portage

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

CARL Portage:
The Canadian Digital Research Infrastructure Landscape

Canadian Data Curation Forum
McMaster University
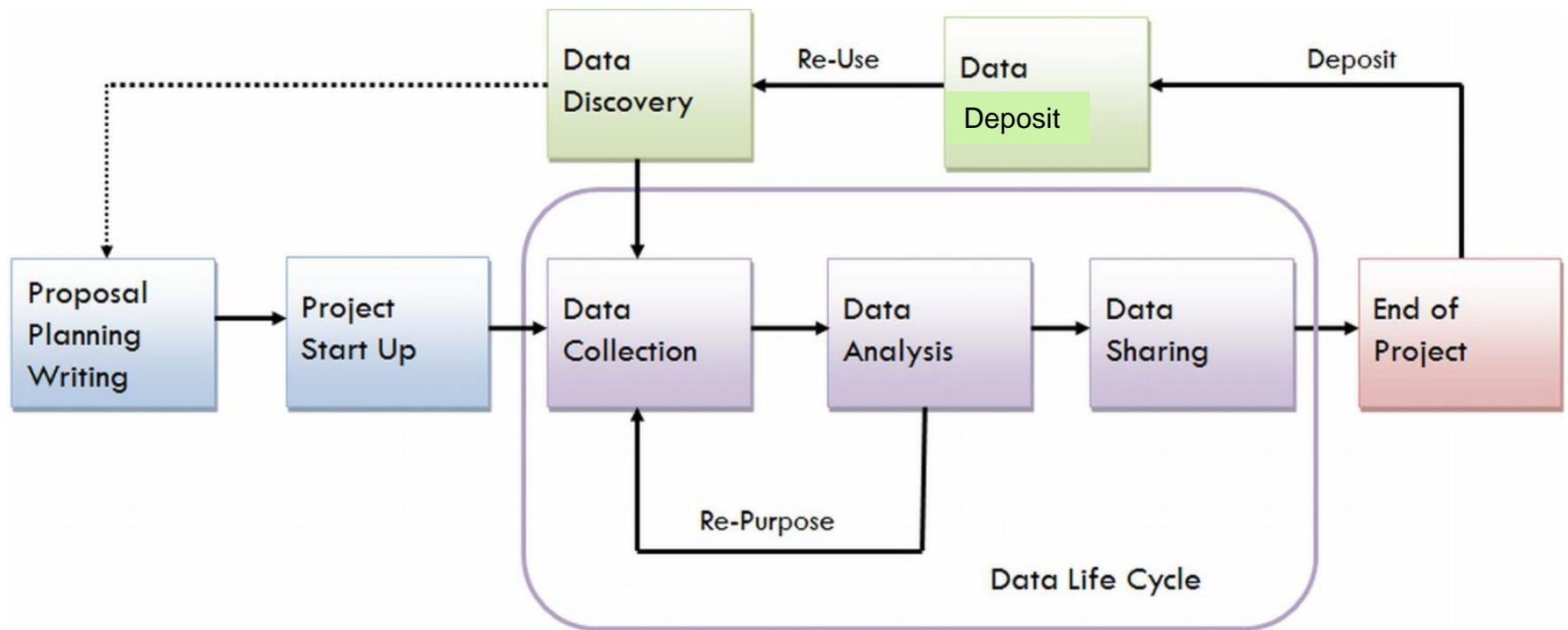October 17, 2019

Jeff Moon, Director, Portage

Curation covers the entire research data life cycle…

SHOULD YOU *CURATE* YOUR DATA?
A FLOW DIAGRAM
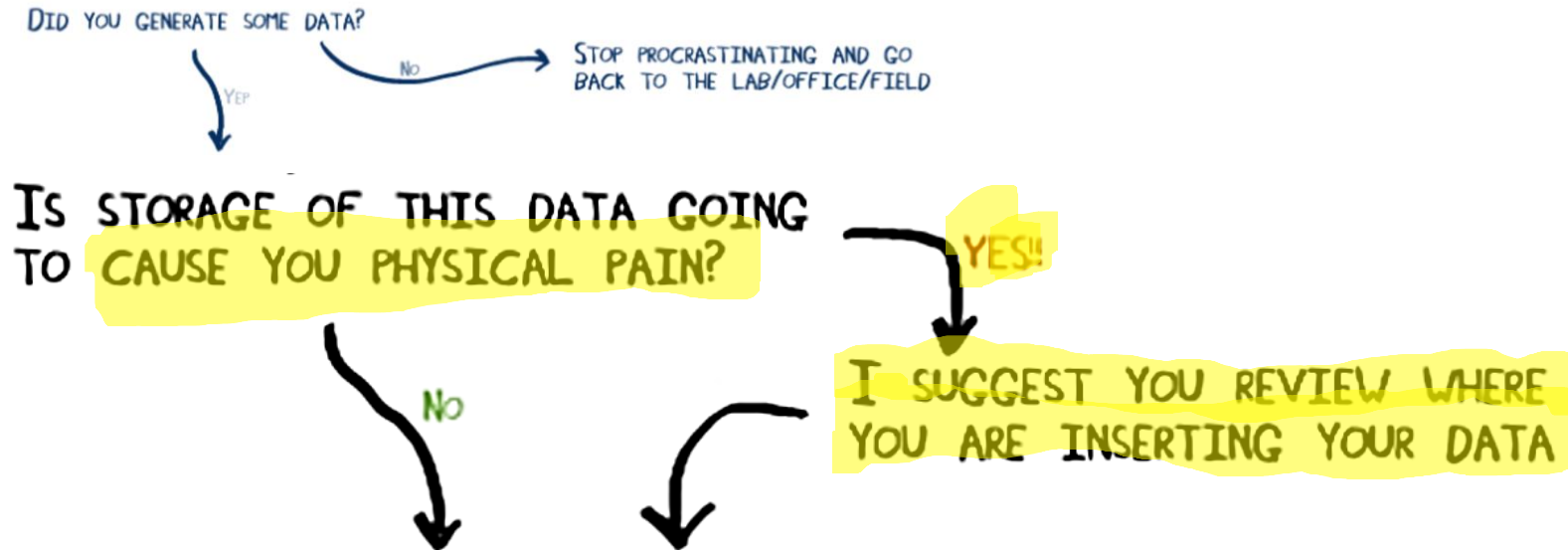
ErrantScience.com

DID YOU GENERATE SOME DATA?

No → STOP PROCRASTINATING AND GO
BACK TO THE LAB/OFFICE/FIELD

YEP

IS STORAGE OF THIS DATA GOING
TO CAUSE YOU PHYSICAL PAIN?

YES!!

NO

I SUGGEST YOU REVIEW WHERE
YOU ARE INSERTING YOUR DATA

SHOULD YOU *CURATE* YOUR DATA?
A FLOW DIAGRAM
ERRANTSCIENCE.COM

DID YOU GENERATE SOME DATA?
No → STOP PROCRASTINATING AND GO BACK TO THE LAB/OFFICE/FIELD
YEP

IS STORAGE OF THIS DATA GOING TO CAUSE YOU PHYSICAL PAIN?
YES! → I SUGGEST YOU REVIEW WHERE YOU ARE INSERTING YOUR DATA
No

IS SAVING THIS DATA GOING TO LEAD TO THE DOWNFALL OF CIVILISATION?
MAYBE
UNLIKELY

HAVE YOU MET CIVILISATION? WOULD THAT REALLY BE A BAD THING?

