CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

computecanada

Globus in Canada: National and University Perspectives

Jason Hlady
University of Saskatchewan / Compute Canada

GlobusWorld
April 25, 2018

Outline

- Compute Canada overview

- Compute Canada Globus File Transfer service

- University of Saskatchewan and Globus

- Federated Research Data Repository (FRDR)
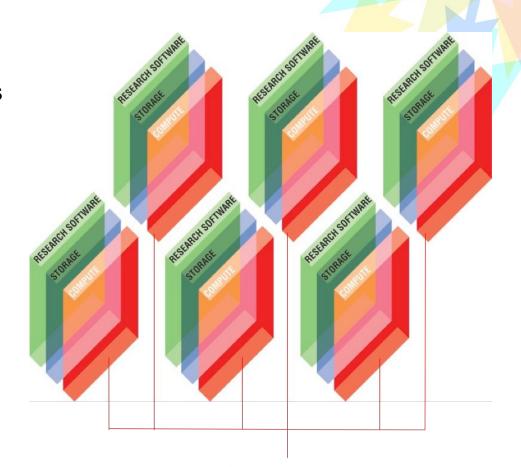 - Globus Publication
  - Globus Search
  - Globus File Transfer

Compute Canada

- CC is a not-for-profit organization that provides digital infrastructure for researchers in Canada
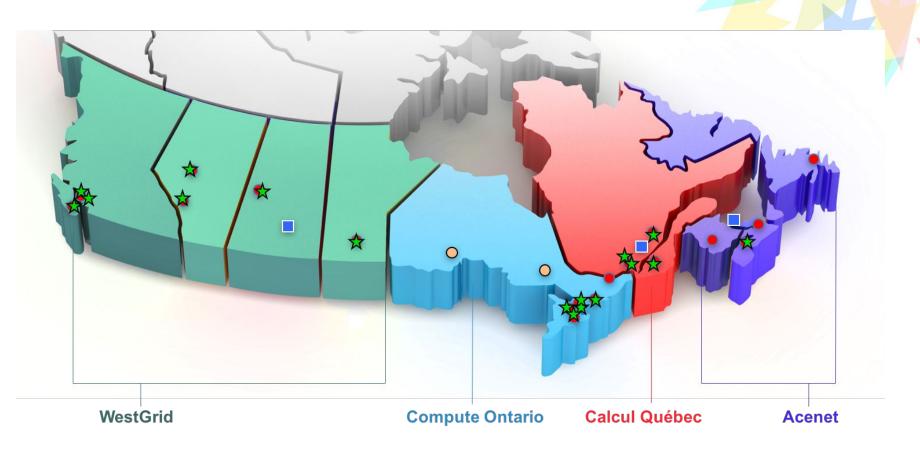
- 50 HPC Systems
- 27 Data Centres
- 200 Experts

- Consolidating, renewing infrastructure in 2017-18

CANARIE Network

CC Globus Deployment

WestGrid   Compute Ontario   Calcul Québec   Acenet

- ◼ Member University
- ◯ Member University + Personnel Site
- ● Member University, Personnel + Infrastructure Site
- ★ Globus Site

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA
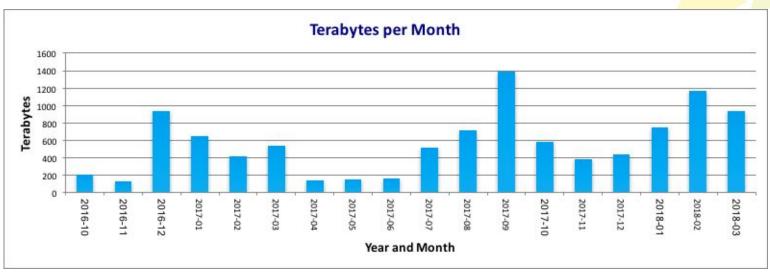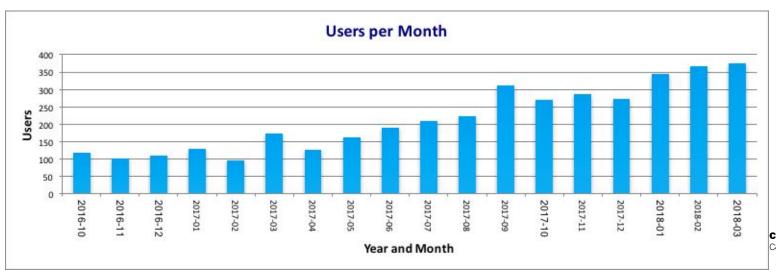
compute | calcul
canada | canada

Compute Canada Globus

- Compute Canada branded file transfer service launched June 2014
  - Compute Canada users have transferred more than 12.8PB of research data
  - 375+ unique users from Compute Canada in March 2017
  - Transferred 7.3PB in 2017 alone, more than all previous years combined

- Compute Canada users leverage Globus Plus features
  - Sharing
  - Personal file transfer

