A partnership in federated research data management infrastructure

Chuck Humphrey
Director, Portage Network
Canadian Association of Research Libraries

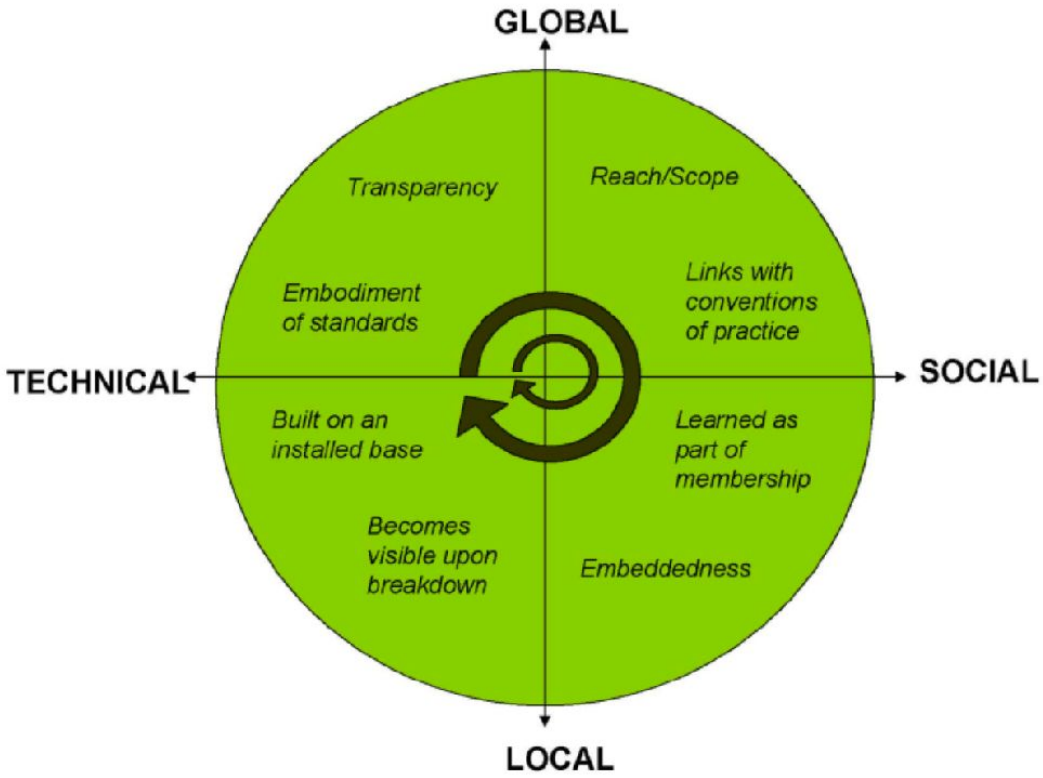
NDS Consortium October 2016

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

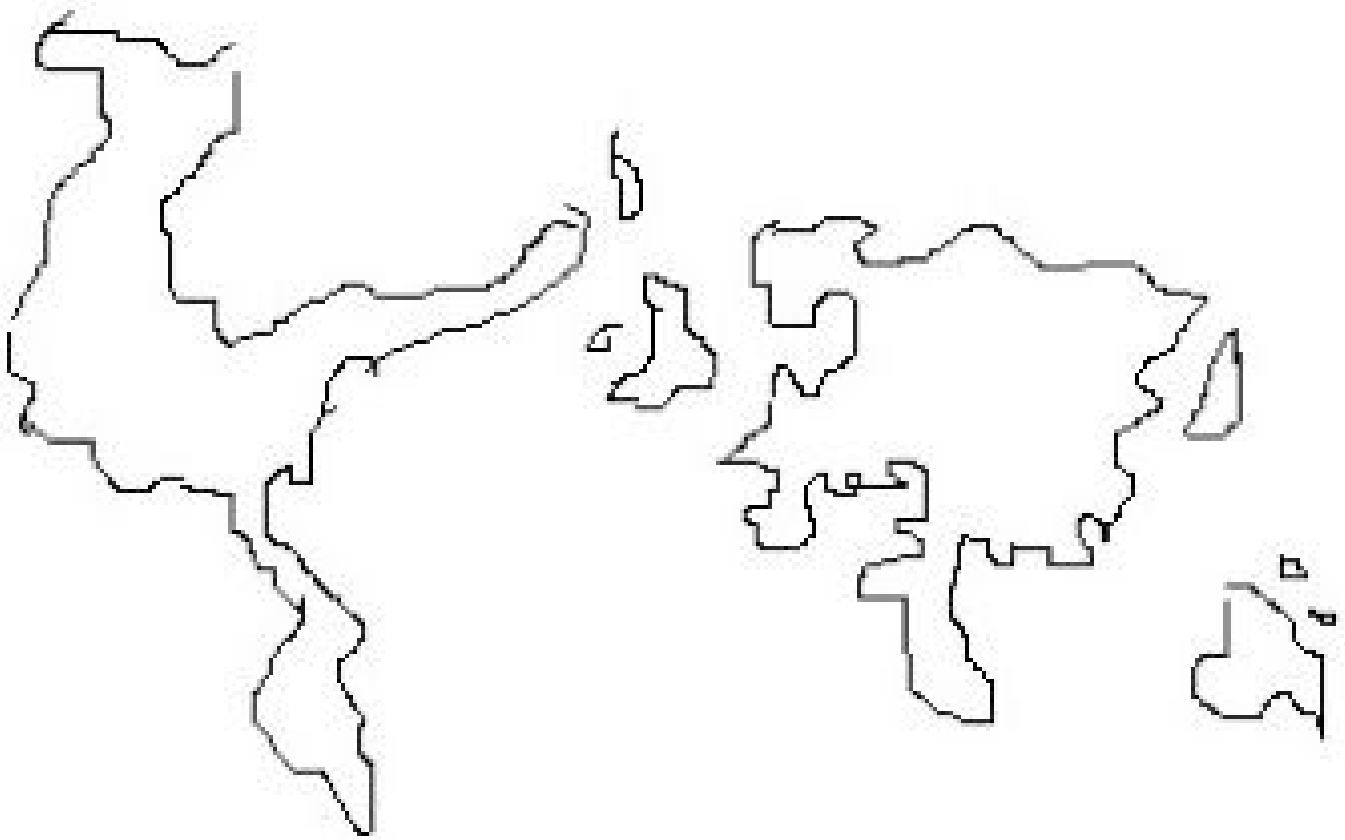
compute canada | calcul canada

Research Data Management Infrastructure

Technology + Expertise + Services supporting data across the research lifecycle, locally & globally



Adapted from **Understanding Infrastructure: Dynamics, Tensions, and Design.** P. Edwards, S. Jackson, G. Bowker, and C. Knobel, January 2007



"Creating the Data World Wide Web: Developments and Constraints," Kevin Schürer, ICPSR OR Meeting, Oct. 2007

Where is Canada?

