Federated Research Data Repository (FRDR)

Data Upload Guide

Overview

Federated Research Data Repository (FRDR) is a bilingual publishing platform for sharing and preserving Canadian research data. It is a curated, general-purpose repository, custom built for large datasets. Can-Peat has no direct affiliation to FRDR, and researchers may upload data to a repository of their choice. Guidance on preparing data, metadata and README files provided here also will be applicable to data deposits outside of FRDR.

If you choose to upload data to FRDR, you will first need to create an account.

For more information on depositing data, you can visit the help section of FRDR.

Data Preparation

Before submitting your dataset to FRDR, it's essential to prepare your files.

Gathering Files

1. Choose whether to provide:
   - Processed data
   - Raw unprocessed data
   - Both
2. Include scripts, code, or software necessary for data generation or reanalysis.

File Format Recommendations

1. FRDR recommends that you deposit models, source code or research software in purpose-built repositories such as GitHub, GitLab or Bitbucket. Please see the FRDR website for more details.
2. Ensure the files are in preservation-friendly formats such as CSV for tabular data.

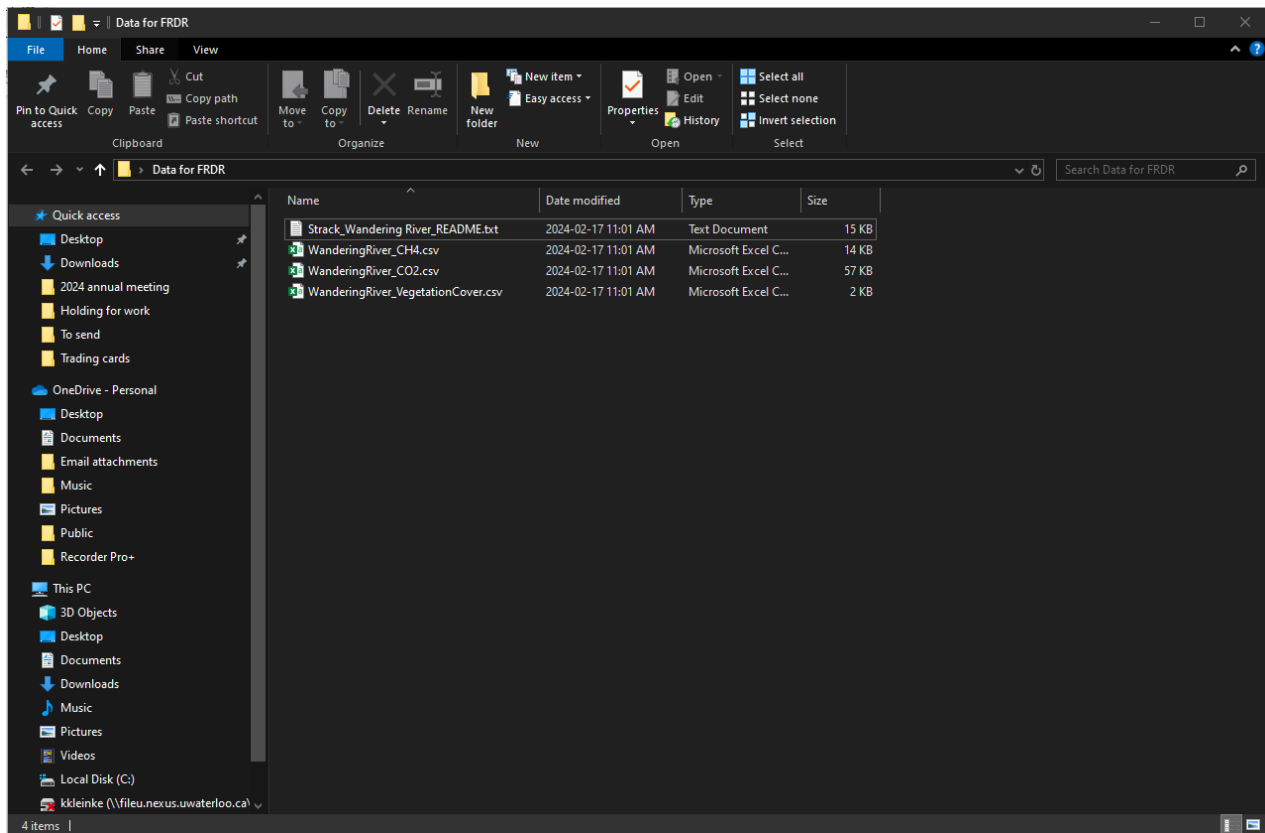In the example, we will use, we have three CSV files with a README (Image 1) file.

