



SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE  
SHARED STEWARDSHIP OF RESEARCH DATA



# **Platform Presentation 1.1: Findability**

## **Federated Research Data Repository (FRDR) – Portage and Compute Canada**

National Data Services Framework Summit  
January 24, 2019

Jeff Moon, Director, Portage

# What is Portage?

The Portage Network is a national initiative of the Canadian Association of Research Libraries (CARL), with the goal of building research data management (RDM) capacity in Canada through a network of experts in a growing community of practice.

## Networks of Expertise

- Pan-Canadian RDM expertise with over 100 in-kind experts
- Contribute to development of RDM resources, tools, training, and policy

## Infrastructure Platforms

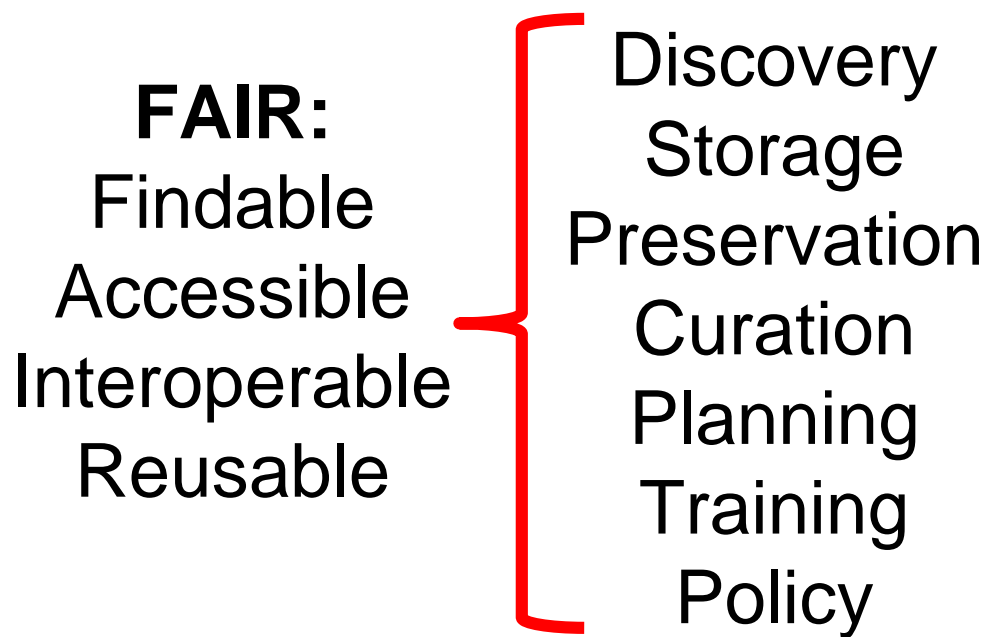
- Working with library consortia, institutions, and other infrastructure partners to assemble essential RDM infrastructure and service components

# Gaps, opportunities...

# and solutions

Principles  Enablers

Solutions...



# Gaps & opportunities... and solutions

Principles

Enablers

Portage solutions...