SHOULD YOU *CURATE* YOUR DATA?
A FLOW DIAGRAM
ErrantScience.com

DID YOU GENERATE SOME DATA?

No → STOP PROCRASTINATING AND GO BACK TO THE LAB/OFFICE/FIELD

YEP ↓

IS STORAGE OF THIS DATA GOING TO CAUSE YOU PHYSICAL PAIN?

YES! → I SUGGEST YOU REVIEW WHERE YOU ARE INSERTING YOUR DATA

No ↓

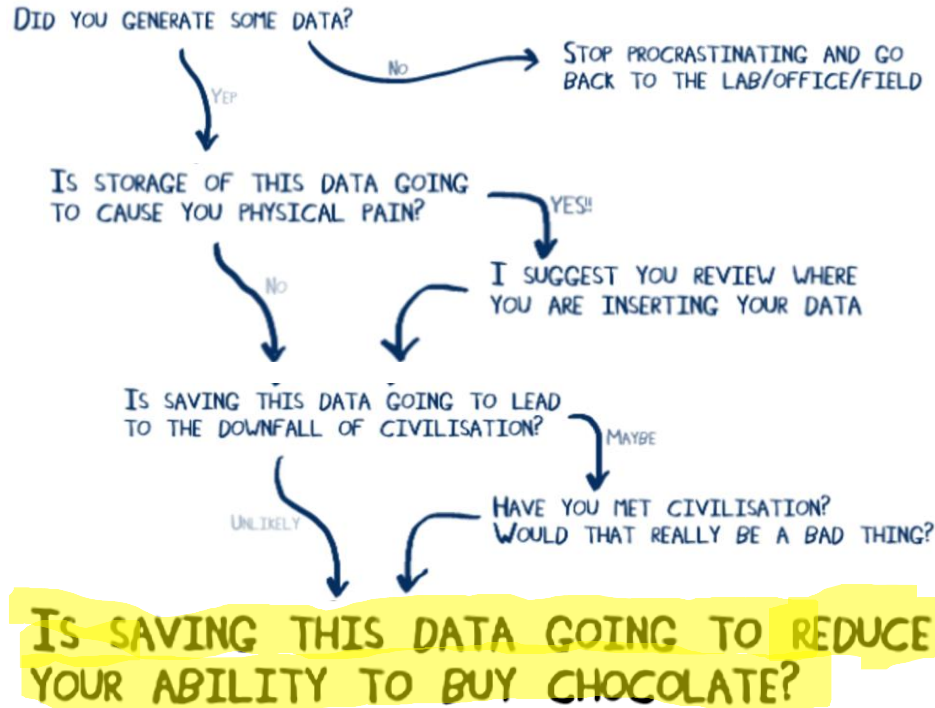IS SAVING THIS DATA GOING TO LEAD TO THE DOWNFALL OF CIVILISATION?

MAYBE → HAVE YOU MET CIVILISATION? WOULD THAT REALLY BE A BAD THING?

UNLIKELY ↓

IS SAVING THIS DATA GOING TO REDUCE YOUR ABILITY TO BUY CHOCOLATE?

https://www.news.com.au/lifestyle/food/eat/the-reason-why-you-shouldnt-store-chocolate-in-the-fridge/news-story/4cdb70a6da223d92fe1b95789ff5377f
https://errantscience.com/blog/2018/03/14/should-you-save-your-data-a-flow-diagram/

SHOULD YOU *CURATE* YOUR DATA?
A FLOW DIAGRAM
ErrantScience.com

Did you generate some data?

No → Stop procrastinating and go back to the lab/office/field

Yep ↓

Is storage of this data going to cause you physical pain?

YES!! → I suggest you review where you are inserting your data

No ↓

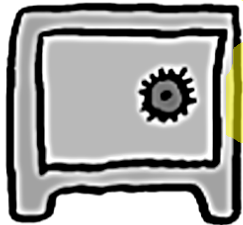Is saving this data going to lead to the downfall of civilisation?

Maybe → Have you met civilisation? Would that really be a bad thing?

Unlikely ↓

Is saving this data going to reduce your ability to buy chocolate?

It might!! → Discard all data immediately, it's not worth risking it.

No ↓

CURATE YOUR DATA

and *METADATA* and Code

https://www.news.com.au/lifestyle/food/eat/the-reason-why-you-shouldnt-store-chocolate-in-the-fridge/news-story/4cdb70a6da223d92fe1b95789ff5377f
https://errantscience.com/blog/2018/03/14/should-you-save-your-data-a-flow-diagram/

Many reasons to curate [manage] research data…

Reproducibility

Verification of results

Extending research

Discoverability

Encour~~~~~~~~~~~~

Complyi~~~~~~~~~~ ~~~ents

FAIR Principles

Avoiding duplication

Aligning with international best practices & standards

Making efficient and effective use of public funds

What keeps me up at night…

Avoiding domination by commercial Interests

Preventing loss of data

data

Data Curation

The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials.

Data Curation

casrai

The activity of managing and promoting the use of data from their point of creation to ensure that they are <mark>fit for contemporary purpose and available for discovery and reuse</mark>. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials.

11 ways to avert a data-storage disaster

Hard-drive failures are inevitable, but data loss doesn't have to be.

Jeffrey M. Perkel
01 APRIL 2019

1. Apply the 3-2-1 rule

2. Talk to the specialists

3. Manage your data

4. Safeguard privacy

5. Automate backing up

6. Protect raw data

7. Make backing up achievable

8. Test backups regularly

9. Expect the unexpected

10. Keep a backup offline

11. Plan ahead

nature
International journal of science

1. **Apply the 3-2-1 rule**

2. **Talk to the specialists**

3. **Manage your data**

4. **Safeguard privacy**

5. **Automate backing up**

6. **Protect raw data**

7. **Make backing up achievable**

8. **Test backups regularly**

9. **Expect the unexpected**

10. **Keep a backup offline**

11. **Plan ahead**

Your institution employs people to think about data full-time, so talk to them, advises Juliane Schneider, who leads data curation at Harvard Catalyst in Boston, Massachusetts. Your research-computing centre might offer free or low-cost institutional backup systems; your librarian can help you to craft a data-management strategy; and your grants office can advise you on funding-agency requirements, including how, and for how long, data must be maintained. "They want to help you keep your data — especially if you have a grant," she says.

1. **Apply the 3-2-1 rule**

2. **Talk to the specialists**

3. **Manage your data**

4. **Safeguard privacy**

5. **Automate backing up**

6. **Protect raw data**

7. **Make backing up achievable**

8. **Test backups regularly**

9. **Expect the unexpected**

10. **Keep a backup offline**

11. **Plan ahead**

Your institution employs people to think about data full-time, so talk to them, advises Juliane Schneider, who leads data curation at Harvard Catalyst in Boston, Massachusetts. Your research-computing centre might offer free or low-cost institutional backup systems; your librarian can help you to craft a data-management strategy; and your grants office can advise you on funding-agency requirements, including how, and for how long, data must be maintained. "They want to help you keep your data — especially if you have a grant," she says.

