

portage

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

Canadian Leadership in RDM: Portage Partnerships & Progress

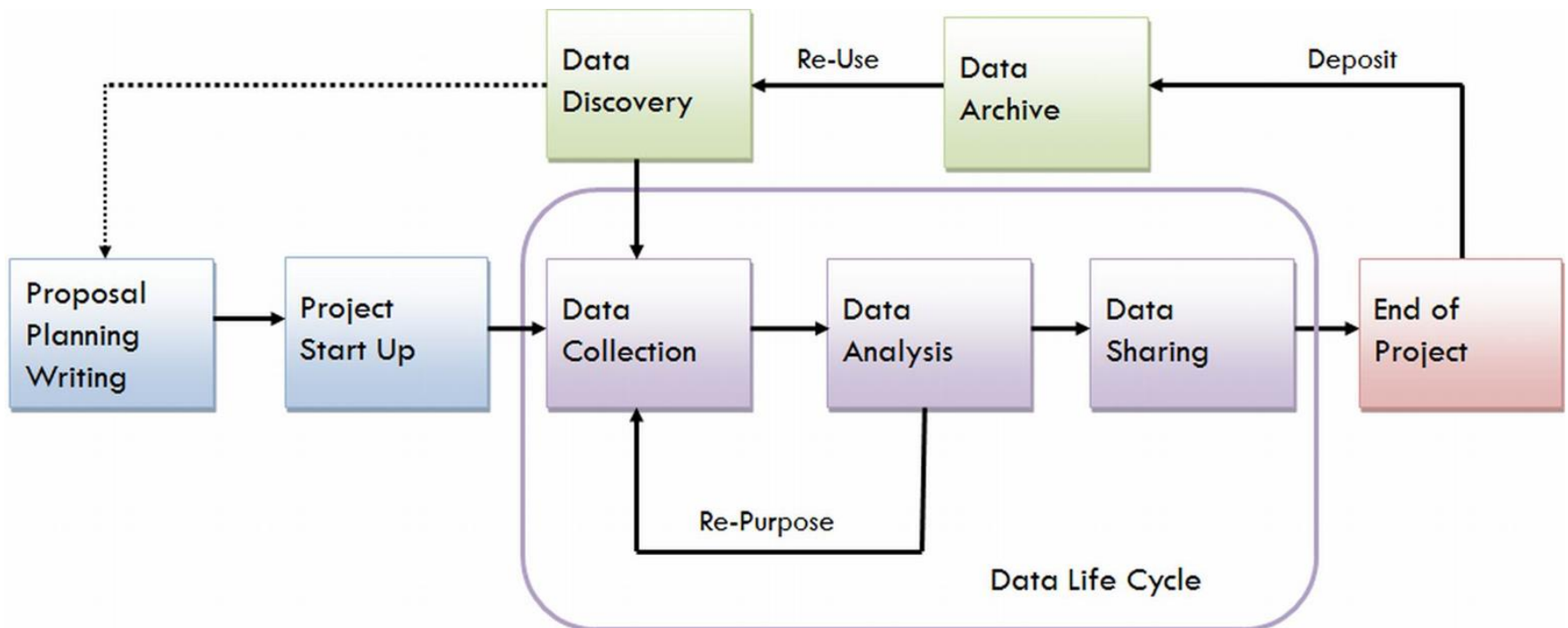
2018 McMaster Research Data Forum:
Common Challenges, Shared Solutions
Dec 12, 2018

Jeff Moon, Director, Portage



What is Research Data Management?

A broad term used to describe the **structure, organization, maintenance, and overall stewardship** of research data.

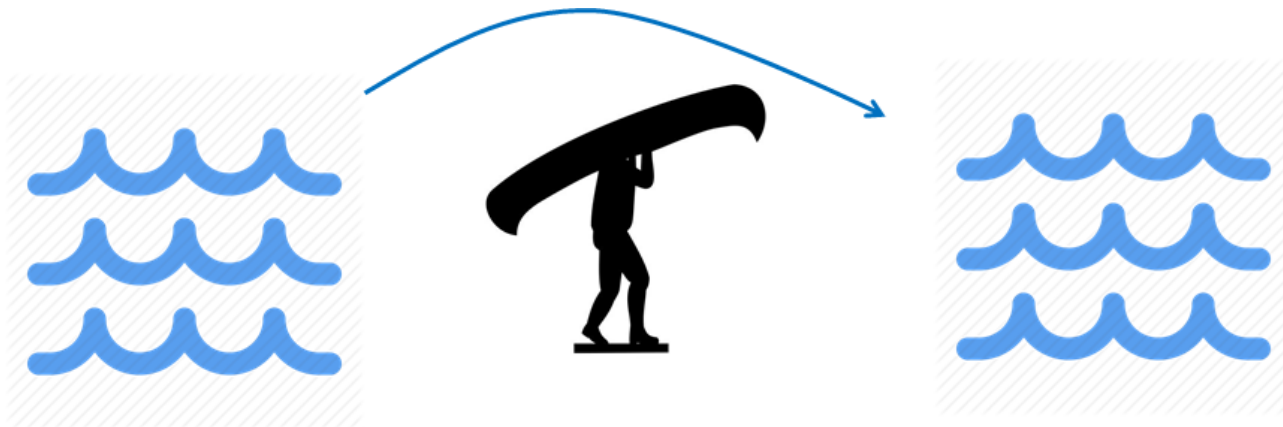
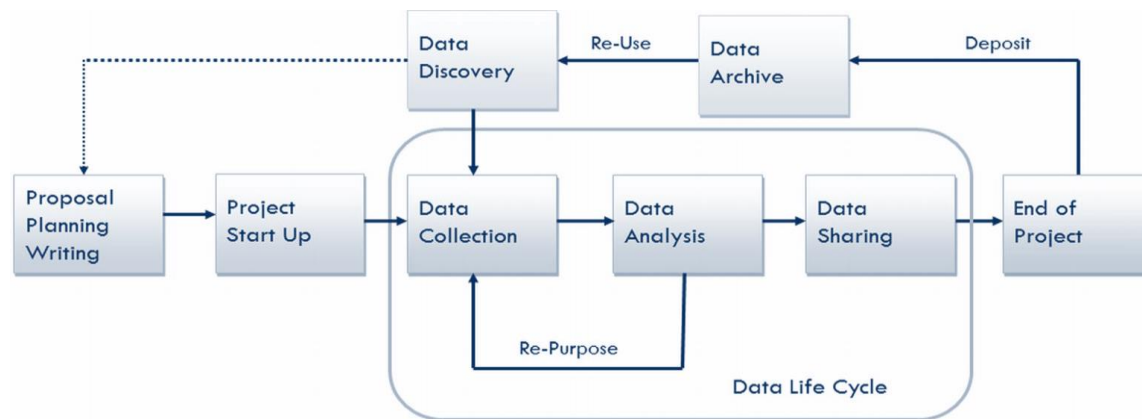


Research Lifecycle

What is Portage?



“Portage is a CARL-sponsored national network of Research Data Management experts **working collaboratively to build capacity and better coordinate RDM activities in Canada**”



Outline

- **Federal RDM policy developments**
- **Further context for RDM in Canada**
- **Goals and outputs of Portage**
- **Developing a shared vision for RDM in Canada**

Outline

- **Federal RDM policy developments**
- Further context for RDM in Canada
- Goals and outputs of Portage
- Developing a shared vision for RDM in Canada

Research Data Management – Recent Policy Developments



Canadian Institutes of Health Research
Instituts de recherche en santé du Canada



NSERC
CRSNG

SSHRC  CRSH

Research Data Management – Recent Policy Developments



Research Excellence
Research Dissemination
Research Impact
Research Best Practice

Research Data Management – Recent Policy Developments



Research Excellence

Research Dissemination

Research Impact

Research Best Practice

Project Management

Reproducibility

Avoiding Duplication

Meeting
Journal Guidelines



Research Data Management – Recent Policy Developments



Research Excellence
Research Dissemination
Research Impact
Research Best Practice

Data Sharing

Citation

Interdisciplinarity

OPEN  ACCESS Freely available online

 PLoS one

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

Research Data Management – Recent Policy Developments

Research Excellence
Research Dissemination

Research Impact

Research Best Practice

Within Science

Within Society

**In Policy
Development**



Research Data Management – Recent Policy Developments



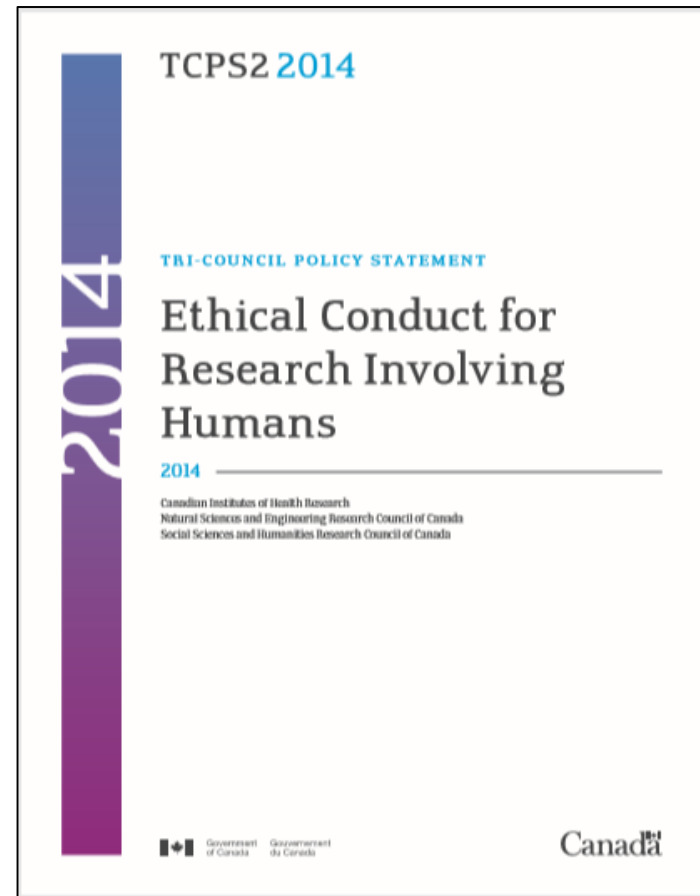
Research Excellence
Research Dissemination
Research Impact
Research Best Practice

Transparency

Trust

Ethics

Responsible Use of Public Funds



Research Data Management – Recent Policy Developments



•2013

- Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada

•2016

- Tri-Agency Statement of Principles on Digital Data Management

•2017-2018

- Draft Tri-Agency Research Data Management Policy

Research Data Management – Recent Policy Developments



Draft Tri-Agency Research Data Management Policy

Online Consultations

OPENED: June 5, 2018

CLOSED: Sept 30, 2018

SSHRC: ResearchData-Donneesderecherche@sshrc-crsh.gc.ca

NSERC: ResearchData-Donneesderecherche@nserc-crsng.gc.ca

CIHR: ResearchData-Donneesderecherche@cihr-irsc.gc.ca

Research Data Management – Recent Policy Developments



**Draft
Tri-Agency
RDM
Policy**

**Institutional
Strategy**



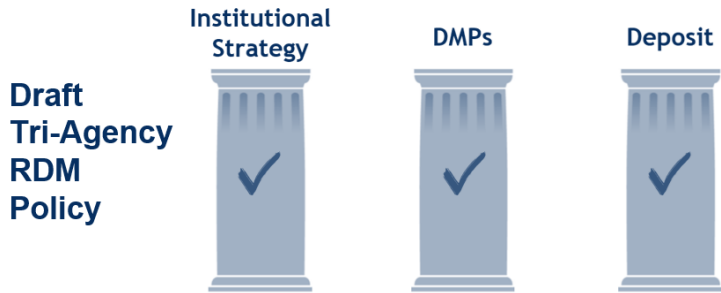
DMPs



Deposit



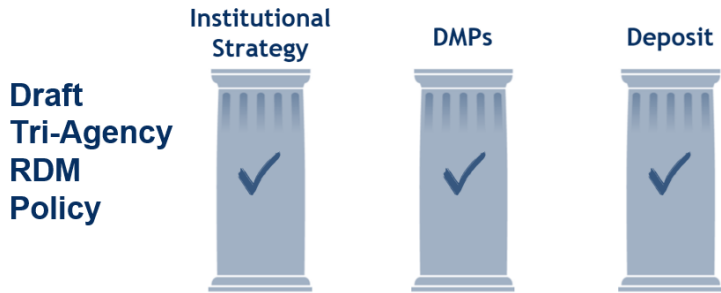
Research Data Management – Recent Policy Developments



1. Institutions: Institutional Strategy

- Institution administering Tri-Agency funds would need an **institutional research data management strategy** outlining how the institution will provide its researchers with an environment that enables and supports world-class RDM practices.
- Strategies would be **posted on the institution's website**, with contact information to direct inquiries about the strategy.

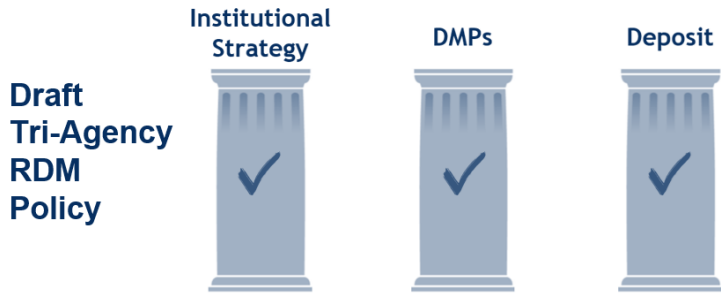
Research Data Management – Recent Policy Developments



Why Require Institutional Strategies?

- Recognizes the role of institutions in RDM
- Opportunity to identify and address gaps at campus level
- Opportunity to investigate local solutions while leveraging regional, provincial, and national services and platforms
- Helps inform Tri-Agencies on institutional data management capacity
- Serves as foundation for potential requirements that follow

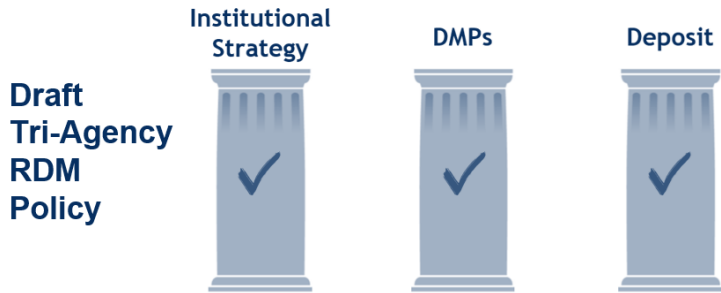
Research Data Management – Recent Policy Developments



2. Researchers: Data Management Plans

- **DMPs required for research projects** supported wholly or in part by Tri-Agency funds. Submission of DMPs would be a condition of the release of grant funds.
- *For specific funding opportunities*, DMPs would be **submitted at time of application** to the appropriate agency, for possible consideration in the adjudication process.

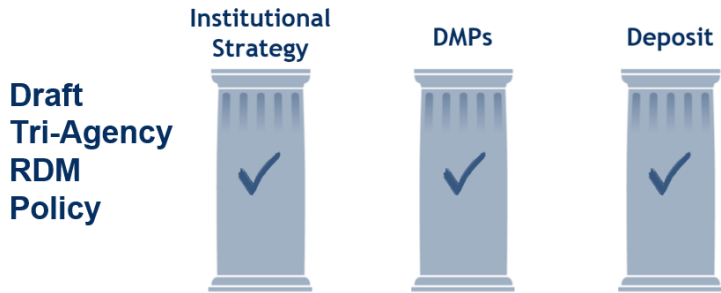
Research Data Management – Recent Policy Developments



Why Require Data Management Plans?

- DMPs are an emerging international best practice
- DMPs help researchers anticipate RDM opportunities and challenges early in the research life cycle
- Researchers report that developing DMPs helps improve research processes, methodologies, and outcomes
- DMPs could feed into research ethics approval processes
- DMPs help anticipate issues related to ownership of data, potential for data sharing, etc.

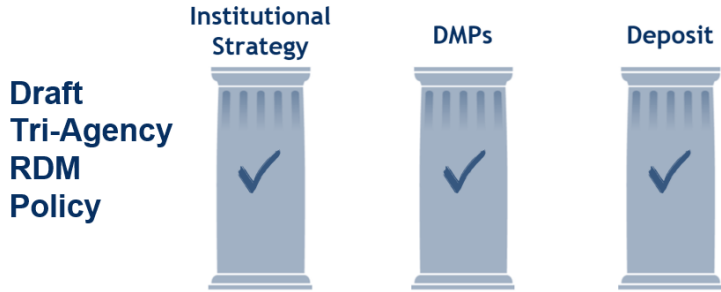
Research Data Management – Recent Policy Developments



3. Researchers: Data Deposit

- Research **data**, **metadata**, and **code** supporting journal publications, pre-prints and other research outputs arising from funded research would be **deposited in an appropriate public repository** or other platform that will ensure safe storage, preservation, curation, and (if applicable) access to the data.

Research Data Management – Recent Policy Developments



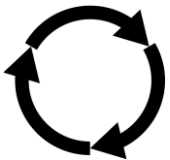
Why Require Data Deposit?



- Data, metadata, and code safe & available to data creator & appropriately to others in future



- Ensure proper use of public funds



- Facilitate reproducibility



- Facilitate appropriate data sharing

Research Data Management – Recent Policy Developments



More on Policy Requirements later...

Outline

- Federal RDM policy developments
- **Further context for RDM in Canada**
- Goals and outputs of Portage
- Developing a shared vision for RDM in Canada

Other factors making Research Data Management important...

**Growing
public
awareness
of data**

**Recognition
of the
value of data in
advancing research**

**Addressing
pressures from
commercial
concerns**

**Need to verify
research results**

RDM

**Government
policy directions**

**Growing demand
for data**

Journal policies

Growing public awareness... RDM in the news



"Now what we're getting is massive amounts of data and really the biggest challenges we have are data management."

