How to generate FAIR drone-based data: Data protocol

– Richard Hann, richard.hann@ntnu.no
Version 1.0 (09.12.2024)

The following is an explanation of the challenges of implementing typical drone data into the SIOS Data Portal and also includes a step-by-step description of how to process data to make it compatible.

Digital elevation models (DEMs) and orthomosaic maps are the most common drone data results. The typical file format for these two results is GeoTIFF. GeoTIFF is a public domain metadata standard that enables georeferencing information to be embedded within an image file. The GeoTIFF format embeds geospatial metadata into image files such as aerial photography, satellite imagery, and digitized maps so that they can be used in GIS applications.

In contrast, the required data format in the SIOS Data Portal – which ensures that data is stored following the FAIR (findability, accessibility, interoperability, and reusability) principle, is NetCDF. NetCDF (network common data form) is a file format for storing multidimensional scientific data (variables) such as temperature, humidity, pressure, wind speed, and direction.

In order to make drone-based GeoTIFF results compatible with the SIOS Data Portal, the following key steps have to be conducted:

1. **GeoTIFF→ NetCDF:** Translate the GeoTIFF data file into NetCDF format.
2. **Metadata:** Prepare the metadata to describe the dataset.
3. **Metadata + NetCDF:** Include the metadata into the NetCDF data file.
4. **Repository:** Upload into a repository that is accessible for SIOS Data Portal.

The suggested method requires the use of Python software. Python is a high-level, easy-to-read programming language known for its clear syntax and versatility. Other programming languages, such as Matlab or C/C++, may also be suitable for this task, but they are not described here. We suggest the use of Anaconda, a popular open-source distribution of Python programming languages, specifically designed for scientific computing, data science, and machine learning. In addition, the GDAL library is used. GDAL is a translator library for raster and vector geospatial data formats.

The following drone-based dataset is used as an example (DOI: 10.18710/IMPEG8).

Additional code can be downloaded from GitHub
https://github.com/imatero/Dronedata_to_CF_NetCDF

Preparations

- Install Anaconda: download
- Open the Anaconda Prompt
- The starting point is a DEM or orthomosaic in GeoTIFF format, e.g. from the example dataset *01_Lagoon Pingo_2020_DEM.tif*

Option 1: GeoTIFF → NetCDF with QGIS

- Install and open QGIS: download
- Add the GeoTIFF Layer:
  - Go to the top menu and click on Layer > Add Layer > Add Raster Layer.
  - In the dialog box that appears, click Browse, locate your GeoTIFF file, and select it.
  - Click Open, and the GeoTIFF will appear in the Layers panel and map canvas.
- To ensure that the data is ready for export, you can inspect the GeoTIFF's metadata.
  - Right-click on the GeoTIFF layer in the Layers panel.
  - Select Properties.
  - Check the Information tab for details about projection, resolution, etc.
- Go to Raster Conversion Tool:
  - Click on the top menu Raster > Conversion > Translate (Convert Format).
- Set Input and Output:
  - In the Translate (Convert Format) dialog:
  - Input Layer: Select your GeoTIFF layer.
  - Output File: Click ... to choose the destination folder and set the file name with the .nc extension (e.g., output.nc).
- Configure NetCDF Settings:
  - Under Output Format, select NetCDF from the dropdown menu.
  - If additional configuration is required (e.g., specifying variables for NetCDF), you may need to click on Advanced Parameters or use the Edit Parameters field.
- Export:
  - Click Run to start the conversion process.
  - The NetCDF file will be saved to the specified location.

Option 2: GeoTIFF → NetCDF with GDAL

- Install the GDAL library

```
conda install -c conda-forge install
```

- Open the Anaconda Prompt
- Navigate to the folder where the GeoTIFF file is located

```
cd C:\<file_directory>, e.g. cd C:\Pingo
```

- Translate the GeoTIFF data file into NetCDF

```
gdal_translate -of NETCDF "02_Lagoon Pingo_2020_
Orthomosaic.tif" "Pingo_Ortho.nc"
```

Metadata

- Compile all relevant metadata according to Attribute Convention for dataset Discovery (ACDD) requirements. More information are given here and here.
- Download the example.csv from the GitHub
- Fill in the data into the metadata.csv
- Keyword selector: link
- Overview of attributes: link

Metadata + NetCDF

- Download the dronedata_to_cf_netcdf.py from the GitHub into your working directory
- Open the Anaconda Prompt
- Install required libraries

```
conda install -c conda-forge netCDF4
conda install -c conda-forge xarray
conda install -c conda-forge subprocess
```

- Open the script *dronedata_to_cf_netcdf.py* in an text editor
- In that document, specify paths to the input and output files
- Save the script
- Run the script

```
python dronedata_to_cf_netcdf.py
```

- The final output will be located in the folder and named e.g. *Pingo_DEM_with_metadata.nc*

Repository

- Use the SIOS dataset validation form to check if your file: here
- If no errors occur, your dataset is ready for publishing
- Find a suitable repository to upload: here
- Your data is now available on SIOS Data Portal