The European common data and metadata model for real-time High Frequency Radar surface current data

Lorenzo Corgnati (1), Carlo Mantovani (1), Anna Rubio (2), Emma Reyes (3), Celine Quentin (4), Simone Cosoli (5,6), Antonio Novellino (7), Julien Mader (2), Annalisa Griffa (1)

(1) CNR-ISMAR, Lerici, Italy; (2) AZTI Marine Research, Pasaia, Spain; (3) ICTS-SOCIB, Palma de Mallorca, Spain; (4) Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France; (5) OGS, Trieste, Italy; (6) Ocean Graduate School and the UWA Oceans Institute, The University of Western Australia, Crawley, Australia; (7) ETT SpA, Genova, Italy

Background of research

High Frequency Radars (HFR) have become invaluable tools in the field of operational oceanography for monitoring surface currents, waves and winds, with direct applications in Search and Rescue, renewable energy, fishery management and monitoring of pollutants and biological quantities. They are increasingly used to support decision-making by coastal ocean users and managers, and it is expected that HFR surface current data will be soon systematically ingested in data assimilation processes. It is then crucial to promote and distribute high quality HFR data for scientific, operational and societal applications.

An appropriate data description complying with an accepted standard, is crucial for enforcing discovery and access. The comprehensive metadata description is a prerequisite for the full implementation of EuroGOOS, providing an inventory of the continuously available data for operational models, and for creating and giving an overview of marine monitoring programmes relevant for the Marine Strategy Framework Directive (MSFD) implementation.

Activity and results

Active international initiatives and ongoing projects aim at fostering and promoting the use of HFR technology in Europe. As part of these efforts, a model for data and metadata was defined and implemented for becoming the official European standard for producing near real-time HFR surface current data and for ensuring efficient and automated HFR data discovery and interoperability. The model has been implemented according to the standards of Open Geospatial Consortium (OGC) for access and delivery of geospatial data, and compliant with the Climate and Forecast Metadata Convention CF-1.6, the OceanSITES convention, the Copernicus-InSituTAC-SRD-1.4 and the INSPIRE directive. The model has been defined following the guidelines of the DATAMEQ working group and it fulfils the recommendations given by the Radiowave Operators Working Group (ROWG).

The model specifies the file format (i.e. netCDF-4 classic model), the global attribute scheme, the dimensions, the coordinate, data and Quality Control (QC) variables and their syntax, the QC procedures and the flagging policy for both radial and total data.

Conclusions

HFR technology is rapidly expanding in Europe, and there is the need for promoting and distributing high quality HFR data for scientific and societal applications.

A common data and metadata model was implemented to ensure efficient and automated HFR data discovery and interoperability across distributed and heterogeneous earth science data systems. A battery of mandatory QC tests was also defined, in order to ensure the delivery of high quality data. Further activities are planned to make the model compliant with the SDC CF extension model.

The data and metadata profile for netCDF-4 classic format HFR data

Profile version history

- 2015
  - RITMARE — CF, ACDD, INSPIRE — **v1.0**
- 2017
  - JERICO-NEXT INCREASE — CF, ACDD, INSPIRE; CMEMS SRD & PUM — **v2.0**
- 2018
  - CMEMS-INSTAC SDC — CF, ACDD, INSPIRE; CMEMS SRD & PUM; SDC CDI — **v2.1**
- 2019
  - CMEMS-INSTAC SDC — CF, ACDD, INSPIRE; CMEMS SRD & PUM; SDC CDI; SDC CF Extension — **v2.2**

Global attributes

The Mandatory Attributes include attributes necessary to comply with CF-1.6, OceanSITES and CMEMS IN-SITU TAC conventions. The global attributes required for the SeaDataCloud (SDC) CDI scheme and the SDC CF extension have been added as mandatory as well.

The Recommended Attributes include attributes necessary to comply with INSPIRE and Unidata Dataset Discovery conventions.