Research Data Mgmt Infrastructure

2011 — National Data Summit

2011
2010 — CARL RDMI Proposal

Shift from a National Institution to National Infrastructure

2007
2006 — Canadian Digital Information Strategy

2005
2004 — National Consultation on Access to Scientific Research Data

2002
2001 — National Data Archive Consultation

Digital research infrastructure

Network + HPC + RDM

2014 Digital Infrastructure Summit II

Leadship Council for DI

2012 Digital Infrastructure Summit I

Convergence

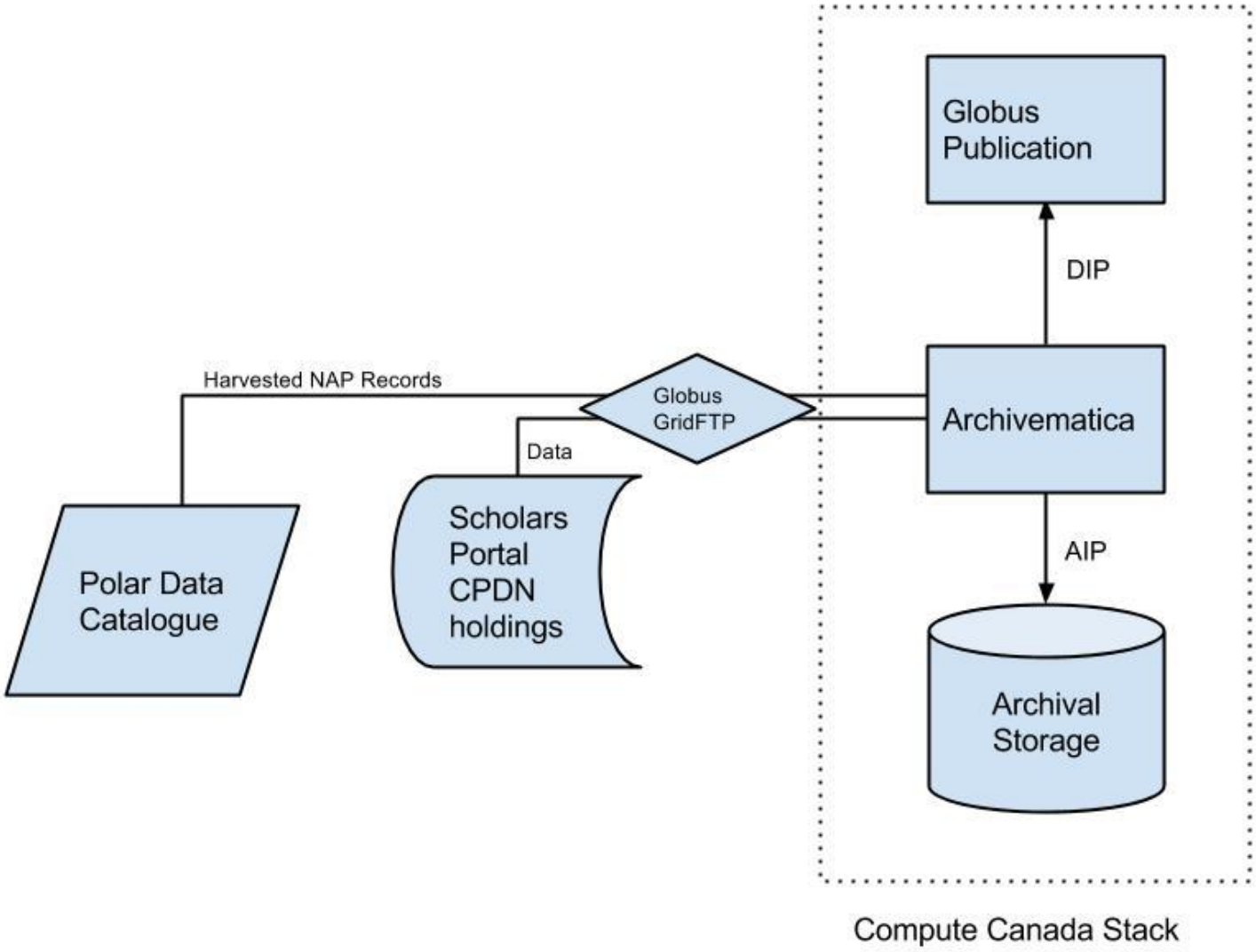
2016	Compute Canada - CARL MOU
2015	CARL Portage Network
2015 2014	RDC-CARL-CC Federated Pilots
2014	CARL ARC Project

