Report of Joint WP2-WP5 workshop:
ASSIMILATING TECHNICAL BEST PRACTICE IMPROVEMENTS TO OPTIMIZE NETWORK DATA FLOW

Date: 05 October 2017
Place: Bergen, Norway

Grant Agreement n° 654410
Project Acronym: JERICO-NEXT
Project Title: Joint European Research Infrastructure network for Coastal Observatory - Novel European eXpertise for coastal observaTories

Coordination: P. Farcy, IFREMER,
JERICO@ifremer.fr

Authors: Leonidas Perivoliotis, Rajesh Nair, Wilhelm Petersen
Involved Institutions: HCMR, OGS, HZG

Date: 08.01.2018

The JERICO-NEXT project is funded by the European Commission's
H2020 Framework Programme under grant agreement No. 654410
Project coordinator: Ifremer, France.

Table of contents

1. Document description

Document information	
Document Name	Report of Joint WP2-WP5 workshop: "Assimilating technical best practice Improvements to optimize network data flow"
Document ID	
Revision	1.0
Revision Date	08.01.2018
Author(s)	Leonidas Perivoliotis (HCMR), Rajesh Nair (OGS), Wilhelm Petersen (HZG)
Security	Public

History			
Revision	Date	Modification	Author
0.5	03.12.2017	First version	Leonidas Perivoliotis
0.6	06.12.2017	Corrections and comments	Wilhelm Petersen
0.7	17.12.2017	Corrections and comments	Rajesh Nair
1.0	08.01.2018	Final version	Leonidas Perivoliotis, Rajesh Nair, Wilhelm Petersen

Diffusion list				
Consortium beneficiaries	X			
Third parties				
Associated Partners				
other				

2. Attendees

Family name	Name	Institution	Country
Anil	Akpinar	Ifremer	France
Cardin	Vanessa	OGS	Italy
Carval	Thierry	Ifremer	France
Charria	Guillaume	Ifremer	France
Cuevas	Antonio	Udec	Chile
Delauney	Laurent	Ifremer	France
Durand	Dominique	Covartec	Norway
Fernadez	Vicente	EuroGOOS	Belgium
Gorringe	Patrick	EuroGOOS	Belgium
Medeot	Nevio	OGS	Italy
Nair	Rajesh	OGS	Italy
Novellino	Antonio	ETT	Italy
Perivoliotis	Leonidas	HCMR	Greece
Petersen	Wilhelm	HZG	Germany
Petihakis	George	HCMR	Greece
Pfeil	Benjamin	BCDC/UiB	Norway
Pouliquen	Sylvie	Ifremer	France
Puillat	Ingrid	Ifremer	France
Ravangan	Elisa	IRIS	Norway
Seppala	Jukka	SYKE	Finland
Sorensen	Kai	NIVA	Norway
Collingridge	Kate	CEFAS	UK
Mader	Julien	AZTI	Spain
Delrio	Joaquin	UPC	Spain
Nolan	Glenn	EUROGOOS	Belgium
Turpin	Victor	CNRS	France

3. Agenda

Thursday, October 05, 2017 (afternoon)

Joint WP2-WP5 Workshop: Assimilating technical Best Practice Improvements to optimize network data flow

From - to	Title
13:50-14:00	Overview of WP2 and WP5 of JERICO-NEXT (R. Nair, OGS/W. Petersen, HZG, L. Perivoliotis HCMR)
14:00-14:15	Biological observations using optics: the data manager’s perspective ***Veronique Creach, CEFAS***
14:15-14:30	Biological observations using optics: the data producer’s perspective ***Jukka Seppala, SYKE***
14:30-15:00	Discussion
15:00-15:15	HF radar observations: the data producer’s perspective. ***Julian Mader, AZTI***
15:15-15:30	HF radar observations: the data manager’s perspective ***Antonio Novellino, ETT***
15:30-16:00	Discussion
16:00-16:20	Coffee break
16:20-16:35	Marine carbonate system observations: the data producer’s perspective ***Kai Sorenson, NIVA***
16:35-16:50	Marine carbonate system observations: the data manager’s perspective ***Benjamin Pfeil, University of Bergen***
16:50-17:20	Discussion
17:20-17:35	Observations using AUVs: the data producer’s perspective ***John Allen, SOCIB (Rajesh Nair, OGS)***
17:35-17:50	Observations using AUVs: the data manager’s perspective ***Thierry Carval, Ifremer***
17:50-18:30	Discussion

4. Executive summary

In the JERICO-NEXT project, WP2 contains the initiatives planned to promote and facilitate the harmonization of the project's observing network from the technical standpoint principally through the promulgation of Best Practice as regards the technologies, methodologies and procedures underpinning actual measurement. On the other hand, WP5 contains the initiatives aimed at providing procedures and methodologies to enable data collected through the project to enter mainstream marine data conduits, complying with the international standards regarding their quality and metadata. The joint WP2-WP5 workshop was designed to examine the possibilities for closer collaboration between the above two WPs in order to better reconcile contrasts arising from differences in the way data are regarded by the project's observing and data management components.

During the workshop, the following kinds of data were targeted for attention: HF-radar data, data relating to biology based on optical measurements, data on the marine carbonate system, and data from AUVs (gliders). For each of the four data types, the perspectives of the data producer and the data manager were presented and discussed with a view towards proposing best practice strategies to mitigate current shortcomings in the way these data are being managed within the JERICO-NEXT network.

5. Statement of decisions

Decision	***WP***	***Content***	***Who***	***when***

6. Main report

Leonidas Perivoliotis from HCMR, the WP5 leader, welcomed the participants of the workshop on behalf of its organizing committee (specifically, himself and the WP2 co-leaders, R. Nair from OGS and W. Petersen from HZG), and opened the event's proceedings with a short presentation of its main aims and topics. In the JERICO-NEXT project, WP2 contains the initiatives planned to promote and facilitate the harmonization of the project's observing network from the technical standpoint while WP5 contains the initiatives aimed at providing procedures and methodologies to facilitate the streaming of gathered data to European data infrastructures. The joint WP2-WP5 workshop was planned to gather the members of the two WP communities in order to try to:

- identify difficulties relative to current data-handling practices employed within the project
- provide clearer terms of reference for handling JERICO-NEXT data
- agree on more appropriate metadata requirements for JERICO-NEXT data, if needed

The following kinds of data were targeted for attention: HF-radar data, data relating to biology based on optical measurements, data on the marine carbonate system, and data from AUVs (gliders). For each of the four data types, the perspectives of the data producer and the data manager were to be presented and discussed with a view towards proposing best practice strategies to mitigate current shortcomings in the way these data are being managed within the JERICO-NEXT network.

