compute canada

National Research Data Repository

Todd Trann
GlobusWorld - April 21, 2016
Compute Canada Regions and Members

- WestGrid
- Compute Ontario
- Calcul Québec
- Acenet

- Member University
- Member University + Personnel Site
- Member University, Personnel + Infrastructure Site
NRDR Project

Building a National Research Data Repository

- In partnership with Globus and CARL (Canadian Association of Research Libraries)
- Using Compute Canada hardware
- Available to any Canadian researcher

Collaborating with existing repository owners

- Indexing their metadata
- Sharing back a search API
- Improving cross-repository search
Features

- **Federated Storage Model:** Storage and repositories can be distributed, and owned operated by organizations / institutions

- **Scalable:** Many files; large files; large total amounts of storage

- **National Data Discovery:** Single search to discover data, regardless of location
Features (cont’d)

• **Data Preservation Pipeline**: Use Archivematica to normalize / preserve data

• **Suitable for broad range of data types**

• **Automatic geographic data replication**: Leverage Compute Canada national data infrastructure being deployed in 2016
System Diagram
Phase 1

Globus Publication
Archivematica
Regional Repository
Institutional Repository

Globus Connect
Metadata Index

CC Storage

Metadata
Compute Canada Cloud
Project Goals

Phase 1

• Stand up an in-Canada instance of Globus Publication service, hosted on Compute Canada hardware

• Demonstrate a researcher can upload a dataset directly into the NRDR prototype

• Demonstrate automated dataset preservation

• Import metadata from existing research data repositories and allow anonymous discovery of and access to the data. Proposed repositories:
  • UBC’s Abacus
  • uAlberta
  • SFU
  • Scholars Portal
Timeline
Phase 1

Stand up instance → Ingest data → Import IR metadata → Demo s & d → Demo data xfer → UI improvements → Preservation Integration → Preservation Scalability → June 2016