CMEMS Service Evolution 21-SE-CALL1

Recommendation and guideline to set up the hardware and software tools to provide HFR data to the HFR Node

July26th, 2017

FOREWORD

This document is the first version of the deliverable D.4.1 for INCREASE WP4. WP4 main focus is the design, development and assessment of the pan-European data management infrastructure to host and make available HFR data and products developed for further use. This task will build on and extend the proof of concept developed within and in collaboration between the EuroGOOS HFR Task Team and the EMODnet Physics.

Index

1 Contributing Experts

A. Novellino (ETT, INCREASE TEAM)
M. Alba (ETT, INCREASE TEAM)
J. Mader (AZTI, INCREASE TEAM)
J.L. Asensio (AZTI, INCREASE TEAM)
A. Rubio (AZTI, INCREASE TEAM)
C. Mantovani (CNR-ISMAR, EuroGOOS HFR Task Team)
L. Corgnati (CNR-ISMAR, EuroGOOS HFR Task Team)
A. Griffa (CNR-ISMAR, EuroGOOS HFR Task Team)
P. Gorringe (EuroGOOS)
V. Fernández(EuroGOOS)

2 Introduction

The accurate monitoring of ocean surface transport, is key for the effective integrated management of coastal areas, where many human activities concentrate. This has been the main driver for the growth of coastal observatories along the global ocean coasts.

Among the different measuring systems, coastal High Frequency Radar (HFR) is the unique technology that offers the means to map ocean surface currents over wide areas (reaching distances from the coast of over 200km) with high spatial (a few kms or higher) and temporal resolution (hourly or higher). Consequently, the European HFR systems are playing an increasing role in the overall operational oceanography marine services. Their inclusion into CMEMS is crucial to ensure the improved management of several related key issues as Marine Safety, Marine Resources, Coastal & Marine Environment, Weather, Climate & Seasonal Forecast.

HFR data is in situ gridded data in time (big data) that has to be managed according its peculiarity, therefore the standard in situ data management infrastructures have to be empowered and updated to allow both Thematic Assembly Centres (TAC) and Marine Forecasting Centres *(*MFC) to assimilate and create new products including HFR data, so it will be necessary to design and develop the hardware and software infrastructures, as well as the data formats, file conventions, file dimension, file naming and labelling for both real time (i.e. the continuous data flow for latest days) and historical (i.e. the complete series) data.

The INCREASE WP4 is focusing on the recommendations and guidelines to set up the hardware and software tools to provide HFR data (derived totals surface currents) to the HFR Node. This Node should serve the CMEMS and should be designed as close as possible to the existing infrastructure for an easy and smooth uptake process.

2.1 Objectives of INCREASE project

INCREASE will set the necessary developments towards the integration of the existing European HFR operational systems into the CMEMS, following **four main objectives:**

- **(i)** **Provide HFR quality controlled real-time surface currents and key derived products**
- **(ii)** **Set the basis for the management of historical data and methodologies for advanced delayed mode quality-control techniques**
- **(iii)** **Boost the use of HFR data for improving CMEMS numerical modelling systems and**
- **(iv)** **Enable an HFR European operational node to ensure the link with operational CMEMS.**

To this end, the work in **INCREASE** will be aimed to enable a homogenised integration of the existing European HFR operational systems into the CMEMS, following five main work lines: i. Define and apply common data and metadata formats and quality control methodologies to ensure the integration of high quality HFR real time data into CMEMS (TAC, MFC); ii. Set the methodologies for reprocessing existing data sets to obtain continuous surface coastal ocean current data sets (QUID, assimilation in reanalysis products); iii. Develop key derived products (gap-filled data, short-term prediction and derived Lagrangian products) of added-value for CMEMS users (TAC); iv. Boost the use of HFR data for improving CMEMS numerical modelling systems (MFC); v. Enable an HFR European operational node to ensure the operational availability of HFR data and data products and the link with operational CMEMS (TAC, MFC)

This will be fulfilled through four technical work packages (WPs) and a WP devoted to the management of the project:

WP1: Towards the integration of HFR observing technology into CMEMS
WP2: Basis for HFR data assimilation into CMEMS models
WP3: HFR Products development
WP4: HFR Node
WP5: Management, the networking and communication activities.

