



Digitalbevaring.dk

Foundations of Research Data Management

(Les bases de la Gestion des Données de Recherche)

Marc-Olivier Croteau et H el ene Tardif
Biblioth ecaires
Direction des biblioth equ

les biblioth equ

Universit  
de Montr al

Outline

- 1- Get organized like a librarian !
- 2- Introduction to Research Data Management (RDM)
- 3- Data Management Plan (DMP)
- 4- Organize research data
- 5-Conclusion

Get organized like a librarian !

Excellence is mundane !

Superlative performance is really a confluence of dozens of small skills or activities, each one learned or stumbled upon, which have been carefully drilled into habit and then are fitted together in a synthesized whole (Chambliss, 1989).

Reference :

Chambliss, D. F. (1989). The mundanity of excellence: An ethnographic report on stratification and olympic swimmers. *Sociological Theory*, 7(1), 70–86. <https://doi.org/10.2307/202063>

Get organized like a librarian !

1. Organize your literature review (with EndNote or Zotero or...)
2. Organize your data (from the beginning...)
 - ⇒ Organize the contents of your files (good practices)
 - ⇒ Organize your files and team collaboration
 - ⇒ Organize data for publication and reuse
3. Organize your writing (using a thesis model + good practices)
4. Organize your future in research (manage your digital reputation as a researcher)

Get organized like a librarian !

Tool box to find everything in one place:

Organiser sa recherche

Dès le début de son parcours de chercheuse ou de chercheur l'approche gagnante est une bonne organisation de son travail :

Préparer sa revue de littérature

- Faire de la [veille informationnelle](#) et utiliser [EndNote](#) ou [Zotero](#)
- Réaliser une [synthèse des connaissances](#)

Organiser ses données de recherche

Rédiger son plan de gestion des données

- [Exemples de PGD](#)
- [Modèle](#) de plan en format MS Word

Gérer ses fichiers

- Bonnes pratiques : [How to name files](#) et [File naming best practices](#)
- Outils de versionnage : [GIT](#), [SVN](#)

Documenter ses données

- Dico de données : [Data dictionaries](#), [How to make a Data Dictionary](#)
- Fichier README en [français](#) ou en [anglais](#)
- Schémas de métadonnées : [Darwin Core / DwC-A](#), [EML-XML](#)

Optimiser sa rédaction

- Modèle de thèse [DOCX](#) ou [LaTeX](#)
- [Rédiger un article scientifique](#)

Planifier son avenir en recherche

- [Identifiants et profils de chercheur.e.s](#) (ORCID, etc.)

In this Libguide : bib.umontreal.ca/informatique-mathematique-sciences-nature/sciences-biologiques?tab=5237710

Outline

1- Get organized like a librarian !

2- Introduction to Research Data Management (RDM)

3- Data Management Plan (DMP)

4- Organize research data

5-Conclusion

Introduction to RDM

- What is research data?
 - Data to produce or validate research results
- Several possible sources
 - observation : e.g. field data
 - experimentation : e.g. chemical analysis
 - simulation : e.g. climate change modeling
 - derived data : data calculated from other data
 - metadata : data about data

Introduction to RDM

What is Research Data Management?

A set of actions taken during a research project that better organizes the storage, sharing, use, documentation and preservation of research data.

Introduction to RDM : The data lifecycle

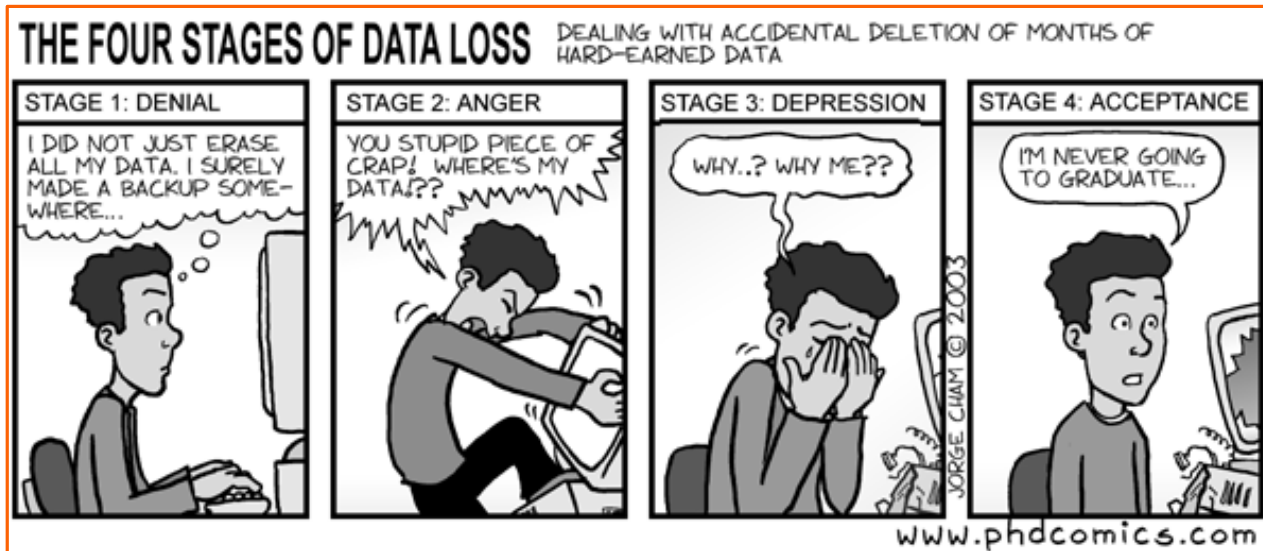
Data management practices cover the entire lifecycle of the data :