Discovery and Identification
site_code (EDIOS Series ID)
platform_code (EDIOS Platform ID)
data_mode
DoA_estimation_method
calibration_type
last_calibration_date
calibration_link
title
summary
source
source_platform_category_code
institution
institution_edmo_code
data_assembly_center
id
Project (EDMERP code)
naming_authority
keywords
keywords_vocabulary
comment

data_language	
data_character_set	
metadata_language	
metadata_character_set	
topic_category	
network	

Geo-spatial-temporal
data_type
feature_type
geospatial_lat_min
geospatial_lat_max
geospatial_lon_min
geospatial_lon_max
geospatial_vertical_min
geospatial_vertical_max
time_coverage_start
time_coverage_end
area
geospatial_lat_units
geospatial_lon_units
geospatial_vertical_resolution

geospatial_vertical_units	
geospatial_vertical_positive	
time_coverage_resolution	
time_coverage_duration	
reference_system	
grid_resolution	
cdm_data_type	

Conventions used
format_version
Conventions
netcdf_version
netcdf_format

Publication information
update_interval
citation
distribution_statement
publisher_name
publisher_email
publisher_url
license
acknowledgment

Provenance
date_created
history
date_modified
date_update
processing_level
contributor_name
contributor_role
contributor_email

Mandatory
Recommended
Suggested

Where **time** is specified as a string, the **ISO8601 standard** is used.

Dimensions and coordinates

Dimensions provide information on the size of the data variables, and tie coordinate variables to data. CF recommends that if any or all of the dimensions of a variable have the interpretations of "date or time" (T), "height or depth" (Z), "latitude" (Y), or "longitude" (X) then those dimensions should appear in the relative order T, Z, Y, X in the variable's definition.

Name	Comment
TIME	Number of time steps.
DEPH	Number of depth levels. Use 1 for HFR data.
LATITUDE	Dimension of the LATITUDE coordinate variable.
LONGITUDE	Dimension of the LONGITUDE coordinate variable.
HEAD	Dimension of the HEAD coordinate variable (bearing away)
RNGE	Dimension of the RNGE coordinate variable (range away).
STRING15	Length in characters of the strings used in the data file
MAXSITE	Maximum number of antennas. Set it as an upper bound.

Syntax

Double <DIM>(<DIM>);

<DIM>:standard_name

<DIM>:units [= "days since 1950-01-01T00:00:00Z"; only for TIME]

<DIM>:axis [= "T" for TIME; = "X" for LONGITUDE and RNGE; "Y" for LATITUDE and HEAD; = "Z" for DEPH]

<DIM>:calendar = "Julian"; [only for TIME]

<DIM>:positive = "down"; [only for DEPH]

<DIM>:long_name

<DIM>:ancillary_variables [= "TIME_SEADATANET_QC" for TIME; = "POSITION_ SEADATANET _QC"; for LATITUDE, LONGITUDE, HEAD and RNGE; = "DEPTH_SEADATANET_QC" FOR DEPH]

Data variables

When an appropriate **CF standard name** is available, it is required to be used; if not the **long_name** attribute has to be used. It is recommended that variable names be a 4-character-capitalized-letters name.

Radial and total velocity data	
EWCT	Surface Eastward Sea Water Velocity
NSCT	Surface Nortward Sea Water Velocity
NARX	Number of Receive Antennas
NATX	Number of Transmit Antennas
SLTR	Receive Antenna Latitudes
SLNR	Receive Antenna Longitudes
SLTT	Transmit Antenna Latitudes
SLNT	Transmit Antenna Longitudes
SCDR	Receive Antenna Codes
SCDT	Transmit Antenna Codes
Radial velocity data	
LATITUDE	Latitude
LONGITUDE	Longitude
RDVA	Radial Sea Water Velocity Away From Instrument
DRVA	Direction Of Radial Vector Away From Instrument
Total velocity data	
EWCS	Standard Deviation Of Surface Eastward Sea Water Velocity
NSCS	Standard Deviation Of Surface Northward Sea Water Velocity
CCOV	Covariance Of Surface Sea Water Velocity
GDOP	Geometrical Dilution Of Precision

Syntax

Float <PARAM>(TIME, DEPH, LATITUDE, LONGITUDE);

<PARAM>:*standard_name*

<PARAM>:units

<PARAM>:_FillValue

<PARAM>:coordinates

<PARAM>:long_name

<PARAM>:valid_range

<PARAM>:*comment*

<PARAM>:*add_offset*

<PARAM>:*scale_factor*

<PARAM>:ancillary_variables

Mandatory
Recommended

QC variables

Since in HFR data the quality control values vary along one or more axes of the data variables, they are provided as separate numeric flag variables, with at least one dimension matching the 'target' variable, and must carry the "flag_values" and "flag_meanings" attributes. QC variables can be linked to a target physical variable, or can be standalone variables reporting the results of specific QC tests.

