

# Data Deposit: Why & How?

James Doiron

RDM Services Coordinator, UofA Libraries  
Academic Director, UofA Research Data Centre  
Co-Chair, Portage DMP Expert Group

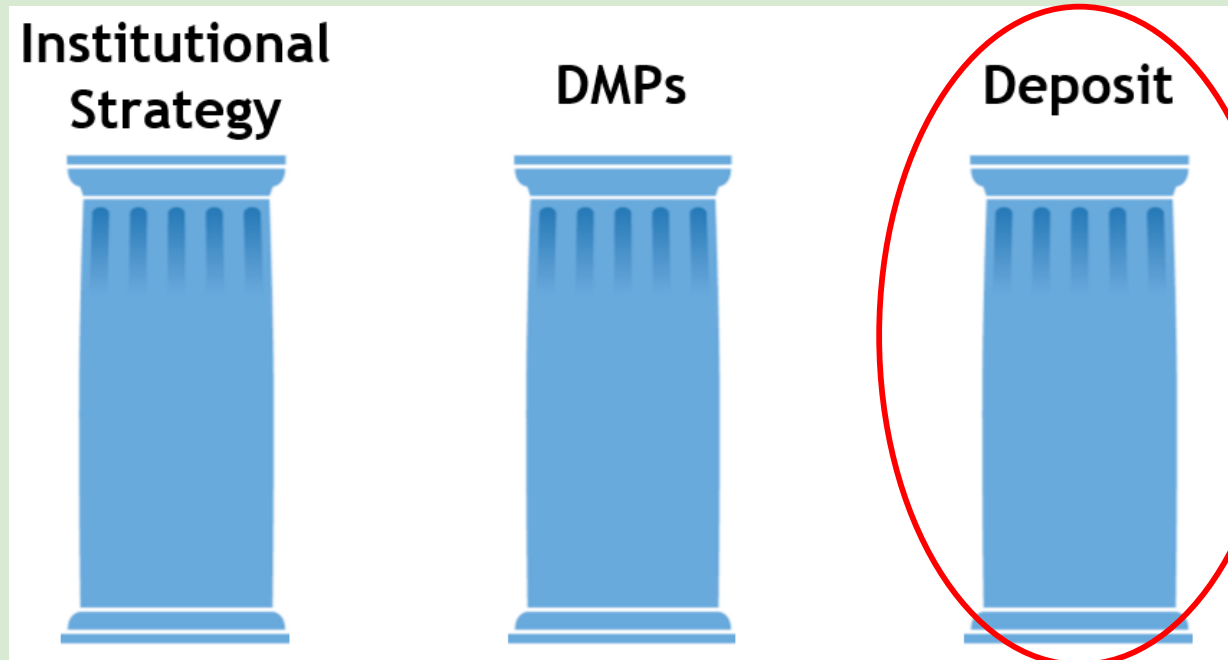
***Shifting Horizons II: Realities of RDM Services***

University of Ottawa  
February 20, 2020

# Tri-Agency RDM Policy (Draft) - 2018

- In the Spring of 2018, the Tri-Agencies released a draft of the “[Tri-Agency Research Data Management Policy](#)”
- Release of a final policy is expected in Spring 2020

The draft policy includes suggested requirements related to three primary areas:



# Tri-Agency RDM Draft Policy - “Take away points”

## Institutional Strategy:

*“Each institution administering tri-agency funds is required to create an institutional research data management strategy”*

## Data Management Plans:

*“...specific funding opportunities may require DMPs to be submitted to the appropriate agency at time of application*

## Data Deposit:

*“Grant recipients are required to deposit into a recognized digital repository all digital research data, metadata and code that directly support the research conclusions in journal publications, pre-prints, and other research outputs that arise from agency-supported research*

# What *is* a ‘research data repository’?

**A research data repository is a technology based platform that allows for research data to be:**

- Deposited & described
- Stored & archived
- Shared & published
- Discovered & reused

**There are different types of repositories including:**

- Proprietary (paid for services)
- Open source (free to use)
- Discipline specific

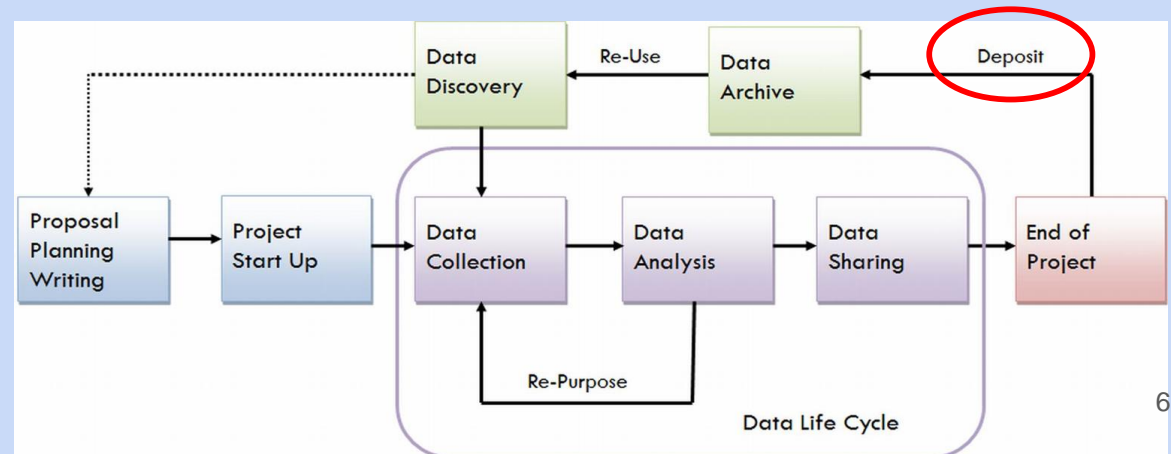
**RESOURCE!** → [Re3data.org](https://re3data.org) is an online registry of data repositories, which can be searched according to subject, content type and country. Find a list of [Canadian research data repositories](#).