2.2 Objectives of this report

This report is the main deliverable for INCREASE WP4.

WP4 is working on:

- T4.1 the definition of the structure of a minimum core data to be disseminated, such as raw and processed data;
- T4.2. Definition of a core package of metadata for the standard data;
- T4.3. Design of the data distribution and interoperability via the CMEMS ISTAC infrastructure and EMODnet Physics portal

The objectives of WP4 are:

- O4.1 design and develop the hardware and software infrastructure for HF Radar data management;
- O4.2 develop and disseminate a common method for data formats, file conventions, file dimension, file naming and labelling for both real time (i.e. the continuous data flow for latest days) and historical (i.e. the complete series) HF Radar data ;
- O4.3 set up THREDDS catalogues

3 General Organization

The Copernicus Marine Service, CMEMS, offers a wide range of oceanographic products: observational products (in situ and satellite) and numerical modelling products to respond to issues emerging in the environmental, business and scientific sectors. Within the Copernicus Marine Service, the In Situ Thematic Centre products are organised by region for the global ocean and the six European seas. However, for the users the quality of the product delivered must be equivalent wherever the data are processed.

In situ data in a given region are collected, quality controlled and distributed into products, which can be NRT (assessed using automated procedures) for real time activities or reprocessed (assessed by scientific teams) for reanalysis and research activities.

The INS TAC is implementing the following functions:

- *Data Acquisition*: Gather data available on international networks or though collaboration with regional and national partners.
- *Quality control*: apply automatic quality control tests that have been agreed at the In Situ TAC level. These procedures are defined by parameter, elaborated in coherence with international agreements.
- *Validation/Assessment*: Assess the consistency of the data over a period of time and an area to detect data that are not coherent with their neighbours but could not be detected by automatic QC.
- *Distribution*: make the data available within CMEMS INSTAC and to the external users.

Nevertheless, the establishment of the HFR data stream and data flow has to be organized in the coordinated framework formed by the exiting main European infrastructures and actors: Data Centres, CMEMS IN SITU TAC, EMODnet, SeaDataNet network of National Oceanographic Data Centres (NODCs).

According the CMEMS definitions the key actors and roles are:

- *National or Institutional Data centres*
  A national or institutional data centre is responsible for assembling data produced by a set of observing systems. The centre collects, controls and distributes data according to its own rules. Examples of these Data Centres are: Global Data Assembly Centres for international programs (Argo, GOSUD, OceanSITES, GTS…), EuroGOOS ROOSs data centres, SeaDataNet National Data Centres (NODC).
- *Production Units.*
  A Production Unit (PU) is responsible of assembling data provided by national or institutional data centres into an integrated dataset. The PU collects and controls data according to CMEMS INSTAC agreed rules and validates the dataset consistency in its area of responsibility.
- *Distribution Units.*

A Distribution Unit (DU) is responsible for assembling data provided by its region Production Units into an integrated dataset.

- *Global Distribution Unit*
  The Global Distribution Unit collects data distributed by Regional Distribution Units for the European seas complemented with Global PUs. The Global DU also acts as a backup for the Regional DUs.

Given the importance of the data type and the diversity with the already available data streams and quality check procedures, the implementation of the HFR data stream has to come together with the development of competence centres which role is to assess, validate, reprocess the HFR data stream in consolidated products. Adopting the CMEMS definitions, INCREASE actors and roles for the HFR data management are:

- *HFR Data Centres*
  An organization that is running a HFR site or is in charge for assembling data as recorded by the system. Given the peculiarity of the system the DC is in charge for the Quality control function, i.e. apply automatic quality controls that have been agreed at the HFR European experts' network (i.e. EuroGOOS HFR TT, JERICONEXT, etc).
- *HFR Production Unit*
  The HFR PU is responsible for assembling data (Acquire Data) as provided by the HFR DC, and for Validation/Assessment. The PU is then in charge for the harmonization of the file naming, data format, etc.
- *HFR (Global) Distribution Unit*
  The HFR DU is in charge for assembling data into an integrated dataset and uniform catalogue. Routinely (i.e. every hour for NRT) the DU distributes all its new data assembled by the Pus.