1. Apply the 3-2-1 rule

2. Talk to the specialists

3. Manage your data

4. Safeguard privacy

5. Automate backing up

6. Protect raw data

7. Make backing up achievable

8. Test backups regularly

9. Expect the unexpected

10. Keep a backup offline

11. Plan ahead

Your institution employs people to think about data full-time, so talk to them, advises Juliane Schneider, who leads data curation at Harvard Catalyst in Boston, Massachusetts. Your research-computing centre might offer free or low-cost institutional backup systems; your librarian can help you to craft a data-management strategy; and your grants office can advise you on funding-agency requirements, including how, and for how long, data must be maintained. "They want to help you keep your data — especially if you have a grant," she says.

1. **Apply the 3-2-1 rule**

2. **Talk to the specialists**

3. **Manage your data**

4. **Safeguard privacy**

5. **Automate backing up**

6. **Protect raw data**

7. **Make backing up achievable**

8. **Test backups regularly**

9. **Expect the unexpected**

10. **Keep a backup offline**

11. **Plan ahead**

Your institution employs people to think about data full-time, so talk to them, advises Juliane Schneider, who leads data curation at Harvard Catalyst in Boston, Massachusetts; your research-computing office can help you opt for low-cost or free storage solutions; your librarian can help you to craft a data-management strategy; and your grants office can advise you on funding-agency requirements, including how, and for how long, data must be maintained. "**They want to help you keep your data — especially if you have a grant**," she says.

CURATION

NSF 19-069

Dear Colleague Letter: Effective Practices for Data

Beth Plale, bplale@nsf.gov   May 20, 2019

Putting data in a form that others can use may require work that goes above and beyond a stated research activity. This additional work may be called "data curation" or "data cleaning." PIs may budget for this expense – that is, they may budget for the work needed to prepare research data for distribution.

https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp

NSF 19-069

Dear Colleague Letter: Effective Practices for Data

Beth Plale, bplale@nsf.gov   May 20, 2019

Putting data in a form that others can use may require work that goes above and beyond a stated research activity. This additional work may be called "data curation" or "data cleaning." PIs may budget for this expense – that is, they may budget for the work needed to prepare research data for distribution.

https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp

NSF 19-069

Dear Colleague Letter: Effective Practices for Data

Beth Plale, bplale@nsf.gov   May 20, 2019

Putting data in a form that others can use may require work that goes above and beyond a stated research activity. This additional work may be called "data curation" or "data cleaning." PIs may budget for this expense – that is, they may budget for the work needed to prepare research data for distribution.

NSF 19-069

Dear Colleague Letter: Effective Practices for Data

Beth Plale, bplale@nsf.gov May 20, 2019

Putting data in a form that others can use may require work that goes above and beyond a stated research activity. This additional work may be called "data curation" or "data cleaning." PIs may budget for this expense – that is, they may *budget for the work needed to prepare research data for distribution*.

National Science Foundation
WHERE DISCOVERIES BEGIN

NSF 19-069

Dear Colleague Letter: Effective Practices for Data

Beth Plale, bplale@nsf.gov   May 20, 2019

Putting data in a form that others can use may require work that goes above and beyond a stated research activity. This additional work may be called "data curation" or "data cleaning." PIs may budget for this expense – that is, they may budget for the work needed to prepare research data for distribution.

As early as January 2013, <u>NSF allowed principal investigators (PIs) to report data products in their biographical sketches</u>. This extension put scientific data sets on a standing equal to traditional publications, such as peer-reviewed journal articles, juried conference papers, book chapters, and monographs.

National Science Foundation
WHERE DISCOVERIES BEGIN

NSF 19-069

Dear Colleague Letter: Effective Practices for Data

Beth Plale, bplale@nsf.gov   May 20, 2019

Putting data in a form that others can use may require work that goes above and beyond a stated research activity. This additional work may be called "data curation" or "data cleaning." PIs may budget for this expense – that is, they may budget for the work needed to prepare research data for distribution.

As early as January 2013, NSF allowed principal investigators (PIs) to report data products in their biographical sketches. This extension put scientific data sets on a standing equal to traditional publications, such as peer-reviewed journal articles, juried conference papers, book chapters, and monographs.

https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp

Against this backdrop

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

2017-2019

Draft Tri-Agency Research Data Management Policy

Institutional Strategy

DMPs

Deposit

2014

2015
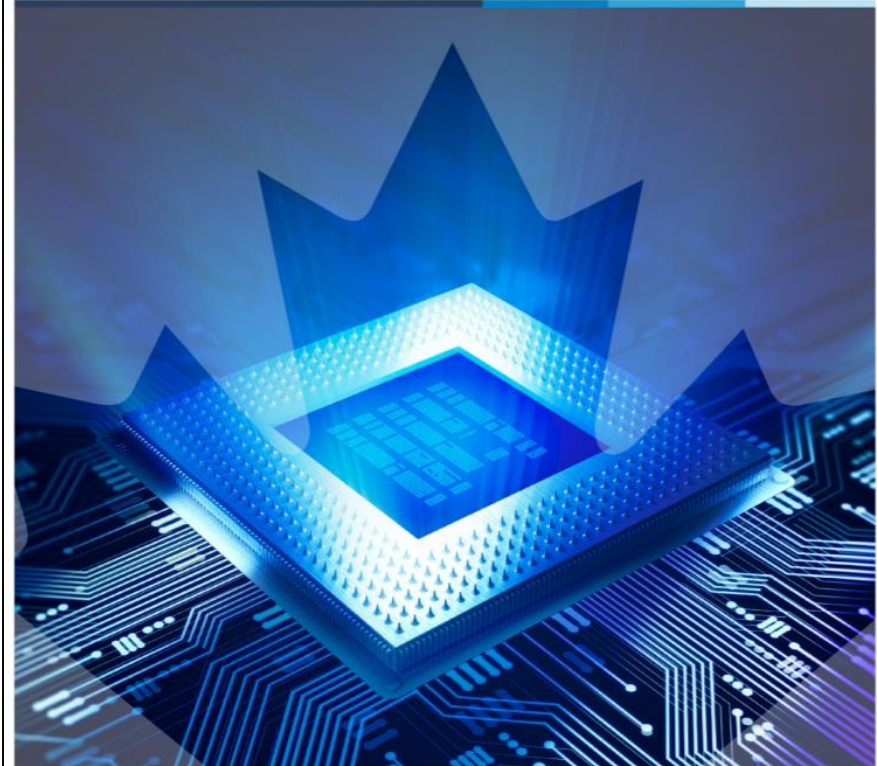
2017: Data Management and ARC position papers