**FAIR:**  
Findable  
Accessible  
Interoperable  
Reusable

Discovery  
Storage  
Preservation  
Curation  
Planning  
**Training**  
Policy

**"Good Enough"  
Guides**



# And a brand-new “Good Enough” Guide for FRDR



# Gaps & opportunities... and solutions

## Principles

**FAIR:**  
Findable  
Accessible  
Interoperable  
Reusable

## Enablers

Discovery  
Storage  
Preservation  
Curation  
Planning  
Training  
Policy

## Portage solutions...

Institutional  
Strategy



DMPs



Deposit



**Tri-Agency  
RDM  
Policy**

# Gaps & opportunities... and solutions

## Principles

**FAIR:**  
Findable  
Accessible  
Interoperable  
Reusable

**Standards, Best Practices,  
& White Papers**

## Enablers

Discovery  
Storage  
Preservation  
Curation  
Planning  
Training  
**Policy**

## Portage solutions...

### Tri-Agency Policy

Institutional  
Strategy



DMPs



Deposit



RDM  
Strategy  
Template



# Gaps & opportunities... and solutions

## Principles

**FAIR:**  
Findable  
Accessible  
Interoperable  
Reusable

## Enablers

Discovery  
Storage  
Preservation  
Curation  
**Planning**  
Training  
Policy

## Portage solutions...

### Tri-Agency Policy

Institutional  
Strategy



DMPs



Deposit



DMP Assistant

Sign In

Sign Up

**DMP Assistant:**  
National, online, bilingual,  
Data Management  
Planning Tool



# Gaps & opportunities... and solutions

## Principles

**FAIR:**  
Findable  
Accessible  
Interoperable  
Reusable

## Enablers

**Discovery**  
**Storage**  
**Preservation**  
**Curation**  
Planning  
Training  
Policy

## Portage solutions...

### Tri-Agency Policy

Institutional  
Strategy



DMPs



Deposit



Useful to take a look at the broader  
Canadian Repository Landscape

# Canadian Repository Landscape



**General  
Repositories**

**Government  
Repositories**

**Domain-specific  
Repositories**



**re3data.org**



# Canadian Repository Landscape



General  
Repositories

Government  
Repositories

Domain-specific  
Repositories



Federated Research Data Repository



National Instance of Dataverse

## National Discovery Service

Portage Data Discovery Service Working Group &  
Compute Canada FRDR Team



Portage Dataverse North Working  
Group and OCUL Scholars Portal  
Dataverse Team



# Gaps & opportunities... and solutions

## Principles

**FAIR:**  
Findable  
Accessible  
Interoperable  
Reusable

## Enablers

**Discovery**  
**Storage**  
**Preservation**  
**Curation**  
Planning  
Training  
Policy

## Portage solutions...

### Tri-Agency Policy

Institutional  
Strategy



DMPs



Deposit



**National Repository  
Solutions:**  
Federated Research Data  
Repository (FRDR) &  
Dataverse

Let's take a look at Dataverse...

# Dataverses in Canada

## Scholars Portal Dataverse

**Algoma**  
UNIVERSITY

Algoma University Dataverse

**Brock**  
University

Brock University Dataverse

 **Carleton**  
UNIVERSITY

Carleton University Dataverse

 **Lakehead**  
UNIVERSITY

Lakehead University Dataverse

 **Laurentian**University  
Université **Laurentienne**

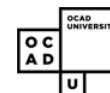
Laurentian University Dataverse

**McMaster**  
University

McMaster University Dataverse

 **NIPISSING**  
UNIVERSITY

Nipissing University Dataverse

 **OCAD**  
UNIVERSITY

OCAD University Dataverse

 **Queens**  
UNIVERSITY

Queen's University Dataverse



Royal Military College Dataverse

**Ryerson**  
University

Ryerson University Dataverse

**TRENT**   
UNIVERSITY

Trent University Dataverse

  
**UNIVERSITY OF CALGARY**  
PRISM Dataverse

**UNIVERSITY**  
*of* **GUELPH**

University of Guelph Dataverse

 **UNIVERSITY OF ONTARIO**  
INSTITUTE OF TECHNOLOGY

UOIT Dataverse

 **uOttawa**

University of Ottawa Dataverse

**University**  
*of* **Regina**

University of Regina Dataverse

 **UNIVERSITY OF**  
**TORONTO**

University of Toronto Dataverse

 **University**  
**of Victoria**

University of Victoria Dataverse

 **UNIVERSITY OF**  
**WATERLOO**

University of Waterloo Dataverse

 **University**  
**of Windsor**

University of Windsor Dataverse

**Western**   
UNIVERSITY - CANADA

Western University Dataverse

**LAURIER**   
Inspiring Lives.

Wilfrid Laurier University  
Dataverse

**YORK**   
UNIVERSITY  
UNIVERSITY

York University Dataverse

# Dataverses in Canada

## Other Dataverse instances

Algoma  
UNIVERSITY

Algoma University Dataverse

Brock  
University

Brock University Dataverse



Carleton University



Laurentian University Dataverse

McMaster  
University

McMaster University Dataverse



Nipissing University



UAL Dataverse



Dalhousie University Dataverse (Dalhousie University)