Biological observations analysing phytoplankton particles
Data producer's perspective:

The collection and dissemination of biology-related data based on optical measurements was the first topic that was tackled. **Veronique Creach** (CEFAS) presented the data manager's perspective for phytoplankton observations using cytometry. She began by describing the different techniques in use and the relevant instrumentation (e.g. Imaging FlowCytobot, CytoSense, FlowCam, FASTCAM, Underwater Vision Profiler), and then reviewed the main tools utilized for processing the relative observations at the present time: manual clustering with CytoClus (CytoBuoy, The Netherlands), the RclusTool package (LISIC/ULCO, France), and EasyClus software (Thomas Rutten Projects, Middelburg, The Netherlands). She highlighted that results obtained from the processing step are very user- and machine-dependent, that the underlying technology tended to change quite quickly, and that there is a strong need for setting guidelines for equipment and sample handling, the processing of raw data, and quality assurance. She added that the processing and analysis procedures needed to be reassessed in the context of new coastal areas being covered in JERICO-NEXT. She also announced that from the 19th to 21st of March 2018, a JERICO-NEXT phytoplankton workshop was being organized in Marseille where a number of relevant standardization issues were to be discussed.

Biological observations using optics
Data producer's perspective:

Jukka Seppala (SYKE) continued with the topic by presenting the methodologies used for observing phytoplankton employing LED fluorometry, spectral fluorescence, spectral reflectance and spectral absorption for quantifying different algae pigments (e.g. chlorophyll-a concentrations) as well as measuring turbidity by light scattering. Furthermore, he presented methods of measuring the photophysiology of the phytoplankton cells by fluorescence induction. In addition to the talk of Veronique Creach he shortly described the kind of data derived from flow-cytometry by pulse shape recording and imaging. He said that it is important to define the raw data for each sensor type and to archive these data, since there are often many different ways of extracting the biological information they represent. He added that there is a need to define best practices to ensure traceable primary calibrations of sensors used in making these kinds of measurements, and to find ways to adequately capture the relative details in the metadata information to improve data inter-comparability. On a closing note, the fact that it is often the manufacturers of instrumentation who establish data formats and processing methods for many such measurements was underlined, and a case was made for greater engagement with this community to enhance the comparability of measurements made with sensors of differing origins.

HF Radar observations

Data producer's perspective:

The second topic to be addressed was HF radar data. **Julien Mader** (AZTI) presented the data producer's view. He began by describing the underlying technologies and the basic steps that should be followed during the planning and installation phases of the equipment, together with the theory behind HF radar measurements. Then, he focused on known issues in the operation of HR radars, such as, for example, problems arising from environmental changes around installations that modify the electromagnetic field in the vicinity of the antennas, often invalidating antenna patterns and calibration parameters.

Data manager's perspective:

The presentation of the data manager's perspective on HF radar data was prepared by **Antonio Novellino** (ETT), but the actual talk was again given by **Julien Mader**. In the talk, the following key areas requiring action to achieve the necessary level of consensus to set up a roadmap for the creation of a European HF radar network were analyzed: data formats, metadata structures, and QC tests and flagging schemes for both the radial and the combined data. The concepts of "data production", "node" and "distribution unit", and their interconnectedness and final link to the major European Data Infrastructures was elaborated. It was suggested that the corresponding radial measurements should accompany the standard measurements coming from HF-radar systems, and that the wave-recording component should also start to be developed.

Marine Carbonate System

Data producer's perspective:

The next topic dealt with was marine carbonate system data. **Kai Sørensen** (NIVA) provided the data producer's perspective on these data. He divided marine carbonate data into two categories: the data gathered by sensors and those obtained from the analysis of discrete samples. Regarding the $p\mathrm{CO}_2$ and pH data generated by sensors, he talked about the basic principles of the measurements themselves, the calibration procedures, the continuing necessity of in-situ data for correcting and converting to final reported values. Other associated variables like total dissolved inorganic carbon and total alkalinity were also discussed. Furthermore, the fact that both automatic and manual measurements, and any computations involving them, needed corrections to account for the difference between in-situ and measurement temperatures was emphasized.

Data manager's perspective:

The data manager's perspective on marine carbonate system data was presented by **Benjamin Pfeil** (University of Bergen) who spoke about the ways these data are being handled at the European and international levels. He said that the data are currently being collected and managed through multiple initiatives that include data repositories, data brokers, integrated networks, research infrastructures and research products. The European Research Infrastructure, ICOS (Integrated Carbon Observation System), the GOOS (Global Ocean Observing System) Biogeochemistry panel, the IOCCP (International Ocean Carbon Coordination Project), EMODNET Chemistry, BiogeoChemical Argo and I3 relevant initiatives were specifically mentioned. The poor inclusion of coastal stations and FerryBox systems in many of the databases was explicitly acknowledged. The need for complementary efforts, more interaction between research infrastructures and networks, and greater interoperability between systems was also stressed. He mentioned that the requirements of quality in terms of precision and accuracy depend on the purpose of the data. To measure the impact of climate changes (e.g. ocean acidification) requires more precise carbon data than monitoring the carbon dynamic in highly biologically active coastal areas.

Glider Observations

Data producer’s perspective:

The last topic of the workshop concerned data coming from gliders. **John Allen** (SOCIB), who had prepared the presentation on the data producer’s perspective for this kind of data, was unable to come to Bergen due to last minute engagements, and his talk was given by **Rajesh Nair** (OGS). After a short introduction on the current capabilities of the technology itself, the presentation focused on the SOCIB toolbox for processing glider observations that generated data on three different quality levels. Then, the current procedures employed at SOCIB for correcting glider data based on field data from other platforms (mainly CTD casts), work that is being implemented in task 5.7 of WP5, was illustrated. It was shown that the corrected data files were accompanied by some new additions to the associated metadata containing information on the applied corrections, and suggestions for a number of other global attributes that could be further included were put forward. The standardization and semi-automation of depth-averaged velocity calculations, combining navigation data, flight models, and compass correction data files, was mentioned as one of the challenges of the moment.

Data manager’s perspective:

Thierry Carval (Ifremer) presented the data manager’s perspective on glider data, portraying the EGO (“Everyone’s Gliding Observatories”) data management system, a product of the consensus, networking and support of the many groups actively participating in that international initiative. He explained that the netcdf CF format had been selected by EGO contributors as the common data format to use for sharing glider observations with specific metadata information included, and that this choice was successively endorsed by Copernicus, SeaDataNet and AtlantOS. Furthermore, all data were being subjected to 14 quality control steps derived from Argo real-time QC protocols, while there were no standard procedures for delayed mode QC in place as yet. An EGO data processing tool (based on Matlab routines) and a netcdf file format checker were also being made freely available to the glider community.

7. Conclusions

The outcomes of the discussions that took place during the workshop are briefly summarized below.

