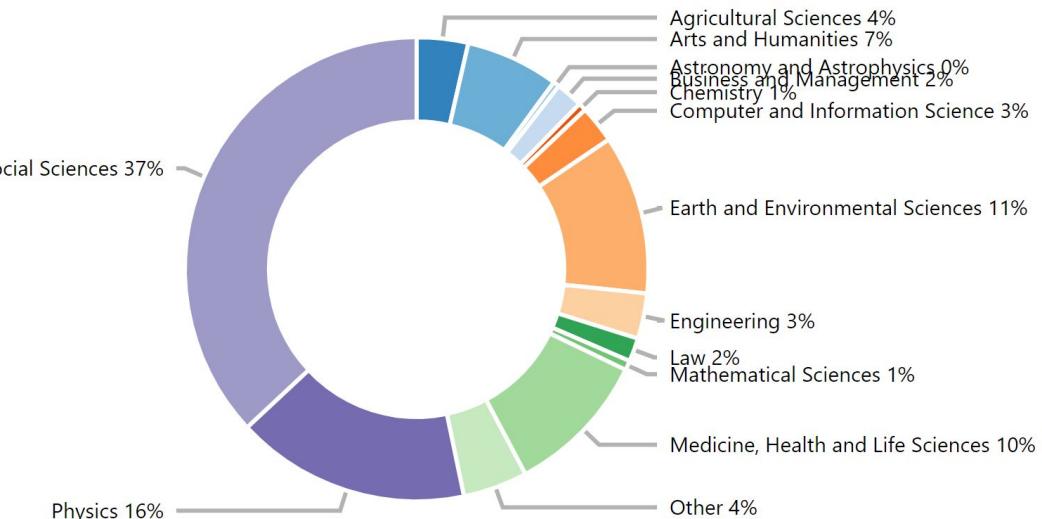
# Background - Published datasets per year

- Slow growth initially
- Rapid increase over past 2 years
- Increased interest in large files, large datasets, other enhancements



# Background - Subject categories

- Social sciences 60+% initially
- Increase in other disciplines deposits
- Increase in audio, video, and image files



# Dataverse features that enable FAIR

- Apply data reuse licence (CC0 default) + templates
- Rich metadata description possible (compliant with standards)
- User-friendly upload of data files, code, documentation
- DOIs minted via Datacite
- Automatic file format identification, file verification (checksums), tabular data transformation
- Dataset and file versioning
- OAI-PMH harvesting protocol; APIs for access, search & upload (SWORD)

# Institutional Dataverses

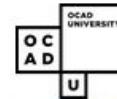
- All data organized by institution
- Library administers Institutional Dataverse space
- Researchers deposit in Institutional Dataverses defined by user affiliation
- Customizable features (branding, featured Dataverses, facets, etc.)



McMaster University Dataverse



Nipissing University Dataverse



OCAD University Dataverse



Ontario Tech University Dataverse



# Institutional branding

Scholars Portal Dataverse

Search ▾ User Guide Support English ▾ Sign Up Log In



Dataverse de l'Université Laval (Université Laval)

Pour déposer des données, consultez notre guide éclair.

Scholars Portal Dataverse > Dataverse de l'Université Laval

[✉ Contact](#) [↗ Share](#)

Ce Dataverse est un dépôt de données institutionnel réservé à la communauté de l'Université Laval. Pour y déposer des données ou pour toute question concernant la gestion des données de recherche communiquez avec nous par courriel à [gdr@bibl.ulaval.ca](mailto:gdr@bibl.ulaval.ca). Vous pouvez également en apprendre davantage en consultant notre page Web à Gestion des données de recherche.

Search this dataverse...



Find Advanced Search

Dataverses (3)

Datasets (2)

Files (183)

Dataverse Category

Chercheur (2)

1 to 5 of 5 Results

Sort ▾

Impact of thinning in naturally regenerated balsam fir stand

Aug 19, 2019 - Marine Duperat Dataverse



Duperat, Marine; Ruel, Jean-Claude; Gardiner, Barry, 2019, "Impact of thinning in naturally regenerated balsam fir stand", <https://doi.org/10.5683/SP2/WZIKSR>, Scholars Portal Dataverse, V1, UNF:6.BZnChBgvrFP6ysC6NQwNmge=[fileUNF]

Data from the submitted article : Impact of thinning in naturally regenerated balsam fir stand.

# Modular Applications / Integrations

Modular configuration of the Dataverse code + APIs for data access allow for custom developments

SP developments include:

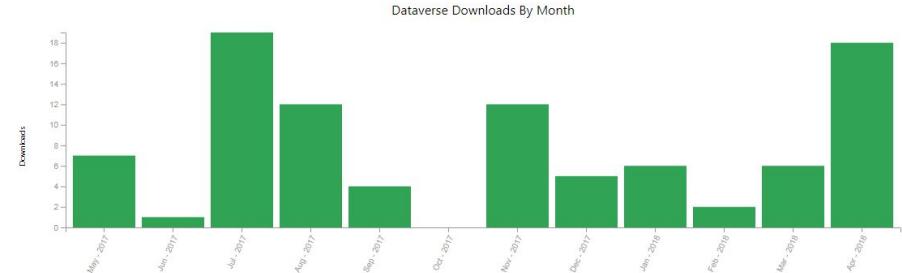
- Online metrics report
- Data Explorer
- Internationalization