- ❖ from planning the investigation to conducting it...
- ❖ ...and from backing up data as it is created and used to long term preservation of data deliverables after the research investigation has concluded.



Introduction to RDM

Why organizing your research data is important ?



- A good organization facilitates research work,
- Ensures continuity in the course of research,
- Avoids the loss or compromise of data,
- Facilitates the validation of results,
- Facilitates the sharing and reuse of data (and collaboration with a view to open science).

Introduction to RDM

The granting agencies / Les organismes subventionnaires

They increasingly recognize the importance of data as a research output. Many organizations require submission of a Data Management Plan (DMP) and data sharing.

*The DMP is a **management tool**. [...] Its objective is to summarize the description and the evolution of the datasets of your research project. It prepares the sharing, reuse and sustainability of data.*



Introduction to RDM

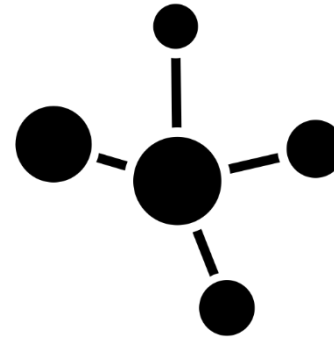
The policy includes suggested requirements related to three primary areas:



Created by popcomarts
from Noun Project



Created by popcomarts
from Noun Project



Created by popcomarts
from Noun Project

Introduction to RDM

Institutional strategy

- Published in February 2023
- Establishes units responsible for specific issues



<https://recherche.umontreal.ca/vrrdci/gestion-des-donnees-de-recherche-la-strategie-institutionnelle/>

Outline

- 1- Get organized like a librarian !
- 2- Introduction to Research Data Management (RDM)
- 3- Data Management Plan (DMP)
- 4- Organize research data
- 5-Conclusion

Introduction to RDM

nature > editorials > article

a natureresearch journal

nature

Subscribe



Search



Login

EDITORIAL · 13 MARCH 2018

Everyone needs a data-management plan



They sound dull, but data-management plans are essential, and funders must explain why.



 PDF version

RELATED ARTICLES

Data management made simple



Data Management Plan (DMP)

Organize data at each stage of the data lifecycle !

- Tool for research project management
- Start writing at the beginning of the research cycle
- Update during research
- Goals = reflection, collaboration, transparency...
- Risk management
- Grant application

Data Management Plan (DMP)

Components of a DMP :

1. Data collection > data types, file formats, naming and version control
2. Documentation > ensure data can be read and interpreted
3. Data storage and backup throughout the research
4. Data preservation strategy for long-term access
5. Provisions for sharing and reuse
6. Sharing of responsibilities and necessary resources
7. Ethical and legal compliance

Tool available :