- The maturity levels as concerns data collection and data processing are different for the four data types that were addressed at the workshop.
- The procedures for the HF Radar and glider data are well implemented and their connections to data infrastructures at the European level are well-established.
- The glider community is by far the most advanced: the parameter naming conventions and QC procedures endorsed by them are in use by other RIs and major research projects, and they also provide detailed guidelines and tools for uploading data to CMEMS and SDN.
- HF radar data are currently linked directly to EMODNET Physics, and there is a drive towards standardization of the relative QC procedures within the community.
- The collection and dissemination of biological data obtained from optical measurements is currently in the very early stages of the long process of standardization from all points of view. Much of the data collected (especially those based on imaging) are very user- and machine-dependent. This imposes serious problems in data comparability, especially between similar measurements collected in different regions by different institutions. There is also still a lot of work that needs to be done as regards appropriate metadata.
- The JERICO NEXT biological data integration plan is based on the data delivery to the EMODNET Biology infrastructure. The processed biological data that matches the EMODNET Biology's data scheme will be fully integrated in this data bank and will be discoverable and accessible though the standard EMODNET Biology tools. However, since a significant part of the JERICO NEXT flow-cytometry data cannot fit this existing scheme, a dedicated to the project data catalogue has been created (http://www.emodnet-biology.eu/data-catalog?module=dataset&show=search&spcolid=910), where detailed metadata information for each data set will be available together with direct links to raw and/or processed data.
- Marine carbonate system data are currently available through a variety of initiatives such as RIs, data infrastructures, integrated networks and research projects. However, the quality of the information on offer can vary, and there is a need for closer collaboration between the different initiatives to avoid duplication of efforts.

8. Annexes and references

JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

JOINT WP2-WP5 WORKSHOP

Assimilating Technical Best Practice Improvements to Optimize Network Data Flow

Rajesh Nair, OGS
Wilhelm Petersen, HZG
Leonidas Perivoliotis, HCMR

WP2 : Harmonization of technologies and methodologies - technical strategy

WP2 of JERICO-NEXT contains the initiatives planned to promote and facilitate the harmonization of the project's observing network from the technical standpoint, principally through the promulgation of Best Practice as regards the technologies, methodologies and procedures underpinning actual measurements.

WP5: Data Management

WP5 of JERICO-NEXT contains the initiatives aimed at providing procedures and methodologies to enable data collected through the project to enter mainstream marine data conduits, complying with the international standards regarding their quality and metadata.

This workshop has been planned to gather the members of the WP2 and WP5 communities of JERICO-NEXT to try to:

- identify difficulties relative to current data-handling practices employed within the project;
- provide clearer terms of reference for handling JERICO-NEXT data;
- agree on more appropriate metadata requirements for JERICO-NEXT data, if needed

Possible areas of cooperation that will be discussed in this workshop

- HF-radar data
- data relating to biology based on optical measurements
- data on marine carbonate system variables
- data from AUVs (gliders)

OUTLINE (the data producer's perspective)

- Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).
- Known requirements for proper deployment and common operating configurations for the described sensors/systems.
- Calibration and operational issues affecting data quality (e.g. limitations of calibration procedures, sampling modes, fouling, pressure effects, …).
- "Meaning" of data acquired (measurement interferences, influences of seasonality and other environmental effects, sensibility to maintenance practices, descriptors used to express the quality of measurements, …).
- Availability and sources of useful documentation in relation to the above topics.

OUTLINE (the data manager's perspective)

- Currently applied protocols (including metadata requirements) for handling data relating to the specific parameter under consideration within JERICO-NEXT.
- Known limitations of the described protocols, and their level of compatibility with other EU and global data management initiatives.
- Issues affecting data dissemination (e.g. scales, units and conversions, processed vs. unprocessed data, data reduction practices, further metadata needs …).
- Proposals/suggestions for improving the JERICO-NEXT terms of reference for handling data relating to the specific parameter under consideration.
- Suggestions for better and more efficient connection with major European Data infrastructures

---

Agenda

From - to	Title
13:50-14:00	Overview of WP2 and WP5 of JERICO-NEXT (R. Nair, OGS/W. Petersen, HZG, L. Perivoliotis HCMR)
14:00-14:15	Biological observations using optics: the data manager's perspective **Veronique Creach, CEFAS**
14:15-14:30	Biological observations using optics: the data producer's perspective **Jukka Seppala, SYKE**
14:30-15:00	Discussion
15:00-15:15	HF radar observations: the data producer's perspective. **Julian Mader, AZTI**
15:15-15:30	HF radar observations: the data manager's perspective **Antonio Novellino, ETT**
15:30-16:00	Discussion
16:00-16:20	Coffee break
16:20-16:35	Marine carbonate system observations: the data producer's perspective **Kai Sorenson, NIVA**
16:35-16:50	Marine carbonate system observations: the data manager's perspective **Benjamin Pfeil, University of Bergen**
16:50-17:20	Discussion
17:20-17:35	Observations using AUVs: the data producer's perspective **John Allen, SOCIB (Rajesh Nair, OGS)**
17:35-17:50	Observations using AUVs: the data manager's perspective **Thierry Carval, Ifremer**
17:50-18:30	Discussion

JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

Biological observations: the data producer's perspectives for phytoplankton

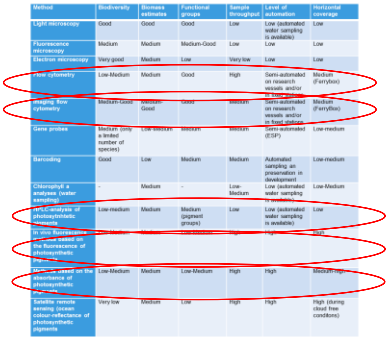
Véronique Créach
Centre for Environment, Fisheries and Aquaculture Science, NR33 0HT Lowestoft, UK

Email: veronique.creach@cefas.co.uk

ASSIMILATING TECHNICAL BEST PRACTICE IMPROVEMENTS TO OPTIMIZE NETWORK DATA FLOW,
Bergen, 5 of October 2018

www.jerico-ri.eu

Biological observations: techniques



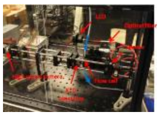
Method	Biodiversity	Biomass estimates	Functional groups	Sample throughput	Level of automation	Horizontal coverage
Light microscopy	Good	Good	Good	Low	Low (automated water sampling is available)	Low
Fluorescence microscopy	Medium	Medium	Medium-Good	Low	Low	Low
Electron microscopy	Very good	Medium	Low	Very low	Low	Low
Flow cytometry	Low-Medium	Medium	Good	High	Semi-automated on research vessels and/or in fixed stations	Medium (Ferrybox)
Imaging flow cytometry	Medium-Good	Medium-Good	Good	Medium	Semi-automated on research vessels and/or in fixed stations	Medium (FerryBox)
Gene probes	Medium (only a limited number of species)	Low-Medium	Medium	Medium	Semi-automated (ESP)	Low-medium
Barcoding	Good	Low	Medium	Medium	Automated sampling an preservation in development	Low-medium
Chlorophyll a analyses (water sampling)	-	Medium	-	Low-Medium	Low (automated water sampling is available)	Low-Medium
HPLC-analysis of photosynthetic pigments	Low-medium	Medium	Medium (pigment groups)	Low	Low (automated water sampling is available)	Low
In vivo fluorescence methods based on the fluorescence of photosynthetic pigments	Low-Medium	Medium	Low-Medium	High	High	High
Methods based on the absorbance of photosynthetic pigments	Low-Medium	Medium	Low-Medium	High	High	Medium-high
Satellite remote sensing (ocean colour-reflectance of photosynthetic pigments	Very low	Medium	Low	High	High	High (during cloud free conditions)

