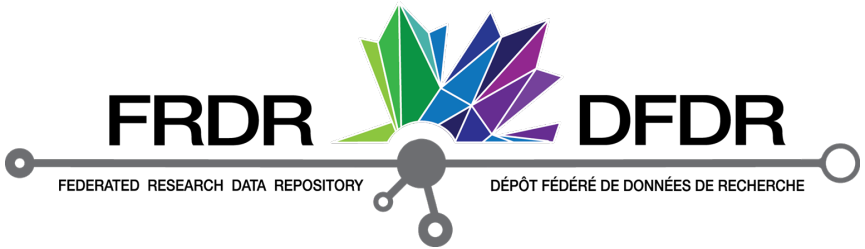
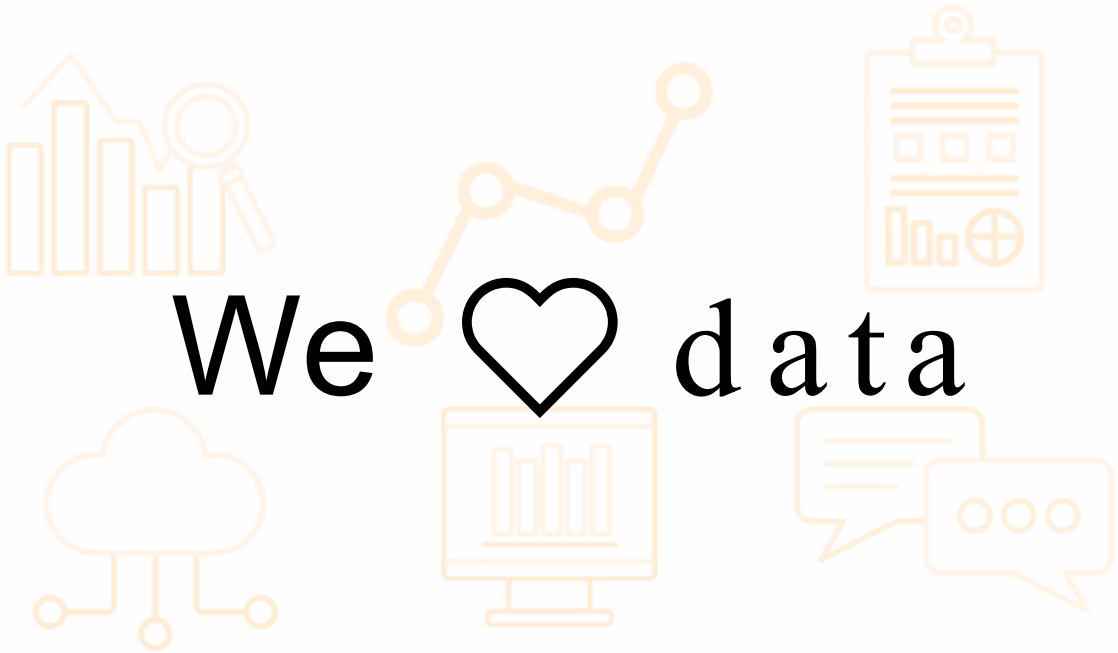
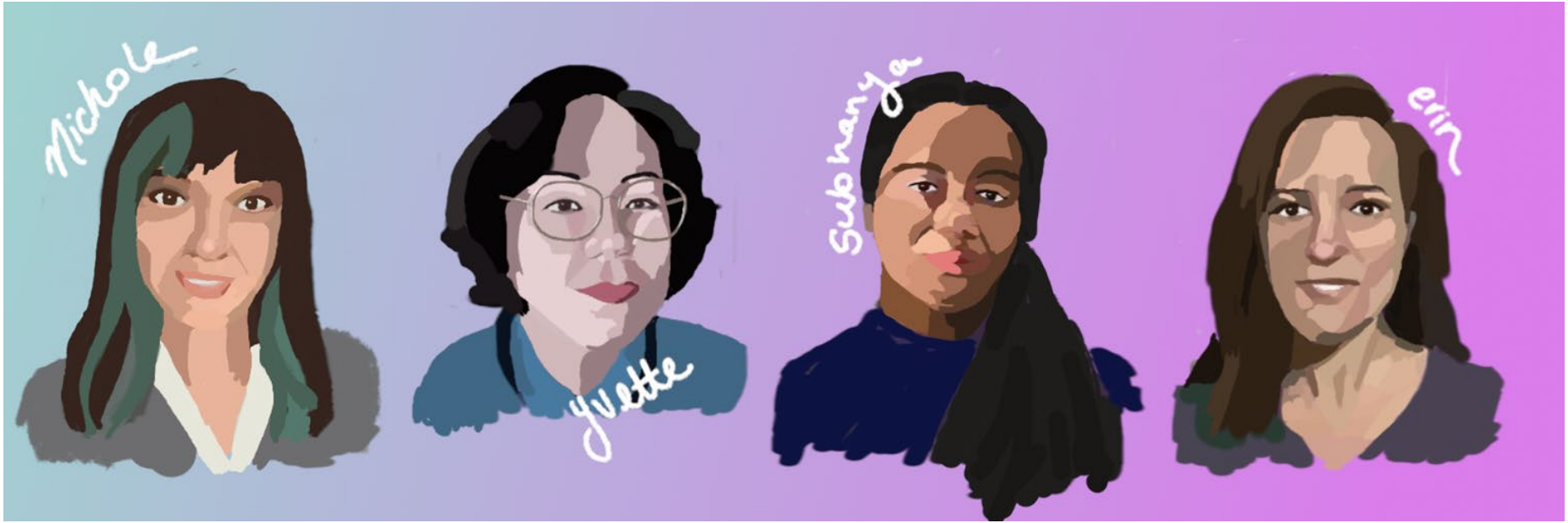
Le Dépôt fédéré de données de recherche
Federated Research Data Repository

