



# Research Data Management:

Organization & Documentation: Best Practices to Structure your Research Project Data

---

JANUARY 18, 2018

A. MICHELLE EDWARDS, OAC

CAROL PERRY, LIBRARY

# Objectives

---

Understand the basic concepts of Research Data Management (RDM)

Identify best practices for organizing and documenting your research data

Ability to apply these practices to your own projects



## Setting the stage

# Discussion

---

What are research data?

How would you define research data?

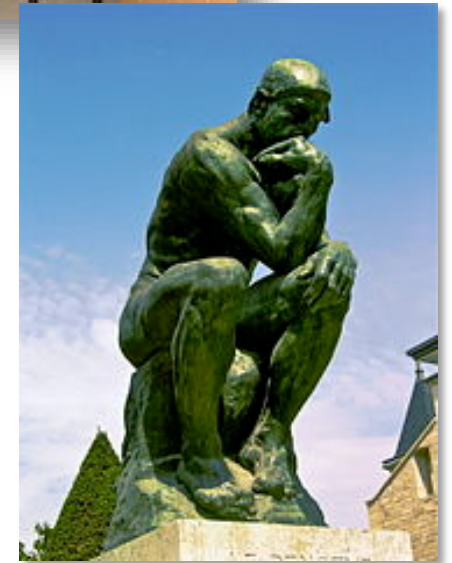
Have you heard of RDM?

# What is data?

122  
Chapter 10  
Pilla Resides.  
Families and individuals alike soon became used to new conditions and accept them un-  
questioningly. By the time the war had closed  
it seemed as if the American body had always  
been at sea-side, hidden after the first time  
of the war. Pilla began to accept again,  
making automatically a slight change  
on solid ground. She went back to work  
and dressed it as a child, as if  
she had been doing it all her life. But she  
did not see the baby sleep  
to sleep; she still handled it as gingerly  
as if it were some kind of a small bird,  
and a breakable bird at that; but she  
did her work thoroughly and that was all  
a cleaner, better - cared - for infant in sleep.  
St. Mary. She was used to making the creature  
every day and getting the result down in  
her diary. But sometimes she asked herself  
particularly why women had ever led  
in even the hardest time a hot / after day.



15.47	16.23	15.99	12.21	14.8	13.68
15.47	15.86	15.99	12.14	14.92	13.68
15.47	16.34	16.11	12.14	14.92	13.68
15.47	15.33	16.11	12.21	14.92	13.68
15.47	16.34	16.11	12.14	15.26	13.68
15.47	15.45	16.11	12.62	15.26	13.68
15.47	15.88	16.11	12.14	15.26	13.68
15.47	15.77	16.11	12.21	14.83	13.68
15.47	15.41	16.11	12.14	14.83	13.68
	15.41	16.1			
	15.45	16.1			
	14.88	15.5			



# Defining research data

---

- Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results
- Can be analogue or digital materials
- Can include lab notebooks and software
- Digital data can be:
  - Born digital - created in a digital form
  - Digitized - converted from analogue to a digital form

Definition c/o [Digital Curation Centre](#)

# What is RDM?

---

A sound strategy and best practices used to.....

- Organize
- Document
- Store
- Analyze
- Secure
- Preserve/Share/Reuse

.....**Your data**

# Why should I care?

**CBCnews | Politics**

Home Opinion World Canada **Politics** Business Health Entertainment Technology & Science Video

Politics Power & Politics CBC SecureDrop

## Massive privacy breach at Public Services reveals workers' salaries

3rd major breach at the department in a year prompts complaints to privacy commissioner

By Dean Beeby, CBC News Posted: Jun 02, 2017 5:00 AM ET | Last Updated: Jun 02, 2017 5:38 AM ET

Stay Connected with CBC News

Mobile Facebook Podcasts Twitter Alerts Newsletter

Marie Lemay is deputy minister of Public Services and Procurement Canada, where a third major privacy breach in the span of about a year exposed the personal information of almost 13,000 workers. CBC News has learned. (CBC News)

## Health, personal records of 2,500 Arizona families with newborns lost

Ken Alltucker, The Republic | azcentral.com 6:14 p.m. MT May 26, 2017

State health officials have warned 2,500 individuals and families with newborns that a box containing sensitive health, financial and personal information has been lost in the mail.

The Arizona Department of Health Services said a newborn-screening program gathered the information for billing purposes.

