



LWRC Seabird Rosette Data Cookbook

Dataset Name: LWRC Seabird Rosette Data

Program / Project: Lake Winnipeg Research Consortium

Document Control

Version History

Version	Author	Type	Date Modified	Comments
1.0	Yanique Campbell	Working Copy	March 9, 2026	Initial version

Document Location

Found on CanWIN under: [LWRC 2008 Seabird Rosette DATA](#)

License

This document is licensed under **CC BY 4.0**, excluding UM branding and images.

1. Overview

This workflow describes how raw CSV files were cleaned and standardized. For this particular workflow, CanWIN's [CSV Studio](#) application was used.

The purpose is to ensure reproducibility in how column headers and metadata rows were processed before the dataset was published on CanWIN.

This workflow is intended for researchers, students, and data curators who need to understand how raw instrument exports were transformed into analysis-ready files on CanWIN.

2. Raw or Input Data

The raw data consisted of CSV files exported from **Seabird Rosette**.

Characteristics:

- Several metadata rows above the actual header row
- Units stored in a separate row
- Occasional non-machine-friendly characters in column names with mixed formatting (underscores, spaces, special characters)
- Metadata elements such as date, and coordinates are present in the metadata rows but not in the data table

Raw file location:

- [Gitlab](#) (Please contact author or data curator to access this repo)

3. Processing Workflow Summary

The following steps summarize how the raw CSV files were cleaned using CSV Studio.

3.1 Reading Input Files

Files were uploaded directly into CSV Studio. No preprocessing was performed prior to upload.

3.2 Preprocessing and Structural Cleanup

Remove Metadata Rows

Files contained one or more metadata rows above the actual header row. These rows were removed using the **Remove Metadata Rows** function.

Extract Date and Lat Long Coordinates

The **Remove Metadata Rows** function also allows for the extraction of certain values from the metadata rows, so that they can be added to the table.

In this case, the date, latitude and longitude were extracted from the metadata rows and added as columns to the table, with the headers:

- `date-time`
- `latitude`
- `longitude`

The `N` and `W` were removed from the lat, long coordinates.

The cleaned file was saved and re-uploaded to ensure a clean starting point.

📘 Note

The **Remove Metadata Rows** function allows for a limited amount of values to be extracted from the metadata rows. The user then provides a column name for each value. This procedure is position based, so it assumes that all files have that particular value, say the date, at the same position in every file. This will likely be true if all the files are produced by the same instrument. The function then allows certain rules to be applied to the value that is being extracted, for example, removing “N” and “W” from the lat long coordinates. Again, this will be applied to all files, assuming that the lat, long values are in the same position.

3.3 Column Header Cleaning

Merge Units Row Into Header Row

Now a properly structured (rectangular) table, the files contained a separate row for units (e.g., `degC`, `mg L-1`, `NTU`).

This row was merged into the header row using the **Merge Header Rows** function.

This ensured that each column had a single, complete header such as:

```
Temperature_degC
Turbidity_NTU
DissolvedOxygen_mg L-1
```

Clean Column Headers

The **Clean Headers** function was used to standardize column names.

This step:

- removed special characters

- extracted units (based on internal units map)
- converted headers to machine-friendly formats
- ensured uniqueness

Examples:

Raw Header	Cleaned Header
<code>Temperature_degC</code>	<code>temperature (degC)</code>
<code>SpecificConductance_uS cm^-1</code>	<code>specificconductance (uS cm-1)</code>

3.4 Additional Processing

Adjust Column Header Names

After cleaning, column names were optionally adjusted for readability using the **Rename columns** function.

This included:

- adding underscores between merged words
- renaming ambiguous columns
- aligning naming conventions across files

Example:

```
specificconductance (uS cm-1) → specific_conductance (uS cm-1)
```

Reorder columns

The column headers were then reordered. This is optional, but it maintains consistency with how the files are served. The date, lat and long are moved to the front of the file. This is done using the **Reorder function**.

3.5 Finalization

The cleaned dataset was exported from CSV Studio as a CSV file with:

- one header row with standardized column names and preserved units
- no metadata rows
- consistent formatting across all files

The files were also merged at this point, and uploaded to CanWIN.

4. Variable Transformations

Original Name	Units (raw)	Units (cleaned)	Final Header	Notes
Depth		m	depth (m)	
Temperature	degC	degC	temperature (degC)	
SpecificConductance	uS cm ⁻¹	uS cm-1	specificconductance (uS cm-1)	
DissolvedOxygen	mg L ⁻¹	mg l-1	dissolved_oxygen (mg l-1)	
DissolvedOxygen	%sat	percent_sat	dissolved_oxygen (percent_sat)	Added underscore between dissolvedoxygen. Removed '_1' from <code>DissolvedOxygen_1</code> before units merge, since there would've been two <code>DissolvedOxygen</code> . The app prevents duplicates by adding '_1' Converted <code>%</code> → <code>percent</code>
Turbidity	NTU	ntu	turbidity (ntu)	
Chlorophyll	RFUB	rfu	chlorophyll (rfu)	Units were RFUB in raw files, correct unit is RFU.
UnderwaterPAR	uE m ⁻² s ⁻¹	uE m-2 s-1	underwater_par (uE m-2 s-1)	
DescentRate	m s ⁻¹	m s-1	descent_rate (m s-1)	
			date_time	Added from metadata rows
			latitude	Added from metadata rows
			longitude	Added from metadata rows

5. Output Files

File Type: CSV

File Location

- [CanWIN page](#)
- [GitLab](#) (Permissions required)

6. Running the Workflow

6.1 Graphical Interface (CSV Studio)

App can be found [here](#).

1. Upload raw CSV file
2. Remove metadata rows
3. Extract date, lat, long (remove **W** & **N** from coordinates)
4. Download structured, tabular files (processed_version 1)
5. Re-upload files
6. Merge units row into header
7. Clean column headers
8. Adjust column names using Rename columns (optional)
9. Reorder columns (optional)
10. Export cleaned files
11. Merge (optional)
12. Export merged file