- Mike Sutor

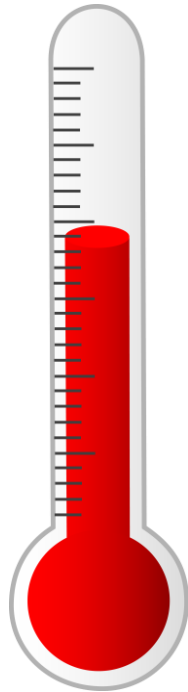
Growing public awareness... RDM in the news

Live Science > Space

'Lost' NASA Tapes Show Humans Sort of Caused Global Warming on the Moon Too

By Brandon Specktor, Senior Writer | June 13, 2018 06:05am ET


'...hundreds of the reels of magnetic tape holding those records went missing nearly 40 years ago, thanks to an archival blunder'




JOURNAL OF GEOPHYSICAL RESEARCH

Planets

AN AGU JOURNAL

Research Article |  Full Access

Examination of the Long-Term Subsurface Warming Observed at the Apollo 15 and 17 Sites Utilizing the Newly Restored Heat Flow Experiment Data From 1975 to 1977

S. Nagihara , W. S. Kiefer, P. T. Taylor, D. R. Williams, Y. Nakamura

First published: 25 April 2018 | <https://doi.org/10.1029/2018JE005579>

<https://geschichte.inf.tu-dresden.de/?q=EN/NORMAL/expo06/content>

http://www.clipartpanda.com/clipart_images/red-thermometer-clip-art-8329215

<https://www.livescience.com/62805-moon-surface-warming-apollo-astronauts.html>

Recognition of the value of data: *within a discipline*

Practice and Research in Nursing (PRN)
Wound Studies, 1999-2009 [Canada]



Scholars Portal **Dataverse**

Queen's University Library Research Data Archive Dataverse
<https://hdl.handle.net/10864/CORX8>

“Many thanks for sharing the raw data. If more research groups and industry would do this we could save more lives and limbs!”

Ronald J. Shannon, MPH
President, Global Health Economic
Projects, LLC
56 Via Da Vinci, Clifton Park, NY 12065

Consensus opinion on compression therapy and venous ulcers

Welsh, Lynn. **Primary Health Care (2014+)**; London Vol. 26, Iss. 4, (Apr 2016): 26.

DOI:10.7748/phc.26.4.26.s28

Recognition of the value of data: *between disciplines*

Five Weird Archives That Scientists Use to Study Past Climates

When tree rings, ice cores, and cave formations can't cut it, try your luck with whale earwax or bat poop.



Data collected on
tree rings
ice cores
cave formations
whale earwax &
bat poop

& sediment in ancient
Roman aqueducts

now being used by
climate scientists to
study global warming

Citation: Wendel, J., and M. Kumar (2018), Five weird archives that scientists use to study past climates, *Eos*, 99, <https://doi.org/10.1029/2018EO095483>. Published on 30 March 2018.

Thanks to **Kevin Lindstrom**, Physical Sciences and Engineering Librarian, Woodward Library, 2198 Health Sciences Mall, University of British Columbia, for pointing out this citation

The need to *verify* research results

Why scientists need to do more about research fraud

Richard P Grant

Scientific misconduct is more than just an academic problem - it has repercussions for real people



Addressing pressures from commercial concerns



ELSEVIER



Research Data
Services

SPRINGER NATURE
Research Data Support

- *Advice on finding repositories*
- *Advice on complying with policies*
- *Check for presence of sensitive information*
- *Advice on data availability statements/data citation*
- *Ensure accessibility of files for end users*
- *Easy upload to Springer Nature Figshare*
- *Ensure files match metadata & assoc. pubs*
- *Organize files into logical structure & collections*
- *Match authors to associated publications*
- *Copy-edit submitted metadata*
- *Generate Digital Object Identifier (DOI)*
- *Categorize content according to industry standard definitions*
- *Enable embargoes with private, anonymous access enabled*
- *Link and synchronize dataset release to associated publications*
- *Enhancement report and feedback from Research Data Editor*

Addressing pressures from commercial concerns



- *Advice on finding repositories*
- *Advice on complying with policies*
- *Check for presence of sensitive information*
- *Advice on data availability statements/data citation*

**“My complaint is they're
"solving" the ‘easy’ RDM
problems, for a fee”**

- *Copy-edit submitted metadata*
- *Generate Digital Object Identifier (DOI)*
- *Categorize content according to industry standard definitions*
- *Enable embargoes with private, anonymous access enabled*
- *Link and synchronize dataset release to associated publications*
- *Enhancement report and feedback from Research Data Editor*



Research Data
Services

SPRINGER NATURE
Research Data Support

And we can now include Google in the mix...

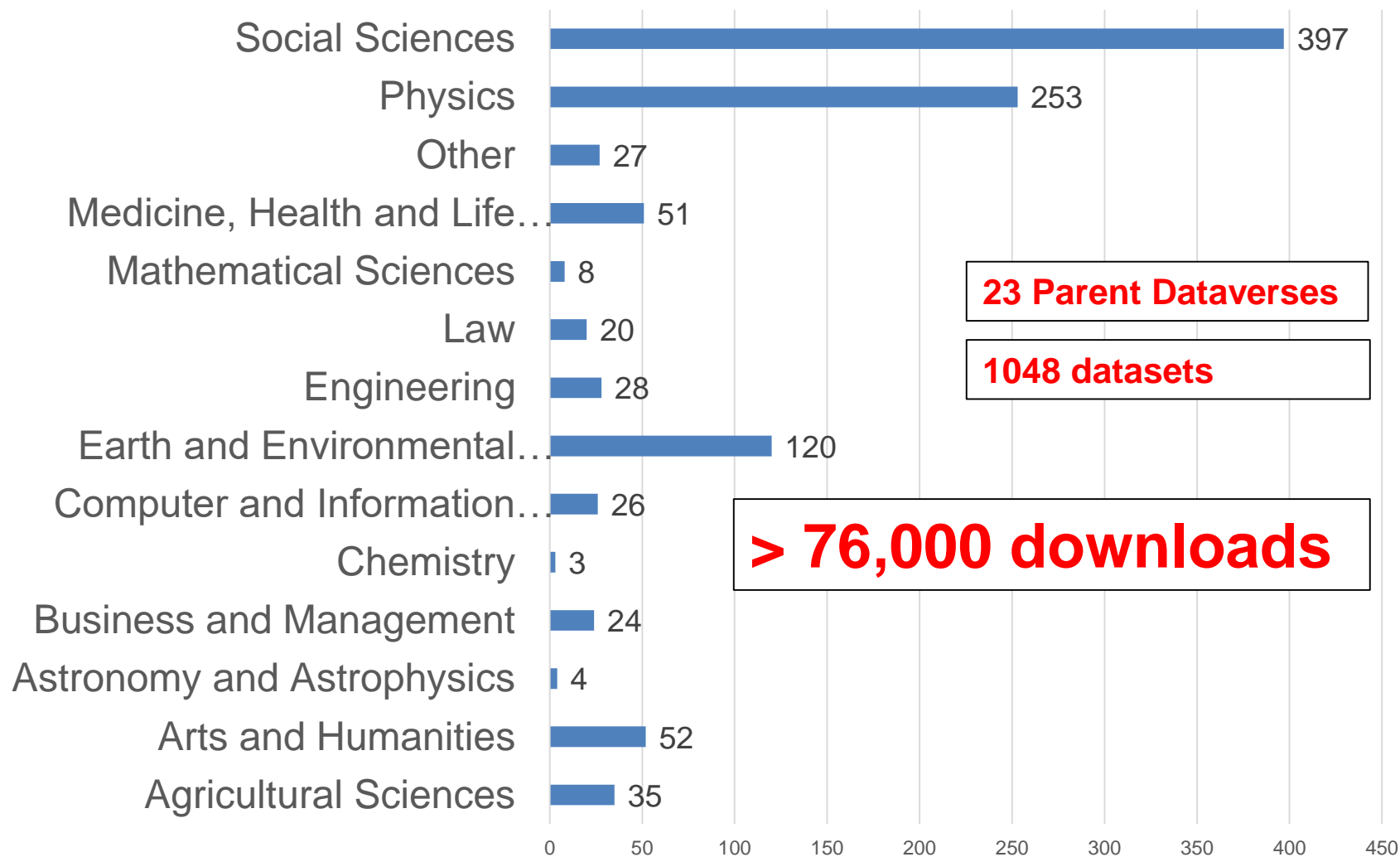
Google Dataset Search Beta

Search for Datasets

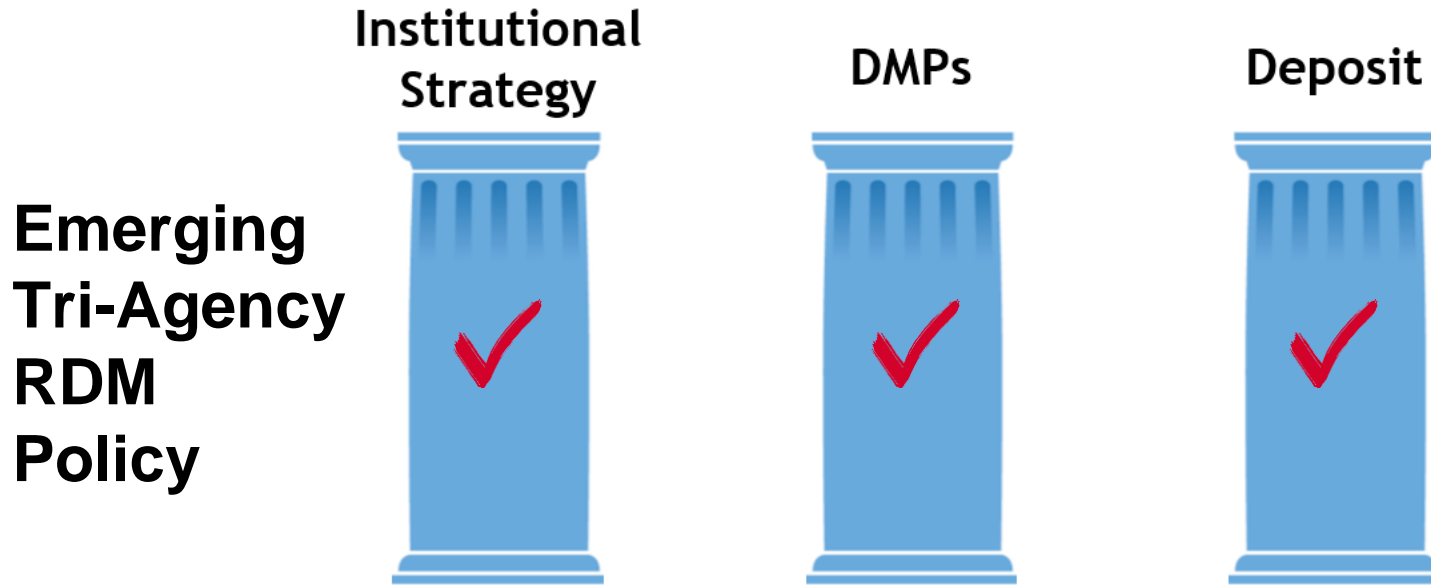


Growing demand for data

Datasets by Discipline: Scholars Portal Dataverse [Aug 23, 2018]



Government policy directions



Drivers also include:

- Trends in **journal publishing**
- Movement toward **open science**
- International initiatives & best practices

Government funding – 2018 Canadian Budget

EQUALITY GROWTH

A Strong Middle Class

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.”

Definition of ‘Digital Research Infrastructure’



“DRI refers to the elements required to perform data-intensive and computationally-intensive research and data management, including high-performance computing, storage, high-speed networks and other tools and resources, including software, standards and data management services.” p. 3

Definition of ‘Digital Research Infrastructure’



CONSULTATION:
DEVELOPING A DIGITAL RESEARCH
INFRASTRUCTURE STRATEGY

“DRI refers to the elements required to perform data-intensive and computationally-intensive research and **data management**, including high-performance computing, storage, high-speed networks and other tools and resources, including software, standards and **data management services.**” p. 3

Outline

- Federal RDM policy developments
- Further context for RDM in Canada
- **Goals and outputs of Portage**
- Developing a shared vision for RDM in Canada

Portage has many partners in RDM...

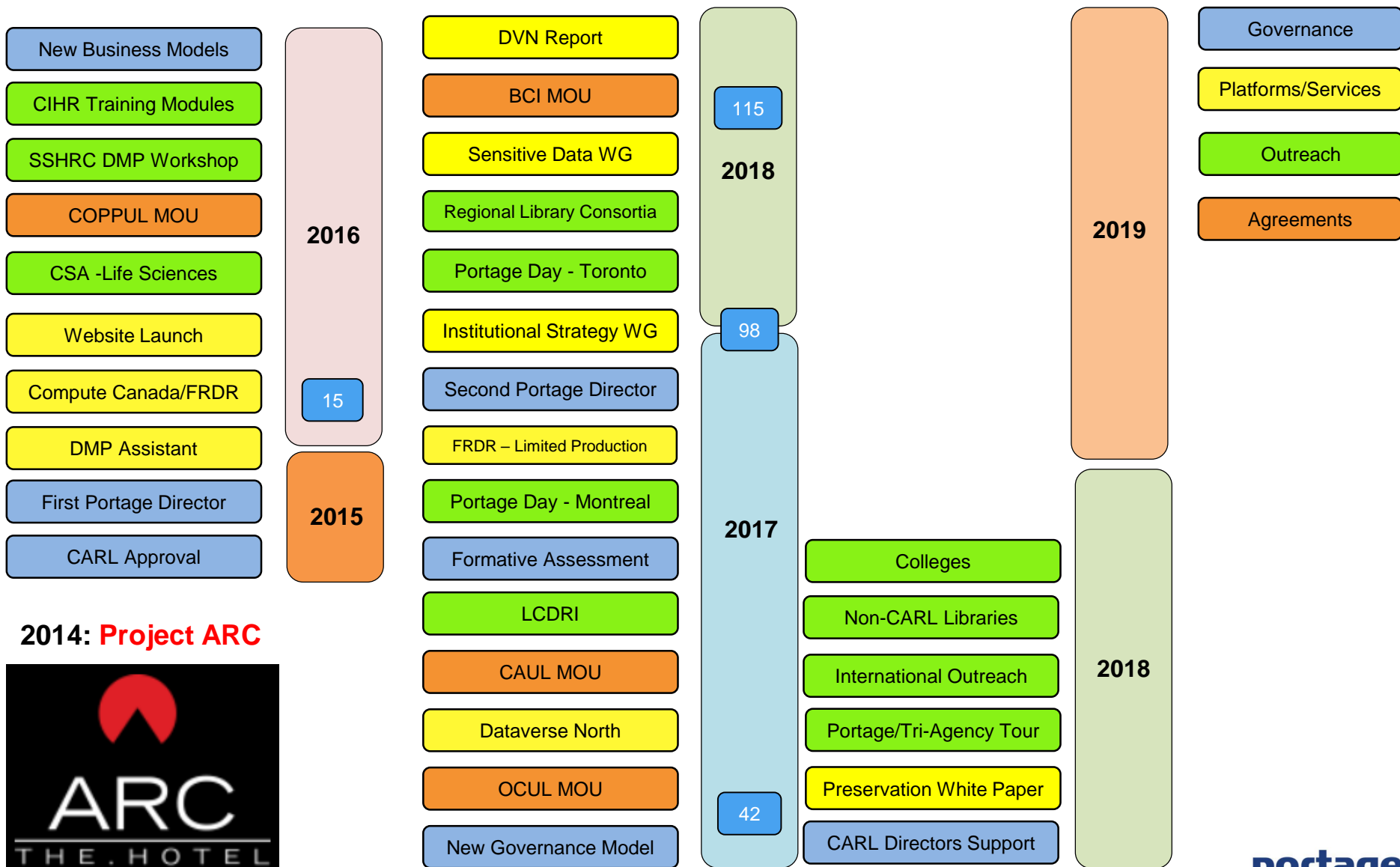


... and many other RDM and DRI stakeholders

Goals of the Portage Network

- Build a **community of practice** for research data management (RDM)
- Engage and advocate for research data management with **stakeholder communities**
- Facilitate and provide leadership in the **development of RDM infrastructure**

Portage origins and progress...



