Rosetta User Manual

29 March 2021

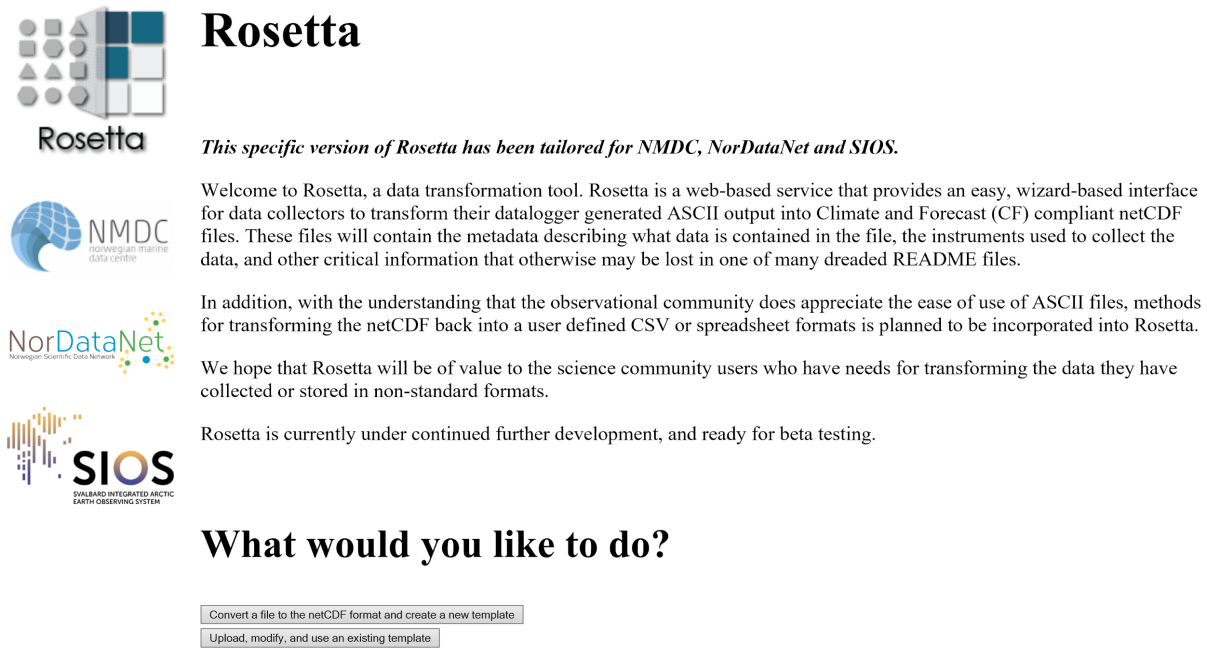
Torill Hamre[1], Frode Monsen[1], Asuka Yamakawa[2], Tor Olaussen[1]
[1]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[2]Institute of Marine Research, Bergen, Norway

Overview

Rosetta is a web application for converting data in ASCII CSV or Excel files to NetCDF format. The conversion is defined in a special file - a template - that defines how (and which parts of) the data file is extracted and converted to spreadsheet NetCDF. If you have several data files formatted in the same way, you can define a conversion template once and reuse it for all of these data files. Rosetta is run in a web browser without installing any plugins. The generated NetCDF file, as well as the generated template, can be downloaded to your computer.

Rosetta can be accessed at http://tomcat.nersc.no/rosetta/

Email support: rosetta@nersc.no



Contents

Introduction

Major functionality

Rosetta provides two main features:

1. Conversion of an ASCII CSV or Excel spreadsheet file to NetCDF/CF format.
2. Definition and storage of a specification (i.e. a template) of the conversion.

Key concepts

Rosetta handles three categories of data:

1. **Profiles** - a dataset collected at a given location and point in time (i.e. fixed latitude, longitude and time) where the observations are made at different altitudes (i.e. variable altitudes above or depths below sea surface)
2. **Time series** - a dataset collected at a given location and altitude (i.e. fixed latitude, longitude and altitude), where observations are made over a time period (i.e. varying time coordinate)
3. **Trajectories** - a dataset collected over time at a given altitude (i.e. fixed altitude or depth) at a series of locations (i.e. varying latitude, longitude, time coordinates).

For each category of data, Rosetta can handle a number of measured or derived parameters. These are stored in the generated NetCDF file together with the coordinate variables (time, altitude, latitude, longitude), description of every parameter and coordinate (variable metadata) and descriptions of the dataset as a whole and (global metadata)

Known limitations

At present, the size of the input data file is limited to 1.2 MB.

Appendix D shows the supported formats for date and time.

The delimiter characters cannot be mixed between any two columns in a data file. For instance, Rosetta will not accept a combination of comma (‘,’) and whitespace (one or more space ‘ ‘) characters in between two columns. Even if the application allows you to check multiple delimiter characters in the “Specify delimiters” screen, only one type of delimiter characters can be used at a time to separate two columns.

Conversion using a predefined template

Step 1: Choose template and data files

From the homepage of Rosetta (http://tomcat.nersc.no/rosetta/) select to start with an existing template by pressing the “Upload, modify and use an existing template” button near the bottom of the page:

What would you like to do?

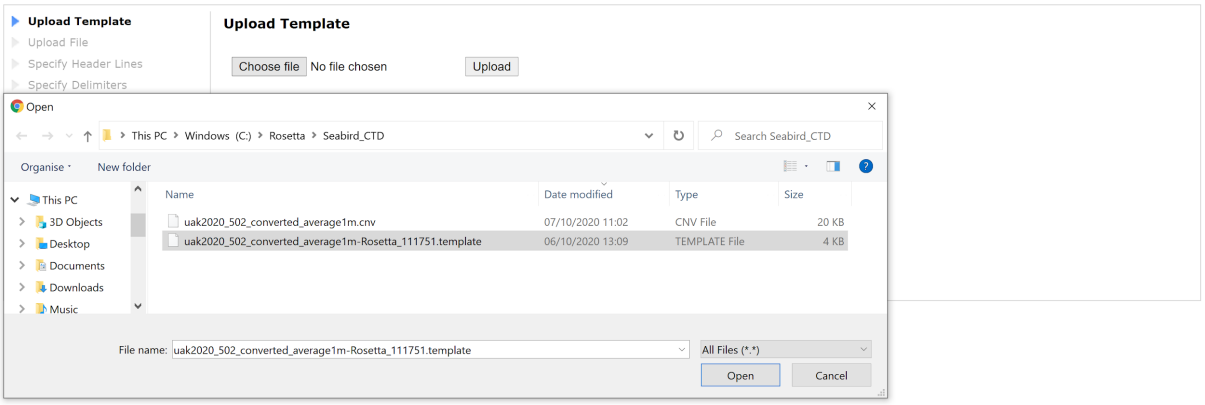
Convert a file to the netCDF format and create a new template

Upload, modify, and use an existing template

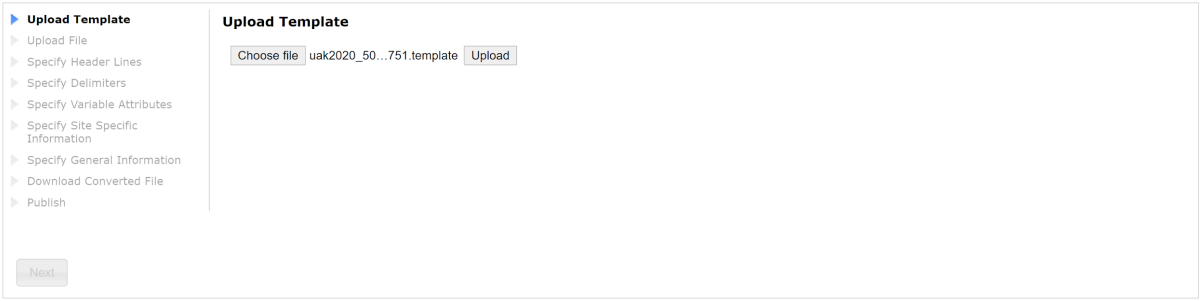
On the new page, click the “Choose file” button. Then select the template file you would like to use from your computer or a network drive, and press the “Upload” button. This will load the definition of the conversion to be performed into Rosetta. After upload is completed, press the “Next” button to proceed with selection of a data file which matches the template. After the data file is successfully uploaded, press “Next” again to move to the next step

Screenshots for uploading files below. (Intermediate screens of uploads completed are left out.)