Data Management Position Paper

For Innovation, Science, and Economic Development Canada

2017-08-31

Advanced Research Computing (ARC) Position Paper

For Innovation, Science, and Economic Development Canada

2017-08-31

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 Expertise
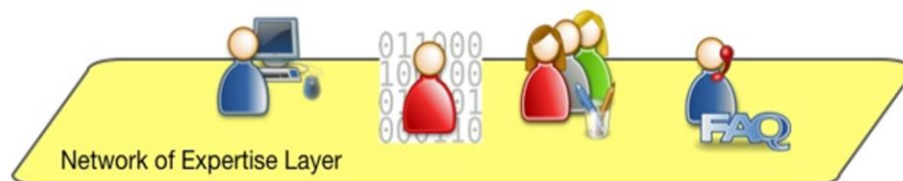
- Pan-Canadian network of RDM expertise
- Contribute to development of RDM resources, tools, training, and policy

Infrastructure Platforms

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

Portage Expert & Working Groups

Expert Groups

- Data Management Planning

- Curation

- Data Discovery

- Preservation

- Training

- Research Intelligence

- Data Repositories

130+

Experts
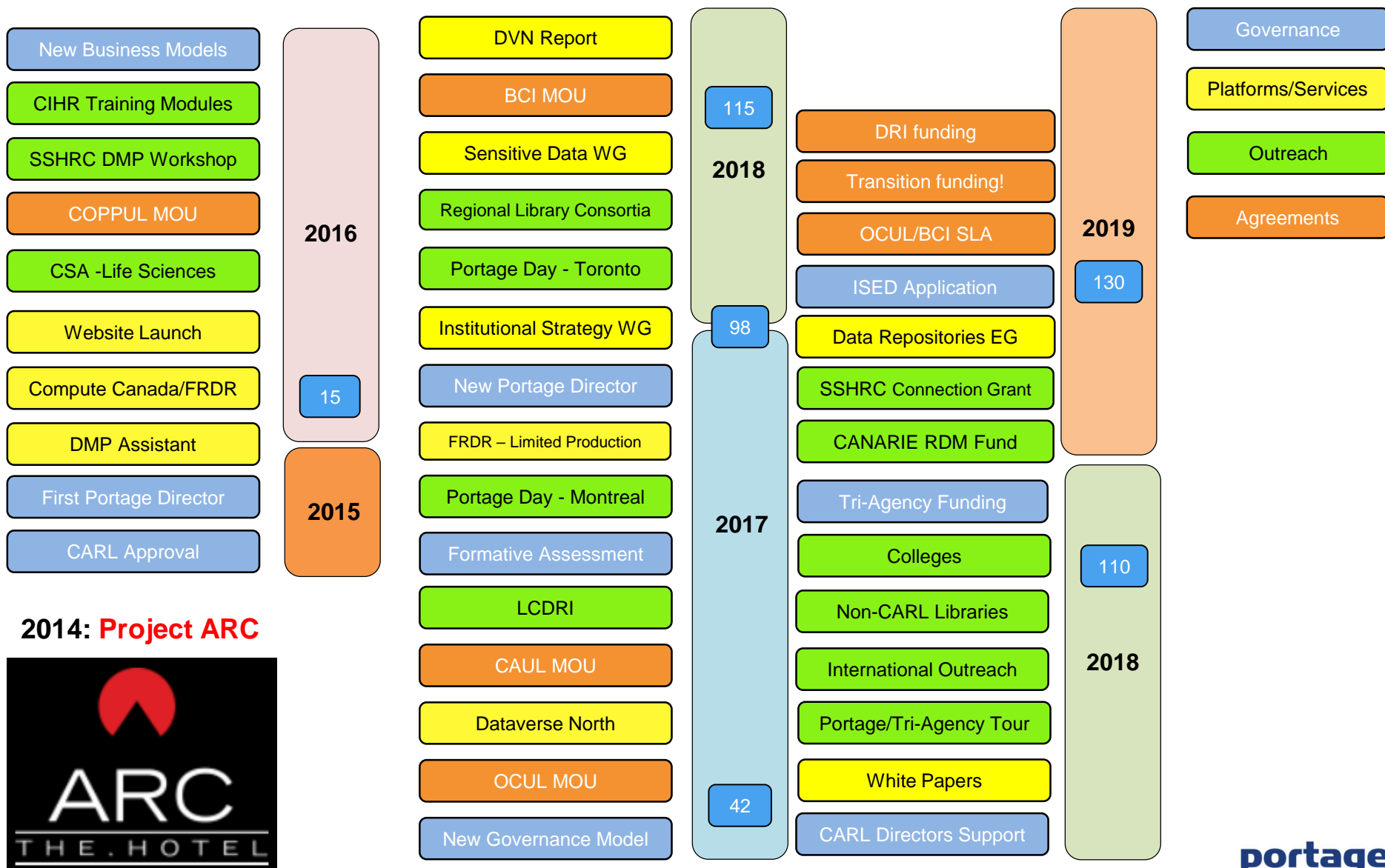
60+

Organizations

Working Groups

- Dataverse North WG

- Responsible Research Data Management Practices for Sensitive Data

- FRDR Service Model WG

- FRDR Discovery Service WG

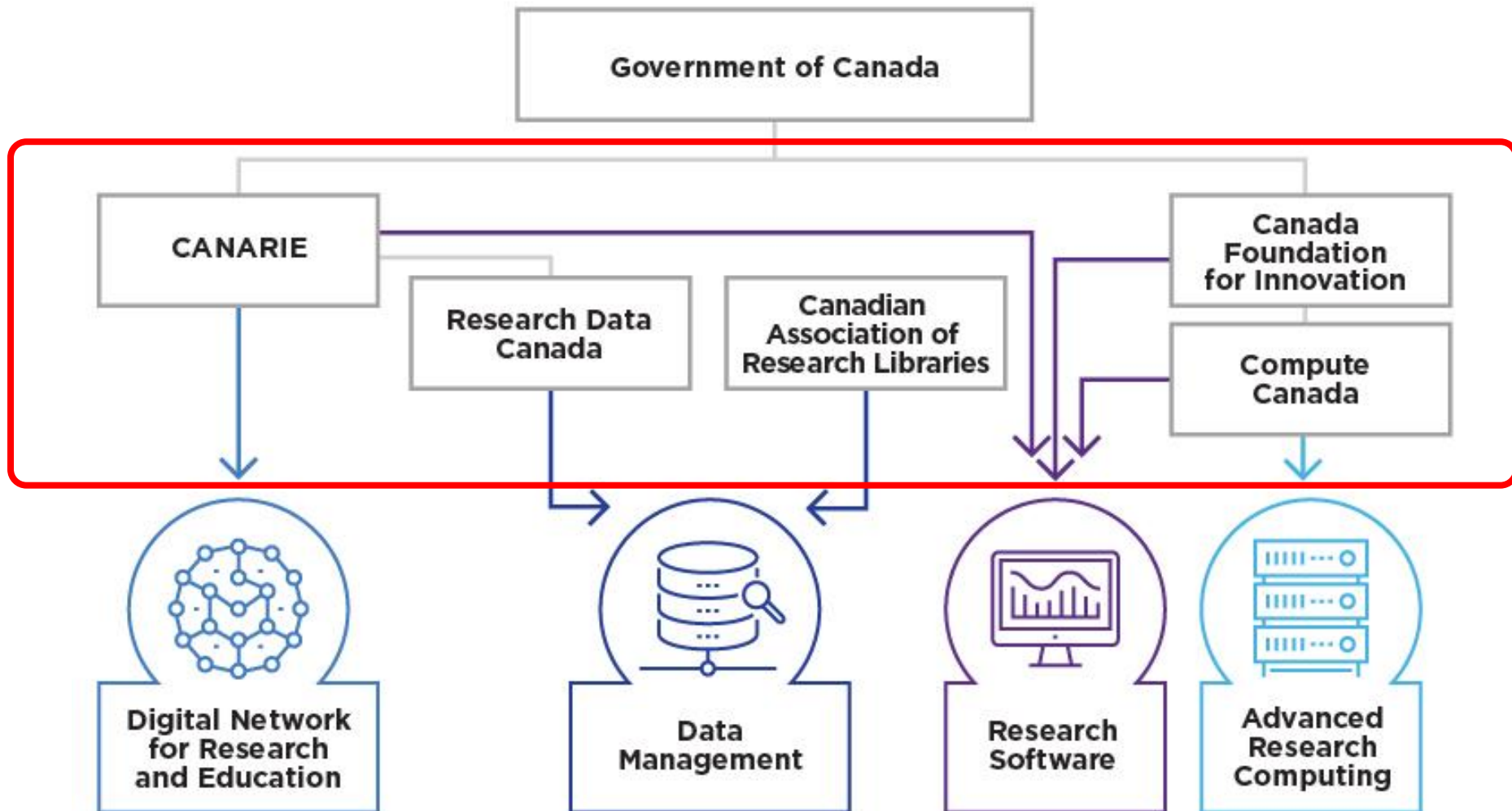- Institutional RDM Strategies

| LOCAL | REGIONAL/ PROVINCIAL | NATIONAL |

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

The Portage Network has been busy!!

2016
- 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

15

2015

2014: Project ARC

2017
- DVN Report
- BCI MOU