Curation Services




We ♡ data

Meet the curators!



Federated Research Data Repository (FRDR)

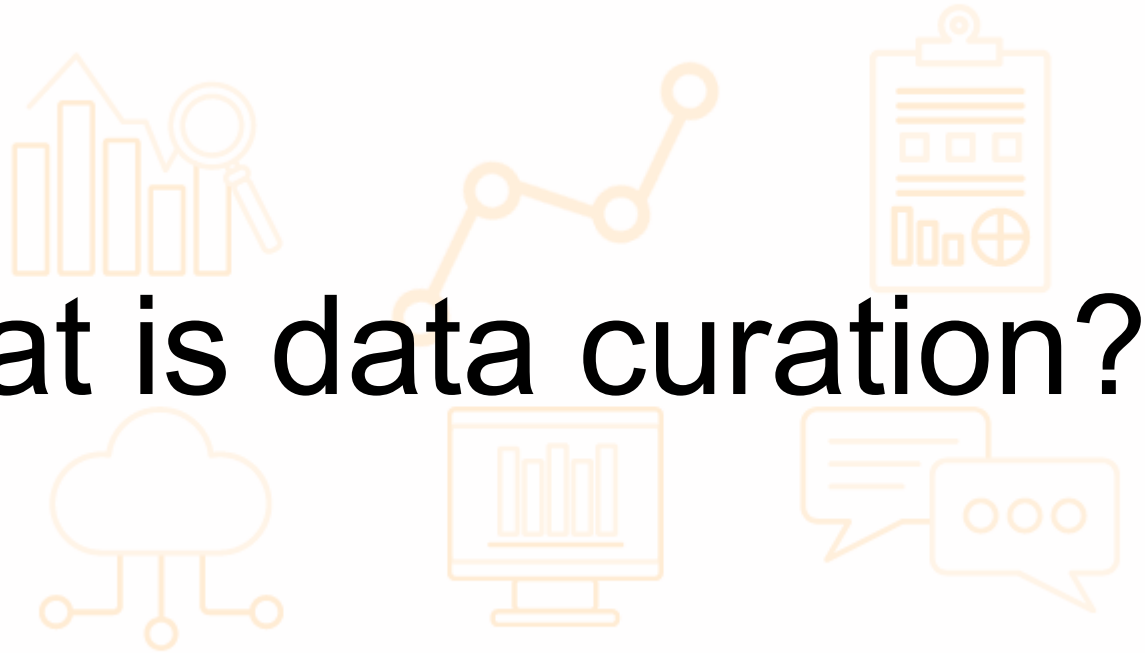
- FRDR is a discovery service and a general-purpose data repository for:
  - Canadian researchers
  - Data in any discipline
  - Datasets of all sizes (well suited for large datasets)
    - Standard allocation of 1TB per researcher
  - Raw or processed data, from past or present studies
  - Data associated with publications, or not

FRDR publication considerations

- Data will be published, so should be relatively unchangeable
- All data will eventually be publicly available
  - Temporary embargoes allowed
  - Data of a sensitive nature must be removed or masked
  - Please consider any data sharing agreements you must adhere to
- Restricted access data not currently accepted, but technical work and policy considerations are in early phases

What happens after deposit?