# Data repository features

There are many different features that various data repositories offer, with some key ones being:

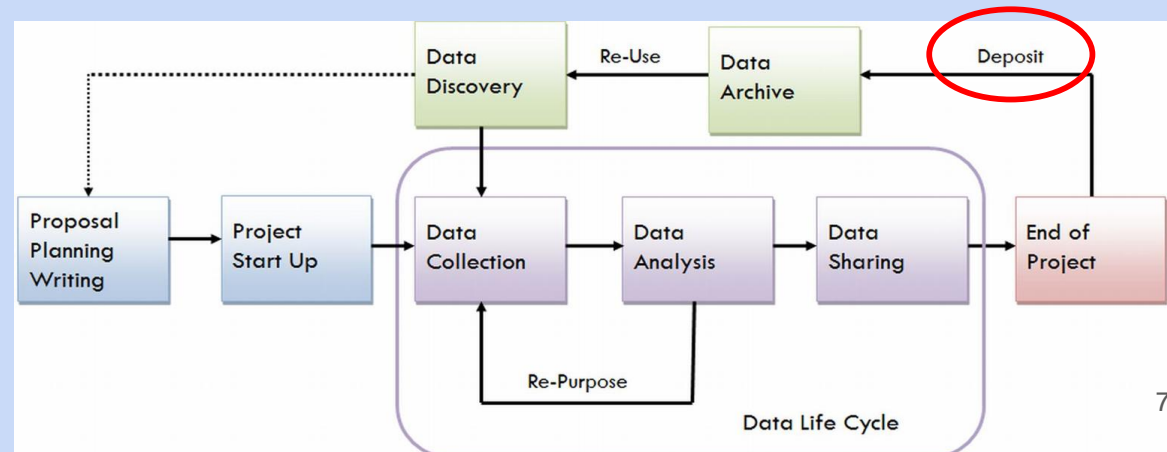
- Digital Object Identifiers (DOIs) → unique persistent identifier
- Generation & tracking of citations
- Data usage metrics
- File versioning
- Automated metadata generation
- Support for sensitive data, including embargo periods/restricted access
- Customized Terms of Use/Copyright agreements
- Preservation support

# Why should I consider depositing my data?



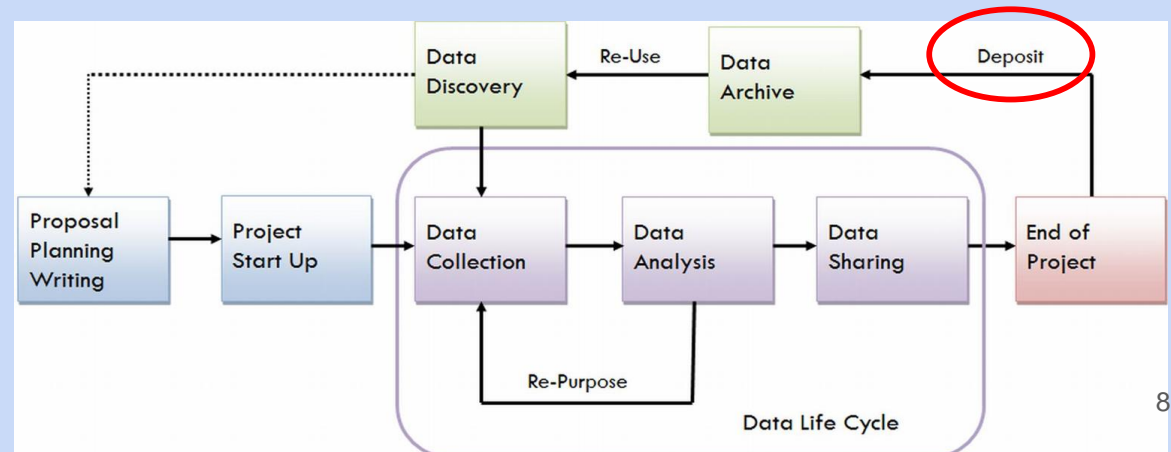
# Why should I consider depositing my data?

- Helps you to meet both funding and journal requirements



# Why should I consider depositing my data?

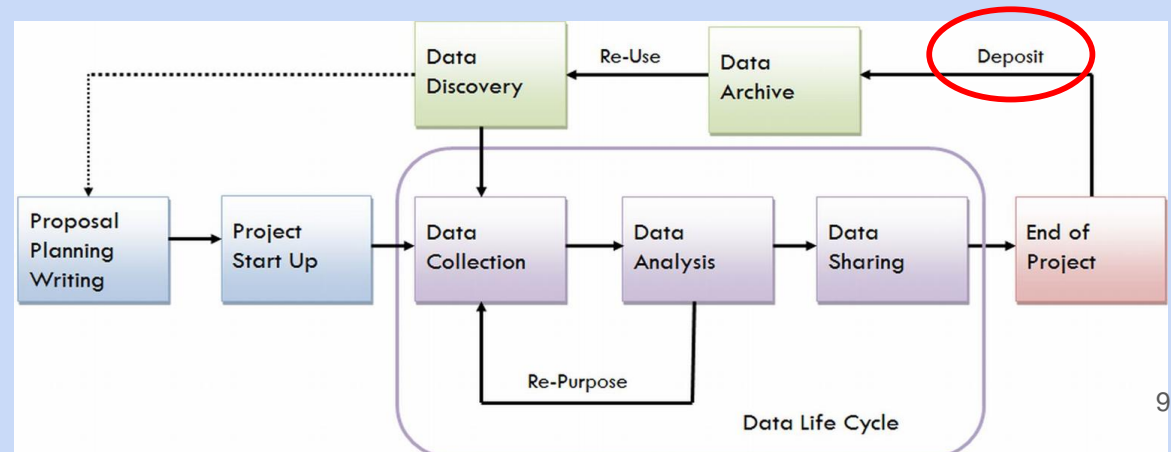
- Helps you to meet both funding and journal requirements
- Increases the impact and visibility of your research