Radial and total velocity data	
Variable name	long_name
TIME_SEADATANET_QC	Time SeaDataNet quality flag
POSITION_SEADATANET_QC	Position SeaDataNet quality flag
DEPTH_SEADATANET_QC	Depth SeaDataNet quality flag
QCflag	Overall Quality Flags
VART_QC	Variance Threshold Quality Flags
CSPD_QC	Velocity Threshold Quality Flags
Radial velocity data	
Variable name	long_name
OWTR_QC	Over-water Quality Flags
MDFL_QC	Median Filter Quality Flags
AVRB_QC	Average Radial Bearing Quality Flags
RDCT_QC	Radial Count Quality Flags
Total velocity data	
GDOP_QC	GDOP Threshold Quality Flags
DDNS_QC	Data Density Threshold Quality Flags

Syntax

Float <QCvar>(TIME, DEPH, LATITUDE, LONGITUDE);

<PARAM>:long_name

<PARAM>:units

<PARAM>:_FillValue

<PARAM>:valid_range

<PARAM>:flag_values

<PARAM>:flag_meanings

<PARAM>:*comment*

<PARAM>:*add_offset*

<PARAM>:*scale_factor*

Mandatory
Recommended

Data types and naming

The data type is a **bigram** used in **filenames** for a quick identification of the file content.

The naming convention for CMEMS-INSTAC data files requires to have the two bigrams '_XX_YY_' as part of the filename, where:

- 'XX' indicates the **type of measurement** (e.g. PR=profiles, TS=timeseries);
- 'YY' indicates the **data type**.

For HFR data the two bigrams 'XX' and 'YY' are defined as:

- **XX=TV for total velocity data files;**
- **XX=RV for radial velocity data files;**
- **YY=HF**

Thus, the two bigrams 'XX_YY' inside the filenames are:

- **'TV_HF' for total velocity data files;**
- **'RV_HF' for radial velocity data files.**

Quality Control

Radial velocity data
QC test (overall)
Syntax
Over-water
Velocity Threshold
Variance Threshold
Median Filter
Average Radial Bearing
Radial Count
Total velocity data
QC test (overall)
Syntax
Data Density Threshold
Velocity Threshold
Variance Threshold
Temporal Derivative
GDOP Threshold

Code	Meaning
0	unknown
1	good data
2	probably good data
3	potentially correctable bad data
4	bad data
5	-
6	-
7	nominal value
8	interpolated value
9	missing value

The overall QC variable will report the quality flags related to the results of all the QC tests: it is a "**good data**" flag if and **only if all QC tests are passed** by the data .

Processing Levels

LEVEL 0	Reconstructed, unprocessed instrument data at full resolution.	Signal received by the antenna before the processing stage.
LEVEL 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced and annotated with ancillary information.	Spectra by antenna channel
LEVEL 1B	Level 1A data that have been processed to sensor units for next processing steps.	Spectra by beam direction
LEVEL 2A	Derived geophysical variables at the same resolution and locations as the Level 1.	HFR radial velocity data
LEVEL 2B	Level 2A data that have been processed with a minimum set of QC.	HFR radial velocity data
LEVEL 2C	Level 2B data that have been reprocessed for advanced QC.	Reprocessed HFR radial velocity data
LEVEL 3A	Variables mapped on uniform space-time grid scales.	HFR total velocity data
LEVEL 3B	Level 3A data that have been processed with a minimum set of QC.	HFR total velocity data
LEVEL 3C	Level 3B data that have been reprocessed for advanced QC.	Reprocessed HFR total velocity data
LEVEL 4	Model output or results from analyses of lower level data.	Energy density maps, residence times, etc.