Ryerson  
University

Ryerson University Dataverse

TRENT  
UNIVERSITY

Trent University Dataverse



PRISM Dataverse

UNIVERSITY  
of GUELPH

University of Guelph Dataverse

Abacus Dataverse Network

UBC, SFU, UNBC, & UVic

University  
of Regina

UNIVERSITY OF  
TORONTO



LIBRARIES

**DATAVERSE**  
RESEARCH DATA REPOSITORY

# Dataverse facts...

## 5 Canadian Dataverse instances:

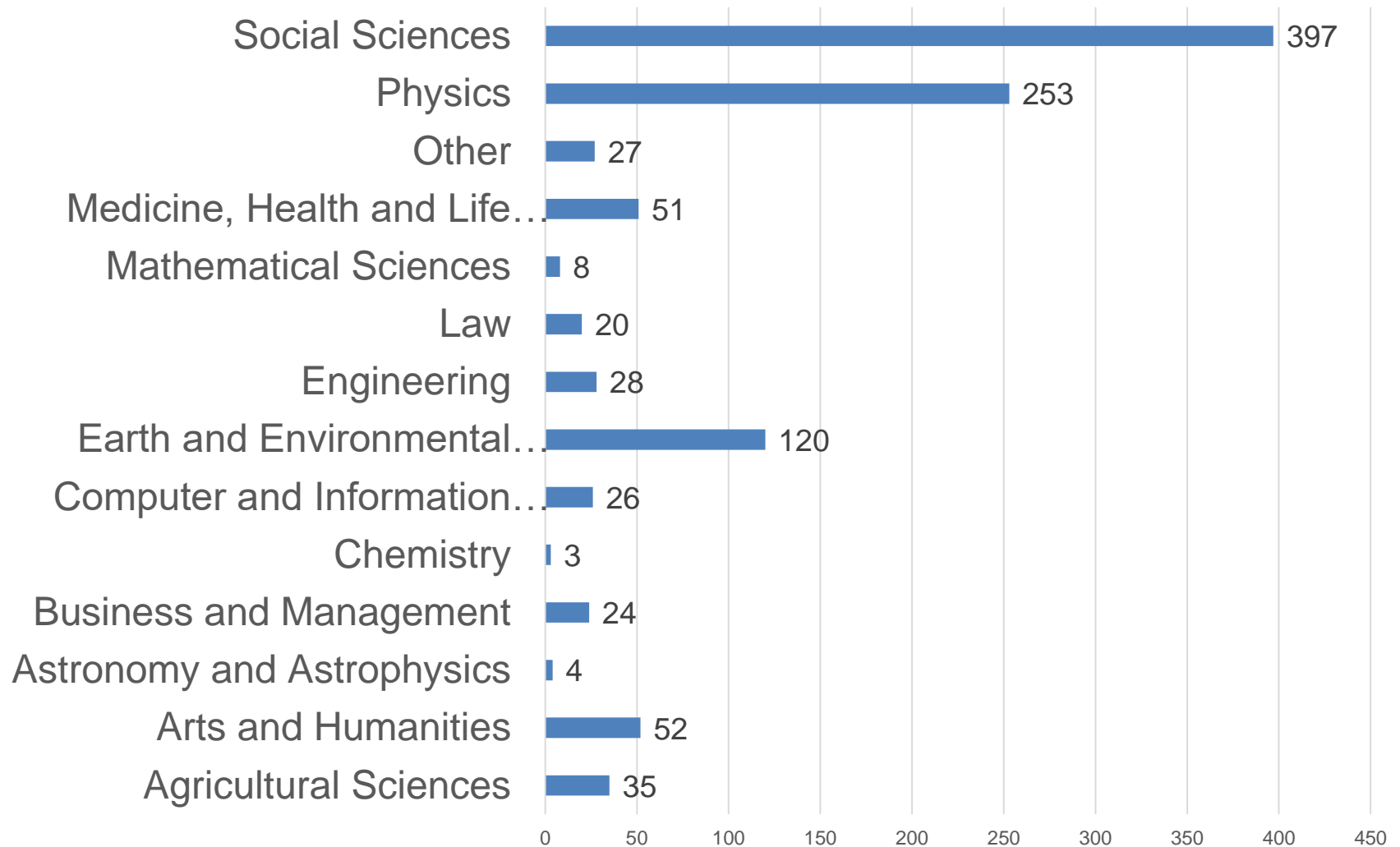
- **3** standalone (U Alberta, Dalhousie, UNB)
- **2** confederated:
  - Abacus: **4** institutions - UBC, SFU, UNBC, & Uvic
  - Scholars Portal Dataverse: **24** institutions