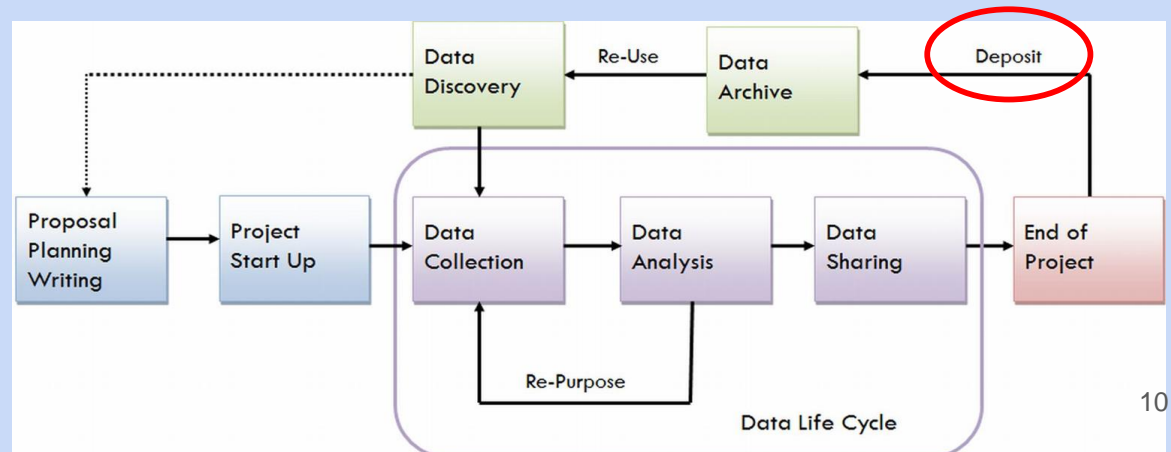
# Why should I consider depositing my data?

- Helps you to meet both funding and journal requirements
- Increases the impact and visibility of your research
- Digital Object Identifier (DOI) = your data are discoverable and citable



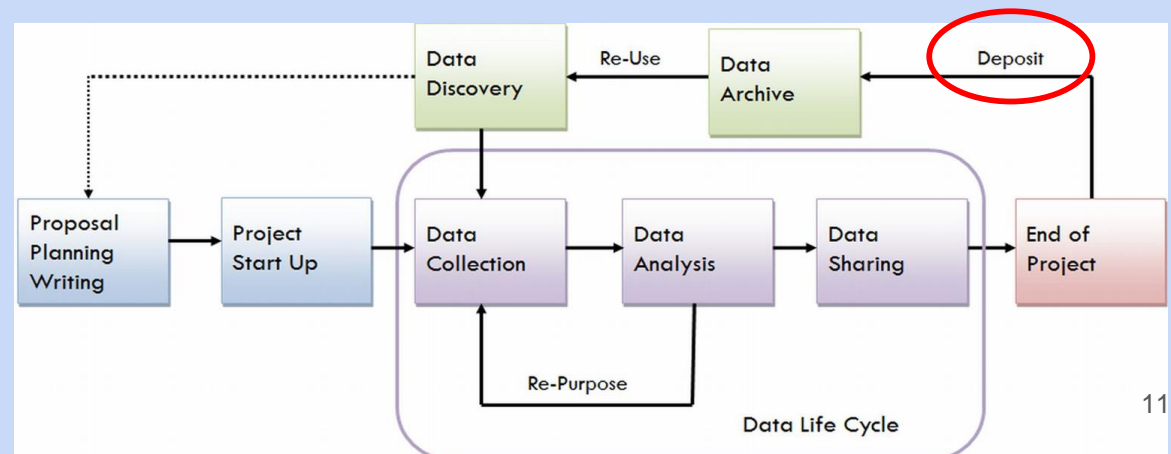
# Why should I consider depositing my data?

- Helps you to meet both funding and journal requirements
- Increases the impact and visibility of your research
- Digital Object Identifier (DOI) = your data are discoverable and citable
- Your data are made available, as appropriate, to others → supporting science



# Why should I consider depositing my data?

- Helps you to meet both funding and journal requirements
- Increases the impact and visibility of your research
- Digital Object Identifier (DOI) = your data are discoverable and citable
- Your data are made available, as appropriate, to others → supporting science
- Obtain metrics on *how* your data are being used



# What is a Digital Object Identifier?

# What is a Digital Object Identifier?

A **digital object identifier** (DOI) is a unique persistent identifier assigned by a registration agency to identify digital content and provide a persistent link to its location

# What is a Digital Object Identifier?

A **digital object identifier** (DOI) is a unique persistent identifier assigned by a registration agency to identify digital content and provide a persistent link to its location

DOIs can be assigned to a wide range of digital content, including both data and scholarly outputs (i.e., peer reviewed publications, research reports, posters, audio/visual materials, etc).

# What is a Digital Object Identifier?

A **digital object identifier** (DOI) is a unique persistent identifier assigned by a registration agency to identify digital content and provide a persistent link to its location

DOIs can be assigned to a wide range of digital content, including both data and scholarly outputs (i.e., peer reviewed publications, research reports, posters, audio/visual materials, etc).

