Portage Network for Research Data Management in Canada

BIREDIAL-ISTEC 2019

Aug 1, 2019

Lee Wilson, Service Manager
Portage/ACENET
lee.wilson@ace-net.ca

What we'll cover…

Background – who we are

Progress – what we've been doing

Future – where we're going

Canadian "data deficit" in the news…

THE GLOBE AND MAIL

IN DEPTH

In the dark: The cost of Canada's data deficit

Highlights gaps in:

What data are collected

How data are collected

&

Who has access to data once collected

ERIC ANDREW-GEE AND TAVIA GRANT
TORONTO
PUBLISHED JANUARY 26, 2019

Makes no mention of progress with research data


Innovation, Science and Economic Development Canada

Funding for DRI: Digital Research Infrastructure

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**.

Network of Experts

Network of Expertise Layer
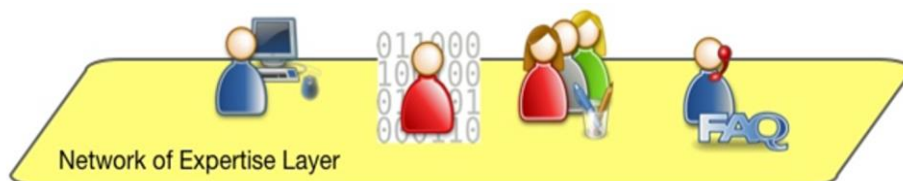
- Pan-Canadian network of RDM expertise

- Contribute to development of RDM resources, tools, training, and policy

Infrastructure Platform Layer

Infrastructure Platforms

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

Grounded in partnerships and collaboration

Portage Expert & Working Groups

Expert Groups

- Data Management Planning
- Curation
- Data Discovery
- Preservation
- Training
- Research Intelligence
- Data Repositories

Working Groups

- Dataverse North WG
- Responsible Research Data Management Practices for Sensitive Data
- FRDR Service Model WG
- FRDR Discovery Service WG
- Institutional RDM Strategies

Over 130 in-kind experts from over 60 organizations / institutions

These groups both *inform* and help *perform* what we do… Portage is community-driven

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

Funders as a catalyst for change

Portage RDM Support for Institutions & Researchers

Tri-Agency Policy

Institutional Strategy ✔
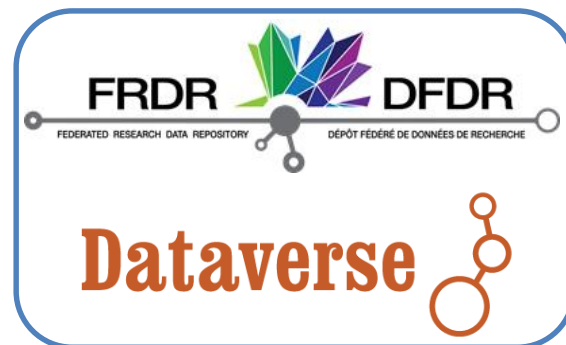
DMPs ✔

Deposit ✔

Portage

RDM Strategy Template

Portage

DMP Assistant

Sign In   Sign Up

National, online, bilingual, Data Management Planning Tool

Portage

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

Dataverse

Institutional RDM Strategies

Institutional Strategy ✓

DMPs

Deposit

RDM Strategy Template

- **Institutions convening diverse teams to create RDM strategies (research services, ethics, libraries, IT, administration)**

- **Teams have fed into broader RDM awareness and culture change on campuses**

- **RDM strategies are setting the stage for future policy development**

- **Portage Research Intelligence Expert Group is conducting a short, targeted survey to measure institutional readiness for this aspect of the policy**

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

Data Management Planning

47 institutional accounts and over 5700 registered users

DMP Assistant

- **Facilitates researcher-compliance with the second pillar of the emerging RDM policy**

- **Migration to DMP 2.0 (DMP Roadmap) promises to improve researcher experience**

- **Ability to 'opt-in' to share DMPs will make strong examples more readily available to researchers**

- **APIs to facilitate sharing among systems will help reduce administrative burden on researchers**

'enter once – use many'

- **DMPs are often our 'first contact' with researchers, and are crucial to the culture change we're advocating for on our campuses**

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

Data Deposit:
Canadian Repository Landscape

Institutional Strategy

DMPs

Deposit

General Repositories

Government Repositories

Domain-specific Repositories

re3data.org

Data Deposit:
Canadian Repository Landscape

Institutional Strategy

DMPs

Deposit ✓

General Repositories

Government Repositories

Domain-specific Repositories

re3data.org

Two National Repository Options

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

Federated Research Data Repository

Dataverse
North
National Instance of Dataverse

National Discovery Service

Portage Data Discovery and FRDR Service Working Groups & Compute Canada FRDR Team

compute | calcul
canada | canada

Portage Dataverse North Working Group and OCUL Scholars Portal Dataverse Team

OCUL Ontario Council of University Libraries

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

FRDR: Federated Research Data Repository

FRDR: Federated Research Data Repository

Federated Research Data Repository

Main Features:

'Big Data' upload/download

Maintains file hierarchies

Scalable storage

National **Repository** Option

Preservation Processing capacity

Currently in Limited Production

Strategic Relevance
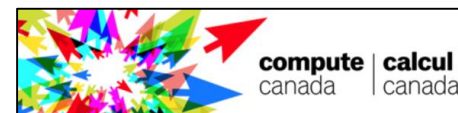
Strategic partnership with Compute Canada (ARC)

Fills key gaps in the RDM ecosystem

Complements existing repository solutions

https://www.frdr.ca/

FRDR facts…

- **Limited Production deposits as of May. 2019:**
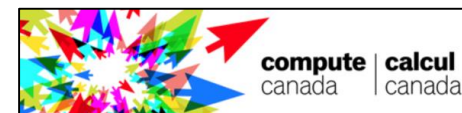 - **18** partnering research groups
  - **86** datasets published (4.5 TB)