## Scholars Portal Dataverse (Jan 2019)

- 280 Dataverses
  - ~100,000 downloads
  - ~1500 datasets
  - ~25,000 files

# Dataverse facts...

**Datasets by Discipline: Scholars Portal Dataverse** [Aug, 2018]





# Gaps & opportunities... and solutions

## Principles

**FAIR:**  
Findable  
Accessible  
Interoperable  
Reusable

## Enablers

**Discovery**  
**Storage**  
**Preservation**  
**Curation**  
Planning  
Training  
Policy

## Portage solutions...

Let's take a closer look at FRDR



Federated Research Data Repository

# What is FRDR?



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

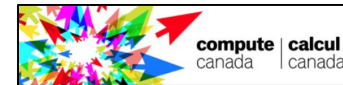
*A scalable, federated repository platform for digital research data management and the discovery of Canadian research data*

# About FRDR



A collaboration among libraries, institutions, HPC, and federal organizations involving many stakeholders across the RDM sphere

- Partnership between **Compute Canada** and the **Canadian Association of Research Libraries (CARL)**
- Partnership between **Compute Canada** and **Globus**



**Globus transfers data  
from one 'secure  
endpoint' to another**



**Globus overcomes  
limitations of  
'http' web protocol**

- **Faster than http**
- **Clean recovery from interruptions**

**Researcher initiates transfer,  
and is notified when  
complete**

# About FRDR



A collaboration among libraries, institutions, HPC, and federal organizations involving many stakeholders across the RDM sphere

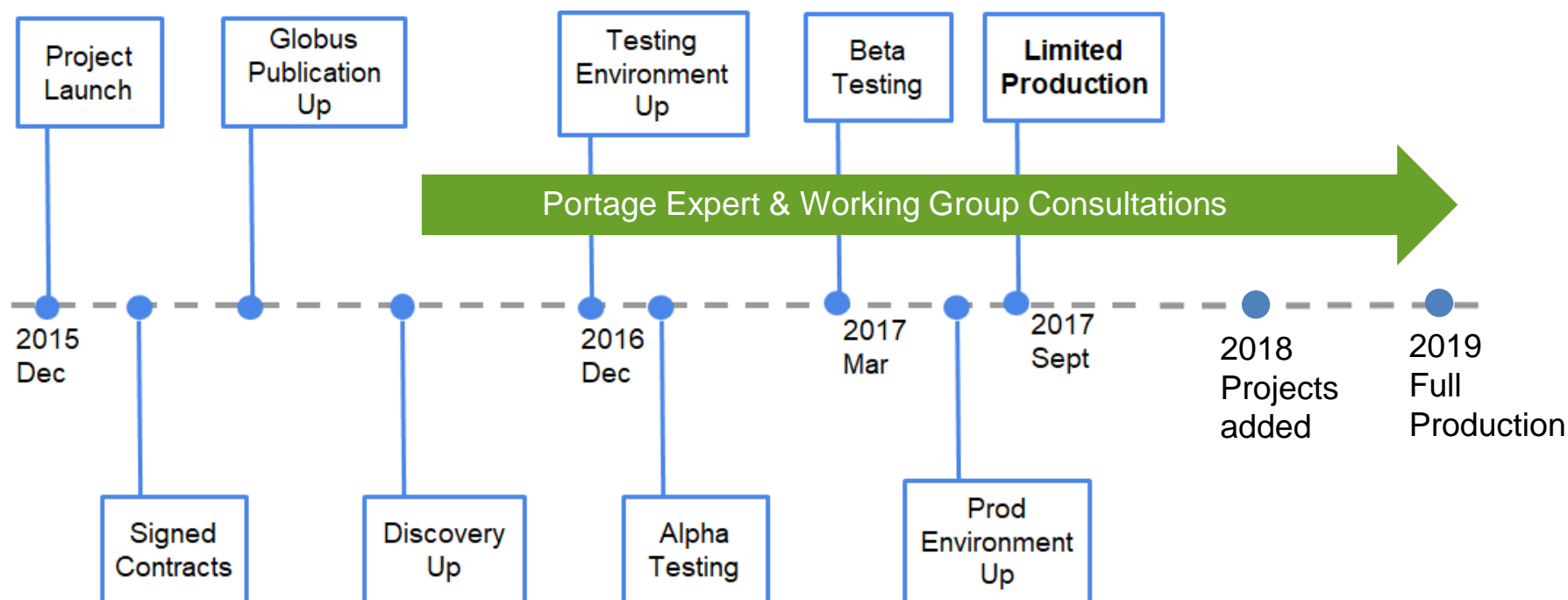
- Partnership between **Compute Canada** and the **Canadian Association of Research Libraries (CARL)**
- Partnership between **Compute Canada** and **Globus**
- Search page code from UBC Open Collection
- Hosted on **Compute Canada** hardware and infrastructure
- Steering and input from **CARL/Portage, Portage Network of Experts**, and **individual institutions**



