Beginning With the End in Mind: Building Documentation and Metadata to support Data Deposit and Preservation

Erin Clary (Curation Coordinator, Portage Network)
Krysha Dukacz (Data Manager, Global Water Futures)

September 29, 2020
The “Rime of the Modern Researcher”

Data, data everywhere… but where is the bit I need?

… can I figure out the coding system?

.... is it in a format I can use?

... is it appropriate to my purpose?

... am I allowed to share it?
The Call for Sharing Publicly Funded Data

Journalists need to know more about COVID-19 transmission, for the sake of all of us
(Sylvia Stead, The Globe and Mail, September 26, 2020)

- No national data standards and needing to source data from provinces, territories and regions.
- Took 10 days to collect data and make it comparable
- Availability of data is spotty, constraints and variables are defined differently
- Variability in data delivery formats: Excel, PDFs, manual copy/paste from mapping application
The Call for Sharing Publicly Funded Data

A new flow for Canadian young hydrologists: Key scientific challenges addressed by research cultural shifts


- Data collection often focused on long term sites, concentrating data control because of costs
- “Hidden” data: unprocessed, fragmented between publications, proprietary, distributed among various government bodies. Data are difficult to find (word of mouth, who you know)
- Support for data sharing, “including appropriate metadata is essential to effective data-sharing”
Funder Driven Requirements - Show me the money

Tri-Agency Statement of Principles on Digital Data Management:

Promote excellence in digital data management practices and data stewardship in agency-funded research

Researchers expected to consider:

<table>
<thead>
<tr>
<th>Data Management Planning</th>
<th>Constraints and Obligations (commercial, legal, ethical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adherence to Standards</td>
<td>Collection and Storage</td>
</tr>
<tr>
<td>Metadata</td>
<td>Preservation, Retention and Sharing</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Acknowledgement and Citation</td>
</tr>
<tr>
<td>Efficient and Cost Effective</td>
<td></td>
</tr>
</tbody>
</table>


Impending Tri-Agency RDM policy

Encourage a culture of strong data management associated with research excellence

Increase the Canadian researchers recognized/rewarded for data as a valued product of research

Equip Canadian researchers to engage in international research

Increase institutional capacity to support the management of the data produced by researchers

Increase ability for research data to be archived

Reproducibility, Discoverability, Reusability

https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97610.html
The Cost of Data

Cost of Data

Personnel (salaries, benefits)

Infrastructure and Logistics (instrumentation, travel costs, processing, incentives)
  Data security, storage, backup
  Data cleaning
  Data Management

Blood, sweat and tears

Sometimes irreplaceable (one-time opportunity to collect)
The Value of Data

Value of Data

- The value of data is in the information that can be derived from its use.
- The information contained in data requires documentation to provide context
- Value is augmented by Reuse

What is Metadata?

“**Metadata** are a subset of core data documentation, which provides standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection” (UK Data Archive).

**Documentation**
- Critical component of a complete dataset

**Metadata**

<table>
<thead>
<tr>
<th>INFORMATION</th>
<th>ABOUT DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WHO?</strong></td>
<td>- Collected the data?</td>
</tr>
<tr>
<td></td>
<td>- Processed the data?</td>
</tr>
<tr>
<td></td>
<td>- To contact for more information?</td>
</tr>
<tr>
<td></td>
<td>- Owns the data?</td>
</tr>
<tr>
<td><strong>WHAT?</strong></td>
<td>- Are the data about?</td>
</tr>
<tr>
<td></td>
<td>- Parameters were measured?</td>
</tr>
<tr>
<td></td>
<td>- Format are the data in?</td>
</tr>
<tr>
<td></td>
<td>- Is the data quality?</td>
</tr>
<tr>
<td></td>
<td>- Are the use constraints?</td>
</tr>
<tr>
<td></td>
<td>- Are the appropriate uses?</td>
</tr>
<tr>
<td><strong>WHEN?</strong></td>
<td>- Were the data collected?</td>
</tr>
<tr>
<td></td>
<td>- Were the data processed?</td>
</tr>
<tr>
<td><strong>WHERE?</strong></td>
<td>- Were the data collected?</td>
</tr>
<tr>
<td></td>
<td>- Are the data held?</td>
</tr>
<tr>
<td><strong>WHY?</strong></td>
<td>- Were the data collected?</td>
</tr>
<tr>
<td><strong>HOW?</strong></td>
<td>- Were the data collected / processed?</td>
</tr>
<tr>
<td></td>
<td>- Do I access the data?</td>
</tr>
<tr>
<td></td>
<td>- Were the data quality assessed?</td>
</tr>
</tbody>
</table>

