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, EMODnet Physics — CF, ACDD, INSPIRE; CMEMS SRD & PUM — **v2.0**
- 2018
  - CMEMS-INSTAC, SDC, EMODnet Physics — CF, ACDD, INSPIRE; CMEMS SRD & PUM; SDC CDI; SDC CF Extension — **v2.1**

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)	data_language	*geospatial_vertical_units*	**Provenance**
platform_code (EDIOS Platform ID)	data_character_set	*geospatial_vertical_positive*	**date_created**
data_mode	metadata_language	*time_coverage_resolution*	**history**
DoA_estimation_method	metadata_character_set	*time_coverage_duration*	**date_modified**
calibration_type	topic_category	*reference_system*	**date_update**
last_calibration_date	network	*grid_resolution*	**processing_level**
calibration_link	**Geo-spatial-temporal**	*cdm_data_type*	**contributor_name**
title	**data_type**	**Conventions used**	**contributor_role**
summary	**feature_type**	**format_version**	**contributor_email**
source	**geospatial_lat_min**	**Conventions**	
source_platform_category_code	**geospatial_lat_max**	*netcdf_version*	
institution	**geospatial_lon_min**	netcdf_format	
institution_edmo_code	**geospatial_lon_max**	**Publication information**	
data_assembly_center	**geospatial_vertical_min**	**update_interval**	
id	**geospatial_vertical_max**	**citation**	
project (EDMERP code)	**time_coverage_start**	**distribution_statement**	
naming_authority	**time_coverage_end**	**publisher_name**	
keywords	*geospatial_lat_units*	**publisher_email**	
keywords_vocabulary	*geospatial_lon_units*	**publisher_url**	
comment	*geospatial_vertical_resolution*	**license**	
		acknowledgment	

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.

Coordinate reference system variable crs is needed.

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.
BEAR	Dimension of the BEAR 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 BEAR; = "Z" for DEPH]

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

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

<DIM>:grid_mapping = "crs"; [for LATITUDE and LONGITUDE]

<DIM>:long_name

<DIM>:ancillary_variables [= "TIME_SEADATANET_QC" for TIME; = "POSITION_SEADATANET_QC"; for LATITUDE, LONGITUDE, BEAR 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 (if data on polar grid)
LONGITUDE	Longitude (if data on polar grid)
RDVA	Radial Sea Water Velocity Away From Instrument
DRVA	Direction Of Radial Vector Away From Instrument
HCSS	Radial Variance of Current Velocity
Total velocity data	
EWCS	Standard Deviation Of Surface Eastward Sea Water Velocity
NSCS	Standard Deviation Of Surface Northward Sea Water Velocity
UACC	Accuracy Of Eastward Surface Sea Water Velocity
VACC	Accuracy Of Northward 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
QWTR_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>:coordinates

<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.**

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 usage metadata document
SDN namespace variable attributes	
:sdn_parameter_name	parameter name from P01 vocabulary
:sdn_parameter_urn	parameter URN from P01 vocabulary
:sdn_uom_name	unit of measurement name from P01 vocabulary
:sdn_uom_urn	unit of measurement URN from P01 vocabulary

SDN namespace variable attributes MUST be present for each coordinate and data variable.

Different QC variable type and flagging scheme with respect to CMEMS-INSTAC requirements

- 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" ;

Different depth variable name with respect to CMEMS-INSTAC requirements

- DEPTH

Time aggregation strategy for distribution to SDC services

When the temporal aggregation is performed, the DEPH variable isrenamed to DEPTH and the flag values and meanings are converted via a mapping table. In that phase the :Conventions and :sdn_conventions_urn attributes are added to the QC variables.

SDC CDI scheme

CDI FIELD	HFR model field	CDI FIELD	HFR model field
cdi-identifier	id	**PROJECTS**	project (EDMERP codes)
METADATA ORGANISATION	institution_edmo_code	**Use Limitation**	textual description
METADATA CREATION-DATE	date_created	**DATASET ACCESS**	"LS" or "UN"
MEASURING AREA TYPE	feature_type	**STATION NAME, CRUISE NAME**	site_code, platform_code
SPATIAL REPRESENTATION	grid_resolution (for total data), geospatial_vertical_resolution , time_coverage_resolution	**EDMED REFERENCE**	EDMED codes
		SPATIAL RESOLUTION	grid_resolution
		Dataset language	data_language
COORDINATE DATUM	reference_system	**Characterset**	"utf8"
NAME OF THE DATASET	title	**Dataset main theme**	"oceans"
DATASET-ID	id	**GEOGRAPHICAL COVERAGE WEST**	geospatial_lon_min
DATASET REVISION-DATE	date_modified	**EAST**	geospatial_lon_max
IDENTIFIER	id	**SOUTH**	geospatial_lat_min
ORIGINATOR ORGANISATION	institution_edmo_code	**NORTH**	geospatial_lat_max
ABSTRACT ON DATASET	summary	**START AND END DATE**	time_coverage_start, time_coverage_end
MANAGING ORGANISATION	institution_edmo_code	**DEPTH OF OBSERVATION**	geospatial_vertical_min, geospatial_vertical_max, geospatial_vertical_units, vertical datum = sea level
RESOURCE MAINTENANCE	update_interval	**VERTICAL DATUM**	institution_edmo_code
PARAMETERS	P02 keywords: RFVL, ACFL	**DISTRIBUTING ORGANISATION**	
INSTRUMENT	L05 code 303	**Dataformat Version**	"CF 4"
PLATFORM	source, source_platform_category_code	**DISTRIBUTION SERVICE**	THREDDS catalog links
		Data Quality Information	processing_level

Quality Control

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

QC flag	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.

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 → OFFICIAL QC AND DATA MODEL MANUAL**
- CMEMS SE INCREASE Deliverable D3.1

Software tools

- Matlab tools for QC and conversion of radial and total data into netCDF data v2.1: https://github.com/LorenzoCorgnati/HFR_Node_tools
- Java tool for QC and conversion of radial and total data into files into netCDF data v2.1: https://github.com/ilasensio/JRadar

HFR are data available in EMODnet Physics:
http://www.emodnet-physics.eu/map/DefaultMap.aspx?sessionid=636583712630081155

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