Biological observations: Instruments

cytobot

FlowCam

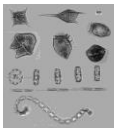
FastCam



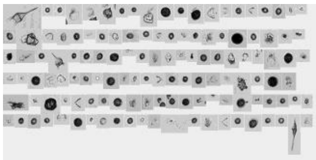
Cytosense

Underwater
Vision Profiler
UVP5

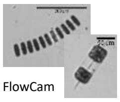
Biological observations: image



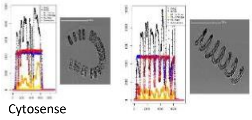
Cytobot



FastCam

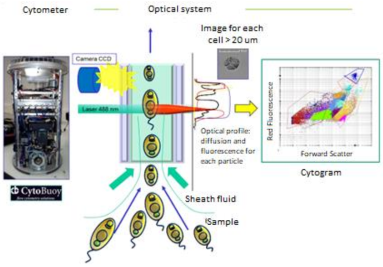


FlowCam



Cytosense

Biological observations: shape-pulse flow cytometry



Biological observations: tool for processing the data

Manual clustering with CytoClus© (CytoBuoy, The Netherlands):

Long, need knowledge of the community in the area of interest and experience

RclusTool package (LISIC/ULCO, France)

The RClusTool is a toolbox based on machine learning, the tool designs automatically clusters of the phytoplankton functional types, with eventually the possibility of correcting the results.

Fast but not yet optimised and need more intercomparison with manual clustering

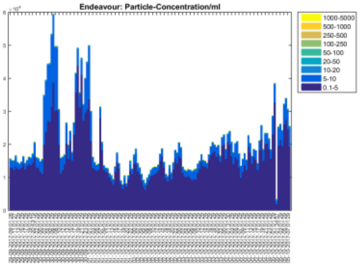
EasyClus© software (Thomas Rutten Projects, Middelburg, The Netherlands)

The EasyClus© software proposes many tools to organize, cluster and handle flow cytometric data (of many types of instruments) and uses the Matlab® environment.

Biological observations

EasyClus© software (Thomas Rutten Projects, Middelburg, The Netherlands):

A live version



Biological observations: Standardisation

Best practices:
Intercalibration workshop (Gothenburg, September 2016)
Sharing cruises :

"Vessel"/Platform Cruise/Coordinator	Year	CTD	PSFCM (C)	PSFCM (D)	Imaging system	Ferry Box	Thermo-salinometer
"Le Carthage" CNRS-MIO	2016		X		CytoSense	X	
"Simon stevin" VLIZ		X	X		2 CytoSenses	X	
\\		X	X	X	CytoSense	X	
Tangesund Observatory		X		X	IFCB + FastCam		
"Le Carthage"			X			X	
"Endeavour" Nephros TV Survey Cefas		X	X		CytoSense	X	
"Endeavour" Pelagic Fish Cefas		X	X		CytoSense	X	
"Endeavour" Cefas	2017	X	X			X	
"Côtes de la Manche" PHYCO CNRS-LOG		X	X	X	CytoSense	X	
"Simon Stevin" VLIZ		X	X	X	3 CytoSenses	X	
"Zirfaea" RWS		X	X	X	2 CytoSenses	X	
"Aranda" SYKE/SMHI		X	X	X	CytoSense + FlowCam + UVP5	X	
"Sepia II" PELRAD CNRS-LOG		X	X		CytoSense		X
"Côtes de la Manche" ETOILE IFREMER/CNRS		X	X	X	CytoSense		X

Biological observations

- The technology changes very quickly and new coastal areas have been added during JERICO-NEXT, analytical procedures need to be reassessed.
- the flow cytometer analysis is user and machine dependant
- setting a control quality procedures for equipment, sample and result analysis under guidelines

Equipment:

- ☐ some parameters related to instrument and the analysis are crucial to report such as the specifications and configuration of the instrument
- ☐ Each FCM needs to be calibrated during each analysis day with beads (fluorescence, size) and algae (from time to time)
- ☐ Performance indicators should be listed in an instrument dependent specification can could be provided by the manufacturer during the maintenance operation

Biological observations: listed in the metadatabase

analysis :

- ☐ follow a procedure according to the specifications of the machine and phytoplankton community
- ☐ mandatory paramaters: volume, trigger level, time of analysis

results :

☐ Mandatory information:

- Total number of phytoplankton particles per ml
- Contribution of the phytoplankton particles to the total particles (%)
- Total number of particles by functional types: picoeukaryotes and Synechococcus, nano- and microphytoplankton per ml per sample
- recognized microalgae (pictures)
- Contribution relative of the main category to total red fluorescence (%)

☐ Optional information:

- Total red fluorescence standardised to total chlorophyll *a* for each sample
- Median size of the phytoplanktonic community
- Number of sub-groups in each main 4 categories and number of phytoplankton particles in the sub-groups)

Biological observation

Comparison of results between experts (manual clustering)

	CNRS LOG / ULCO		CNRS MIO		CEFAS	
	phyto counts	%contrib flr	phyto counts	%contrib flr	phyto counts	%contrib flr
Smp2	43298 (94.4%)		41278 (90.0%)		41484 (90.75%)	
- Synechococcus	1773	0,1	945	0.1	1728	0,1
- Picoeukaryotes	2361	0,5	1332	0.4	832	0,3
- Nanoeukaryotes	38842	91,5	37332	45.1	37718	89,6
- Microeukaryotes	322	8,0	1669	54.4	1206	10,0
Smp3	15357 (89.22%)		12557 (72.9%)		13621 (79.14%)	
- Synechococcus	3964	0,2	2478	0.1	3899	0,2
- Picoeukaryotes	4622	0,7	3939	0.2	4038	1,6
- Nanoeukaryotes	5506	40,6	4219	23.1	3933	39,5
- Microeukaryotes	1265	58,5	1921	76.6	1751	58,7
Smp4	6073 (34.09%)		6086 (34.2%)		5589 (23.88%)	
- Synechococcus	3030	0,3	2303	0.1	2805	0,3
- Picoeukaryotes	1344	0,6	1018	0.1	915	0,3
- Nanoeukaryotes	1270	7,6	2326	7.0	1365	5,9
- Microeukaryotes	429	91,4	439	92.8	504	93,4

From 0.66 to 0.98

Biological observations: next

☐ Comparison the manual clustering with automatic clustering

↓

Redefine the phytoplankton functional types to report in the database

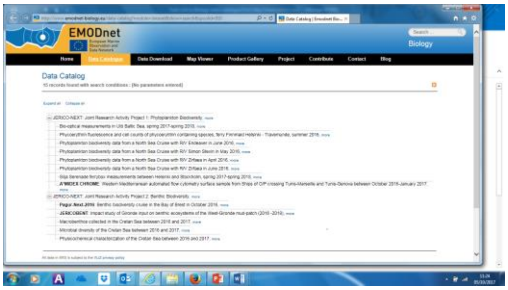
☐ Design a template for reporting mandatory parameters