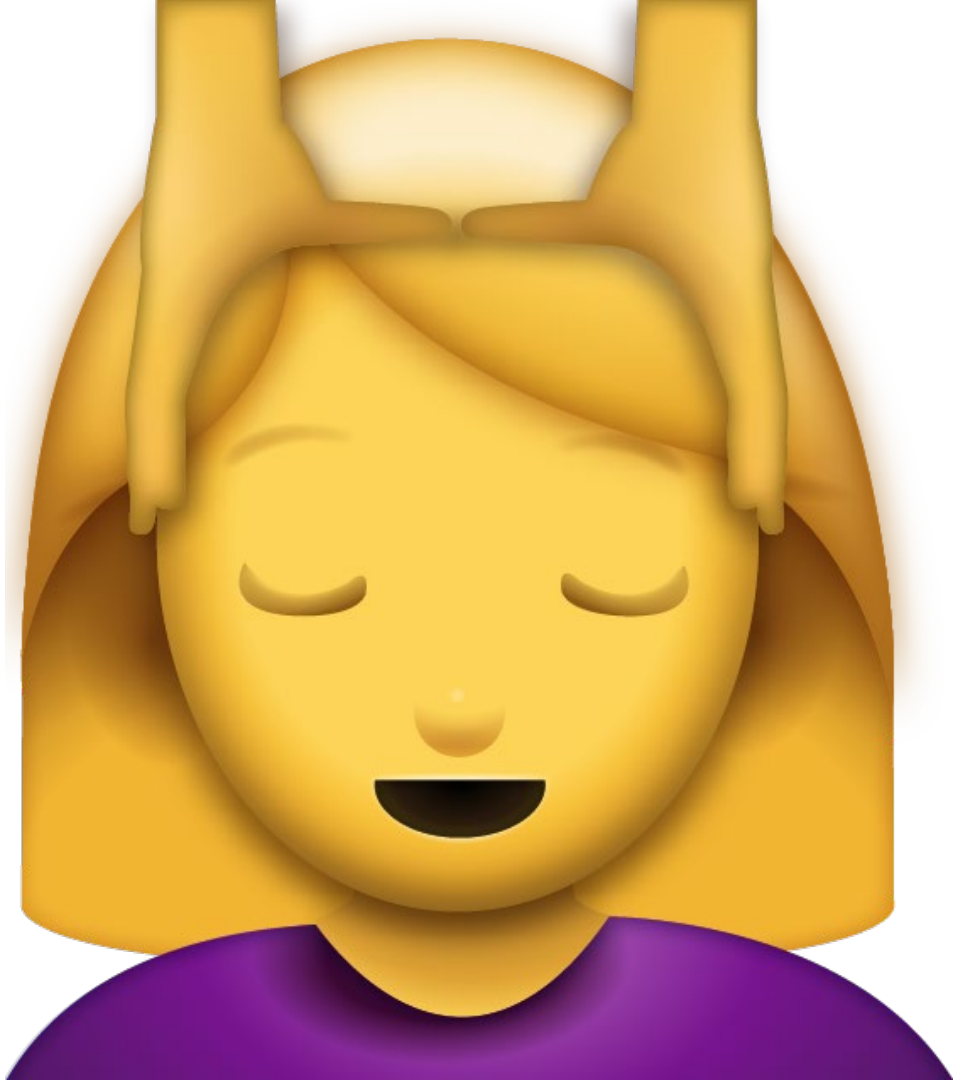
- Confirmation email with DOI (not yet registered)
- Dataset no longer available to edit
- Review by a member of our curation team
  - Approximately 2-5 business days
- Dataset publication!



What is data curation?

Start with the data.

Make it FAIRer.