The misplaced records may include health information for the program and their mothers, as well as the...

(Photo: Kristyna Wentz-Graff/Milwaukee, Wisconsin, Journal Sentinel)

f 81 CONNECT TWEET in 15 LINKEDIN COMMENT 1 EMAIL

**Global NEWS** Toronto Change Location News & Radio Programs Newscasts, Radio, and Videos

World Canada Local Politics Smart Living Money Entertainment Health Tech Science Spor

HEALTH May 25, 2017 8:42 pm Updated: May 26, 2017 11:27 am

## Saskatoon Health Region sees increase in privacy breaches and complaints

By Jacqueline Wilson Reporter Global News

Comments Facebook 4 Twitter Email Print

PLAY VIDEO

**SECURITY EXECUTIVES**

## Data Breach Digest: Cybersecurity trends we've seen thus far in 2017, what's still to come

BY MICHAEL BRUEMMER ON JUN 1 2017

**Oh Man, I Accidentally Deleted Production Data From the Database**



# Why should I care?

---

Ethical/Legal obligations

Publication needs

Reuse data later

Mitigate risks

- file corruption
- lost data
- hard drive failure
- obsolete software
- inability to identify files
- human error
- unforeseen disaster

# Meeting legal/ethical requirements

---

Ethics board approval for certain types of research

Funding agencies

Public policies & legislation

Journal policies



# Why should I care?

---

Can you find the data you need?

Can you understand the data you have?

Can others understand the data you have?

Are you required to handle your data in a specific way?

Are you required to share or preserve it?

How will you share/preserve your data?

Where will you share/preserve it?

# Research life cycle



# Research life cycle



# Create a plan

---

## Elements:

Collecting/creating data

Documenting your work

Managing your files

Storing, securing & backing up files

Preservation

Access, sharing & reuse

# Where to start?

## DMP Assistant

Canadian Data Management Planning online bilingual tool for preparing data management plans.

Guides users through online form with questions and prompts

Completed plan can be downloaded or exported in a variety of formats

The screenshot shows the Portage DMP Assistant website. At the top left is the Portage logo with the tagline "Shared stewardship of research data". To the right are navigation links for "Version française", "Home", "About", "Help", and "Terms of Use". The main content area features a description of DMP Assistant as a bilingual tool for preparing data management plans. Below this is a four-step process: 1. Sign up with DMP Assistant, 2. Sign in and select a template, 3. Answer questions with guidance, and 4. Revisit the tool to review answers. On the right side, there are two large buttons: "Sign in" (for existing users) and "Sign up" (for new users). A note at the bottom right states that the website is currently working on single sign-in authentication and that users should create a new account for now.

**portage**  
Shared stewardship of research data

[Version française](#) | [Home](#) | [About](#) | [Help](#) | [Terms of Use](#)

DMP Assistant is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

- Step 1** Sign up with DMP Assistant
- Step 2** Sign in and select a template under Organizations. The Portage template is the default.
- Step 3** Answer the questions that are relevant to your work. Guidance and examples are provided.
- Step 4** Revisit the tool throughout your research to review or revise your answers.

**Sign in** +

If you have an existing account with DMP Assistant or previous version of DMP Builder.

**Sign up** +

New to DMP Assistant? Sign up today.

Please note that we are currently working on single sign-in authentication. For now, please create a new DMP Assistant account. You will have the option to link your DMP Assistant account to your campus ID when that feature becomes available.

## Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, you can choose the **Portage Data Stewardship Template**. The Portage Data Stewardship Template is based on internationally accepted standards and best practices. It has been prepared and is maintained by a group of research data management experts from research libraries across Canada.

To see institutional questions and/or guidance, select your organization.

You may leave blank or select a different organization to your own. If you leave blank, def

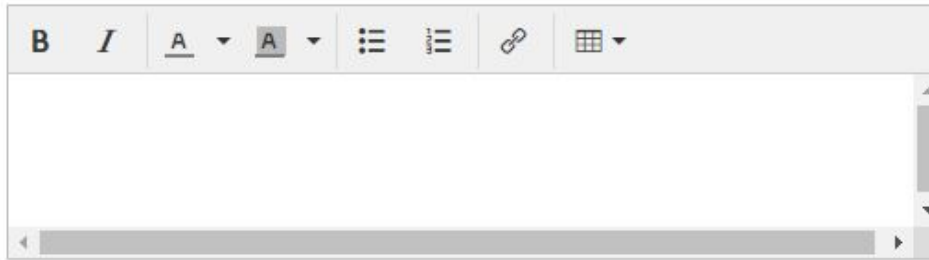
[Create plan](#)

Portage

McGill University  
Queen's University  
Ryerson University  
Scholars Portal  
University of Calgary  
**University of Guelph**  
University of Lethbridge



What documentation will be needed for the data to be read and interpreted correctly in the future?