DOI example:

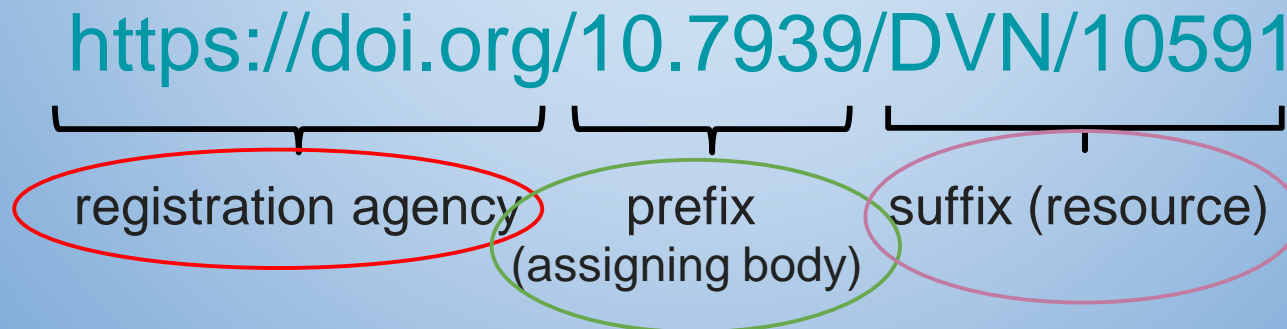
<https://doi.org/10.7939/DVN/10591>

# What is a Digital Object Identifier?

A **digital object identifier** (DOI) is a unique persistent identifier assigned by a registration agency to identify digital content and provide a persistent link to its location

DOIs can be assigned to a wide range of digital content, including both data and scholarly outputs (i.e., peer reviewed publications, research reports, posters, audio/visual materials, etc).

DOI example:





# What does a DOI do?

**DOIs** help data producers take control of the management of their research in that they:

# What does a DOI do?

**DOIs** help data producers take control of the management of their research in that they:

- Support the discovery & appropriate sharing of data

# What does a DOI do?

**DOIs** help data producers take control of the management of their research in that they:

- Support the discovery & appropriate sharing of data
- Support data producers in receiving credit for their data through data citations

# What does a DOI do?

**DOIs** help data producers take control of the management of their research in that they:

- Support the discovery & appropriate sharing of data
- Support data producers in receiving credit for their data through data citations
- Make research data easier to access, reuse and verify

# What does a DOI do?

**DOI's** help data producers take control of the management of their research in that they:

- Support the discovery & appropriate sharing of data
- Support data producers in receiving credit for their data through data citations
- Make research data easier to access, reuse and verify
- Help to meet funder requirements → data deposit and assignment of DOIs are becoming increasingly expected, or even required, in academia

# DOIs assignment

DOIs can be assigned to discrete *versions* of data:

- Full data
- Sample data
- Different years of data
- Tables generated from microdata
- Analytic files

# DOIs assignment

Outputs generated by any given version of data then appropriately cite *that* version

# DOIs assignment

Outputs generated by any given version of data then appropriately cite *that* version

Example: Jaywalking Activities Monthly Evaluation Survey



# DOIs assignment

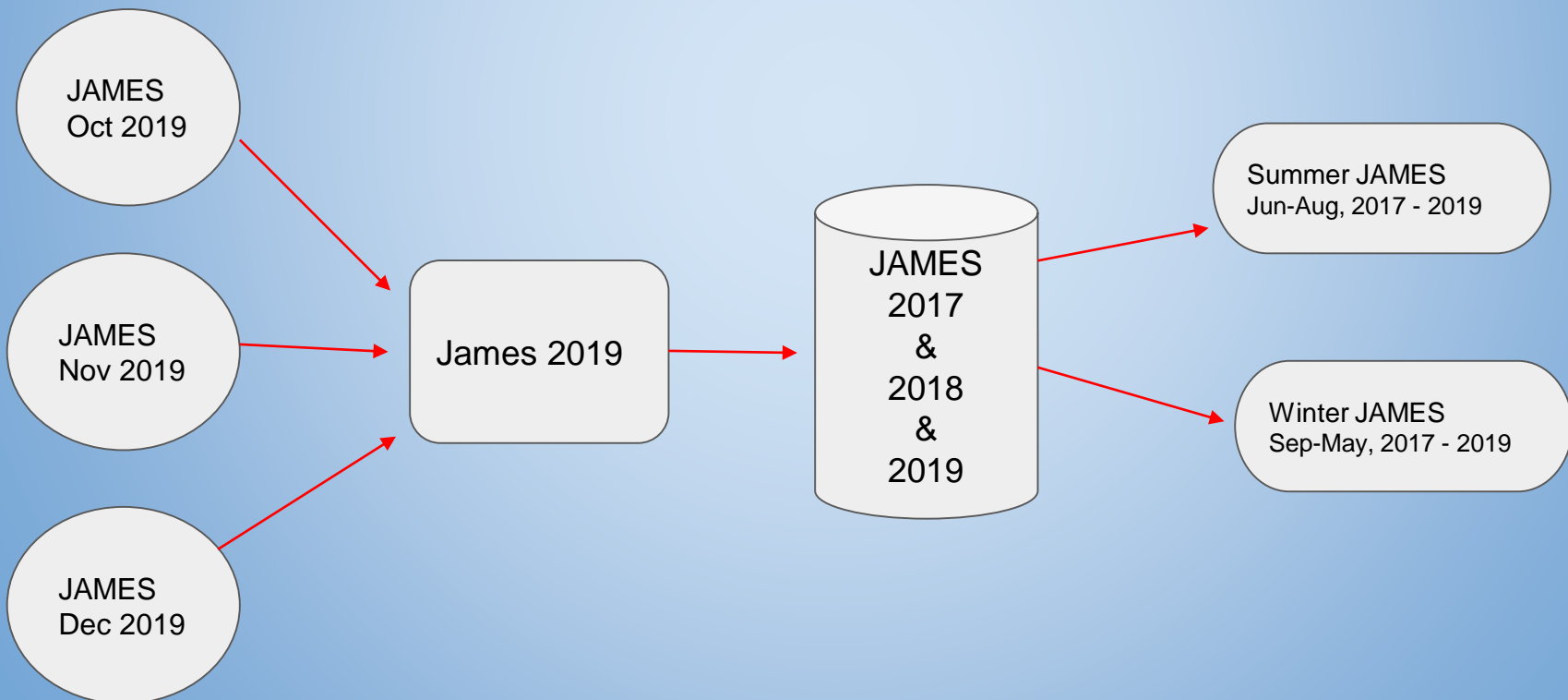
Outputs generated by any given version of data then appropriately cite *that* version