First CARL-CC Collaboration

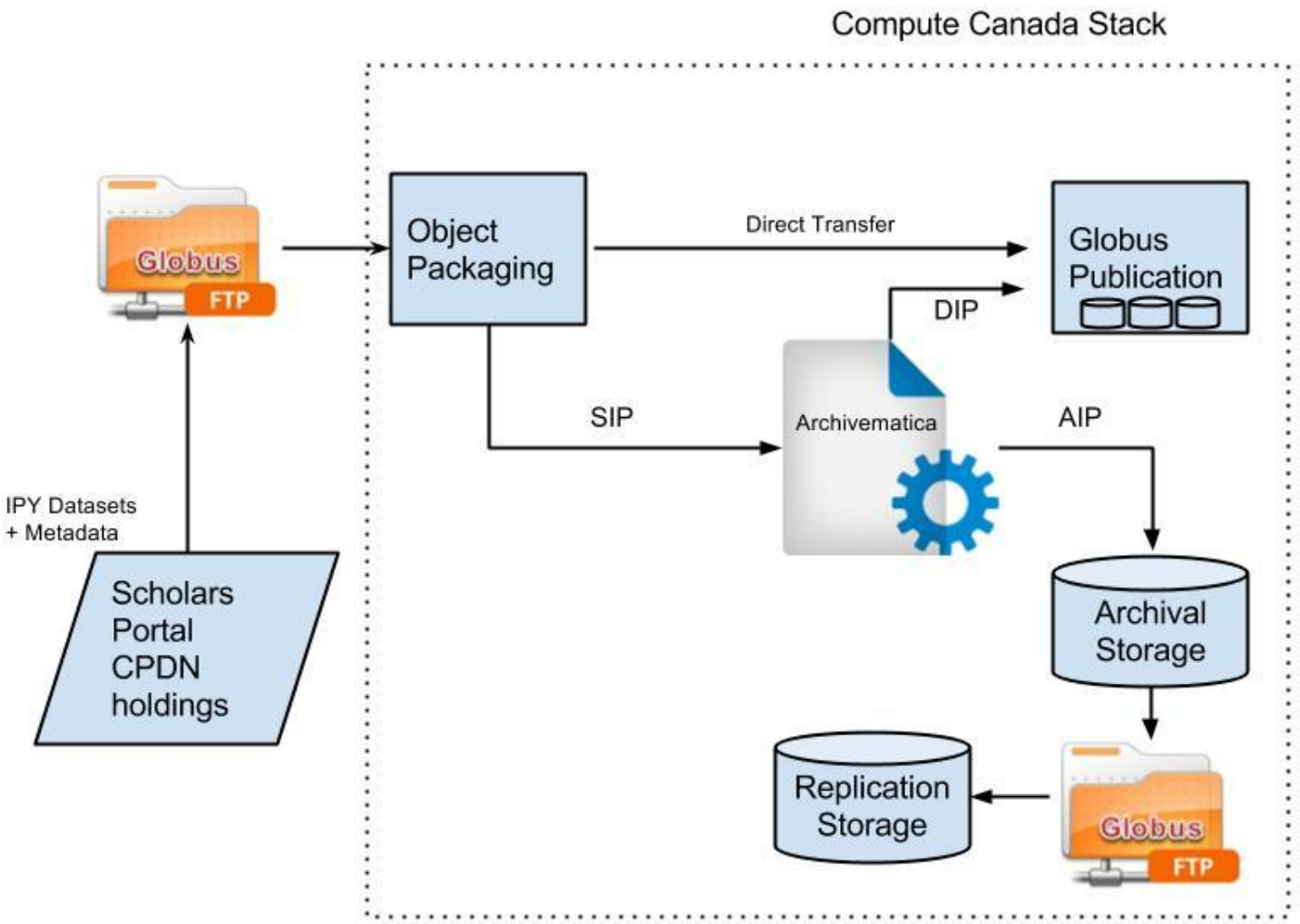
CARL(CPDN)-CC RDC Federated Pilot

- Started with the realization that CARL and CC were both tackling RDMI problems from different perspectives.
- Both organizations work on campuses across Canada serving researchers, but tended to work with researchers on different problems.
- While the desire to work together was strong - it took some time just to learn to speak the same language.
- Federated pilot (coordinated by RDC) gave us a way to work together. At the same time, something useful came out!

CPDN CC Federated Pilot Diagram 1



CPDN CC Federated Pilot Diagram 2



Complementary Roles

CARL-Portage and Compute Canada bring complementary skills and views to the partnership.

- Traditionally, CC has focused on data transfer and use (at scale), but very little on true data management and preservation.
- CARL has expertise in curation, preservation, stewardship, etc. that CC lacks.
- Combine the expertise of CARL and CC!

Collaborative Framework

- We started working from a single, high-level model but have allowed for the possibility of other partners joining to add to the overall functionality to this model.
- The framework is a modular design based on a few major functions divided into sub-functions.
- Each sub-function is implemented preferably as a microservice but could be a separate system.
- This allows for many contributors, some of whom may have their own microservices or systems for specific sub-functions.
- The partners will work together using agreed upon standards and protocols.