Need to Know for each dataset:

- Where it is
- How to access it
- What it can be used for
- Known issues/Quality
- Collection Methods
- Ethical /Privacy issues
- Licensing
- How to cite

ftp.ncddc.noaa.gov/pub/Metadata/DISL_June30
Beginning with the End in Mind…

How can the value of data be protected…

- Figure out what is needed in the end and plan out the steps to get there.
- Make a plan - Formal Data Management Plan (DMP) or informal process
- Reuse already contained in your research plan.

DMP considers:

- Ethics
- Data collection
- Licensing
- Metadata
- Documentation
- Sharing
- Storage

DMP exemplars will be available on the Portage Network site soon. Excerpt courtesy of Bhaleka Persaud and the University of Waterloo Ecohydrology Research Group.
Benefits of Well-documented Data

<table>
<thead>
<tr>
<th>Researcher</th>
<th>Publisher</th>
<th>Funders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>Rigorous peer review</td>
<td>Transparency</td>
</tr>
<tr>
<td>Ease of collaborative work</td>
<td>Validation and reproducibility</td>
<td>Accountability</td>
</tr>
<tr>
<td>Credit and impact</td>
<td>Defensible authority</td>
<td>Return on Investment</td>
</tr>
</tbody>
</table>
Principles to consider

- DOIs for Publication
- Funding Compliance
- Retraction Avoidance

Sharing with others (just metadata piece or whole datasets)

Personal accomplishment and reliability, journal publication

Today, I have accomplished something in my career of which I am more proud than even passing my defence, or tenure & promotion. Today, I opened a program that I wrote a year ago, & it still totally makes sense to me AND does exactly what I think it should.

*Tweet used with permission
Tools for Building Metadata

- Standard Operating Procedures
- Log files or Field Notes
- File Organization
- File Naming
- Analysis Scripts
- README
- Metadata standards (General and Disciplinary)
- Collaborative platforms
Standard Operating Procedures

Standard Operating Procedures (SOPs) provide step by step instructions on how to successfully complete a task. This enables consistency over time and among different people. Common examples include:

- **Field Visit SOP** may include what to bring, tasks that need to be completed at each site (download data, check sensors, take sample, potential issues/solutions)
- **Instrument Setup** - location, connectivity, calibration, expected output/QA/QC, troubleshooting
- **Data Management** - Folder Structures; Protocols for uploading field data; Storage locations (3-2-1 backup); File Naming, Transition expectations; Data Acquisition steps; Data upload
Preparing Rating Curves for Discharge Estimation

Discharge cannot be measured continuously with the sensors used in the STREAMPULSE project. At monitoring sites, a set of discharge rating curves is available to estimate discharge from water level data.

Velocity should be measured in various flow stages from low summer base flow to high flow in order to achieve the best estimation of peak discharge. Velocity can be estimated using either of the two ways listed below.

Preparation of rating curves:

1. Choose a site with steady flow and measure water velocity at different water levels.
2. Measure velocity for each water level, noting the corresponding discharge.
3. Plot velocity against discharge to obtain a rating curve.
4. Use the rating curve to estimate discharge from water level data.

Preparation of velocity rating curves:

1. Measure water level and velocity at the site.
2. Plot water level against velocity to obtain a rating curve.
3. Use the rating curve to estimate velocity from water level data.
4. Use the velocity data to estimate discharge from the rating curve.

Field Notes and Log files

- Various methods such as collector apps, field books, field notes, site reports
- Can be manual or electronic

See also:
Great comparison of Electronic Lab Notebooks from Harvard Biomedical Data Management. [https://datamanagement.hms.harvard.edu/electronic-lab-notebooks](https://datamanagement.hms.harvard.edu/electronic-lab-notebooks)
Field Notes and Log files - Examples and Resources

Field Notes capture the who, what, where, why, and how of data collection.

Standard Operating Procedures can be included for easy reference.

Field book printed on Rite in the Rain paper and bound for a cohesive record.
Good ideas:

- Use a **ToFile** folder to collect items that don’t have a place yet or if you are short on time.
- Be aware of **maximum path length**.
- Avoid spaces and special characters which programs and code may not interpret correctly.
- For teams and labs, create an **empty file structure that can be shared** to enhance consistency or use a script to generate the structure.