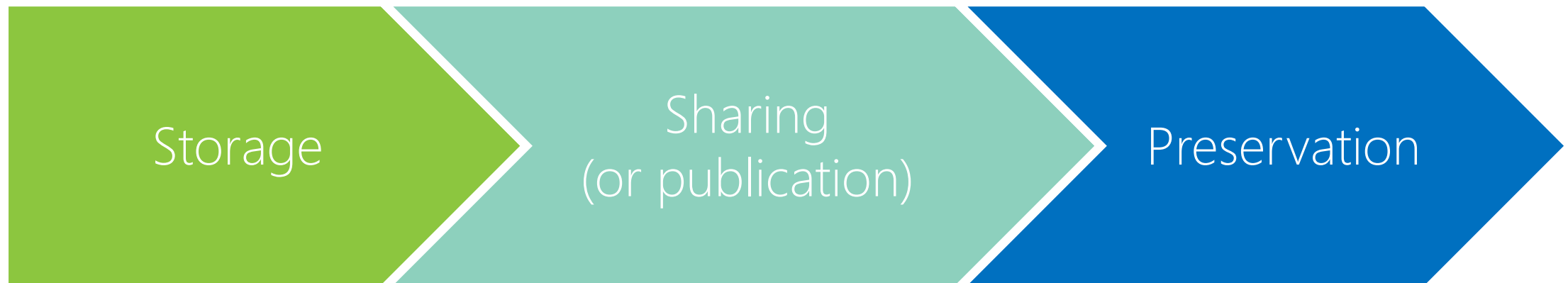
[UdeM MS Word form](#)

Outline

- 1- Get organized like a librarian !
- 2- Introduction to Research Data Management (RDM)
- 3- Data Management Plan (DMP)
- 4- Organize research data
- 5-Conclusion

Organize research data

- Organize data for publication and reuse
- Organize the contents of your files
- Organize your files and team collaboration



Organize research data : Data collection



Digitalbevaring.dk

Organize the contents of your files :

- The goal of data entry is to create data that is valid or has gone through a quality assurance process ...
- ... and which are organized in such a way as to facilitate their use or archiving.
- Commonly used tools: Excel, R, Google Docs forms, Open Office.

Organize research data : Data collection

Organize from the start of taking or obtaining data

	A	B	C	D	E	F	G	H	I	J
1	Site_ID	Year	Month	Site_type	Reservoir_number	System_type	location_sampled	Latitude_GPS	Longitude_GPS	MeHg
2	SW33	2016	Aug_16	natural	natural	GW	groundwater	51,36618	-63,87583	0,021072679
3	SW35	2016	Aug_16	natural	natural	GW	groundwater	51,21284	-63,92261	0,080348182
4	SW40	2016	Aug_16	natural	natural	GW				
5	L001	2016	Aug_16	natural	natural	lake				
6	L001-K-epi	2016	Aug_16	natural	natural	lake				
7	L001-K-hypo	2016	Aug_16	natural	natural	lake				
8	L002	2016	Aug_16	natural	natural	lake				
9	L002-K-hypo	2016	Aug_16	natural	natural	lake				
10	LR1	2016	Aug_16	natural	natural	LR_DS				
11	LR2	2016	Aug_16	natural	natural	LR_DS				
12	LR3	2016	Aug_16	natural	natural	LR_DS				
13	R01-07	2016	Aug_16	reservoir	Romaine_1	bay				
14	R01-07-HK	2016	Aug_16	reservoir	Romaine_1	bay				
15	R01-08	2016	Aug_16	reservoir	Romaine_1	main_channel				
16	R01-08.4	2016	Aug_16	reservoir	Romaine_1	main_channel				
17	R01-09	2016	Aug_16	reservoir	Romaine_1	main_channel				
18	R01-09-HK	2016	Aug_16	reservoir	Romaine_1	main_channel				
19	R02-46	2016	Aug_16	reservoir	Romaine_2	bay				
20	R02-46-3	2016	Aug_16	reservoir	Romaine_2	main_channel				
21	R02-46-4	2016	Aug_16	reservoir	Romaine_2	bay				
22	R02-48	2016	Aug_16	reservoir	Romaine_2	main_channel				
23	R02-49	2016	Aug_16	reservoir	Romaine_2	main_channel				

	A	B	C
1	Column name	Definition	
2	Site_ID	Site identification code	
3	Year	Year of field sampling campaign (2016, 2017, 2018)	
4	Month	Month of field sampling campaign (Aug_16, Jun_17)	
5	Site_type	Non-flooded or reservoir site	
6	Reservoir_number	Number of the reservoir for reservoir complex	
7	System_type	Aquatic system type (lake, headwater (GW), upstream river (river_US) & downstream river (river_DS) in relation to the reservoir complex, tributaries, lakes, groundwater, river and reservoir sites)	
8	Sampling_depth	Sampling depth of water (surface (epilimnion or groundwater))	
9	Latitude_GPS	Latitude of the sampled site measured with GPS	
10	Longitude_GPS	Longitude of the sampled site measured with GPS	
11	Depth	Depth of sampling site measured with a depth gauge	
12	Temperature	Water temperature of sampling site measured with a thermometer	
13	pH	pH of sampling site measured with a pH meter	
14	DO	Dissolved oxygen concentration measured with a DO meter	
15	%DO	Percentage of dissolved oxygen concentration measured with a DO meter	