SDC CDI scheme

CDI FIELD	HFR model field
cdi-identifier	id
METADATA ORGANISATION	institution_edmo_code
METADATA CREATION-DATE	date_created
MEASURING AREA TYPE	feature_type
SPATIAL REPRESENTATION	grid_resolution (for total data), geospatial_vertical_resolution , time_coverage_resolution
COORDINATE DATUM	reference_system
NAME OF THE DATASET	title
DATASET-ID	id
DATASET REVISION-DATE	date_modified
IDENTIFIER	id
ORIGINATOR ORGANISATION	institution_edmo_code
ABSTRACT ON DATASET	summary
MANAGING ORGANISATION	institution_edmo_code
RESOURCE MAINTENANCE	update_interval
PARAMETERS	P02 keywords: RFVL, ACFL
INSTRUMENT	L05 code 303
PLATFORM	source, source_platform_category_code

CDI FIELD	HFR model field
PROJECTS	project (EDMERP codes)
Use Limitation	textual description
DATASET ACCESS	"LS" or "UN"
STATION NAME, CRUISE NAME	site_code, platform_code
EDMED REFERENCE	EDMED codes
SPATIAL RESOLUTION	grid_resolution
Dataset language	data_language
Characterset	"utf8"
Dataset main theme	"oceans"
GEOGRAPHICAL COVERAGE WEST	geospatial_lon_min
EAST	geospatial_lon_max
SOUTH	geospatial_lat_min
NORTH	geospatial_lat_max
START AND END DATE	time_coverage_start, time_coverage_end
DEPTH OF OBSERVATION	geospatial_vertical_min, geospatial_vertical_max,
VERTICAL DATUM	geospatial_vertical_units, vertical datum = sea level
DISTRIBUTING ORGANISATION	institution_edmo_code
Dataformat Version	"CF 4"
DISTRIBUTION SERVICE	THREDDS catalog links
Data Quality Information	processing_level

SDC CF extension

SDN namespace variables:

- char SDN_CRUISE --> site_code
- char SDN_STATION --> platform_code
- char SDN_LOCAL_CDI_ID --> id
- int SDN_EDMO_CODE --> institution_edmo_code
- char SDN_REFERENCE --> link to a single landing page
- char SDN_XLINKS --> array of text strings containing URIs pointing to a web resource such as a usage metadata document

SDN namespace variable attributes:

- :sdn_parameter_name --> from P01 vocabulary
- :sdn_parameter_urn--> URN from P01 vocabulary
- :sdn_uom_name --> from P06 vocabulary
- :sdn_uom_urn--> URN from P06 vocabulary
- :sdn_convention_urn only for QC ancillary variables

Other specific variable attributes

:Conventions for QC ancillary variables --> reference to the encoding convention used for the flag ('SeaDataNet measurand qualifier flags' in the case of SDC CF extension).

Different QC variable type and flagging scheme

- QC variable type is 'byte'.
- :flag_values = 48b, 49b, 50b, 51b, 52b, 53b, 54b, 55b, 56b, 57b, 65b ;
- :flag_meanings = "no_quality_control good_value probably_good_probably_bad_value bad_value changed_value value_below_detection value_in_excess interpolated_value missing_value value_phenomenon_uncertain" ;

INSPIRE compliance options

The INSPIRE directive uses the Lambert Azimuthal Equal Area (ETRS89-LAEA) coordinate reference system instead of WGS84. In order to become INSPIRE compliant, the addition of a projected grid is required, which allows the data to remain unchanged. This is currently under review; the inclusion of this is not mandated under the CF compliance.

This means that the following variables are required:

- INSPIRE_LONGITUDE --> conversion from WGS84
- INSPIRE_LATITUDE --> conversion from WGS84
- INSPIRE_crs --> verbatim for ETRS89-LAEA

Resources & Tools

Documentation

- JERICO-Next Deliverable D2.1
- JERICO-Next Deliverable D3.2
- JERICO-Next Deliverable D3.3
- JERICO-Next Deliverable D5.13
- JERICO-Next Deliverable D5.14 (upcoming)
- CMEMS SE INCREASE Deliverable D3.1

Software

- Tools for netCDF data v2.0 generation: https://github.com/LorenzoCorgnati/HFR_Combiner_TirLiq
- Tools for netCDF data v2.1 generation: https://github.com/LorenzoCorgnati/HFR_Combiner_TirLiq/tree/CMEMS-INSTAC
- Jradar tool for transformation of ASCII radial and total velocity files into netCDF data v1.0

For questions, information, collaboration please contact:
lorenzo.corgnati@sp.ismar.cnr.it