Dataverse Report

McMaster University Dataverse ▾

May - 2017 to Apr - 2018

Download

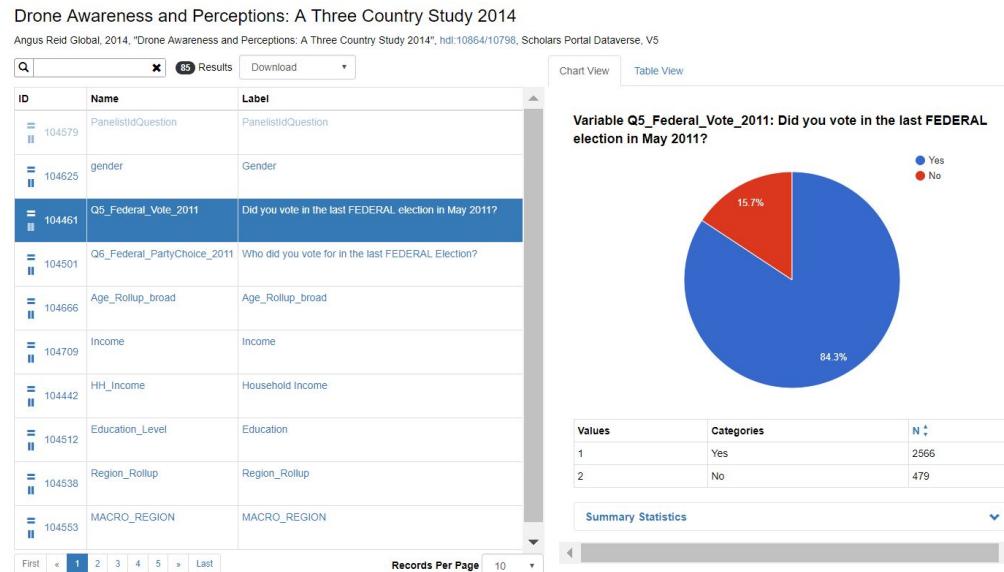


<https://dataverse.scholarsportal.info/metrics/>

Other integrations: OSF, Archivematica, TwoRavens, WorldMap, OJS 2.x, and others...

# Data Explorer

- Online tool to explore tabular data files
- Chart view and table view
- Developed at Scholars Portal
- Installed by other Dataverse instances



# Data Explorer

## Forum Research Political Poll - Provincial Issues (Quebec) 2018

Français

Forum Research Inc, 2018, "Forum Research Political Poll - Provincial Issues (Quebec) 2018", <https://doi.org/10.5683/SP2/8NLNFX>, Scholars Portal Dataverse, V1, UNF:6:M3oloqagYn7P3ouGyQD4lg==

x **23** Results [Download](#)

Chart View

Table View

229117	Q8B	Q8B. What party did you vote for in the last Quebec provincial election held in April 2014?	5
229107	D1	D1. What is your gender?	3
229100	D2	D2. How old are you?	6
229096	D3	D3. Do you have children under the age of 18 in the household?	2
229113	D4	D4. What is the highest level of education you completed?	4
229106	D5	D5. Have I reached you on a landline or on a cell phone?	2
229116	D6	D6. In addition to your cell phone, do you also have a landline in your household or not?	2

⊕ D2. How old are you?	⊕ 5. Regardless of which party you would vote for, who do you think would make the best Premier of Quebec?	Under 25 years	25 to 34	35 to 44	45 to 54	55 to 64	65 and over	Total	N
		Philippe Couillard	0.9	0.7	2.1	3.8	4.2		
Jean-Francois Lisée	0.9	2.2	2.5	2.5	3.3	5.8	17.2	166	
Francois Legault	0.6	2.1	4.7	7.0	8.5	12.7	35.6	344	
None of the above	1.1	3.6	4.0	3.3	3.2	3.3	18.7	180	
Don't know	0.2	1.3	0.6	1.5	1.8	2.3	7.7	74	
<b>Total</b>	<b>3.8</b>	<b>9.9</b>	<b>13.9</b>	<b>18.1</b>	<b>21.0</b>	<b>33.2</b>	<b>100.0</b>		
<b>N</b>	<b>37</b>	<b>96</b>	<b>134</b>	<b>175</b>	<b>203</b>	<b>320</b>			<b>965</b>

Z-Score **-2<** <0< **>2**

# Internationalization

Scholars Portal Dataverse

Fichiers Métadonnées Conditions Versions

Recherche ▾ À propos Guide d'utilisation Soutien Français ▾ English ▾ Amber Leahy ▾

Métadonnées de la référence bibliographique ▾

Identifiant pérenne de l'ensemble de données	doi:10.5072/FK2/CYAV9Y
Date de publication	2018-10-18
Titre	Dataset #1
Auteur	Vezina, Marie-Helene (U d
Personne-resource	Utiliser le bouton de cou
Description	Vezina, Marie-Helene (U d
Sujet	Astronomie et astrophysiqu
Langue	Assamais; Aragonais; Arab
Collaborateur	Chargé de la collecte de d
Déposant	Vezina, Marie-Helene
Date de dépôt	2018-10-18

Métadonnées géospatiales ▾

 **Dataverses (112)**

 **Ensembles de données (137)**

 **Fichiers (437)**

Collaborateur

Type

Chargé de la collecte de données

Sélectionner...

Chargé de la collecte de données

Intendant des données

Gestionnaire de données

Éditeur

Bailleur de fonds

Établissement hôte

Chef de projet

Nom

Nom, prénom ou nom de l'organisation

Numéro de la subvention

Affiliation

Adresse URL

Entrer l'adresse URL complète commen

Adresse URL du logo

Entrer l'adresse URL complète de l'imag

+ Ajouter des données ▾

1 à 10 de 249 résultats

Dataset #1

Oct 18, 2018 - Amber Leahy Dataverse

Tri ▾

# CANARIE RDM grant program funding

“Dataverse for the Canadian Research Community: Developing reusable and scalable tools for data deposit, curation, and sharing”

- Led by Scholars Portal and UTL with support from CARL and Portage
- November 2018 - March 2020
- Aim is to enhance Dataverse to address needs of a broad range of researchers in Canada:
  - Improved scalability and support for large(r) files
  - Integrations with Canadian cloud storage and authentication providers
  - Better support for data curation workflows (Data Curation Tool)
- Check out [our blog post](#)



# CANARIE project roadmap

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

# Data Curation Tool

- Aim to improve support for data curation workflows and help adopt standards and best practices
- External web application that connects to Dataverse to create and edit metadata at the variable level
- Works with files uploaded through tabular ingest process
- Blog post: [Introducing the Data Curation Tool](#)

# Data Curation Tool

Labour Force Survey (Curated)						
LFS2016-01_PUMF_EN.tab						
Tester, Curation, 2019, "Labour Force Survey (Curated)", <a href="https://doi.org/10.5072/FK2/YLJJAY">https://doi.org/10.5072/FK2/YLJJAY</a> , Scholars Portal Dataverse, V1, UNF:6:EerkQFr2ySzwCu4oV5jExA== [fileUNF]						
<a href="#">Hide Groups</a>	<a href="#">Add Group</a> +					<a href="#">Download</a> <a href="#">Save</a>
All Variables	<input type="checkbox"/> Search					Items per page: 25 <a href="#">▼</a> 1 - 25 of 75 <a href="#">◀</a> <a href="#">▶</a>
Number of hours	<input type="checkbox"/>	ID	Name	Label	Weight	View
	<input type="checkbox"/>	v14110	REC_NUM	Order of record in file		
	<input type="checkbox"/>	v14130	SURVYEAR	Survey year		
	<input type="checkbox"/>	v14106	SURVMNTH	Survey month		
	<input type="checkbox"/>	v14127	LFSSTAT	Labour force status		
	<input type="checkbox"/>	v14165	PROV	Province		
	<input type="checkbox"/>	v14168	CMA	3 largest CMAs		
	<input type="checkbox"/>	v14155	AGE_12	Five-year age group of respondent		
	<input type="checkbox"/>	v14136	AGE_6	Age in 2 and 3 year groups		