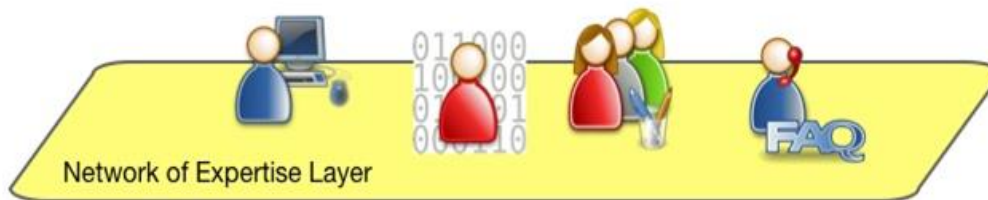
Goals of the Portage Network

- Build a **community of practice** for research data management (RDM)

Building a Community of Practice



Institutions



Network of Experts
RDM Specialists/Librarians

Expert Groups

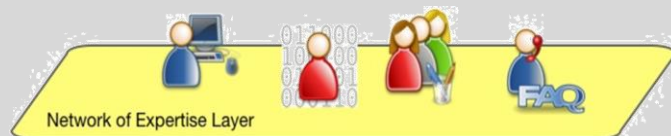
- Data Management Planning
- Curation
- Data Discovery
- Preservation
- Training
- Research Intelligence

Working Groups

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

➤ **110+ Experts from across Canada**

Broad representation across Canada



**22
CARL
Library
Institutions**

Brock University	Acadia University
Carleton University	Cape Breton University
Dalhousie University	Concordia University
McGill University	HEC Montréal
McMaster University	Mount Saint Vincent University
Queen's University	Saint Mary's University
Ryerson University	Trent University
Simon Fraser University	University of Notre Dame
Université de Montréal	University of Prince Edward Island
University of Alberta	Wilfrid Laurier University
University of British Columbia	AceNet
University of Calgary	Canadian Institutes of Health Research
University of Manitoba	CAREB
University of New Brunswick	CASRAI-CA
University of Ottawa	Compute Canada
University of Saskatchewan	Council of Prairie and Pacific University Libraries
University of Toronto	Cybera
University of Waterloo	First Nations Information Governance Centre
University of Western Ontario	Métis Nation Council
University of Windsor	National Research Council Canada
York University	Research Data Canada
Library and Archives Canada	Scholar's Portal
	Secretariat on Responsible Conduct of Research
	Statistics Canada
	BCI
	CARL

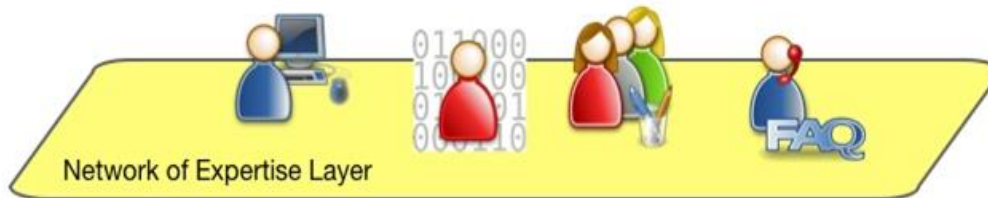
**26
Other
Institutions
&
Organizations**

Building a Community of Practice



Institutions

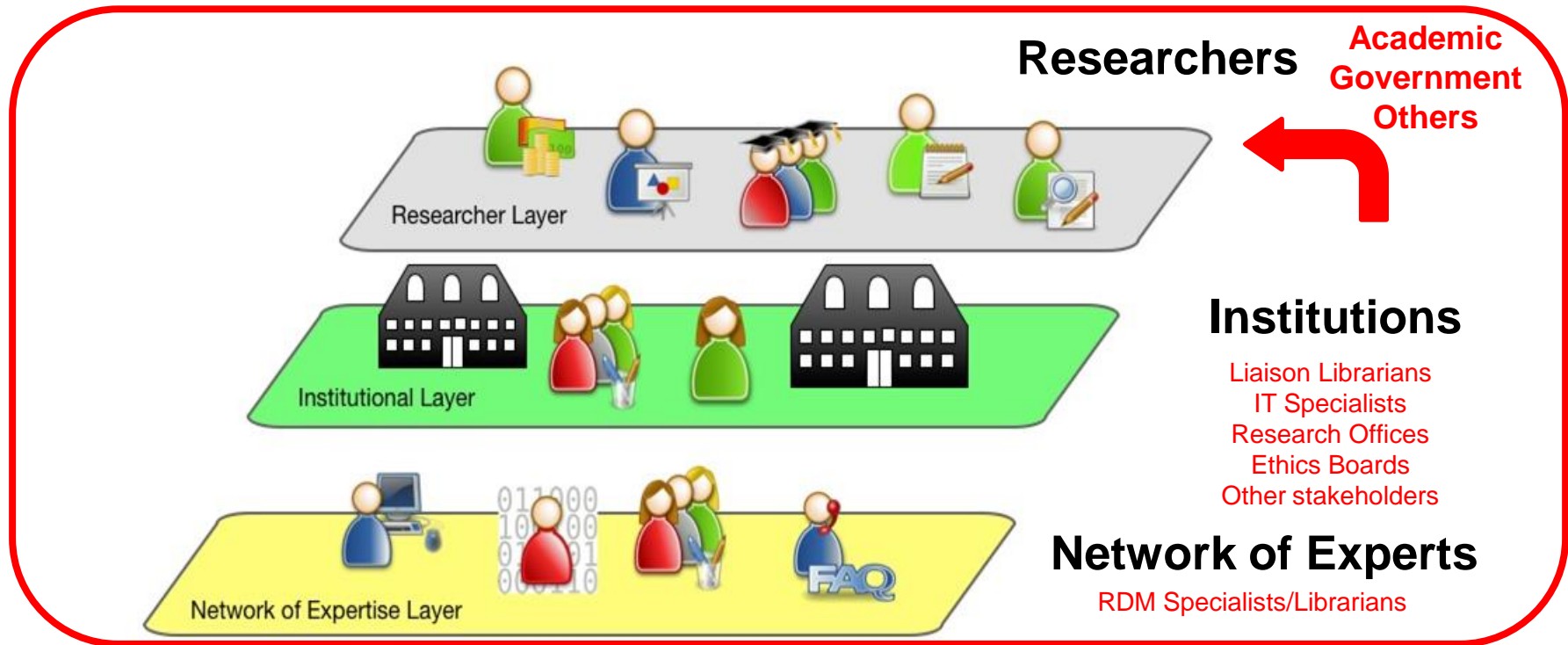
Liaison Librarians
IT Specialists
Research Offices
Ethics Boards
Other stakeholders



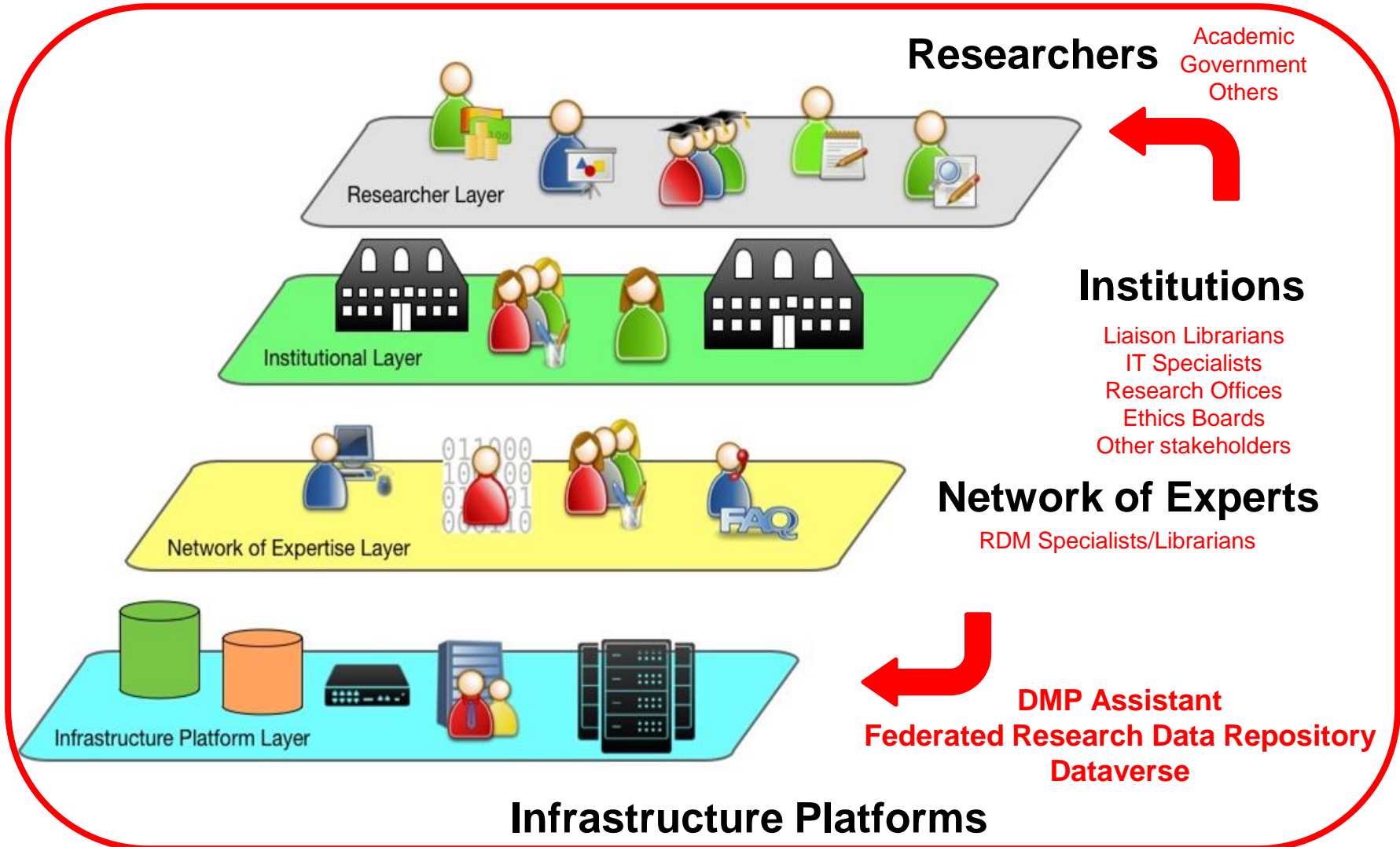
Network of Experts

RDM Specialists/Librarians

Building a Community of Practice



Building a Community of Practice



Network of Experts Publications



Portage Comments on Draft Tri-Agency Research Data Management Policy

Original Tri-Agency [DRAFT POLICY](#) & [FAQ](#)

Background

The Portage Network, an initiative of the Canadian Association of Research Libraries (CARL), is dedicated to the advancement of research data management (RDM) in Canada. Portage, through its nation-wide network of over 115 experts from over 40 organizations, works proactively with community partners to develop national RDM services and infrastructure in support of all researchers in Canada.

Portage had the privilege of working closely with Tri-Agency representatives during the pre-consultation phase of policy development, traveling the country to meet with 14 different stakeholder groups in 4 cities in Jan-Feb 2018. These meetings provided an ideal opportunity both to share Tri-Agency policy requirements with researchers and to describe Portage services and infrastructure designed to help researchers fulfill each requirement. Additionally, these meetings gave Portage, alongside our Tri-Agency colleagues, the chance to hear feedback from stakeholder groups and to respond together to questions that arose.

Portage also took the opportunity to convene three Town Hall meetings in July 2018, engaging over 32 participants. Feedback from these Town Halls and other less formal discussions was considered when crafting the Portage response provided below. A summary of comments from the community is provided in **Appendix A**.

Selected Portage Publications

- **Portage Network** Graduate Student Researcher
- **Portage Program** Research Libraries
- **The SSHRC DMC** Director of the
- **Towards a Collaborative** paper presented by Humphrey, Kate ([PDF](#))
- **Multi-Stakeholder** presented at IABIS Bourne-Tyson,
- **Research Data Management** (Presentation),

Strategic Documents

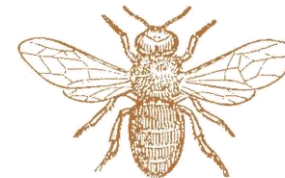
- **Formative Assessment** 2017 ([PDF](#))
- **Portage Business Plan: 2017 and 2018** ([PDF](#))
- **Portage: Organizational Framework**, April 2015 ([PDF](#))

- **Portage Network Response to Industry Canada's Consultation on Developing a Digital Research Infrastructure Strategy**, September 2015 ([PDF](#))
- **CARL Response to the Draft Tri-Agency Statement of Principles on Digital Data Management**, September 2015 ([PDF](#))

ave
tive
ey
ng
a

Management

Busy Bees – brief guides for busy people



"Good Enough" Research Data Management

(a brief guide for busy people)

This brief guide presents a set of good data management practices that busy people can adopt, regardless of their data management skills.

1

Save your raw data in original format

- 1.1 Don't overwrite your original data with a cleaned version.
- 1.2 Protect your original data by locking them or making them read-only.
- 1.3 Refer to this original data if things go wrong (as they often do).

2

Backup your data

- 2.1 Use the 3-2-1 rule: 3 copies of your data, on two different media, and one off-site.
- 2.2 Do not backup or store on a commercial cloud service (Dropbox, Google Drive, etc.).

3

Describe your data

- 3.1 **Machine Friendly:** Describe your dataset with a metadata standard for discovery.
- 3.2 **Human Friendly:** Describe your variables, so your colleagues will understand what you meant. Data without good metadata is useless. Give your variables clear names.

4

Process your data

- 4.1 Make each column a variable.
- 4.2 Make each row an observation.
- 4.3 Store units (e.g. kg or cm) as metadata (in their own column).
- 4.4 Document each step processing your data in a README file.

5

Archive and preserve your data

- 5.1 Submit final data files assigning a persistent identifier (DOI).
- 5.2 Provide good metadata so others could find your data (e.g. Darwin Core, DDI, etc.).

portage

This guide was reproduced by the Portage Training Expert Group (Eugene Barsky, University of British Columbia). It can be modified and re-used under the CC-BY license.

"Good Enough" Dataverse

(a brief guide for busy people)

This brief guide presents key information for researchers to understand its features, and how to access and use it.

1

What is Dataverse?

- 1.1 A free, open source web application developed at Harvard University.
- 1.2 Stores, shares, cites, explores, and analyzes research data.
- 1.3 Facilitates making your data available to others, as appropriate.
- 1.4 A robust repository environment for depositing your research data.

2

Why should I use Dataverse (DV)?

- 2.1 Helps you meet both funding and requirements.
- 2.2 Stores many types of data, including:
 - Tabular data (CSV, SPSS, R, etc.)
 - Documentation (pdf, doc, text)
 - Geospatial data
 - Multimedia file (audio-visual)

3

What are the benefits?

- 3.1 Creates a Digital Object Identifier (DOI) so your data is a citable research output for publications and CV.
- 3.2 Increases the impact and visibility of your research.
- 3.3 Stability and longevity of your data as created at time of deposit.
- 3.4 Statistics and metrics to track and understand how your data is used worldwide.

How to use

- 4.1 Choose your data to upload.
- 4.2 Custom your data.
- 4.3 Modify your data.
- 4.4 Your metadata.

5

Process your data

- 5.1 Find out if your institution has a local Dataverse - you can check the Portage site to see if your institution is using Dataverse.
- 5.2 If needed, check with your local Data contact for help.
- 5.3 Create your Dataverse account.

6

What's the next step?

- 6.1 Ensure that your data follows best practices for formatting and metadata.
- 6.2 Upload your data and add your metadata.
- 6.3 Set any access restrictions or terms for your data.
- 6.4 Create a "guestbook" to track the use of your data.
- 6.5 Publish your data!

portage

This guide was produced by the Portage Dataverse North Training Working Group. It can be modified and re-used under the CC-BY license.

"Good Enough" Data Curation

(A brief guide for busy people)

CREATING

Good curation begins with developing strategies for collecting data, as well as creating documentation and metadata to explain and contextualize them. Data curators can help create and review strategies.

MAINTAINING

Ongoing data maintenance supports quality assurance. Practices include implementing a secure storage plan and protective measures such as file backups and file integrity checks.

REUSING

To encourage reuse, data curators may augment metadata, transform file formats to improve interoperability, and link to related datasets to increase discoverability.

Data curation is the active management of research data throughout a project to produce datasets that are FAIR: Findable, Accessible, Interoperable, and Reusable.

USING

Applying robust quality control and assurance practices during data analysis -- such as file versioning and metadata review -- facilitates and expedites data archiving and reuse.

SHARING

Data curators facilitate the deposit of data, code, and metadata into reputable repositories that enable discovery and access. This includes applying appropriate anonymization or access controls.

ARCHIVING

Data curators select data for long-term preservation and then facilitate bundling data, code, and metadata into a single package for transfer to an archive.

portage
SHARED STEWARDSHIP OF RESEARCH DATA

This guide was created by the Portage Curation Expert Group and can be modified and re-used freely under the CC-BY license.



Busy Bees – aussi disponible en français



De bonnes pratiques en gestion des données de recherche

(Un guide sommaire pour gens occupés)

Ce petit guide présente un ensemble de bonnes pratiques qui s'appliquent indépendamment de leurs compétences ou de leur niveau d'expertise.

1

Enregistrer vos données brutes dans leur format original

- 1.1 N'écrasez pas vos données avec une version corrigée.
- 1.2 Protégez vos données originales en verrouillant vos fichiers (avec un mot de passe) ou en définissant un accès en lecture seule.
- 1.3 Référez-vous à ces données originales lorsque les choses tournent mal (comme c'est souvent le cas).

Sauvegarder vos données

- 2.1 Utilisez la règle des trois copies de données : conservez-en une sur votre ordinateur, une sur un serveur sécurisé et une sur un support externe.
- 2.2 Ne sauvegardez pas vos données sur un support externe non sécurisé.

3

Décrire vos données

3.1 **Pour qu'elles soient facilement lisibles par les ordinateurs :** Décrivez votre ensemble de données selon une norme de métadonnées facilitant la découverte.

3.2 **D'une manière conviviale pour les humains :** Décrivez vos variables de façon à ce qu'elles soient facilement compréhensibles pour vos collègues. Des données sans bonnes métadonnées sont inutilisables. Donnez des noms clairs à vos variables.

4

Traiter vos données

- 4.1 Faites de chaque colonne une variable.
- 4.2 Faites de chaque rangée une observation.
- 4.3 Stockez les unités (tels que kg ou cm) en tant que métadonnées (avec leur propre colonne).
- 4.4 Documentez chaque étape du traitement de vos données dans un fichier « LISEZ-MOI ».

Archiver et partager vos données

- 5.1 Soumettez vos données dans un dépôt permanent (DOI).
- 5.2 Fournissez de votre étude de façon à ce qu'elles soient facilement lisibles par les autres (utilisez des métadonnées telles que Darwin Core).

portage

Ce guide a été produit par le Groupe d'experts sur la formation des données de recherche de Portage, en collaboration avec le Groupe de travail sur la formation des données de recherche de l'Université de la Colombie-Britannique. Il peut être réutilisé en vertu de la licence d'attribution de Creative Commons (CC-BY).

Dataverse : l'essentiel

(Un guide sommaire pour gens occupés)

Ce bref guide présente des informations essentielles sur Dataverse | chercheurs de comprendre comment y accéder et l'utiliser.

1

Qu'est-ce que Dataverse?