- Globus is Compute Canada's backbone service for migrating PB of data from defunded, legacy systems to new infrastructure

Compute Canada Globus Usage

Globus and University of Saskatchewan / CLS

- Canadian Light Source: only Canadian Synchrotron

- USask / CLS became Globus campus to help manage and move large amounts of data generated at CLS and USask
  - predominant data transfer for CLS users still sneakernet

- CLS
  - Alternate Globus IDP using CLS credentials
  - Production DTN purchased, test DTN deployed
  - CLS look-and-feel Globus transfer page
  - A new way for researchers to move their data from CLS to their home institution

Globus and University of Saskatchewan

- USask provides research storage service DATASTORE
  - Access for small data by NFS, CIFS for convenience
  - Access by Globus
    - speed
    - automation
    - connection with Compute Canada
    - more complex data management

- Integrated USask IDP using CAF/CILogon
  - USask look-and-feel customized Globus page May 2018
  - Production DTNs connecting to DATASTORE, research clusters, Compute Canada

- Large-scale research projects leveraging the service

Federated Research Data Repository

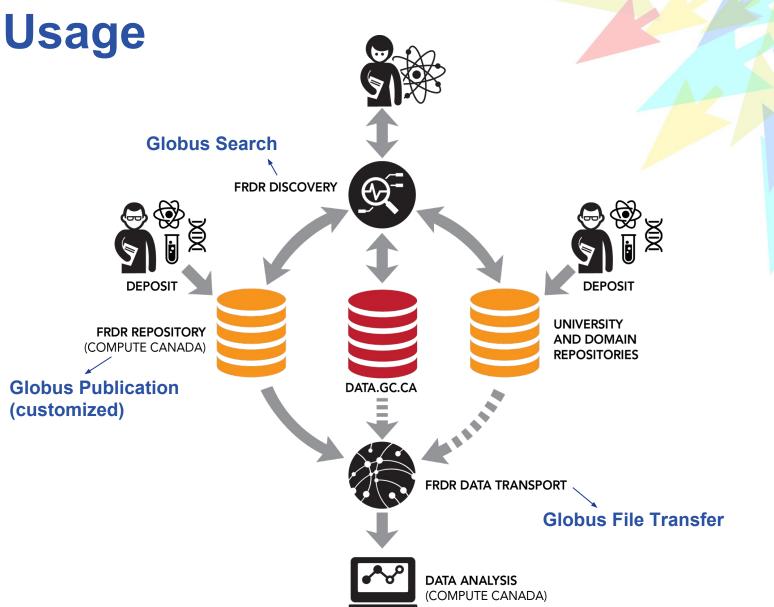FRDR = Federated Research Data Repository (DFDR en français)

Designing infrastructure for a national-scale research data repository

- Partnership between Compute Canada and CARL (Canadian Association of Research Libraries)
- Partnership between Compute Canada and Globus
- Using Compute Canada hardware

Discovery platform

- Leveraging Globus Search
- Indexing items in existing repositories, improving discovery

Usage

FRDR **DFDR**

FEDERATED RESEARCH DATA REPOSITORY   DÉPÔT FÉDÉRÉ DE DONNÉES DE RECHERCHE

Find and Share Canadian Research Data

Search   🔍   Deposit Data ⬇

Advanced Search

About FRDR

The Federated Research Data Repository (FRDR) is a joint initiative led by the Canadian Association of Research Libraries (CARL) and Compute Canada to provide Canadian researchers a place to deposit large data sets and to improve the discovery of existing Canadian research data.

Learn More »

Deposit Data

Any researcher affiliated with a Canadian institution can deposit data in the Federated Research Data Repository (FRDR). The platform efficiently moves large data sets and does preservation and archiving automatically. Curators will assist researchers to record metadata to facilitate discovery and data reuse.

Learn More »