Example: Jaywalking Activities Monthly Evaluation Survey  
A.K.A. - “**JAMES**”

# DOIs assignment

Outputs generated by any given version of data then appropriately cite *that* version

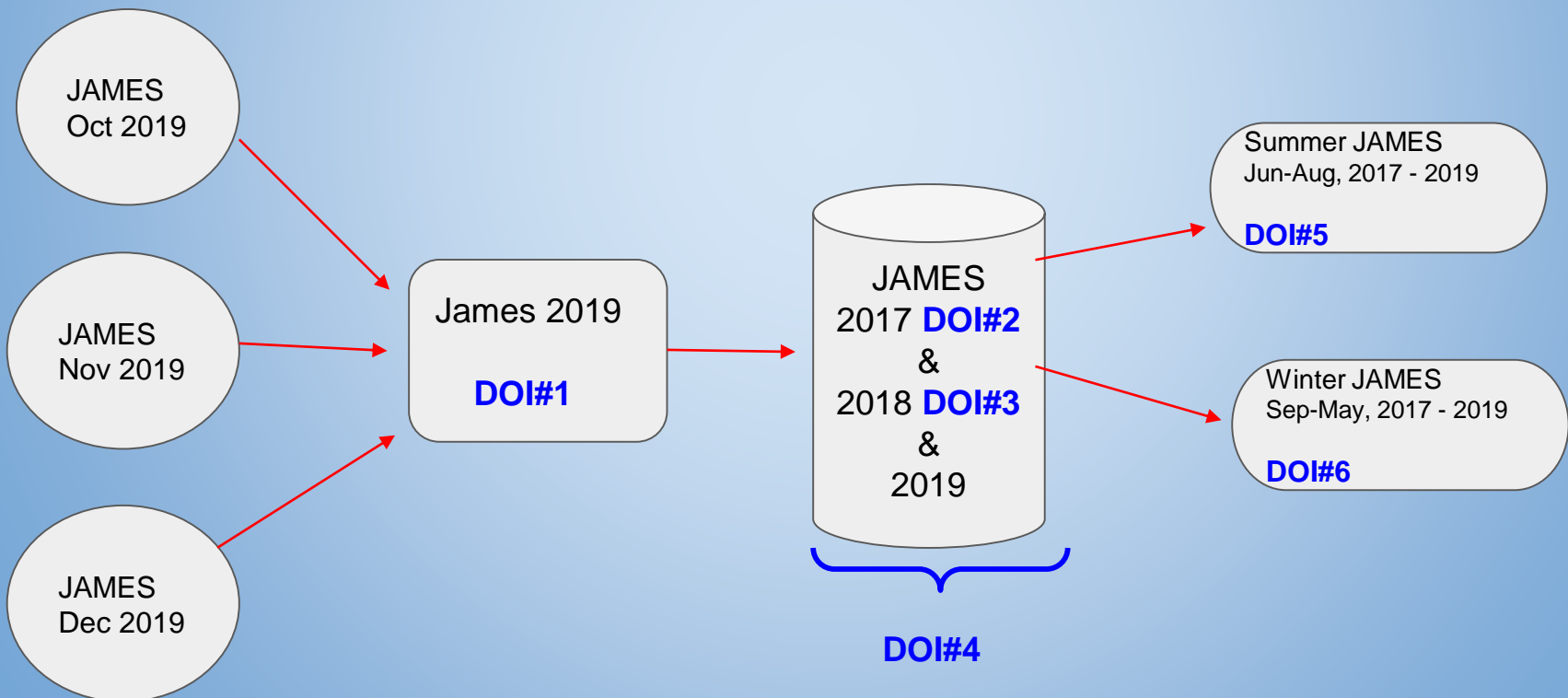
Example: Jaywalking Activities Monthly Evaluation Survey  
A.K.A. - “JAMES”



# DOIs assignment

Outputs generated by any given version of data then appropriately cite *that* version

Example: Jaywalking Activities Monthly Evaluation Survey  
A.K.A. - "JAMES"



# DOIs - supporting data usage metrics

Without DOIs our understanding of data usage and impact is extremely limited:

- # of downloads → does not tell us how data are used
- Possible citations, but not easily tracked

A DOI supports data usage metrics by providing:

- Persistent identifier to the data
- Ready-made citation which includes the DOI
- Bi-directional discovery → data to outputs and back to data
- Ability to 'scrape' the web using the DOI