- 1.1 Une application Web gratuite et libre développée à l'Université Harvard.
- 1.2 Un environnement de dépôt robuste pour le dépôt de vos données de recherche.
- 1.3 Permet de conserver, d'échanger, de citer, d'explorer et d'analyser les données de recherche.
- 1.4 Facilite la mise à disposition de vos données, au besoin.

2

Pourquoi devrais-je utiliser Dataverse?

- 2.1 Aide à répondre aux exigences des organismes subventionnaires et des éditeurs de périodiques.
- 2.2 Stocke de nombreux types de données, notamment :
 - Données tabulaires (CSV, SPSS, R, etc.)
 - Documents (pdf, doc, texte, code)
 - Données géospatiales
 - Fichiers multimédias (audiovisuels)

3

Quels sont les avantages?

- 3.1 Crée un identifiant numérique d'objet (DOI) qui permet de citer vos données à partir d'une publication ou d'un CV.
- 3.2 Augmente l'impact et la visibilité de votre recherche.
- 3.3 Assure la stabilité et la longévité de vos données telles qu'elles étaient au moment du dépôt.
- 3.4 Offre des statistiques et d'autres mesures pour suivre et comprendre l'utilisation de vos données partout dans le monde.

Mettre à jour vos données

- 4.1 Définissez vos paramètres de mise à jour de vos données et ouvrez à la modification.
- 4.2 Personnalisez les paramètres de mise à jour de vos données.
- 4.3 Modifiez les liens vers vos données en tout temps.
- 4.4 Vos métadonnées sont accessibles, donc réutilisables.

5

Comment commencer?

- 5.1 Vérifiez si votre établissement utilise déjà Dataverse. Pour ce faire, vous pouvez consulter le site de [Portage](#) pour voir si votre établissement figure dans la liste.
- 5.2 Au besoin, demandez de l'aide à la personne responsable des données de votre établissement.
- 5.3 Créez votre compte Dataverse.

6

Quelle est la prochaine étape?

- 6.1 Assurez-vous que vos données respectent les meilleures pratiques en matière de formatage de données et d'utilisation des métadonnées.
- 6.2 Téléchargez vos données et ajoutez vos métadonnées.
- 6.3 Établissez les restrictions d'accès ou les conditions d'utilisation de vos données, s'il y a lieu.
- 6.4 Créez un « registre des visiteurs » pour suivre l'utilisation de vos données.
- 6.5 Publiez vos données!

portage

comptes canadiens

Ce guide a été produit par le Groupe de travail sur la formation des données de recherche de Portage, en collaboration avec le Groupe de travail sur la formation des données de recherche de l'Université de la Colombie-Britannique. Il peut être réutilisé en vertu de la licence d'attribution de Creative Commons (CC-BY).

Préparation à la conservation des données

(Un guide sommaire pour gens occupés)

CRÉER

Un bon travail de conservation commence par l'élaboration de stratégies de collecte de données, ainsi que la création de la documentation et des métadonnées qui les décrivent et les mettent en contexte. Les conservateurs de données peuvent aider à créer et à revoir des stratégies.

METTRE À JOUR

La mise à jour continue des données soutient l'assurance de la qualité. Les pratiques comprennent la mise en œuvre d'un plan de stockage sécurisé et de mesures de protection comme la création de fichiers de sauvegarde et la vérification d'intégrité des fichiers.

RÉUTILISER

Pour promouvoir la réutilisation des données les conservateurs peuvent ajouter des métadonnées, transformer les formats de fichiers pour améliorer l'interopérabilité, ainsi qu'ajouter des liens vers des ensembles de données connexes pour accroître la réutilisabilité.

La conservation des données est la gestion active des données de recherche tout au long d'un projet visant à produire des ensembles de données qui suivent les principes FAIR : facilement repérables, accessibles, interopérables et réutilisables.

UTILISER

L'application de solides pratiques de contrôle et d'assurance de la qualité durant l'analyse des données, comme par exemple le contrôle des versions et des métadonnées, facilite et accélère l'archivage et la réutilisation des données.

PARTAGER

Les conservateurs de données facilitent le dépôt de données, du code et des métadonnées dans des dépôts fiables qui permettent l'exploration et l'accès. Cela inclut la mise en place de mesures appropriées d'anonymisation ou de contrôle d'accès.

ARCHIVER

Les conservateurs de données sélectionnent les données pour la conservation à long terme et regroupent ensuite les données, le code et les métadonnées en un seul paquet pour le transfert à une archive.

portage

SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE

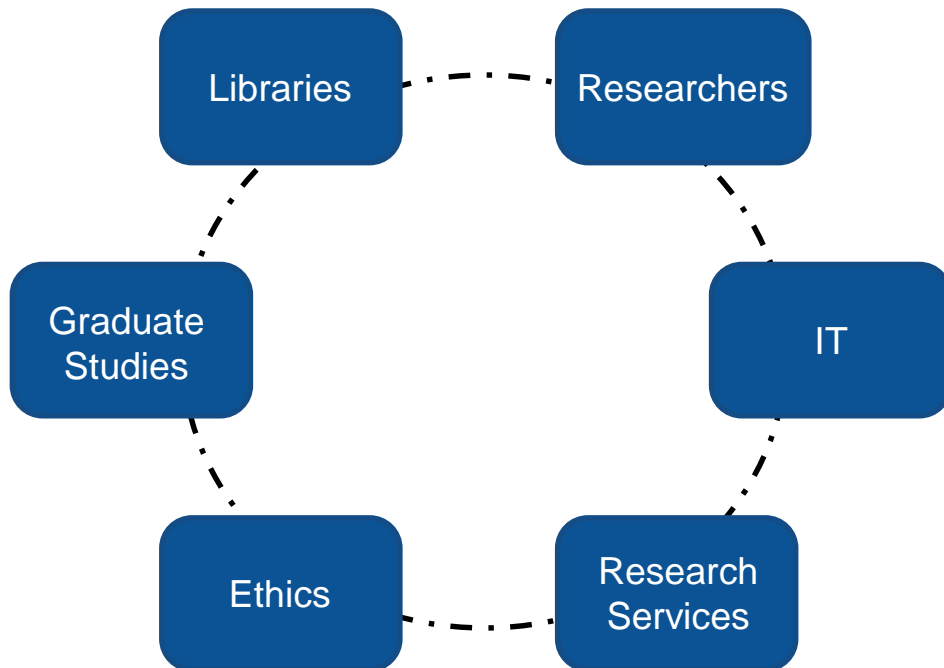
Le présent guide a été produit par le Groupe d'experts sur l'organisation des données de Portage et peut être modifié et réutilisé en toute liberté en vertu de la licence d'attribution de Creative Commons (CC-BY).



Goals of the Portage Network

- Build a **community of practice** for research data management (RDM)
- Engage and advocate for research data management with **stakeholder communities**

Levels of Stakeholder Engagement

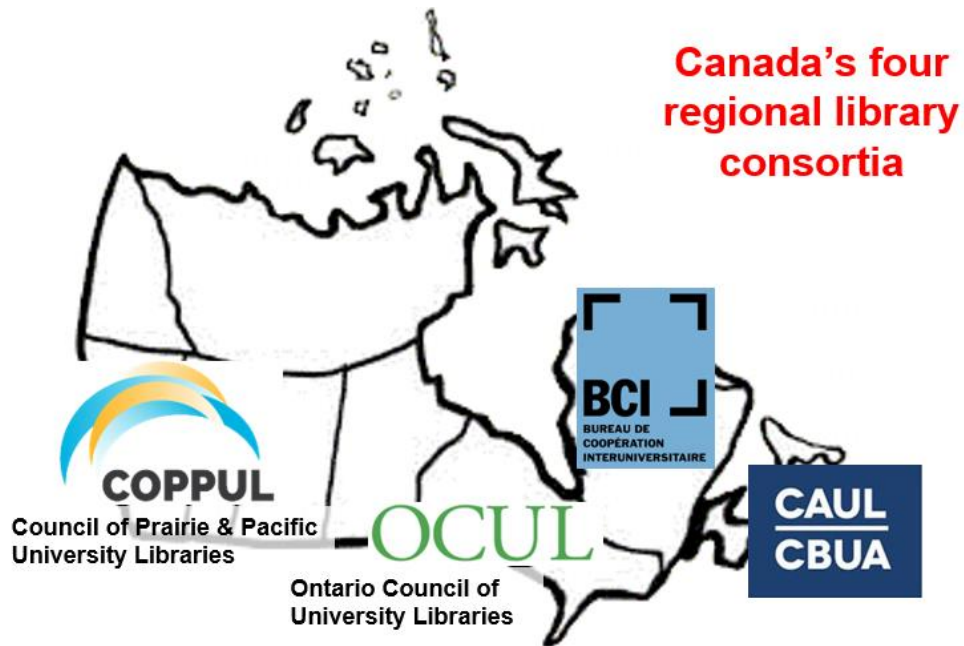
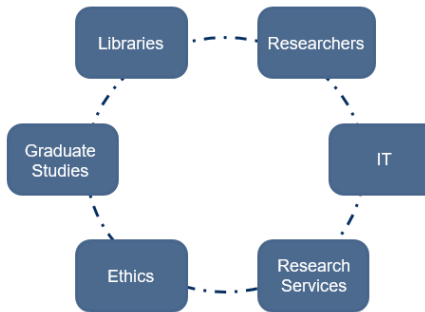


Levels of Stakeholder Engagement

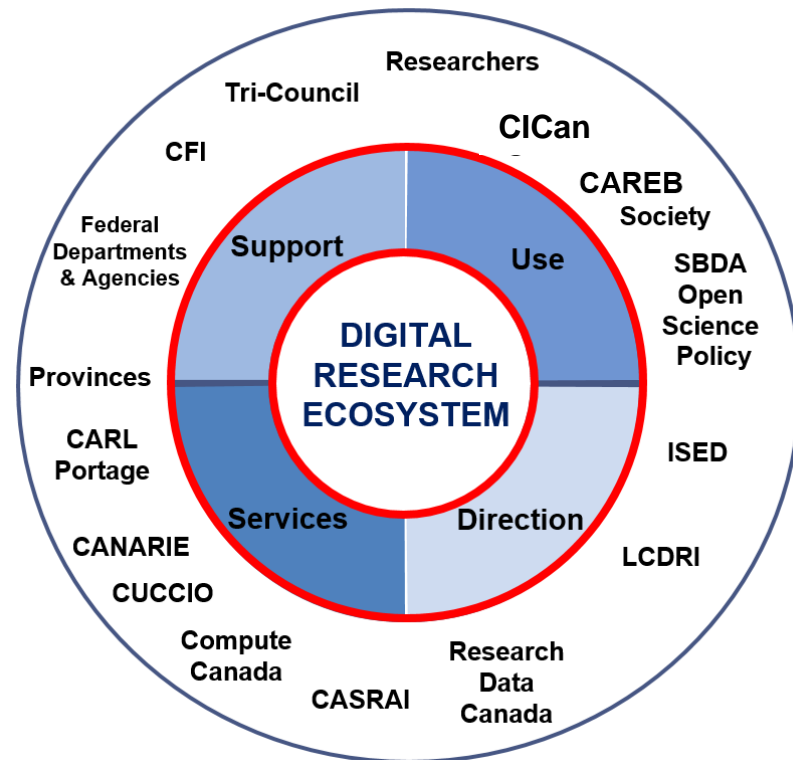
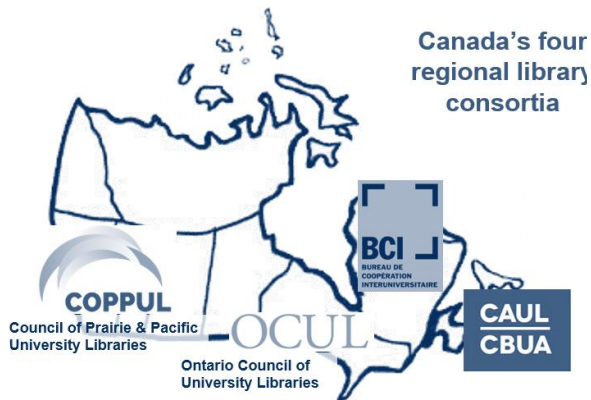
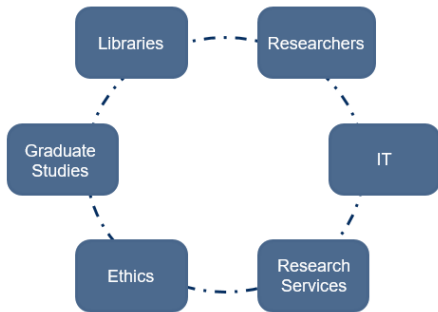
LOCAL

REGIONAL/
PROVINCIAL

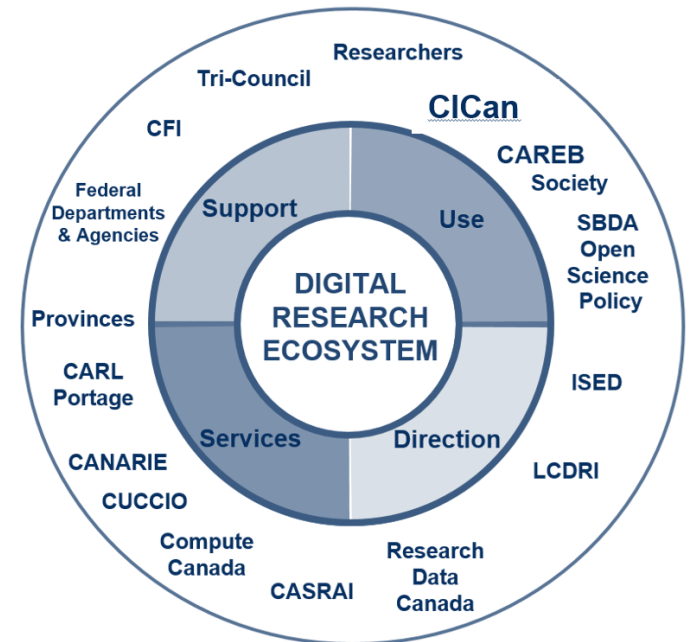
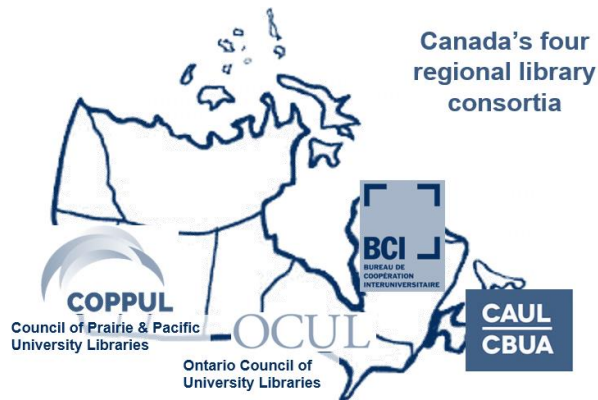
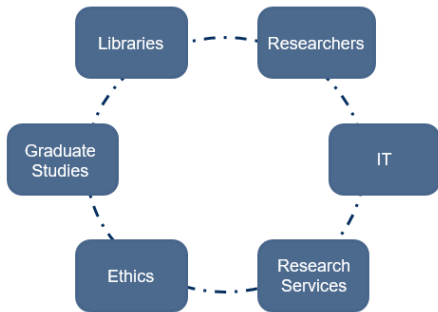
NATIONAL



Levels of Stakeholder Engagement



Levels of Stakeholder Engagement



Goals of the Portage Network

- Build a **community of practice** for research data management (RDM)
- Engage and advocate for research data management with **stakeholder communities**
- Facilitate and provide leadership in the **development of RDM infrastructure**

Portage and RDM Infrastructure

Institutional
Strategy



DMPs



Deposit



Initial focus on the emerging
Tri-Agency RDM Policy

Portage Responses to Tri-Agency Policy

Institutional Strategy



DMPs



Deposit

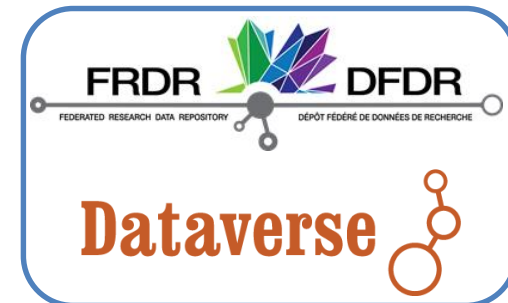


✓ Institutional RDM Strategies

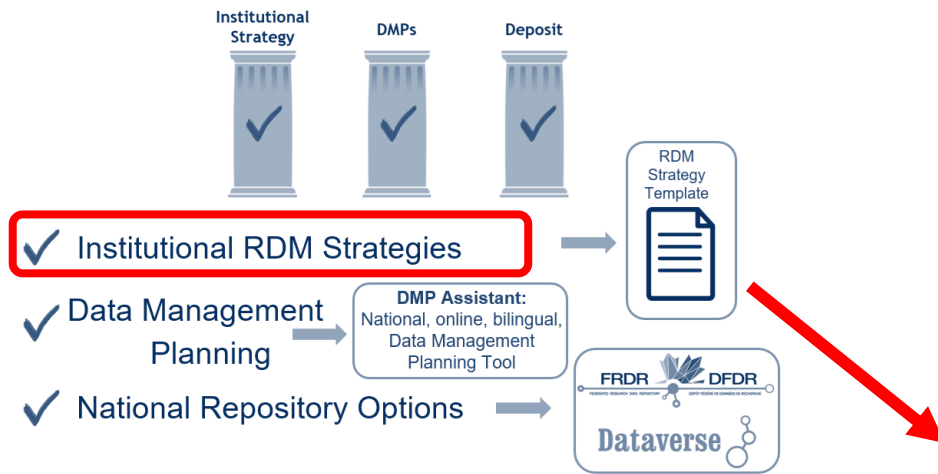
✓ Data Management Planning

✓ National Repository Options

DMP Assistant:
National, online, bilingual,
Data Management
Planning Tool



Pillar 1: Institutional RDM Strategies



Expert Groups

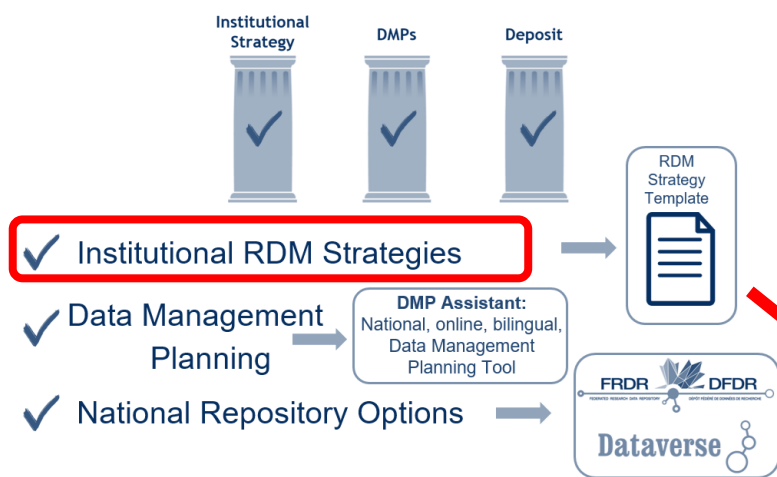
- Data Management Planning
- Curation
- Data Discovery
- Preservation
- Training
- Research Intelligence

