

Research Data Management and Historians

12/14/2022

Ian Milligan

Associate Professor of History &
Associate Vice-President, Research Oversight & Analysis

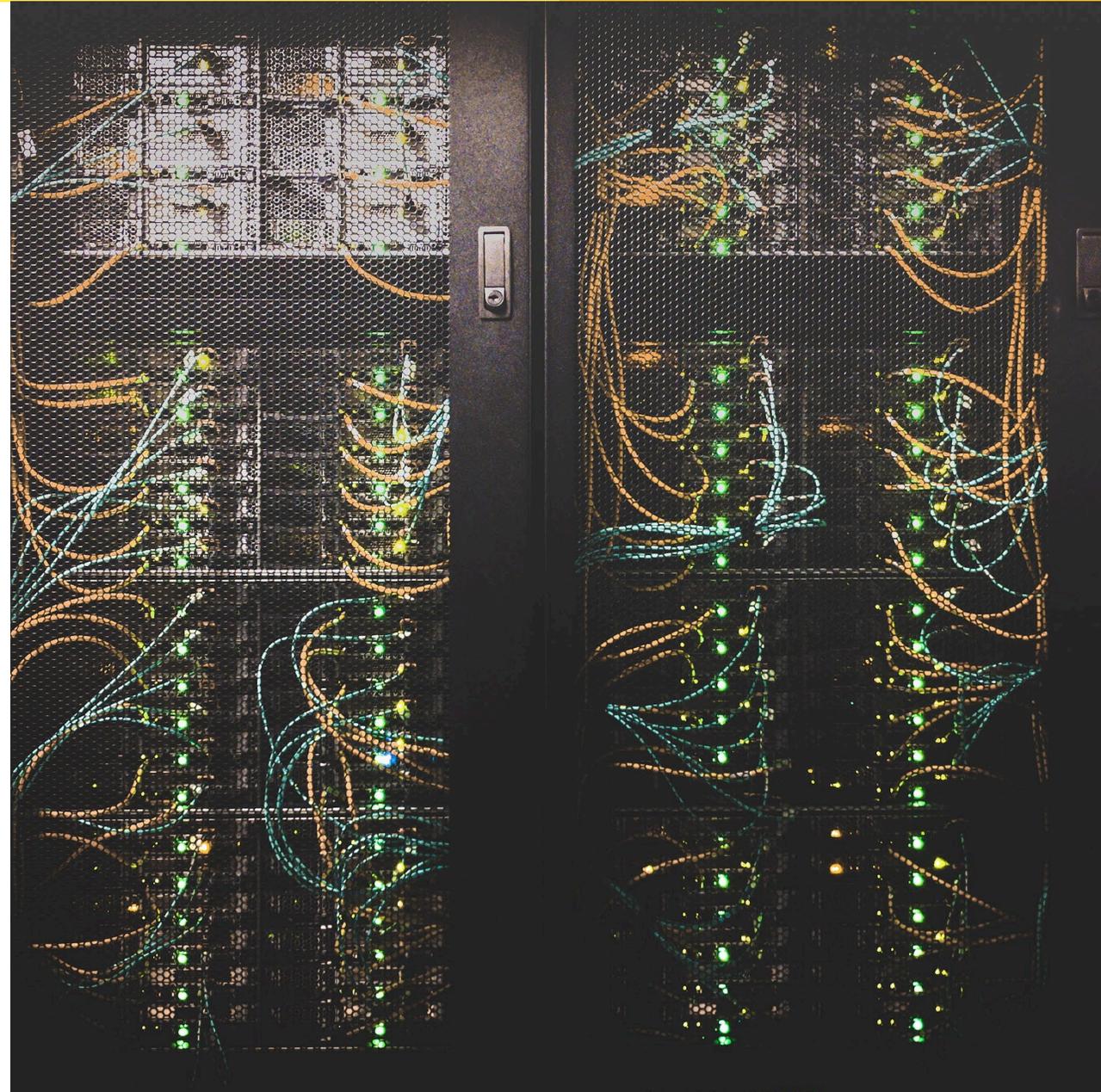


Photo by [Taylor Vick](#) on [Unsplash](#)

Outline for Today's Talk

- Why should you care about data?
- What's going on in Canada?
- Types of data
- How to share
- What does the future hold for RDM and historians?
- Pourquoi devriez-vous vous soucier des données ?
- Que se passe-t-il au Canada?
- Type de données
- Comment partager
- Quel avenir pour RDM et les historiens ?

LET'S BEGIN WITH TWO VIGNETTES:

A HISTORIAN IN 2001 VS 2021

COMMENÇONS PAR DEUX SCENARIOS:

UN HISTORIEN EN 2001 VS 2021

In 2001 / En 2001

- Choose a Project
 - Explore the secondary literature, find a historiographical niche
 - Find primary documents
 - Travel to Archive
 - In-Person Research for Months
 - Complement by Months in the Microfilm room
- Choisissez un projet
 - Découvrir la littérature
 - Trouver des sources primaires
 - Se rendre aux archives
 - Recherche en personne
 - Microfilm

Research was labourious, characterized by information scarcity.

La recherche était laborieuse, caractérisée par la rareté de l'information.



**HISTORIANS COLLECTED THEIR OWN DATA (OR
THOUGHT THEY DID)**

**LES HISTORIENS ONT COLLECTÉ LEURS
PROPRES DONNÉES (OU PENSAIENT L'AVOIR
FAIT)**

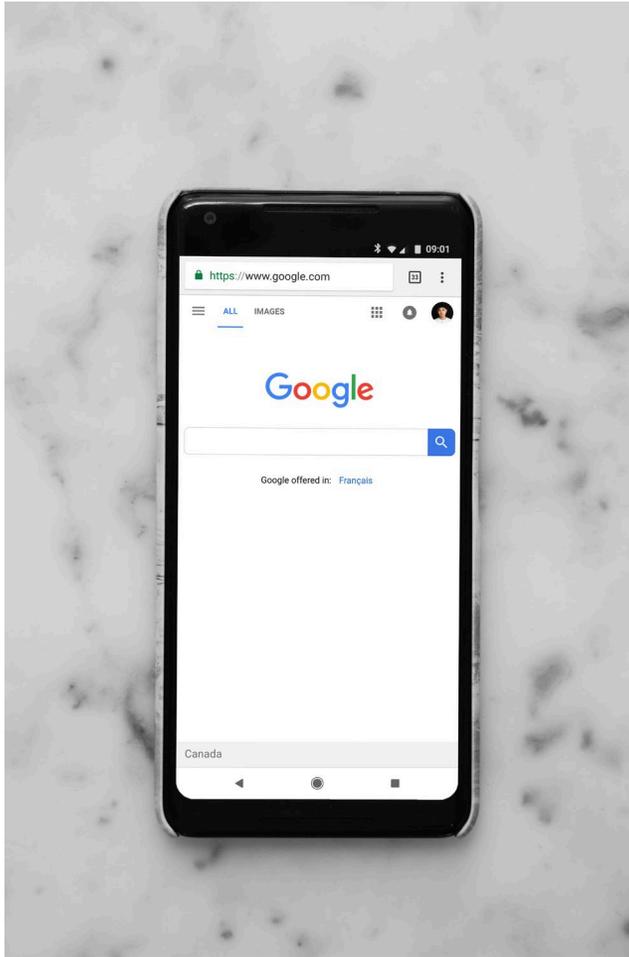
Flash forward 20 years / En avant 20 ans.