- Sensitive Data WG
- Regional Library Consortia
- Portage Day - Toronto
- Institutional Strategy WG
- New Portage Director
- FRDR – Limited Production
- Portage Day - Montreal
- Formative Assessment
- LCDRI
- CAUL MOU
- Dataverse North
- OCUL MOU
- New Governance Model

98
42

2018
115

- DRI funding
- Transition funding!
- OCUL/BCI SLA
- ISED Application
- Data Repositories EG
- SSHRC Connection Grant
- CANARIE RDM Fund
- Tri-Agency Funding
- Colleges
- Non-CARL Libraries
- International Outreach
- Portage/Tri-Agency Tour
- White Papers
- CARL Directors Support

2019
130

2018
110

Legend
- Governance
- Platforms/Services
- Outreach
- Agreements

DRI
&
RDM

Current DRI Landscape

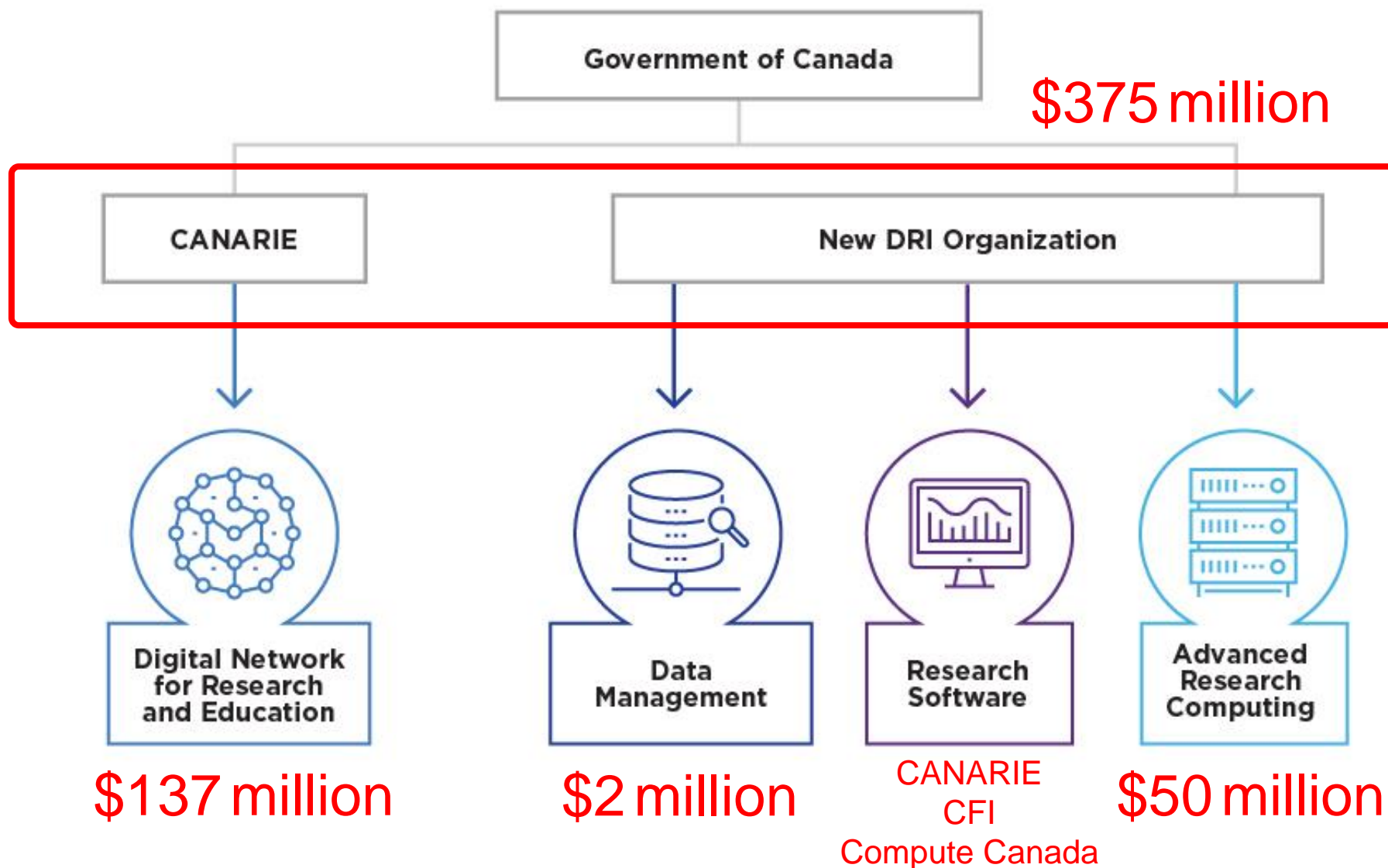# 2018: Federal Budget

Innovation, Science, & Economic Development

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

Moving toward a new DRI organization

$375 million

$137 million

$2 million

CANARIE
CFI
Compute Canada

$50 million

Finally, on August 8th, 2019

Innovation, Science & Economic Development Canada
approved the DRI proposal

- $375 million toward a New DRI Organization

- $50 million to immediately increase the national advanced research computing (ARC) platform's capacity at the five national host sites.