Find Data

Search the Federated Research Data Repository (FRDR) to find research data sets from researchers affiliated with Canadian institutions. Data deposited to other repositories across Canada can also be found by searching in FRDR. See the growing list of collaborating repositories by clicking Learn More.
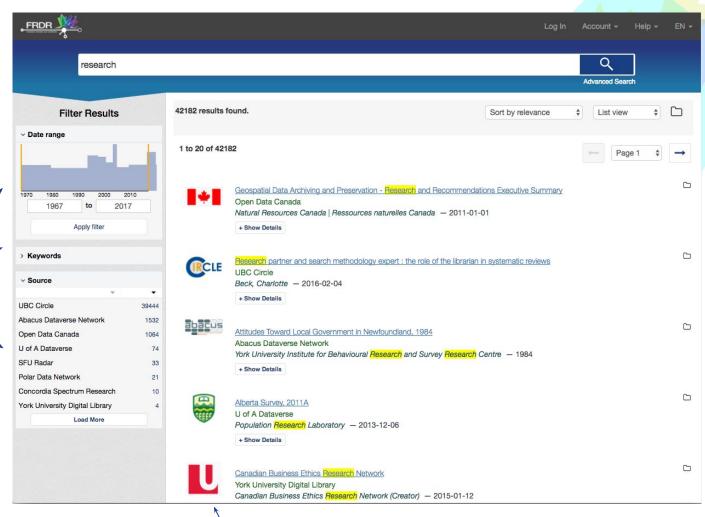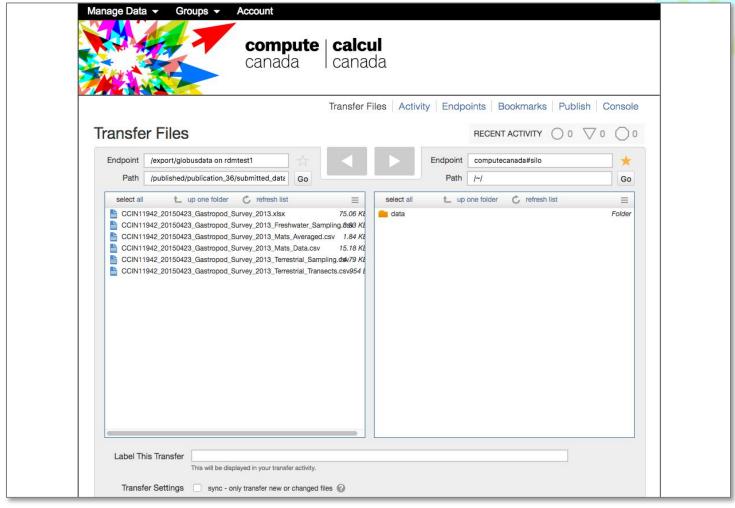
Learn More »

Search Results

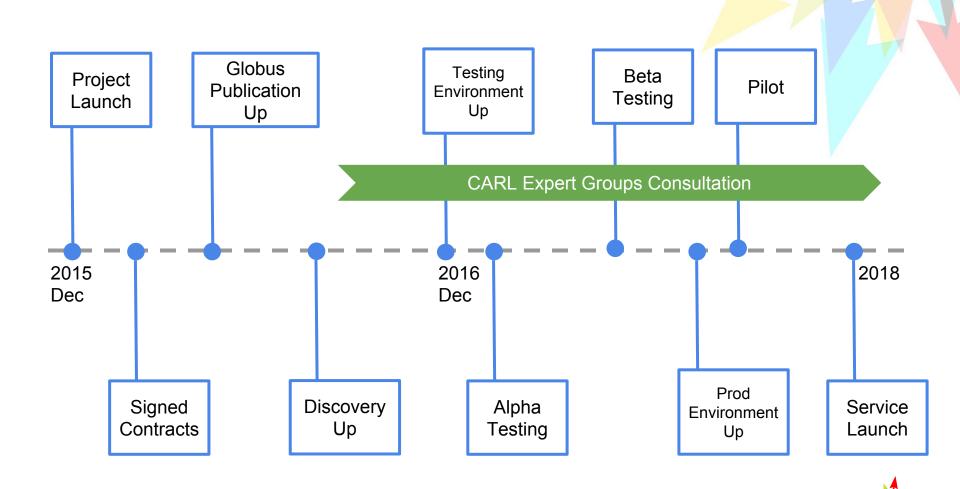# Dataset File Transfer

Globus Transfer

FRDR service

- FRDR Launched Limited Production September 2018

- A service based on FRDR offers to Canadian researchers:
    - A place to deposit large data sets
    - A place to deposit data sets if researcher does not have a local option
    - Preservation for long term usability of data sets
    - Archiving (and automatic geographic replication) of data sets to fit in with researcher backup strategy
    - Provide permanent identifiers (DOIs) for data sets
    - A way for researchers to discover data sets regardless of which repository they were deposited
- Storage and repositories can be distributed, owned and operated by organizations / institutions.

Project Timeline

Project Launch

Globus Publication Up

Testing Environment Up

Beta Testing

Pilot

CARL Expert Groups Consultation

2015 Dec

2016 Dec

2018

Signed Contracts

Discovery Up

Alpha Testing

Prod Environment Up

Service Launch

Questions?

- www.computecanada.ca

- FRDR Public web site: http://frdr.thedev.ca

- Access the repository at:

  https://www.frdr.ca/repo/

  https://demo.frdr.ca/repo/

Acknowledgements

- Steering committee: Dugan O'Neil, Jeff Moon, Lee Wilson, Jay Brodeur, Umar Qasim, John Simpson, Jason Hlady

- CARL/Portage experts: DEG / CEG / PEG

- UBC and the Open Collections team: Eugene Barsky, Schuyler Lindberg

- Compute Canada: Cloud East and Cloud West teams, communications team, translators, support

- FRDR Development team: Alex Garnett, Keith Jeffrey, Todd Trann, Mike Winter, Adam McKenzie