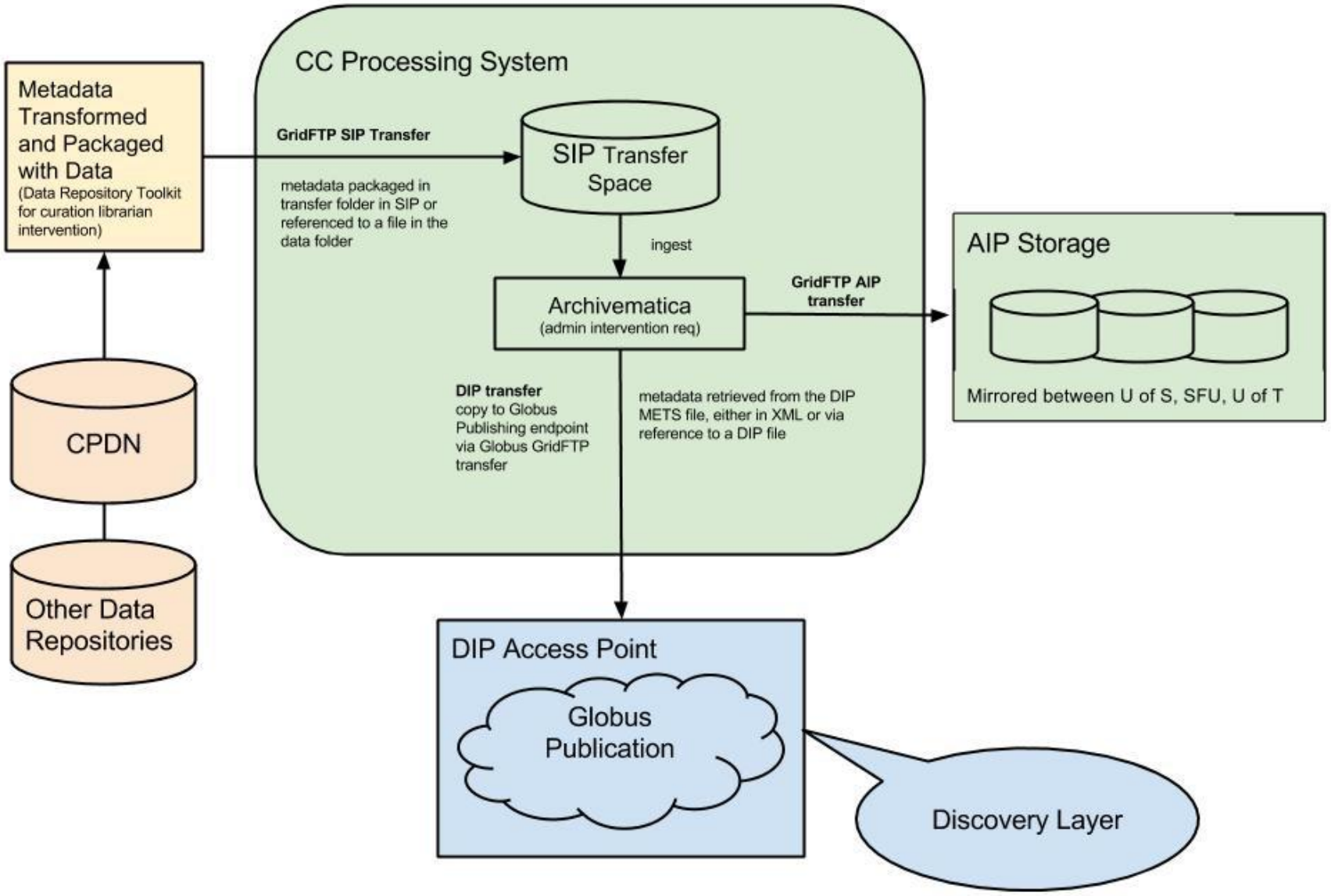
Operating principles

- Use existing tools & code when possible
  - Example: Globus Data Publication
- Don’t start from scratch if possible
  - Example: Archivematica
- If we borrow, we contribute back
  - Example: UBC Digital Collections User Interface

Important Features

- Federated storage model
- Scalable model
- Federated data discovery
- Preservation pipeline
- Geographic replication
- Bulk ingestion