4 Data management infrastructure and organization

The DU distributes all its data and metadata products in NetCDF files on a THREDDS Data Server.

THREDDS catalogues and federative data repositories infrastructure have been developed and tested to make this high valuable data ready for the CMEMS processing core systems in an automatic (machine-to-machine) and supervised (monitoring tools) manner

4.1 TDS catalogue structure

The TDS is open source and runs inside the open source Tomcat Servlet container. The TDS provides catalog, metadata, and data access services for scientific data. Every TDS publishes THREDDS catalogs that advertise the datasets and services it makes available. THREDDS catalogs are XML documents that list datasets and the data access services available for the datasets. Catalogs may contain metadata to document details about the datasets. TDS configuration files provide the TDS with information about which datasets and data collections are available and what services are provided for the datasets.

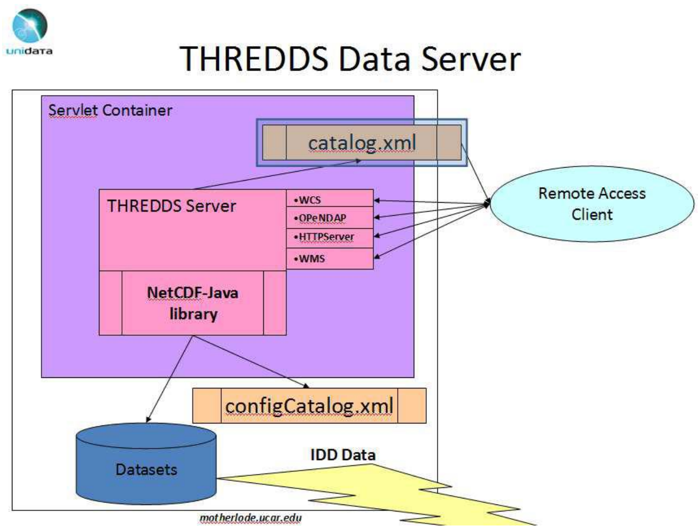

Figure 1. Schematic view of a TDS

The available remote data access protocols include OPeNDAP, OGC WCS, OGC WMS, and HTTP. It has to be noticed that the ncISO service allows THREDDS catalogs to be translated into ISO metadata records.

The CDM provides data access through the netCDF-Java API to a variety of data formats (e.g., netCDF, HDF, GRIB). Layered above the basic data access, the CDM uses the metadata contained in datasets to provide a higher-level interface to geoscience specific features of datasets, in particular, providing geolocation and data subsetting in coordinate space.

The TDS uses the CDM/netCDF-Java to read datasets in various formats. The CDM also provides the foundation for all the services made available through the TDS.

A pluggable framework allows other developers to add readers for their own specialized formats. The CDM also provides standard APIs for georeferencing coordinate systems, and specialized queries for scientific feature types like Grid, Point, and Radial datasets, and so it represents the best suitable available technology to manage HFR data products.

The TDS catalogue is described below:

Last day
- *one catalogue – one file per HFR containing daily observations*

Last 30 days
- *one catalogue – one file per HFR containing 30 days of observations (sliding window)*

Monthly
- *one catalogue per month – one file per HFR containing the observations for a given month*

History
- *one catalogue per year – one file per HFR containing the observations for a given year*

All data are accessible with the following services:
- OPENDAP (reference http://opendap.org/)
- WMS (reference http://www.opengeospatial.org/standards/wms)
- NetcdfSubset (reference http://www.unidata.ucar.edu/software/thredds/current/tds/reference/NetcdfSubsetServiceReference.html)

The Data format is described in Section 4.3.

4.2 Naming conventions

4.2.1 Last day catalogue

The catalogue contains the last day of HFR data (days of observation).
There is one catalogue per HFR site (two or more antennas), each catalogue contains one file containing the observation of the day (sliding window)

Proposed file naming convention:

- RR_LATEST_XX_CODE_YYYYMMDD.nc
  - RR: region bigram (see table 1)
  - LATEST: fixed name
  - XX_XX: data type (TL_HR for totals, RL_HF for radials)
  - CODE: platform code (see Subsection 4.2.1)
  - YYYYMMDD: year, month, day of observation (today)
  - .nc: NetCDF file name suffix

Example: IR_LATEST_TL__HR_BasqueHFR_20170530.nc , for a total file from the HFR radar system in the SE Bay of Biscay.