- $137 million to CANARIE, the Canadian not-for-profit responsible for providing the national network for education and research, will be receiving a contribution of $137 million with a renewed mandate, including an enhanced mandate on cybersecurity and dedicated funding for Northern connectivity.

http://www.ic.gc.ca/eic/site/136.nsf/eng/h_00000.html

Finally, on August 8th, 2019

Innovation, Science & Economic Development Canada
approved the DRI proposal

- **$375 million toward a New DRI Organization**

- $50 million to immediately increase the national advanced research computing (ARC) platform's capacity at the five national host sites.

- $137 million to CANARIE, the Canadian not-for-profit responsible for providing the national network for education and research, will be receiving a contribution of $137 million with a renewed mandate, including an enhanced mandate on cybersecurity and dedicated funding for Northern connectivity.

http://www.ic.gc.ca/eic/site/136.nsf/eng/h_00000.html

# On August 22, the 'Applicant Board' was announced

New DRI Organization

An applicant board has been established to undertake the work necessary to incorporate the new organization, including ongoing community consultations to address a number of outstanding governance issues prior to the establishment of the inaugural board in the new year.

- **Dr. Guillaume Bourque**, Director of the Canadian Centre for Computational Genomics, Professor, Human Genetics, McGill University,
- **Ms. Barbara Kieley**, ICD.D, Retired Partner Ernst & Young focus on public sector digital service transformation projects,
- **Mr. Peter MacKinnon** OC, QC, President Emeritus of the University of Saskatchewan, and former interim president of Dalhousie University and Athabasca University, and;
- **Ms. Lori MacMullen**, Executive Director, Canadian University Council of Chief Information Officers (CUCCIO).

Creating a new not-for-profit organization…
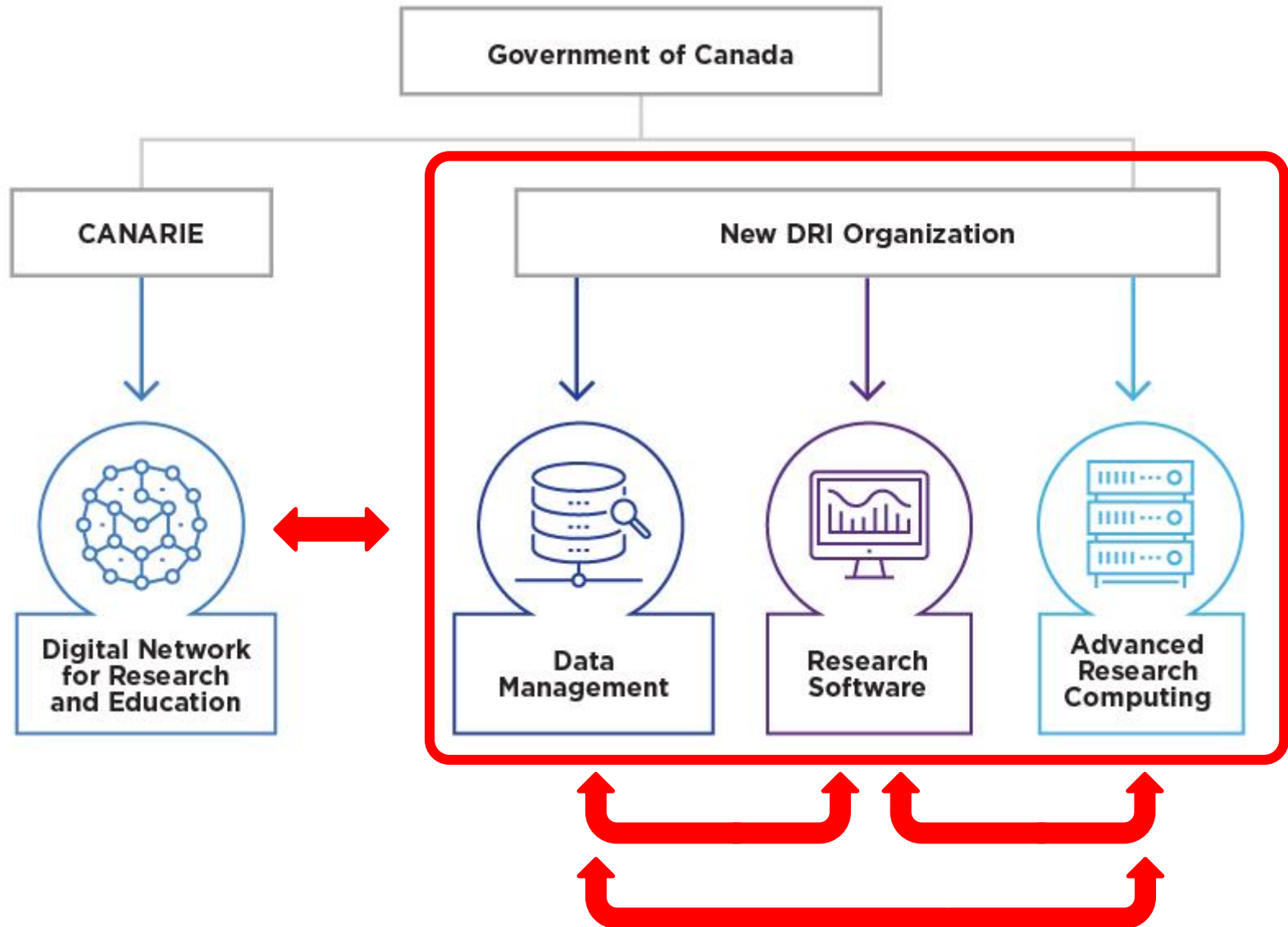
Converging streams of activity…
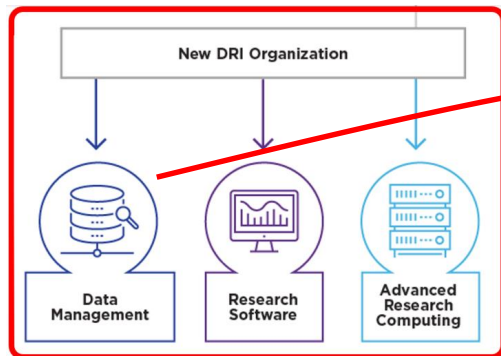
Forging a new Digital Research Strategy for Canada

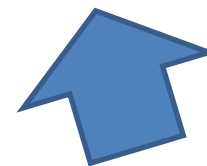Helping institutions and researchers address the challenges of RDM

$375 Million
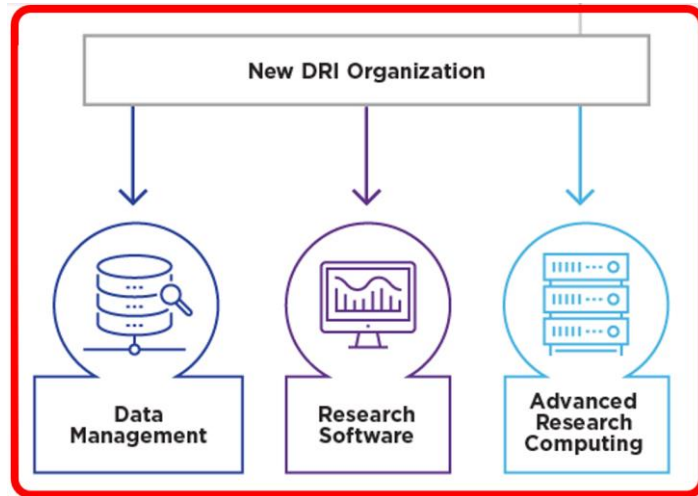
$2 Million
(1-year transition)