☐ Create a network of experts for specific areas to check analyses and clustering

↓

Exchange of files and samples (*Chlorophyll*)

Biological observations: EMODnet Biology



EMODnet

Biology

Home · Data Catalogue · Data Download · Map Viewer · Product Gallery · Project · Contribute · Contact · Blog

Data Catalog

15 records found with search conditions : (No parameters entered)

Expand all · Collapse all

- JERICO-NEXT: Joint Research Activity Project 1: Phytoplankton Biodiversity. more
  - Bio-optical measurements in Utö Baltic Sea, spring 2017-spring 2018. more
  - Phycoerythrin fluorescence and cell counts of phycoerythrin containing species, ferry Finnmaid Helsinki - Travemünde, summer 2016. more
  - Phytoplankton biodiversity data from a North Sea Cruise with R/V Endeavour in June 2016. more
  - Phytoplankton biodiversity data from a North Sea Cruise with R/V Simon Stevin in May 2016. more
  - Phytoplankton biodiversity data from a North Sea Cruise with R/V Zirfaea in April 2016. more
  - Phytoplankton biodiversity data from a North Sea Cruise with R/V Zirfaea in June 2016. more
  - Silja Serenade ferrybox measurements between Helsinki and Stockholm, spring 2017-spring 2018. more
  - **A'MIDEX CHROME**: Western Mediterranean automated flow cytometry surface sample from Ships of O/P crossing Tunis-Marseille and Tunis-Genova between October 2016-January 2017. more
- JERICO-NEXT: Joint Research Activity Project 2: Benthic Biodiversity. more
  - **Pagur Next 2016**: Benthic biodiversity cruise in the Bay of Brest in October 2016. more
  - **JERICOBENT**: Impact study of Gironde input on benthic ecosystems of the West-Gironde mud-patch (2016-2019). more
  - Macrobenthos collected in the Cretan Sea between 2016 and 2017. more
  - Microbial diversity of the Cretan Sea between 2016 and 2017. more
  - Physicochemical characterization of the Cretan Sea between 2016 and 2017. more

---

Biological observations

19th-21st March 2018: JERICO-NEXT phytoplankton workshop in Marseille

22nd-23th March 2018:

Improving the visibility of ocean data from new technologies: a case study of high frequency flow cytometry.
(*Euromarine workshop if funded*)

Session 1: How to harmonise flow cytometry data: from individual scientist to pan-European research network.
Session 2: Use of flow cytometry information by users from different fields
Session 3: Integrating of new types of phytoplankton data in Europe's ocean observing infrastructure
Session 4: practical workshop on clustering and identification of phytoplankton functional types.

Biological observations

Thank you

ASSIMILATING TECHNICAL BEST PRACTICE IMPROVEMENTS TO OPTIMIZE NETWORK DATA FLOW,
Bergen, 5 of October 2018

JERICO-NEXT

JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

HORIZON 2020

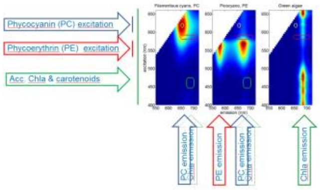
Biological observations using optics: the data producer's perspective

Jukka Seppälä
Finnish Environment Institute, SYKE
Jukka.seppala@ymparisto.fi

A lot of information/figures reused from JERICO-NEXT D2.2 & D3.1, thanks to those who contributed!!

LED Fluorometry

- By selecting appropriate LEDs and filters, excitation and emission wavebands may be matched with the fluorescence properties of different compounds; like Chlorophyll a , phycoerythrin, phycocyanin, CDOM.
- Small size sensors (100 g->, diam. 3-10 cm), relatively low price (1000-5000€), suitable for all oceanographic platforms.
- Well established technology, TRL 8-9, but lack of harmonization.
- Instruments from different manufacturers have different optical setups & different calibration practices by users -> challenging the consistency of data.



LED Fluorometry

- Key issues related to the long or short term stability of the instruments, due to biofouling, condensation of water inside or deteriorating of the optics. As the LED fluorimeters are single channel instruments, resolving blanks, biofouling, drift or other interferences requires discrete sampling, additional measures or good knowledge of the system.
- Fluorometers are most often providing accurate description of fluorescence intensity, but the interpretation of this signal as concentration of pigments or cell numbers is not straightforward due to photobiological processes in living cells affecting the fluorescence yield.

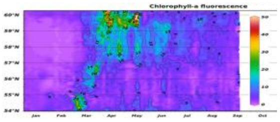
$F = [Chla] \cdot R$

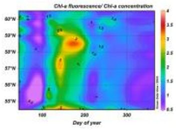
R varies 2-4 fold for single species,
and up to 50-fold between different species.

$$F\,(\lambda_{ex/em}) = \underbrace{\boxed{[Chla]}}_{\text{Biomass}} \cdot \underbrace{\boxed{E_{ex}}}_{\text{Instrument}} \cdot \underbrace{\boxed{\bar{a}_{PSII}*}}_{\text{Species}} \cdot \underbrace{\boxed{Q_a*(\lambda_{em})}}_{\text{Pigmentation}} \cdot \underbrace{\boxed{\phi_F}}_{\text{Physiology}}$$

LED Fluorometry

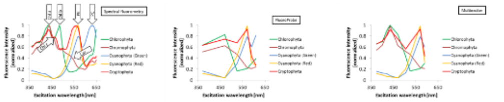
- Primary calibration of LED fluorimeters has not been agreed and the results are most often in relative units (or in µg/L but without any traceability of calibration)
- Fluorescence typically considered as a semi-quantitative proxy of concentrations
- Validation with field samples is an important step in analysing fluorescence data
- Most common validation method is linear regression, but this tend to fail, e.g. when the changes in the phytoplankton physiology are "larger" than changes in biomass (example day-night shifts in non-photochemical quenching).
- Alternative methods for validation are (and will be more) available, but no guidelines/decision-tree what to use (user/event/location specific validation).