In 2021/En 2021

- **Find archival sources by**
 - Googling
 - Visit websites
 - Consider what is and what is not online
 - Travel to archives and take thousands of photographs to read at your home
- **Trouver des sources d'archives par**
 - Recherche sur Google
 - Visiter des sites Web
 - Considérez ce qui est et ce qui n'est pas en ligne
 - Voyagez dans les archives et prenez des milliers de photos à lire chez vous

Same with other primary sources/La même chose avec d'autres sources primaires



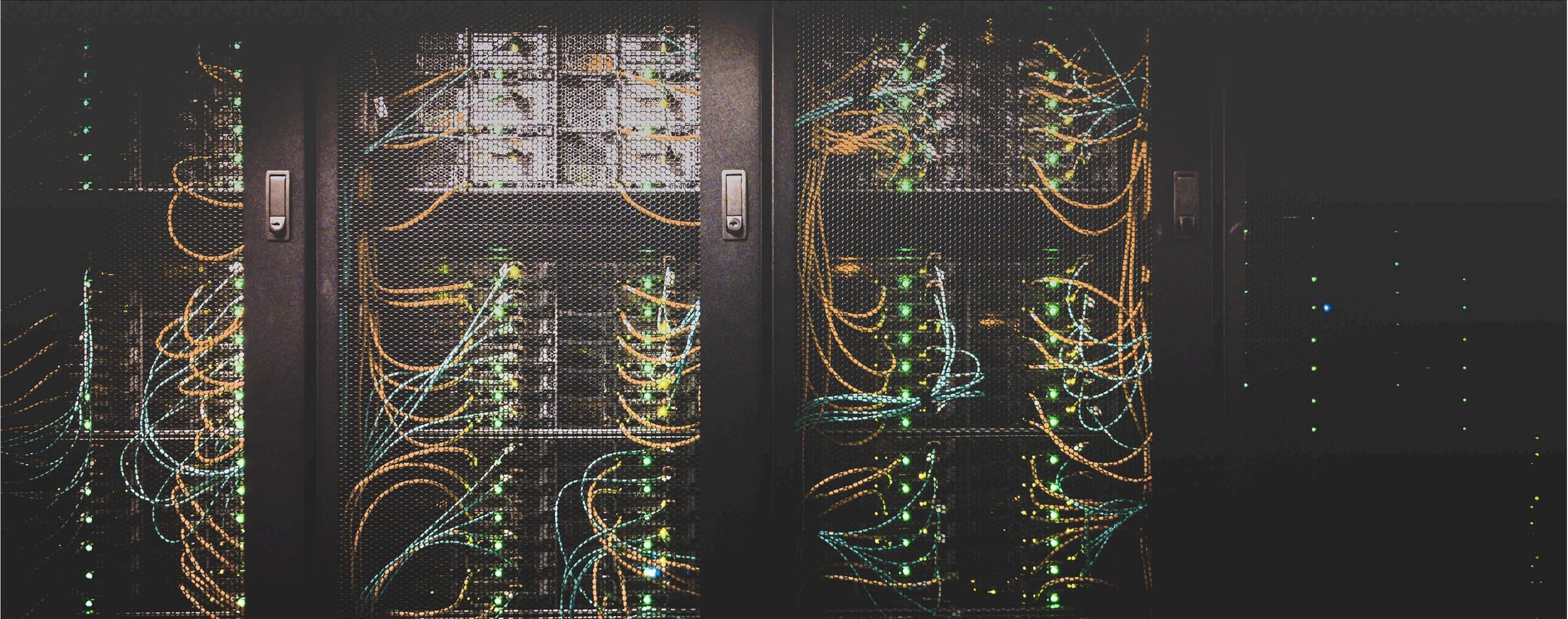
- **Historians approach sources through:**
 - Search
 - Keyword search rather than expert indexes or reading
 - Consulting digitized rather than the most relevant
- **Les historiens étudient les sources à travers**
 - rechercher
 - recherche par mot clé non indexée ou par lecture
 - la consultation numérique n'est pas la plus pertinente

**(AND, IN A FEW YEARS, THEY WILL
HOPEFULLY USE THE DATA OF OTHERS TO
AVOID DUPLICATION)**

**(ET, DANS QUELQUES ANNÉES, ILS
UTILISERONT, ESPÉRONS-LE, LES DONNÉES
DES AUTRES POUR ÉVITER LES DOUBLONS)**

The “Matthew Effect” of Historical Research.

L'« effet Matthieu » de la recherche historique.



The research approach of a historian has dramatically transformed.

L'approche de recherche d'un historien s'est radicalement transformée.



All parts of a historians work has changed in 20 years.

Toutes les parties du travail d'un historien ont changé en 20 ans.



And it is thanks to data!

et c'est grâce aux données numériques !



**ALL OF THIS CAN ACCELERATE THANKS TO
FEDERAL POLICY.**

**TOUT CELA PEUT S'ACCÉLÉRER GRÂCE À LA
POLITIQUE FÉDÉRALE.**

WHAT'S HAPPENING IN CANADA? QUE SE PASSE-T-IL AU CANADA?

SSHRC + Data / CRSH + Données

- Research data includes quantitative social, political and economic data sets; qualitative information in digital format; experimental research data; still and moving image and sound databases; and other digital objects used for analytical purposes.
- Ces données comprennent des ensembles quantitatifs de données sociales, politiques et économiques, des renseignements qualitatifs sous forme numérique, des données de recherche expérimentale, des bases de données d'images et de sons fixes et mobiles, ainsi que d'autres objets numériques utilisés à des fins d'examen analytique.



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

Canada 

SSHRC + Data / CRSH + données

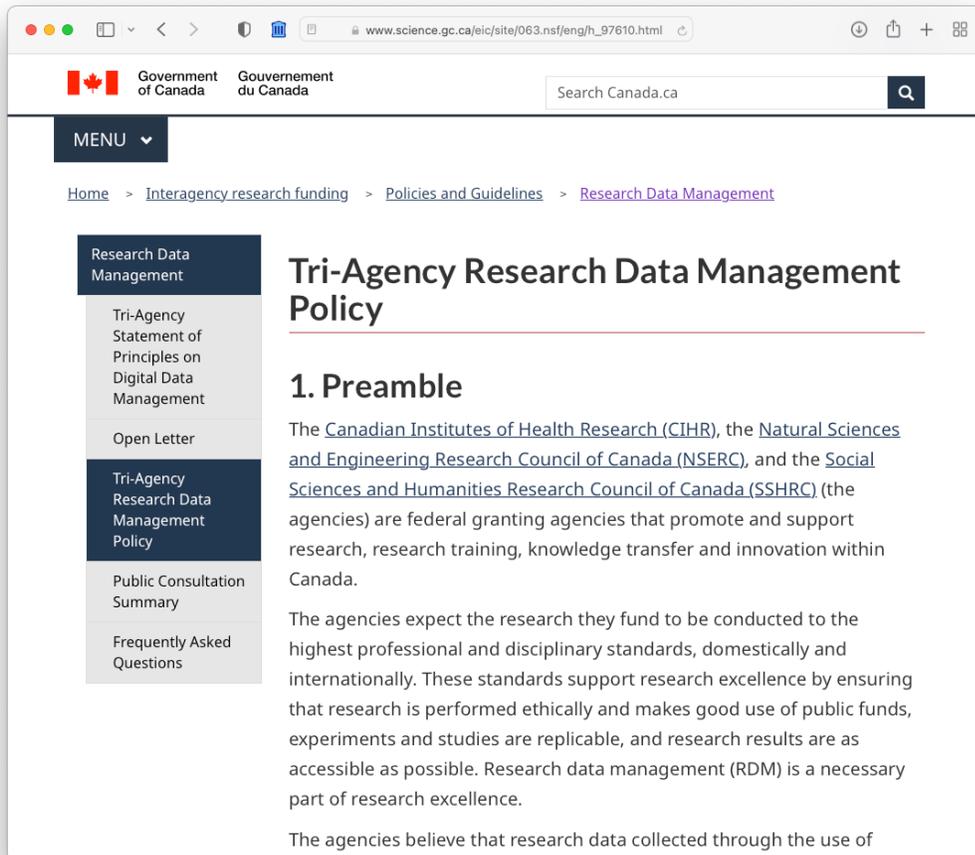
- Such data come in many forms:
 - Oral history transcripts/recordings
 - Spreadsheets of historical information
 - GIS data
- And some ambiguity
 - Machine learning models?
 - Archival photographs?
- Ces données se présentent sous plusieurs formes :
 - Transcriptions/enregistrements d'histoires orales
 - Feuilles de calcul d'informations historiques
 - Données SIG
- Et une certaine ambiguïté
 - Modèles d'apprentissage automatique ?
 - Des photos d'archives ?