Working Groups

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



Pillar 1: Institutional RDM Strategies



Expert Groups

Data Management Planning
Curation
Data Discovery
Preservation
Training
Research Intelligence

Strategy Components

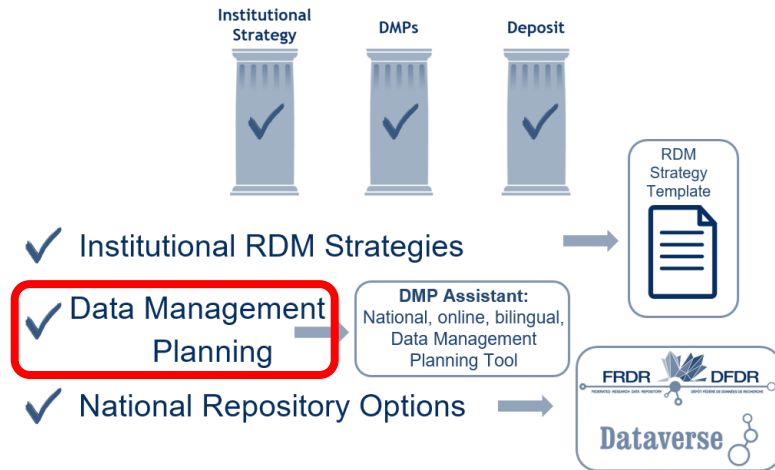
+ associated guidance

Working Groups

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

1. **Raise awareness:** about benefits of best practices and emerging policy requirements.
2. **Assess institutional readiness:** to manage and realize the potential of data holdings.
3. **Formalize RDM practices:** through the development and/or adoption of guidelines, best practices, and policies – setting the tone and underscoring institutional commitment and expectations.
4. **Define a roadmap:** to help with capacity-building over the medium term, to ensure research excellence, efficiency, and transparency.

Pillar 2: Data Management Planning



Data Management Planning

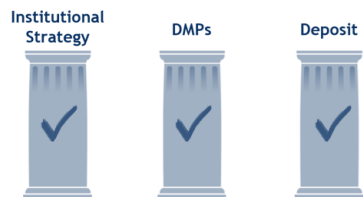
Expert Groups

- Curation
- Data Discovery
- Preservation
- Training**
- Research Intelligence**

Working Groups

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

Pillar 2: Data Management Planning



✓ Institutional RDM Strategies

✓ Data Management Planning

✓ National Repository Options

DMP Assistant:
National, online, bilingual,
Data Management
Planning Tool



Data Management Planning

Expert Groups

Curation

Data Discovery

Preservation

Training

Research Intelligence

Working Groups

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

DMP Assistant

Sign In

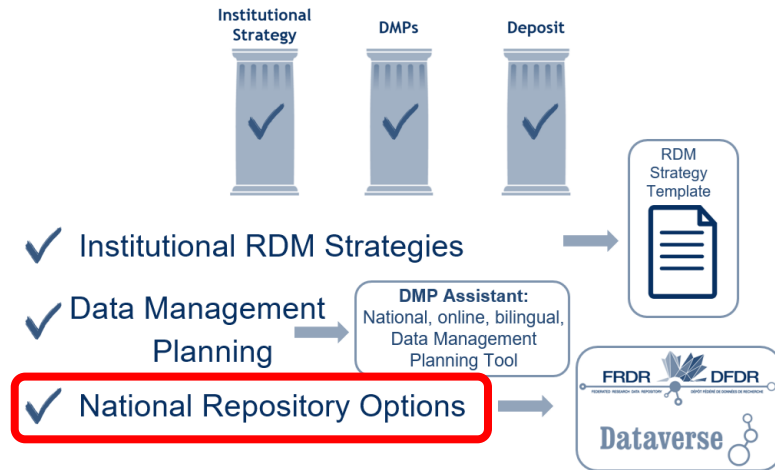
Sign Up



<https://assistant.portagenetwork.ca/en>

1. Data collection
 2. Documentation & metadata
 3. Backup & storage
 4. Preservation
 5. Data sharing & reuse
 6. Responsibilities & resources
 7. Ethics & legal compliance
- DMP Assistant 2.0
Coming Soon

Pillar 3: Data Deposit into a Repository



Data Management Planning

Curation

Data Discovery

Preservation

Training

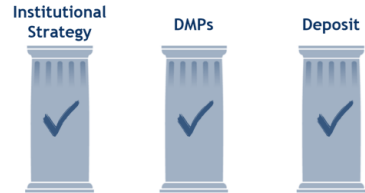
Research Intelligence

Expert Groups

Working Groups

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

Pillar 3: Data Deposit into a Repository



✓ Institutional RDM Strategies

✓ Data Management Planning

✓ National Repository Options

DMP Assistant:
National, online, bilingual,
Data Management
Planning Tool



Data Management Planning

Curation

Data Discovery

Preservation

Training

Research Intelligence

Expert Groups

Working Groups

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



Canadian Repository Landscape



**General
Repositories**

**Government
Repositories**

**Domain-specific
Repositories**



re3data.org



Benefits of Data Repositories



From draft Tri-Agency RDM Policy:

“The repository will ensure safe storage, preservation, and curation of the data....Whenever possible, these data, metadata and code should be linked to the publication with a persistent digital identifier.”

Also:

- Promotes best practices for data stewardship
- Researchers not responsible for the ongoing stewardship of their research data in perpetuity
- Increases discoverability and reusability of research data; collect once, reuse many times
- Provides open or controlled access to research data without subscription or licensing fees associated with commercial publishers

Canadian Repository Landscape



**General
Repositories**

**Government
Repositories**

**Domain-specific
Repositories**



Two National Repository Options



Federated Research Data Repository

National Discovery Service

Portage Data Discovery Service Working Group &
Compute Canada FRDR Team



National Instance of Dataverse

Portage Dataverse North Working
Group and OCUL Scholars Portal
Dataverse Team



Pillar 3: Data Deposit into a Repository



Federated Research Data Repository

Main Features:

1. National Repository Option

'Big Data' upload/download

Maintains file hierarchies

Scalable

2. National Discovery Platform

3. Preservation Processing capacity

Currently in Limited Production



Stone Tools, Diet & Sociality at the Dawn of Humanity
U. Calgary



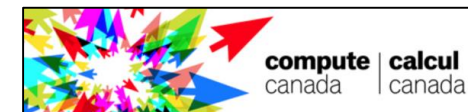
The Water Institute
U. Of Waterloo



"Big data solutions for the marine acoustic community"
Dalhousie University



Mountain Legacy Project:
Capturing Change in Canada's Mountains
U. Vic



Pillar 3: Data Deposit into a Repository

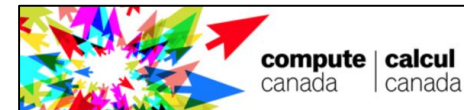


Limited Production by the numbers:

- Working with 18 Limited Production partners
- 52 datasets published (4.2 Tb)
- 32 datasets in progress (2.7 Tb)

Google Analytics for <https://www.frdr.ca/> (May-Oct 2018)

- Users: 2,937/Avg. per month: 490
- Sessions: 5,001/Avg. per month: 834
- Pageviews: 18,057/Avg. per month: 3,010



Pillar 3: Data Deposit into a Repository



Federated Research Data Repository

Main Features:

1. National Repository Option
 - 'Big Data' upload/download
 - Maintains file hierarchies
 - Scalable
2. National Discovery Platform
3. Preservation Processing capacity

Currently in Limited Production



Main Features:

























Mature platform
Widely adopted
User-friendly
Brandable

National Repository Option



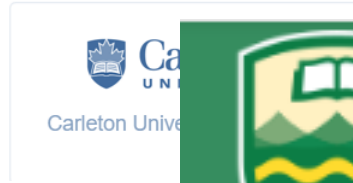
Pillar 3: Data Deposit into a Repository

Dataverses at OCUL Scholars Portal:

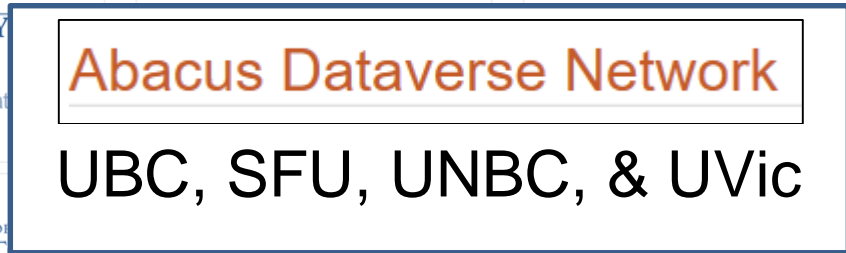
 <p>Algoma University Algoma University Dataverse</p>	 <p>Brock University Brock University Dataverse</p>	 <p>Carleton University Carleton University Dataverse</p>	 <p>Lakehead University Lakehead University Dataverse</p>
 <p>Laurentian University Université Laurentienne Laurentian University Dataverse</p>	 <p>McMaster University McMaster University Dataverse</p>	 <p>NIPISSING UNIVERSITY Nipissing University Dataverse</p>	 <p>OCAD UNIVERSITY OCAD University Dataverse</p>
 <p>Queen's University Queen's University Dataverse</p>	 <p>Royal Military College Royal Military College Dataverse</p>	 <p>Ryerson University Ryerson University Dataverse</p>	 <p>TRENT UNIVERSITY Trent University Dataverse</p>
 <p>UNIVERSITY OF CALGARY PRISM Dataverse</p>	 <p>UNIVERSITY OF GUELPH University of Guelph Dataverse</p>	 <p>UNIVERSITY OF ONTARIO INSTITUTE OF TECHNOLOGY UOIT Dataverse</p>	 <p>uOttawa University of Ottawa Dataverse</p>
 <p>University of Regina University of Regina Dataverse</p>	 <p>UNIVERSITY OF TORONTO University of Toronto Dataverse</p>	 <p>UNIVERSITY OF WATERLOO University of Waterloo Dataverse</p>	 <p>University of Windsor University of Windsor Dataverse</p>
 <p>University of Windsor University of Windsor Dataverse</p>	 <p>Western University Western University Dataverse</p>	 <p>LAURIER Inspiring Lives. Wilfrid Laurier University Dataverse</p>	 <p>YORK UNIVERSITY York University Dataverse</p>

Pillar 3: Data Deposit into a Repository

... Dalhousie, Abacus, & UNB



University of Alberta Libraries



Pillar 3: Data Deposit into a Repository

Who to contact?

Interested in joining Dataverse
North?

Contact:

dataverse@scholarsportal.info

for more information!



University of Windsor Dataverse

Western University Dataverse

Wilfrid Laurier University
Dataverse

York University Dataverse

Portage RDM Support for Institutions & Researchers

Tri-Agencies
&
Portage

Institutional
Strategy



DMPs



Deposit



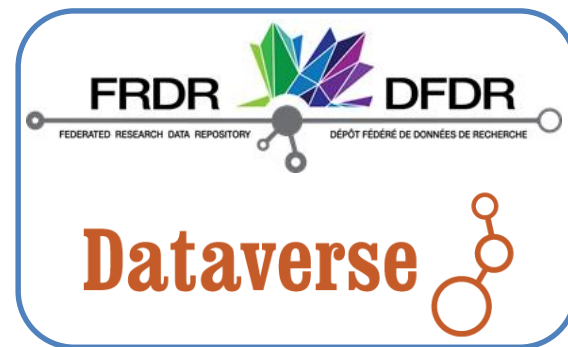
RDM
Strategy
Template



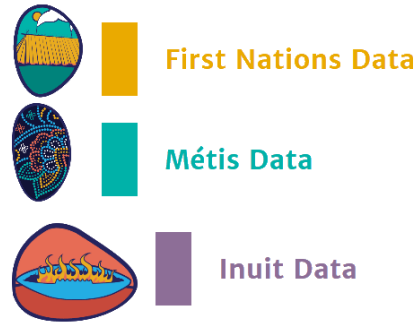
DMP Assistant
Sign In Sign Up

National, online,
bilingual, Data
Management
Planning Tool

Portage



Circling back to another Portage initiative...

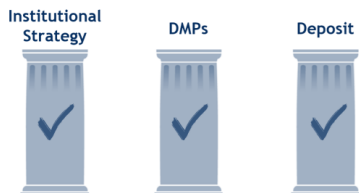


Sensitive Data



General
Data
Protection
Regulation

Portage 'Sensitive Data' Working Group



RDM
Strategy
Template



- ✓ Institutional RDM Strategies WG
- ✓ DMP Assistant
- ✓ National Repository Options

National, online,
bilingual, Data
Management Planning
Tool



Data Management Planning

Curation

Data Discovery

Preservation

Training

Research Intelligence

Expert Groups

Working Groups

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



Four priority outcomes identified...



Deposit-friendly text for ethics applications & informed consent



Data access agreements



Training materials



Environmental scan of best practices for Indigenous data

Goals of the Portage Network

- Build a **community of practice** for research data management (RDM)
- Engage and advocate for research data management with **stakeholder communities**
- Facilitate and provide leadership in the **development of RDM infrastructure**

Vision for a Federated RDM Architecture

Outline

- Federal RDM policy developments
- Further context for RDM in Canada
- Goals and outputs of Portage
- **Developing a shared vision for RDM in Canada**