Image 1: Folder having three data files and README

Standardization

You'll need to make sure your data uses standard variable names and units. We recommend using variable names and units as defined by FLUXNET, the National Forest Inventory for biomass, and Can-Peat created variable names for chamber-based measurements and greenhouse gas fluxes to assist in separation of fluxes measured with eddy covariance and chambers when possible. This promotes interoperability among datasets and aids in future data compilation.

In this example (Image 2), you can see there are two columns for fluxes. One for the units used in the paper (g CO_2 per square metre per day) and one for standard units (g C per square metre per day) to ensure users can interpret the data and compare to other datasets correctly.

Image 2: Showing carbon dioxide flux data (used units in paper and standard units)

File Structuring

Depending on the complexity of your dataset, you will need to structure your files (Image 3). We recommend having folders for code, data, and outputs. Place the README file in the main folder.

- **Code:** Include all processing and analysis scripts.
- **Data:** Separate folders for raw and processed data.
- **Outputs:** Include folders for figures and models.

```
├── Code
│   ├── process_raw_data.r
│   ├── analysis_1.r
│   └── analysis_2.r
├── Data
│   ├── Raw_data
│   │   ├── file_a.raw
│   │   └── file_b.raw
│   └── Processed_data
│       ├── file_a.csv
│       └── file_b.csv
├── Outputs
│   ├── Figures
│   └── Models
└── README.txt
```
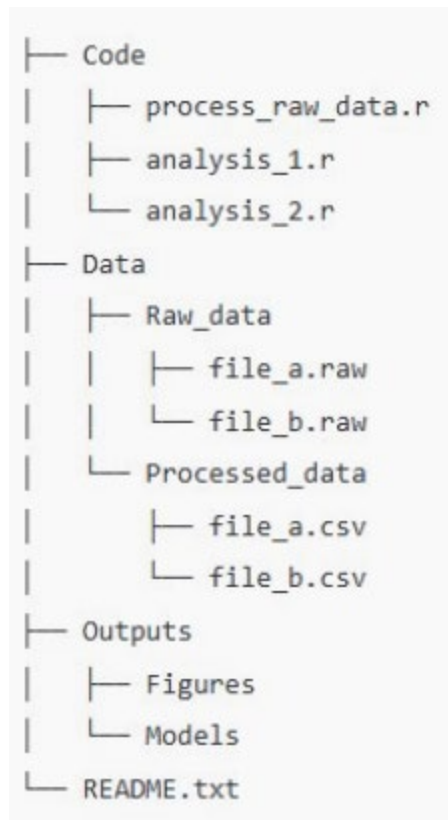
Image 3: File structuring

Naming Conventions

Ensure your files are named both logically and descriptively but keep it brief. Include important information such as the location, data type or version number. Avoid spaces or special characters. In this example (Image 1) we have WanderingRiver_CO2 as well as WanderingRiver_CH4, and WanderingRiver_VegetationCover. Make sure you describe your naming convention in your README file.

README File

The README file acts as a codebook to ensure that your data can be interpreted correctly. We will be using an altered version of the template provided by FRDR.