Code	Regional TAC
AR	Arctic
BO	Baltic Sea (BOOS)
BS	Black Sea
GL	Global Ocean
IR	Iberia Biscay Ireland (IBI ROOS)
MO	Mediterranean Sea (MONGOOS)
NO	North West Shelf (NOOS)

Table 1

4.2.2 Last 30 days directory catalogue

The catalogue "latest" contains the latest 30 days of HFR data (days of observation).
There is one catalogue per HFR site (two or more antennas), each catalogue contains one file containing 30 days of observation (sliding window)

Proposed file naming convention for the "latest" catalogue:

- RR_LATEST_XX_CODE.nc
  - RR: region bigram (see table 1)
  - LATEST: fixed name
  - XX_XX: data type (TL_HR for totals, RL_HF for radials)
  - CODE: platform code1
  - .nc: NetCDF file name suffix

Example: IR_LATEST_TL_HR_BasqueHFR.nc

4.2.3 Monthly catalogue

The catalogue "monthly" contains monthly data aggregation files.
It is regularly updated, at least once a month. It contains the last five years of data (per observation date). The older data are available in the history catalogue

Proposed file naming convention:

- RR_YYYYMM_XX_CODE.nc
  - RR: region bigram (see table 1)
  - YYYYMM : measurement date (YearMonth : 200901)
  - XX_XX: data type (TL_HR for totals, RL_HF for radials)
  - CODE: platform code
  - .nc : NetCDF file name suffix

Example: IR_201701_TL_HR_ BasqueHFR.nc
This file contains data for January 2017

4.2.4 History catalogue

The catalogue "history" contains yearly files of data aggregations. It is regularly updated but there is no fixed schedule for updates, at last once a year.

Proposed file naming convention:

- RR_YYYY_XX_CODE.nc
  - RR: region bigram (see table 1)
  - XX_XX: data type (TL_HR for totals, RL_HF for radials)
  - CODE: platform code
  - YYYY: year of observation

- .nc : NetCDF file name suffix

Example: IR_2016_TL_HR_BasqueHFR.nc
This file contains data for the year 2016

4.2.5 Platform code convention

There is a common effort between CMEMS INS TAC, EMODnet Physics, SeaDataNet to have platforms with a unique code. This code is mandatory and appears in the global attributes of the NetCDF files. The common agreement is that when available, the platform code should be the WMO platform code. When it is not available the code is the unique code assigned by the HFR DU.
The following naming convention is recommended:
_RR_XXXX

- RR: regional bigram (table 1)
- XXXX: platform name assigned by the DU where the suffix is HFR

e.g. IR_BasqueHFR

4.3 Data Format

The European common data and metadata model for real-time HFR data uses NetCDF-4 (Network Common Data Form). The implementation of NetCDF is based on the community-supported Climate and Forecast Metadata Convention (CF) version 1.6.

The definition of the European common data and metadata model for real-time HFR data follows the guidelines of the DATAMEQ working group and adds some requirements to CF 1.6 standard to make it easier to share in-situ data, to make it simpler for the Global Data Assembly Centers (GDACs) to aggregate data from multiple sites, and to ensure that the data can be created and understood by basic NetCDF utilities.

Section 5 presents a description of the data format as defined by the EuroGOOS European HF Task Team and under the WP5 of the JERICO-NEXT project.

The JERICO-NEXT data model for HFR is extended by adding EDMO code instead of free text for organisation information.

EDMO - European Directory of Marine Organizations - contains up-to-date addresses and activity profiles of research institutes, data holding centres, monitoring agencies, governmental and private organisations, that are engaged in oceanographic and marine research activities, data & information management and/or data acquisition activities. The EDMO was developed within the SeaDataNet projects and networks and currently it lists and describes more than 3.000 organisations.