# DOIs - discovery, citations & data usage

DOI example: <https://doi.org/10.7939/DVN/10591>

**Dataverse** Search ▾ User Guide Support Sign Up Log In

## Canadian Heart Health Survey

UAL Dataverse > Canadian Heart Health Survey > Canadian Heart Health Survey

Metrics 131 Downloads [Contact](#) [Share](#)

**Canadian Heart Health Survey** Version 2.0

Canadian Heart Health Database Centre, 2015, "Canadian Heart Health Survey", <https://doi.org/10.7939/DVN/10591>, UAL Dataverse, V2 [Cite Dataset ▾](#)

[Learn about Data Citation Standards.](#)

**Description**

The Canadian Heart Health Database (CHHDB) is an integration of data from ten Provincial Heart Health Surveys conducted between 1986 and 1992. The Provincial Heart Health Surveys were conducted as part of the Canadian Heart Health Initiative and have been a collaborative effort among provincial departments of health, Health Canada and The Heart and Stroke Foundation of Canada. The origin of the collaborative approach to cardiovascular disease prevention (CVD) lies in a report prepared by the Federal Provincial Working Group on Cardiovascular Disease Prevention and Control. The Canadian Heart Health Database consists of two sets of integrated data: core information collected by all ten provincial surveys, and family history information collected by only four provinces: Ontario, Quebec, Saskatchewan and Alberta. The objective of the program was to estimate at the provincial level the

Canadian Heart Health Database Centre, 2015, "Canadian Heart Health Survey", <https://doi.org/10.7939/DVN/10591>, UAL Dataverse, V2

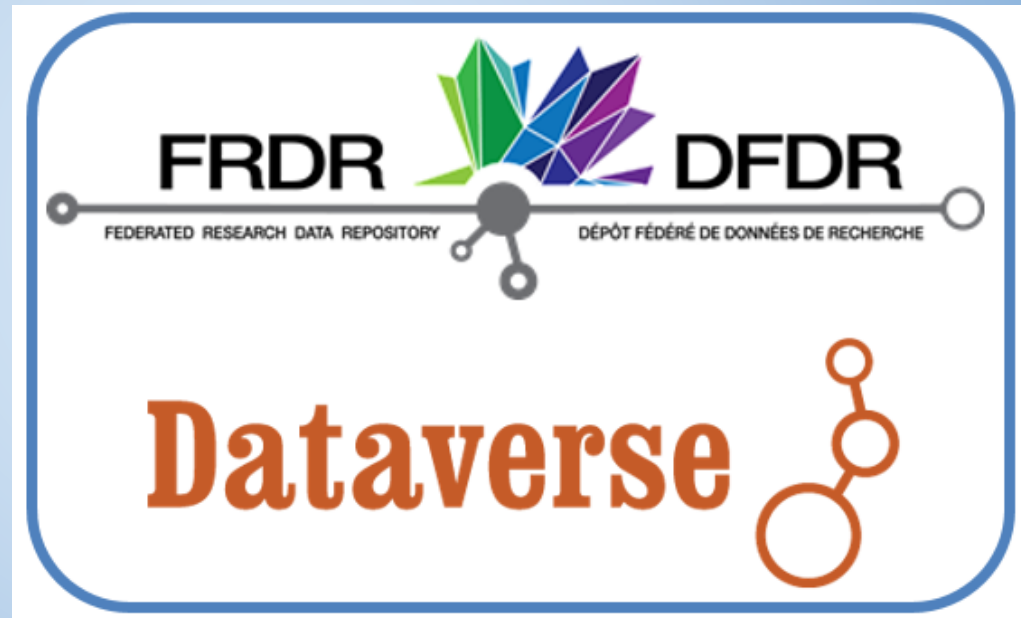
Files Metadata Terms Versions

Search this dataset... [Find](#)

1 to 10 of 14 Files

# Two Repository options:

- Federated Research Data Repository
- Dataverse





## Federated Research Data Repository

### **Main Features:**

- ‘Big Data’ upload/download
- Maintains file hierarchies
- National **Discovery** Platform
- National **Repository** Option
- Preservation** Processing capacity
- Scalable solution