Image 4: README file

In this example (Image 4) you will see that some of the fields are mandatory, indicated by an asterisk, while some are optional and may not apply to your data. If these fields do not apply to your data, you can leave them blank. There will also be instructions through indicated by the pound symbol. Ensure you delete these before finalizing your README file.

Structure of the README

The README file is divided into sections.

General information

This section is mandatory and includes the title of your dataset, author information including their contact details, date or range of data collection (this can be an approximation), and information about your funding source (Image 4).

Sharing and access restrictions

You will need to choose a license for your dataset. In this example (Image 5), we are using the license CC BY 4.0.

Image 5: Sharing and access restrictions

This license has the least restrictions. Users are free to share and adapt the dataset as long as they give appropriate credit to the original. You can find more details about licenses at the website here (Image 6). There is also a tool to help you choose your licenses if you are

unsure (Image 7). Note that FRDR cannot managed restricted data past what licenses can dictate. Short-term embargos are available to delay releasing the data publicly. In this section you can also indicate if your data is linked or derived from another source.
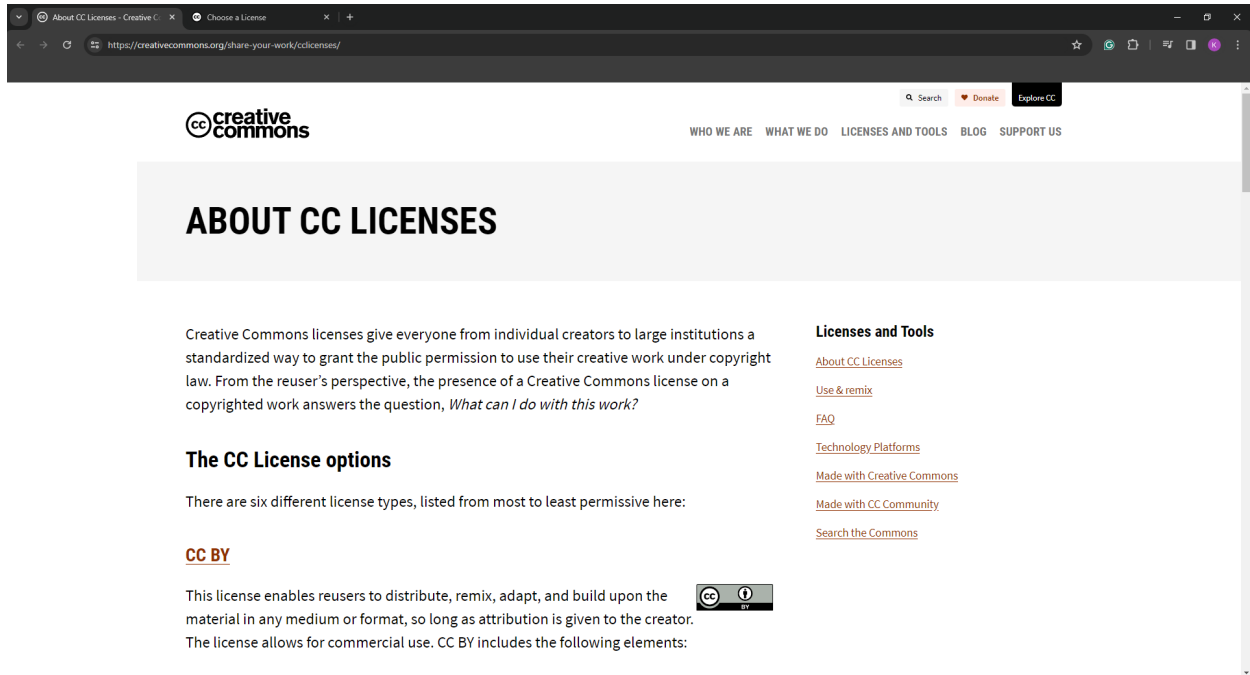
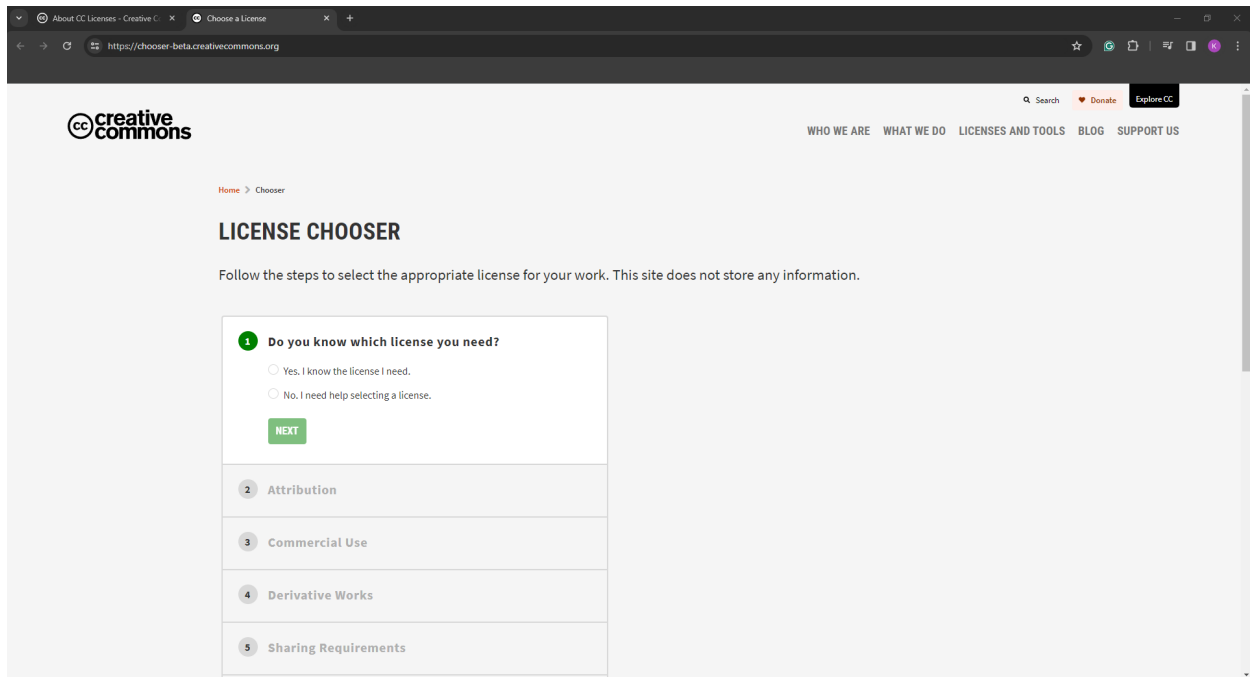

Image 6: Licensing options

Image 7: Tool to choose the license

Data and file overview

Data and file overview is to describe your file structure from earlier. Ensure you use the full file name and provide a short description of what each file contains (Image 8). For example, we have Wandering River_CH4 for which the description is: instantaneous measurements of methane flux at restored and unrestored plots alongside concurrently measured water table and soil temperature data.

```
--------------------
DATA & FILE OVERVIEW
--------------------

1. File List*

    A. Filename: WanderingRiver_CH4.csv
Short description: Instantaneous measurements of methane flux at restored and unrestored plots alongside concurrently measured water table and soil temperature data.

    B. Filename: WanderingRiver_CO2.csv
Short description: Instantaneous measurements of net ecosystem carbon dioxide exchange, ecosystem respiration and calculated gross primary production alongside concurrently measured water table and soil temperature data.

    C. Filename: WanderingRiver_VegetationCover.csv
Short description: Visually estimated cover to species level in each sampling plot reported in WanderRiver_CO2 and WanderingRiver_CH4 in August 2012.

2. Relationship between files, if important: The variable STUDY_PLOT appears in each file and uses the same name to refer to each study plot allowing comparison among files.

3. Additional related data collected that was not included in the current data package: None

4. Are there multiple versions of the dataset?* no
    A. If yes, name of file(s) that was updated:
            i. Why was the file updated?
            ii. When was the file updated?
```

Image 8: Data and file overview

Methods in collection and processing the data

If you have a published paper describing your methods in detail, we recommend including a very brief description in the README and then linking to your paper for more details (Image 9).

```
---------------------------
METHODOLOGICAL INFORMATION
---------------------------

1. Description of methods used for collection/generation of data*: Carbon dioxide and methane fluxes were measured using the closed chamber technique using chamber 60 cm x 60 cm x 30 cm tall. Net ecosystem exchange of carbon dioxide was measured in full light conditions and under artificial shade with photon flux density of photosynthetically active radiation measured at each light level. Ecosystem respiration was measured as carbon dioxide flux under dark conditions. Gross primary production was calculated as the difference between net ecosystem exchange and ecosystem respiration. Methane flux was measured under dark conditions. Water table was measured in a standpipe adjacent to each carbon flux measurement location. Soil temperature was measured with a K-type thermocouple probe. Vegetation cover was visually estimated. Further methodological details can be found in Strack et al. 2014.

2. Methods for processing the data*: Carbon dioxide and methane fluxes were calculated from the linear change in concentration in the chamber headspace during the closure period. Details on quality control procedures can be found in Strack et al. 2014.

3. Instrument- or software-specific information needed to interpret the data, if appropriate: Details of instruments used for determination of carbon dioxide and methane concentrations are given in Strack et al. 2014. Conversion of concentration change over time to flux was completed in Microsoft Excel.

4. Standards and calibration information, if appropriate: See Strack et al. 2014.

5. Environmental/experimental conditions: Measurements were collected in a former horticultural peat extraction area that had undergone restoration using the moss layer transfer technique in 2008 (3-4 years prior to measurement). Twelve plots were arranged along hydrological gradients that had developed at the restored sector. A additional three plots were measured in the adjacent unrestored area that was no longer used for peat extraction but had received no restoration treatments. The study site was formerly a bog prior to peat extraction.

6. Describe any quality-assurance procedures performed on the data, if appropriate: See Strack et al. 2014.

7. People involved with sample collection, processing, analysis and/or submission: Maria Strack, Bin Xu, Magnus Keith, Tariq Munir, Jordanna Branham
```

Image 9: Methods in collection and processing the data

Codebook

In this section, you will describe the variables in each of your data sheets. Ensure you describe any codes such as N/As for missing values (Image 10). In this example, you can see that we have a variable list with the name of each of our columns and a description of what this means. This will be vital in anyone interpreting your data.

```
------------------------------------------------------------
DATA-SPECIFIC INFORMATION FOR: WanderingRiver_CH4.csv
------------------------------------------------------------

1. Number of variables: 15

2. Number of cases/rows: 122

3. Missing data codes:
        N/A              DATA IS MISSING. Check NOTES variable for further details.


4. Variable List*:

Name: STUDY_SITE
Description: Name of study site. All rows in data set are from Wandering River Restored that includes both restored and unrestored plots.

Name: DATE
Description: Date of measurement given in format YYYY-MM-DD

Name: STUDY_PLOT
Description: Text string for study plot names used by research team. WRC A, WRC B, WRC C are unrestored plots. All plots that start with WRR have been restored by moss layer transfer technique in 2008. Four transects were created (A-D)
each with triplicate plots. Plots are named by WRR followed by transect letter and replicate number (e.g., WRR A1)

Name: TREATMENT
Description: Categorical variable indicating status of study plot as "Unrestored" or "Restored". Stored as text string.

Name: TS_2
Description: Soil temperature at 2 cm depth measured in degrees Celsius

Name: TS_5
Description: Soil temperature at 5 cm depth measured in degrees Celsius

Name: TS_10
Description: Soil temperature at 10 cm depth measured in degrees Celsius

Name: TS_15
Description: Soil temperature at 15 cm depth measured in degrees Celsius

Name: TS_20
Description: Soil temperature at 20 cm depth measured in degrees Celsius

Name: TS_25
Description: Soil temperature at 25 cm depth measured in degrees Celsius

Name: TS_30
Description: Soil temperature at 30 cm depth measured in degrees Celsius

Name: WTD
Description: Water table position relative to the surface of the peat measured in cm. Negative values indicate a depth below the surface of the peat.

Name: CH4_CHAMBER_ORIG
Description: Methane flux measured with the closed chamber method in g CH4 per square metre per day.
```

Image 10: Codebook

Secondary Use Considerations

Before moving on to the submission of your data, please consider if you have any secondary use of data or code such as redistributing or publishing from a third-party source. If you are, ensure that you have permission to publish the data or code to FRDR.

Submitting to FRDR

When ready to submit, navigate to frdr-dfdr.ca/repo/ and click "Deposit Data."

Submission Process

- **Login**: You will need to log in to begin the submission progress. You can save and exit your progress at any time at the bottom to save a draft of your submission that you can return to (Image 11).

Image 11: Login

- **Accept Terms of Use:** First you need to accept terms of use by clicking "I accept" (Image 12).
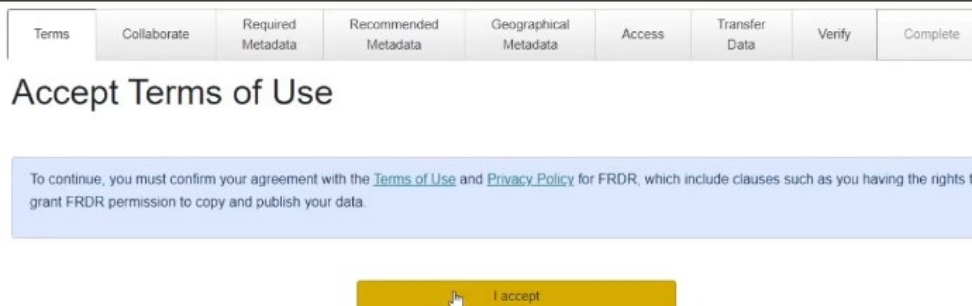

Image 12: Accept terms of use

- **Adding collaborators:** This is not to reference those who contributed to the dataset but to add any users to edit your metadata, add and remove files, manage reviewers for the submission, or to submit your dataset. This could be used for example if your supervisor or co-authors would like to review the submission before it is finalized (Image 13).

Image 13: Adding collaborators

- **Required metadata:** Some fields, such as title, will be shared with the metadata. Ensure these shared fields are the same as your README file when completing your metadata (Image 14). For the description, include the purpose, nature, and scope of the dataset. For your keywords, we strongly recommend using 5 words from our Can-Peat list. Standardizing keywords will improve data findability. For example, if you write CO_2 but users search carbon dioxide, your data may not come up in the search. Our current list of keywords includes carbon dioxide, methane, nitrous oxide, soil carbon, peat properties, biomass, vegetation, wildlife, meteorology, hydrology, biodiversity, traditional knowledge, remote sensing, mapping, modeling, conservation, reclamation/restoration, policy, land-use, effects of disturbance, carbon accounting. **We also recommend include "peatland" as a keyword**. FRDR is a Canadian database but in international submissions we also recommend including "Canada". To help fill out the field of research, you can explore the Canadian Research and Development Classification field names.

Image 14: Metadata

- **Authors of the dataset:** Fill out the authors of the dataset (Image 15). This should be the same as in your README file.

Image 15: Authors

- **License selection**: Select the license you have chosen. You can click "More" to see the full list. For custom licensing or any questions, you can always contact the FRDR support.

Image 16: License selection

Recommended metadata

Although this section (Image 17) is not required by FRDR. We strongly recommend filling out all applicable fields to provide the most informative and useful metadata for your dataset.

Image 17: Recommended metadata

For this dataset we will be using the collection period instead of the time period covered. The collection period refers to the range of dates when the data was collected while the time period covered is the range of dates when the data refers to. For example, the time period covered would be used if you had soil cores that dated back past the time of collection.

Next include all funders, and award information if applicable (Image 18).

Image 18: Funding details

Contributors are for others who assisted in the creation of the dataset other than the authors (Image 19). In this example, we have three individuals who assisted with data collection in the field. You can also include data manager, project manager, research group, sponsor, supervisor, or other. Fill out the related identifier with any links related to the dataset. Mostly commonly used will be "this dataset is cited by …". This option should be used for publications that analyze the dataset. Use the doi to reference the appropriate resource.

Notes should be used to include Local Contexts notices related to your dataset if applicable. You can use the decision tree in the Can-Peat Local Contexts Guide to help you decide when to set up a subproject in the Local Context Hub and create Notices. If you have done this, the links provided with your notices can be copied and added to this Notes section of the metadata. You can also use this section for any important information not included in the metadata. For example, you can include references to related items that do not have a doi or URL. This section can also include requests for those wishing to re-use the dataset such as "Please contact the authors to discuss collaboration opportunities".

Image 19: Additional information

Geographical metadata

We strongly recommend filling out the geographical metadata section. This will allow your data to be discovered via map-based searches. There are three options to geographical describe your metadata. For better integration into maps, ensure you fill out the geographic point or bounding box.

Image 20: Geographical metadata

Final Steps

The metadata is now complete (Image 21). The next sections will be quick to fill out. Make sure to list the correct person as the contact for the dataset. This may or may not be you. If you are a student, make sure to discuss with your supervisor on who should be the contact. If this dataset needs to be formally reviewed by another party such as a journal editor or colleague, you can select a review period. This will embargo the dataset until the external reviewer reviews the submission. In this case, we do not require an external review.

Image 21: Access, dataset review and embargo

We also do not require an embargo period. FRDR only guarantees preservation of the dataset for 10 years. If you would like the dataset to be preserved longer, ensure to leave a comment describing the importance of the dataset. In this example, we are requested longer preservation (Image 22) as historic measurement of greenhouse gas exchange following restoration will be important for monitoring of restoration and impacts of climate change.
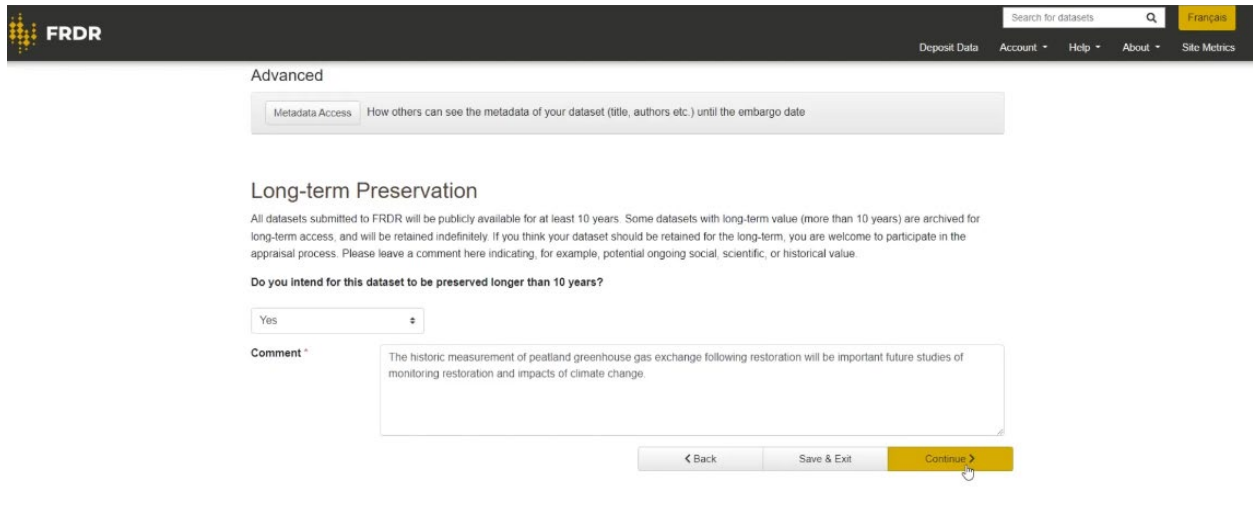
Image 22: Request to preserve data longer than 10 years.

Next you will need to transfer the data files in your dataset (Image 23). Make sure to include your README file. Lastly, verify that all the information is correct.
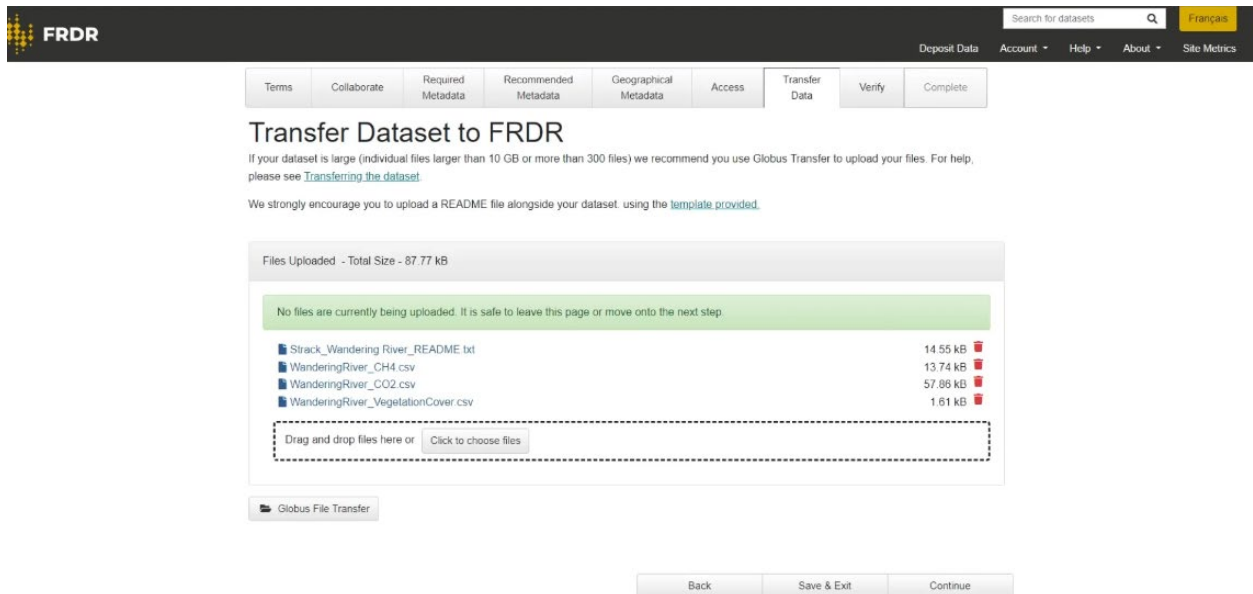

Image 23: Transfer the data files in your dataset.

Note that a verification email will be sent to the contact person you provided (Image 24).

Image 24: Verify dataset.

When you are ready to submit, click Finish (Image 25). Once submitted, your data submissions will be reviewed by data curators at FRDR and they may request additional information to improve clarity for data users. However, you will receive a doi for your dataset as soon as you submit.

Image 25: Finish the data upload.

Conclusion

By following this structured approach, you can effectively prepare and submit your dataset to FRDR. Following suggested data formatting and documentation steps will be useful for data deposits to other repositories as well, ensuring that your data is well-organized and

easily interpretable. Adding Local Contexts Notices is a tool that can help advance Indigenous data sovereignty.