Federated RDM Architecture

Active Storage

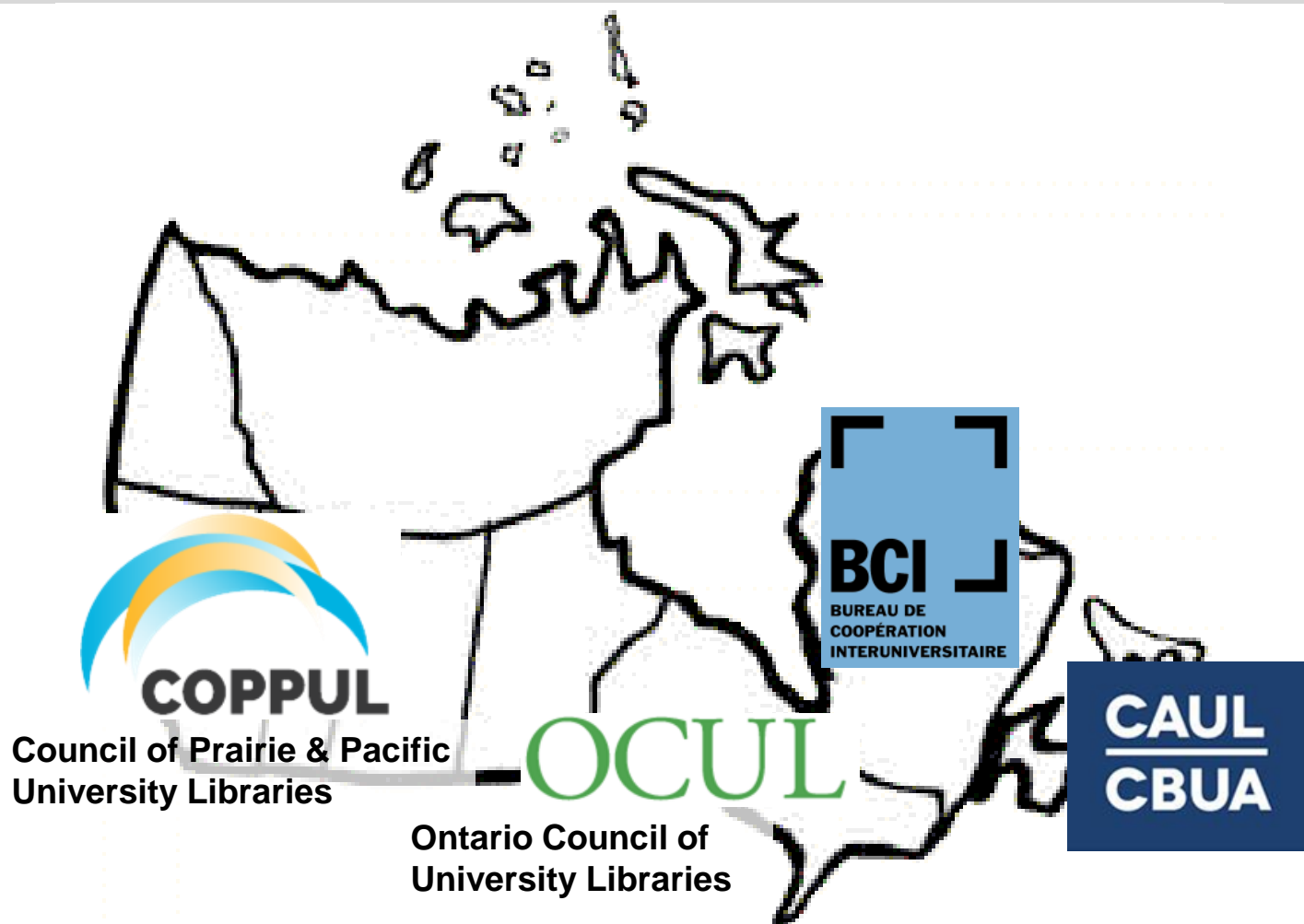
Repository Storage

Archival (Preservation) Storage

Data Management Planning

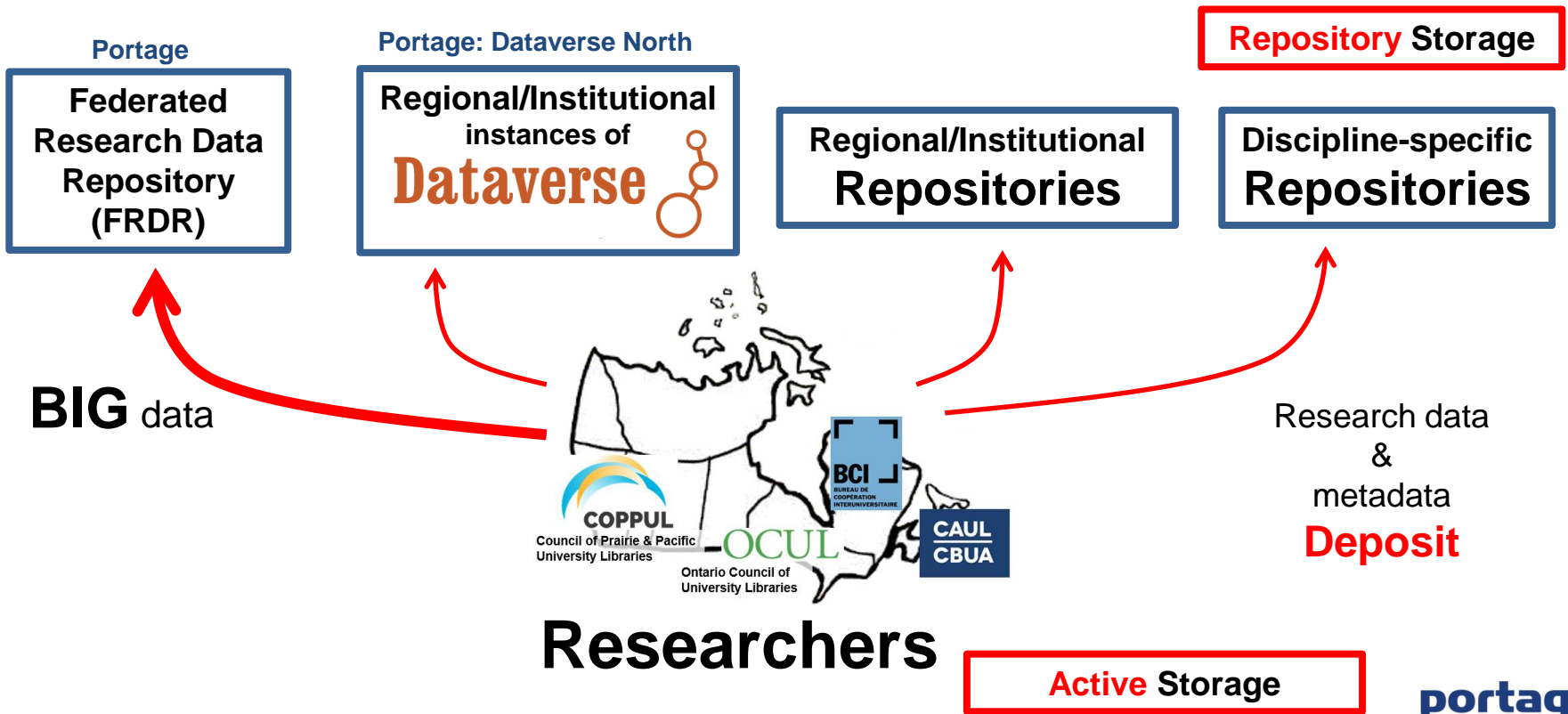
Data Curation

Federated RDM Architecture

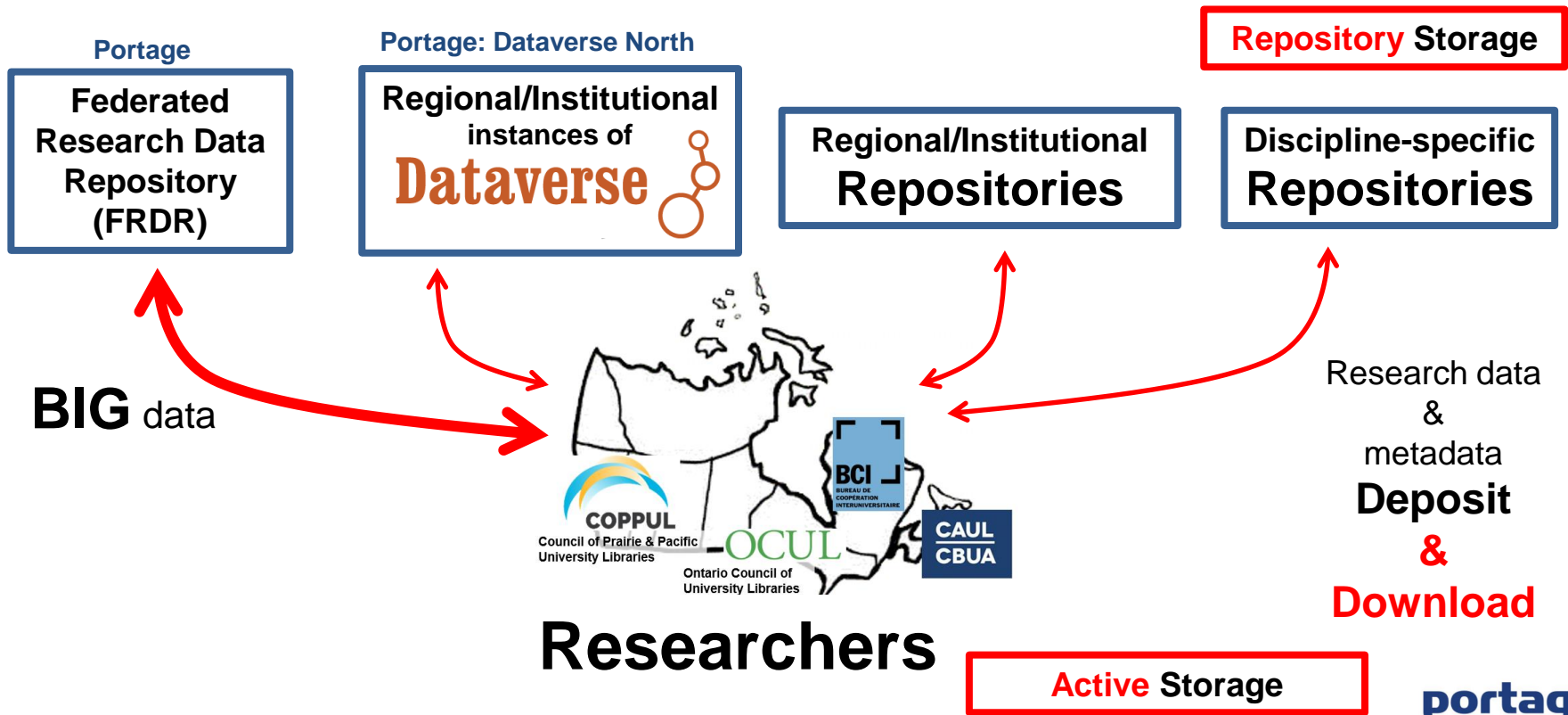


Researchers

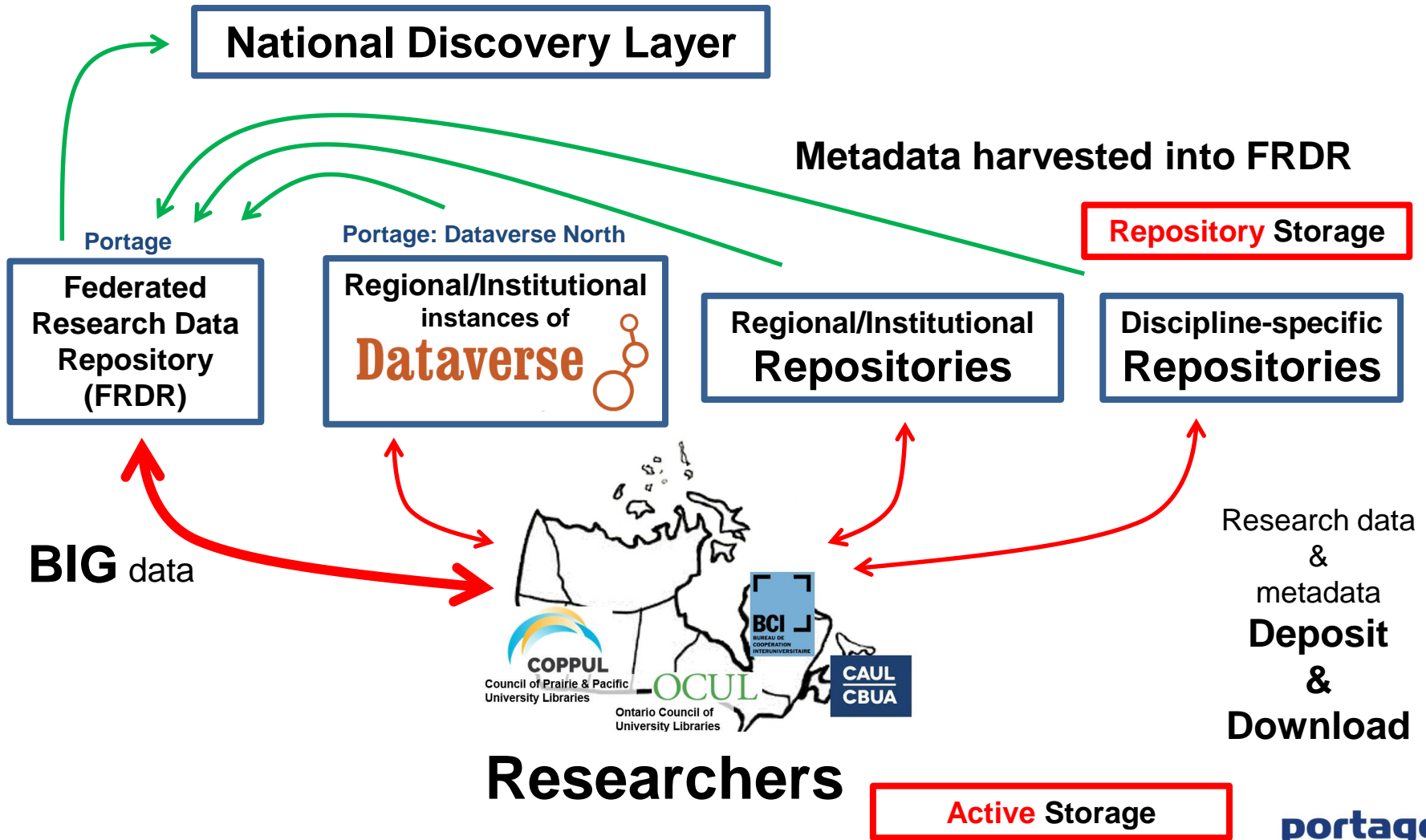
Federated RDM Architecture



Federated RDM Architecture



Federated RDM Architecture



Discovery: Collaborating repositories



**Metadata harvested
from 40 repositories
indexing
over 90,000 datasets**

**All searchable via
<https://www.frdr.ca>**



LAU



Deposit Data

Advanced search

National Discovery Layer

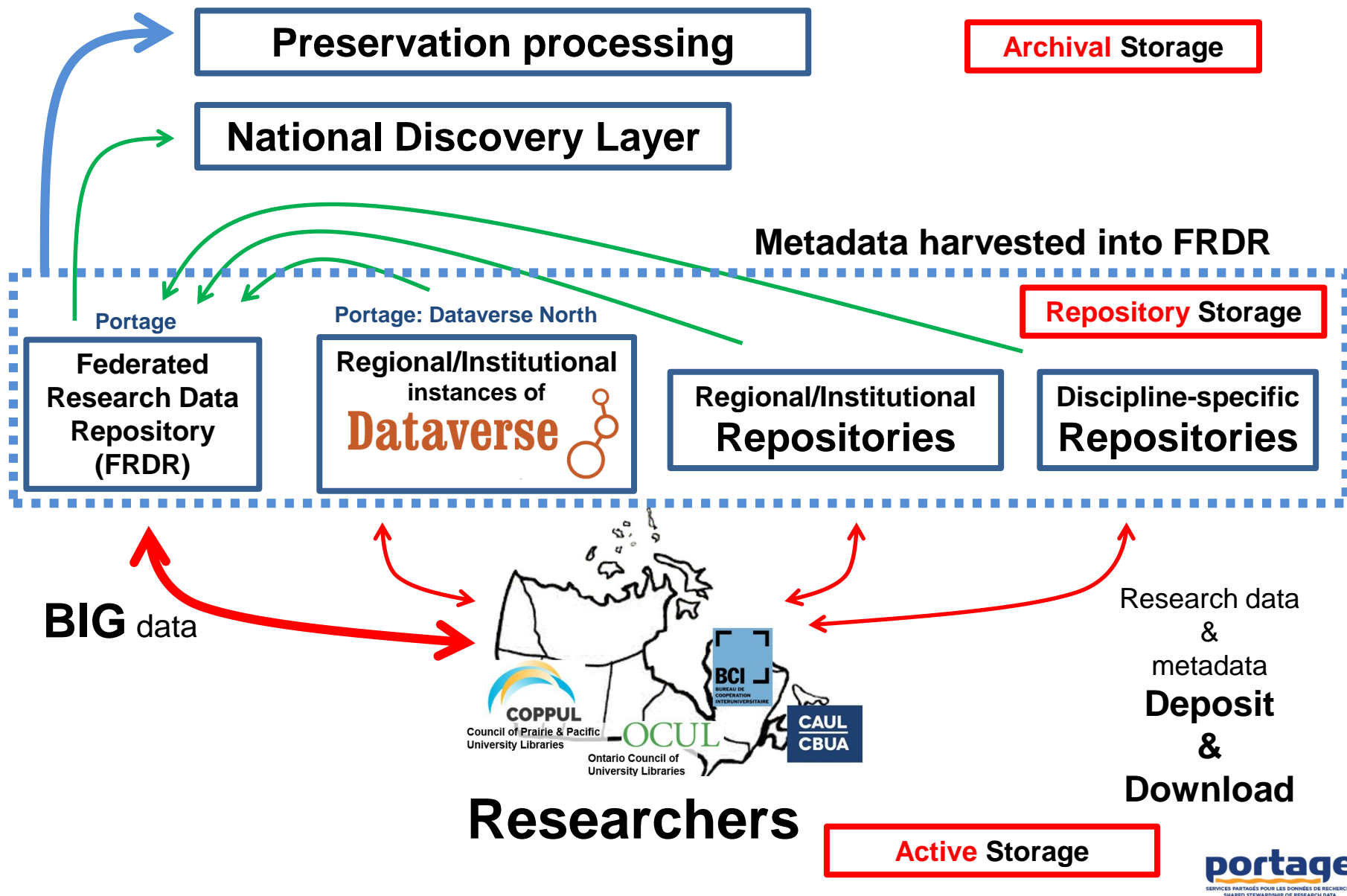


- FRDR's harvester indexes data repositories from across Canada **to make research data held in many repositories discoverable from a single platform**
- Bilingual user interface for search and discovery
- Indexing 40+ repositories; 90,000 metadata records
- Metadata \neq Data

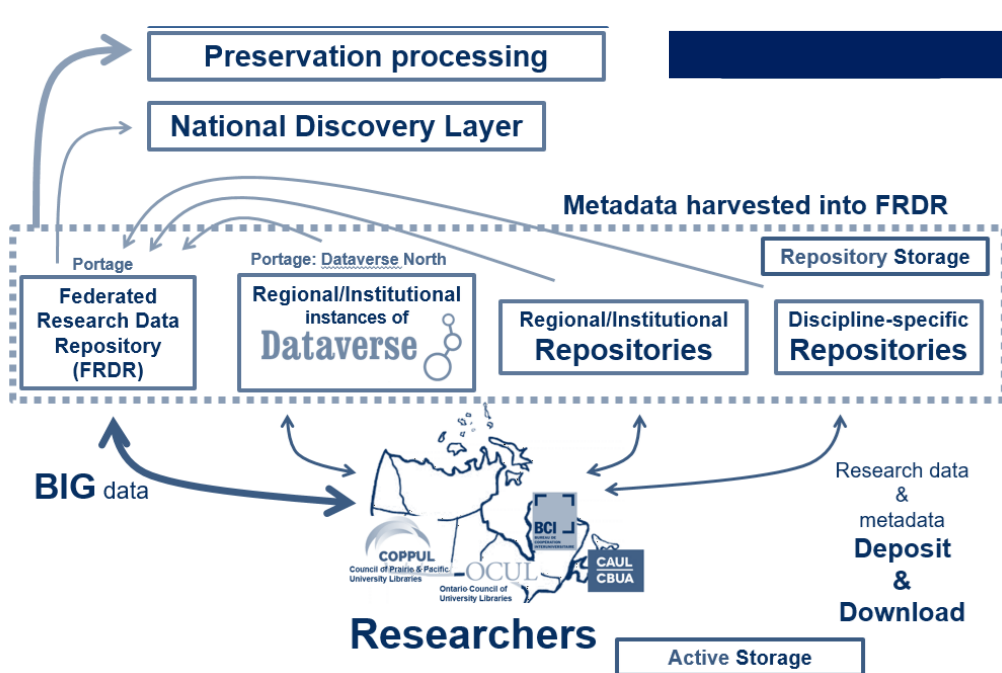
Goals:

- Improve discovery of research data/metadata across Canada
- Drive traffic to existing repository sites
- Break down repository siloes

Federated RDM Architecture



Federated RDM Architecture



Archival Storage

Preservation Processing via
Archivematica
 ↓
 Convert to 'future-friendly' formats
 (e.g. docx → PDF/A)
 &
 Generation of
**Archival Information Packages
 (AIPs)**
 ↓
 Storage in a
 Network of
Preservation Service Providers



ONTARIO LIBRARY
 RESEARCH CLOUD

Data Storage Continuum



CONTINUUM OF RESEARCH DATA STORAGE

Research
Lifecycle
from a
storage
perspective

**ACTIVE
STORAGE**

ACCESS:

Controlled

**STORAGE
PURPOSE:**

*Working copy
Short-term
for duration of
project*

USE:

**Completing
research**

Data Storage Continuum

CONTINUUM OF RESEARCH DATA STORAGE

Research
Lifecycle
from a
storage
perspective

**ACTIVE
STORAGE**

**REPOSITORY
STORAGE**

ACCESS:

Controlled

Open

(as appropriate)

**STORAGE
PURPOSE:**

Working copy
*Short-term
for duration of
project*

Dissemination copy
*Medium-term,
beyond duration of
project*

USE:

**Completing
research**

Discovery & Access

Data Storage Continuum

CONTINUUM OF RESEARCH DATA STORAGE

Research Lifecycle from a storage perspective

ACCESS:

STORAGE PURPOSE:

USE:

ACTIVE STORAGE

Controlled

*Working copy
Short-term
for duration of project*

Completing research

REPOSITORY STORAGE

**Open
(as appropriate)**

*Dissemination copy
Medium-term,
beyond duration of project*

Discovery & Access

Preservation Processing

ARCHIVAL STORAGE

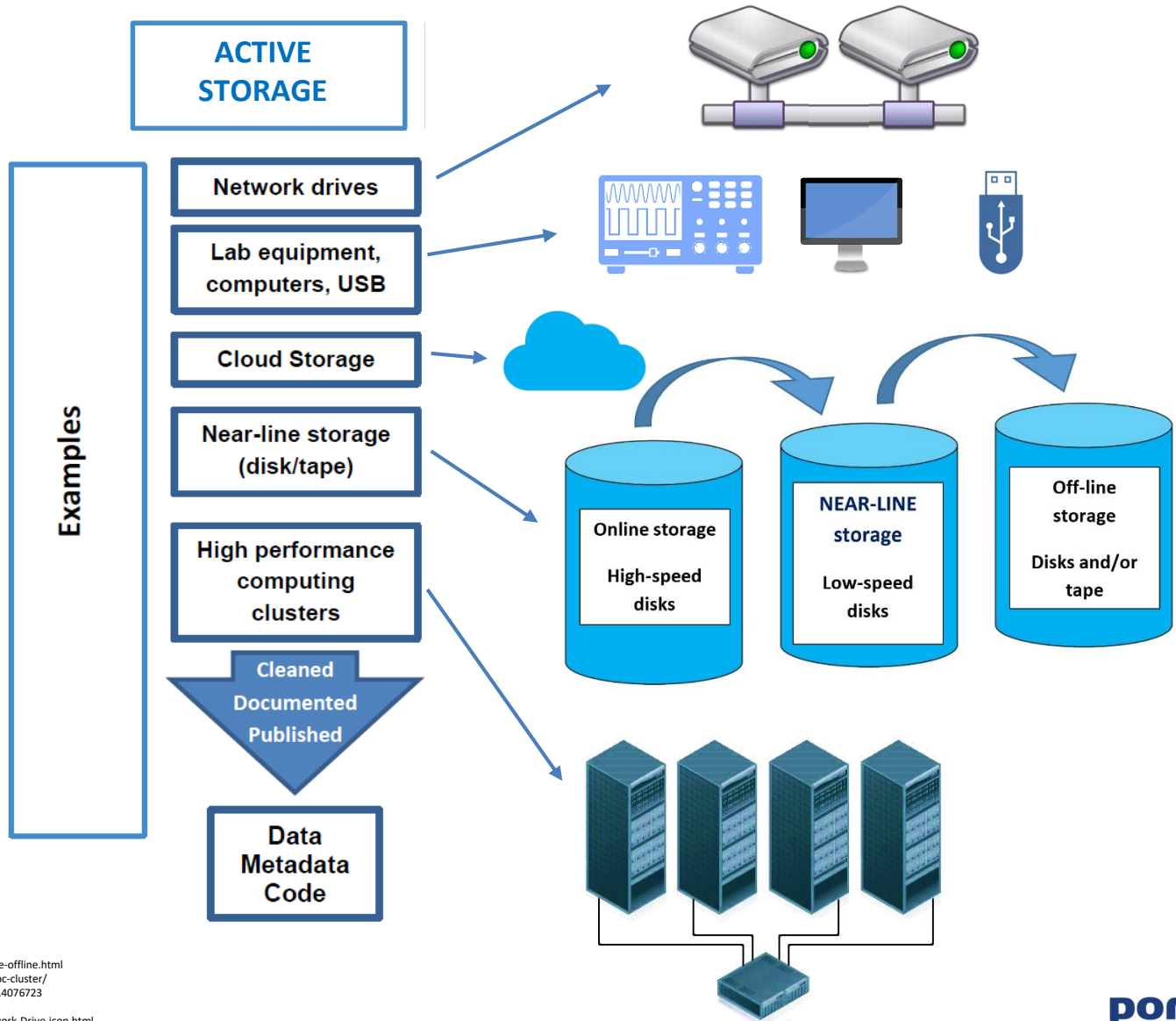
**Open
(as appropriate)**

*Preservation copy
Long-term*

**Disaster recovery/
Copy of last resort**

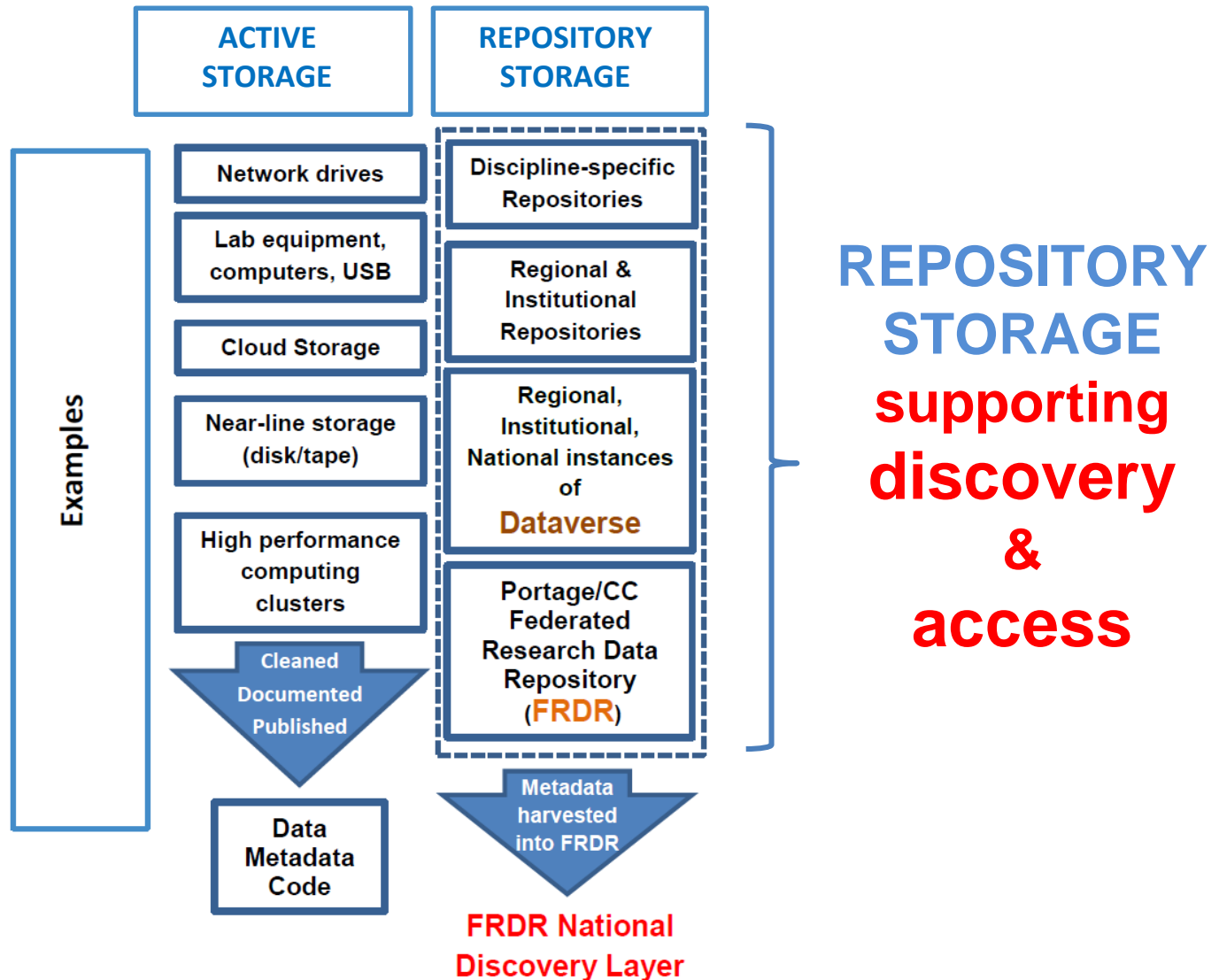
Examples of each type of storage...

ACTIVE STORAGE
supporting
completion
of research



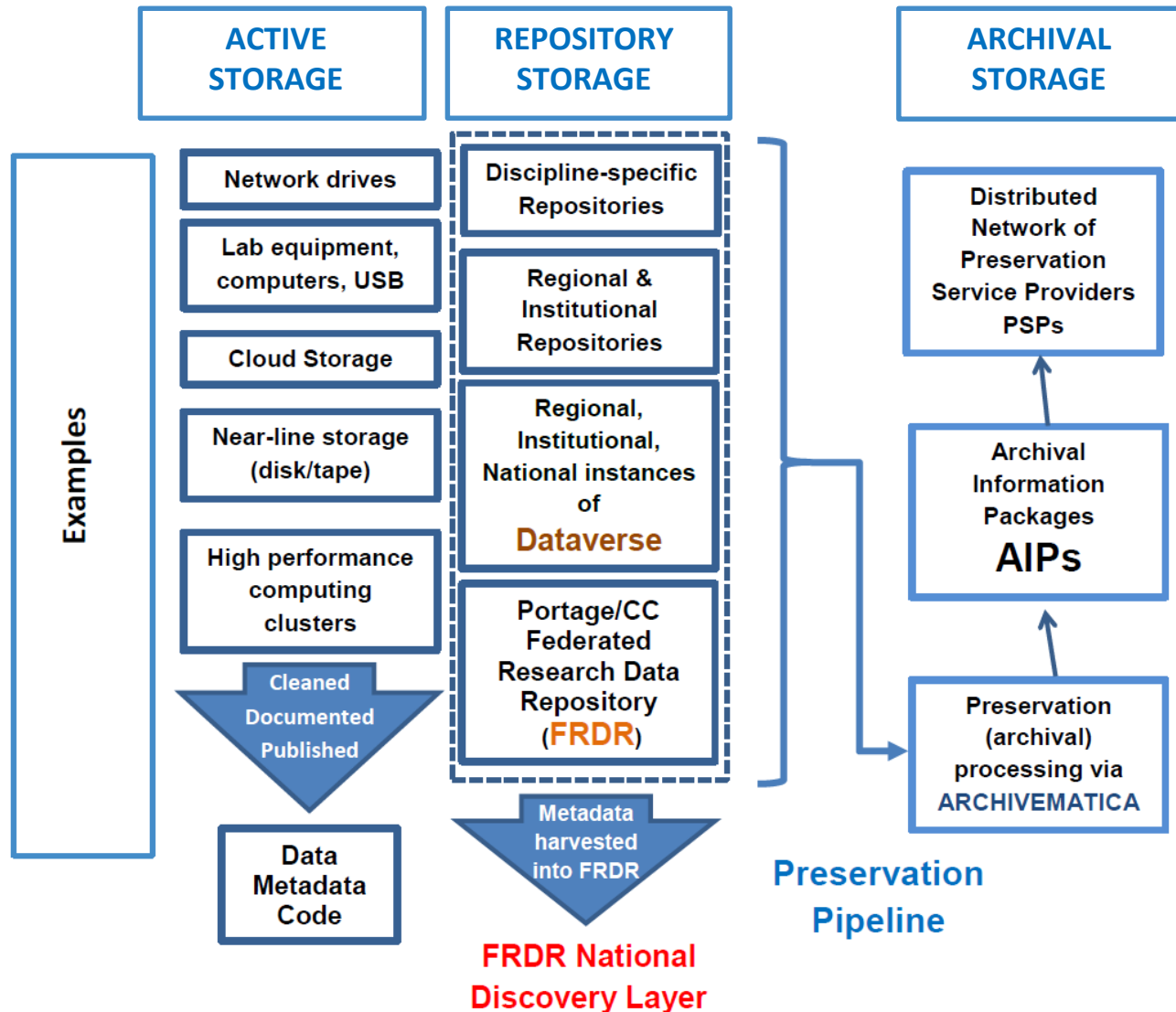
<http://weeratapwutti.blogspot.com/2014/12/online-storage-nearline-storage-offline.html>
<https://www.happyware.com/it-solutions/cluster-solutions/cluster-typology/hpc-cluster/>
<https://www.vectorstock.com/royalty-free-vector/oscilloscope-icon-vector-14076723>
<https://101clipart.com/clipart-computer-monitor/>
<http://www.iconarchive.com/show/refresh-cl-icons-by-tpdkdesign.net/Network-Drive-icon.html>

Examples of each type of storage...

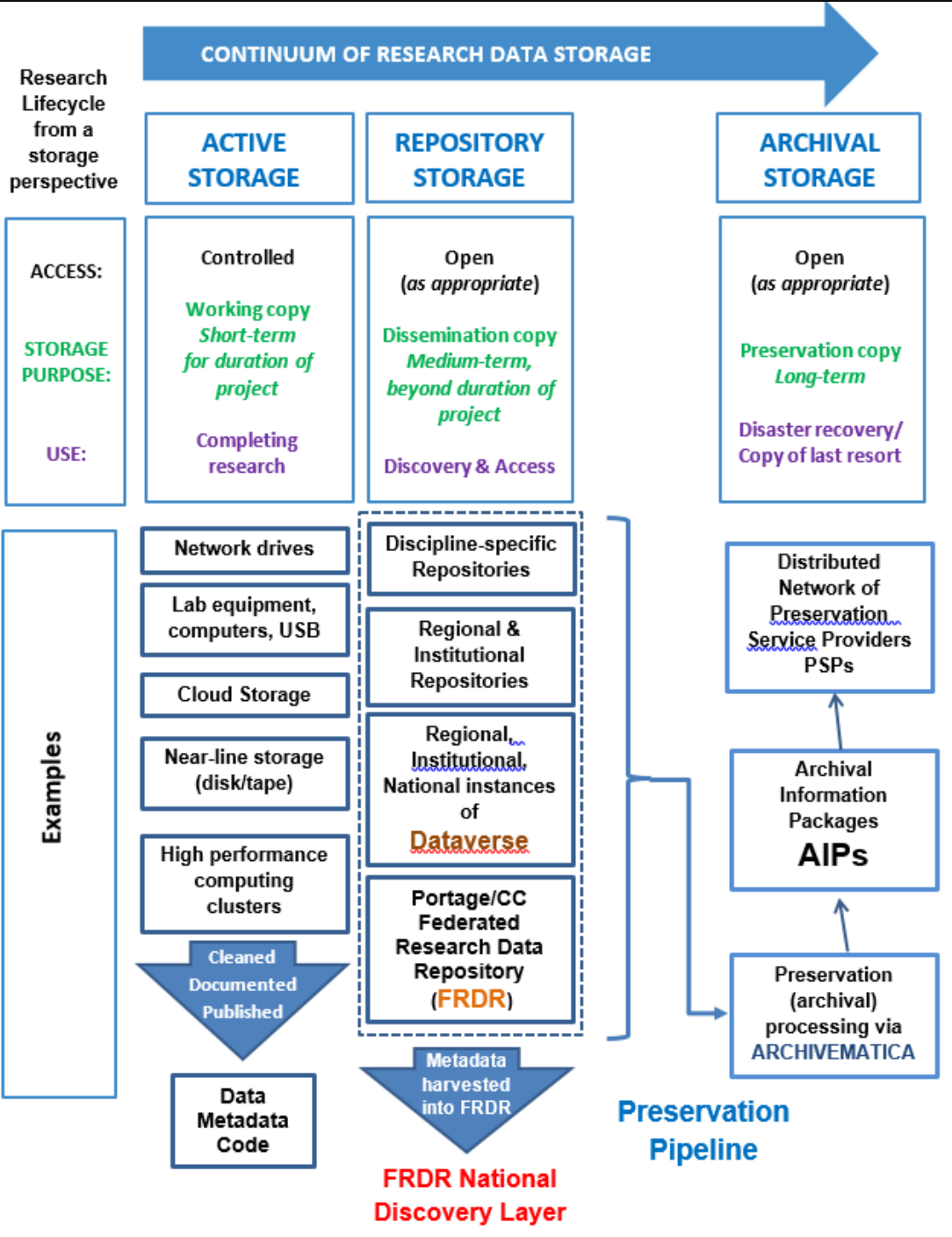


Examples of each type of storage...