# 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](#)

Add Group



Search

All Variables

<input type="checkbox"/> ID	Name
<input type="checkbox"/> v14158	SEX
<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

Sex of respondent

[Download](#)

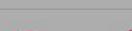
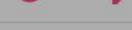
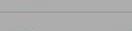
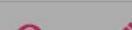
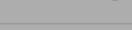
[Save](#)

Items per page: 25

1 - 25 of 75

Weight

View



# 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 [edit](#)

ID Name

v14110 REC\_N

v14130 SURVY

v14106 SURVM

v14127 LFSSTA

v14165 PROV

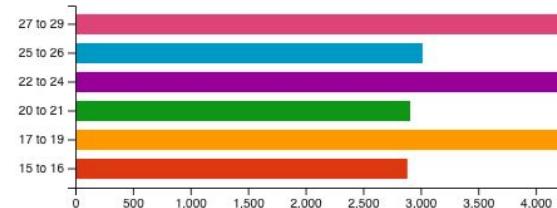
v14168 CMA

v14155 AGE\_12

v14136 AGE\_6

v14158 SEX

## AGE\_6: Age in 2 and 3 year groups



Values	Categories	Count	Count Percentage(%)	Weighted Count
1	15 to 16	2,885	13.417	
2	17 to 19	4,224	19.644	
3	20 to 21	2,909	13.528	
4	22 to 24	4,219	19.621	
5	25 to 26	3,017	14.031	
6	27 to 29	4,249	19.76	

Age in 2 and 3 year groups

Sex of respondent

# Data Curation Tool Codebook

**Data Curation Tool Testing Dataset (ICPSR doi:10.5072/FK2/0TYIHL)**  
**(DCT Testing Dataset)**

**View:** [Part 1: Document Description](#)  
[Part 2: Study Description](#)  
[Part 3: Data Files Description](#)  
[Part 4: Variable Description](#)  
[Part 5: Other Study-Related Materials](#)  
[Entire Codebook](#)

## Document Description

### Citation

*Title:* Data Curation Tool Testing Dataset  
*Identification Number:* doi:10.5072/FK2/0TYIHL  
*Distributor:* Scholars Portal Dataverse  
*Date of Distribution:* 2019-06-14  
*Version:* 2  
*Bibliographic Citation:* Lubitch, Victoria; Leahey, Amber, 2019, "Data Curation Tool Testing Dataset", <https://doi.org/10.5072/FK2/0TYIHL>

## Study Description

### Citation

*Title:* Data Curation Tool Testing Dataset  
*Subtitle:* DDI Test  
*Alternative Title:* DCT Testing Dataset  
*Identification Number:* doi:10.5072/FK2/0TYIHL  
*Authoring Entity:* Lubitch, Victoria (University of Toronto)  
Leahey, Amber (Scholars Portal)  
*Producer:* Leahey, Amber  
*Date of Production:* 2019-05-22  
*Grant Number:* 4445555  
*Distributor:* Scholars Portal Dataverse  
*Date of Distribution:* 2019-06-14

## Study Scope

*Keywords:* Astronomy and Astrophysics, test, smoking  
*Topic Classification:* Metadata

# Try it out!

bit

Code on GitHub:

<https://github.com/scholarsportal/Dataverse-Data-Curation-Tool>

Available in Scholars Portal Dataverse October 31st.

# Portage Network's Dataverse North (DVN)

Business Models Working Group key recommendations:

- Establish National Dataverse Service
  - Hosted by SP/UTL
  - Sustainable library service
- Coordinate governance through Portage Network DVN

Guiding benefits/principles:

- Realizing true economies of scale
- Pooling resources
- Strengthen technical capacity
- Improve funding opportunities
- Support local and regional efforts



# Building a national service

## *National multi-stakeholder engagement*

1. Developing partnerships and agreements with other regional library consortia
  - a. Agreement with BCI to extend service to 19 university libraries in Quebec
  - b. Working on agreements with COPPUL and CAUL
2. Engaging with Portage's Dataverse North Working Group
  - a. Policy subgroup
3. Seeking national funding opportunities

# Building a national service

## *Local connections and touchpoints*

1. Scholars Portal Dataverse community
  - a. Support email listserv
  - b. Monthly meetings
  - c. Testing and feedback
2. Training and outreach to help build local capacity
  - a. Introductory [YouTube videos](#)
  - b. [User Guide](#) with updated FAQ
  - c. New [Scholars Portal Dataverse Blog](#)
  - d. Dataverse North Training Group [Materials](#)

# Building a national service

Key projects funded by this investment:

- Federated Research Data Repository & National Discovery Service (in partnership with Compute Canada Federation including regional partners, University of Saskatchewan, Simon Fraser University, and University of Toronto);
- National Instance of Dataverse (in partnership with OCUL Scholars Portal, University of Toronto)
- Data Management Planning Assistant (in partnership with University of Alberta and CASRAI).

**CANARIE announces \$2M in funding to CARL Portage and its partners to strengthen Research Data Management ecosystem in Canada**

Posted on **August 27, 2019** in **Press Releases**

<https://www.canarie.ca/canarie-announces-2m-in-funding-to-carl-portage-and-its-partners-to-strengthen-rdm-ecosystem-in-canada/>

# Considerations for building a national service

- Lowering barriers to access for researchers
- Building capacity locally and at Scholars Portal
- Defining national governance structure and policies
- Strengthening Canadian Dataverse community

# Get in touch

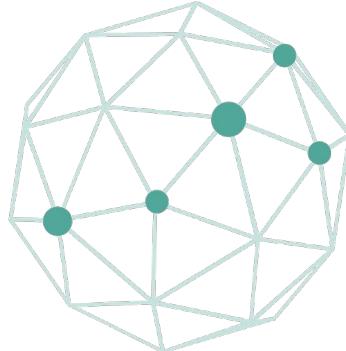
Dataverse team: [dataverse@scholarsportal.info](mailto:dataverse@scholarsportal.info)