Innovation, Science, & Economic Development

All this was successfully launched ahead of the September 10 election call…

Progress toward the New DRI Organization…

New DRI Organization

Data Management — Research Software — Advanced Research Computing

Submitted May 6 ✓

Initial Signatories

Government Approval
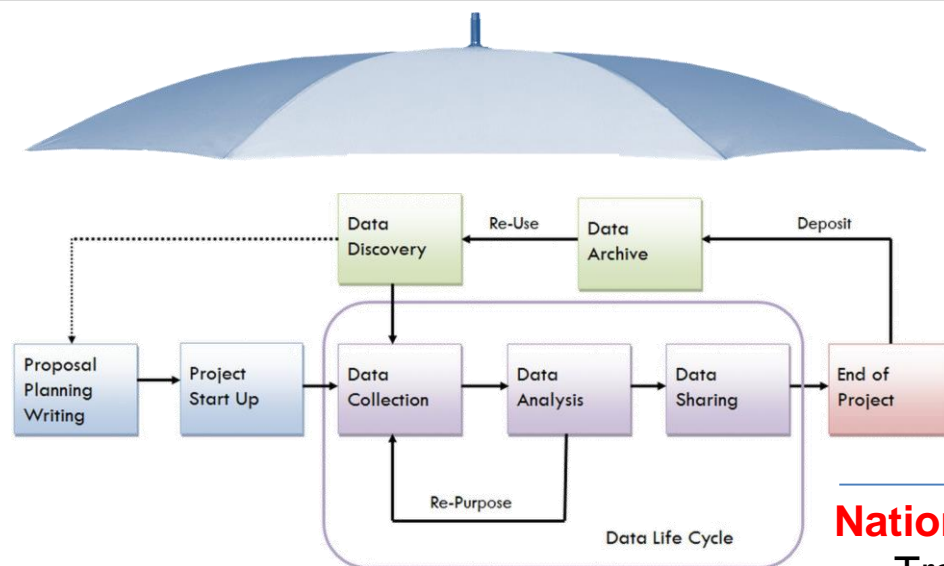
Announced Aug 8 ✓

Applicant Board

Inaugural Board

2021?

New DRI Organization

…and progress on the RDM front

Helping
Researchers achieve
FAIR Principles
throughout the
Research Life Cycle

**In partnership with our
Network of Experts and growing
Community of Practice**

CARL ABRC

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

Data Management

$2 Million
(1-year transition)

National Portage Secretariat Support:
- Training Coordinator,
- Curation Coordinator,
- Discovery and Metadata Coordinator,
- Preservation Coordinator
- Project Manager
- Communications Officer
- translation, website, travel, training & outreach

**Institutional partners
are also key to our
progress**

Institutional Partners

Institutional Partners

Project	Partner
Dataverse North	**OCUL** Ontario Council of University Libraries

- **National instance of Dataverse**

- **Improved Geospatial visualization & discovery**

CARL ABRC

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

Institutional Partners

Project

Dataverse North

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

Partner

OCUL Ontario Council of University Libraries

UNIVERSITY OF SASKATCHEWAN

UNIVERSITY OF WATERLOO

compute | calcul canada | canada

SFU

CARL ABRC

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

- **National instance of Dataverse**

- **Improved Geospatial visualization & discovery**

- **Use-case narratives for Dataverse, FRDR, & domain repositories**

- **Improving metadata for FRDR discovery layer**

- **Expanded metadata harvesting to expand discovery**

Institutional Partners

Project   **Partner**

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

- **DMP Exemplars for disciplines – opt-in to share**

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

Institutional Partners

Project   **Partner**

- **Glossary of Research Management Information (RMI) terms**

- **Interoperable DMP template(s)**

Institutional Partners

Project **Partner**
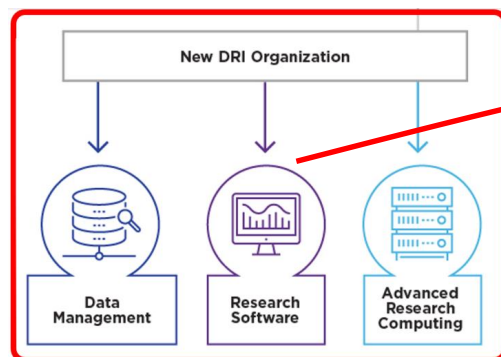
Portage is also working closely with RDC…
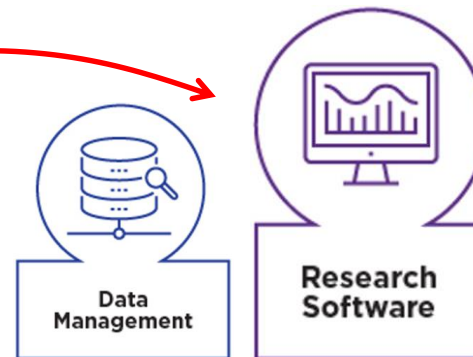
National Data Services
Framework Summit 2021

National Data Services
Framework 2019

National Data Services
Framework Summit 2017

Webinars

**…and collaborating
on other training
initiatives**

2018
CANARIE Launches Funding Call for Research Data Management
Software

Posted on May 8, 2018 in Press Releases

2019
RDM Competitive Funding Call 2

Evolution of Existing RDM Platforms, Repositories, and Services
for National and International Interoperability

canarie

RDC 🍁 DRC
Research Data Canada – Données de recherche Canada

Let's circle back to RDM and Curation

CDCF 2019 Participants

Portage Expert & Working Groups …and Community of Practice

Expert Groups

- Data Management Planning
- Curation
- Data Discovery
- Preservation
- Training
- Research Intelligence
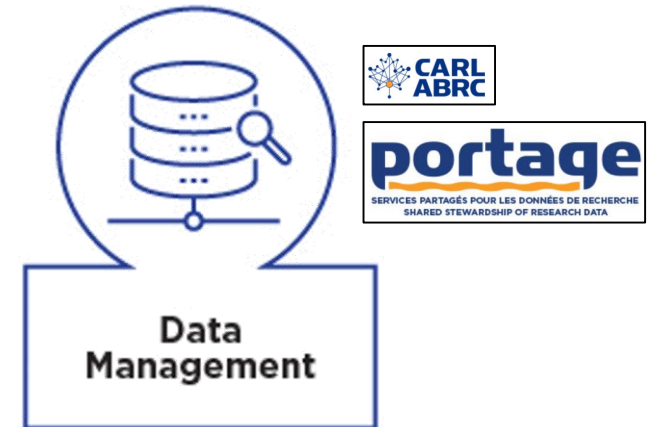- Data Repositories
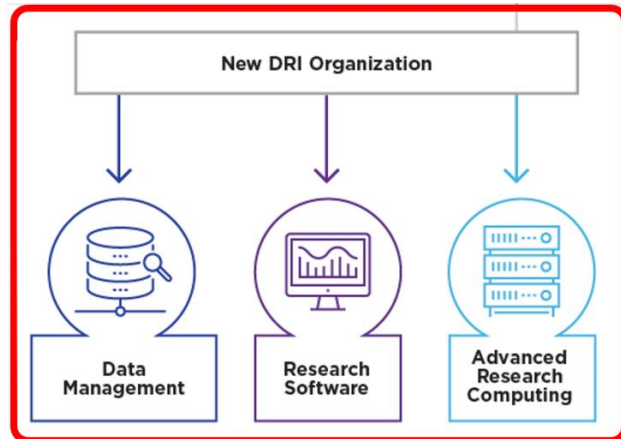
130+

