

Documentation of data

From metadata to tools

Øystein Godøy, Markus Fiebig, Torill Hamre

Topics

- Metadata
 - Discovery level
 - Use level
- Data structure and formatting
- Provenance
- ***Tools***

Tools

- Why tools?

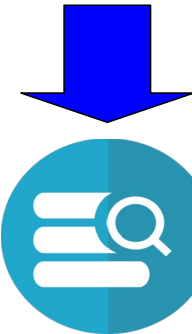
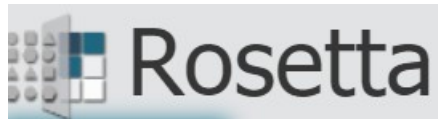
- To add documentation to dataset files
 - Discovery metadata
 - Usage metadata
- To prepare datasets in standard formats
- Makes data easier to reuse (also by others)



```
WELL-18.CSV - Notepad
File Edit Format View Help
Data file for DataLogger.
=====
COMPANY : <company name>
COMP STATUS: Done
DATE : 03/06/2009
TIME : 09:54:45
FILENAME : C:\WELL-18.CSV
CREATED BY : SWS Diver-Office 2.0.1.2
===== BEGINNING OF DATA =====
[Logger settings]
Instrument type =Micro-Diver=15
Status =Started =0
Serial number =..00-C2999 215.
Instrument number =
Location =MW-18
Sample period =M10
Sample method =T
Number of channels =2
[channel 1]
Identification =WATER HEAD
Reference level =400.0 cm
Range =1750.0 m
Master level =0 m
Altitude =0 m
[channel 2]
Identification =TEMPERATURE
Reference level =-20.00 °C
Range =100.00 °C
```

- What tools?

- GIS (commercial, free)
- Data processing and analysis tools (commercial, free)
- Rosetta (free)



Tools

- Rosetta
 - Open source tool, written in Java
 - <https://github.com/Unidata/rosetta>
 - Developed by UniData
 - Web-based application
 - Converts ASCII files to NetCDF
 - Can read metadata from header(s)
 - Can add user defined metadata
 - Saves “setup” as templates for future



Rosetta

Rosetta is Beta software under active development, use at your own risk. This specific version of Rosetta has been tailored for NMDC.

Welcome to Rosetta, a data transformation tool. Rosetta is a web-based service that provides an easy, wizard-based interface for data collectors to transform their datalogger generated ASCII output into Climate and Forecast (CF) compliant netCDF files. These files will contain the metadata describing what data is contained in the file, the instruments used to collect the data, and other critical information that otherwise may be lost in one of many dreaded README files.

In addition, with the understanding that the observational community does appreciate the ease of use of ASCII files, methods for transforming the netCDF back into a user defined CSV or spreadsheet formats is planned to be incorporated into Rosetta.

We hope that Rosetta will be of value to the science community users who have needs for transforming the data they have collected or stored in non-standard formats.

Rosetta is currently under continued further development, and ready for beta testing.



What would you like to do?

Convert a file to the netCDF format and create a new template

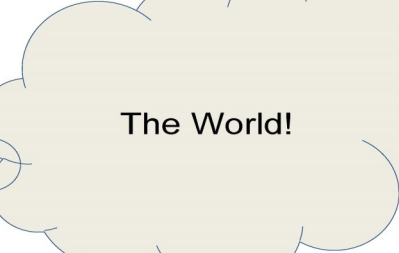
Upload, modify, and use an existing template

Version : 0.5.1-NERSC

Build Date: Fri Aug 03 11:25:13 CEST 2018



INTAROS



The World!

[illegible]

The World!

Rosetta

- Index - Overview - Features
- Admin File
- Search - Admin - Users
- Security - Overview
- Security - Admin - Accounts
- Security - File - Search
- Security - File - Administration
- Download - Download File

Download Converted File

[Download](#)
[Download](#)

[Downloaded from Rosetta](#)
[Downloaded from Rosetta](#)

Download Converted File

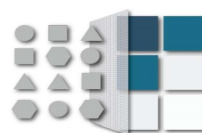
[Download](#)

[Download](#)

[Download](#)


Download
Converted File

Publish to a data server



Rosetta

Transformation Workflow



The screenshot shows the 'Rosetta' web application interface. The main heading is 'Rosetta'. Below it, there are several sections with expandable/collapsible headers:

- Specify General Information** (expanded):
 - Basic Observation Platform:
 - Station ID:
 - Station Name:
 - Station Address:
 - Station Phone:
 - Station Email:
 - Basic General Information:
 - Version:
 - Comments:
 - Deployment Date:
 - Deployment Time:
- Specify Observation Parameters** (collapsed):
 - Observation Type:
 - Observation Frequency:
 - Observation Duration:
 - Observation Start Time:
 - Observation End Time:
- Specify Metadata** (collapsed):
 - Metadata Type:
 - Metadata Frequency:
 - Metadata Duration:
 - Metadata Start Time:
 - Metadata End Time:

At the bottom of the form, there are two buttons: 'Submit' and 'Cancel'.

Below the form, there is a section for 'Support' with the following text:

Support is provided for Rosetta. You can contact us at support@rosetta.com or visit our website at <https://www.rosetta.com>.

Below the support section, there is a footer with the following text:

© 2023 Rosetta. All rights reserved. | Privacy Policy | Terms of Service | Contact Us

Below the footer, there is a logo for 'unidata' and the text 'The Data Center for the Earth and Space Sciences'.