Federated Research Data Repository

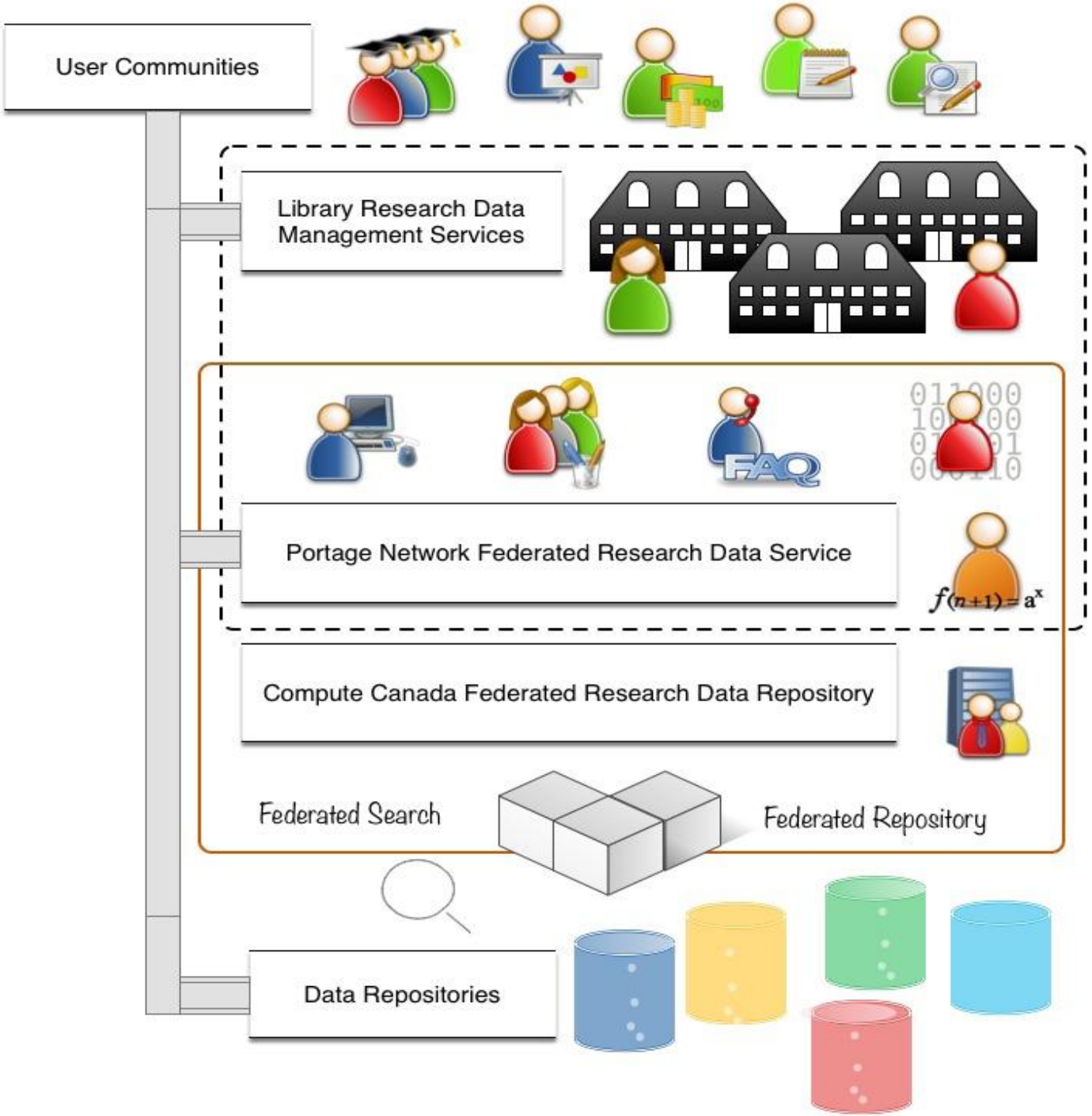


Collaboration Profile + Next Step

- Worked together on successful technology pilot (convened by Research Data Canada).
- Implementation tested with Canadian-funded International Polar Year research data (50-year archival commitment required).
- Signed 2-year MOA (2016 & 2017) to work on a platform for RDM for Canada.
- Tremendous progress towards a platform which is:
  - Federated storage
  - Federated service
  - Scaleable
  - Nationally integrated for data discovery
  - Capable of preservation
  - Suitable for a broad range of data types
- Development team in place, operations team is needed.

Federated Research Data Service

- Currently, no single organization has the mandate or funding to provide a comprehensive, open research data service in Canada.
- Portage and Compute Canada are working collaboratively to develop a platform to support a federated research data service in Canada.
- Requirements for a federated research data service:
  - RDM expert consultation on Data Management Plans (DMPs), curation, discovery, preservation
  - Data curation
  - Training on RDM services and platform
  - Privacy protection
  - Software licenses and their renewal
  - Evergreening servers and storage
  - Ongoing platform development, coordination, administration