A screenshot of a rich text editor interface. The top toolbar includes buttons for bold (B), italic (I), text color (A with a dropdown arrow), background color (A with a dropdown arrow), bulleted list, numbered list, link, and table. Below the toolbar is a large, empty text area with a vertical scrollbar on the right and a horizontal scrollbar at the bottom.

Save

Not answered yet

General guidance and information



Local guidance links and information



Guidance **Share note**

### Portage +

Typically, good documentation includes information about the study, data-level descriptions, and any other contextual information required to make the data usable by other researchers. Other elements you should document, as applicable, include: research methodology used, variable definitions, vocabularies, classification systems, units of measurement, assumptions made, format and file type of the data, a description of the data capture and collection methods, explanation of data coding and analysis performed (including syntax files), and details of who has worked on the project and performed each task, etc.

### University of Guelph guidance +

The University of Guelph Library has created a [Data Management Planning Checklist](#) which can be used to identify and keep track of the data management practices that you will utilize throughout the data life-cycle, including what information and tools will be used to document your work.

For more information see [Documenting Your Work](#) or contact us at [lib.research@uoguelph.ca](mailto:lib.research@uoguelph.ca).

# Collecting data

---

**What** ... will you collect?

**How** .... will you collect it? .... Organize it? .... Handle it?

**Where/how** .... will you store it?

**Who** .... will have access to it? ..... Be responsible for data consistency and quality?




























Are there requirements on how you handle your data ?

Have you documented these requirements for referral during the project?

Have you signed any forms/agreements that constrain your options to share your data if desired?

# Do you have everything under control?

Sample directory:

 June 2016 BRDC Access Report.	9/14/2016 2:34 PM	Adobe Acrobat D...	139 KB
 Cisco_WebEx_Add-On(2)	8/17/2016 3:15 PM	Application	881 KB
 Cisco_WebEx_Add-On(1)	8/17/2016 3:08 PM	Application	881 KB
 Watson et al. 2012	8/15/2016 3:34 PM	Adobe Acrobat D...	790 KB
 Lyons et al. 2011	8/15/2016 3:34 PM	Adobe Acrobat D...	584 KB
 Lyons et al. 2008	8/15/2016 3:34 PM	Adobe Acrobat D...	295 KB
 Scanned from a Xerox Multifunction Printer	8/9/2016 12:41 PM	Adobe Acrobat D...	39 KB
 Portage_discovery_white_paper_EN	8/8/2016 2:11 PM	Adobe Acrobat D...	1,073 KB
 zg63std	7/22/2016 10:38 AM	Application	9,324 KB
 SecureDownloadManager	7/21/2016 3:21 PM	Text Document	0 KB
 100484827198	7/21/2016 2:42 PM	Secure Download ...	1 KB
 SDM_EN	7/21/2016 2:42 PM	Windows Installer ...	756 KB
 From Coast to Coast Canadian Collaboration in a Changing RDM...	6/28/2016 9:05 AM	Adobe Acrobat D...	968 KB
 L06_DataProtectionBackups	4/27/2016 3:07 PM	Microsoft PowerP...	981 KB
 L01_DataManagement	4/27/2016 3:05 PM	Microsoft PowerP...	12,407 KB
 WinDirStatPortable_1.1.2.80_Rev_3.paf	4/15/2016 9:52 AM	Application	948 KB
 windirstat1_1_2-src	4/5/2016 9:11 AM	ZipGenius Zip File	607 KB
 windirstat1_1_2-src	4/5/2016 9:09 AM	ZipGenius 7z File	255 KB
 owssvr	2/23/2016 1:33 PM	Microsoft Excel W...	1 KB
 COLOURBOX1582191_organize	2/8/2016 3:23 PM	JPEG image	3,439 KB
 COLOURBOX_SAMPLE541068	2/8/2016 9:56 AM	JPEG image	292 KB
 savedrecs (1)	10/30/2015 10:57 ...	Text Document	2 KB
 savedrecs	10/30/2015 10:57 ...	Text Document	9 KB
 hlud04ww	9/10/2015 3:07 PM	Application	26,476 KB
 ccsetup509	9/10/2015 1:38 PM	Application	6,512 KB
 meeting.collab	9/9/2015 10:08 AM	COLLAB File	11 KB
 Atrium_overview_20150825_CEBedits	9/2/2015 2:12 PM	Microsoft Word D...	20 KB