Add General Metadata

[illegible]

Choose data collection method

[illegible]

Select the data header

[illegible]

Define Data Variables and Metadata

[illegible]

Define Coordinate Variables and Metadata

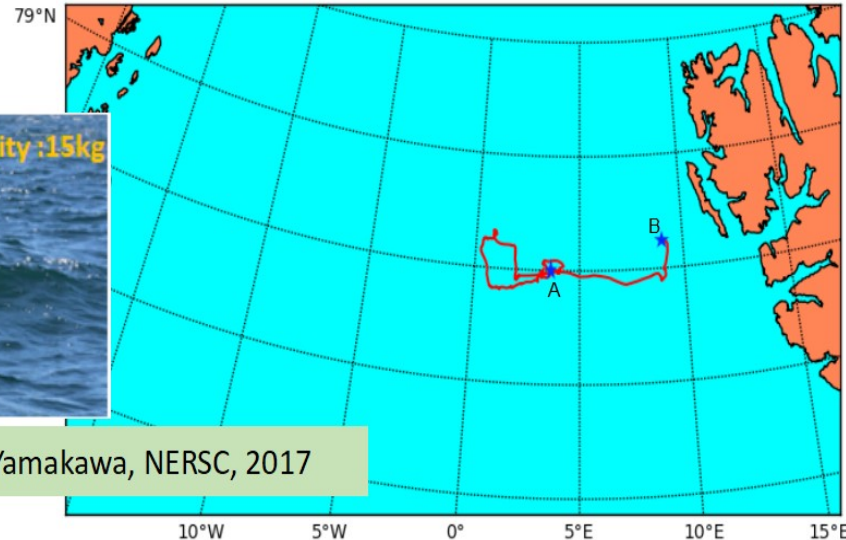
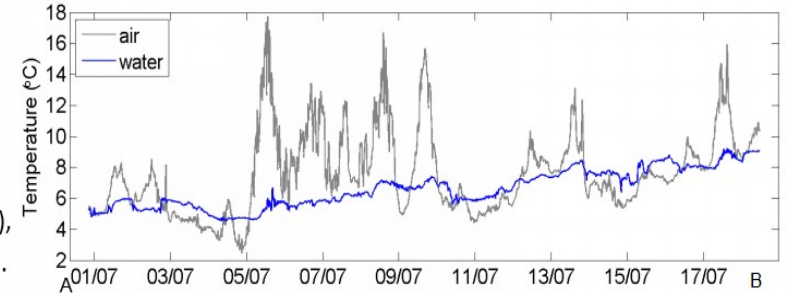
Tools

- Example

- Sailbuoy data 2016
- Converted from ASCII to NetCDF with Rosetta
- Published in NMDC through Thredds Data Server (TDS)
- Searchable through NMDC Data Portal at <https://nmdc.no/>

Ocean Acidification data in 2016 in the Fram Strait (RFF Iskantseilas)

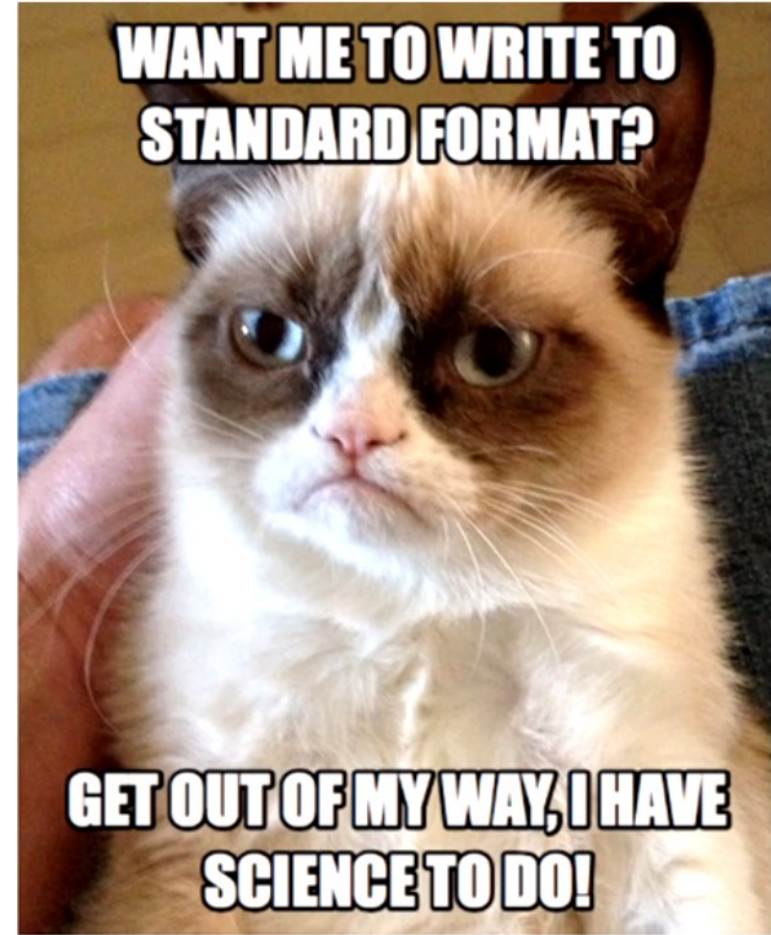
- Collected by a remotely controlled sailbuoy
- Parameters:
 - temperature,
 - conductivity,
 - pH
 - partial pressure of carbon dioxide ($p\text{CO}_2$),
 - dissolved oxygen (O_2).
- Every 10 minutes
- 30 June - 18 July 2016



Courtesy of J. Ullgren and A. Yamakawa, NERSC, 2017

Tools

- What should motivate a tool developer?
 - To serve the needs of the users!
 - Be like a good football referee (i.e. “invisible”)
 - Because, how much fun really is format conversion?
 - Luckily we can use tools like Rosetta
 - Most of us would rather do something else...





**STOP HERE
TO REFUEL**
(TAKE OUT AVAILABLE)