

Handling of acoustic data in integrated data management systems

Torill Hamre

Nansen Environmental and Remote Sensing Center (NERSC)

UAK Winter School 2018, UNIS, Longyearbyen, Svalbard







Topics

- Data lifecycle
 - Plan for data management
 - Data collection
 - Getting data into your system
 - Preparing for publishing your data
 - Uploading data to a repository
 - Integrating your data with other sources
- Data portals for acoustic data
- Why data management is important
- Summary and resources

Data lifecycle

- Elements of data management (DataONE Best Practices)
- Start thinking about and doing data management early!
- Start with making a Plan
- Follow up with actions in each step
- Good tools will help
 - <u>DataONE investigator toolkit</u>
 - <u>Rosetta file conversion (Unidata)</u>
 - <u>NetCDF Utilities (Unidata)</u>
 - <u>GDAL (Geospatial Data</u>
 <u>Abstraction Library</u>)
- The most important part of this process is YOU!







DataONE - Data Observation Network for Earth

Data lifecycle

- Data lifecycles can vary with project type
 - Primary data collection (red)
 - Data synthesis / Assessment (green)
- Your data lifecycle model may be different, but must include
 - PLAN
 - DESCRIBE
 - PRESERVE

to support data reuse

- Exercise for this afternoon:
 - Describe what does your data lifecycle model looks like





Plan for data collection/compilation

- Data Management Plan
 - What data will be collected?
 - By which instruments?
 - Where is the documentation?
 - What context was data collected in?
 - Methods applied for processing and QC?
 - What formats will we get data in?
 - What formats do we want to work with?
 - What formats do funders expect?
 - Where to store data for sharing?
 - How much time&money to spend?
- Scale plan and DM to project size
 - Automate where possible
 - Use best practices







Data collection



Getting data into your system

- Passive acoustic data often recorded as WAV files
- Use own/other scientists' scripts for reading the data
 - Python: <u>scipy.io.wavfile.read() Examples</u>
 - Matlab: <u>audioread function</u>
 - R: <u>readWave function</u>
- Ingest through processing and analysis tools, e.g.
 - <u>Audacity</u> (matlab)
 - <u>PAMGuide</u> (matlab & R)
 - <u>TRITON</u> (matlab)
- More on this in the afternoon (lecture+exercise)











Preparing for publishing your data

- Compile both metadata and documentation
- Metadata
 - Discovery metadata, e.g. area covered, time period, parameters, ...
 - Usage metadata, e.g. processing history, units, data quality, ...
 - Avoid making new standards; utilize the existing ones
- Decide on suitable tools for formatting
 - Depends on format used in processing & analysis tools
 - Select a standard data format and metadata standard
 - Data formats should be self-describing
 - Files should include all needed metadata





Preparing for publishing your data

- Rosetta
 - Open source tool, written in Java, <u>https://github.com/Unidata/rosetta</u>
- Web-based application that converts ASCII files to NetCDF
 - Can read metadata from header(s), and add user defined metadata
- Saves "setup" as templates for future use



Rosetta

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

MMDC on

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.

norwegian marine data centre 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.

Data Services and Tools for Geoscience



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.



S 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





and Metadata

Define Coordinate Variables and Metadata

Preparing for publishing your data

- 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)
- What tools are available?
 - GIS (commercial, free)
 - Data processing and analysis tools (commercial, free)
 - Open source libraries and community utilities
 - Rosetta (free) for ASCII to NetCDF











INTAROS

Preparing for publishing your data

- 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 <u>https://nmdc.no/</u>



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



Uploading data to a repository

• Decide repository

- Must offer a DOI
- Must support metadata standards and search APIs
- Should be easy to upload
- Should offer human support
- Should have long term funding
- Metadata must include:
 - Area and time range
 - Summary of content
 - PI(s) and their e-mails
 - Parameters contained
 - Links to documentation
 - Data license
 - Funding organisation









ROS

Uploading data to a repository

SIOS Data Access Point

- Once uploaded
 - Easy to search
 - free text
 - area
 - time rang
 - keywords ┈ "

orwegian leteorological hstitute	MET Norway Thredds Service THREDDS Data Server	
TITLE	met-arome-arctic-2p5km-extracted	
Metadata key	Metadata value	Z
Available M	1etadata	I
Searc	ch results	
Number o	of datasets found: 457	
Home / Se	earch results	
	LAST UPDATED: NOVEMBER 21, 2018	