- Increases findability and efficiency.
- Prompts for critical data management elements (RawData, Licensing etc.).
In terms of a filing system, we suggest starting with the following file structure somewhat like the following:

- Clubs (e.g., NREM GSO, SASA, Grebe)
- Courses (e.g., NREM507, SUSTAG509)
- FieldNotes
- Miscellaneous
- Personal
- Project
  - Analyses
    - AnalysisX
    - AnalysisY...
  - GrantDocumentation-Reports
  - Images
  - MeetingNotes
  - Presentations
  - Proposal
  - RawData
  - Readings
  - Thesis/Dissertation
    - Chapter 1
    - Chapter2....

For further guidance: NIST Electronic File Organization Tips:
File Naming Guidelines

- Balance between concise and descriptive. May contain:
  - Project name, site name, type of data or analysis
  - Date (YYYYMMDD format)
  - Version number (v_01, v_02, etc.)

- Avoid using:
  - Spaces (use - or _ instead)
  - Special characters (~ ! @ # $ % ^ & * ( ) ; < > ? , [ ] { } ' " and | )
  - Potentially sensitive or restricted information

- Lead numbers with a zero (e.g, 001, 002 ... 010, 011 ... 100, 101)
- Try to make filenames unique (directory structure may change over time)
- Document your naming conventions to help with consistency

Resources: UBC [File Naming Guidelines](https://example.com), Stanford [Best Practices for File Naming](https://example.com)
File Naming - Example

Uploading data for core sites
Core sites can upload raw datalogger files and/or pre-formatted data.

File naming
Name your upload file - REGIONID_SITEID_YYYY-MM-DD_LOGGERID_xxx - where

REGIONID is the name of your region (PI, FL, NC, WI, or AZ),
SITEID is your unique site name,
YYYY-MM-DD is the download date,
LOGGERID is the logger routing code,
CS: CR1000 data file
HD: Hobo DO logger
HP: Hobo water pressure logger
HA: Hobo air pressure logger
HP: Hobo light pendant logger
EM: Eureka Manta logger
XX: Calibrated and formatted data
and .xxx is the file extension.
Extension must be .dat if LOGGERID is CS, Otherwise must be .csv.
Note: LOGGERID may include numbers if you have several of the same logger type at one site. For example, LOGGERID could be CS4 or HA12.
Example filename: NC_fmo_2017-12-06_HP.csv

Calibrated and formatted data
You can upload raw data (from the datalogger) and/or calibrated data (e.g., turbidity in NTU, water level or discharge, etc.) at the same time.

If you modify a datalogger file to generate calibrated and derived variables, you must save it as a .csv with the _XX.csv extension (see LOGGERID above),
one header row followed directly by data rows, one row per timestamp,
the first column as a Date-Time stamp converted to UTC standard time and formatted as YYYY-MM-DD HH:MM:SS, and additional columns for each data variable.

E.g., http://pulseofstreams.weebly.com/uploading_data.html
Scripts and syntax files

- If possible use a scripted language to process and analyze your data
- Promote transparency and reproducibility
- Don’t alter your raw data directly
  - Input raw data -> generate new file as output
- Include comments
  - These make each line of code human readable and the function is clear
- Keep track of any external dependencies
  - List name and version of packages or libraries you use

Analysis Scripts - Example

Digital Elevation Map (DEM) and derived products, slope and aspect. Runtime around 5-10 minutes

After running 'src_borders.R' we now have a set of polygons ('snrc_std.shp'), loaded as snrc.sf, below with which to split the data into blocks (ie mapsheets). The path of this shapefile is stored in the cfg.borders metadata list, which can be loaded from the file 'borders.rds' located in the data.dir directory.

```r
# download again to rewrite existing files?
force.download = FALSE

# load the helper functions
library(here)
source(here('utility_functions.R'))

# load the borders info and shapefiles for mapsheets
cfg.borders = readRDS(here('data', 'borders.rds'))
snrc.sf = st_read(cfg.borders$out$fname$shp['snrc'])
snrc.codes = cfg.borders$out$code

As with the 'borders' collection, we will fill in the metadata as we go, and save a copy to disk at the end

# create the source data directory, metadata list and its path on disk
collection = 'dem'
Cfg = MF8_metadata(collection)
```