Meghan Goodchild: [meghan@scholarsportal.info](mailto:meghan@scholarsportal.info)

Kaitlin Newson: [kaitlin@scholarsportal.info](mailto:kaitlin@scholarsportal.info)

# CIOOS

CANADIAN INTEGRATED OCEAN  
OBSERVING SYSTEM



SYSTÈME INTÉGRÉ D'OBSERVATION  
DES OCÉANS DU CANADA

Oct 18, 2019

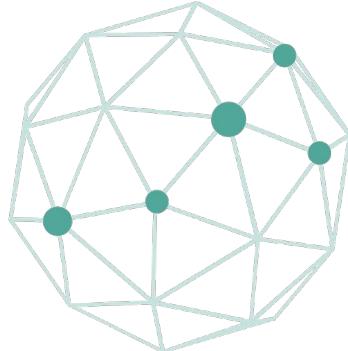
## CIOOS Update Canadian Data Curation Forum

Brad Covey, Reyna Jenkyns

[coveyb@ogsl.ca](mailto:coveyb@ogsl.ca), [reyna@oceannetworks.ca](mailto:reyna@oceannetworks.ca)

# CIOOS

CANADIAN INTEGRATED OCEAN  
OBSERVING SYSTEM

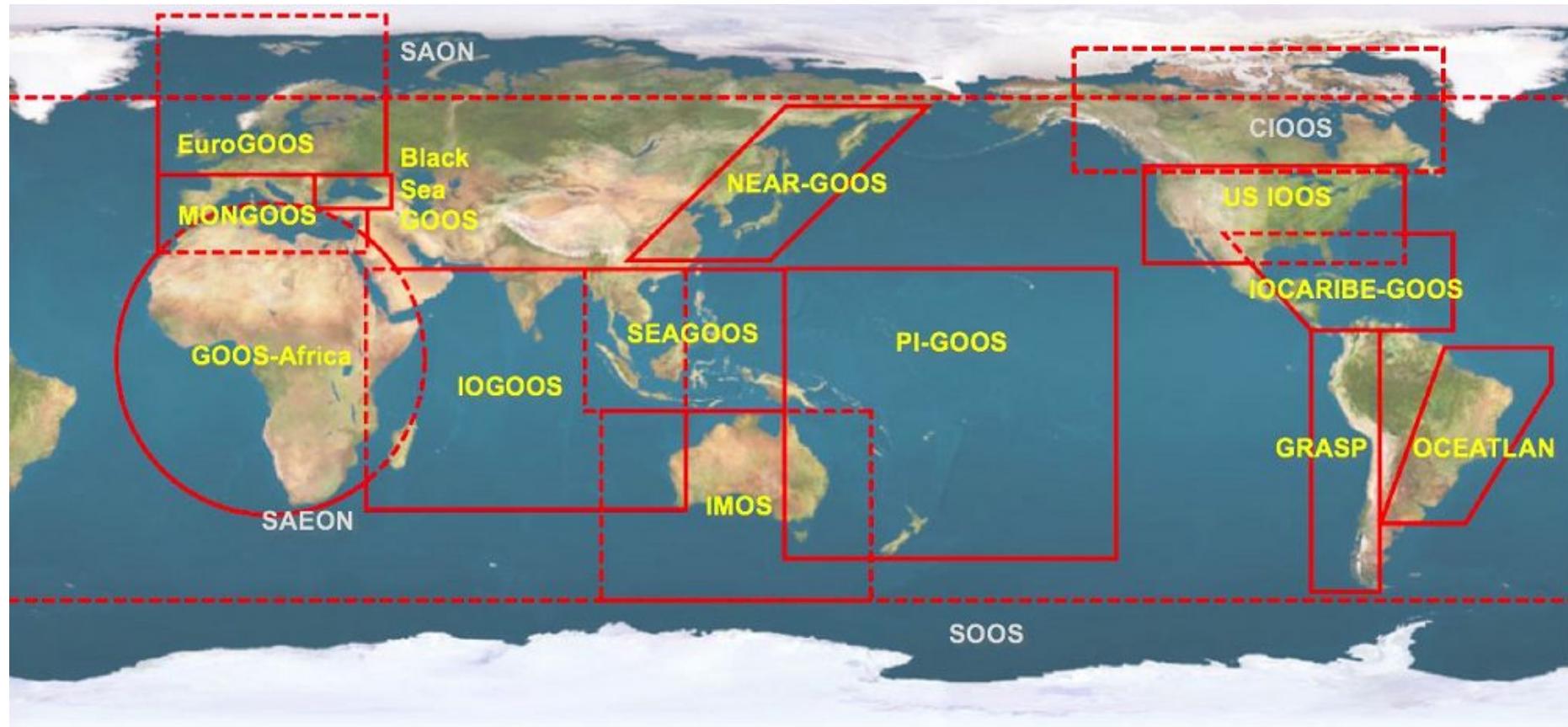


# SIOOC

SYSTÈME INTÉGRÉ D'OBSERVATION  
DES OCÉANS DU CANADA

Discussion Questions

<http://bit.ly/2VOFCV6>



# Stage t-4: Community of Practice on Ocean Data Management (ODM CoP)

The CoP on Ocean Data Management connects Canadian ocean data centres to share and mobilize expertise and best practices, promote cooperation and alignment, and develop a shared vision for ocean data management in Canada.

<http://www.oceansdata.ca>



Pêches et Océans  
Canada

Fisheries and Oceans  
Canada

# Stage t-3

- Expert Forum in  
Montreal, Nov 2015



## Towards a Unified Vision for Ocean Data Management in Canada: Results of an Expert Forum

Published May, 2016



# Stage t-2

**Task Team** of federal and non-federal ocean observing experts formed (November 2016)

MEOPAR renewed for five years with advancing ocean data in Canada a significant part of their strategic objectives.