**MOST HISTORIANS *S T I L L* DIDN'T THINK
ABOUT DATA...**

**LA PLUPART DES HISTORIENS NE PENSAIENT
TOUJOURS PAS AUX DONNÉES...**

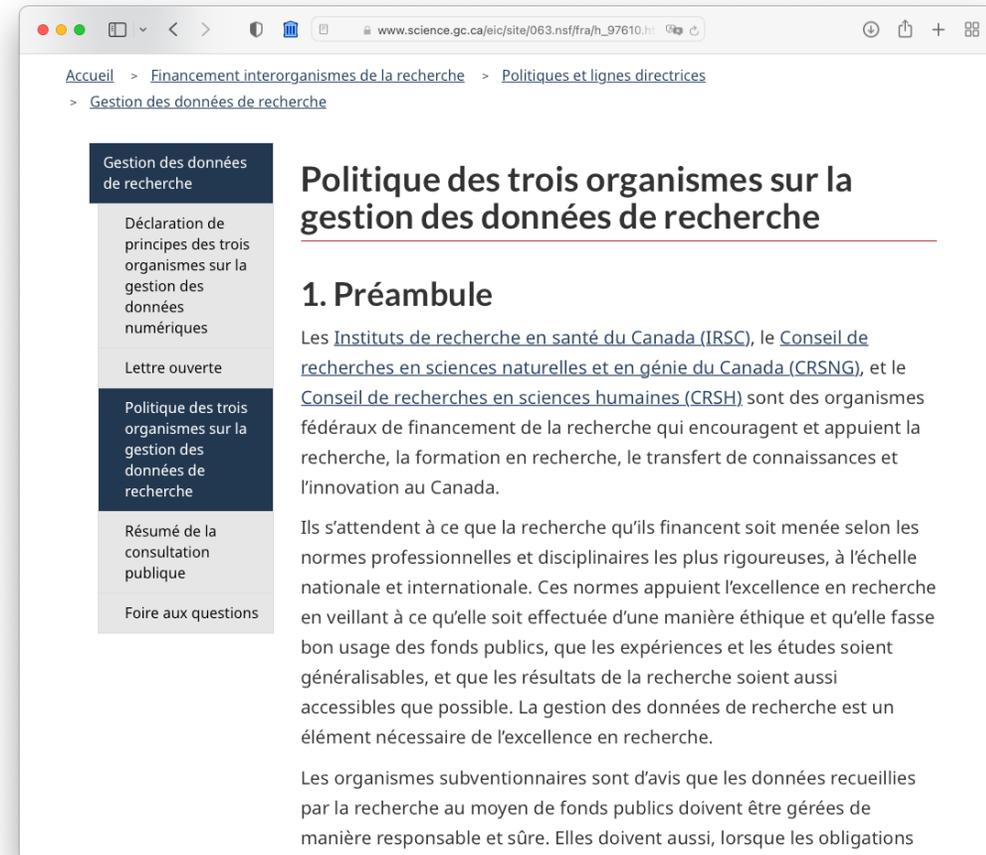
Tri-Agency Research Data Management Policy

Politique des trois organismes sur la gestion des données de recherche



The screenshot shows the Government of Canada website with the following structure:

- Header: Government of Canada / Gouvernement du Canada, Search Canada.ca
- Navigation: Home > Interagency research funding > Policies and Guidelines > Research Data Management
- Left sidebar menu:
 - Research Data Management
 - Tri-Agency Statement of Principles on Digital Data Management
 - Open Letter
 - Tri-Agency Research Data Management Policy**
 - Public Consultation Summary
 - Frequently Asked Questions
- Main content:
 - ## Tri-Agency Research Data Management Policy
 - ### 1. Preamble
 - The [Canadian Institutes of Health Research \(CIHR\)](#), the [Natural Sciences and Engineering Research Council of Canada \(NSERC\)](#), and the [Social Sciences and Humanities Research Council of Canada \(SSHRC\)](#) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada.
 - The agencies expect the research they fund to be conducted to the highest professional and disciplinary standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible. Research data management (RDM) is a necessary part of research excellence.
 - The agencies believe that research data collected through the use of



The screenshot shows the content of the policy page with the following structure:

- Header: Accueil > Financement interorganismes de la recherche > Politiques et lignes directrices > Gestion des données de recherche
- Left sidebar menu:
 - Gestion des données de recherche
 - Déclaration de principes des trois organismes sur la gestion des données numériques
 - Lettre ouverte
 - Politique des trois organismes sur la gestion des données de recherche**
 - Résumé de la consultation publique
 - Foire aux questions
- Main content:
 - ## Politique des trois organismes sur la gestion des données de recherche
 - ### 1. Préambule
 - Les [Instituts de recherche en santé du Canada \(IRSC\)](#), le [Conseil de recherches en sciences naturelles et en génie du Canada \(CRSNG\)](#), et le [Conseil de recherches en sciences humaines \(CRSH\)](#) sont des organismes fédéraux de financement de la recherche qui encouragent et appuient la recherche, la formation en recherche, le transfert de connaissances et l'innovation au Canada.
 - Ils s'attendent à ce que la recherche qu'ils financent soit menée selon les normes professionnelles et disciplinaires les plus rigoureuses, à l'échelle nationale et internationale. Ces normes appuient l'excellence en recherche en veillant à ce qu'elle soit effectuée d'une manière éthique et qu'elle fasse bon usage des fonds publics, que les expériences et les études soient généralisables, et que les résultats de la recherche soient aussi accessibles que possible. La gestion des données de recherche est un élément nécessaire de l'excellence en recherche.
 - Les organismes subventionnaires sont d'avis que les données recueillies par la recherche au moyen de fonds publics doivent être gérées de manière responsable et sûre. Elles doivent aussi, lorsque les obligations

RDM: A Definition

- *Research data management (RDM) refers to the processes applied through the lifecycle of a research project to guide the collection, documentation, storage, sharing and preservation of research data.*
- -Tri Agency FAQ on RDM Policy

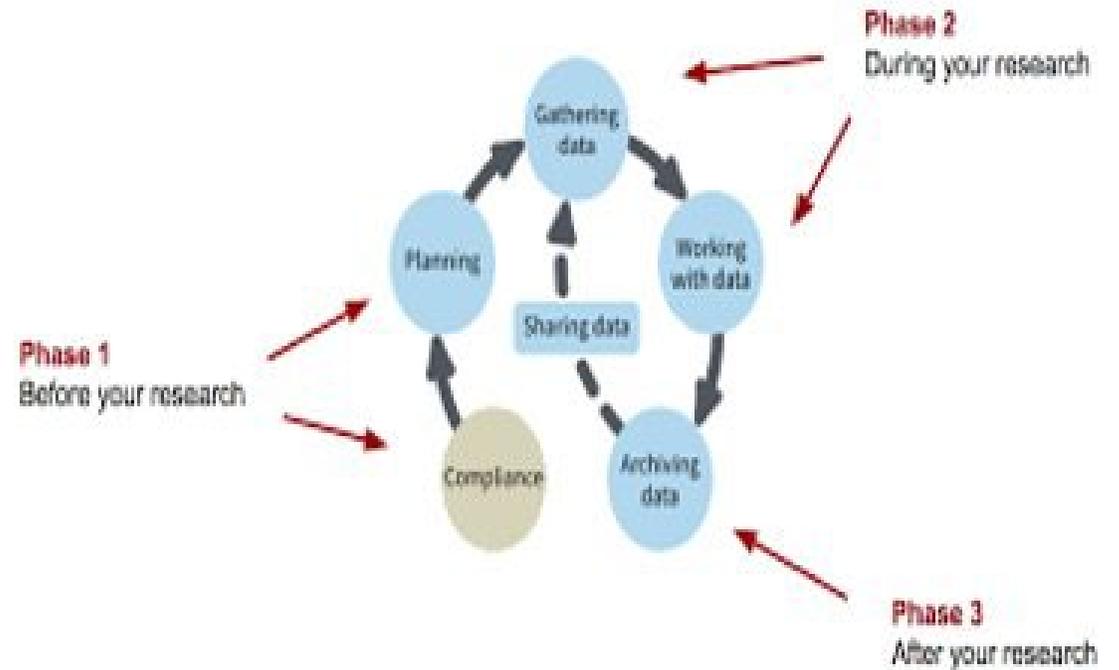


Image source: York University

Tri-Agency Research Data Management Policy

- The agencies believe that research data collected through the use of public funds should be responsibly and securely managed and be, where ethical, legal and commercial obligations allow, available for reuse by others. To this end, the agencies support the FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles for research data management and stewardship.
- Les organismes subventionnaires sont d'avis que les données recueillies par la recherche au moyen de fonds publics doivent être gérées de manière responsable et sûre. Elles doivent aussi, lorsque les obligations éthiques, juridiques et commerciales le permettent, être disponibles pour être réutilisées par d'autres. À cette fin, les organismes subventionnaires appuient les principes directeurs FAIR (traduction de l'anglais « findable, accessible, interoperable and reusable » : repérables, accessibles, interopérables et réutilisables) pour la gestion et la gérance des données de recherche.