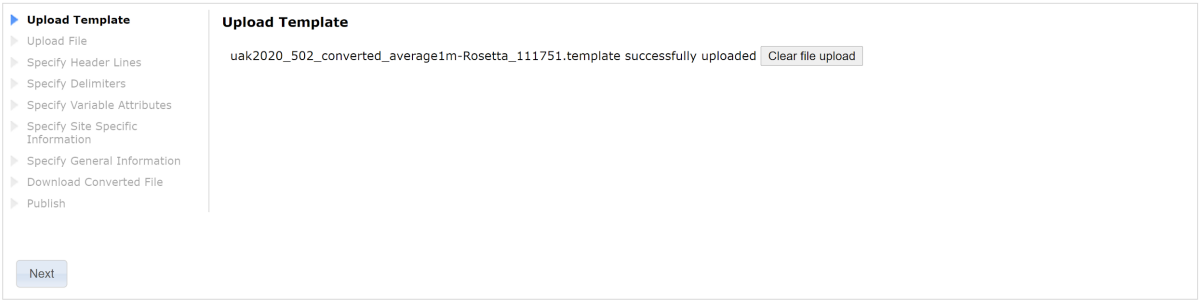
Rosetta



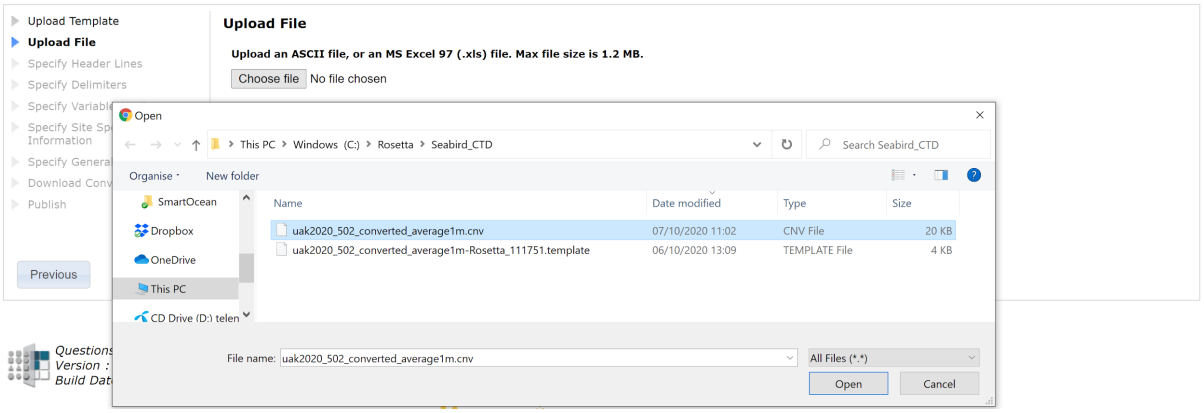
Rosetta



Rosetta

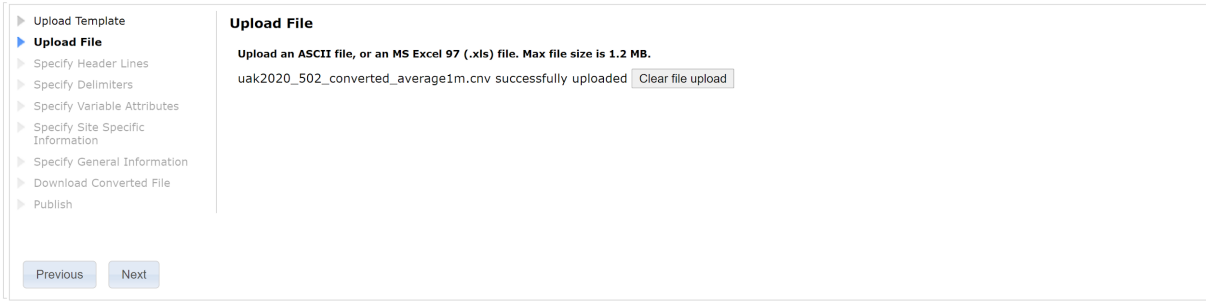


Rosetta



Rosetta

Rosetta



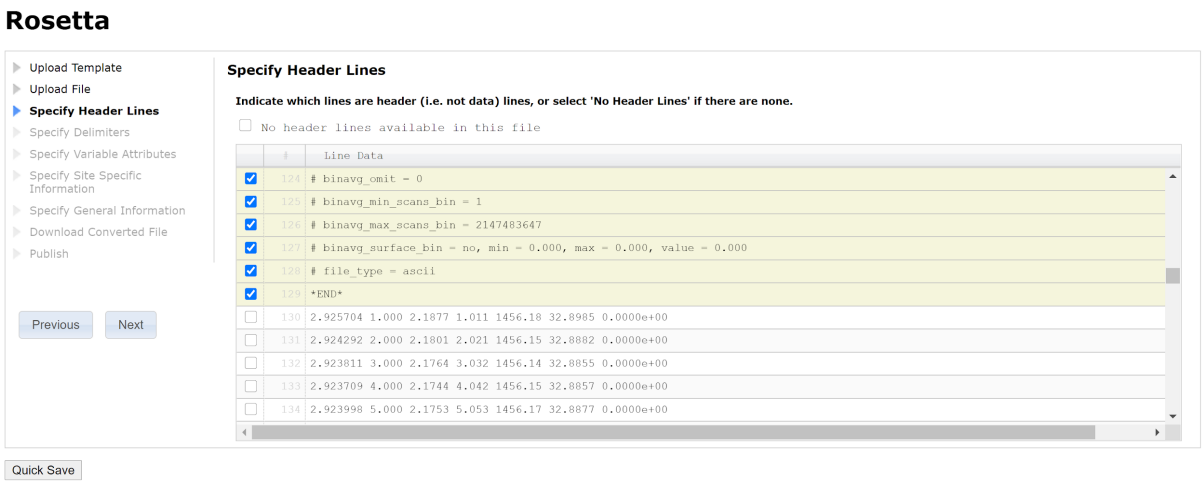
After successfully uploading template and data files:

- uak2020_502_converted_average1m-Rosetta_111751.template
- uak2020_502_converted_average1m.cnv

you can proceed by pressing the “Next” button again.

Step 2: Specify which lines in the data file are headerlines

The uploaded template specifies which lines in the data files constitute the header. Typically the header is located in the beginning of the data file, and occupies from a single line to some tens of lines. An empty header is also possible (i.e. all lines in the file contain data). Rosetta shows a screen like below where the header lines are marked, and then the beginning of the data found in the file.

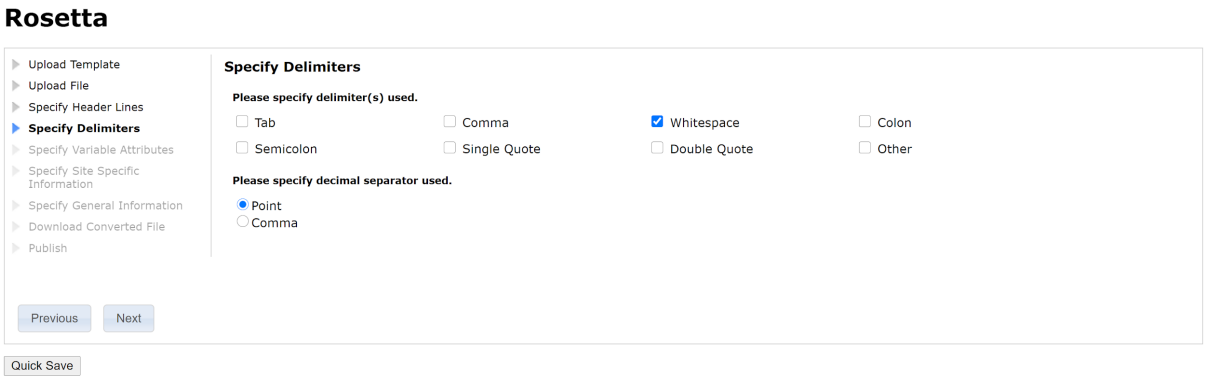


The header lines are marked a background colour, making it is easy to separate from the lines containing data. Lines holding data are always shown on a white background.

The example used above has many header lines, so you will need to scroll down to see the data lines. In case your data file has more or fewer header lines than specified by the template, you can check or uncheck these lines to either mark them as header lines or include them as data lines. If you modify the number of header lines, you can use the “Quick Save” button to store a copy of your new template. Once you have verified that the header lines are correctly marked up, press “Next” to continue.

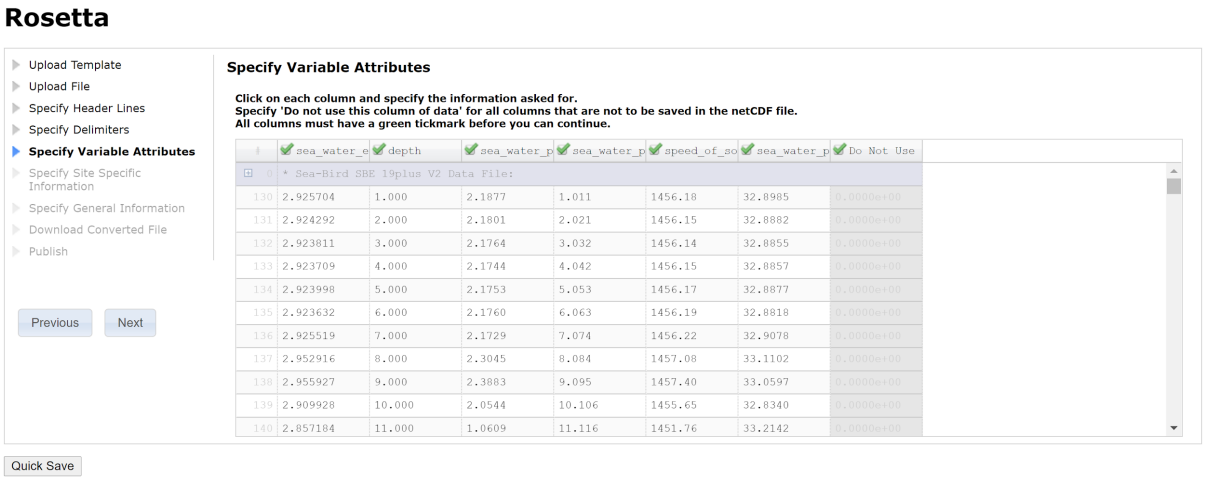
Step 3: Specify delimiter and decimal separator

Rosetta accepts different characters as delimiters between the columns in the data lines. The Nansen Center version of Rosetta also accepts comma (“,”) as the decimal separator. You can choose the characters in the screen below. Note that these characters must match the data file uploaded in step 1. As in the previous step, you can “Quick Save” the template if you make any changes. After checking that the given delimiters and decimal separators are correct for your data file. press “Next” to continue to the next step.



Step 4: Specify variables

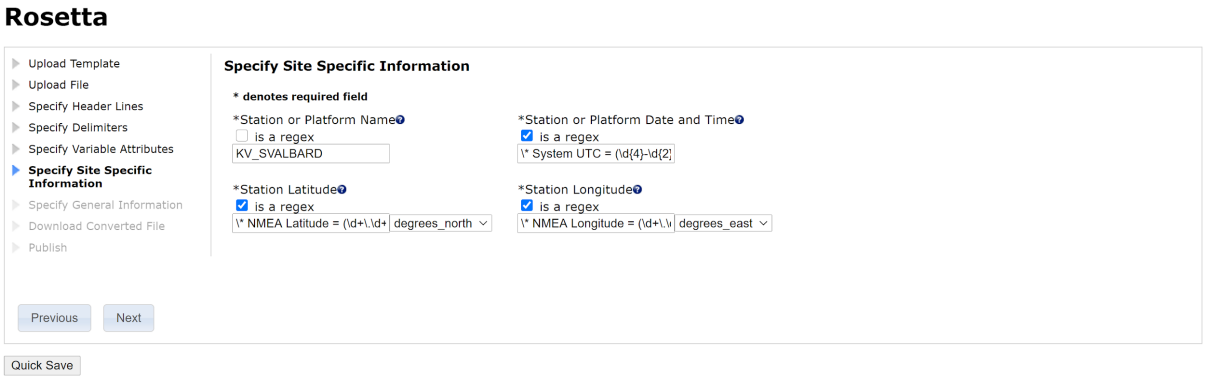
Rosetta now uses the specification of header lines, delimiter and decimal separators to split the data lines into columns for the variables of the dataset. This split is shown in the following screen:



Since the uploaded template matches the chosen data file, no action is needed here. The template holds the specification of each variable to include in the conversion, and marks those columns that are not converted. In this case, the last data column will not be included in the NetCDF file to be generated in the last step.

Step 5: Specify site specific information

Rosetta lets you specify the name of the site, the location, time and altitude (or depth) where the dataset was collected. This can be given in the following screen.



The uploaded template defines the dataset as a profile. This means that the geographical location and time can be given in addition to the name of the site where the data were collected. In this case the site is a vessel, Norwegian Coast Guard vessel KV Svalbard. The remaining site variables, location and time, are read from from the header of the data files, using a string matching pattern called a regular expression.

The regular expressions used in this template works like this:

- * Escapes the asterisk to find an actual * symbol.
- () It is the information inside the parentheses that will be included in the NetCDF file.
- \d Represents a number character (i.e. a digit).
- {} Indicates how many times the character in front of it is repeated.
- + Lets the character in front of it repeat 1 or more times.

The Nansen Center Rosetta version has implemented regular expressions as a means to extract descriptive information (i.e. metadata) for a dataset from a data file. This allows a

template to be used to automatically extract metadata that will go into the output NetCDF file instead of having to type this information into the screen.

It is also possible to fill in the fields with the actual location and time instead of using regular expressions. This will be prefered if e.g. the text in the header is not in the correct format, if the information is not available in the header, or if the information is the same across all the files you want to use the template for.

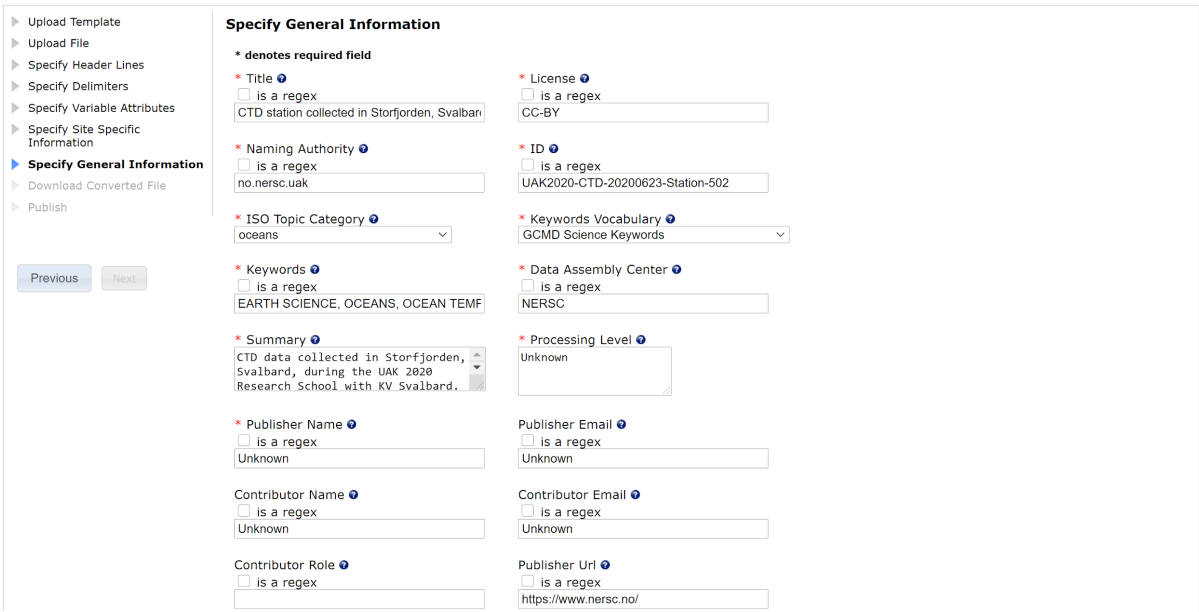
Step 6: Specify general information

Additional descriptions (metadata) for the dataset can be given in the screen below. This allows you to specify important information such as

- Title of the dataset
- The data license that you publish the dataset under
- A short summary of the dataset
- Keywords to make it more easily searchable once published in a data repository
- Name and email for the scientists producing and publishing the data
- Name of the organisation (data centre) publishing the data

Note that in this template, some of the mandatory fields are filled with dummy values. This includes, among others, the names and emails of publishers and contributors. The reason for this was the template was created for use in a research school where different students contributed to collecting different CTD stations together with the course instructors. The use of dummy values showed the students where they needed to fill in information when they prepared the data for publication after the fieldwork had been completed.

Rosetta

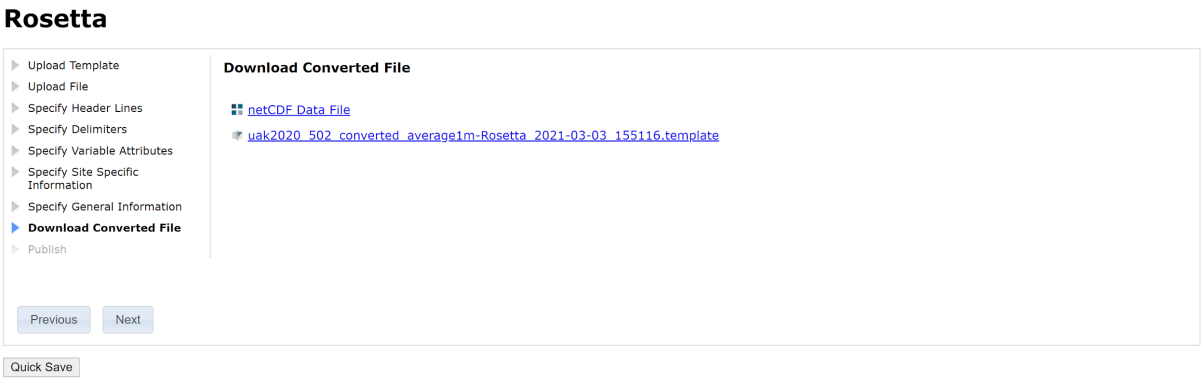


Once you have filled in the remaining descriptions in this screen, press “Next” to go to the final step of the conversion.

Note: If you have a template where all needed information is filled in, you still have to place your cursor in one of the fields and then move it to another field to enable the “Next” button. For instance, pressing the Tab key will move the cursor to the next field.

Step 7: Download resulting files

Using the template and the additional information provided in steps 2-6, Rosetta can now convert the input data file into a NetCDF file. After the conversion is completed, you can download the generated NetCDF file and the updated template file from the screen below.



The template can be reused to convert other data files with the same structure for header and data lines. This will save time for later conversions.

Creating your own template

If no predefined template fits your ASCII data file, you can create a new template.

This section shows you how to create a template for a time series of observations from a meteorological station. The beginning of the ASCII data file is as follows:

```
Filnavn:Florida_2019-11-01_2020-01-31_1580479916.csv
Lat:60.38
Lon:05.33
Area:Florida-Bergen-Norge
Dato,,id,Solskinstid,Lufttemperatur,"Relativ luftfuktighet",Vindretning,Vindstyrke
2019-11-01,00:00:00,0,5,9999.99,161,2.3
2019-11-01,00:10:00,0,5.1,9999.99,158,2.6
2019-11-01,00:20:00,0,5.1,9999.99,169,2.8
2019-11-01,00:30:00,0,5.1,9999.99,163,2.3
2019-11-01,00:40:00,0,5.1,9999.99,159,2.2
2019-11-01,00:50:00,0,5.1,9999.99,164,1.8
2019-11-01,01:00:00,0,5.1,9999.99,166,2.5
2019-11-01,01:10:00,0,5,9999.99,154,2.2
2019-11-01,01:20:00,0,5,9999.99,161,2
2019-11-01,01:30:00,0,5,9999.99,155,2
...
```

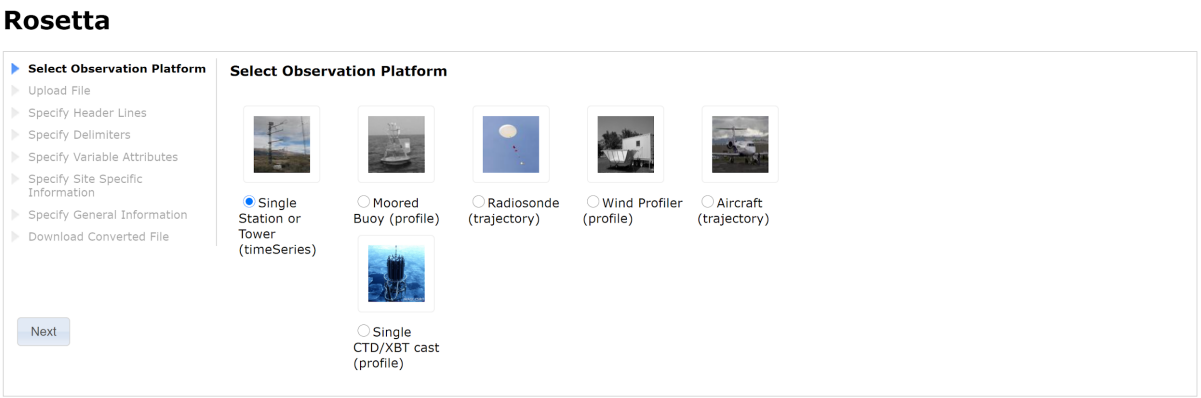
This input file has seven data columns separated by comma (‘,’). The first two columns (numbered 0 and 1 by Rosetta) contain the date and time of day (UTC) of the observation. The next five columns contain the number of seconds of sunshine on this date, air temperature, relative humidity, wind direction and wind strength.

Step 1: Choose to create a new template

From the homepage of Rosetta (http://tomcat.nersc.no/rosetta/) select to create a new template by pressing the “Convert a file to the netCDF format and create a new template” button near the bottom of the page:

Step 2: Select the observation platform that collected your data

From the screen below, select the observation platform for your data. Then press the “Next” button to move to the next step.



The difference between the observation platforms are:

Timeseries:
- Constant Longitude, Latitude and Altitude.
- Time is the coordinate variable.

Profile:
- Constant Longitude, Latitude and Time.
- Altitude is the coordinate variable.

Trajectory:
- Constant Altitude.
- Longitude, Latitude and Time are coordinate variables.

Values for the constants are specified in Step 7: Specify site specific information.

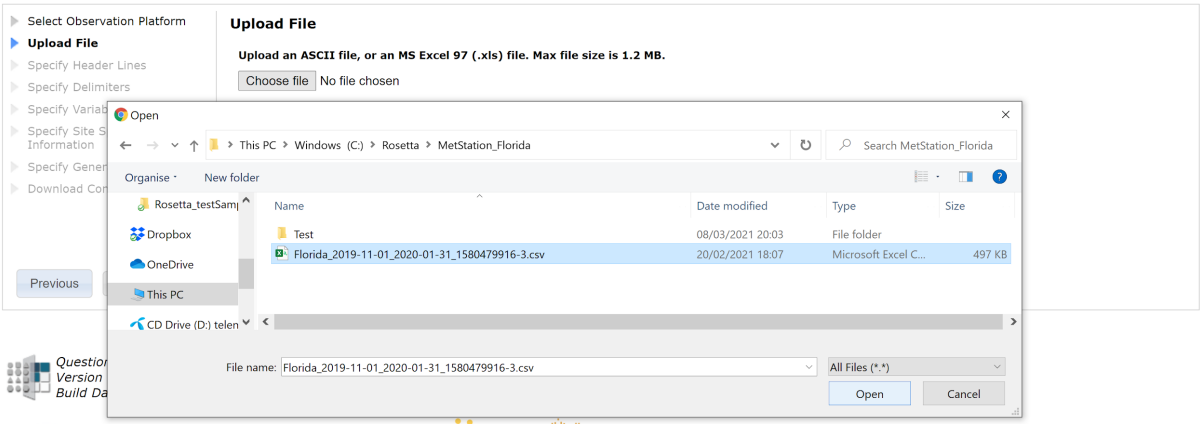
Values for each coordinate variable that vary, must be given in one of the columns in the input data file.

Step 3: Upload your data file

Select the data file by pressing the “Choose file” button and navigate to where it is located on your computer or network drive. Then press the “Upload” to let Rosetta read the data file into memory, and then press “Next” to continue with creating a new template.

Screenshots from the upload process:

Rosetta



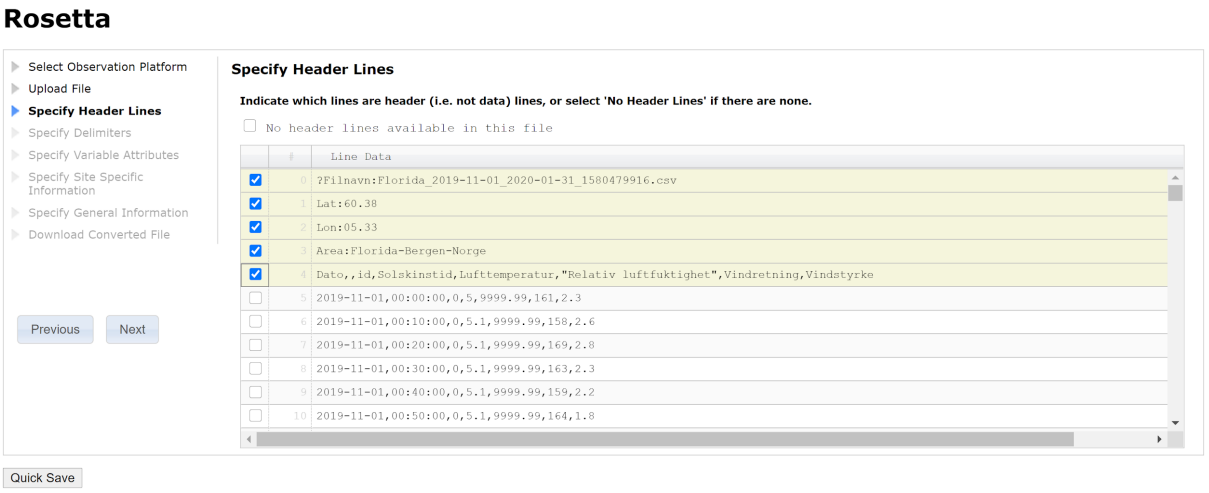
Rosetta

Rosetta

Step 4: Define header lines

The next screen allows you to mark the header lines of the data files just loaded. Check each line that Rosetta should skip when looking for data. If every line of the data file contains data values, press the “No header lines available in this file” option. When header lines are defined, press “Next” to continue with template definition.

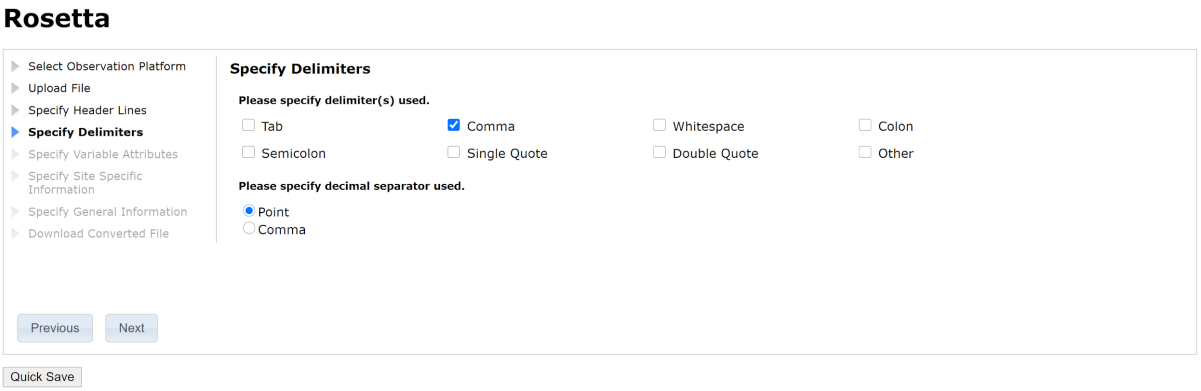
Screenshots from defining the header of the data file:



If you want to store the part of the template defined so far, press “Quick Save” before moving on.

Step 5: Specify delimiter and decimal separator

To be able to split the data lines into columns, and convert the text strings representing data values into numeric or time variables, Rosetta needs to know what characters are used to delimit columns and which is used as a decimal separator. Select from the screen shown below and the press “Next” to continue to the next step.



Step 6: Specify variables

Using the provided separators Rosetta split each data line into columns and displays the first lines in the screen shown below. To begin with, these columns are given dummy names such as “column 0”, “column 1” and so on. As you begin to define what parameters are found in the file, the columns will be renamed accordingly.

In the data file above the columns contain the following:

- Column 0: date of measurement, formatted as YYYY-MM-DD
- Column 1: time of day, formatted as HH:MM:SS
- Column 2: seconds of sunshine that day, a floating point value
- Column 3: air temperature, a floating point value
- Column 4: relative humidity, a floating point value
- Column 5: wind direction in degrees (relative to north), a floating point value
- Column 6: wind speed, a floating point value

We want to include five of these columns in the output NetCDF file, and show the definition of a few of these below. The remaining two columns will be skipped, which we indicate by checking the “Do not use this column of data” option in the “Enter variable attributes” screen.

To define the content of “column 0”, click on its name in the table header, check the “Assign a variable name” option in the “Enter Variable Attributes” screen and fill in the chosen name (in our case “date”). The variable name has to match one of the alternatives that pops up when you start to write. Then press the Tab key to register your choice. This will open the next part of the screen.

This column defines the time axis of the data, so check “Yes” for “Is this column a coordinate variable?” and select the appropriate type (“Date only (year, month, and/or day”). Next, select “Text” as the data type and enter the required metadata for the variable. All required metadata elements are marked with a red asterisk ‘*’ in Rosetta. For units, enter the string “YYYY-MM-DD” (without quotes!) or check the “show unit builder” option to get a list of allowed values for a unit for this variable. When you have finished, press “Done” to get back to the table. Now you will see that the column you just defined has changed name and been given a green checkmark to indicate that it is fully defined in the template.

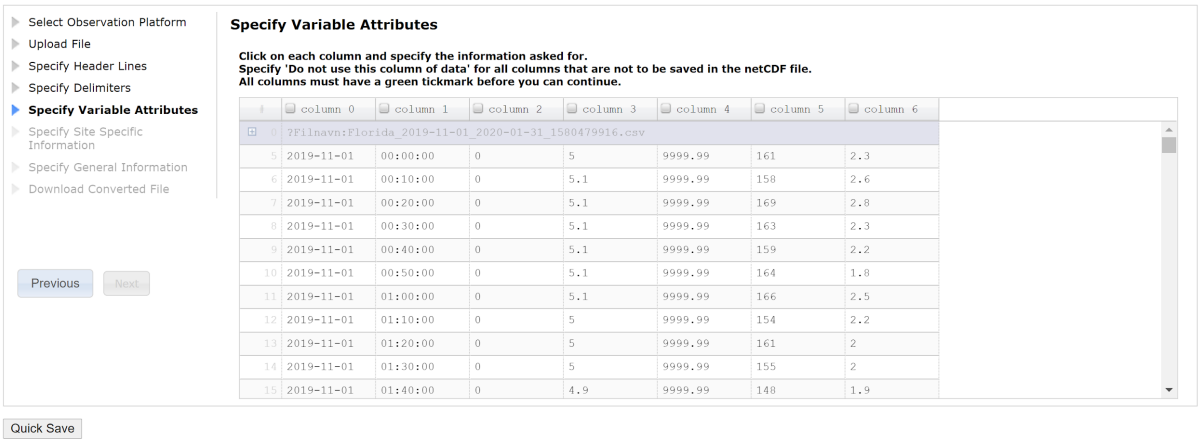
Once a variable has been successfully defined, you will see a checkbox with the label “use metadata from another column?“ when opening subsequent columns. Only click this option if you already have filled in a column of the same type because it will use all the input from the column you choose to copy from.

To define the content of “column 3”, click on its name in the table header, check the “Assign a variable name” option and fill in the chosen name (in our case “air_temperature”). This column is not a coordinate variable so we check “No” for the question “Is this column a coordinate variable?”. This column contains floating point values, so check “Float (decimal)” for data type. Then fill the required metadata. The correct string for units is “degree_C” (without the quotes); this string can be entered by hand or copied from the unit builder by checking this and selecting from a list of allowed values for air temperature.

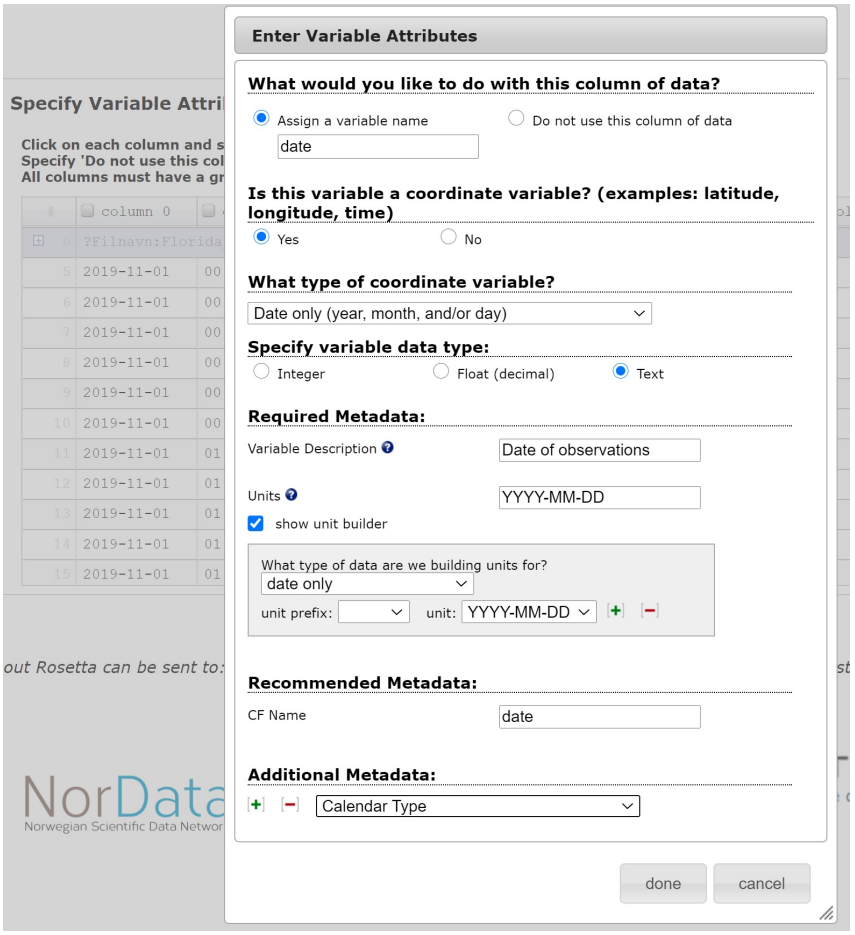
After defining the remaining three columns we want to include in the output NetCDF file, the table header should contain green tick marks for all columns, and we can press “Next” to get to the next step of the conversion process. (The columns that we wanted to skip are named “Do Not Use” in the table.)

Screenshots from the process of specifying variables:

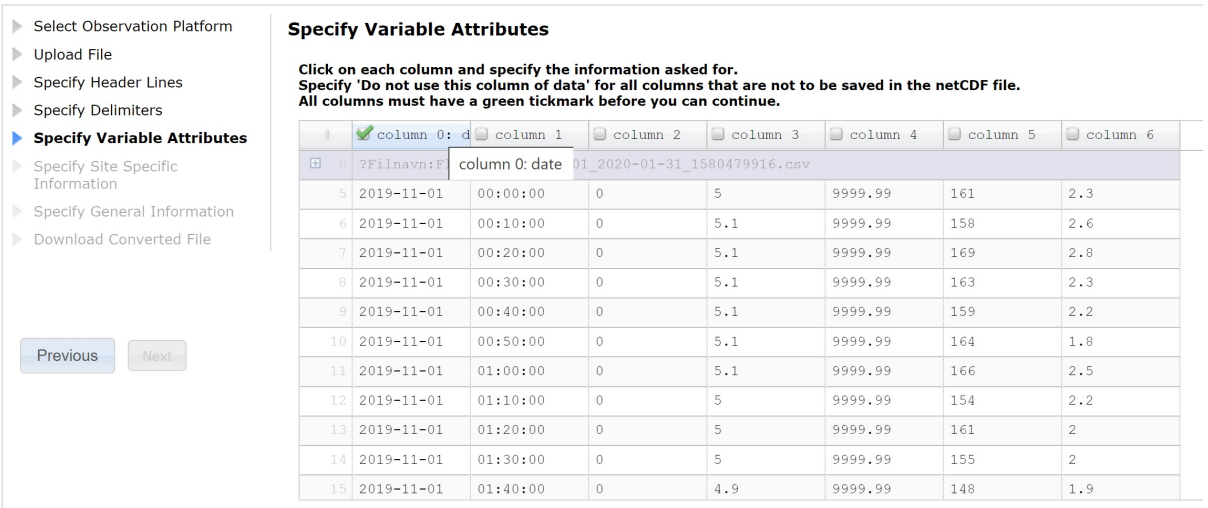
Rosetta

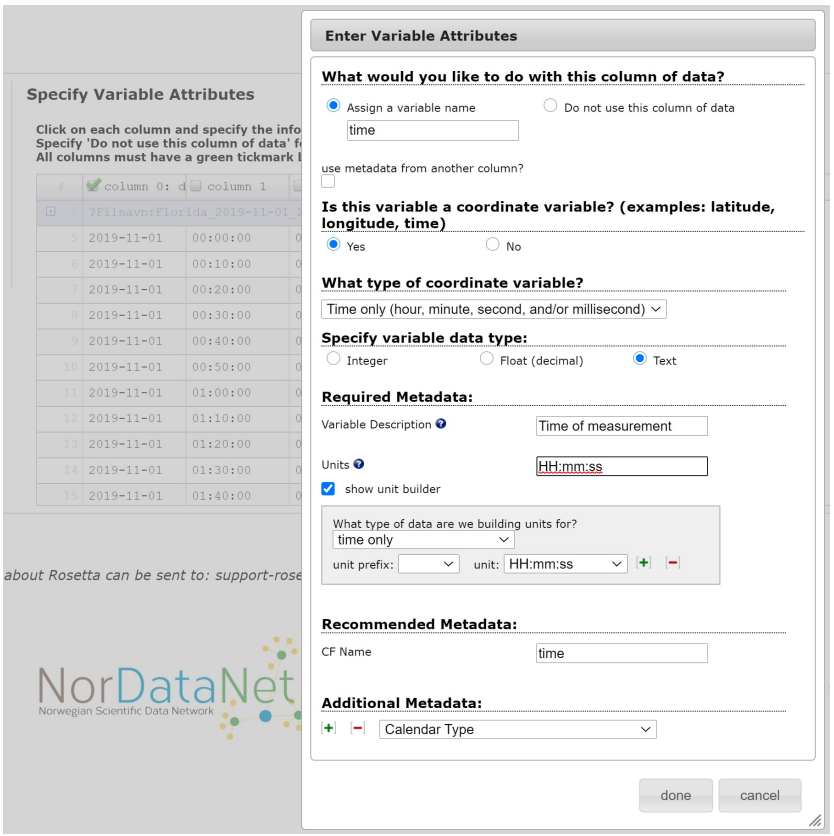


#	column 0	column 1	column 2	column 3	column 4	column 5	column 6
0	7Filnavn:Florida_2019-11-01_2020-01-31_1580479916.csv						
5	2019-11-01	00:00:00	0	5	9999.99	161	2.3
6	2019-11-01	00:10:00	0	5.1	9999.99	158	2.6
7	2019-11-01	00:20:00	0	5.1	9999.99	169	2.8
8	2019-11-01	00:30:00	0	5.1	9999.99	163	2.3
9	2019-11-01	00:40:00	0	5.1	9999.99	159	2.2
10	2019-11-01	00:50:00	0	5.1	9999.99	164	1.8
11	2019-11-01	01:00:00	0	5.1	9999.99	166	2.5
12	2019-11-01	01:10:00	0	5	9999.99	154	2.2
13	2019-11-01	01:20:00	0	5	9999.99	161	2
14	2019-11-01	01:30:00	0	5	9999.99	155	2
15	2019-11-01	01:40:00	0	4.9	9999.99	148	1.9

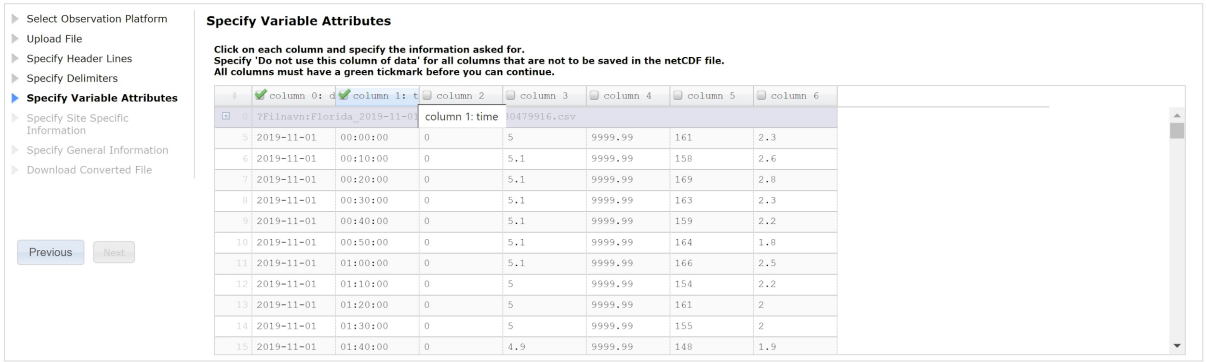

Specify Variable Attri
Click on each column and s
Specify 'Do not use this col
All columns must have a gr
column 0
?Filnavn:Florida
2019-11-01
out Rosetta can be sent to:
NorData
Norwegian Scientific Data Networ
Enter Variable Attributes
What would you like to do with this column of data?
Assign a variable name
Do not use this column of data
date
Is this variable a coordinate variable? (examples: latitude, longitude, time)
Yes
No
What type of coordinate variable?
Date only (year, month, and/or day)
Specify variable data type:
Integer
Float (decimal)
Text
Required Metadata:
Variable Description
Date of observations
Units
YYYY-MM-DD
show unit builder
What type of data are we building units for?
date only
unit prefix:
unit:
YYYY-MM-DD
Recommended Metadata:
CF Name
date
Additional Metadata:
Calendar Type
done
cancel

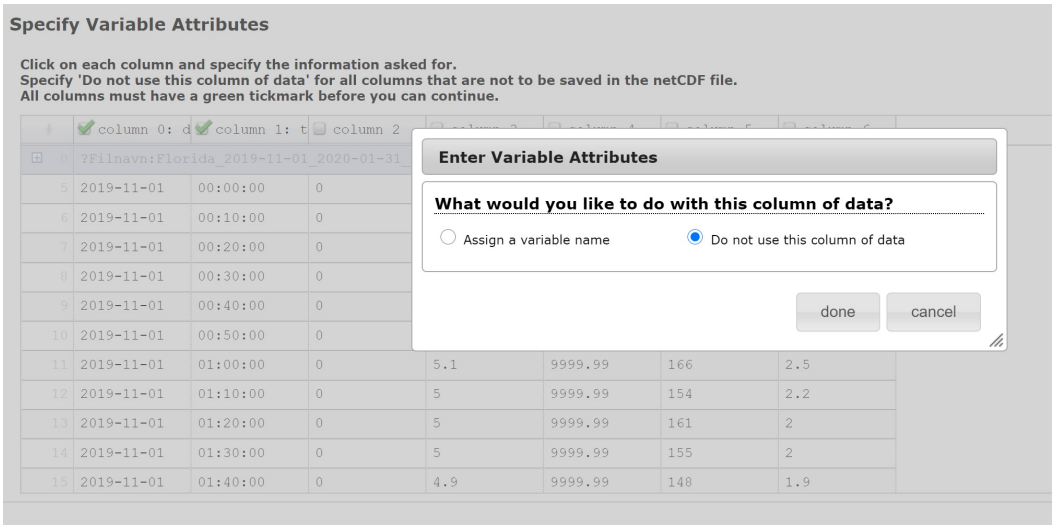
Rosetta


Select Observation Platform
Upload File
Specify Header Lines
Specify Delimiters
Specify Variable Attributes
Specify Site Specific Information
Specify General Information
Download Converted File
Previous
Next
Specify Variable Attributes
Click on each column and specify the information asked for.
Specify 'Do not use this column of data' for all columns that are not to be saved in the netCDF file.
All columns must have a green tickmark before you can continue.
column 0: date
column 1
column 2
column 3
column 4
column 5
column 6
?Filnavn:F
01_2020-01-31_1580479916.csv
2019-11-01 00:00:00 0 5 9999.99 161 2.3
2019-11-01 00:10:00 0 5.1 9999.99 158 2.6
2019-11-01 00:20:00 0 5.1 9999.99 169 2.8
2019-11-01 00:30:00 0 5.1 9999.99 163 2.3
2019-11-01 00:40:00 0 5.1 9999.99 159 2.2
2019-11-01 00:50:00 0 5.1 9999.99 164 1.8
2019-11-01 01:00:00 0 5.1 9999.99 166 2.5
2019-11-01 01:10:00 0 5 9999.99 154 2.2
2019-11-01 01:20:00 0 5 9999.99 161 2
2019-11-01 01:30:00 0 5 9999.99 155 2
2019-11-01 01:40:00 0 4.9 9999.99 148 1.9

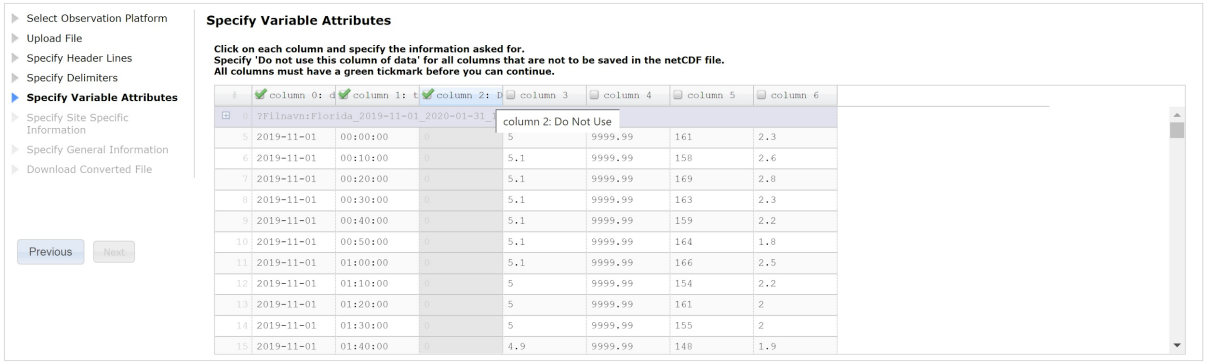

Specify Variable Attributes
Click on each column and specify the info
Specify 'Do not use this column of data' f
All columns must have a green tickmark
column 0: d
column 1
?Filnavn:Florida_2019-11-01_
2019-11-01 00:00:00
2019-11-01 00:10:00
2019-11-01 00:20:00
2019-11-01 00:30:00
2019-11-01 00:40:00
2019-11-01 00:50:00
2019-11-01 01:00:00
2019-11-01 01:10:00
2019-11-01 01:20:00
2019-11-01 01:30:00
2019-11-01 01:40:00
about Rosetta can be sent to: support-rose
NorDataNet
Norwegian Scientific Data Network
Enter Variable Attributes
What would you like to do with this column of data?
Assign a variable name
Do not use this column of data
time
use metadata from another column?
Is this variable a coordinate variable? (examples: latitude, longitude, time)
Yes
No
What type of coordinate variable?
Time only (hour, minute, second, and/or millisecond)
Specify variable data type:
Integer
Float (decimal)
Text
Required Metadata:
Variable Description
Time of measurement
Units
HH:mm:ss
show unit builder
What type of data are we building units for?
time only
unit prefix:
unit:
HH:mm:ss
Recommended Metadata:
CF Name
time
Additional Metadata:
Calendar Type
done
cancel

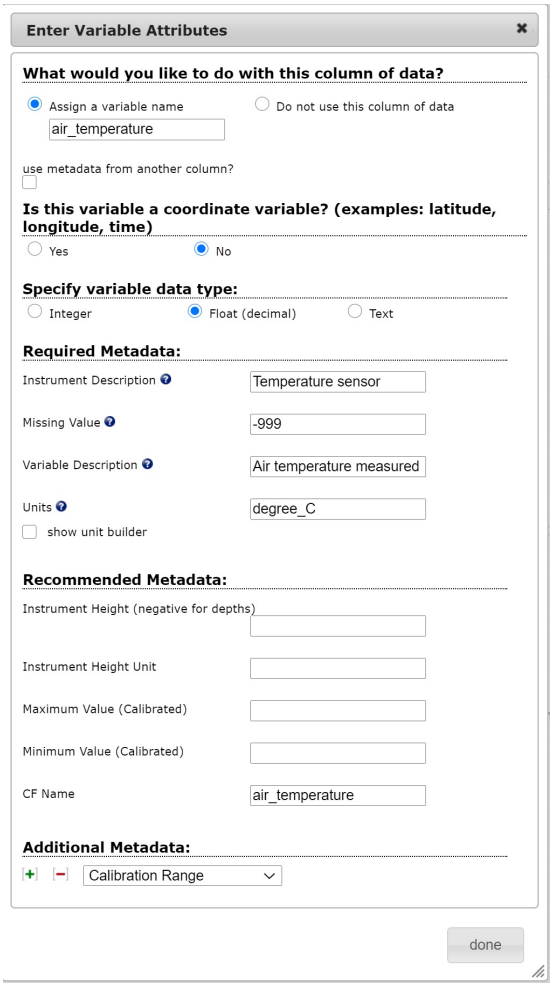
Rosetta


Select Observation Platform
Upload File
Specify Header Lines
Specify Delimiters
Specify Variable Attributes
Specify Site Specific Information
Specify General Information
Download Converted File
Previous
Next
Specify Variable Attributes
Click on each column and specify the information asked for.
Specify 'Do not use this column of data' for all columns that are not to be saved in the netCDF file.
All columns must have a green tickmark before you can continue.
column 0: d
column 1: t
column 2
column 3
column 4
column 5
column 6
?Filnavn:Florida_2019-11-01
0479916.csv
column 1: time
2019-11-01 00:00:00 0 5 9999.99 161 2.3
2019-11-01 00:10:00 0 5.1 9999.99 158 2.6
2019-11-01 00:20:00 0 5.1 9999.99 169 2.8
2019-11-01 00:30:00 0 5.1 9999.99 163 2.3
2019-11-01 00:40:00 0 5.1 9999.99 159 2.2
2019-11-01 00:50:00 0 5.1 9999.99 164 1.8
2019-11-01 01:00:00 0 5.1 9999.99 166 2.5
2019-11-01 01:10:00 0 5 9999.99 154 2.2
2019-11-01 01:20:00 0 5 9999.99 161 2
2019-11-01 01:30:00 0 5 9999.99 155 2
2019-11-01 01:40:00 0 4.9 9999.99 148 1.9

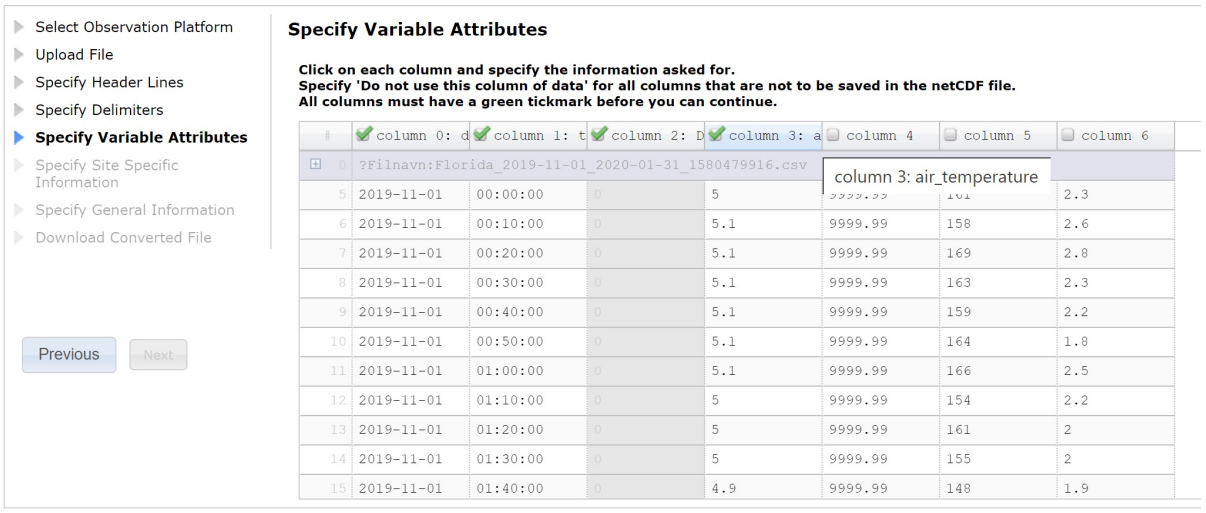

Specify Variable Attributes
Click on each column and specify the information asked for.
Specify 'Do not use this column of data' for all columns that are not to be saved in the netCDF file.
All columns must have a green tickmark before you can continue.
column 0: d
column 1: t
column 2
?Filnavn:Florida_2019-11-01_2020-01-31_
2019-11-01 00:00:00 0
2019-11-01 00:10:00 0
2019-11-01 00:20:00 0
2019-11-01 00:30:00 0
2019-11-01 00:40:00 0
2019-11-01 00:50:00 0
2019-11-01 01:00:00 0 5.1 9999.99 166 2.5
2019-11-01 01:10:00 0 5 9999.99 154 2.2
2019-11-01 01:20:00 0 5 9999.99 161 2
2019-11-01 01:30:00 0 5 9999.99 155 2
2019-11-01 01:40:00 0 4.9 9999.99 148 1.9
Enter Variable Attributes
What would you like to do with this column of data?
Assign a variable name
Do not use this column of data
done
cancel

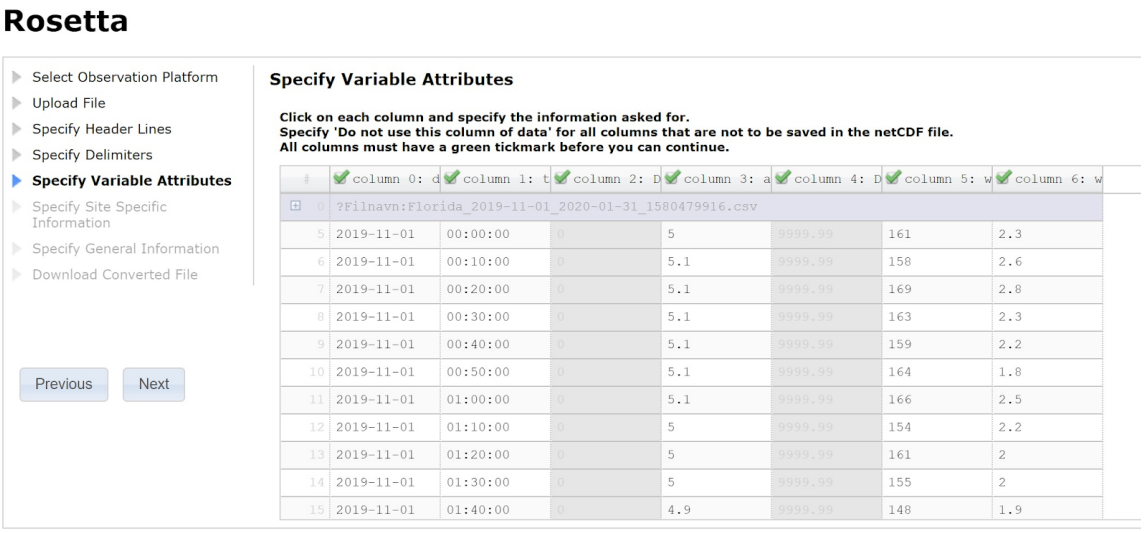
Rosetta





Rosetta





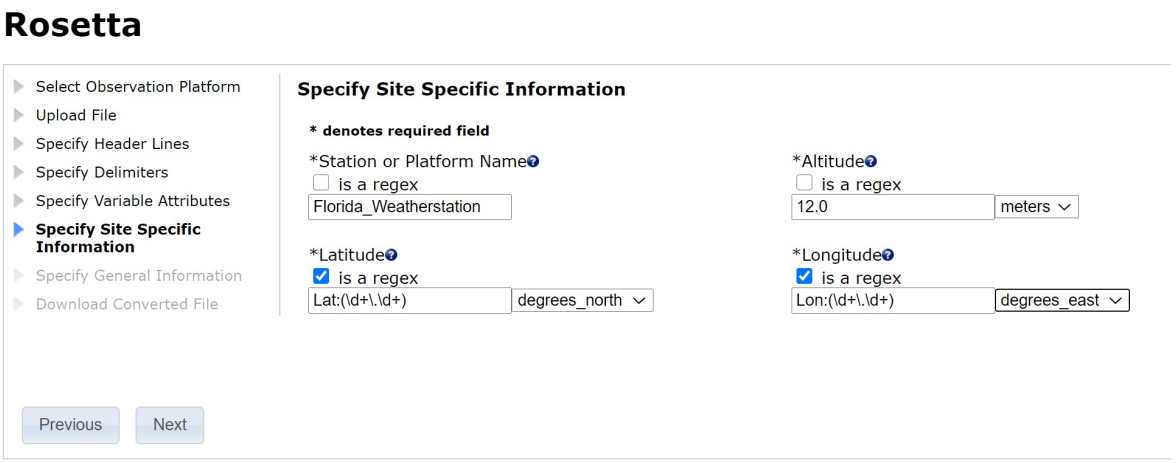
Step 7: Specify site specific information

Our data file contains a time series of meteorological observations. Hence, the site specific information to provide is the name of the site (i.e. the station name) and its location (latitude and longitude, in decimal degrees). The latter information is found in the header of the data file, so we can use two regular expressions to extract the location from the file. These regular expressions are:

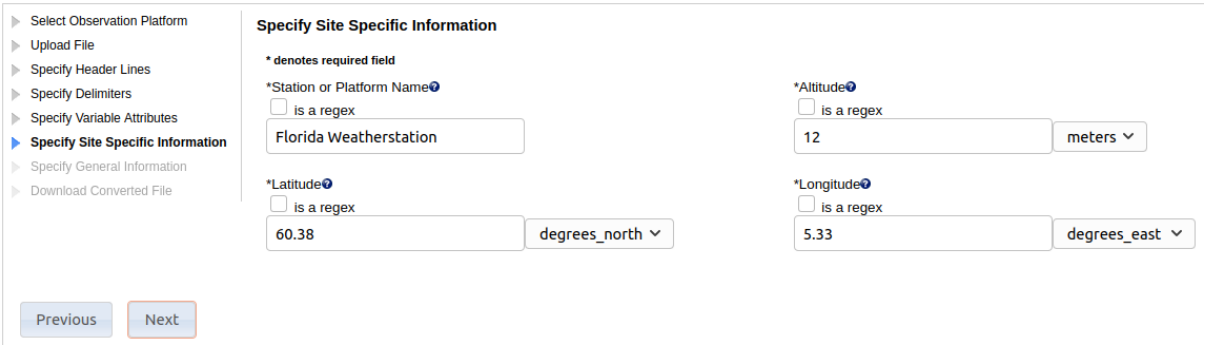
- Lat:(\d+\.\d+)
- Lon:(\d+\.\d+)

The name of the site we enter as a string. The altitude of the weather station has been extracted from the meteorological station's web page.

Screen for specifying the site information:



Alternative without regular expressions:



The values for Latitude and Longitude must be in the decimal degrees.

For Profile datasets, the date and time must be in ISO 8601 format. E.g. 2017-06-30T13:23:00Z or 2017-06-30.

Step 8: Specify general information

Additional descriptions (metadata) for the dataset can be given in the screen below. This allows you to specify important information such as

- Title of the dataset
- The data license that you publish the dataset under
- A short summary of the dataset
- Keywords to make it more easily searchable once published in a data repository
- Name and email for the scientists producing and publishing the data
- Name of the organisation (data centre) publishing the data

The text strings (or regular expressions) that are specified in these fields will be included in the new template, and used to fill the corresponding metadata elements in the output NetCDF file. All mandatory fields are marked with a red asterisk ('*').When all mandatory fields are filled, the "Next" button is enabled. Press this button and you proceed to the next step, which performs the conversion.

Screenshot for general information:

Specify General Information

*** denotes required field**

* Title
is a regex
Florida (Bergen) weather station data Nov-Dec 2019

* License
is a regex
NLOD

* Naming Authority
is a regex
no.uib.gfi

* ID
is a regex
Florida_2019-11-01_2020-01-31_1580479916

* ISO Topic Category
climatologyMeteorologyAtmosphere

* Keywords Vocabulary
GCMD Science Keywords

* Keywords
is a regex
Bergen Florida wind temperature

* Data Assembly Center
is a regex
Geofysisk intitutt (UIB)

* Summary
Data from met station at Florida in Bergen, Norway, during November and December in 2019.

* Processing Level
3

* Publisher Name
is a regex
Geofysisk intitutt (UIB)

Publisher Email
is a regex

Contributor Name
is a regex

Contributor Email
is a regex

Contributor Role
is a regex

Publisher Url
is a regex

Related Url
is a regex

Data Set Language
is a regex

Platform
is a regex

Source
is a regex

Project
is a regex

Citation

History
is a regex

Data Set Progress
is a regex

References
is a regex

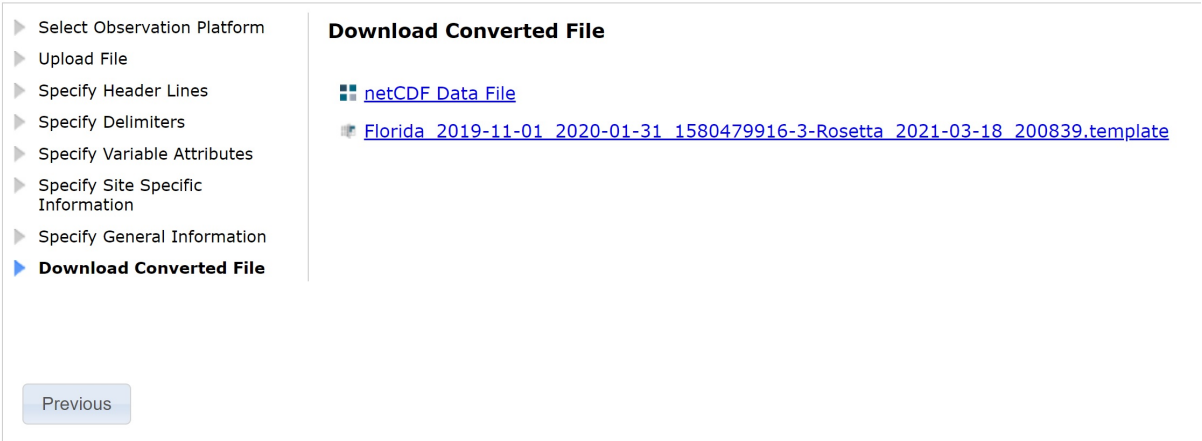
Comment
is a regex

Step 9: Download resulting files

This step uses the newly created template and converts the chosen columns of the ASCII data file to NetCDF format and adds the specified metadata. When conversion is completed Rosetta presents two download links allowing you to download the generated files.

Screenshot for NetCDF and template file download:

Rosetta



Resources

Sample templates and ASCII data files are available from https://tinyurl.com/Rosetta-NERSC

Summary

Rosetta enables you to convert ASCII CSV or Excel spreadsheet files to NetCDF, without having to write a single line of code. You can use a predefined template or make your own to specify how and what parts of the data file should be included in the generated NetCDF file. A template can be reused to convert other data files with the same structure, or used as the basis for making a new template to convert data files with a similar structure. Additional descriptions (e.g. site metadata) can be added interactively or extracted from the header of the input data file. For the latter, Rosetta uses string pattern matching techniques (regular expressions) to automatically extract e.g. geographic positions or date and times from the header of the data file.

Acknowledgements

The Rosetta conversion tool was originally developed by the Unidata Community Program managed by the University Corporation for Atmospheric Research (UCAR),, which operates the National Center for Atmospheric Research in the US. The development of Rosetta at the Nansen Center has been supported by the Research Council of Norway projects Norwegian Marine Data Centre (NMDC) (project no. 208849) and Norwegian Scientific Data Network (NorDataNet) (project no. 245967). Other research projects have contributed to Rosetta development by providing example data files and giving feedback on the implemented functionality, among others, Svalbard Integrated Arctic Earth Observing System – Knowledge Centre (SIOS -KC) (project no. 291644), Useful Arctic Knowledge - Partnership for Research and Education (UAK) (project nr. 274891), and Integrated Arctic Observation System (INTAROS) (Grant Agreement no. 727890).

References

Unidata, 2021. Rosetta format conversion tool, https://doi.org/10.5065/D6N878N2.

Appendix A Example files for profile data

The following snippet shows a .cnv file containing CTD data collected by a Seabird SBE 19plus V2 sensor, quality controlled and averaged to 2 meter intervals.

```
* Sea-Bird SBE 19plus V2 Data File:
* FileName = C:\Documents and Settings\Administrator\Desktop\UAK2020\uak2020_502.hex
* Software Version Seasave V 7.26.7.110
* Temperature SN = 6579
* Conductivity SN = 6579
* System UpLoad Time = Jan 06 1980 15:00:22
* NMEA Latitude = 76.6462
* NMEA Longitude = 19.4062
…
** Ship: KV_SVALBARD
** Cruise: UAK 2020
** Station: 502
** Depth: 198
** Operator: TSS
* Real-Time Sample Interval = 0.2500 seconds
* System UTC = 2020-06-23T10:38:00Z
# nquan = 7
# nvalues = 188
# units = specified
# name 0 = c0S/m: Conductivity [S/m]
# name 1 = depSM: Depth [salt water, m], lat = 76.00
# name 2 = potemp090C: Potential Temperature [ITS-90, deg C]
# name 3 = prdM: Pressure, Strain Gauge [db]
# name 4 = svCM: Sound Velocity [Chen-Millero, m/s]
# name 5 = sal00: Salinity, Practical [PSU]
# name 6 = flag: flag
# span 0 =   2.756824,   2.974668
# span 1 =      1.000,    188.000
# span 2 =    -1.8962,     2.3883
# span 3 =      1.011,    190.069
# span 4 =    1442.56,    1457.40
# span 5 =    32.8340,    35.3753
# span 6 = 0.0000e+00, 0.0000e+00
# interval = meters: 1
# start_time = Jun 23 2020 10:38:00 [System UTC, header]
# bad_flag = -9.990e-29
…
# binavg_min_scans_bin = 1
# binavg_max_scans_bin = 2147483647
# binavg_surface_bin = no, min = 0.000, max = 0.000, value = 0.000
# file_type = ascii
*END*
   2.925704      1.000      2.1877      1.011    1456.18    32.8985 0.0000e+00
   2.924292      2.000      2.1801      2.021    1456.15    32.8882 0.0000e+00
   2.923811      3.000      2.1764      3.032    1456.14    32.8855 0.0000e+00
   2.923709      4.000      2.1744      4.042    1456.15    32.8857 0.0000e+00
   2.923998      5.000      2.1753      5.053    1456.17    32.8877 0.0000e+00
…
```

This data file can be downloaded from Rosetta-NERSC-Resources in a folder named Seabird_CTD. A template that can be used to convert this ASCII data file to NetCDF is found at Rosetta-NERSC-Resources in a folder named Seabird_CTD.

Appendix B Example files for time series data

The following snippet shows a CSV file with observations from the Florida meteorological station at the Geophysical Institute, University of Bergen, Norway. Data were downloaded from https://veret.gfi.uib.no/?action=download.

```
Filnavn:Florida_2019-11-01_2020-01-31_1580479916.csv
Lat:60.38
Lon:05.33
Area:Florida-Bergen-Norge
Dato,tid,Solskinstid,Lufttemperatur,"Relativ luftfuktighet",Vindretning,Vindstyrke
2019-11-01,00:00:00,0,5,9999.99,161,2.3
2019-11-01,00:10:00,0,5.1,9999.99,158,2.6
2019-11-01,00:20:00,0,5.1,9999.99,169,2.8
2019-11-01,00:30:00,0,5.1,9999.99,163,2.3
2019-11-01,00:40:00,0,5.1,9999.99,159,2.2
2019-11-01,00:50:00,0,5.1,9999.99,164,1.8
2019-11-01,01:00:00,0,5.1,9999.99,166,2.5
2019-11-01,01:10:00,0,5,9999.99,154,2.2
2019-11-01,01:20:00,0,5,9999.99,161,2
2019-11-01,01:30:00,0,5,9999.99,155,2
2019-11-01,01:40:00,0,4.9,9999.99,148,1.9
2019-11-01,01:50:00,0,4.9,9999.99,157,2.2
2019-11-01,02:00:00,0,4.9,9999.99,161,2.3
2019-11-01,02:10:00,0,5,9999.99,163,2.3
2019-11-01,02:20:00,0,5.1,9999.99,171,2.3
2019-11-01,02:30:00,0,5.3,9999.99,174,2.8
2019-11-01,02:40:00,0,5.4,9999.99,159,2
2019-11-01,02:50:00,0,5.4,9999.99,165,2.3
2019-11-01,03:00:00,0,5.3,9999.99,165,3
2019-11-01,03:10:00,0,5.2,9999.99,173,2.8
2019-11-01,03:20:00,0,5.4,9999.99,171,3.1
2019-11-01,03:30:00,0,5.3,9999.99,167,2.1
2019-11-01,03:40:00,0,5.5,9999.99,171,2.3
2019-11-01,03:50:00,0,5.5,9999.99,168,2.5
2019-11-01,04:00:00,0,5.5,100,169,2.9
2019-11-01,04:10:00,0,5.4,99,178,3.4
2019-11-01,04:20:00,0,5.2,100,182,2.4
2019-11-01,04:30:00,0,5.2,9999.99,180,2.2
2019-11-01,04:40:00,0,5.1,9999.99,163,2.2
2019-11-01,04:50:00,0,5.1,9999.99,143,2.2
2019-11-01,05:00:00,0,5.4,99,162,3
2019-11-01,05:10:00,0,5.5,97,156,3.2
2019-11-01,05:20:00,0,5.4,96,152,3
2019-11-01,05:30:00,0,5.6,94,154,3.7
2019-11-01,05:40:00,0,5.5,93,154,3.3
2019-11-01,05:50:00,0,5.6,92,178,4.1
2019-11-01,06:00:00,0,5.6,90,174,4.6
2019-11-01,06:10:00,0,5.6,91,171,4.6
2019-11-01,06:20:00,0,5.4,92,160,4.3
2019-11-01,06:30:00,0,5.3,92,161,4.5
2019-11-01,06:40:00,0,5.4,92,159,4.6
2019-11-01,06:50:00,0,5.5,89,152,4.6
2019-11-01,07:00:00,0,5.5,89,152,4.3
2019-11-01,07:10:00,0,5.2,91,158,4
2019-11-01,07:20:00,0,5,93,162,3.7
…
```

This data file and a template that can be used for conversion can be downloaded from Rosetta-NERSC-Resources in a folder named MetStation_Florida.

Appendix C Example files for trajectory data

The following snippet of a CSV file with data collected by an unmanned surface vessel called a SailBuoy, from a field experiment in the Fram Strait in summer 2016.

```
% Seaguard II   Product number 5650   Serial number 1801. On SailBuoy ?SB Iskant? deployed
in Fram Strait 2016.


% Rec.Year     Hour    Record Number  CO2 optode # 69       #69     #69     #69     #69     #69
       #69     #69     #69     #69     Conductivity sensor #1376    #1376  #1376  #1376
#1376  Optode sensor #2537   #2537  #2537  Optode sensor #2551   #2551  #2551  pH Optode
5499#37 5499#37 System Par.    System Par.
% Rec.Year     Hour    Record Number  pCO2[uatm]     CO2[mg/l]      Temperature[Deg.C]
CalPhase[Deg]  DPhase[Deg]    C1RPh[Deg]     C2RPh[Deg]     C1Amp[mV]      C2Amp[mV]
RawTemp[mV]    Conductivity[mS/cm]   Temperature[Deg.C]    Salinity[PSU]  Density[kg/m3]
Soundspeed[m/s]       O2Concentration[uM]   AirSaturation[%]      Temperature[Deg.C]
O2Concentration[uM]   AirSaturation[%]      Temperature[Deg.C]    pH      Temperature[Deg.C]
       GPS Latitude[Deg]     GPS Longitude[Deg]
30.06.2016     20:40:00       6      489.968 1.367999       5.265263       31.39719
31.39719       64.20319       32.806 389.075 513.9634       611.1559       33.12206
5.41011 34.19186       1026.987       1471.365       398.479 100.6301       5.283304
388.5434       98.63229       5.486095       7.018903       5.262424       77.99095
3.181263
30.06.2016     20:50:00       7      469.0541       1.308763       5.282877       31.12162
       31.12162       63.93534       32.81372       390.5771       513.9194       610.6235
       33.14514       5.324608       34.30579       1027.087       1471.165       398.0173
       100.5288       5.289223       389.0276       98.67171       5.453039       7.439208
       5.280493       77.99055       3.191327
30.06.2016     21:00:00       8      460.9776       1.285676       5.294542       31.01051
       31.01051       63.81671       32.8062 390.9192       513.8643       610.2708
33.24416       5.316055       34.42833       1027.185       1471.286       396.8074
100.2513       5.300133       389.012 98.68472       5.459761       7.499966       5.292583
       77.99001       3.201878
30.06.2016     21:10:00       9      464.2966       1.293792       5.318565       31.04537
       31.04537       63.85447       32.8091 391.0192       513.7015       609.5444
33.27413       5.316055       34.46277       1027.212       1471.33 395.6103       100.006
5.322386       388.5174       98.6216 5.484468       7.512468       5.316317       77.98944
       3.211182
30.06.2016     21:20:00       10     461.3003       1.284232       5.344294       30.99577
       30.99577       63.80964       32.81387       390.7969       513.5674       608.7663
       33.29895       5.328033       34.47895       1027.224       1471.399       395.3341
       99.98319       5.340738       388.9698       98.65065       5.450502       7.528384
       5.342283       77.98876       3.220687
30.06.2016     21:30:00       11     461.5434       1.28485 5.345528       30.99853
30.99853       63.80807       32.80954       391.0323       513.4227       608.7289
33.34467       5.350273       34.50854       1027.245       1471.526       394.5273
99.79166       5.345636       388.9748       98.60802       5.433096       7.539204
5.344091       77.9881 3.23087
30.06.2016     21:40:00       12     461.0642       1.284667       5.321012       31.00154
       31.00154       63.81398       32.81244       390.5168       513.6245       609.4705
       33.35756       5.319489       34.55513       1027.285       1471.461       395.7533
       100.0234       5.315092       389.309 98.60985       5.400303       7.538395
5.317787       77.98846       3.245447
30.06.2016     21:50:00       13     454.957 1.281549       5.024206       31.0331 31.0331
63.81392       32.78082       390.4458       515.4869       618.4335       33.17656
5.062538       34.61255       1027.361       1470.495       404.0652       101.3898
5.034569       390.1268       98.635 5.328398       7.545804       5.022024       77.98921
       3.257905
30.06.2016     22:00:00       14     451.0114       1.27373 4.953966       31.00669
31.00669       63.79381       32.78712       390.5508       515.5406       620.5507
```

33.16626 5.093399 34.5686 1027.323 1470.564 403.0811 101.1635
5.042488 390.8219 98.60456 5.247053 7.545156 4.947441
77.98887 3.267775
30.06.2016 22:10:00 15 453.7019 1.277775 5.029291 31.01432
31.01432 63.79123 32.77691 390.8596 515.6187 618.2801
33.16979 5.007637 34.66184 1027.406 1470.334 407.1436
102.1033 5.012196 390.9528 98.58065 5.224575 7.556883
5.032427 77.9884 3.273687
30.06.2016 22:20:00 16 426.8588 1.21277 4.791449 30.73537
30.73537 63.49145 32.75608 392.0937 517.1923 625.4441
32.99825 4.82901 34.64877 1027.417 1469.59 418.2342 104.3086
4.799192 391.8062 98.59593 5.145813 7.55701 4.789618 77.98972
3.285247
30.06.2016 22:30:00 17 428.0244 1.215877 4.795938 30.75008
30.75008 63.51055 32.76047 391.8707 517.1356 625.309
32.99034 4.808372 34.66111 1027.429 1469.522 419.5575
104.6409 4.80003 392.7053 98.57622 5.049099 7.565363 4.793711
77.9903 3.296528
30.06.2016 22:40:00 18 427.5335 1.213138 4.825987 30.73169
30.73169 63.49197 32.76029 392.1042 516.959 624.4048
33.01128 4.827293 34.66574 1027.43 1469.605 418.5255 104.4574
4.827374 393.3568 98.61485 5.000025 7.572043 4.824236
77.99181 3.307652
30.06.2016 22:50:00 19 425.7151 1.206877 4.850676 30.69656
30.69656 63.45789 32.76134 392.2543 516.6379 623.6617
33.02962 4.84964 34.66377 1027.426 1469.693 417.306 104.2228
4.853235 393.3505 98.60178 4.995514 7.574779 4.848297
77.99421 3.315795
30.06.2016 23:00:00 20 431.6272 1.22097 4.909824 30.75713
30.75713 63.52243 32.7653 391.677 516.2448 621.8806 33.0755 4.908058
34.65616 1027.413 1469.922 415.4022 103.9058 4.912235
393.3346 98.56558 4.982847 7.583844 4.907745 77.9948
3.327235
30.06.2016 23:10:00 21 432.0412 1.219992 4.957491 30.74467
30.74467 63.52188 32.77721 391.765 516.0535 620.4445
33.11085 4.942413 34.6614 1027.414 1470.069 413.9754 103.6734
4.958746 393.4167 98.52766 4.959871 7.580613 4.955214
77.99739 3.335092
30.06.2016 23:20:00 22 432.6718 1.220666 4.982155 30.74403
30.74403 63.51588 32.77184 391.9126 515.7735 619.7012
33.13191 4.988754 34.63756 1027.389 1470.227 412.3542
103.3473 4.98871 393.9305 98.59568 4.936054 7.578448
4.98041 77.99919 3.338833
30.06.2016 23:30:00 23 431.4367 1.218943 4.942936 30.74178
30.74178 63.50959 32.76781 391.3005 515.9627 620.8831
33.11352 4.957863 34.6484 1027.401 1470.115 412.2912
103.2319 4.95133 393.7526 98.5284 4.927113 7.584423 4.940372
78.00155 3.324332
30.06.2016 23:40:00 24 435.3734 1.231664 4.907658 30.8102
30.8102 63.58142 32.77122 391.1663 516.2208 621.9459 33.09903
4.921806 34.66917 1027.422 1469.995 412.8964 103.2953
4.918356 394.5448 98.57632 4.868181 7.58095 4.905445
78.00131 3.314612
30.06.2016 23:50:00 25 454.3575 1.273791 5.153452 30.97586
30.97586 63.75793 32.78207 390.1699 514.7718 614.5336
33.27711 5.108833 34.68077 1027.41 1470.769 402.6834
101.2634 5.119089 393.8926 98.52522 4.912109 7.581867
5.157036 78.00317 3.301947
01.07.2016 00:00:00 26 467.9089 1.29929 5.41437 31.0564 31.0564 63.86175
32.80535 389.4678 513.2097 606.6459 33.45393 5.302368
34.68371 1027.389 1471.556 395.5664 100.1384 5.378284

```
        393.4456        98.53271        4.958979        7.576968        5.411978        78.00467
        3.290077
01.07.2016      00:10:00        27      454.3582        1.270552        5.222803        30.94937
        30.94937        63.7481 32.79873        389.5665        514.0494        612.439 33.36375
        5.245888        34.63843        1027.36 1471.27 398.209 100.4245        5.230064
393.1877        98.51691        4.978189        7.587081        5.219887        78.00387
3.29543
…
```

This data file and a template that can be used for conversion can be downloaded from Rosetta-NERSC-Resources in a folder named Sailbuoy_ArcticOceanAcidification,

Appendix D Example Date/Time formats

Date formats

yyyy-MM-dd

```
2019-11-01
2020-01-05
2021-03-17
```

YYYY/MM/dd

```
2019/11/01
2020/01/05
2021/03/17
```

MM-dd-YYYY

```
11-01-2019
01-05-2020
03-17-2019
```

dd-MM-YYYY

```
01-11-2019
05-01-2020
17-03-2021
```

YYYY-DDD

```
2019-305          # November 1 is day 305 in a non leap year.
2020-005
2020-077          # March 17 is day 77 in a leap year.
```

Time formats

HH:mm:ss # Uppercase HH is 24 hours 00-23, lowercase hh is 12 hours 01-12

```
00:00:00
02:10:31
13:20:12
```

hh:mm:ssZ

```
00:00:00+01       # CET is universal time +01 hour
```

```
02:10:31+01
13:20:12+01
```

hh:mm:ssZ

```
00:00:00Z         # Universal time can be marked as Z or +00 or +0000
02:10:31Z
13:20:12Z
```

Data files and templates that use these data formats can be downloaded from Rosetta-NERSC-Resources-Alternative_timestamps.