# Stage t-1

DFO funded **Investigative Evaluations (IEs)** to move CIOOS from the concept state to the design stage

- Observation and Data
- Cyber Infrastructure
- Visualization Tools and Platforms

**Integrated Ocean Observing System**  
Investigative Evaluations  
Cyberinfrastructure  
2017-11-30

**PRINCIPAL INVESTIGATORS:**  
Richard Kelly<sup>1</sup>, Mike Smil<sup>1</sup>

**PROJECT MANAGER:**  
Shayla Fitzsimmons<sup>2</sup>

**CONTRIBUTORS:**  
Scott Bruce<sup>1</sup>  
Craig Bulger<sup>1</sup>  
Brad Covey<sup>1</sup>  
Richard Davis<sup>3</sup>  
Ryan Gosse<sup>3</sup>  
Dwight Owens<sup>3</sup>  
Benoit Pironno<sup>2</sup>

**AFFILIATIONS:**  
<sup>1</sup>Fisheries and Marine Institute of Memorial University of Newfoundland  
<sup>2</sup>School of Information Management, Dalhousie University  
<sup>3</sup>Department of Oceanography, Dalhousie University  
Ocean Networks Canada<sup>2</sup>, University of Victoria

**MARINE INSTITUTE**  
**MEMORIAL UNIVERSITY**  
**DALHOUSIE UNIVERSITY**  
**University of Victoria**



# Request for Proposals: Canadian Integrated Ocean Observing System

# CANADIAN INTEGRATED OCEAN OBSERVING SYSTEM

**DEADLINE: June 7, 2018 at 4PM Atlantic.** Submit as a PDF to Alexa Reedman at [alexa.reedman@meopar.ca](mailto:alexa.reedman@meopar.ca)  
All interested applicants must contact Alexa Reedman before submitting their proposal.



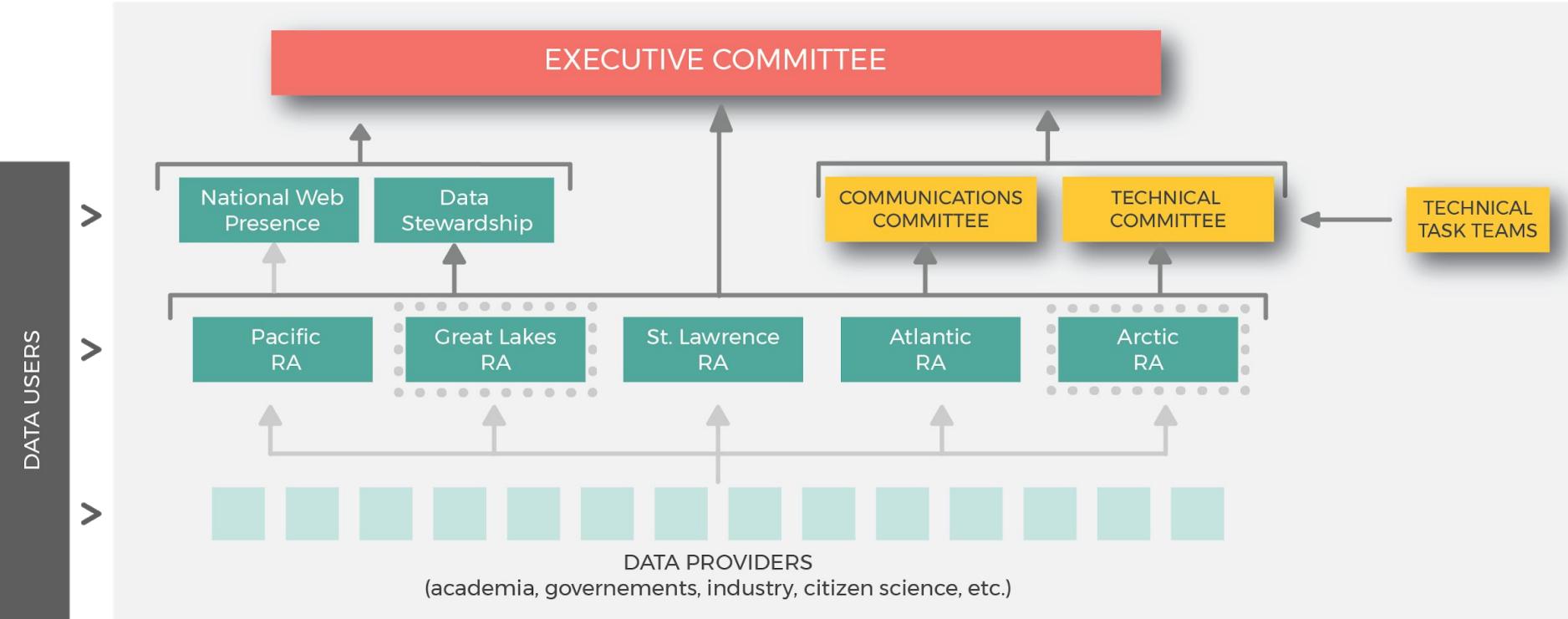
# CIOOS Face2Face Kickoff Victoria, BC January, 2019