Metadata information	
Table 1. Description of the dataset	
Title of dataset	Mercury concentrations and water chemical variables measured in La Romaine hydroelectric reservoir complex in the summer of 2016, 2017 and 2018.
URL of dataset	DOI to insert here (once on Zenodo)
Abstract	This dataset includes chemical and physical data measured in situ in water in natural and recently dammed portions of La Romaine River watershed in Northern Quebec, Canada. Samples were collected from 2016 to 2018 in August, as well as in June 2017. For most sites, the samples were collected close to the water surface (30 cm deep) with a peristaltic pump following the clean hands dirty hands sampling protocol to avoid any contamination by trace metals. Filtered samples were collected using an in-line Whatman 0.45 µm filtration capsule attached to the tubing. For June 2017, physicochemical variables collected with a YSI multiprobe, greenhouse gas partial pressure (methane and carbon dioxide), total nutrient concentrations and dissolved metals (iron and manganese) are also included. We used this dataset to explore the distribution in time and space of various forms of mercury and pools of carbon across the dammed watershed of La Romaine and in different aquatic systems (e.g. tributaries, lakes, groundwater, river and reservoir sites).
Keywords	Mercury, methylmercury, dissolved organic carbon, PARAFAC, metals, hydroelectric reservoir
Dataset lead author	[Redacted]
Position of data author	graduate student
Address of data author	Département de sciences biologiques, Université de Montréal, Complexe des sciences, 1375 Avenue Thérèse-Lavoie-Roux, Montréal, QC H2V 0B3, Canada
Email address of data author	[Redacted]
Primary contact person for dataset	[Redacted]

Organize research data : Documentation

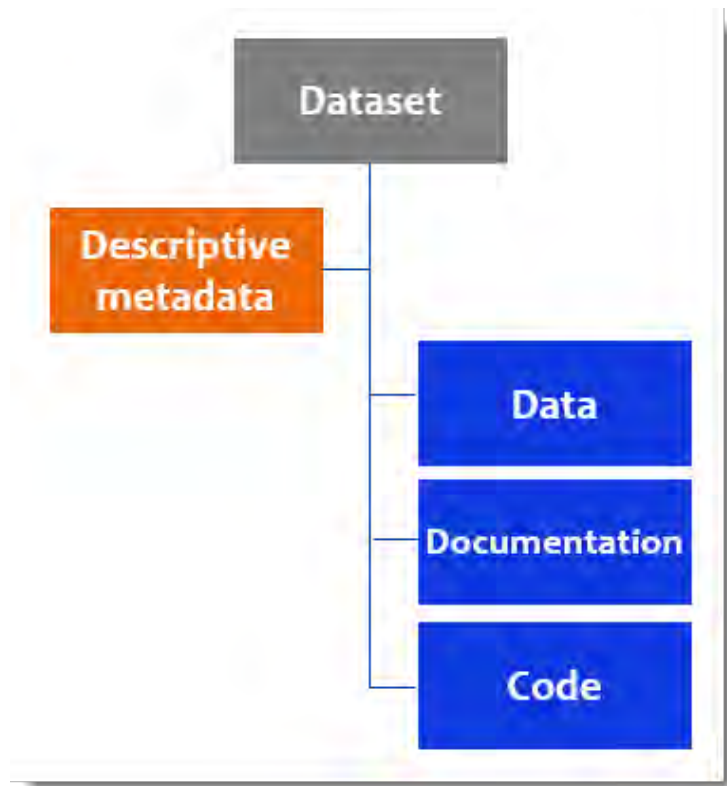


Digitalbevaring.dk

Organize for publication and reuse :

- Build documentation throughout the project.
 - Codebook
 - README File
 - Source codes (scripts that will read and process the data)
 - Metadata (which standard to use?)

Organize research data : Documentation



Metadata

Data that provides information on the nature of certain other data in order to facilitate their understanding and management.

In other words: Data about data

For humans and for machines!