Spectral fluorescence

- Phytoplankton fluorescence excitation spectra shows the spectral shape of the light absorption by accessory pigments.
- Consistent differences between spectra of different taxonomic pigment groups: green algae (Chlorophyta); brown algae (e.g. Dinophyta and Bacillariophyta) cryptomonads (Cryptophyta), Cyanobacteria (Cyanophyta)
- Two major brands Multiexciter (JFE Advantech Co, Ltd, Japan), FluoroProbe & AlgaeOnlineAnalyzer (bbe Moldaenke GmbH, Germany)
- Weight 1.6 – 6.4 kg, price 20k€-> . May be used in "Logging" or "Online" modes; profiling & flow-through.
- High TLR (8-9), but the agreed traceable calibration is lacking decreasing the value of instruments



Spectral fluorescence

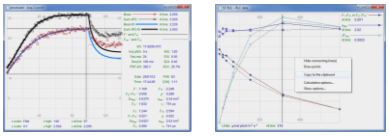
- Primary data is (uncalibrated) fluorescence intensity at different wavebands.
- Secondary instrument output is Chlorophyll a concentration in different taxonomic groups, based on 1) linear unmixing and 2) laboratory measured norm-spectra (fingerprints).
- Several reasons the method above is biased: selection of correct spectral groups, no co-varying spectral groups allowed, diversity of spectral properties within taxonomic group, variability in fluorescence quantum yield. -> Chlorophyll a concentrations largely biased by selection of fingerprints.
- Other data analysis methods available, but they are rather unstandardized

Fluorescence induction

- Fluorescence induction measurements result in several parameters describing the state of photophysiology of the phytoplankton cells
- Two main techniques in measuring variable fluorescence: single turnover technique like Fast Repetition Rate Fluorometry (FRRF) and multiple turnover technique like Pulse Amplitude Modulation (PAM) fluorometry.
- In Jerico-Next: FastOcean (Chelsea Technologies Group Ltd, UK) FRRF sensor with accessories can be used as profiler, as bench top model and in flow-through system & PhytoPAM (Heinz Walz GmbH, Germany) PAM sensor is bench top model.
- Sensors are relatively large and expensive (20k€->), limiting their use.
- Basic technology mature, accessories not necessarily well tested

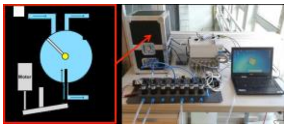
Fluorescence induction

- Complex data flow when most advanced systems used
  - Fluorescence induction curves - µs scale, some hundreds of raw data points per curve + model output
  - Fluorescence induction curves carried out in different light steps
  - Modelled data summarizing the light curves
- Meaning of all parameters not well described (e.g. additional wavebands)
- Between model (FRRF vs. PAM) comparison difficult.
- Calibration protocol established (may need revisit?)
- Fluorescence induction methods provide estimates on electron transport rate, and with some assumptions this may be converted to rates of oxygen evolution or carbon fixation, but still there are large uncertainties in this conversion



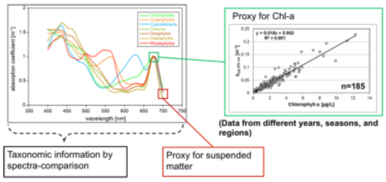
Spectral absorption

- HyAbs (Nexos development) & Oscar (Trios Gmbh, Germany): integrating cavity, which allows sensitive measurements due to a long optical path length and eliminates errors introduced by light scattering by particles.
- Light transmission difference between the sample (seawater) and the reference (purified water) can be measured and used together with reflectivity, temperature, salinity and calibration factors to calculate the spectal absorption coefficient of the water constituents in units [m^{-1}].
- Absorption coefficient may be decomposed mathematically into different components (phytoplankton, inorganic particles, organic particles, CDOM)
- TRL level 5-6, price 20k€+, for flow-through systems / profiling.
- Needs extensive cleaning procedure (which need to be automated for commercial sensor)



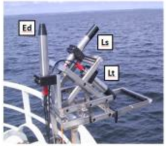
Spectral absorption

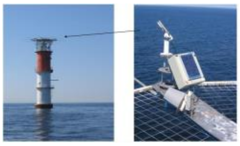
- In principle, data in physical units, comparable to other techniques.
- For reliable measurements, the reflectivity of the integrating cavity has to be determined by a calibration measurement (OSCAR - Nigrosin, HyAbs – solid standard)
- Primary data absorption coefficient spectra in the range of the visible light (400-710 nm)
- 1st Secondary data: absorption spectra of different in-water constituents
- 2nd Secondary data: [Chla], [TSM], [DOC], algal pigment classes



Spectral reflectance

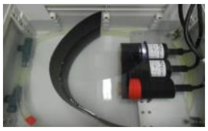
- Spectral radiance and irradiance sensors used to estimate light reflectance above water , e.g. to validate satellite products.
  - irradiance sensor towards zenith providing the total downwelling light Ed
  - radiance sensor providing upwelling light from the sea and sea surface, Lt
  - radiance sensor providing sky contributions from the upwelling component, Ls
- Established technology, measurements in physical units (µmolq m^{-2} s^{-1} nm^{-1}) , may be used in stationary or moving platforms
- Traceable calibration of sensors available
- Challenges to achieve ideal conditions, such as weather and sea conditions, sensor angles relative to sun, and selection of the Fresnel Coefficient, and subsequent data flagging.





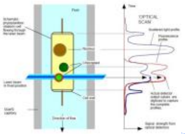
Turbidity and scattering

- For turbidity, measure of light scattering, the standard is a Nephelometric laboratory method based on a 90° (+/-2.5°) scattering at 860 nm (+/- 10-15 nm) wavelength detection.
- Turbidity is a proxy for the total suspended material .
- Other techniques: 1) total beam attenuation (*c*) at different wavelength and resolving total scattering (*b*) and absorption (*a*) at 9 wavelengths and 2) backscattering coefficient (*bb*) at different wavelengths with fixed angles.
- Calibration traceable but interpretation of data sometimes challenging due to different measuring geometries, wavelengths etc. data from different instruments not directly comparable
- Metadata information of the sensor specifications of e.g. wavelength and scattering angles are important to report since this influence the optical signal (turbidity) from different particles types



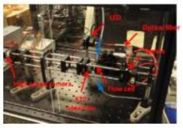
Pulse shape-recording flow cytometry

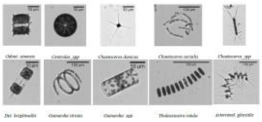
- Three interoperating systems: Optical, Fluidics and Electronics.
- PSR FCM from the Cytobuoy® : automatically recording the optical pulse shape of every particle passing through a laser beam.
- Benchtop, continuous (e.g. ferrybox) and submersible (e.g. buoy)
- Use of calibration beads for laser alignment, size calibration and fluorescence calibration.
- Different setups (lasers, power) -> intercomparison of machines required
- Raw data need to be retained. One issue with the data reliability is the significance of abundance estimation for rare species/particles
- Manual/automated/semi clustering & classification methods -> affect interpretation of data (user dependent)



Imaging flow cytometry

- Imaging FlowCytoBot (McLane Research Laboratories), FlowCAM (Fluid Imaging Technologies), FastCAM prototype (IFREMER - LDCM), Underwater Vision Profiler UVP5 (Hydroptic)
- Images are the main product for all instruments but due to instrument–specific differences in optical and fluidic characteristics, various results in terms of image resolution (magnification/size of particles analysed) and measurements (features). Specific training sets are built for each instrument.
- Standardization of analytical and data processing as well as data management need more development
- Classification based on training sets involving taxonomic expertise.
- Importance of keeping raw data, images





For further discussion

- Defining raw data for each sensor type
  - Storing raw data is very important especially in the cases when there exist many different ways of calculating biological information
- Defining best practices for traceable primary calibration and it's inclusion in metedata
  - raw data is relatively useless if there is no traceability of data!! Without traceability raw data from different sources/instruments cannot be compared!
- Often the "data" issue is driven by manufacturers with a major market within non-scientific monitoring activity.
  - As scientific community we call for traceability and correct measuring practices, while bulk selling point is to provide instruments which replace lab-based measurements; i.e. why should fluorometer manufacturer sell an instrument measuring at relative scale to please us, while majority of customers like the (biased) µg/L scale.
  - -> collaboration with manufacturers
  - -> providing demonstrations of the added value of traceability & best practices i.e. showing examples of consistent and comparable multisensor data-sets

JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

HF radar observations:
the data producer's perspective.

Julien Mader, AZTI, jmader@azti.es
Carlo Mantovini, CNR, Italy
Jochen Horstmann, HZG, Germany

OUTLINE (the data producer's perspective)

- Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).
- Known requirements for proper deployment and common operating configurations for the described sensors/systems.
- Calibration and operational issues affecting data quality (e.g. limitations of calibration procedures, sampling modes, fouling, pressure effects, …).
- "Meaning" of data acquired (measurement interferences, influences of seasonality and other environmental effects, sensibility to maintenance practices, descriptors used to express the quality of measurements, …).
- Availability and sources of useful documentation in relation to the above topics.

OUTLINE (the data producer's perspective)

- Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Hardware overview

- Land based remote sensing instrument
- HF = High Frequency (from 3 to 30 MHz)
- One receiving and one transmitting station (plus electronics)
- different antennas configurations (depending on frequency and signal processing technique)

OUTLINE (the data producer's perspective)

- Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Hardware overview

From University of Hamburg

4-element square array receiver

OUTLINE (the data producer's perspective)

- Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Hardware overview

From University of Hamburg

16-element linear array receiver

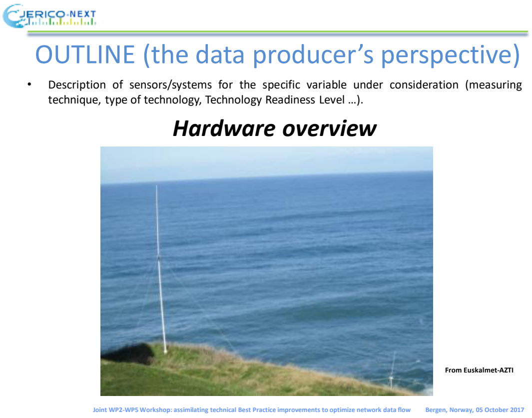
OUTLINE (the data producer's perspective)

- Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Hardware overview

From CNR-ISMAR

Compact transmitting and 3-element receiving antenna

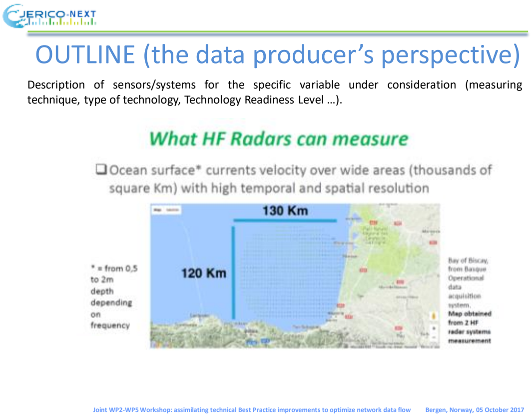


OUTLINE (the data producer's perspective)

- Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Hardware overview

From Euskalmet-AZTI



OUTLINE (the data producer's perspective)

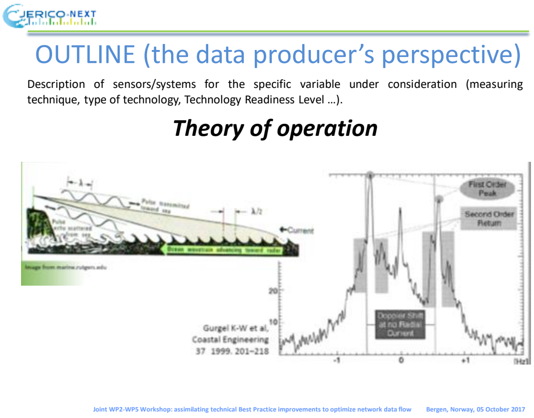
Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

What HF Radars can measure

☐ Ocean surface* currents velocity over wide areas (thousands of square Km) with high temporal and spatial resolution

* = from 0,5 to 2m depth depending on frequency

Bay of Biscay, from Basque Operational data acquisition system. **Map obtained from 2 HF radar systems measurement**



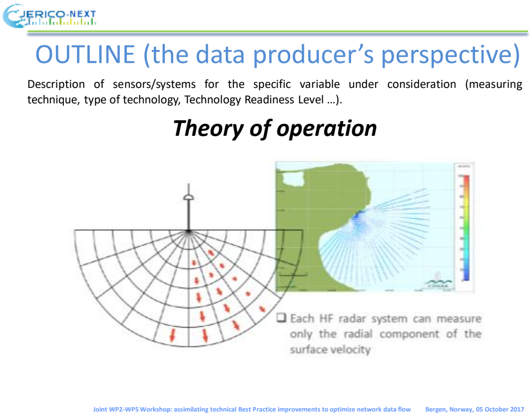
OUTLINE (the data producer's perspective)

Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Theory of operation

Image from marine.rutgers.edu

Gurgel K-W et al, Coastal Engineering 37 1999. 201–218

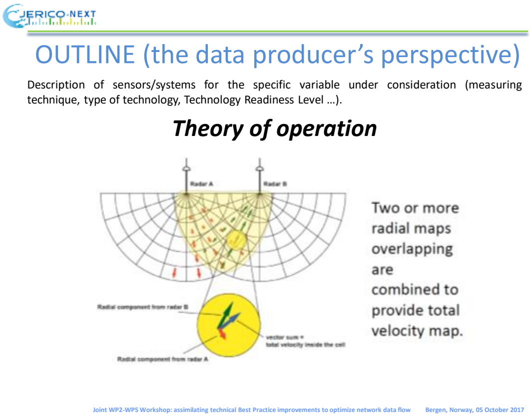


OUTLINE (the data producer's perspective)

Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Theory of operation

☐ Each HF radar system can measure only the radial component of the surface velocity

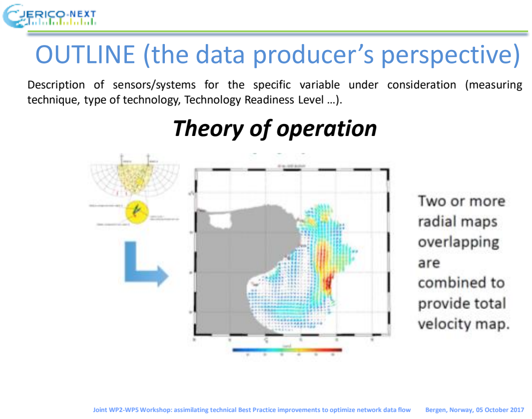


OUTLINE (the data producer's perspective)

Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Theory of operation

Two or more radial maps overlapping are combined to provide total velocity map.