# Organizing data

---

Develop a project folder structure

Assign project acronym or ID

e.g. Beef Cattle Diet Study =BCDS

Structure :

begin folder name with project acronym or ID

follow the logic of the project

clear, concise and logical folder naming

use underscore in between naming segments

Create a readme (overview) file which describes project structure and contents

# Folder structure

---

BCDS

bcds\_readme.txt

BCDS\_2015

BCDS\_2016

BCDS\_2017

OR

bcds\_feed\_suppl

bcds\_silage

bcds\_hay

OR

bcds\_farm\_a

bcds\_farm\_b

bcds\_farm\_c

# Readme file – sample

---

Beef Cattle Diet Survey 2016

BCDS = Beef Cattle Diet Survey

Folders:

*bcds\_feed\_suppl = feed containing supplements data – all farms*

*bcds\_hay = hay feed data - all farms*

Files:

*raw = rawdata files – all farms*

Data collected daily stored in monthly files January 1 2016-December 31 2016.

Date = ISO date yyyy/mm/dd

UOM = units of measure

Note: Farm 3 data not collected January 15 2016 and January 16 2016.

Out of range values are shown as -6999



## Workshop Scenario

The French Blast Research Group received an NSERC grant in 2016-2017 to continue an ongoing feeding trial. The research involves three horses from a number of different farms, where each horse is weighed(kg) at the beginning of the trial, placed on one of three feed regimens (hay, pasture, or silage), their total feed intake(kg) is measured for 2 consecutive weeks, and their weights (kg) are taken at the same time.

Additional information that is gathered includes: the name of the horse's owner, the annual income of the owner, and feed costs of the owner for their horses.

You have been hired by the French Blast Research Group to manage their data collections. They have 5 years of horse feeding trial data that was collected from a number of horse farms across Ontario since 2011. You have been hired to review, manage, and preserve their research trial data according to best practices and requirements of the NSERC grant.

# Exercise 1

---

Work in groups of 2 or 3

Use the scenario to identify the information you will be gathering.

Create new folder directory structure for the data collected from this research project during the last 5 years

Create a readme file for the project



# Review

Did you encounter any problems?

# File naming

---

Be consistent

Be descriptive

Be brief – less than 25 characters preferred

Begin file name with project acronym or ID (file name independent of location)

Use an underscore between different name elements

## **Don't use:**

- Unusual characters !@#\$%^&\*()\_+
- Space between name segments

# Exercise 2

---

Rename the files and outline which folder they would be placed in

# Review

Did you encounter any problems?

# Conclusion

---

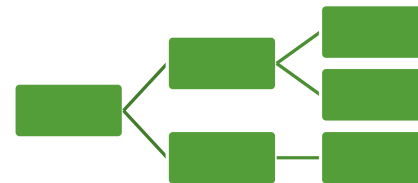
Research projects are cyclical in nature



Start planning how you will handle your data at the beginning of the project



Developing an organizational structure for your project documents should be well thought out before the data has been collected.



# Contact information:

---

Michelle Edwards  
OAC Statistics Consultant  
[oadstats@uoguelph.ca](mailto:oadstats@uoguelph.ca)

<http://www.uoguelph.ca/oac/stats>

Carol Perry  
Associate Librarian Research & Scholarship  
[lib.research@uoguelph.ca](mailto:lib.research@uoguelph.ca)

# Photo credits

---

All photos used under CC BY – SA 2.0 licenses

<https://www.flickr.com/photos/smerikal/5532818874/in/photostream/> - title image2

[https://upload.wikimedia.org/wikipedia/commons/9/93/Cutting\\_silage\\_at\\_Folsetter\\_1.jpg](https://upload.wikimedia.org/wikipedia/commons/9/93/Cutting_silage_at_Folsetter_1.jpg) - title image1

<https://www.flickr.com/photos/91035846@N05/9437703460/> - title image3