Organize research data : Documentation

```
CCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAA
CTTGACAACCTGGAACCTTCTAGGAGACGACCAAATTTACAATGTAATCGTCACTGCCACGCCTTCG
TAATAATTTCTTTATAGTAATACCAATCATGATCGGTGGTTTCGGAAACTGACTAGTCCCCTCATAAT
CGGCGCCCCCGACATAGCATTCCCCGTATAAACATAAGCTTCTGACTACTTCCCCCATCATTCTT
TTACTTCTAGCATCCTCCACAGTAGAAGCTGGAGCAGGAACAGGGTGAACAGTATATCCCCCTCTCGCTG
GTAACCTAGCCCATGCCGGTGCTTCAGTAGACCTAGCCATCTTCTCCCTCCACTTAGCAGGTGTTTCCTC
TATCCTAGGTGCTATTAACCTTATTACAACCGCCATCAACATAAAACCCCCAACCTCTCCCAATACCAA
ACCCCCCTATTCGTATGATCAGTCCTTATTACCGCCGTCCTTCTCCTACTCTCTCTCCCAGTCCTCGCTG
CTGGCATTACTATACTACTAACAGACCGAAACCTAAACACTACGTTCTTTGACCCAGCTGGAGGAGGAGA
CCCAGTCCTGTACCAACACCTCTTCTGATTCTTCGGCCATCCAGAAGTCTATATCCTCATTTTAC
```

Roselin familier – Génétique

Il s'agit d'une séquence nucléotidique de *Carpodacus mexicanus* (clone 6b).
(A = Adénine, G = Guanine, C = Cytosine, T = Thymin : bases d'acide nucléique).

>Seq1 [organism=*Carpodacus mexicanus*] *C.mexicanus* clone 6b actin (act) mRNA, partial cds



Organize research data : Data storage and backup

Data horror stories ...

Have you ever lost or not been able to access data for any of these reasons?

- My lab was destroyed by fire/flood/eathquake
- My laptop was lost/stolen
- There was a hardware failure
- The file was corrupt
- I couldn't find software to open the file
- I couldn't find a computer to open the storage device

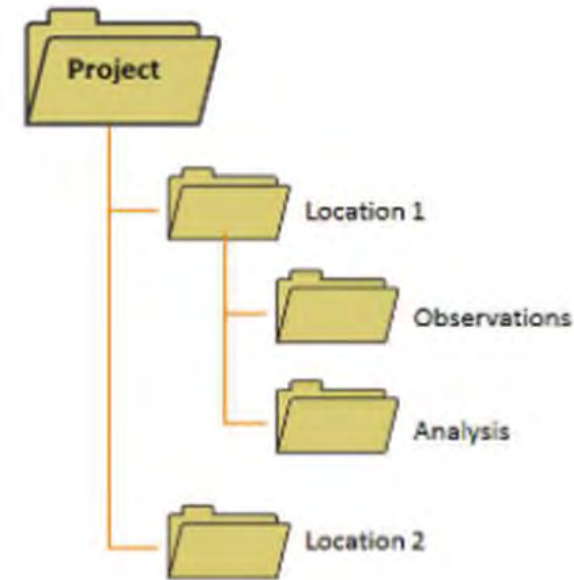


<https://tolkien2008.wordpress.com/2013/11/22/incendie-a-luniversite-de-montreal-22-novembre-1919/>,
consulté le 20 novembre 2022

Organize research data : Data storage and backup

Organize files and team collaboration during research:

- Evaluate how much space is needed et who needs access
- Agree on a naming convention and consider using a versioning tool (like GitHub)
- Regular backup and if possible automated
- 3-2-1 backup rule (3 copies of your data, on 2 different media and 1 backup copy offsite)
- Use an electronic Lab Notebook (ELN) to promote the traceability, publication and sharing of research data.



Organize research data : Sharing and reuse

- Sharing data
- From a FAIR perspective
- As mentioned in the Policy

"to provide appropriate access to the data where ethical, cultural, legal and commercial requirements allow, and in accordance with the FAIR principles and the standards of their disciplines."

From data collection to publication

FAIR principles (FAIR Data)



Organize research data : Sharing and reuse

Data repository / Dépôt de données

- Space for publishing research results.
- Often provides a unique and permanent identifier, such as the DOI.
- Rarely for storage and sharing between collaborators during research.
- Borealis UdeM (for deposit and distribution with or without restriction).



Extra

Some attractions of the Bibliothèque des sciences on the Campus MIL

Salle d'enregistrement sonore



Atelier de fabrication numérique



Studio d'édition numérique



Conclusion

Librarians are here to help !

- Consulting service with the Data Management Plan.
- Assistance in the choice of a data repository.
- UdeM Dataverse (Borealis) :
 - Assistance in using
 - Enhancement of descriptive metadata;
 - Support for data curation;
 - Support for standardization of data during research