Data in FRDR at Koch DC, Lewis MA (2020) Raster datasets relevant to mountain pine beetle outbreak ecology in the province of British Columbia. Federated Research Data Repository. [https://doi.org/10.20383/101.0283](https://doi.org/10.20383/101.0283) with code used to derive the files in GitHub at [https://github.com/deankoch/rasterbc_src](https://github.com/deankoch/rasterbc_src) and code to import a subset of data into R for modelling at [https://github.com/deankoch/rasterbc](https://github.com/deankoch/rasterbc).
ReadMe files

- General information
  - Dataset name, project summary, contact information, date range, funding sources,
- Data and file overview
  - File name, description of content, format, notes
- Sharing and access information
  - Limits of use, how to access data, citation
- Methodological information
  - Instrumentation, links to protocols and manuals, QA/QC, processing details
- Data-specific information
  - Variable names, units, missing data codes
- Details of anomalies necessary for interpreting the data and its suitability for use

Further guidance is available in UBC’s [Quick Guide: Creating a README for your dataset](#) and Cornell University’s [Guide to writing "readme" style metadata](#) (template available for download).
ReadMe files - Examples and Resources

# bpwtpR: Buffalo Pound Water Treatment Plant (BPWTP) data integration & reporting

## Summary

This project was completed as part of the MITACS Accelerate program (IT-1756) and the CREFI (Global Water Futures) through MITACS sponsored by Venkiteswaran and Buehler.

### Highly Qualified Personnel
- Megan L. Larsen, Wilfrid Laurier University, mlarsen@wlu.ca, ORCID 0000-0002-4155-1329
- Arti Goyal, Trent University, angel13@trentu.ca, ORCID 0000-0002-6674-7671

### Principle Investigators
- Jason Venkiteswaran, Wilfrid Laurier University, jvenkiteswaran@wlu.ca, ORCID 0000-0002-4155-1329

### Data sources
- Provide links to any data used from external providers and/or submodules and their associated management scheme (e.g. ISO-14001, MWH).

### Funding sources
- MITACS Accelerate Program (IT-1756)
- CREFI/MITACS

### Keywords

List keywords separated by commas

### Additional information and support
- Sensitive Data Flag: Human Participants: NO
- Sensitive Data Flag: Indigenous Partnerships: NO
- Sensitive Data Flag: Government Partnerships: NO
- Access Restrictions

---

*Github example courtesy of Dr. Jason Venkiteswaran (Wilfrid Laurier University)

---

### Variables

Potential variables include:

- Date-Time (UTC)
- DO (mg/L)
- Saturation DO (mg/L)
- Water Temperature (°C)
- Water Pressure (kPa)
- Air Temperature (°C)
- Air Pressure (kPa)
- Depth (m)
- Discharge (m3/s)
- Velocity (m/s)
- Light, PAR (µmol/m2/s)
- Light, lux
- Specific Conductivity (µS/cm or µS/cm)
- pH
- fDOM (frac)
- fDOM (mV from sensor)
- Turbidity (NTU)
- Turbidity (mV from sensor)
- Nitrate (mg/L)
- CO2 (ppm)
### Metadata Standards - General Purpose

<table>
<thead>
<tr>
<th>Element</th>
<th>Notes</th>
<th>E.g., <a href="https://doi.org/10.20383/101.0193">https://doi.org/10.20383/101.0193</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Be concise and descriptive. Choose something appropriate for use in a data citation.</td>
<td>Unmanned aerial vehicle structure from motion and lidar data for sub-canopy snow depth mapping</td>
</tr>
<tr>
<td>Author</td>
<td>Dataset creators. Be prepared to provide author affiliations and ORCID(s).</td>
<td>Harder, Phillip. University of Saskatchewan. <a href="https://orcid.org/0000-0003-2144-2767">https://orcid.org/0000-0003-2144-2767</a></td>
</tr>
<tr>
<td>Description</td>
<td>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.</td>
<td>Unmanned Aerial Vehicles (UAV) have had recent widespread application to capture high resolution information on snow processes and the data herein was collected to address the sub-canopy snow depth challenge...</td>
</tr>
<tr>
<td>Subject</td>
<td>Keywords that describe the dataset.</td>
<td>Unmanned aerial vehicle; lidar; snow depth, point cloud, digital surface model</td>
</tr>
<tr>
<td>Spatial coverage</td>
<td>Place where data was gathered. Place name, geographic coordinates, and/or bounding box.</td>
<td>Fortress Mountain Snow Laboratory (Fortress), Kananaskis, AB; 50.833 -115.220</td>
</tr>
<tr>
<td>Related identifier</td>
<td>Link to associated publications, code, protocols, methods, and other information that gives context to the data.</td>
<td><a href="https://doi.org/10.5194/tc-14-1919-2020">https://doi.org/10.5194/tc-14-1919-2020</a></td>
</tr>
<tr>
<td>Rights</td>
<td>Terms of use. Note: the terms you can assign might be limited if your data are derived from external sources.</td>
<td>Creative Commons Attribution 4.0 International (CC BY 4.0)</td>
</tr>
<tr>
<td>Funder</td>
<td>Granting agency or funder(s) that supported your research.</td>
<td>Natural Sciences and Engineering Research Council of Canada (NSERC)</td>
</tr>
</tbody>
</table>
General Purpose Elements Aid Discovery
## Metadata Standards - Disciplinary