Findable, Accessible, Interoperable, Reuseable

FAIR as a continuum

Hydrometeorological observations at three boreal forest sites (aspen, jack pine, and black spruce) located in central Saskatchewan, Canada

Contact Dataset Administrator

Full metadata record

DC Field	Value	Language
dc.contributor.author	Ahmed, Hafiz Faizan	
datacite.creator.affiliation	University of Saskatchewan	en_US
datacite.creator.nameIdentifier	https://orcid.org/0000-0002-7002-082X	en_US
dc.contributor.author	Helgason, Warren	
datacite.creator.affiliation	University of Saskatchewan	en_US
datacite.creator.nameIdentifier		en_US
dc.contributor.author	Barr, Alan G.	
datacite.creator.affiliation	Environment and Climate Change Canada	en_US
datacite.creator.nameIdentifier		en_US
dc.contributor.author	Black, T. Andrew	
datacite.creator.affiliation	University of British Columbia	en_US
datacite.creator.nameIdentifier		en_US
dc.date.accessioned	2020-10-29T00:30:41Z	
dc.date.available	2020-10-29T00:30:41Z	

Findable, Accessible, Interoperable, Reuseable

F

Is the data easy for both humans and computers to **find** ?

persistent **identifier**

rich **metadata**

indexed in a searchable resource.

A

Once the user finds the required data, can they **access** it?

mediated **without specialised or proprietary** tools or communication methods

provide the **exact conditions** under which the data are accessible

I

Can the data be integrated with other data?

formal, accessible, shared language for knowledge representation

references to other (meta)data

R

Is the data **reproducible** ?

adequate **context**

accurate and relevant attributes

clear and accessible **data usage licence**

detailed **provenance**

paraphrased from Bishop, B. W., & Hank, C. (2018). Measuring FAIR principles to inform fitness for use.

Machine & human discoverable & useable

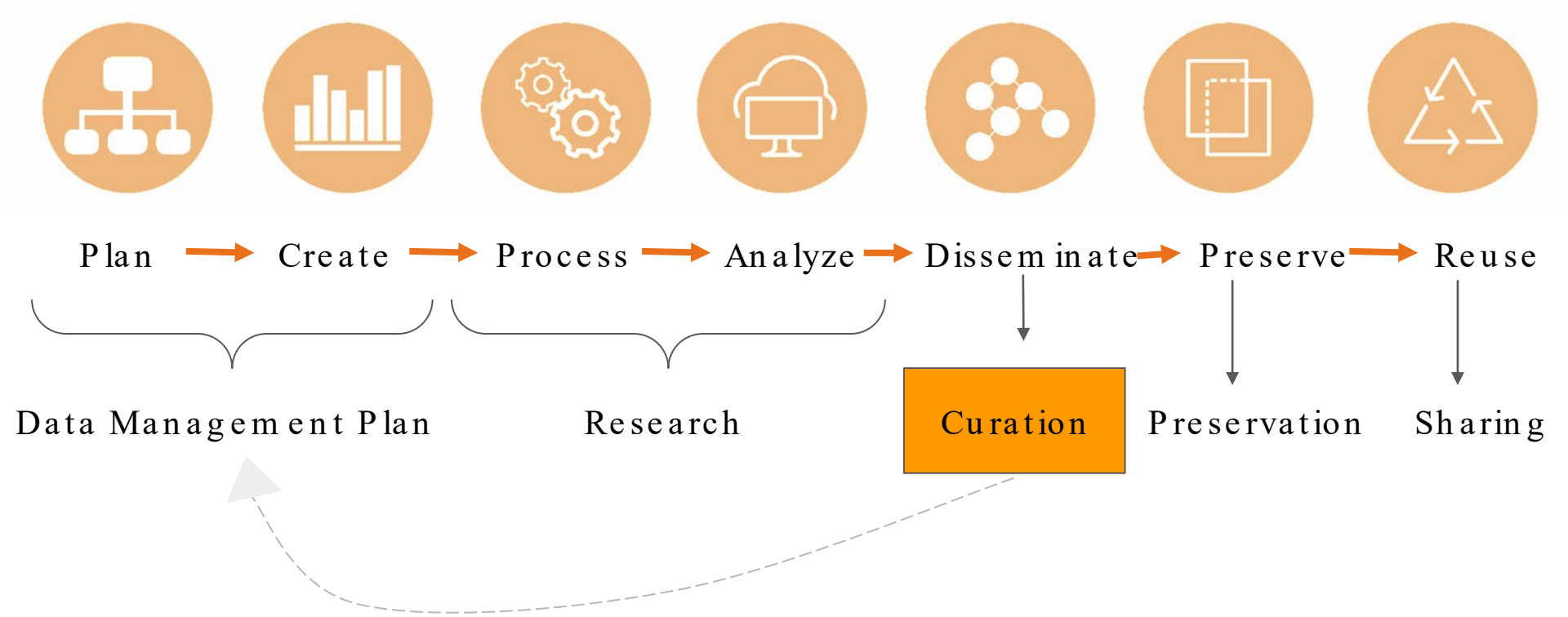
Machines are good at discovering data that have rich metadata, controlled vocabularies, persistent identifiers, and are indexed in a searchable repository.