- Show on majCatalog http://thredds.met.no/thredds/catalog/aromearcticlatest/catalog.html
- Access to dat^{Dataset:} AROME Arctic latest/arome_arctic_extracted_2_5km_latest.nc
 - Data size: 12.81 Gbytes
 - ID: aromearcticlatest/arome_arctic_extracted_2_5km_latest.nc

Access:

- 1. **OPENDAP:** /thredds/dodsC/aromearcticlatest/arome_arctic_extracted_2_5km_latest.nc
- 2. HTTPServer: /thredds/fileServer/aromearcticlatest/arome_arctic_extracted_2_5km_latest.nc
- 3. WMS: /thredds/wms/aromearcticlatest/arome_arctic_extracted_2_5km_latest.nc
- 4. WCS: /thredds/wcs/aromearcticlatest/arome_arctic_extracted_2_5km_latest.nc

Dates:

• 2018-12-04T15:15:36Z (modified)

Viewers:

- Godiva2 (browser-based)
- NetCDF-Java ToolsUI (webstart)





- Acoustic Data Portal: https://iqoe.org/acoustic-data-portal
 - Data access, literature, marine animal sounds, ...



NOAA passive acoustic data viewer

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION

NOAA > NESDIS > NCEI (formerly NGDC) > Maps

Layers

Passive Acoustic Data

Additional Filters 🗙 Reset

Passive acoustic data are used by NOAA and other agencies and institutions for a wide range of activities including monitoring living marine resources, monitoring of earthquake and geological activity, and assessing impacts of anthropogenic noise on marine life.

Information on the archive is available on the <u>NCEI Passive Acoustic Data Archive</u> page.

The NOAA Ocean Noise Reference Station Network, through a unique collaborative effort across offices in NOAA's OAR, NMFS, and NOS and the Department of Interior's National Park Service, is NOAA's first ever acoustic monitoring system deployed broadly throughout the US exclusive economic zone, and enables the collection of consistent and comparable multi-year acoustic data sets covering all major regions of the US.

If you have any questions on the collection and use of the Ocean Noise Reference Station Network data, please contact NRS inquiries@noaa.gov

Online Resources:

- Discovery of Sound in the Sea
- Ocean Biogeographic Information
 System



TILL

IMOS Acoustic Data Viewer





<u>Alaska Ocean Observing System (AOOS)</u>







Alaska Ocean Observing System (AOOS)







Why data management is important

- For the researcher
 - Good organisation helps in finding needed information quickly
 - like the files you created three months ago
 - and associated documentation...
 - and scripts used to process, analyse and visualise...
 - Supports reuse
 - other scientists can repeat your analysis
 - verify that the same input gives the same output
 - perhaps suggest enhancements for new projects with you!
 - provide training for next generation







Why data management is important

- For the researcher
 - Improve data quality
 - more efficient data handling
 - needed metadata is in place
 - procedures (and scripts) for quality control are well documented
 - Data are securely stored
 - reusable by yourself and others
 - available for next generation
 - citable (e.g. by DOIs)
 - can lead to new research







Why data management is important

- For funders
 - Data collection is costly
 - Must maximize availabliity and reuse
 - Faciliate for longer time series of important climate parameters
 - Enables new projects to reuse earlier data
 - Also benefits the scientists (citation, DOI, impact rating)
 - Sustainable solutions for long term storage
 - Standardisation to facilitate interoperable DM systems
 - In brief, return on investment (ROI)





Summary and resources

- Key points
 - Planning data management is crucial
 - The plan must be followed up and acted upon
 - Many templates and tools to help you
 - Planning a little up front can save a lot of effort later
 - First, decide on the lifecycle model for your datasets
 - Then, plan accordingly and follow up; revise plan as needed
 - Upload data to a repository to facilitate reuse
 - Investigate current data portals for acoustic data
- Resources:
 - Strasser C, Cook R, Michener W, Budden A., 2012 <u>Primer on data management: what</u> <u>you always wanted to know</u>. Technical report, DataONE.
 - DataONE Data Management modules, <u>https://www.dataone.org/education-modules</u>
 - Acoustic Data Portal: <u>https://iqoe.org/acoustic-data-portal</u>