The adoption of a controlled vocabulary guarantees better traceability and mapping of the data providers in the future.

5 European HFR Data format

During the INCREASE HFR Experts Workshop, a first discussion towards a definition of a standard concerning data formats and QC procedures was held. The main goal was to define a European Standard HFR Data format which met the needs of both the HFR community and CMEMS operational services. In particular, the main points analysed for achieving a common consensus were: data format, metadata structure, QC flagging scheme and QC tests. The European common data and metadata model for real-time HFR data was finally developed and defined in JERICO-NEXT D5.15 deliverable (the reader is referred to Corgnati et al., 2017[1] , available at http://www.jerico-ri.eu/download/jerico-next-deliverables/JERICO-NEXT-Deliverable-5.13_V1.pdf, for more details on the data and metadata model)

The main elements of this model are summarized as follows:

-Data Format: Data is produced using netCDF-4 classic model format, in order to apply the state of the art version. For the specific purposes of the INCREASE project, the netCDF-4 classic data will be then converted in netCDF-3.6.1 by the central HFR node to be developed by WP4. This double data production will meet both HFR community needs and CMEMS IN-SITU TAC needs.

-Metadata Structure: the CMEMS IN-SITU TAC reference conventions for the metadata attributes are CF-1.6 and OceanSITES. Thus, Mandatory Attributes have been defined and include attributes necessary to comply with CF-1.6 and OceanSITES conventions. For the data variables, the SDN P09 vocabulary is used.

-Standards on dimension, naming, definition and syntax of coordinate variables and data variables, including QC variables derived from the established QC tests, have been defined and implemented.

-A preliminary set of mandatory and recommended QC tests. These sets of QC tests are the required ones for labelling the data as Level 2B (for radial velocity) and Level 3B (for total velocity) data. Please refer to Table 8 of INCREASE D1.1 for the processing level definition.

- Reference QC flagging scheme to be used it the one from CMEMS IN-SITU TAC, i.e. the OceanSITES one. The CMEMS IN-SITU TAC strategy for quality flagging states that a “good data” flag is to be assigned to a data if and only if all QC tests are passed by the data. This strategy is adopted as a standard by the HFR community. Moreover, each file will include a gridded variable reporting QC flags for each velocity vector (for radial files and for total files). Finally, each file will include a QC variable for each performed QC test, in order to have an overall file quality flagging and all the specific flagging for the QC tests.

[1] Corgnati L., Mantovani, C., Novellino, A., Rubio, A., Mader, J., Reyes, E., Griffa, A., Asensio, J.L., Gorringe, P., Quentin, C., Breitbach, G. (2017). Recommendation Report 1 for HFR data implementation in European marine data infrastructures.

6 Conclusions

HFR is today recognized internationally as a cost-effective solution to provide mapping of ocean surface currents over wide areas with high spatial and temporal resolution that are needed for many applications for issues related to ocean surface drift or hydrodynamical characterization. Other R&D lines open interesting perspective for other variables like wave or surface wind data. The European HFR systems are playing an increasing role in the overall operational oceanography marine services (Rubio et al, 2017[2]). So, the basic and advanced products developed in INCREASE, based on the real-time 2D monitoring of shelf/slope surface circulation, will impact directly on key issues of CMEMS (Marine Safety, Marine Resources, Coastal & Marine Environment, Weather, Climate & Seasonal Forecast).
Direct products will be potentially implemented in In Situ TAC. Others will be used in Sea Level TAC for data intercomparison and integration, and in different MFCs for quality assessment (QUID, real-time indicators) or through data assimilation.

The performed inventory of the operational HFR systems in Europe (Mader et al, 2016[3]) includes 51 sites (in 20 networks) with potential impact in CMEMS. The MFC meshes overlap so one HFR station could impact in different MFC areas. The potential impact of the currently available data in the INCREASE catalogue is distributed as follow: MED-MFC (17 stations), IBI-MFC (15 stations), NWS-MFC (3 stations). These numbers will grow at a mid/long term scale, because countries like Portugal, France or UK are establishing plans for developing their networks.
Moreover, the covered areas allow a fundamental assessment in the buffer zone between CMEMS and downstream coastal tools. The products based on the real-time 2D monitoring of shelf/slope surface circulation will deliver key information for assessing the boundary conditions applied in the coastal models of intermediate users

The INCREASE project will provide demonstrators of key solutions for the implementation of new products based on HFRs and this deliverable is developing the "Recommendation and guideline to set up the hardware and software tools to provide HFR data to the HFR Node".