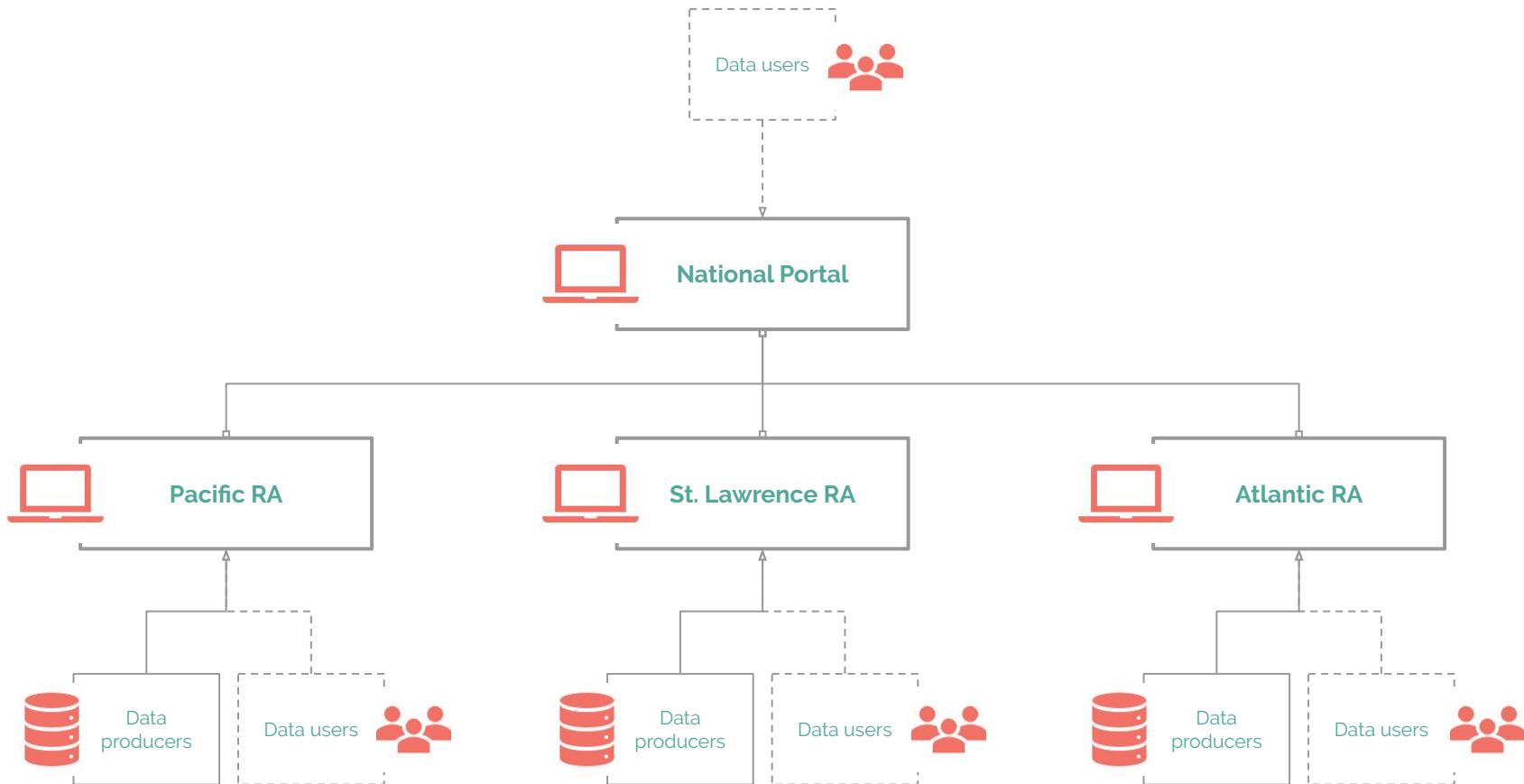
# CIOOS

CANADIAN INTEGRATED  
OCEAN OBSERVING SYSTEM

Functional / Reporting link  
Data / Metadata Aggregation link  
Future Regional Association

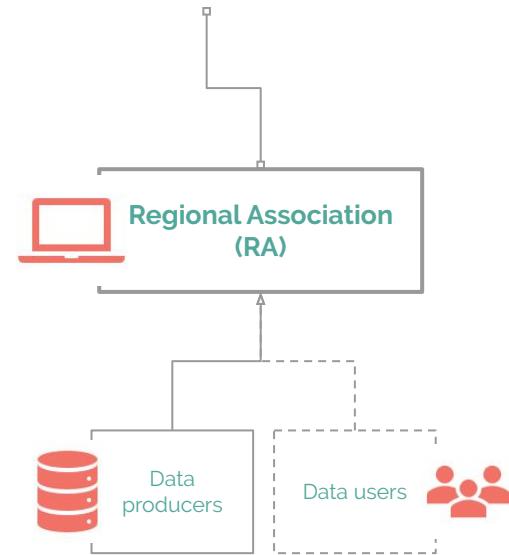


# About CIOOS



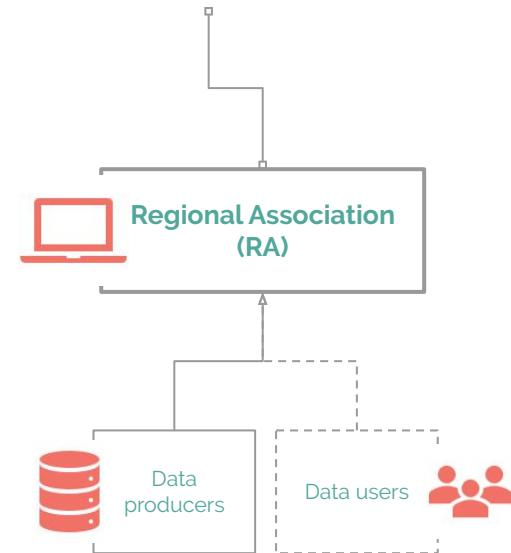
# Outreach and Engagement

- ✓ Meet regional data producers (academic, private, NGOs, ...)
- ✓ Meet regional data users
- ✓ Foster their participation in CIOOS
- ✓ Communicate and collaborate continuously with partners and other RAs.  
*+ data = + robust information system*



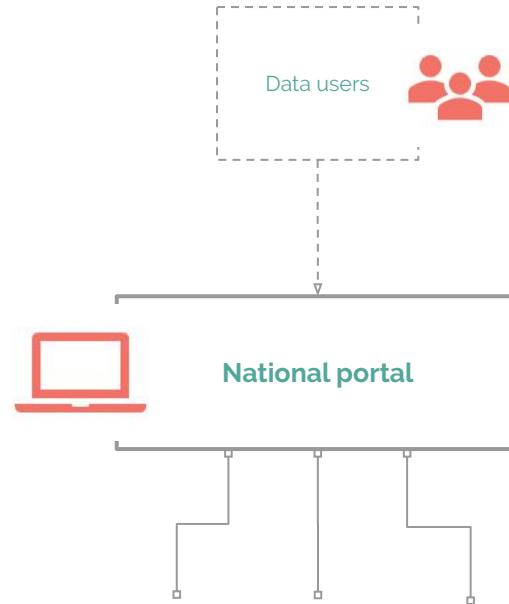
# Data and metadata

- ✓ Adopt metadata standards common to the 3 RAs
- ✓ Develop a metadata validation tool
- ✓ Aggregate data for access and download
- ✓ Visualization, search, and filter tools
  - Cartographic visualization tool common to the 3 RAs;
  - Regional and inter-regional data catalogue (ex. common dataset for St. Lawrence & Atlantic RA)



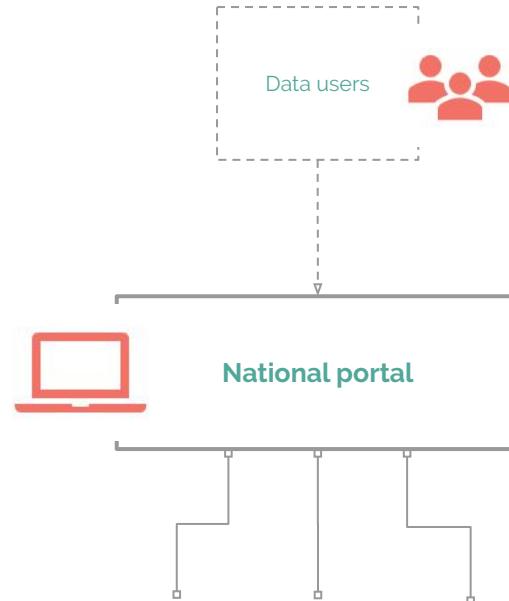
# National Portal

- ✓ Meet and collaborate with national and international partners
- ✓ Develop a common look and feel between the 3 RAs and the national portal
- ✓ Aggregate metadata and redirect to each RA for data access
- ✓ Develop and maintain the national web presence