Machines are good at using data that are in open formats, well-structured (e.g. tidy data)

Humans are good at discovering data that have descriptive metadata, including fields that may not be indexed. README files are essential!

Humans are good at using data that have clearly-defined variables, instructions, license terms, citations, comments etc.

Curation throughout the research process



Curation as part of Research Data Management

- Ensures the dependability, accuracy, and validity of data
  - Enhances the replicability of research
- Promotes organization
- Provides transparency and accountability

Curation's added value

- Reinforces all the added value of RDM, more broadly
  - Improves transparency
  - Renders research more likely to be reproducible
  - And more likely to be properly cited
- Sharing data means more citations: more sharing, more knowledge
  - Up to 25% citation advantage when data is shared alongside an article!
- Cost-savings

Curation at FRDR

Check files

Understand dataset

Request changes

Augment metadata

Transform file formats

Evaluate FAIRness

Document changes

Icon from FreePik

In practice

- Arrangement and description
- Code review
- Discovery services
- Documentation
- Evaluate FAIRness
- File format transformations
- File inventory
- Metadata

For more info see Johnston et al. (2016) *Definitions of Data Curation Activities used by the Data Curation Network*. http://hdl.handle.net/11299/188638

Check & Understand

Files in Dataset - Total Size - 893.38 KB

01_wow_art.tiff	149.14 KB
02_mysterious_art.tiff	123.54 KB
03_landscape_art.tiff	620.66 KB
data.csv	0.02 KB
README.txt	0.01 KB

Request & Augment

dc.contributor.author	Rancourt, Y.G.	
datacite.creator.affiliation	Portage	en_US
datacite.creator.nameIdentifier	https://orcid.org/0000-0002-9895-8668	en_US
dc.description	A selection of Paint 3D works, created for this demo.	en_US
dc.description.provenance	Submitted by Y.G. Rancourt	en
dc.rights	Creative Commons Attribution 4.0 International (CC BY 4.0)	en_US
dc.rights.uri	https://creativecommons.org/licenses/by/4.0/	en_US
dc.subject	test	en_US
dc.subject	data	en_US
dc.subject	art	en_US
dc.subject	Paint 3D	en_US
dc.title	TEST Paint 3D Masterpieces	en_US

Transform & Evaluate

Document changes & publish datasets

Photo by Pineapple Supply Co. on Unsplash

Tips & Tricks

Documentation

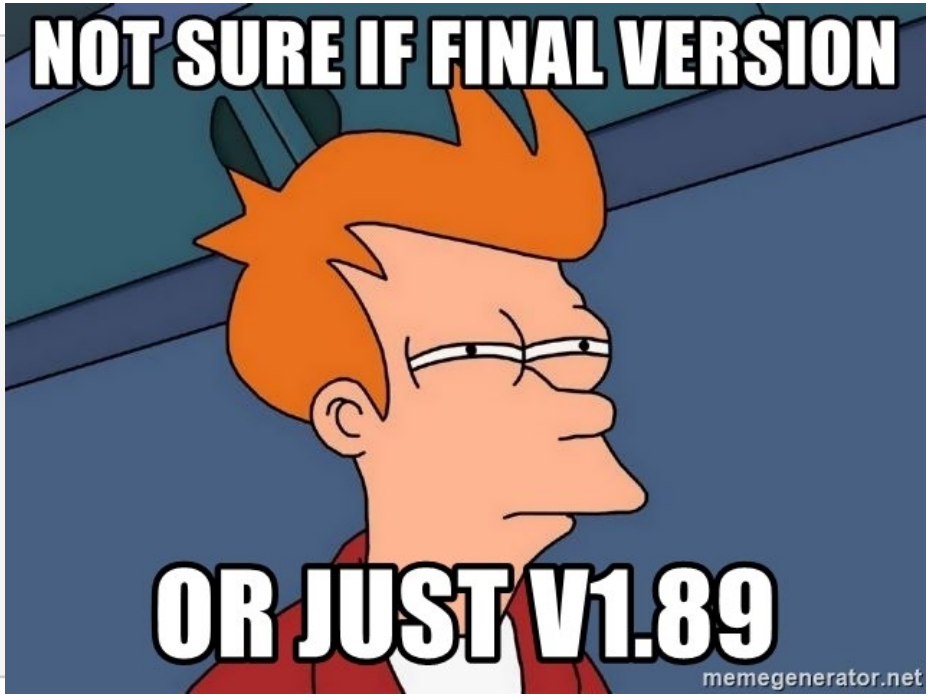
```
This README.txt file was generated on YYYY-MM-DD by NAME

## PLEASE NOTE: Help text is preceded by ## and should be deleted before
saving this file!
## This is a template, and you are free to modify it. Please remove any
sections or items that are not applicable to your dataset. Please add
additional information not currently represented in the template that is
needed to correctly interpret or reuse your data.

--------------------
GENERAL INFORMATION
--------------------

1. Title of Dataset:

2. Author Information
        A. Principal Investigator Contact Information
                Name:
                Institution:
                Email:

        B. Associate or Co-investigator Contact Information
                Name:
                Institution:
                Email:


3. Date of data collection (single date, range, approximate date):

## suggested format YYYY-MM-DD

4. Geographic location of data collection:

## latitude, longitude, or city/region, province, country, as appropriate

5. Information about funding sources that supported the collection of the
data:
```