Researchers, Policy Analysts, & the Canadian Public

↑

Data Management & Liaison Librarians in Canadian Post-Secondary Institutions

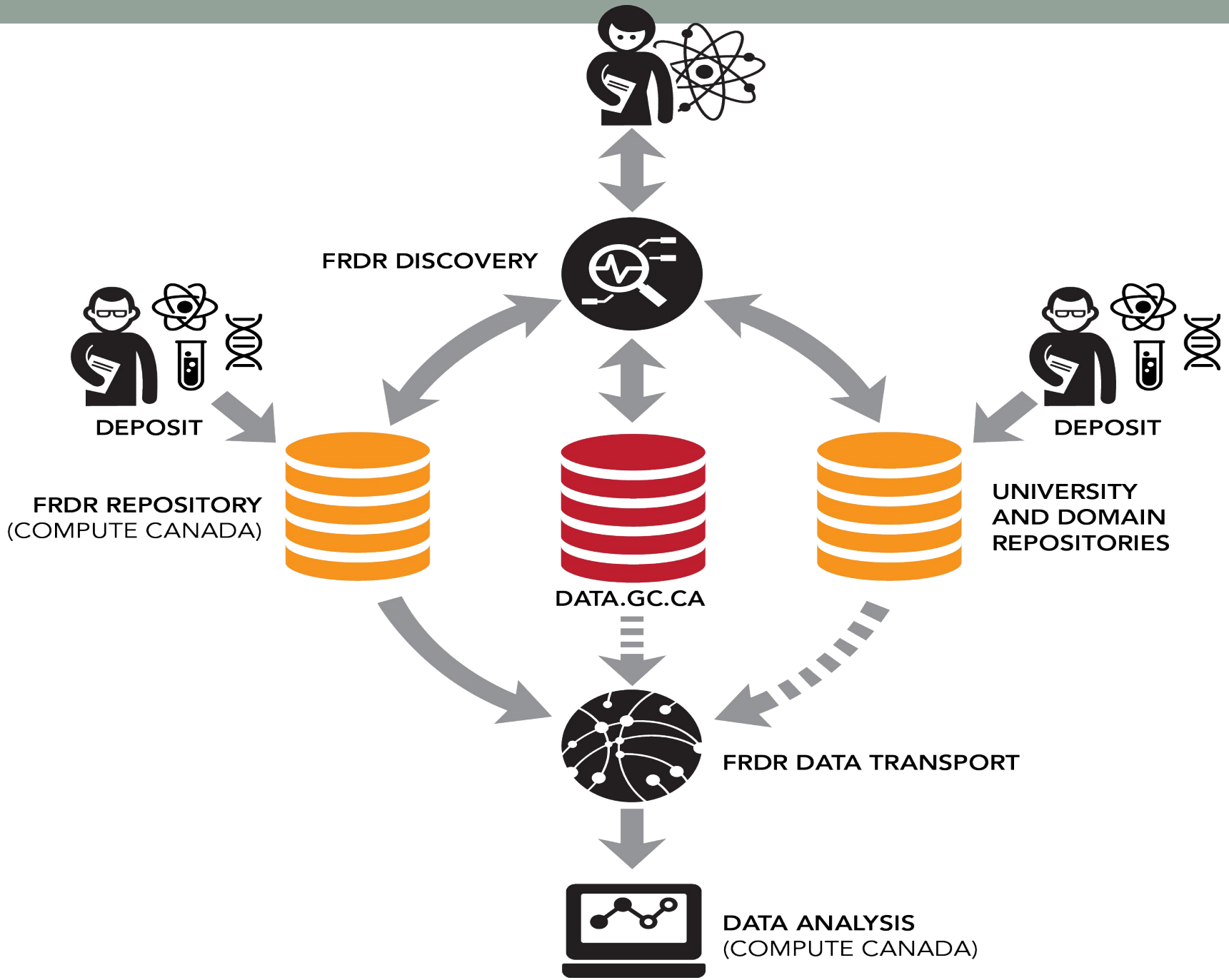
↑

Portage Secretariat (3) and Federated Research Data Service Staff (10)

↑

Contract infrastructure service provider (6)

A total of 19 FTEs are required for the Canadian Federated Research Data Service.


FRDR DISCOVERY
DEPOSIT
DEPOSIT
FRDR REPOSITORY
(COMPUTE CANADA)
DATA.GC.CA
UNIVERSITY
AND DOMAIN
REPOSITORIES
FRDR DATA TRANSPORT
DATA ANALYSIS
(COMPUTE CANADA)

Thank you