# National Portal

- ✓ The national web presence will include:
  - General information about CIOOS
  - Links to each of the RAs
  - Data discovery through a CKAN catalog with metadata aggregated from the RAs
  - Data discovery through a cartographic metadata visualization



# Essential Ocean Variables - Phase 1

## Primary

- Temperature
- Salinity
- Oxygen
- Water Level
- Currents
- Nutrients

## Derived/related

- Seawater Density
- Potential Temperature
- Potential Density
- Speed of Sound
- Nutrient Concentrations
  - Nitrate ( $\text{NO}_3^-$ ) & Nitrite ( $\text{NO}_2^-$ )
  - Phosphate ( $\text{PO}_4$ )
  - Silicic Acid ( $\text{Si(OH)}_4$ )

# Metadata Standards and Best Practices

**Scope:** ISO 19115 profile development, controlled vocabulary and terminology adoptions, persistent identifiers applications, timestamp and datum conventions, data policy/licensing guidance, dataset granularity definitions, metadata workflows and tools, quality control flags, and documentation guidelines.

## Current status:

- dataset granularity practices surveyed and recommendations determined (more flexibility for legacy datasets at this stage)
- ISO 19115 - profile definition nearing completion
- ISO 19157 - gaining familiarity for integration of data quality information

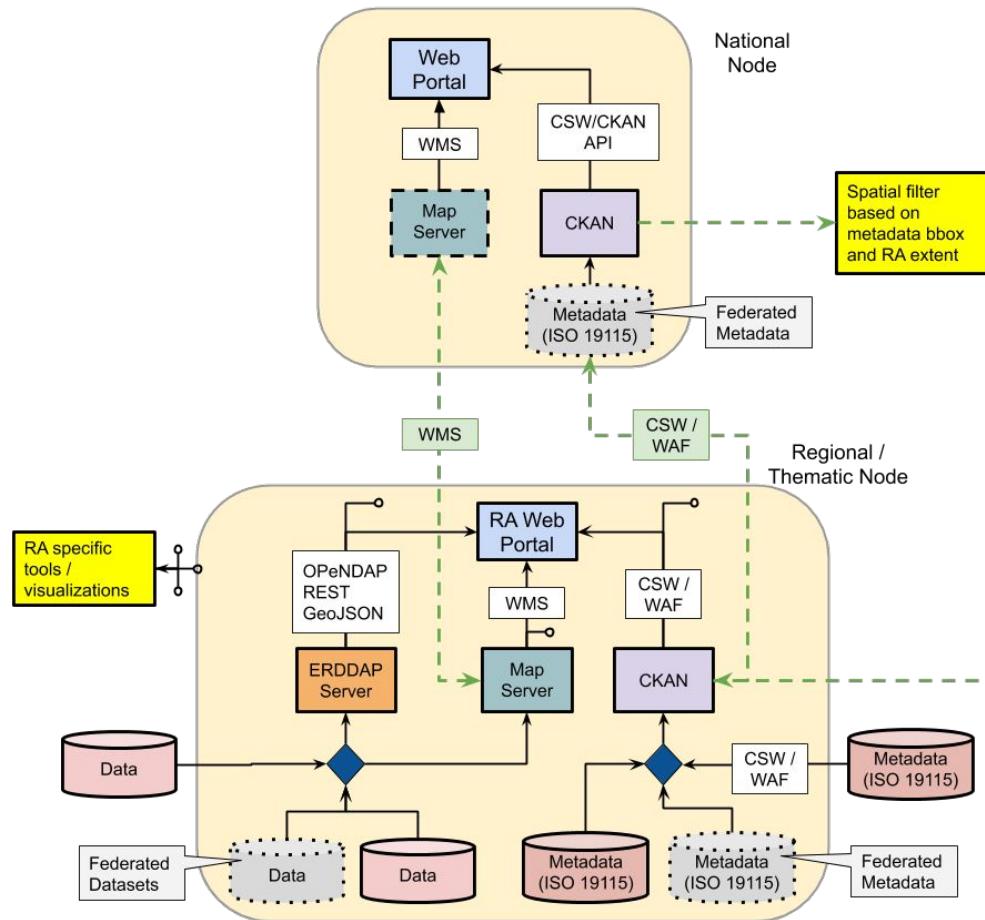
# Metadata Standards and Best Practices

## Specific Challenges

- bilingual support
- instrument/platform provenance and controlled vocabularies - relates also back to dataset granularity
- compatibility with North American Profile (NAP) which is based on the ISO 19115:2003 while we are using ISO 19115: 2014
- how to adequately and consistently represent lineage and data quality
- interplay with ERDDAP and CKAN

Focussed on instrument-based measurements, although physical samples for target EOVs in scope (e.g., O<sub>2</sub> from water samples). Human/machine observation/annotation based measurements out of scope for pilot phase.

# Solution Architecture



# Status

- Hardware allocated
- CKAN catalogues deployed
  - Bilingual support integrated
  - Still to be customized and populated

# Open Questions

- Catalogue aggregation
  - Customized CKAN aggregator?
  - External aggregation tool?
- Metadata creation tools
- Metadata validation tools - evaluating an extension of the IOOS compliance checker with our custom profile

# Asset map

- Goal is to direct the user to better tool / interface for data visualization / extraction
- Built on top of CKAN API with geoextent
- Embedded on the landing page of all Websites
  - National and each of the 3 RAs



# Use Case Development

- Personals with profiles and user stories
  - Early Career Researcher
  - Grad Student
  - Marine Pilot
  - Journalist
  - Application Programmer
  - Government Bureaucrat
  - DFO Officer
  - Maritime Forces Oceanographer

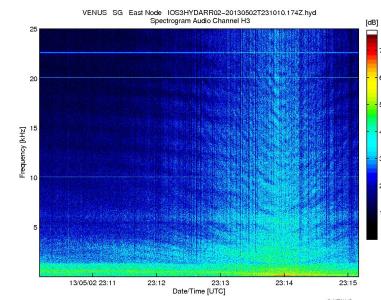


# Data Policy Development

- Using Framework of CoreTrustSeal requirements, beginning with Licences (R2), Continuity of Access (R3), Confidentiality/Ethics (R4), Appraisal (R8), Documented Storage Procedures (R9), Preservation Plan (R10), Data Discovery and Identification (R13), Data Reuse(R14)
- Raises questions with regard to which items require national agreement and which items are better suited to regionally (and even institutionally) varying approaches
- Framework for periodic revisals will be needed as CIOOS evolves

