computecanada

Compute Canada: Experiences with Globus as National Research Infrastructure
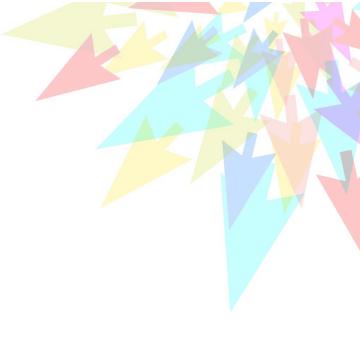
Jason Hlady
University of Saskatchewan / Compute Canada

GlobusWorld
April 11, 2017

Outline

- Compute Canada overview

- Compute Canada Globus File Transfer service

- 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
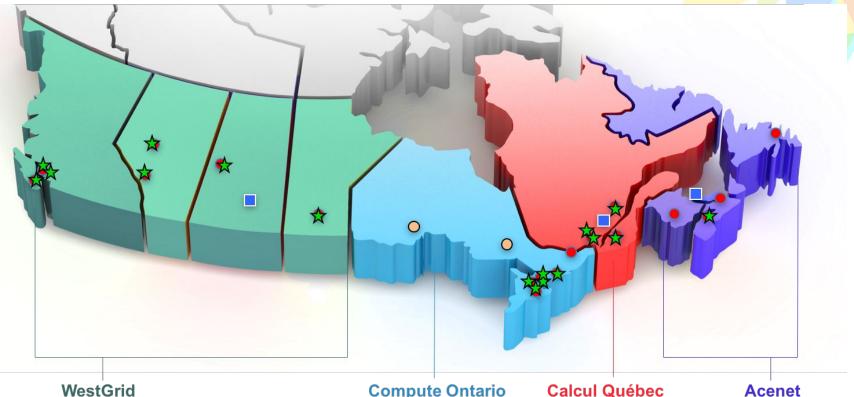
- 50 HPC Systems
- 27 Data Centres
- 200 Experts

- Consolidating, renewing infrastructure in 2017

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
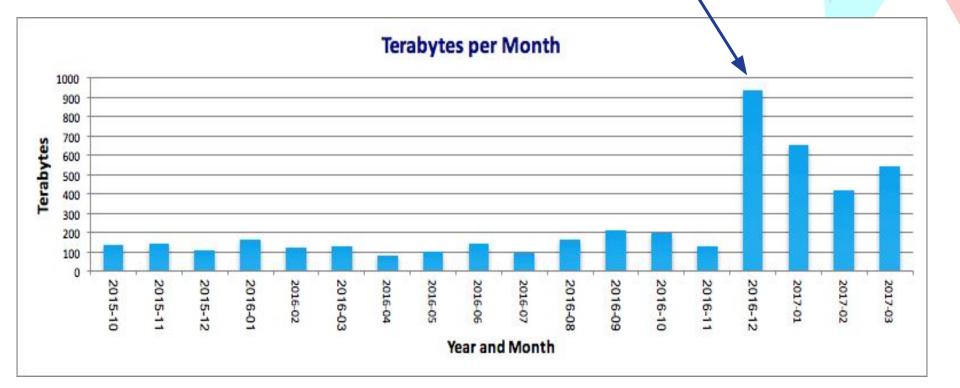
compute | calcul
canada | canada

Compute Canada Globus

- Compute Canada branded file transfer service launched June 2014
  - Compute Canada users have transferred more than 5.5PB of research data
  - 174 unique users from Compute Canada in Mar 2017