The CMEMS INSTAC is a decentralized architecture, the centralized HFR Node is proposed as a new component of this federative In Situ TAC infrastructure, with activities in all ROOSes where HFR systems are operating.

[2] Rubio A, Mader J, Corgnati L, Mantovani C, Griffa A, Novellino A, Quentin C, Wyatt L, Schulz-Stellenfleth J, Horstmann J, Lorente P, Zambianchi E, Hartnett M, Fernandes C, Zervakis V, Gorringe P, Melet A and Puillat I (2017) HF Radar Activity in European Coastal Seas: Next Steps toward a Pan-European HF Radar Network. Front. Mar.Sci.4:8. doi: 10.3389/fmars.2017.00008. http://journal.frontiersin.org/article/10.3389/fmars.2017.00008

[3] Mader J., Rubio A., Asensio J.L, Novellino A., Alba M., Corgnati L., Mantovani C., Griffa, A., Gorringe P., Fernandez V. (2016). The European HF Radar Inventory, EuroGOOS publications. http://eurogoos.eu/download/publications/EU_HFRadar_inventory.pdf

The development of this HFR Node goes towards three main steps:

- Set up a data server that is able to provide a comprehensive catalogue for European HFR radar.
- Create different adapters for the harmonization of data and metadata of HFR data coming from different sources
- Create a catalogue of HFR data that is compliant with CMEMS

The implementation of a such federative structure in situ thematic data assembly has to be coordinated at European level (Competence centre - HFR data node) and can be based on a hierarchical infrastructure to facilitate the management and integration of any potential data provider according a couple of simple and very effective rules:

- If the data centre can set up the data flow according the defined standards, the PU only has to link and include the new catalogue and data stream
- If the data centre cannot setup the data flow (because of lack of experience, technical capacity etc.), the PU has to work on harvesting the data from the provider, harmonize and format these data and make them available from the DU.

The integration and assessment of the HFR data in a centralized data system will allow:

- A second level quality check assessment
- Harmonized data products
- More efficient implementation of tools for downstream services

7 Useful Links

In the following the useful links to the reference conventions applied by the European common data and metadata model for real-time HFR data are listed.

- JERICONEXT Deliverable 5.13: http://www.jerico-ri.eu/download/jerico-next-deliverables/JERICO-NEXT-Deliverable-5.13_V1.pdf
- NetCDF Best Practices: unidata.ucar.edu/software/netcdf/docs/BestPractices.html
- NetCDF Climate and Forecast Metadata Convention 1.6, including the CF standard names: cfconventions.org
- Udunits package as implemented by CF: unidata.ucar.edu/software/udunits/
- ISO8601 description: http://en.wikipedia.org/wiki/ISO_8601
- Unidata NetCDF Attribute Convention for Dataset Discovery (ACDD): http://wiki.esipfed.org/index.php/Category:Attribute_Conventions_Dataset_Discovery https://podaac.jpl.nasa.gov/PO.DAAC_DataManagementPractices
- INSPIRE convention: http://inspire.ec.europa.eu/index.cfm/pageid/101 http://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf
- NODC NetCDF Templates: http://www.nodc.noaa.gov/data/formats/netcdf/
- Online CF-1.6 compliance checker: http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl?cfversion=1.6
- The SeaDataNet P09 controlled vocabulary: http://vocab.nerc.ac.uk/collection/P09/current/
- THREDDS: unidata.ucar.edu/projects/THREDDS/CDM/CDM-TDS.htm
- EPSG, used for the coordinate reference frames: http://www.epsg.org/
- QARTOD manual: https://ioos.noaa.gov/wp-content/uploads/2016/06/HFR_QARTOD_Manual_05_26_16.pdf
- OceanSITES convention manual: http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf