

INTAROS metadata harmonization for ocean mooring data

Arnfinn Morvik
arnfinn.morvik@hi.no



Harmonizing data and metadata

- Background:
 - Large project with various data sources and data providers
- Goal:
 - Define standard metadata and datafile format for publishing and dataset sharing



Metadata harmonization

- Why harmonize metadata within INTAROS?
 - Several different instruments for point measurements
 - microCAT
 - RBR Concentro
 - AADI SeaGuard SW
 -
 - Different instruments for profiling measurements (ADCP)
 - Teledyne RDI LR-ADCP
 - Signature 250
 -
 - Instrument data handled by different institutions/scientists



Metadata

- <https://en.wikipedia.org/wiki/Metadata>:
 - Data about data
- Metadata for publishing
 - Search metadata used to describe the dataset when it is published at a data portal.
- Specific metadata for the dataset
 - Description of methods, instruments, parameters, units, etc.



Metadata harmonization

- Mandatory metadata
- Mandatory parameter list from similar instruments
- Harmonize parameter names
- Harmonize variable types
- Harmonize units
 - Units to use for different parameters
 - How to name the unit
- Use of standard vocabularies



File format harmonization

NetCDF

- Machine independent data format
 - Self-describing
 - Portable
 - Scalable
 - Appendable
 - Shareable
 - Archivable
- <https://www.unidata.ucar.edu/netcdf/>



Metadata harmonization - INTAROS

- Mandatory global attributes

<ul style="list-style-type: none">• Title• Source• Institution• Temporal extent• Spatial extent• Feature type• Conventions• License	CF
<ul style="list-style-type: none">• Keywords• Keyword vocabulary• Summary• Project• Id	ACDD
<ul style="list-style-type: none">• Citation• Primary investigator	OceanSITES
<ul style="list-style-type: none">• Cruise details	INTAROS

time_coverage_start
time_coverage_end

geospatial_lat_min
geospatial_lat_max
geospatial_lon_min
geospatial_lon_max
geospatial_vertical_min
geospatial_vertical_max
geospatial_vertical_positive



Metadata harmonization - INTAROS

- Recommended global attributes

<ul style="list-style-type: none">Data modePlatform codeData assembly centerNetCDF version	OceanSITES
<ul style="list-style-type: none">Comments	CF
<ul style="list-style-type: none">ContributorsCreatorsPublisher	ACDD

creator_name
creator_email
creator_url

Full list at:

<https://docs.google.com/spreadsheets/d/1Za6IXGQDmPsMDNas6R0i5l4ip0Kdc-W0/edit#gid=1636839350>



NetCDF Conventions

ACDD 1.3

https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3

CF 1.8

<https://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html>

OceanSITES

http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf



Metadata harmonization

- Examples of vocabularies
 - SeaDataNet
 - BODC
- Definition of a parameter
 - BODC P01
- Definition of a unit
 - BODC P06
- List of vocabularies can be found here:
<https://vocab.seadatanet.org/search>



NetCDF file format harmonization

parameter	type	dimension
instrument_id	char	(numInstruments, shortStr)
instrument_owner	char	(numInstruments, shortStr)
latitude	float32	(numInstruments,)
longitude	float32	(numInstruments,)
depth	float32	(numInstruments,)
instrument_description	char	(numInstruments, maxStr)
data_processing	char	(numInstruments, maxStr)
timeseries_id	char	(numInstruments, shortStr)
time	double	(numInstruments, obs)
var	float32/int	(numInstruments, obs)

CF Standard Name Table

<https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html>



NetCDF file creation tools

- Rosetta
 - <https://www.unidata.ucar.edu/software/rosetta/>
 - <http://tomcat.nersc.no/rosetta/>
- Python
 - netCDF4
 - Xarray
- R
 - RNetCDF



NetCDF Metadata test

- Online tool for compliance test of NetCDF files
 - ACDD 1.1
 - ACDD 1.3
 - CF 1.6
 - CF 1.7
 - CF 1.8

<https://compliance.ioos.us/index.html>



Data publishing

Metadata for publishing

- Retrieved from global attributes

Data in open format

- NetCDF

License

- Creative Commons: <https://creativecommons.org/licenses/>

Data paper

- Data Science Journal: <https://datascience.codata.org/>



Data publishing portals



EMODnet Product Catalogue