# Future Directions

- Expansion of Essential Ocean Variables and platforms
  - **biological variables** (e.g., OTN, OBIS, passive acoustics)
  - Gliders, High Frequency radar
- Inclusion of other 'data' (beyond instruments) like
  - model data,
  - observations (human expert, citizen scientist, automated)
  - physical samples (fluid, biological, geological)
- Continued refinement of **use cases** and **data products**
- Increased support for DataCite DOIs, ORCIDs, RORs

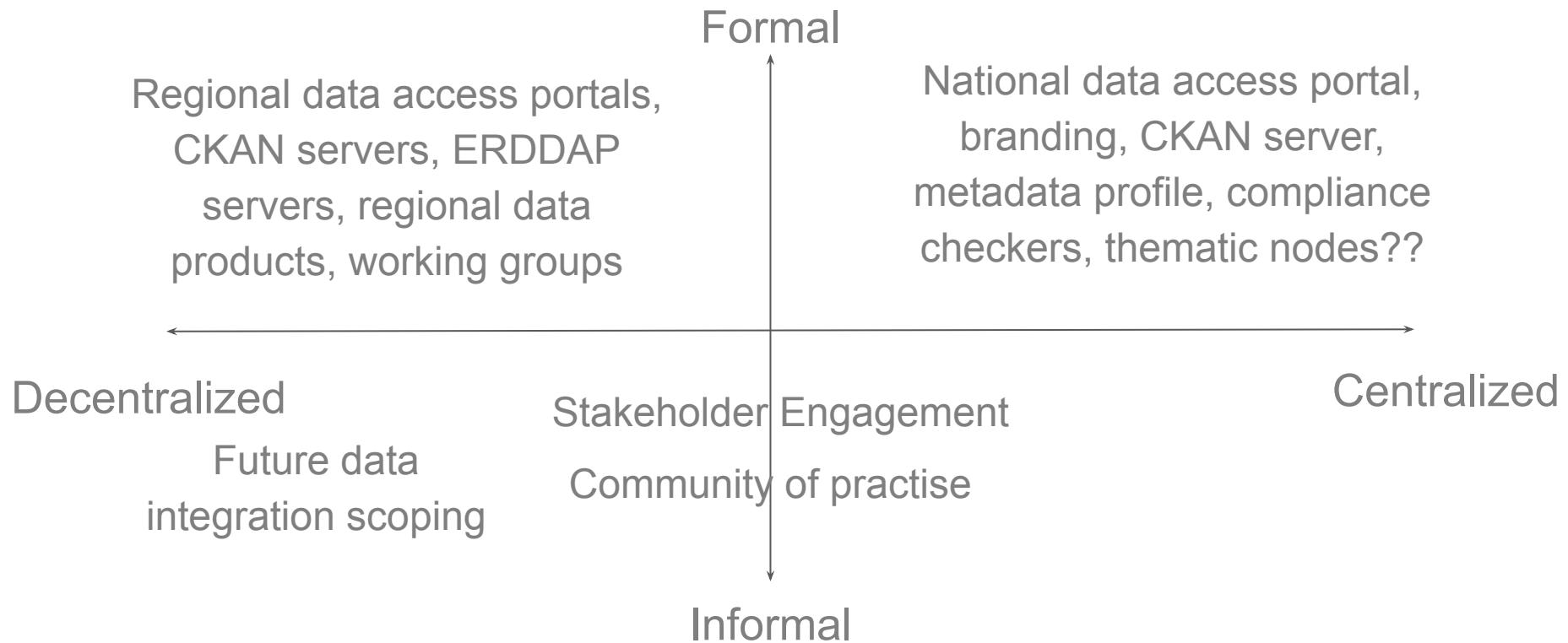


# Future Directions

- Data storage with **FRDR** for some Regional Associations
- Integration of **additional data sources**
  - Development of a DMP template (at least 2 options - ingestion vs harvesting)
  - Defined and documented workflows for data integration
  - Training and resources for data contributors
- Expanding Within Arctic and Great Lakes
- Coordination with other GOOS entities (building relationships IOOS & DOOS already)
- **FAIR compliance checker** on metadata catalogs? (used by DataONE)

1	Check	Category	Description	Item
2	Abstract	F	Checks to see if at least one Abstract concept exists.	abstract
3	Keyword	F	Checks to see if at least one Keyword concept exists.	keyword
4	MetadataIdentifier	F	Checks to see if at least one Metadata Identifier concept exists.	MetadataIdentifier
5	ResourceAuthorOriginator	F	Checks to see if at least one Resource Author / Originator concept exists.	ResourceAuthorOr
6	ResourceAuthorOriginatorIdentifier	F	Checks to see if at least one Resource Author / Originator Identifier concept exists.	ResourceAuthorOr
7	ResourceAuthorOriginatorIdentifierType	F	Checks to see if at least one Resource Author / Originator Identifier Type concept exists.	ResourceAuthorOr
8	ResourceCreationRevisionDate	F	Checks to see if at least one Resource Creation/Revision Date concept exists.	ResourceCreation

# Integration Approach





# Lessons Learned/Challenges

- Integrating across 3 **heterogenous RA contexts** in terms of existing data holdings, data partner relationship models, infrastructure and work culture
- **Project management** - currently distributed, but developing framework for an integrated approach
- **Integrating DFO data** in terms of clear communication pathways and non-centralized data sources
- **Human Resources** - extra work load on existing staff, hiring new staff, scheduling meetings with key personnel
- **Interdependencies and status awareness** of working group activities



# Discussion Questions - <http://bit.ly/2VOFCV6>

**Question 1:** What should be the points of collaboration between CIOOS and the broader Canadian data curation community? It's clear a great deal of work has been done, how do we leverage this and avoid duplication?

**Question 2:** In anticipation of the Tri-Agency Data Policy this winter, how can CIOOS be integrated into the institutional strategies as a recommended data repository for oceanographic data produced by researchers at their institution, and/or increase linkages to oceanographic datasets in their collections? What are the perceived barriers or obstacles to that approach?

**Question 3:** How should CIOOS communicate with the Canadian data curation community moving forward?