Experts

60+

Organizations

Working Groups

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

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

Curation is baked into all we do…

Submitted May 6 ✓

Initial Signatories

+ Writing Group

Government Approval

Announced Aug 8 ✓

Applicant Board

+ Advisory Group

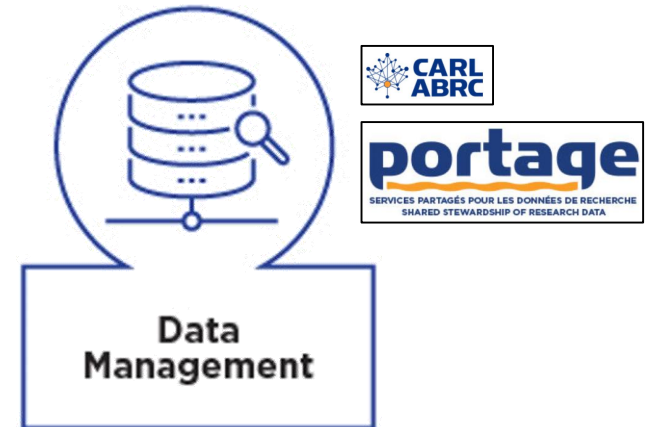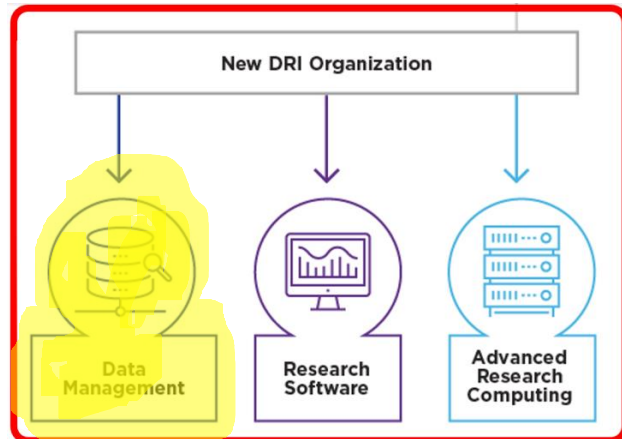## Inaugural Board

New DRI Organization

2021?

National Secretariat Support:
- Training Coordinator,
- Curation Coordinator,
- Discovery and Metadata Coordinator,
- Preservation Coordinator
- Project Manager
- Communications Officer
- translation, website, travel, training, and outreach

National Platform Support:
- FRDR → full production
- Dataverse North → augment storage
- DMP Assistant → migration to new merged codebase, and development of a service plan
- DMP Integration sub-project with CASRAI

Curation is baked into all we do…

New DRI Organization

Data Management · Research Software · Advanced Research Computing

Submitted May 6 ✓

Initial Signatories

+ **Writing Group**

Government Approval

Announced Aug 8 ✓

Applicant Board

+ **Advisory Group**

Inaugural Board

New DRI Organization

2021?

Data Management

CARL ABRC

portage — SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE / SHARED STEWARDSHIP OF RESEARCH DATA
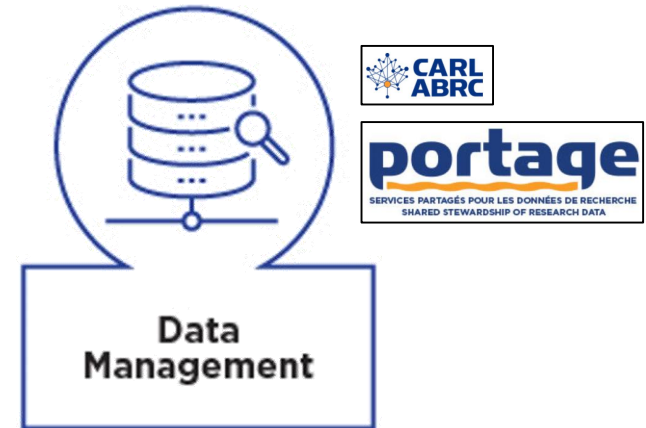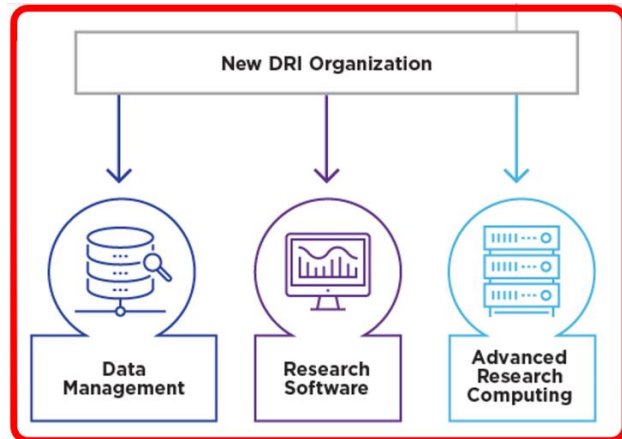
National Secretariat Support:
- Training Coordinator,
- Curation Coordinator,
- Discovery and Metadata Coordinator,
- Preservation Coordinator
- Project Manager
- Communications Officer
- translation, website, travel, training, and outreach

National Platform Support:
- FRDR → full production
- Dataverse North → augment storage
- DMP Assistant → migration to new merged codebase, and development of a service plan
- DMP Integration sub-project with CASRAI

Curation is baked into all we do…

New DRI Organization

Data Management · Research Software · Advanced Research Computing

Submitted May 6 ✓

Initial Signatories

+ Writing Group

Government Approval

Announced Aug 8 ✓

Applicant Board

+ Advisory Group

Inaugural Board

New DRI Organization

2021?

Data Management

CARL ABRC

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

National Secretariat Support:
- Training Coordinator,
- Curation Coordinator,
- Discovery and Metadata Coordinator,
- Preservation Coordinator
- Project Manager
- Communications Officer
- translation, website, travel, training, and outreach

National Platform Support:
- FRDR → full production
- Dataverse North → augment storage
- DMP Assistant → migration to new merged codebase, and development of a service plan
- DMP Integration sub-project with CASRAI

CDCF discussion & outputs

There's a lot going on with RDM…

… and lots more work to do!

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

Acknowledgements…

Innovation, Science, &
Economic Development

Questions?

portagenetwork.ca

Jeff Moon, Director, CARL Portage

portagedir@carl-abrc.ca