FRDR facts…

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

- **Limited Production deposits** **as of May. 2019:**
  - **18** partnering research groups
  - **86** datasets published (**4.5 TB**)

Globus

compute | calcul
canada | canada

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

Globus File Transfer

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

transfer

2 Globus moves the data for you

secure endpoint A

secure endpoint B

1 you submit a transfer request

3 Globus notifies you once the transfer is complete

Globus overcomes limitations of 'https' web protocol

- **Faster than https**
- **Clean recovery from interuptions**
- **No practical limit on file sizes or number of files**
- **Retains file hierarchies**

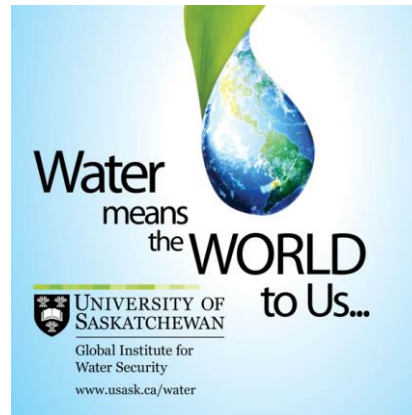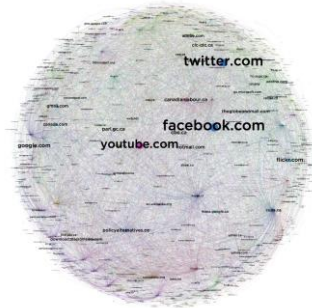**Researcher initiates transfer, and is notified when complete**

compute | calcul
canada | canada

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

https://www.globus.org/data-transfer

FRDR Limited Production Projects

MERIDIAN

"Big data solutions for the marine acoustic community"

Dalhousie University

Web Archives for Longitudinal Knowledge

SuperDARN CANADA

Water means the WORLD to Us...

UNIVERSITY OF SASKATCHEWAN
Global Institute for Water Security
www.usask.ca/water

universitycapebreton.ca
Eastern and Central Europeans come ashore...

MOUNTAIN LEGACY PROJECT

MOUNTAIN LEGACY PROJECT

"Capturing Change in Canada's Mountains"
U. Victoria

the water INSTITUTE University of Waterloo

The Water Institute
U. Of Waterloo

"Breaking boundaries in water research"

CARISMA

Canadian Array for Realtime Investigations of Magnetic Activity

U. Alberta

S.D.S.
STONE TOOLS, DIET, AND SOCIALITY

compute | calcul
canada | canada

U. Calgary

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

Dataverse

Dataverse

Main Features:

Mature platform
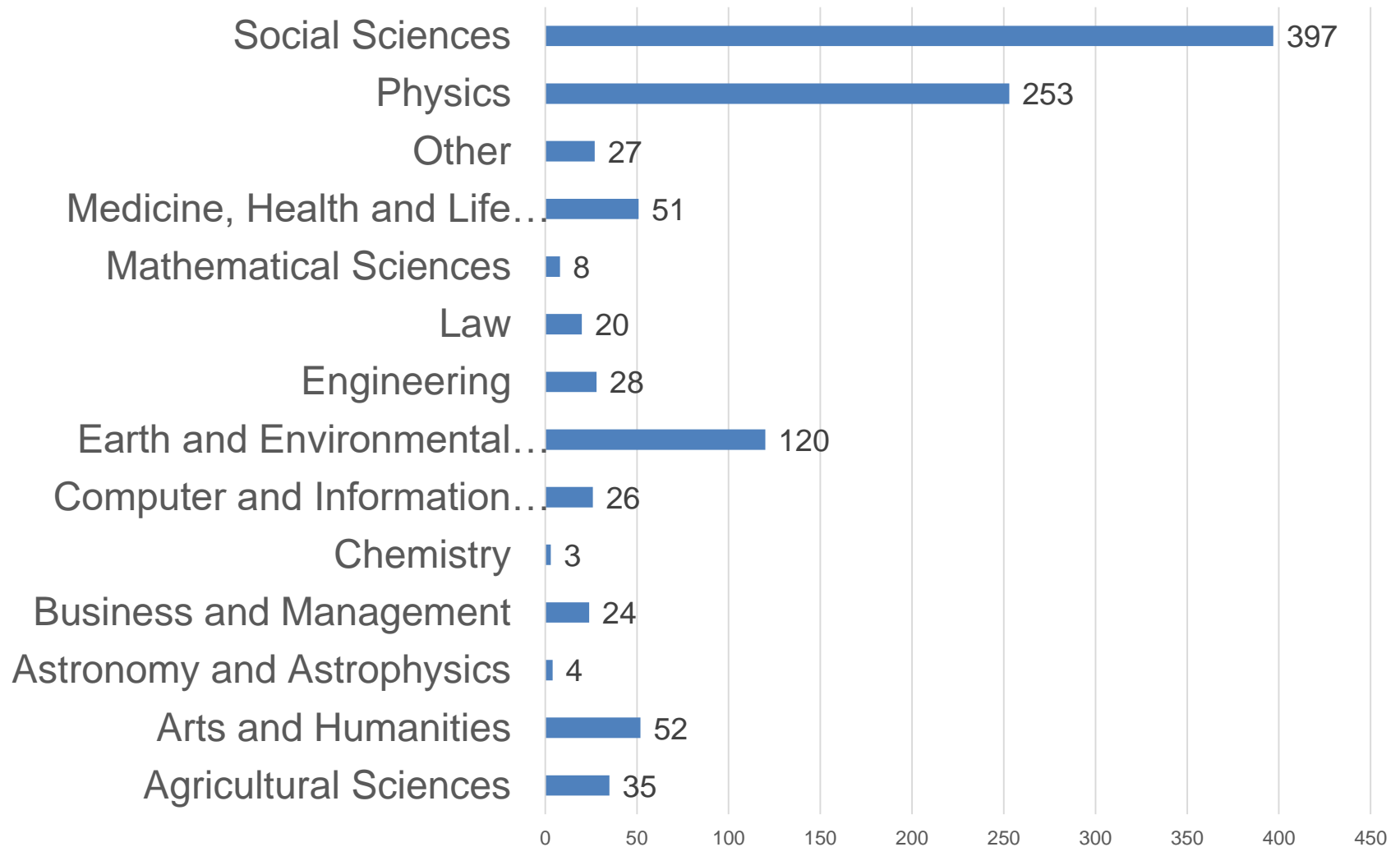
Widely adopted

User-friendly

Brandable

Localized service

Developing National Hosted Instance

Dataverse used widely in Canada with 1000's of data files deposited

Dataverse

Dataverse facts…

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

Discipline	Count
Social Sciences	397
Physics	253
Other	27
Medicine, Health and Life…	51
Mathematical Sciences	8
Law	20
Engineering	28
Earth and Environmental…	120
Computer and Information…	26
Chemistry	3
Business and Management	24
Astronomy and Astrophysics	4
Arts and Humanities	52
Agricultural Sciences	35

Dataverses in Canada

6 Canadian Dataverse instances:
- **4** standalone (U Alberta, Dalhousie, UNB, U Manitoba)
- **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

< 2 TB

Recently signed an SLA to allow Quebec universities to use Scholars Portal Dataverse

portage

National Discovery Layer

Globe & Mail, 28 January, 2019:

"Some researchers would like to see a **national open data portal** that incorporates not just the holdings of the federal government, but the numbers gathered by cities and provinces as well."

"There needs to be a **central clearinghouse of publically available datasets in Canada**"

Fuyuki Kurasawa, Director, Global Digital Citizenship Lab, York University.
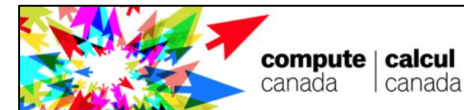
National Discovery Layer

National Discovery Layer

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

All searchable via
https://www.frdr.ca

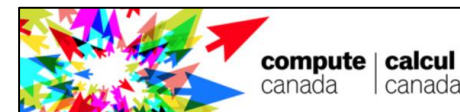**Improve discovery of research data/metadata across Canada**

Break down repository siloes

Drive traffic to existing repository sites

- **FRDR searches between Nov 2017- April 2019**
  - *Users*:   6,792
  - *Sessions*:   11,923
  - *Pageviews*:   41,553

compute | calcul
canada | canada

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

National Discovery Layer

All searchable via
https://www.frdr.ca

Improve discovery of research data/metadata across Canada

Break down repository siloes

Drive traffic to existing repository sites

Data Deposit
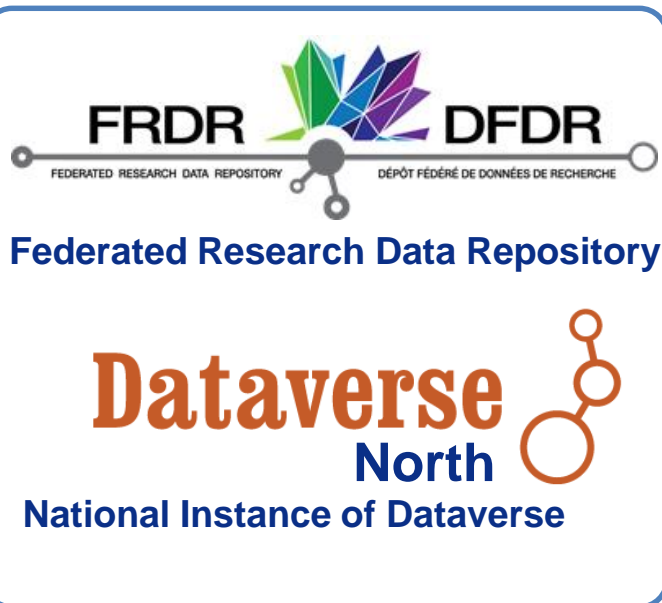
Repository Updates

Federated Research Data Repository

Dataverse North
National Instance of Dataverse

**New Data Repositories Expert Group → DREG**

- **Progress toward a national instance of Dataverse (Service Level Agreement template developed)**

- **Improved Geospatial visualization & discovery through CANARIE RDM funding**

- **Developing 'use-case' narratives for Dataverse, FRDR, & Domain repositories**

- **New Working Group to improve metadata for FRDR discovery layer**

- **Bringing on new Canadian repositories for discovery & working with international initiatives**

Data Deposit

Repository Updates

Federated Research Data Repository

Dataverse North
National Instance of Dataverse

New Data Repositories Expert Group → DREG

- **Progress toward a national instance of Dataverse (Service Level Agreement template developed)**

- **Improved Geospatial visualization & discovery through CANARIE RDM funding**

- **Developing 'use-case' narratives for Dataverse, FRDR, & Domain repositories**

- **New Working Group to improve metadata for FRDR discovery layer**

- **Bringing on new Canadian repositories for discovery & working with international initiatives**