- 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

Migration from 3PB Silo system begins

Terabytes per Month

Federated Research Data Repository

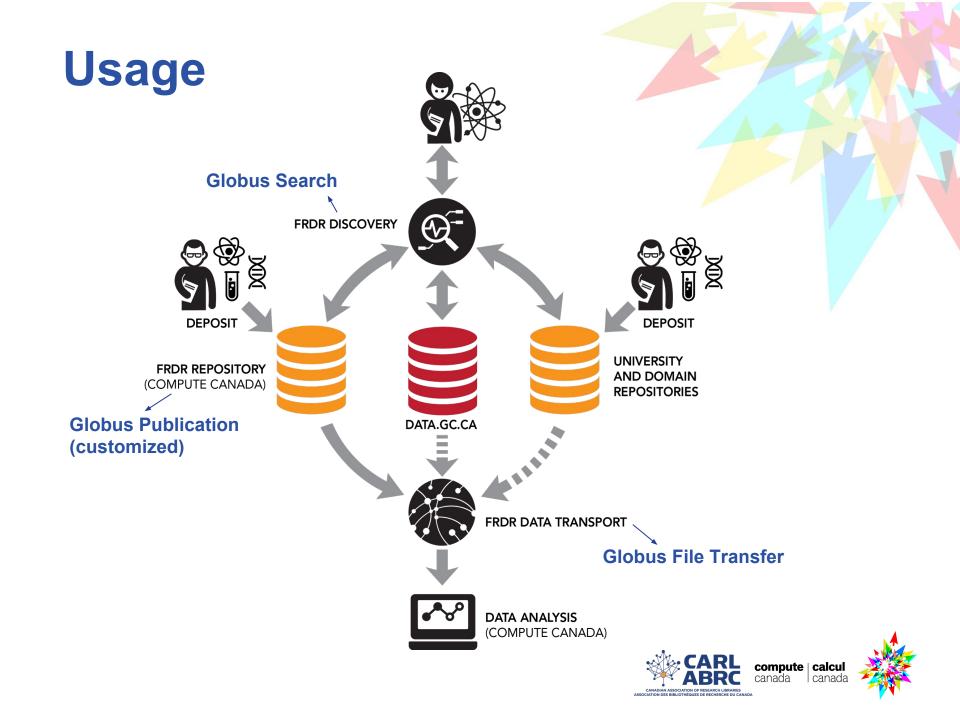FRDR = Federated Research Data Repository (DFDR en français)

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

Globus Search

FRDR DISCOVERY

DEPOSIT

FRDR REPOSITORY
(COMPUTE CANADA)

Globus Publication
(customized)

DATA.GC.CA

DEPOSIT

UNIVERSITY
AND DOMAIN
REPOSITORIES

FRDR DATA TRANSPORT

Globus File Transfer

DATA ANALYSIS
(COMPUTE CANADA)

Find and Share Canadian Research Data

FRDR DFDR
FEDERATED RESEARCH DATA REPOSITORY   DÉPÔT FÉDÉRÉ DE DONNÉES DE RECHERCHE

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.
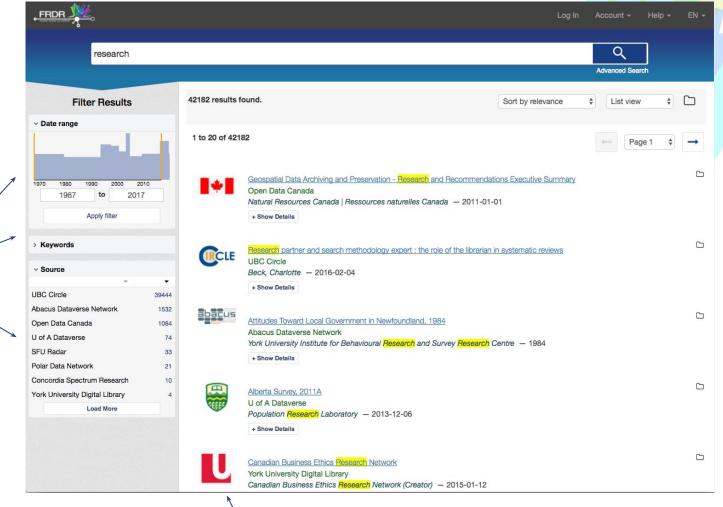
Learn More »

Search Results

Search Facets

Source Repositories

Dataset File Transfer

Globus Transfer

Project Timeline

Questions?

- www.computecanada.ca

- FRDR Public web site:

    www.computecanada.ca/rdm

- Additional info available in trailing slides

Features Coming This Year*

- Branding!

- Batch upload of data sets

- API for automated deposit

- Ingestion in place for existing data collections

- Scaling up capacity for preservation and archiving

- Increasing the number of repositories being indexed

- Metrics and reports for items, collections and communities

- Advanced search in discovery

- Geo-spatial search in discovery

- ORCID integration

* subject to change, this is not an approved road map

What will FRDR support?

- A service based on FRDR could offer 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. Multiple CC sites could run endpoints.

Part of the Bigger Picture

- CARL has applied for independent funding to operate a National Research Data Service

- Portage: involved in many RDM services, including FRDR

- Has support of RDC, LCDI

- Leveraging a broad host of librarian experts, resources, tools

- Partnering with infrastructure providers (CC) to provide FRDR service

  - Infrastructure and system administration

  - Technical and data management support

  - Development and ongoing feature extension

- Could offer FRDR services January 2018

Acknowledgements

- Steering committee: Dugan O'Neil, Jason Hlady, Chuck Humphrey, Steve Marks

- 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