Globus Connect



# Timeline



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

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

Discovery      Storage      Preservation      Curation

# What is FRDR?



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

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

Discovery      **Storage**      Preservation      Curation

‘Big Data-capable’

Scalable, distributed storage

Option for those *without* a local repository platform, or  
for *highly data-intensive* research projects



# Storage Metrics...



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

- **Limited Production **deposits** as of Jan. 2019:**
  - 18 partnering groups
  - 75 datasets published (4.4 TB)
  - 7 in progress (3.8 TB)

# What is FRDR?



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

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

Discovery      Storage      **Preservation**      Curation

Supports preservation for long term usability

Uses Archivematica to process datasets of any size

Fits preservation processing into broader plans for a network of preservation service providers

# What is FRDR?



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

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

Discovery      Storage      Preservation      **Curation**

Envisions a national network of curation expertise

Uses a federated approach with FTEs embedded at institutions

Facilitates curation for a broad range of disciplines

# What is FRDR?

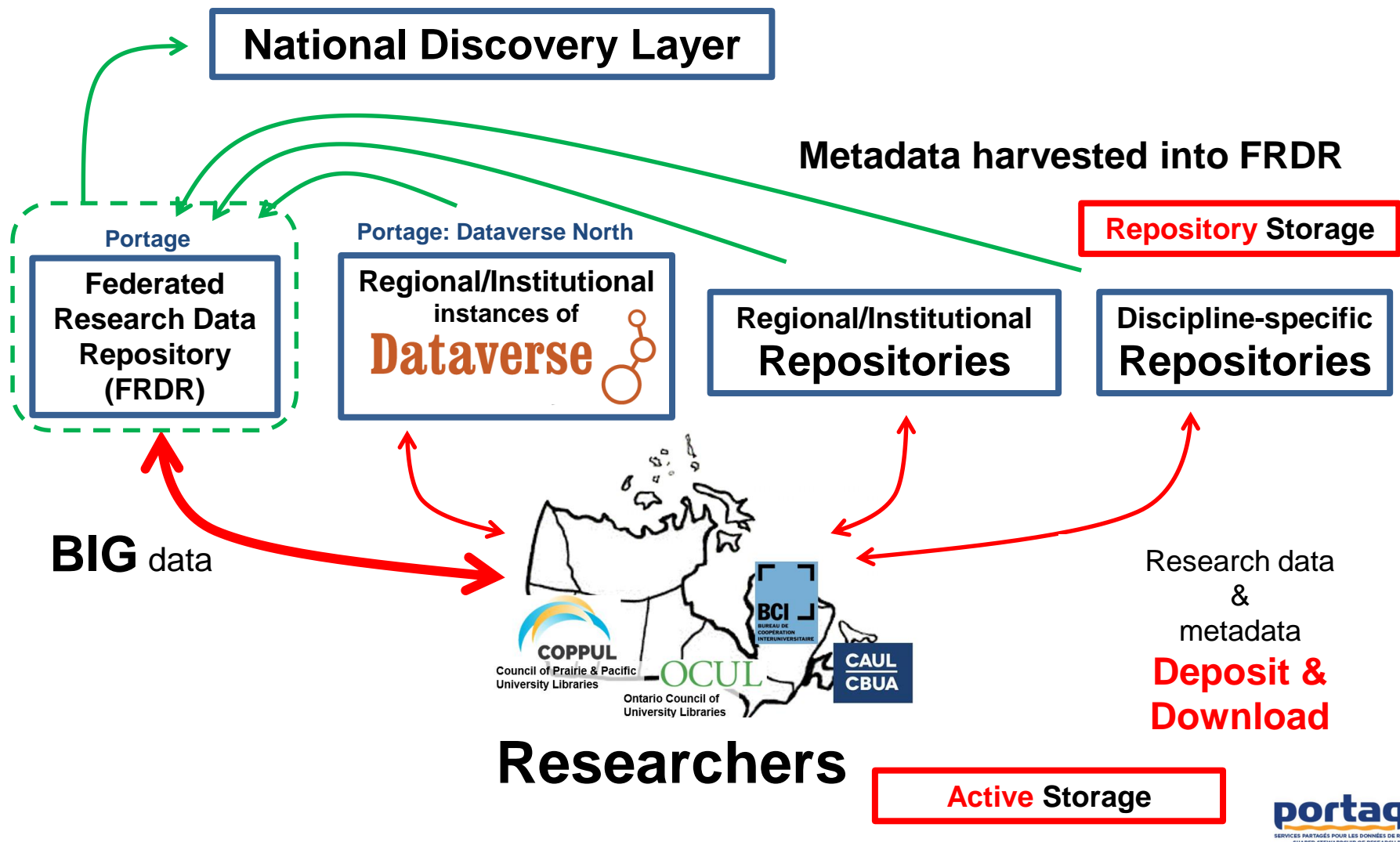


**FRDR = Federated Research Data Repository (DFDR en français)**

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

**Discovery**      Storage      Preservation      Curation

# National Discovery Layer



FRDR's harvester gathers and indexes metadata from repositories across Canada to make research data held in many repositories discoverable from a single platform

- Supports internationally recognized standards<sup>\*</sup>, with more being added as needed
- Provides a metadata crosswalk & uses harvested search terms as 'facets'