Geodisy: Geospatial Discovery

Funded by CANARIE (RDM-059)
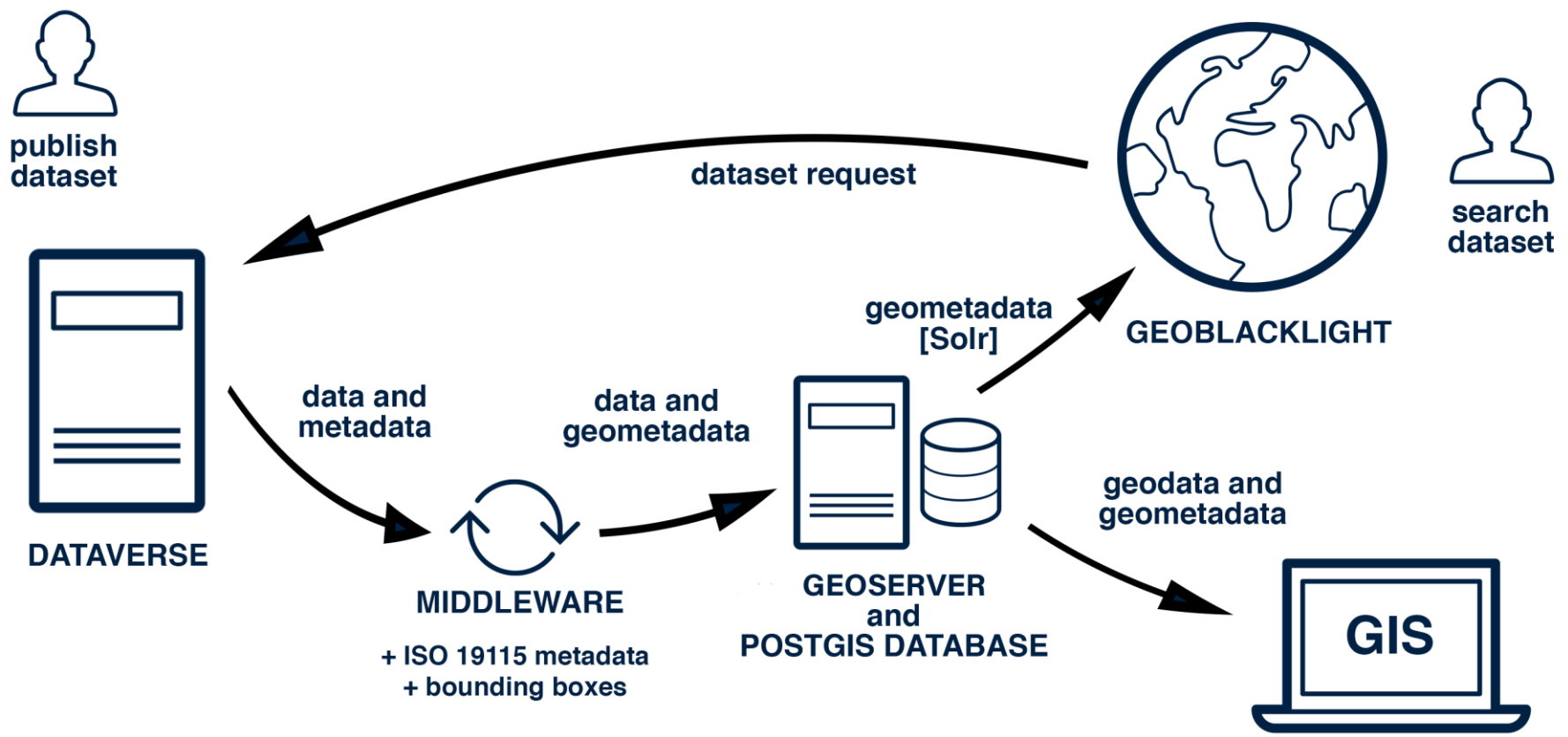
New FRDR Working Group to improve discovery

Mapping researcher-supplied subject keywords to the **FAST** schema using **OpenRefine**
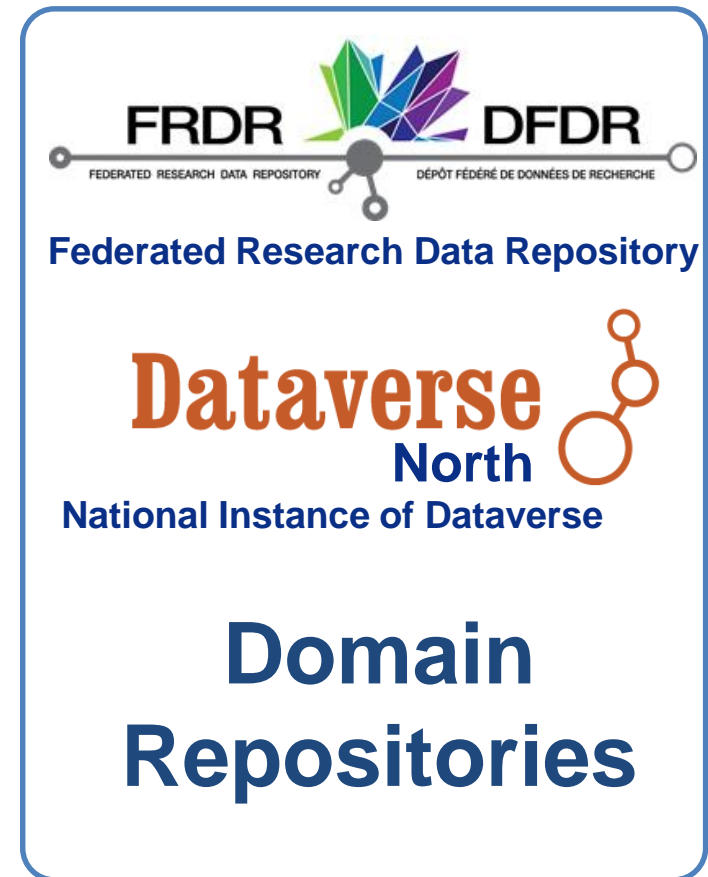

FAST and the FRDR

RVM: Répertoire de vedettes-matière

Developed by l'Université Laval

(French equivalent of LCSH)

→ Have goal of mapping the RVM to FAST over the next year

'DREG'

Coordinating work of current national repository platforms (Dataverse & FRDR) and developing connections/collaborations with domain repositories

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

Federated Research Data Repository

Dataverse
North
National Instance of Dataverse

Domain Repositories

Continuum of Training

Webinars

Coming in the future! Potential webinar topics include: RDM 101; DMPs 101; Institutional Repositories; Dealing with Sensitive Data; Data Curation 101; Data Preservation 101; How to Develop an Institutional Strategy; and others.

On-Demand Training

On-demand assistance is available to those organizing RDM training on Canadian academic campuses who would like help developing and delivering training and guidance materials in all areas of RDM.

Supporting national integration of Digital Research Infrastructure (DRI)

RDM in transition…

2018 Canadian Budget

Harnessing Big Data

"The Government proposes to provide **$572.5 million** over five years, with $52 million per year ongoing, to implement a **Digital Research Infrastructure Strategy** that will deliver more open and equitable access to advanced computing and big data resources to researchers across Canada."
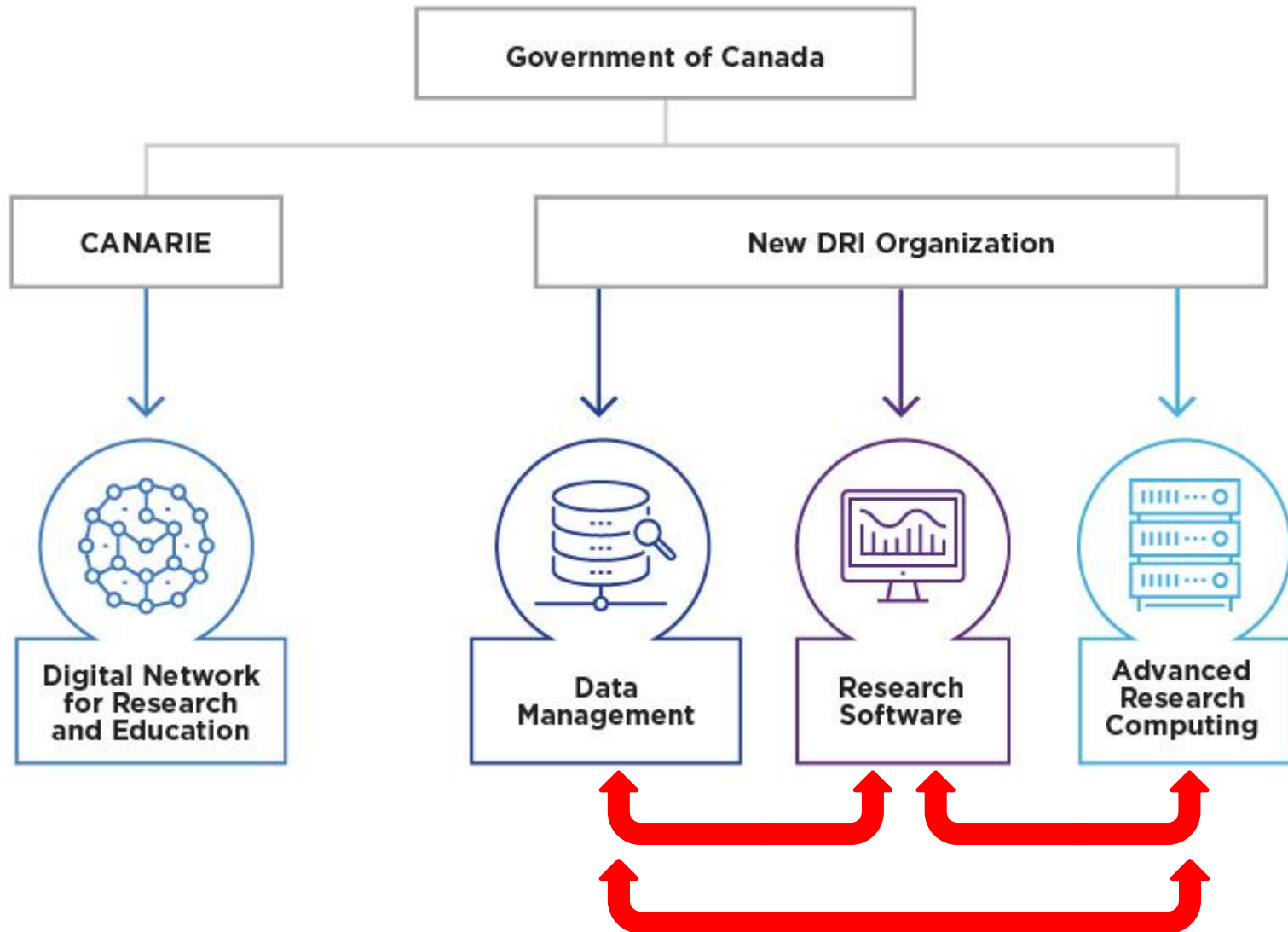
https://www.budget.gc.ca/2018/docs/plan/budget-2018-en.pdf
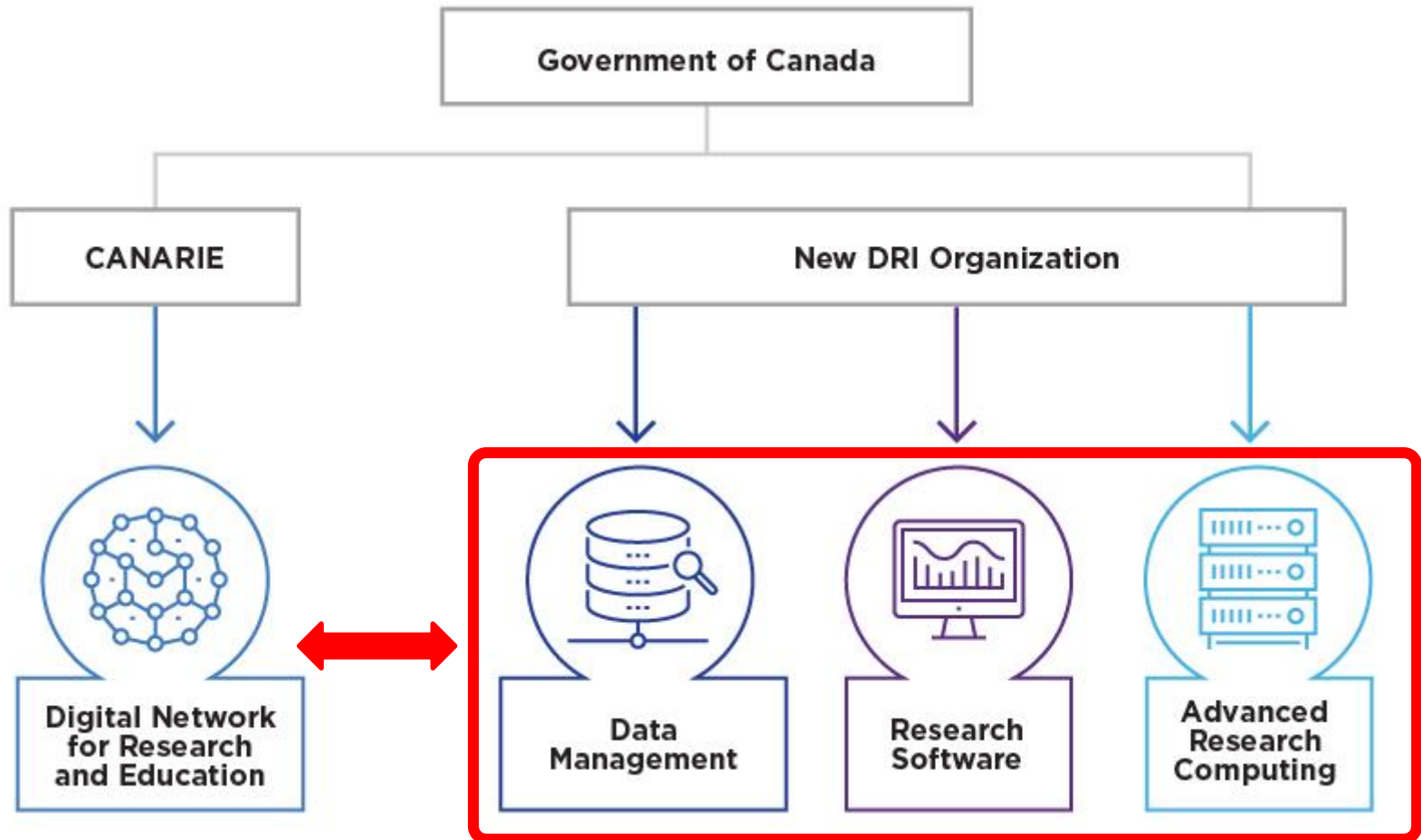
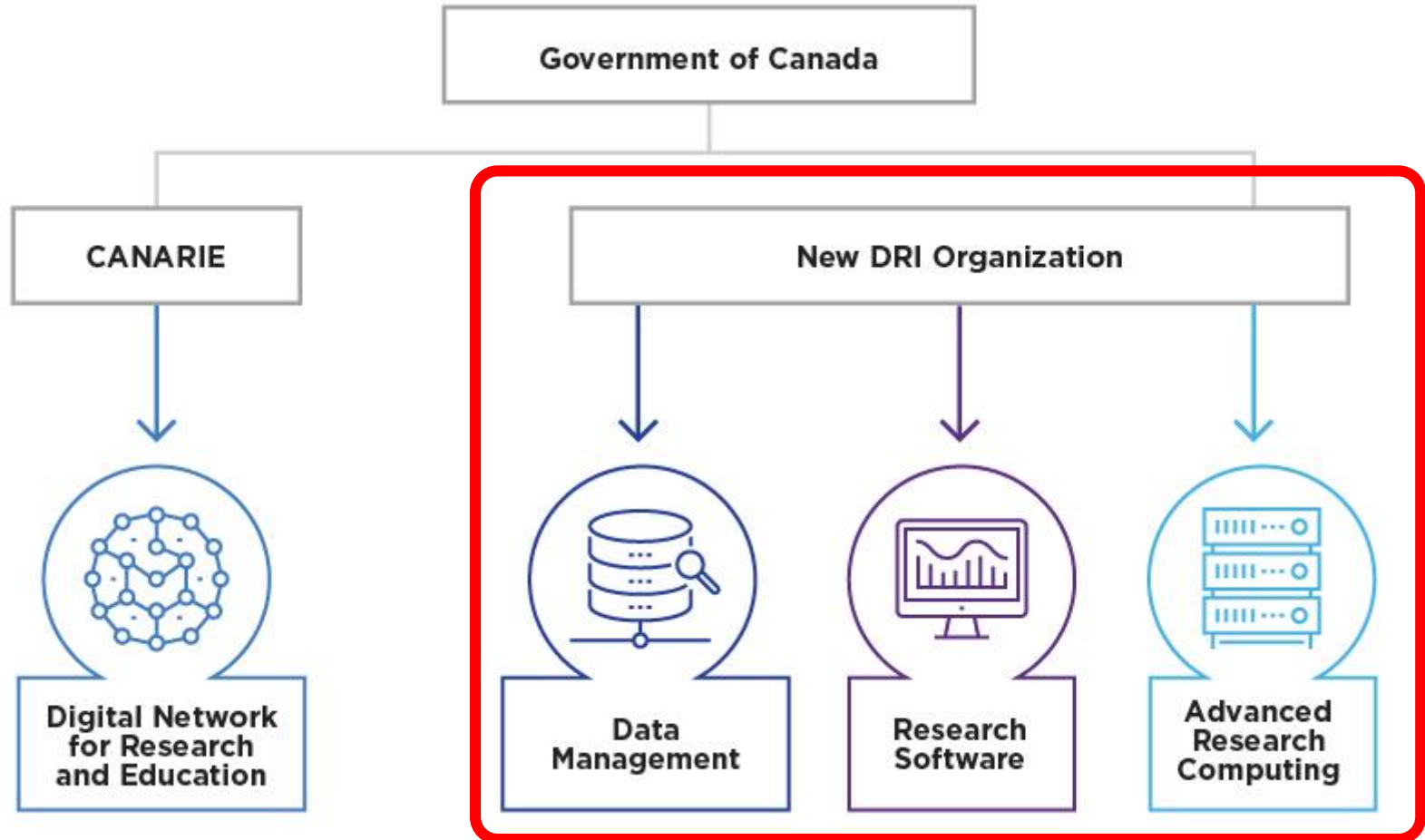# Current DRI Landscape
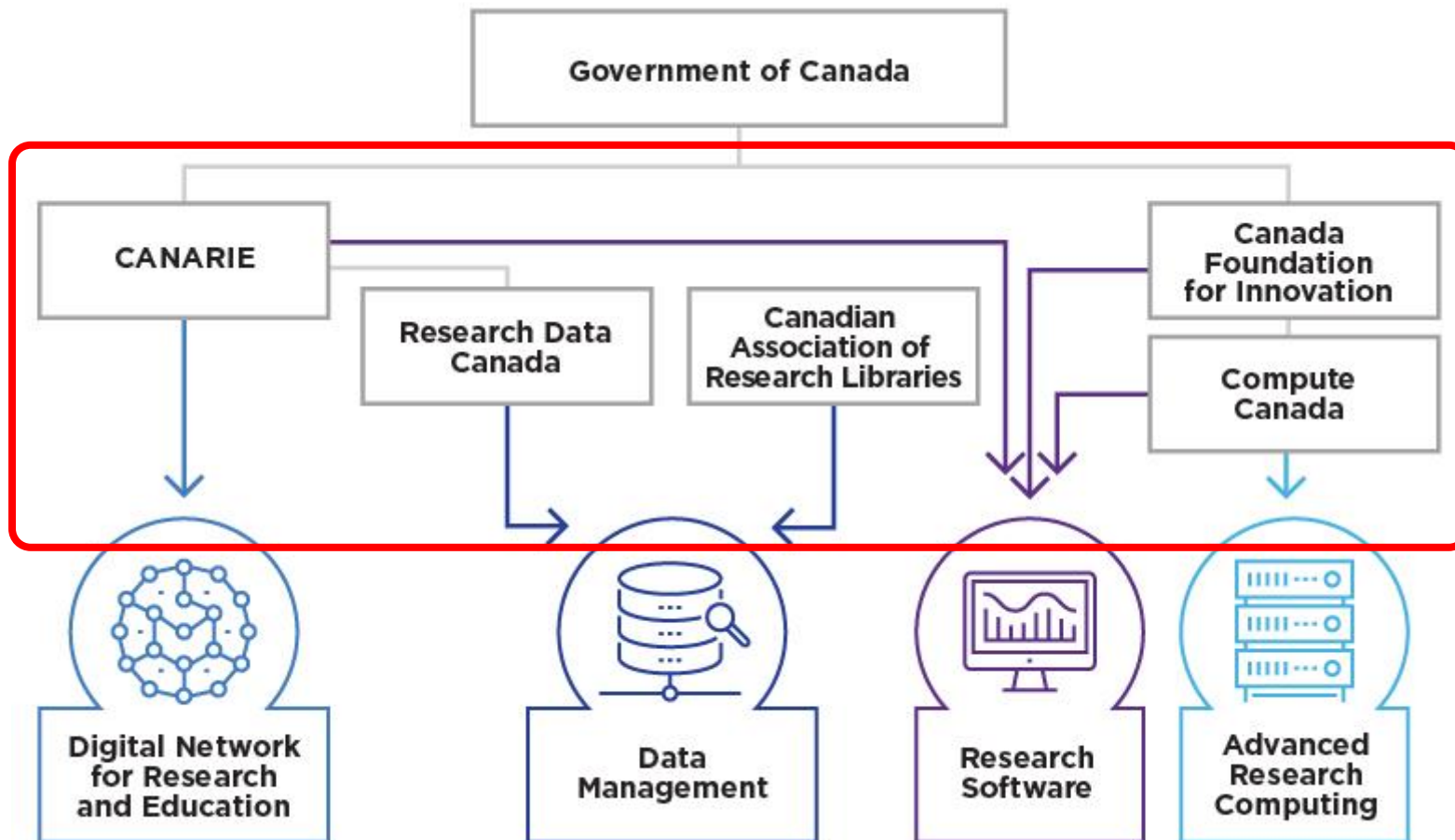
Future-state DRI Landscape
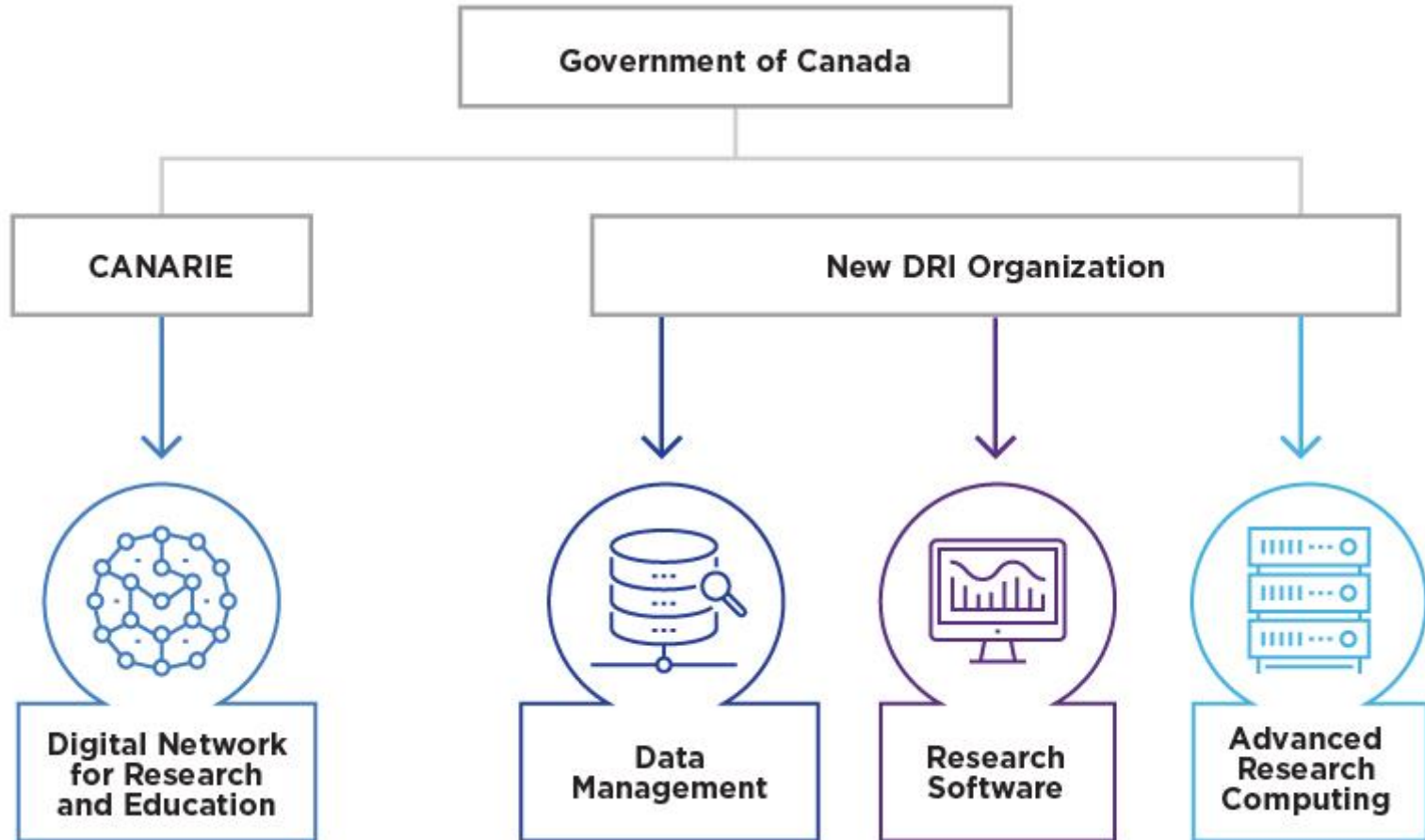
Future-state DRI Landscape

Future-state DRI Landscape

Creating a new not-for-profit organization…

Applicant Guide

All applications must include the following elements:

1. Program Application Form

2. Completed proposal that contains the following sections:
   a. Executive Summary
   b. Implementation
   c. Collaboration and Engagement
   d. Governance and Management

3. Appendices*
   a. An outline of engagement activities over the next 2 years (until approximately March 2021)
   b. Organizational chart for the governance structure
   c. Competency matrix for the planned Board of Directors, if available
   d. Resumes and biographies for all listed applicants on the Application Form
   e. Letters of Support

… and do it all in 18 business days

Ahead of the looming election call…

http://natoassociation.ca/get-involved-the-canadian-federal-elections/

A sprint to the starting line

https://commons.wikimedia.org/wiki/File:Numbers_of_starting_line_track.jpg

Engage DRI

Feedback from 50+

& 30 letters of support

Collaborating for the future of Canada's Digital Research Infrastructure

Comments and questions may be directed to the email address info@engagedri.ca.

Summary of process…

Application Signatories ✓ **May 6 deadline**

⬇

Government Approval in Advance of Writ being dropped

- -

⬇

Applicant Board   **Advisory Group**

And there is a lot that needs to happen during the transition period

⬇

Inaugural Board

⬇

New Organization
for
Digital Research Infrastructure

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

A. National Data Stewardship Support

Enhance the Portage Secretariat through National Coordinator roles for Curation, Preservation, Training, and Discovery & Metadata

B. RDM Platform Support and Development

Help Develop & Operationalize Several National, Collaborative RDM Platforms and Services

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

A. National Data Stewardship Support

- Portage Director
    - Project Manager – finances & reporting
    - Communications Officer (0.5 FTE) – comms strategy
    - Training Coordinator
    - Service Manager
        - Curation Coordinator
        - Preservation Coordinator (0.5 FTE)
        - Discovery & Metadata Coordinator

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

B. RDM Platform Support and Development

- **FRDR to full Production**
    - Project Manager
    - Developer
    - Platform maintenance, development, & storage
- **Dataverse North** (a Portage Working Group)
    - Augmented storage capacity
- **DMP Assistant**
    - Platform migration & service development
        - Programmer
        - Program Management
    - RMI - DMP Integration Project
        - Metadata specialist

More to come from Portage & others!

Questions?

portagenetwork.ca

Lee Wilson, Service Manager, CARL Portage

lee.wilson@ace-net.ca