### **Strategic Relevance**

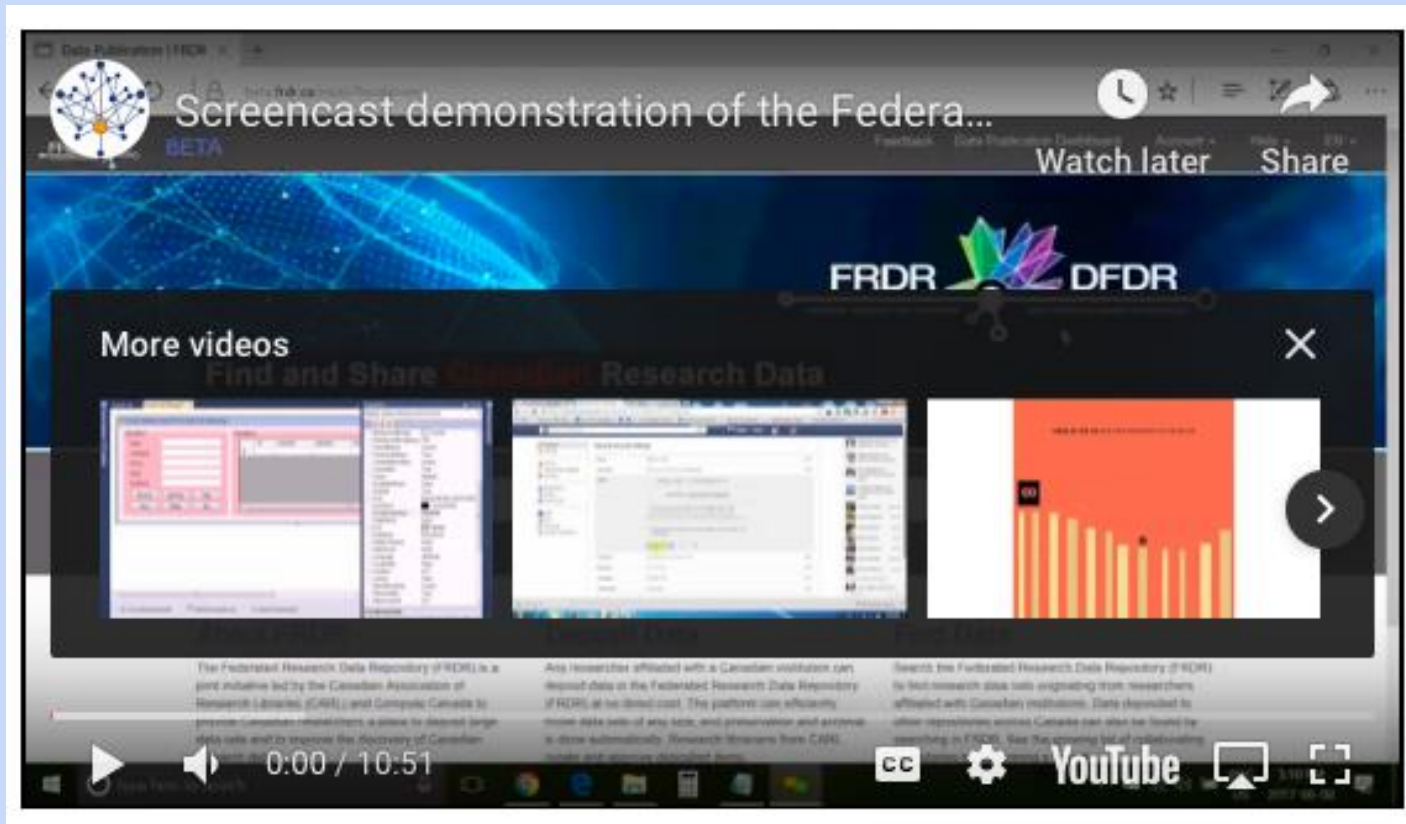
Fits into broader Portage vision for  
**Preservation**

**Strategic partnership: Portage,  
Compute Canada, and CARL**

Fills key gaps in the RDM ecosystem:

**Single platform from which research data can be ingested, curated,  
preserved, discovered, cited & shared**

## Demonstration Video



[www.frdr.ca](http://www.frdr.ca)





## Demonstration version

A screenshot of the FRDR/DFDR demo website. The page has a dark header with the FRDR/DFDR logo on the left, the word 'DEMO' in the center, and navigation links for 'Feedback', 'Log In', 'Help', 'About', and 'EN' on the right. The main content area features a blue background with a network of glowing nodes and lines. A large white box with a black border contains the text 'Find and Share Canadian Research Data'. Below this is a search bar with the placeholder text 'Search' and a magnifying glass icon, followed by a 'Deposit Data' button with a cloud icon. The page is divided into two columns: 'Find Data' and 'Deposit Data'. The 'Find Data' section includes a paragraph about searching for research datasets and a 'Learn more' link. The 'Deposit Data' section includes a paragraph about depositing data into FRDR and a 'Learn more' link.

FRDR DEMO Feedback Log In Help About EN

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

Find and Share **Canadian** Research Data

Search  Deposit Data 

Advanced search

### Find Data

Search FRDR to find research datasets originating from researchers affiliated with Canadian institutions. Data deposited to other repositories across Canada can also be found by searching in FRDR. View the growing [list of collaborating repositories](#).

[Learn more](#)

### Deposit Data

Any researcher affiliated with a Canadian institution can deposit data into FRDR. The platform can efficiently ingest datasets of any size, and preservation processing is done automatically. Data professionals from the Portage Network and institutions across Canada work with researchers to curate and approve deposited items.

[Learn more](#)

<https://demo.frdr-dfdr.ca/repo/>

## **Main Features:**

**Mature platform**

**Open source**

**Widely adopted**

**Bilingual**

**User-friendly**

**Brandable**

**File versioning**

**Supports many data types**

**International community support**

## **Strategic Relevance**

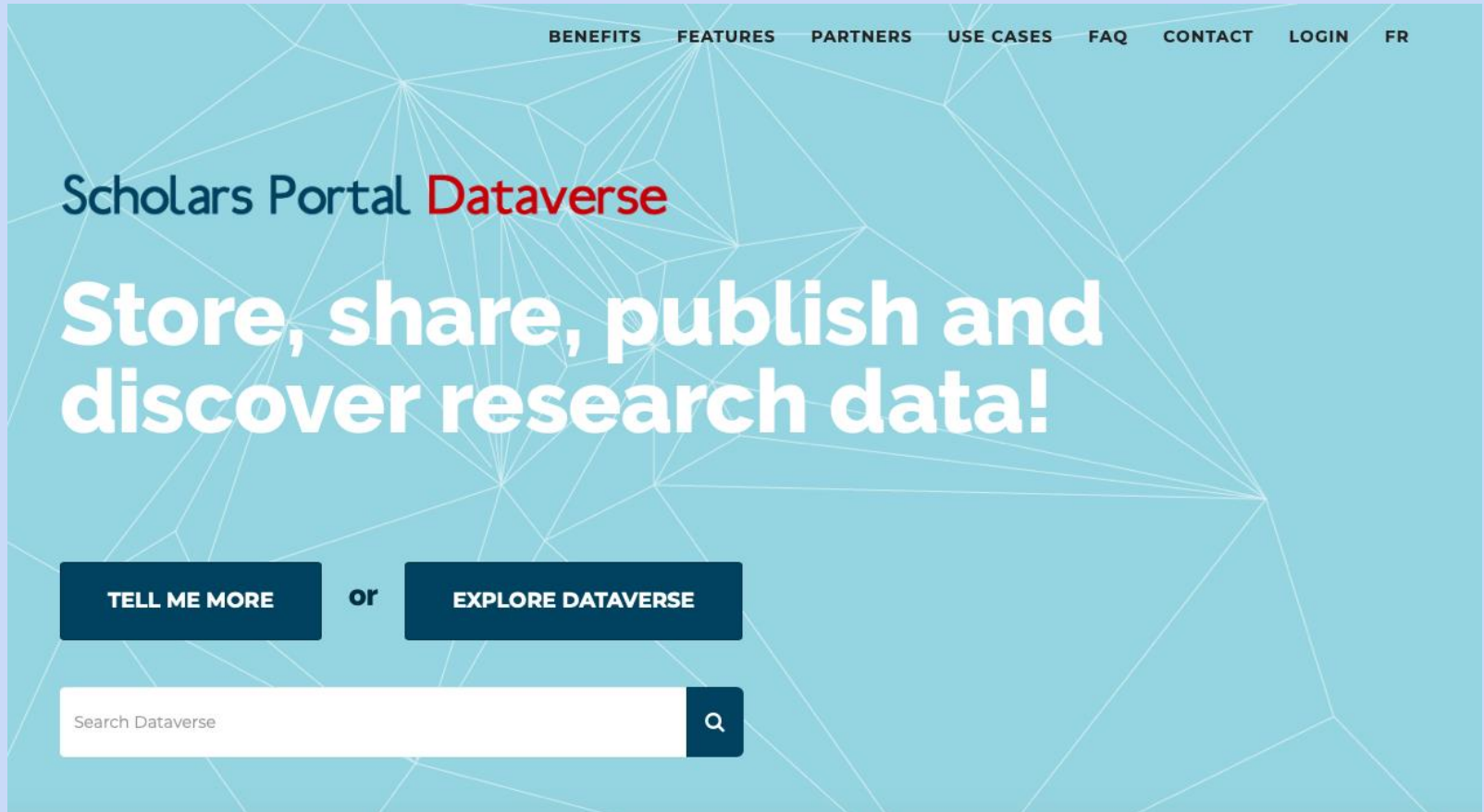
**Emerging National Repository**

**Option: “Dataverse North”**

**Cost model built to support  
institutions of any size**

**Agreements currently in place both  
with BCI & COPPUL**

## Scholars Portal Dataverse




The screenshot shows the homepage of the Scholars Portal Dataverse. At the top, there is a navigation menu with links for BENEFITS, FEATURES, PARTNERS, USE CASES, FAQ, CONTACT, LOGIN, and FR. The main heading reads "Scholars Portal Dataverse" in a dark blue font, with "Dataverse" in red. Below this, a large white text block states "Store, share, publish and discover research data!". At the bottom of the main content area, there are two dark blue buttons: "TELL ME MORE" and "EXPLORE DATAVERSE", separated by the word "or". Below the buttons is a search bar with the placeholder text "Search Dataverse" and a magnifying glass icon.

<https://dataverse.scholarsportal.info>

## UOttawa Dataverse

Scholars Portal **Dataverse** Search ▾ User Guide Support English ▾ Log In

 **uOttawa**

[uOttawa - Dataverse](#) (Université d'Ottawa / University of Ottawa)

[Scholars Portal Dataverse](#) > **uOttawa - Dataverse**

[Contact](#) [Share](#)

Search this dataverse...  [Advanced Search](#)

[Dataverses \(9\)](#)  
 [Datasets \(22\)](#)  
 [Files \(140\)](#)


**Dataverse Category**  
[Organization or Institution \(1\)](#)  
[Research Group \(1\)](#)  
[Research Project \(1\)](#)  
[Researcher \(1\)](#)

**Publication Year**  
[2019 \(8\)](#)  
[2014 \(7\)](#)  
[2017 \(5\)](#)  
[2013 \(3\)](#)  
[2015 \(3\)](#)

**1 to 10 of 31 Results**

**Dramatic Growth of Open Access Dec. 31, 2019**


Jan 3, 2020 - [Dramatic Growth of Open Access Dataverse](#)

 Morrison, Heather, 2020, "Dramatic Growth of Open Access Dec. 31, 2019", <https://doi.org/10.5683/SP2/CHLOKU>, Scholars Portal Dataverse, V1

Quarterly growth statistics for macro-level open access indicators from 2004 - 2019 in excel and csv format.

**OA APC - Subject Analysis – Statistic Frequency & Coding**

Dec 18, 2019 - [Open access article processing charges Dataverse](#)

 Morrison, Heather; Villamizar, César; Mondésir, Guinsly; Calvé-Genest, Alexis, 2019, "OA APC - Subject Analysis – Statistic Frequency & Coding", <https://doi.org/10.5683/SP2/4WKTZF>, Scholars Portal Dataverse, V1, UNF:6:7jt5ajlhjqik0HosFpQkA== [fileUNF]

Data for 2014 OA APC subject analysis

<https://dataverse.scholarsportal.info/dataverse/ottawa>



# Questions & Discussion

James Doiron

RDM Services Coordinator, UofA Libraries  
Academic Director, UofA Research Data Centre  
Co-Chair, Portage DMP Expert Group

[jdoiron@ualberta.ca](mailto:jdoiron@ualberta.ca)