- Uses the same harvester to index data housed in the FRDR repository
- Generates JSON-LD<sup>\*\*</sup> for the search backend

<sup>\*</sup>Currently supported standards: OAI-PMH, CKAN, CSW, Marklogic

Open Archives Initiative Protocol for Metadata Harvesting

CKAN, the world's leading Open Source data portal platform

<sup>\*\*</sup>JavaScript Object Notation for Linked Data

# Discovery: Collaborating repositories



**Harvests metadata  
from 40 repositories  
indexing  
over 100,000 datasets**

**All searchable via  
<https://www.frdr.ca>**



LAU

**Find and Share **Canadian** Research Data**



Search



Deposit Data

Advanced search

western

# Benefits of a National Discovery Layer



**Improve discovery of research data/metadata**

**Drive traffic to existing repository sites**

**Break down repository siloes**





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

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

**Discovery**      Storage      Preservation      Curation

→ • **FRDR searches over 6 months, May-Oct 2018<sup>1</sup>**

- |                     |         |                        |
|---------------------|---------|------------------------|
| • <b>Users:</b>     | ~3000   | Avg. per month: ~500   |
| • <b>Sessions:</b>  | ~5,000  | Avg. per month: ~800   |
| • <b>Pageviews:</b> | ~18,000 | Avg. per month: ~3,000 |

<sup>1</sup> From Google Analytics

# Challenges & Opportunities

## Challenges

**Securing funding for full production**

Ongoing discussions with ISED and other stakeholders.

**Lack of standardized metadata practices among repositories**

Portage experts are engaged in a project to map subject keywords to the Online Computer Library Center's (OCLC) Faceted Application of Subject Terminology (FAST) using OpenRefine.

# Challenges & Opportunities

## Opportunities

Compute Canada and Portage working on this.

Complementing the features of, and collaborating with, a national instance of Dataverse.

CANARIE RDM Funding to improve geospatial data discovery and visualization in both the Dataverse and FRDR repositories.

*Led by UBC with support from Compute Canada and Scholar's Portal.*

**Providing Canadian researchers with a standard 'storage allocation'**

**Providing a robust, scalable, and complementary national repository option**

**Improved Geospatial Discovery**

# Progress toward the FAIR Principles

FAIR  
Principles

← Enablers

**FAIR:**  
**Findable**  
Accessible  
Interoperable  
Reusable

Discovery  
Storage  
Preservation  
Curation  
Planning  
Training  
Policy

Solutions

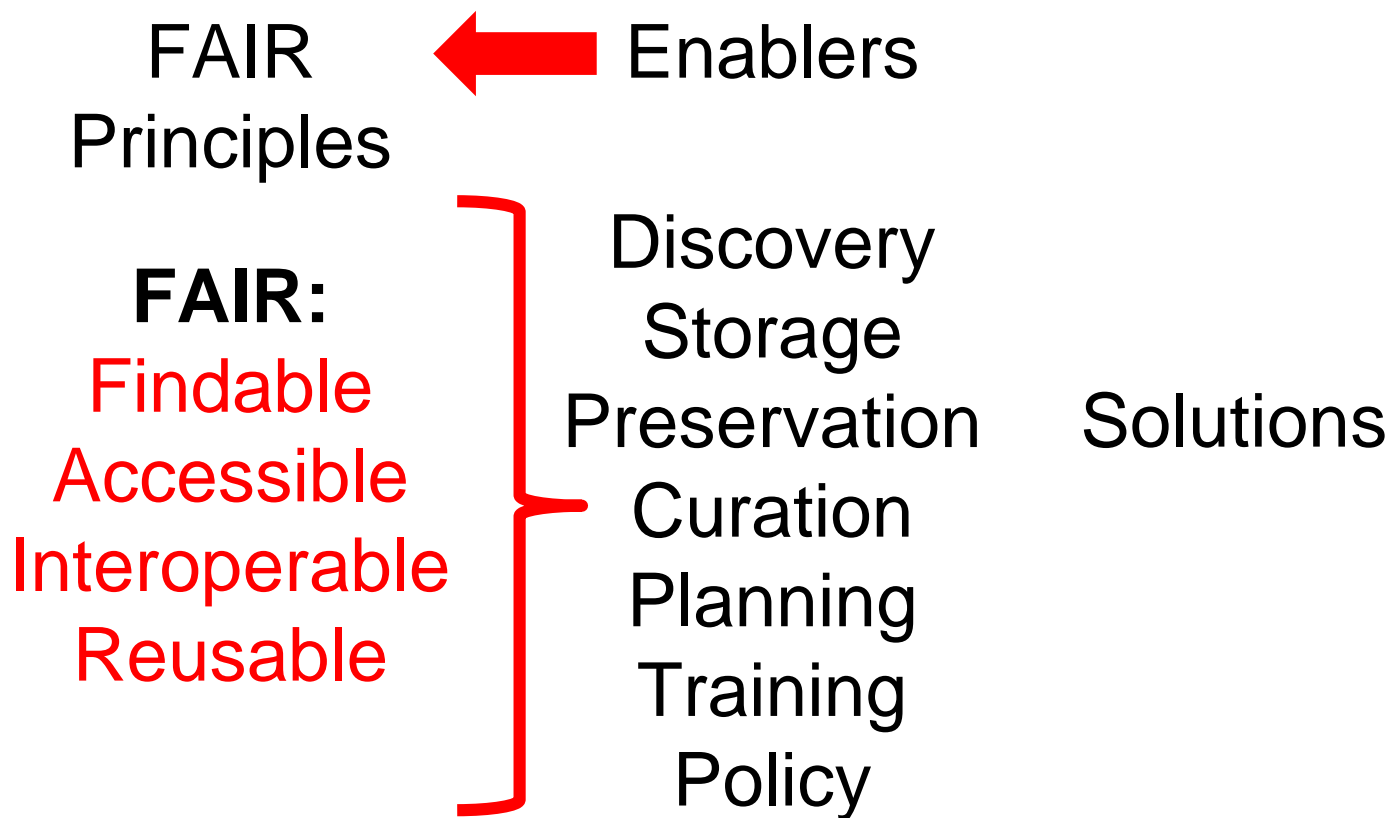


National  
Discovery  
Layer

# Progress toward the FAIR Principles



and others  
in the RDM  
landscape



# A lot of contributors to RDM solutions



... and many other RDM and DRI stakeholders

# Questions?



[portagenetwork.ca](http://portagenetwork.ca)

Jeff Moon, Portage Director  
[portagedir@carl-abrc.ca](mailto:portagedir@carl-abrc.ca)