Government
of Canada

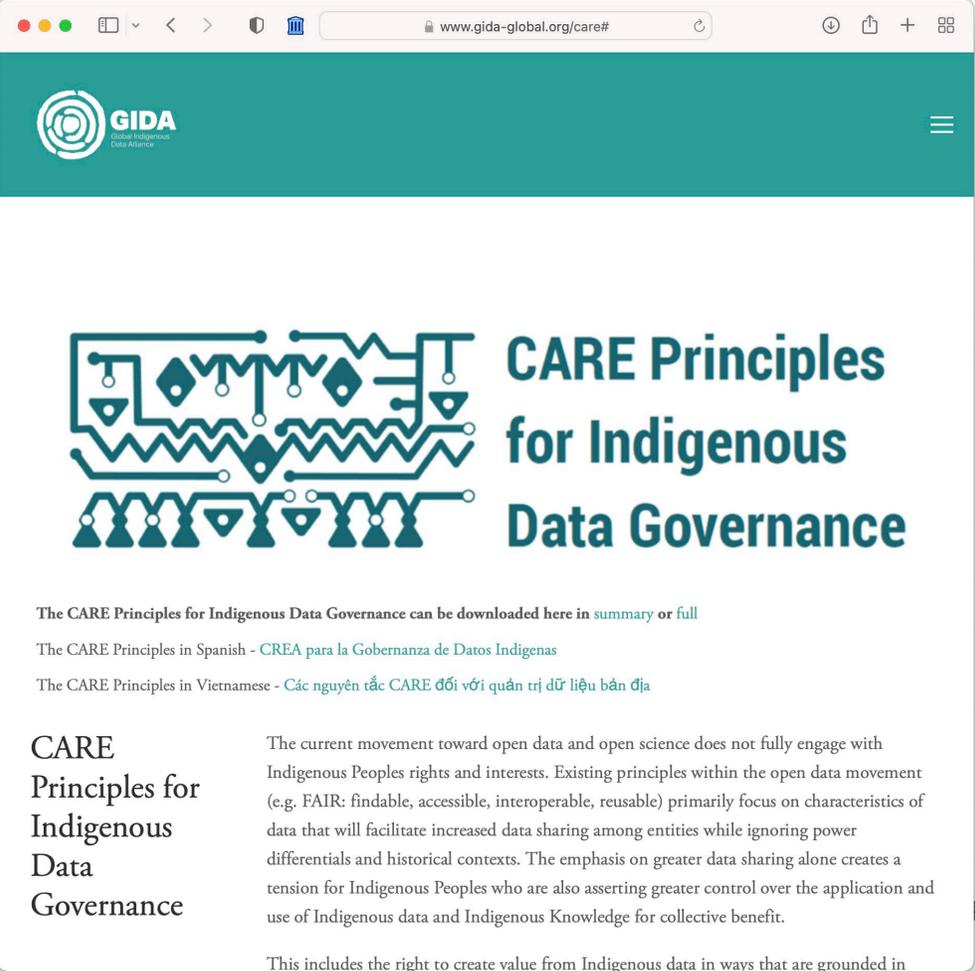
Gouvernement
du Canada

SSHRC + Data / CRSH + Données

- And it's coming soon!
- SSHRC Partnership Grants Stage 2 (Summer 2023) – The “initial” opportunity that will require a DMP.
- A strong Canadian RDM environment will begin to take shape.
- Et c'est pour bientôt !
- Subventions de partenariat du CRSH Étape 2 (Été 2023) - L'opportunité "initiale" qui nécessitera un DMP.
- Un solide environnement canadien de GDR commencera à prendre forme.

First Nations Data / Données sur les Premières Nations

- The CARE Principles



The screenshot shows a web browser window with the URL www.gida-global.org/care#. The page features the GIDA logo (Global Indigenous Data Alliance) and a stylized graphic of a circuit board with traditional Indigenous patterns. The main heading is "CARE Principles for Indigenous Data Governance". Below the heading, there are links for downloading the principles in summary or full, and translations in Spanish and Vietnamese. The text on the page discusses the current movement toward open data and open science, noting that existing principles like FAIR focus on characteristics of data that facilitate sharing while ignoring power differentials and historical contexts. It emphasizes the need for greater control over the application and use of Indigenous data and knowledge for collective benefit.



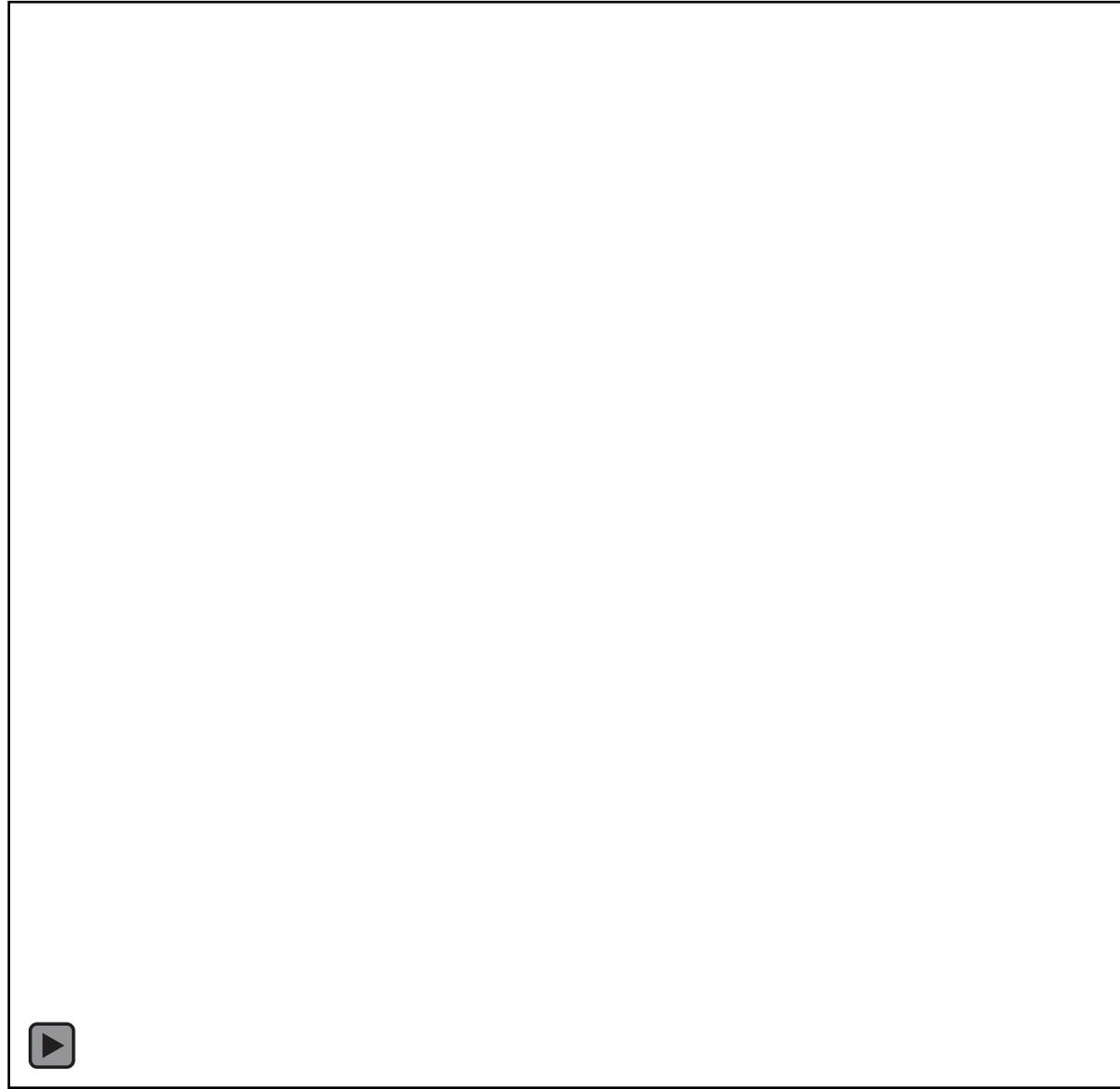
Data / Données

- Not all data needs to be openly shared – there is a colonialist viewpoint behind “all data must be shared” – but decisions to share, not share, and beyond are all taken **deliberately**.
- Toutes les données n'ont pas besoin d'être partagées ouvertement - il y a un point de vue colonialiste derrière "toutes les données doivent être partagées" - mais les décisions de partager, de ne pas partager et au-delà sont toutes prises délibérément.

At Waterloo, for example / A Waterloo (par exemple)

- Beyond boilerplate, we want a strategy that **sets researchers up for success.**
 - Stewarding your data, describing it, preserving it, and sharing it (if appropriate) – beyond buzzwords, they're things that help foster **research excellence**
 - Thinking about your data architecture at the beginning prevents headaches later down the road (incl. Security, IP debates, etc.)
 - You can access your older data and understand it
 - You can share it with other people so they can build on your work – and vice versa
 - We don't want to just have researchers putting data on disk for no reason.
- Au-delà du passe-partout, nous voulons une stratégie qui prépare les chercheurs au succès.
 - Gérer vos données, les décrire, les préserver et les partager (le cas échéant) - au-delà des mots à la mode, ce sont des choses qui contribuent à favoriser l'excellence en recherche
 - Penser à votre architecture de données dès le début évite les maux de tête plus tard sur la route (y compris la sécurité, les débats IP, etc.)
 - Vous pouvez accéder à vos anciennes données et les comprendre
 - Vous pouvez le partager avec d'autres personnes afin qu'elles puissent s'appuyer sur votre travail - et vice versa
 - Nous ne voulons pas que les chercheurs mettent des données sur disque sans raison.

Data Management Plans / Plans de gestion des données



PRESENTATION TITLE



At Waterloo / A Waterloo

- So far it's clear that **researchers** want:
 - Available at the time of need (not just when they apply, publish, or compliance)
 - Tailored to their discipline
 - Viewed as an asset (i.e. valued in T & P, etc.)
- **Researchers often want to share their data, they just don't have the time, rewards, or incentives to do so.**
- Jusqu'à présent, il est clair que les chercheurs veulent :
 - Disponible au moment du besoin (pas seulement lorsqu'ils postulent, publient ou se conforment)
 - Adapté à leur discipline
 - Considéré comme un atout (i.e. valorisé en T&P, etc.)
- Les chercheurs veulent souvent partager leurs données, mais ils n'ont tout simplement pas le temps, les récompenses ou les incitations pour le faire.

WHAT DO WE DO WITH DATA?

QUE FAIT-ON DES DONNÉES ?

Independently Understandable Data



FILE FORMAT/FORMAT DE
FICHER



DESCRIPTION/LA
DESCRIPTION



METADATA/MÉTADONNÉES



DATA STORAGE/STOCKAGE
DE DONNÉES

File Format / format de fichier

- “Recommended Preservation Formats for Electronic Records,” Smithsonian Institution Archives.
 - Word Processing/Text
 - PDF preferred
 - RTF, TXT, XML (schema) – acceptable
 - Spreadshet
 - PDF preferred
 - CSV – acceptable, TSV, TXT, XML
 - Presentation
 - PDF preferred
 - Original – acceptable
 - Images
 - TIFF – preferred
 - JPG, DNG, PNG, JP2, etc. – acceptable.
- « Formats de conservation recommandés pour les enregistrements électroniques », Archives de la Smithsonian Institution.
 - Traitement de texte/Texte
 - PDF préféré
 - RTF, TXT, XML (schéma) – acceptable
 - Feuille de calcul
 - PDF préféré
 - CSV – accepté, TSV, TXT, XML
 - Présentation
 - PDF préféré
 - Original - acceptable
 - Images
 - TIFF – préféré
 - JPG, DNG, PNG, JP2, etc. – acceptables.

EMULATION VS. MIGRATION

ÉMULATION VS MIGRATION

Description / description

- **Who, What, When, Why, Where, How**
- How was data collected?
- What did you collect?
- How did you structure it?
- Who can access it?
- What are the columns?
- What codes might you use?
- What variables did you use?
- Qui, quoi, quand, pourquoi, où, comment
- Comment les données ont-elles été collectées ?
- Qu'avez-vous collecté ?
- Comment l'avez-vous structuré ?
- Qui peut y accéder ?
- Quelles sont les colonnes ?
- Quels codes pourriez-vous utiliser ?
- Quelles variables as-tu utilisé ?

Metadata / métadonnées

Not Secure — www.yoyodesign.org/d

DCMI: Domains and Ranges for DCMI Properties

1. Introduction

Ce document emploie la terminologie du modèle abstrait DCMI [DCAM]. Le modèle abstrait (DCAM) décrit ainsi les types de relation qui sont l'objet principal de ce document :

Chaque *propriété* peut être liée à une ou plusieurs *classes* par une relation *a pour domaine*. Là où une *propriété* est déclarée avoir une telle relation avec une *classe* et qu'une *ressource décrite* est liée à une *valeur* par cette *propriété*, il s'ensuit que la *ressource décrite* est une instance de cette *classe*.

Chaque *propriété* peut être liée à une ou plusieurs *classes* par une relation *a pour image*. Là où une *propriété* est déclarée avoir une telle relation avec une *classe* et qu'une *ressource décrite* est liée à une *valeur* par cette *propriété*, il s'ensuit que la *valeur* est une instance de cette *classe*.

En pratique, cela signifie que le *domaine* (*domain*) désigne la *classe* des *ressources* pour la description desquelles on devrait utiliser cette *propriété*, tandis que l'*image* (*range*) désigne la *classe* des *ressources* qui devraient être utilisées comme *valeurs* de cette *propriété*.

Les types de relation DCAM *a pour domaine* et *a pour image* sont les mêmes que pour les propriétés RDF Schema [RDFS] `rdfs:range` et `rdfs:domain`.

2. Classes

Cette section décrit un ensemble de *classes* qui sont nécessaires pour décrire les relations *a pour domaine* et *a pour image* des *propriétés* DCMI.

Les relations entre les classes sont résumées dans le diagramme de classe UML de la figure 1 :

```
classDiagram
    class Resource
    class Class
    class PhysicalResource
    class MethodOfInstruction
    class MethodOfAccrual
    class FileFormat
    class PhysicalMedium
    class MediaFormat
    class Media Type
    class Media TypeOrExtent
    class LocationPeriodOrJurisdiction
    class Location
    class Jurisdiction
    class SizeOrDuration
    class LinguisticSystem
    class Frequency
    class Collection
    class Literal

    Resource <|-- Class
    Resource <|-- PhysicalResource
    Resource <|-- MethodOfInstruction
    Resource <|-- MethodOfAccrual
    Resource <|-- FileFormat
    Resource <|-- PhysicalMedium
    Resource <|-- MediaFormat
    Resource <|-- Media Type
    Media TypeOrExtent <|-- Media Type
    Media TypeOrExtent <|-- FileFormat
    LocationPeriodOrJurisdiction <|-- Location
    LocationPeriodOrJurisdiction <|-- Jurisdiction
    SizeOrDuration <|-- Media TypeOrExtent
```

www.dublincore.org/specifications/du

DCMI: Domains and Ranges for DCMI Properties

1. Introduction

This document uses the terminology of the DCMI Abstract Model [DCAM]. The relationship types with which this document is principally concerned are described by the DCAM as follows:

- Each *property* may be related to one or more *classes* by a *has domain* relationship. Where it is stated that a *property* has such a relationship with a *class* and a *described resource* is related to a *value* by that *property*, it follows that the *described resource* is an instance of that *class*.
- Each *property* may be related to one or more *classes* by a *has range* relationship. Where it is stated that a *property* has such a relationship with a *class* and a *described resource* is related to a *value* by that *property*, it follows that the *value* is an instance of that *class*.

In practice, this means that the *domain* indicates the *class* of *resources* that the *property* should be used to describe, while the *range* indicates the *class* of *resources* that should be used as *values* for that *property*.

The DCAM relationship types *has domain* and *has range* are the same as the RDF Schema [RDFS] properties, `rdfs:range` and `rdfs:domain`.

2. Classes

This section describes a set of *classes* that are needed in order to describe *has domain* and *has range* relationships for DCMI *properties*.

The relationships between the classes are summarized in the UML class diagram in Figure 1:

```
classDiagram
    class Resource
    class Class
    class PhysicalResource
    class MethodOfInstruction
    class MethodOfAccrual
    class FileFormat
    class PhysicalMedium
    class MediaFormat
    class Media Type
    class Media TypeOrExtent
    class LocationPeriodOrJurisdiction
    class Location
    class Jurisdiction
    class SizeOrDuration
    class LinguisticSystem
    class Frequency
    class Collection
    class Literal

    Resource <|-- Class
    Resource <|-- PhysicalResource
    Resource <|-- MethodOfInstruction
    Resource <|-- MethodOfAccrual
    Resource <|-- FileFormat
    Resource <|-- PhysicalMedium
    Resource <|-- MediaFormat
    Resource <|-- Media Type
    Media TypeOrExtent <|-- Media Type
    Media TypeOrExtent <|-- FileFormat
    LocationPeriodOrJurisdiction <|-- Location
    LocationPeriodOrJurisdiction <|-- Jurisdiction
    SizeOrDuration <|-- Media TypeOrExtent
```

Metadata / métadonnées

- Title
 - Creator – author/researcher
 - Subject
 - Description
 - Abstract
 - Type
 - Date
 - License
- Titre
 - Créateur – auteur/chercheur
 - Question
 - La description
 - Sommaire
 - Taper
 - Date
 - Licence

Data Storage / stockage de données

- Compute Canada / Digital Research Alliance of Canada
 - Rapid Access Service
 - Dataverse
 - Zenodo
 - FRDR
 - Institutional repositories
- Calcul Canada / Alliance de recherche numérique du Canada
 - Service d'accès rapide
 - Dataverse
 - Zénodo
 - FRDR
 - Dépôts institutionnels

Data Storage / stockage de données

The screenshot shows a web browser window displaying the Scholars Portal Dataverse interface. The URL in the address bar is `dataverse.scholarsportal.info/dataset.xhtml?persistentId=hdl:10864/11301`. The page title is "Derivative data for the Canadian Political Parties and Interest Groups collection" (Version 10.0). The page includes a citation for Milligan, Ian; Ruest, Nick; Lin, Jimmy (2015) and provides options to "Access Dataset", "Contact Owner", and "Share". It also displays "Dataset Metrics" showing 2,247 Downloads. The "Description" section states that the data is for the Canadian Political Parties and Interest Groups collection and provides citation instructions. The "Notes" section mentions a CC BY 2.0 CA license. Navigation tabs for "Files", "Metadata", "Terms", and "Versions" are visible, along with an "Export Metadata" button.

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

Web Archives for Historical Research Group Dataverse

Scholars Portal Dataverse > Web Archives for Historical Research Group Dataverse >

Derivative data for the Canadian Political Parties and Interest Groups collection

Version 10.0

 Milligan, Ian; Ruest, Nick; Lin, Jimmy, 2015, "Derivative data for the Canadian Political Parties and Interest Groups collection", <https://hdl.handle.net/10864/11301>, Scholars Portal Dataverse, V10

[Cite Dataset ▾](#) [Learn about Data Citation Standards.](#)

[Access Dataset ▾](#)

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

Dataset Metrics [?](#)

2,247 Downloads [?](#)

Description [?](#) This contains derivative data for the Canadian Political Parties and Interest Groups collection.

If you cite this material, please use:

University of Toronto Libraries, Canadian Political Parties and Interest Groups, Archive-It Collection 227, Canadian Action Party, <http://wayback.archive-it.org/227/20051004191340/http://canadianactionparty.ca/Default2.asp>

Notes [?](#) Type: License Notes: CC BY 2.0 CA <https://creativecommons.org/licenses/by/2.0/ca/>;

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

[Export Metadata ▾](#)

Citation Metadata [^](#)

Dataset Persistent ID [?](#) hdl:10864/11301

Data Storage / stockage de données

The screenshot shows a web browser window with the URL `dataverse.scholarsportal.info/dataset.xhtml?persistentId=hdl:10864/11301`. The browser tabs include Google Translate, Describing Data - Research Data Management, and Network Data for the Web Archives for Longitudin... The page header features the 'Scholars Portal Dataverse' logo and navigation links for Search, User Guide, Support, English, and Log In.

The main content area is titled 'Citation Metadata' and contains the following information:

| | |
|------------------------------|--|
| Dataset Persistent ID | hdl:10864/11301 |
| Publication Date | 2015-12-01 |
| Title | Derivative data for the Canadian Political Parties and Interest Groups collection |
| Author | Milligan, Ian (University of Waterloo) Ruest, Nick (York University) Lin, Jimmy (University of Waterloo) |
| Contact | Use email button above to contact. Ian Milligan (University of Waterloo) |
| Description | This contains derivative data for the Canadian Political Parties and Interest Groups collection. If you cite this material, please use: University of Toronto Libraries, Canadian Political Parties and Interest Groups, Archive-It Collection 227, Canadian Action Party, http://wayback.archive-it.org/227/20051004191340/http://canadianactionparty.ca/Default2.asp |
| Notes | Type: License Notes: CC BY 2.0 CA https://creativecommons.org/licenses/by/2.0/ca/ ; |
| Producer | Ian Milligan (University of Waterloo) https://uwaterloo.ca/web-archive-group/ |
| Grant Information | 435-2015-0011 |
| Deposit Date | 2015-12-01 |
| Time Period Covered | Start: 2005 ; End: 2015 |
| Date of Collection | Start: 2005 ; End: 2015 |

The bottom of the page shows a 'Geospatial Metadata' section which is currently collapsed.

What to do? Que faire?

- Traditionally, historians would.... **use it for their work and then put it in a filing cabinet to then throw out when they retire or move.**
 - Fears of being “scooped”
 - Fruit of “their” labour
- It was essentially unpublishable and unsharable, too.
- Traditionnellement, les historiens seraient.... l'utiliser pour leur travail, puis le mettre dans un classeur pour ensuite le jeter lorsqu'ils prennent leur retraite ou déménagent.
 - Craintes d'être « écopés »
 - Fruit de "leur" travail
- Il était essentiellement non publiable et non partageable, aussi.

The digital age is changing all of this... L'ère numérique change tout cela.



In the historical profession / dans le métier historique

- Increasingly an argument that **online databases** and **visualizations** make scholarly arguments and need to be recognized as valuable contributions.
- De plus en plus un argument selon lequel **les bases de données** et les **visualisations en ligne** constituent des arguments scientifiques et doivent être reconnus comme des contributions précieuses.

Digital History and Argument

Organizers

Stephen Robertson
George Mason University

Lincoln A. Mullen
George Mason University

Workshop participants

Edward Ayers
University of Richmond
Edward Baptist
Cornell University
Cameron Blevins
Northeastern University
Diane Harris Cline
George Washington University
Ryan Cordell
Northeastern University
Kalani Craig
Indiana University Bloomington
Seth Denbo
American Historical Association
Fred Gibbs
University of New Mexico
Jennifer Guiliano
Indiana Univ. Purdue Univ. Indianapolis
Jo Guldi
Southern Methodist University
Jason A. Heppler
University of Nebraska Omaha
Michael Jarvis
University of Rochester
Micki Kaufman
City University of New York

Sharon M. Leon
Michigan State University
Matthew Lincoln
Getty Research Institute
Austin Mason
Carleton College
Jeffrey W. McClurken
University of Mary Washington
Ian Milligan
University of Waterloo
Michelle Moravec
Rosemont College
Robert K. Nelson
University of Richmond
Angel David Nieves
Yale University
Lynn Ramey
Vanderbilt University
Janneken Smucker
West Chester University
William Thomas III
University of Nebraska Lincoln
Lauren Tilton
University of Richmond

How to cite this white paper: Arguing with Digital History working group, "Digital History and Argument," white paper, Roy Rosenzweig Center for History and New Media (November 13, 2017): <https://rrchnm.org/argument-white-paper/>.

November 13, 2017



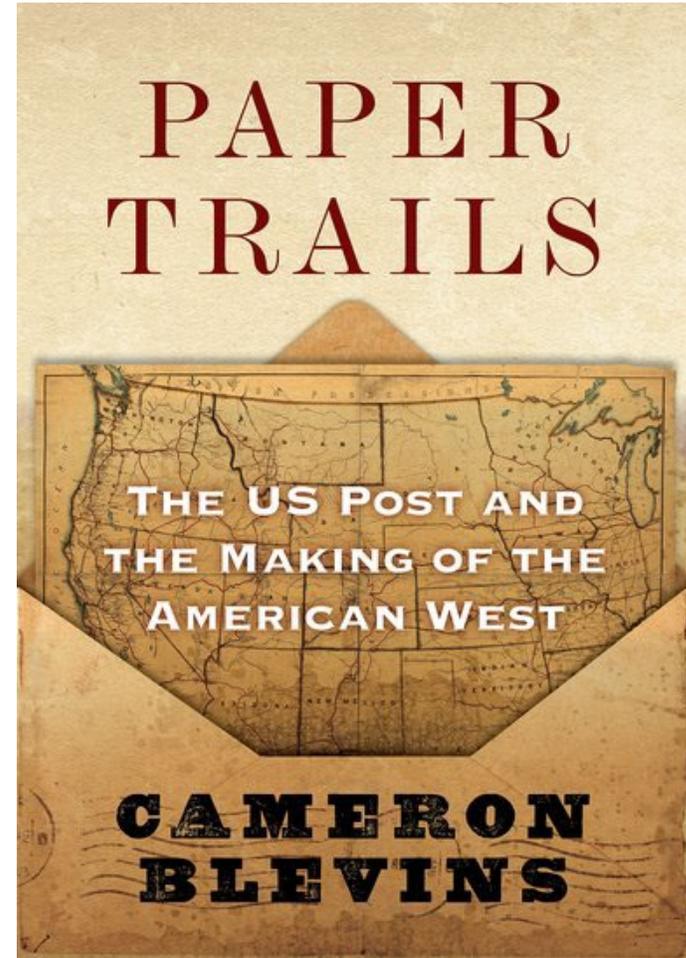
This white paper was published by the Roy Rosenzweig Center for History and New Media at George Mason University and generously funded by the Andrew W. Mellon Foundation.

In the historical profession / dans le métier historique

- Les historiens sont chargés d'expliquer la valeur de leurs productions au reste de leur col professionnel.
- Les historiens sont chargés d'expliquer la valeur de leurs productions au reste de leur col professionnel.

In the historical profession / dans le métier historique

- An exemplar project / un projet exemplaire: Cameron Blevins, *Paper Trails: The US Post and the Making of the American West*.



In the historical profession / dar

- The “Gossamer Network” Companion Website
- And the “US Post Offices” GitHub Site.



US Post Offices
1789

The screenshot shows the Harvard Dataverse interface for the 'US Post Offices' dataset. The browser address bar shows 'dataverse.harvard.edu'. The page title is 'US Post Offices' and the version is 'Version 1.0'. The citation information is: 'Blevins, Cameron; Helbock, Richard W., 2021, "US Post Offices", https://doi.org/10.7910/DVN/NUKCNA, Harvard Dataverse, V1, UNF:6:8ROml5/4qA8jHrt62PpyA== [fileUNF]'. The page includes buttons for 'Access Dataset', 'Contact Owner', and 'Share'. The 'Dataset Metrics' section shows '1,243 Downloads'. The 'Description' section provides a detailed overview of the dataset, mentioning that it contains 166,140 post offices that operated in the United States between 1639-2000. The 'Subject' is 'Arts and Humanities; Social Sciences' and the 'Keyword' is 'post offices, federal government, historical geography, united states history'.

In the historical profession / dans l

- Kellen Funk and Lincoln Mullen’s *American Historical Review* article “The Spine of American Law: Digital Text Analysis and U.S. Legal Practice.”
- Article de *American Historical Review* de Kellen Funk et Lincoln Mullen « The Spine of American Law : Digital Text Analysis and U.S. Legal Practice »

The screenshot shows the article page for "The Spine of American Law: Digital Text Analysis and U.S. Legal Practice" by Kellen Funk and Lincoln Mullen. The page is from the *American Historical Review*, Volume 123, Issue 1, February 2018. The article contents include the abstract, supplementary information, and notes. There is a link to supplementary data (zip file). The page also features a metrics section with 12 citations, 2,429 views, and 28 altmetrics. There are also links for email alerts: article activity alert, advance article alerts, and new issue alert.

academic.oup.com

PDF Split View Cite Permissions Share

Supplementary data

Supplementary Data - zip file

The American Historical Review

Never miss the latest content
Sign up for new issue email alerts

Advertisement

CITATIONS 12 VIEWS 2,429 ALTMETRIC 28

More metrics information

Email alerts

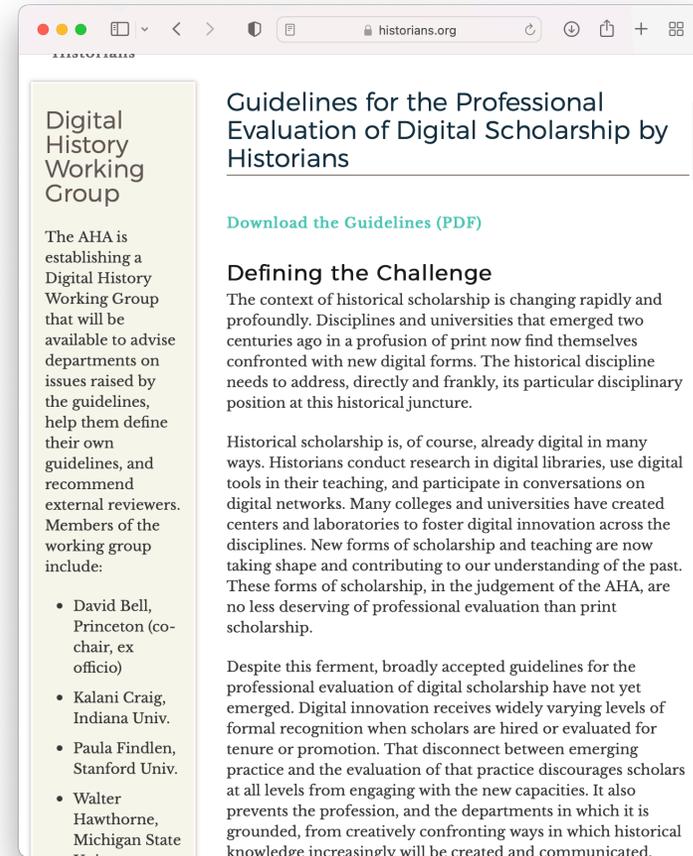
Article activity alert
Advance article alerts
New issue alert

**BUT THESE ARE MORE THE EXCEPTION THAN
THE NORM...**

**CE SONT PLUS DES EXCEPTIONS QUE LA
NORME...**

Guidelines / Des lignes directrices

- *The Guidelines on Evaluating Digital Scholarship*
- Growing acceptance of digital scholarship at the AHA and professional publication
- The rise of the “digital dissertation”
- Les lignes directrices sur l'évaluation des bourses d'études numériques
- Acceptation croissante de l'érudition numérique à l'AHA et publication professionnelle
- L'essor de la « dissertation numérique »

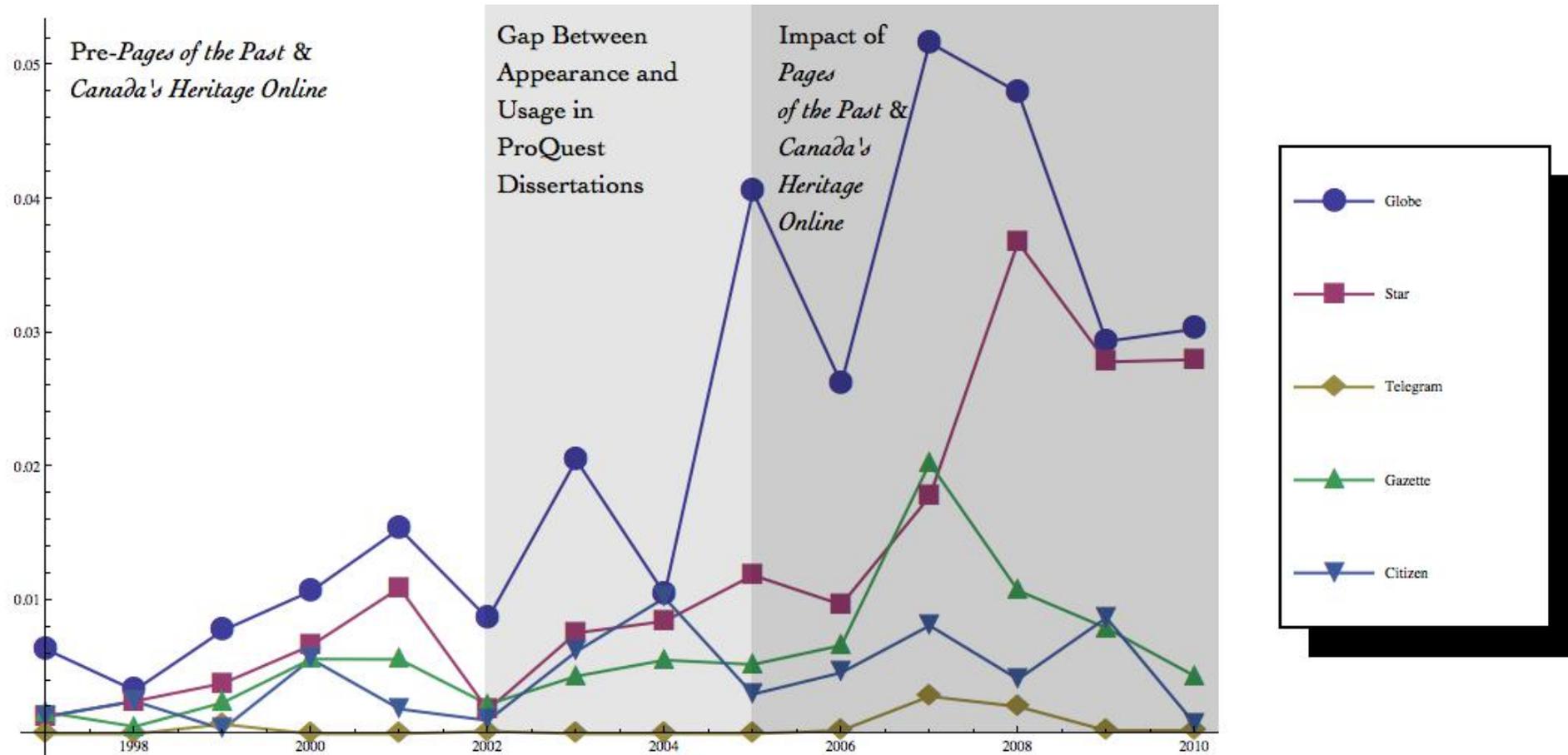


Concerns / Préoccupations

- Hesitation to cite datasets
 - Data provision is not SERVICE – it is RESEARCH!
- Sustainability
 - What happens if a Press reorganizes its website?
 - A faculty member retires?
 - GitHub collapses?
- Hésitation à citer des ensembles de données
 - La fourniture de données n'est pas SERVICE - c'est RECHERCHE !
- Durabilité
 - Que se passe-t-il si une Presse réorganise son site Internet ?
 - Un membre du corps professoral prend sa retraite?
 - GitHub s'effondre ?

And digitization is uneven...

Et la numérisation est inégale.



In other words / En d'autres termes

The more something is digitized the more it is used / Plus il est numérisé, plus il est utilisé.

The mediation of a source impacts its use / La manière dont une source est médiatisée a un impact sur l'utilisation.

Also globally / Aussi globalement



“The Transnational and the Text-Searchable”

The Transnational and the Text-Searchable: Digitized
Sources and the Shadows They Cast

LARA PUTNAM

THE TRANSNATIONAL TURN IS accelerating simultaneously with the digital turn, and it is no coincidence. Source digitization has transformed historians' practice in ways that facilitate border-crossing research in particular. Web-based full-text search decouples data from place. In doing so, it dissolves the structural constraints that kept history bound to political-territorial units long after the intellectual liabilities of that bond were well known. Digital search has become the unacknowledged handmaiden of transnational history. It is time to take stock of what that partnership enables—and what it obscures.

Digital search offers disintermediated discovery. Algorithms fetch for us, doing away with the need for intermediaries like brick-and-mortar stores (if you are Amazon.com) or regional expertise, bibliographies, and immersive reading (if you are a historian). For the first time, historians can *find* without knowing where to look. As a result, at an unprecedented rate we are finding connections in unexpected places: powering publication on mobile ideas and international audiences; circuits, networks, and border-crossing flows. Technology has exploded the scope and speed of discovery. But our ability to read accurately the sources we find, and evaluate their significance, cannot magically accelerate apace. The more far-flung the locales linked through our discoveries, the less consistent our contextual knowledge. The place-specific learning that historical research in a pre-digital world required is no longer baked into the process. We make rookie mistakes.

Moreover, historical inquiry that is powered by patterns in the digitizable detritus of the modern world will tend to foreground certain kinds of actors and certain aspects of their lives, pulling toward “a kind of international provincialism” that fails to note key local and national dynamics.¹ We risk overemphasizing the importance of

The "Side Glance"; Le coup d'œil latéral



- Use Wikipedia/HathiTrust/Google Books to "side glance" around the world.
- Utilisez Wikipedia/HathiTrust/Google Books pour « jeter un coup d'œil de côté » à travers le monde.

Should historians travel? Les historiens doivent-ils voyager ?



FOOD FOR THOUGHT

Could you be a historian
without visiting the place
you study?

Pourriez-vous être historien
sans visiter le lieu que vous
étudiez ?

What do we gain? What do
we lose?

Que gagnons-nous ? Que
perdons-nous ?

What about our
relationships? Et nos
relations?

Yet in sum...

- A Transformation is underway.
 - **The digital mediates the past, just as archives do.**
 - Le numérique médiatise le passé, tout comme les archives.
 - **The research approach of a historian has dramatically transformed.**
 - L'approche de recherche d'un historien s'est radicalement transformée.



Digital Literacy/Littératie numérique



- **All** historians have been transformed by the digital.
- **Tous** les historiens transformés par le numérique.
- **All** of our sources are mediated through technology.
- **Toutes** les sources sont médiatisées par la technologie.

Interdisciplinarity / interdisciplinaire

- Using a search engine is interdisciplinary.
- L'utilisation d'un moteur de recherche est interdisciplinarité.
- Talk to librarians!
- Parlez aux bibliothécaires !



Beyond history/Au-delà de l'histoire

- **All students need**
 - Knowledge of algorithms
 - Knowledge of context
 - Digital literacy skills
- **Tous les élèves ont besoin:**
 - connaissance des algorithmes
 - connaissance du contexte
 - compétences en littératie numérique



**THROUGH GOOD RDM PRACTICE, WE CAN STOP
REINVENTING THE WHEEL...**

**GRÂCE À DE BONNES PRATIQUES RDM, NOUS
POUVONS ARRÊTER DE RÉINVENTER LA
ROUE...**

Thank you very much. Merci beaucoup.



UNIVERSITY OF WATERLOO



YOU+WATERLOO

Our greatest impact happens together.