<table>
<thead>
<tr>
<th>Standard</th>
<th>Use for</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 19115</td>
<td>Geographic information in many disciplines (e.g., Glaciology, Hydrogeology, Oceanography, Climatology)</td>
<td>mdEditor <a href="https://www.mdeditor.org/">https://www.mdeditor.org/</a> (also supports FGDC)</td>
</tr>
<tr>
<td>Ecological Metadata Language (EML)</td>
<td>Ecology and environmental sciences</td>
<td>Morpho <a href="https://knb.ecoinformatics.org/tools/morpho">https://knb.ecoinformatics.org/tools/morpho</a></td>
</tr>
<tr>
<td>Data Documentation Initiative (DDI)</td>
<td>Social, behavioral, and economic sciences</td>
<td>Many options available at <a href="https://ddialliance.org/resources/tools">https://ddialliance.org/resources/tools</a></td>
</tr>
<tr>
<td>Crystallographic Information Framework (CIF)</td>
<td>Crystallographic and related structured data</td>
<td>Many options available at <a href="https://www.iucr.org/resources/cif/software">https://www.iucr.org/resources/cif/software</a></td>
</tr>
</tbody>
</table>

Collaborative platforms

Collaborative tools provide a platform for sharing procedures, data, code and other project related information. The helps to create a common understanding and identify authoritative versions.

- Google (forms and documents)
- Open Science Framework
- GitHub
- CEDAR Workbench

*GitHub example courtesy of Dr. Jason Venkiteswaran (Wilfrid Laurier University) [https://jjvenky.github.io/rules/](https://jjvenky.github.io/rules/) ; [https://github.com/biogeochem/musical-funicular](https://github.com/biogeochem/musical-funicular)*
<table>
<thead>
<tr>
<th>Product</th>
<th>Notes</th>
<th>Storage limit (free)</th>
<th>Getting Started</th>
</tr>
</thead>
</table>
| **Google Drive** | - Set up shared Drive for your research project  
- Develop shared documentation, and adjust settings to allow offline editing of documents  
- Use forms to standardize metadata acquisition                                                                                     | 15 GB                | [https://www.google.ca/forms/about/](https://www.google.ca/forms/about/)                           |
| **OSF**        | - Option to set global storage location to Montréal so your documents are hosted on Canadian servers  
- Keep your full project private to your team, or choose to share individual components of the project more widely  
- Use storage add-ons to connect your project to Drive, GitHub and other external services | Unlimited            | OSF guidance at [https://help.osf.io/hc/en-us](https://help.osf.io/hc/en-us)  
OSF instructional videos [https://osf.io/3yw4y/wiki/home/](https://osf.io/3yw4y/wiki/home/)  
UBC Research Commons [upcoming OSF workshop](https://osf.io/hc/en-us), Oct 5 (open to public). |
| **GitHub**     | - Can be used for more than software - also use for creating and maintaining documentation  
- Built for collaboration and version control  
- Free basic accounts for teams (unlimited public/private repositories, unlimited collaborators) [https://github.com/pricing](https://github.com/pricing) | 500 Mb               | Jason Brodeur’s [upcoming webinar](https://www.google.ca/forms/about/) for the GWF/Portage webinar series, Oct 6. 1:00 PM EDT.  
Git cheat sheets [https://training.github.com/](https://training.github.com/) (Eng & Fra) and GitHub’s beginner guide [Hello World](https://librarycarpentry.org/lc-git/)  
Library Carpentry Introduction to Git [https://librarycarpentry.org/lc-git/](https://librarycarpentry.org/lc-git/) |
| **CEDAR Workbench** | - Create rich web-based metadata-acquisition forms  
- Pre-configure forms, and pre-set your team’s defaults  
- Templates can be versioned  
- APIs allow you to import existing metadata or export metadata  
- Developed for biomedical sciences, but you can create your own templates                                                                 | N/A                  | CEDAR user guide [https://metadatacentre.github.io/cedar-manual/](https://metadatacentre.github.io/cedar-manual/) |
In Summary:

● Have a data management plan that addresses documentation and metadata
● Budget time
  ○ Add a buffer so metadata isn’t edged out by other competing priorities!
● Budget resources
  ○ Metadata creation takes people and time. If you are writing a new grant proposal, consider budgeting staff time for metadata and other data management tasks
  ○ OpenAIRE RDM Costing Tool
    https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs
● Integrate into workflow
● Collaborative effort
Checklist for success ...

- Use SOPs to help you manage things consistently.
- Follow best practice guidance to name and structure files.
- Create a README, codebook, and other documentation necessary to understand your data and interpret it correctly.
  - Document data collection instruments, methods, quality control measures, and specialized software needed to view or manipulate data.
  - Describe the contents of your directories and/or files.
  - Provide attribution to any external data sources.
- For tabular data, define all variables and allowable values including null values. Include units of measure where appropriate.
- Scripts and code should be well commented and dependencies accounted for.
- Provide links to associated publications, code in external repositories, and any other information that provides more context to your data.
- Select a license that respects the constraints of any data you may have reused (for help choosing a Creative Commons license, see https://chooser-beta.creativecommons.org/)
Resources list

● Data Management Planning
  ○ DMP Assistant, bilingual tool for preparing data management plans (DMPs)
    https://assistant.portagenetwork.ca/
  ○ OpenAire. How to identify and assess Research Data Management (RDM) costs.
    https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs
    https://doi.org/10.1890/0012-9623-90.2.205
  ○ FAIR Principles: https://www.go-fair.org/fair-principles/
  ○ CARE Principles: https://www.gida-global.org/care
  ○ DataONE Data Management Skill Building Hub https://dataoneorg.github.io/Education/

● Standard Operating Procedures - Documenting expectations and best practices
  ○ DataONE, Best Practices https://dataoneorg.github.io/Education/bestpractices/ for ideas on
    elements of RDM that may be useful to document
  ○ StreamPulse examples: Click “Show SOPs“ button on https://data.streampulse.org/ for SOP
    examples.
Resources list, continued

- Lab Notebooks / Field Notes
  - Pain E (2019, Sep 3) How to Keep a Lab Notebook. Science. [https://doi.org/10.1126/science.caredit.aaz3678](https://doi.org/10.1126/science.caredit.aaz3678)

- File Organization and File Naming
  - University of British Columbia. File Naming Guidelines. [https://researchdata.library.ubc.ca/files/2019/01/FileName_Guidelines_20140410_v03.pdf](https://researchdata.library.ubc.ca/files/2019/01/FileName_Guidelines_20140410_v03.pdf)
Resources list, continues

- **READMEs**
    - En français [https://doi.org/10.5281/zenodo.4058961](https://doi.org/10.5281/zenodo.4058961)
  - Cornell University. Research Data Management Service Group. Guide to writing "readme" style metadata. [https://data.research.cornell.edu/content/readme](https://data.research.cornell.edu/content/readme) (Template available for download).

- **Metadata Standards**
  - List of disciplinary metadata standards, brief descriptions, and tools
    - Digital Curation Centre. Disciplinary Metadata. [https://www.dcc.ac.uk/guidance/standards/metadata](https://www.dcc.ac.uk/guidance/standards/metadata)
  - Search for standards and policies by domain or subject.
    - FAIRSharing. [https://fairsharing.org/standards/](https://fairsharing.org/standards/)

- **Licensing**
  - For Data. Choosing a Creative Commons License. [https://chooser-beta.creativecommons.org/](https://chooser-beta.creativecommons.org/)
  - Choose an Open Source License. [https://choosealicense.com/](https://choosealicense.com/)
  - For Software. Open Source Initiative. Licenses and Standards. [https://opensource.org/licenses](https://opensource.org/licenses)