ARCHIVAL STORAGE
supporting
long-term
preservation

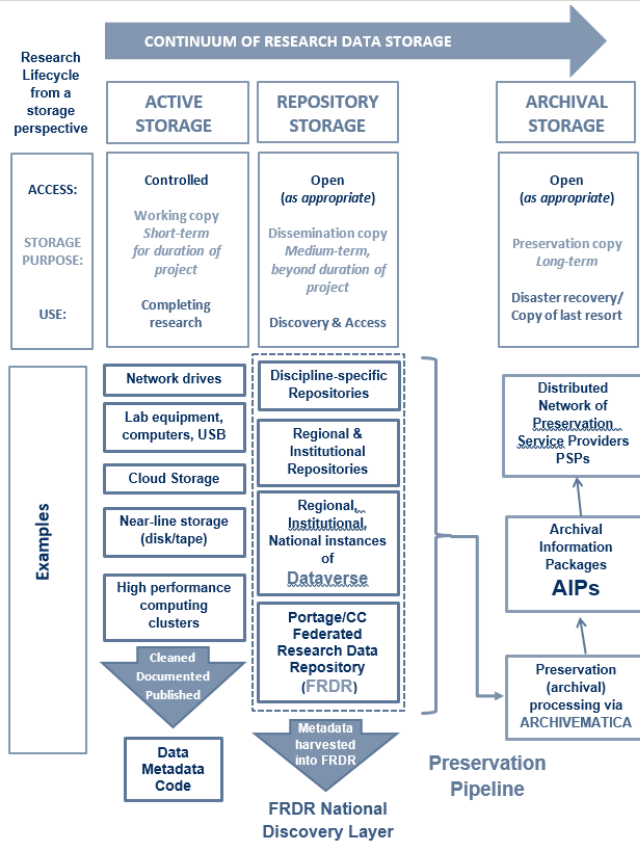


Data Storage Continuum



A useful way of framing the storage picture for RDM

But other activities happen along the way...



And the whole can be described as
GOOD DATA STEWARDSHIP

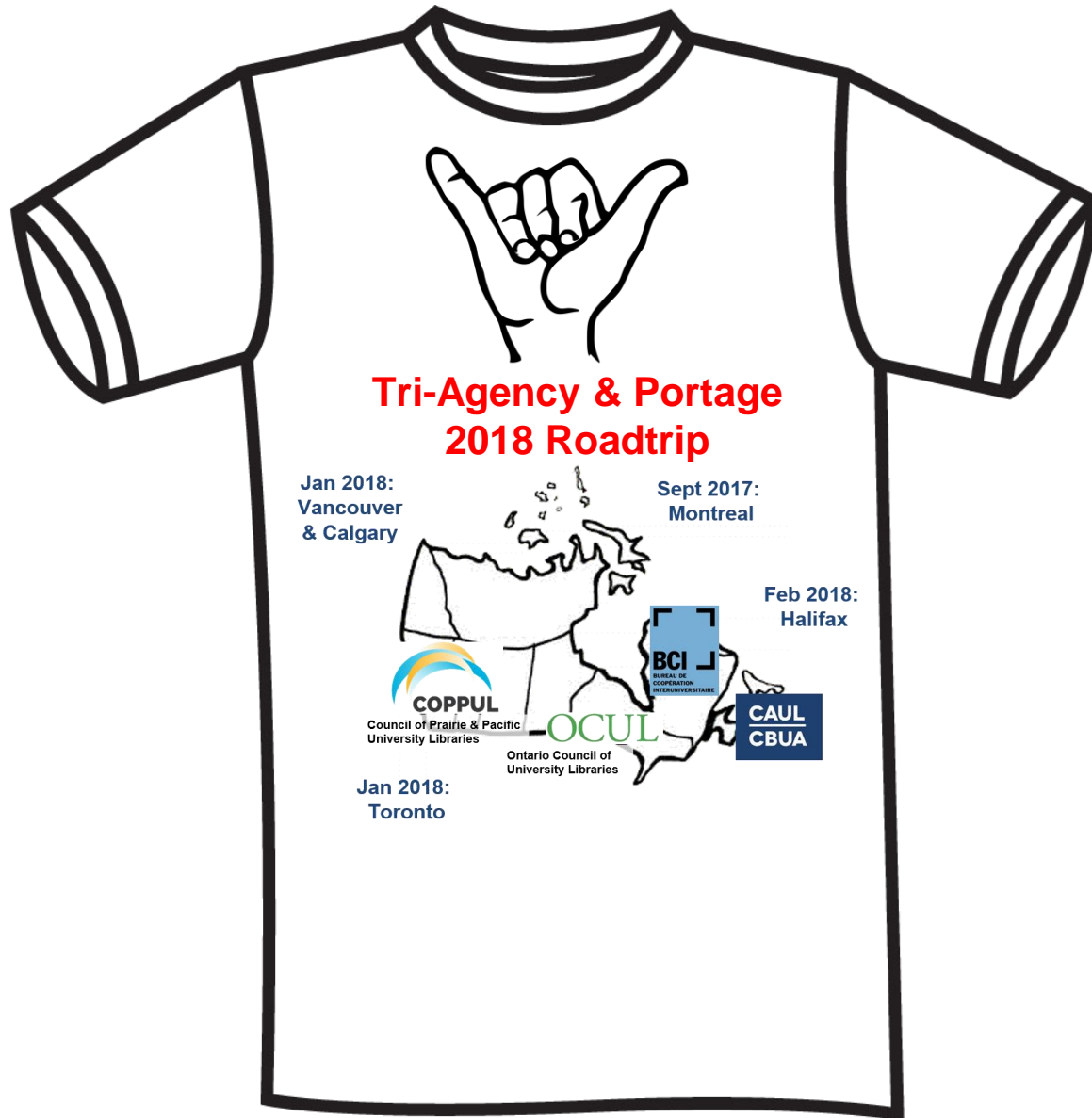
DATA MANAGEMENT PLANNING

DATA CURATION

Staying in touch with RDM stakeholders...

Portage Outreach

Portage Outreach with the Tri-Agencies



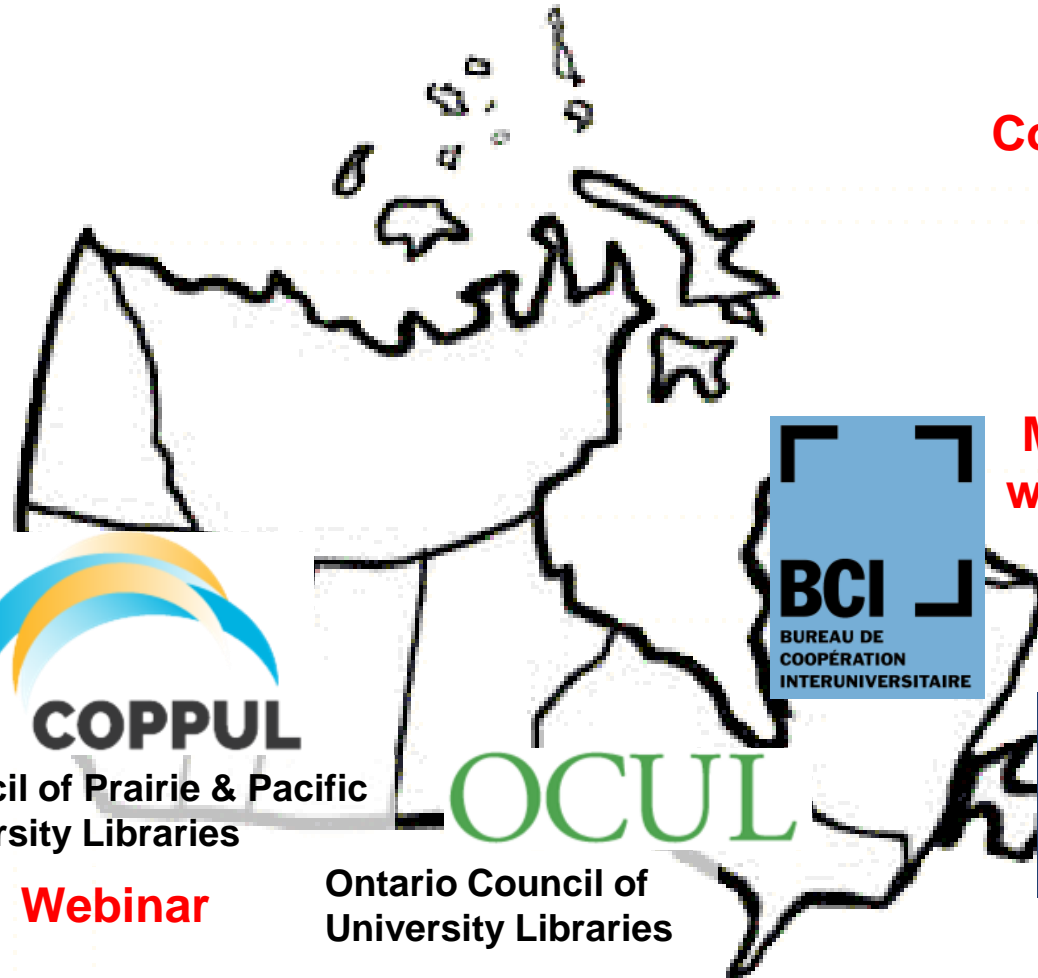
Portage Outreach: *national and diverse*



Regular updates & calls
Steering Committee
Advisory Committee
Infrastructure Committee

CANARIE RDM Funding Call

Regular
Regional
Consortia calls



COPPUL
Council of Prairie & Pacific
University Libraries

Webinar

Ontario Council of
University Libraries

OCUL

BCI
BUREAU DE
COOPÉRATION
INTERUNIVERSITAIRE

Meetings
with BCI &
FRQ

**CAUL
CBUA**

Webinar

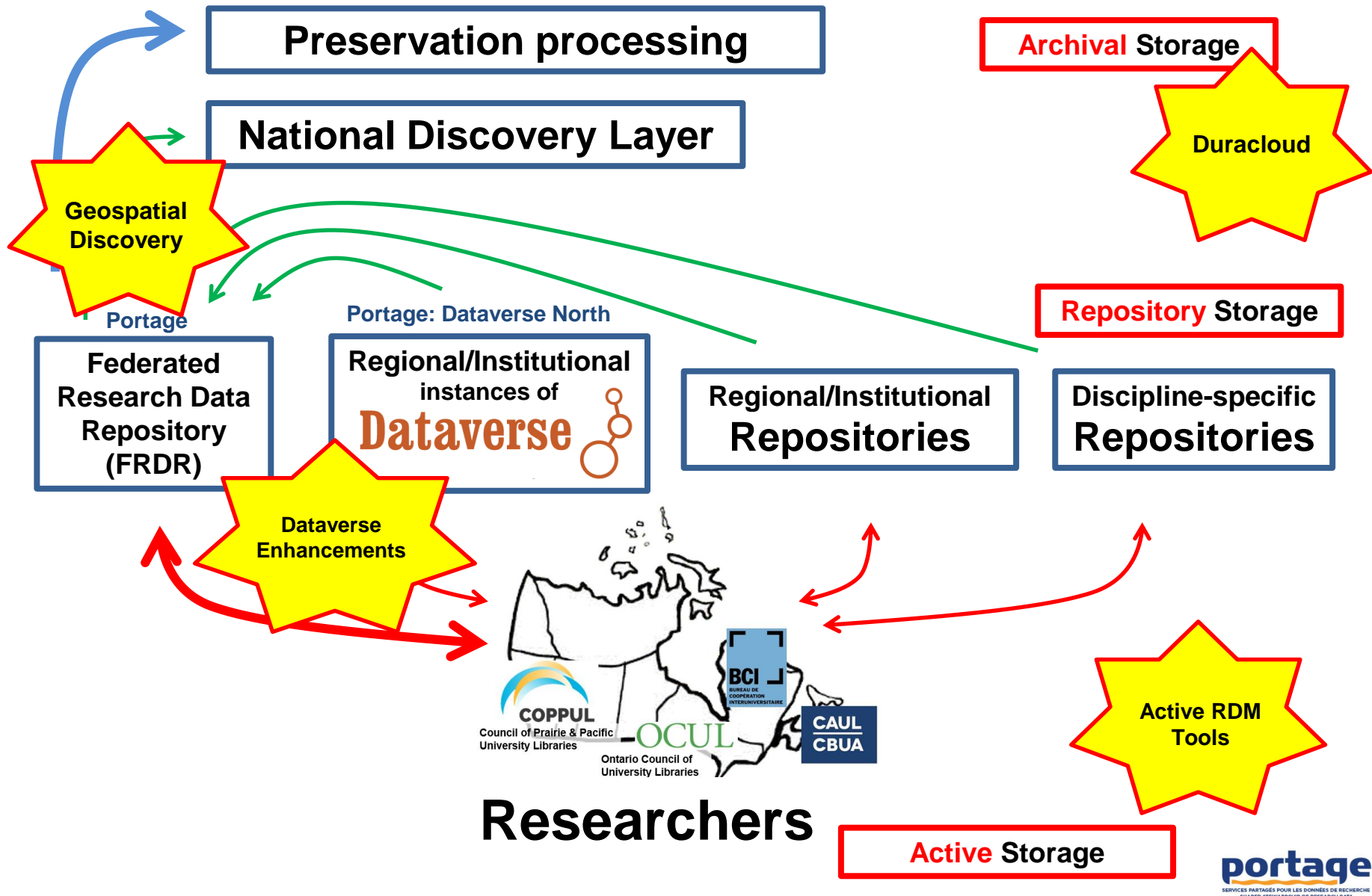
CICan Applied
Research
Symposium

RDM Workshops @
WLU and Lakehead

Successful CANARIE RDM Proposals

- **A National Geospatial Discovery Layer** to improve the discoverability of geospatial metadata harvested by FRDR and create connections between Dataverse and a GeoServer to enhance Dataverse's ability to visualize geospatial information and generate geocoded metadata for subsequent harvesting by FRDR. Proposal is led by UBC with support from Compute Canada and Scholar's Portal.
- **Improved integration with 'Active' RDM tools** through metadata aggregation and enhancement plug ins/APIs (for example, HubZero and Open Science Framework). Proposal to be led by the Compute Canada RDM team at USask (the same team that developed FRDR).
- **Dataverse enhancements for national repository deployment** - enhance features in Dataverse to make it a better fit for use as a national repository option, including integration with Shibboleth, with the OLRC, and simplified workflows for uploading data, especially large datasets (i.e., integration with Globus File Transfer and Compute Canada storage infrastructure). Proposal led by Scholar's Portal.
- **Data Preservation:** Enhance DuraCloud to be deployed outside the AWS system and add integration for multiple Canadian research data repositories and storage platforms. Led by U of Toronto.

Federated RDM Architecture



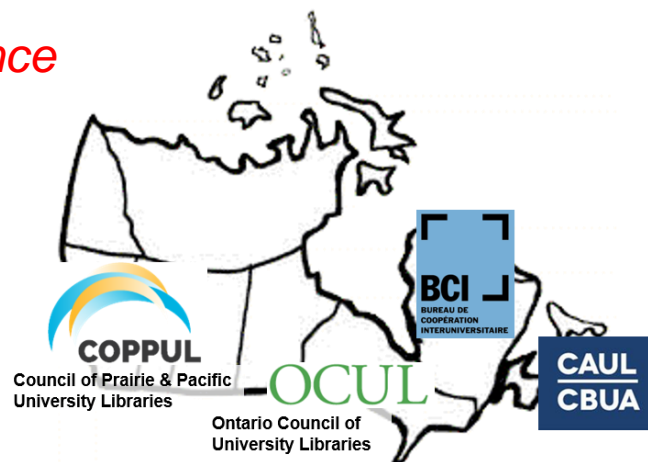
Portage Outreach: *lots more on the go!*

Conferences, Symposia, AGMs

- G7 Research Summit
- Data Curation Workshop
- CNC CODATA AGM
- IASSIST & RDA
- Colleges – Applied Research Symposium
- College VPR Summit
- Statistics Canada @ 100 panel
- NDSF Summit
- OLA Super Conference
- Open Science Conference
- IASSIST

Regular and Targeted Meetings

- Portage-ISED meetings
- Portage-Compute Canada meetings
- Portage-RDC meetings
 - DM 'Dream Team' meeting
 - regular monthly calls with RDC
- Portage-Regional Consortia calls
- Portage-Scholars Portal calls



Portage: Partnerships & Progress



Dataverse



CANARIE RDM Funding



Curation Outreach Preservation Discovery

Training Policy Development

Sensitive Data

- Build a Community of Data Curators (CEG)
- Add Geospatial discovery features for FRDR & Dataverse
- Launch “RDM 101” training modules (TEG)
- Improve tools for “active” RDM (CCF/Radium)
- Offer FRDR & Dataverse as national repository options
- Transition to ‘new organization’ under ISED funding

More to come!

New Business Models

CIHR Training Modules

SSHRC DMP Workshop

COPPUL MOU

CSA -Life Sciences

Website Launch

Compute Canada/FRDR

DMP Assistant

First Portage Director

CARL Approval

2016

15

2015

DVN Report

BCI MOU

Sensitive Data WG

Regional Library Consortia

Portage Day - Toronto

Institutional Strategy WG

Second Portage Director

FRDR – Limited Production

Portage Day - Montreal

Formative Assessment

LCDRI

CAUL MOU

Dataverse North

OCUL MOU

New Governance Model

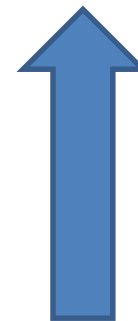
115

2018

98

2017

42



More to come!

2019

2018

Governance

Platforms/Services

Outreach

Agreements

Colleges

Non-CARL Libraries

International Outreach

Portage/Tri-Agency Tour

Preservation White Paper

CARL Directors Support

2014: **Project ARC**



Questions?



portagenetwork.ca

Jeff Moon, Portage Director
portagedir@carl-abrc.ca