Metadata

Element	Notes	E.g.
Title	Be concise and descriptive. Choose something appropriate for use in a data citation.	3D mapping for surveying Alpha Onias III using an Unmanned Aerial Vehicle (UAV) system vehicle
Author	Dataset creators. Be prepared to provide author affiliations and ORCIDs.	Dax, Jadzia. University of Toronto. https://orcid.org/0000-0003-2144-3423
Description	It may help to think of the dataset description as you would an article abstract. This is a high-level summary that addresses the nature and scope of the dataset.	Unmanned Aerial Vehicles (UAV) have had recent widespread application to capture high resolution information on volcanic gasses and the data herein was collected to address viability for terraforming.
Subject	Keywords that describe the dataset.	Unmanned aerial vehicle; sulphur; non-toxic gas; onias; methane;
Related identifier	Link to associated publications, code, protocols, methods, and other information that gives context to the data.	https://doi.org/10.5194/tc-14-1919-2359
Rights	Terms of use. The terms you can assign might be limited if your data are derived from external sources.	Creative Commons Attribution 4.0 International (CC BY 4.0)
Funder	Granting agency or funder(s) that supported your research.	Natural Sciences and Engineering Research Council of Canada (NSERC)

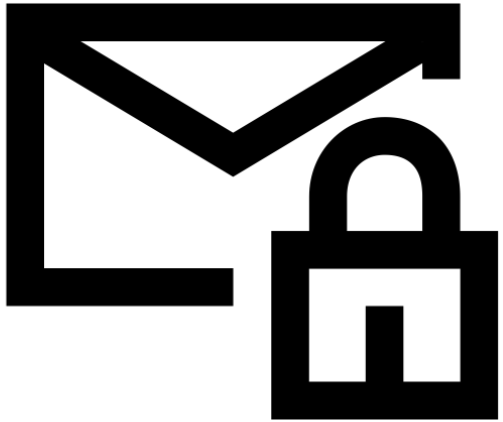
File naming, format & structure

```
├── Documentation
|   ├── site_information.csv
|   ├── site_1.shp
|   └── site_2.shp
├── Data
|   ├── year_01
|   |   ├── site_1.csv
|   |   └── site_2.csv
|   └── year_02
|       ├── site_1.csv
|       └── site_2.csv
└── README.txt
```



Potential sensitivities

- Data from third party source
- Sensitive location information
- Data from human participants



Created by Chameleon Design
from Noun Project

Thank you! Questions?

FRDR Curation Team:
Erin Clary, Nichole DeMichelis, Yvette Rancourt, Subhanya Sivajothy

support@frdr-dfdr.ca

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

compute canada | calcul canada

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES | ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

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

Innovation, Science and Economic Development Canada
Innovation, Sciences et Développement économique Canada

NDRIO | NOIRN
New Digital Research Infrastructure Organization
Nouvelle organisation d'infrastructure de recherche numérique