OUTLINE (the data producer's perspective)

Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

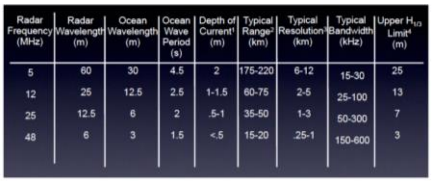
Theory of operation

Two or more radial maps overlapping are combined to provide total velocity map.

OUTLINE (the data producer's perspective)

Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Theory of operation



Radar Frequency (MHz)	Radar Wavelength (m)	Ocean Wavelength (m)	Ocean Wave Period (s)	Depth of Current^1 (m)	Typical Range^2 (km)	Typical Resolution^3 (km)	Typical Bandwidth (kHz)	Upper $H_{1/3}$ Limit^4 (m)
5	60	30	4.5	2	175-220	6-12	15-30	25
12	25	12.5	2.5	1-1.5	60-75	2-5	25-100	13
25	12.5	6	2	.5-1	35-50	1-3	50-300	7
48	6	3	1.5	<.5	15-20	.25-1	150-600	3

OUTLINE (the data producer's perspective)

Description of sensors/systems for the specific variable under consideration (measuring technique, type of technology, Technology Readiness Level …).

Conclusions

Some advantages:

☐ Land based –> low maintenance cost
☐ Continuous monitoring of the sea state in automated way
☐ Wide area covered

Some limitations:

☐ Radio frequency bands are busy -> radio interferences
☐ Possible gaps in space and time due to bad S/N ratio
☐ Only surface measurements

OUTLINE (the data producer's perspective)

Known requirements for proper deployment and common operating configurations for the described sensors/systems.

Planning and installation phase

- Selection of the desired resolution, range and coverage of the HF-radar
- Selection of the major parameter of interest (in the majority of today's setups, these are ocean surface currents)
- Depending on these criteria, the operating frequency as well as the number of systems and their relative locations can be defined.
- To identify potential installation sites, taking into account: relative location (of the sites to each other), available space (depending on the type of system), infrastructure availability and status (power supply, accessibility), and sources of possible interaction (e.g. other nearby antennas, metal fences, etc.).
- Possible sites should be chosen to satisfy logistical prerequisites first, before going on to fulfill the specific requirements in relation to the particular application, the coverage and the resolution.
- It is recommended to monitor the HF-spectrum at the selected sites in order to identify any interference issues and to plan appropriate countermeasures, e.g. selecting the most suitable frequencies

---

OUTLINE (the data producer's perspective)

Calibration and operational issues affecting data quality (e.g. limitations of calibration procedures, sampling modes, fouling, pressure effects, …).

Main Operational Issues

- Various factors affect the radar performance directly, and therefore the accuracy of the measurements, or lead to an interruption of the data flow
- Generally, data coverage is not regular for a number of reasons. Spatial and temporal data gaps may occur at the outer edge, as well as inside the measurement domain.
- This can be due to several environmental and/or electromagnetic causes: the lack of Bragg scattering ocean waves or severe ocean wave conditions, low salinity environments, and the occurrence of radio interference.
- The most frequent problems arise from environmental changes, which lead to changes of the electromagnetic field in the vicinity of the antennas and therefore to invalid antenna patterns and calibration parameters.
- Changes of antenna patterns are more significant for direction-finding systems than for phased array systems.

OUTLINE (the data producer's perspective)

Calibration and operational issues affecting data quality (e.g. limitations of calibration procedures, sampling modes, fouling, pressure effects, …).

Main Operational Issues

- Another problem is the quality loss or failure of antennas due to the environment. This happens more frequently to phased array systems as significantly more antennas are involved. For phased array systems, the performance is strongly affected if the Tx array and/or antennas close to the center of the Rx array are compromised in some way. Usually, these problems arise from damaged or broken cables, connectors or radials caused by wildlife or vandalism.
- A breakdown of the internet connection can lead to measurement gaps in the long-term record. The stability of the power supply, particularly at very remote sites, can also be a problem. These can lead to permanent data gaps but typically do not occur very often, and can be mitigated by using UPSs.
- Further dangers to operational integrity include malfunctions or downtime arising from air-conditioning failures, electromagnetic interferences, lightning strikes, accidental fires, coastal erosion and inherent system weaknesses

---

OUTLINE (the data producer's perspective)

"Meaning" of data acquired (measurement interferences, influences of seasonality and other environmental effects, sensibility to maintenance practices, descriptors used to express the quality of measurements, …).

Quality assessment is being discussed in WP2
Many associated variables are included in the mandatory ones for enable QA/QC

OUTLINE (the data producer’s perspective)

Availability and sources of useful documentation in relation to the above topics.

- Best practices in Deployment and Operation: Capacity Building, Lucy R Wyatt, ACORN, Australia
- DEPLOYMENT & MAINTENANCE of a High-Frequency Radar for Ocean Surface Current Mapping: BEST PRACTICES. Feb 2008. Radiowave Operators Working Group
- Guidelines for Assessing HF Radar Capabilities and Performance , George Voulgaris, 2011. University of South Carolina
- CODAR SeaSonde QA/QC Remote Monitoring Checklist
- CODAR SeaSonde QA/QC Setup and Diagnostics

JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

HF Radar observation: the data manager's perspective

Antonio Novellino
ETT
Antonio.Novellino@ettsolutions.com

Currently applied protocols (including metadata requirements) for handling data relating to the specific parameter under consideration within JERICO-NEXT.

Analysis of four key points for achieving a common consensus and set up a roadmap:

- Data format
- Metadata structure
- QC flagging scheme
- QC tests

Final scheme of processing levels for HFR

Processing Level	Definition	Products
LEVEL 0	Reconstructed, unprocessed instrument/payload data at full resolution; any and all communications artifacts, e.g. synchronization frames, communications headers, duplicate data removed.	Signal received by the antenna before the processing stage. (No access to these data in Codar systems)
LEVEL 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing.	Spectra by antenna channel
LEVEL 1B	Level 1A data that have been processed to sensor units for next processing steps. Not all instruments will have data equivalent to Level 1B.	Spectra by beam direction
LEVEL 2A	Derived geophysical variables at the same resolution and locations as the Level 1 source data.	Radial velocity data
LEVEL 2B	Level 2A data that have been processed with a minimum set of QC .	Radial velocity data
LEVEL 3A	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency	HFR total velocity data
LEVEL 3B	Level 3A data that have been processed with a minimum set of QC .	HFR total velocity data
LEVEL 4	Model output or results from analyses of lower level data, e.g. variables derived from multiple measurements	Energy density maps, residence times, etc.

Currently applied protocols (including metadata requirements) for handling data relating to the specific parameter under consideration within JERICO-NEXT.

Basic products: data format and QA/QC

data format	metadata structure	QC flagging scheme	QC tests
• **netCDF-4** data, and **netCDF-3.6.1**. • CMEMS IN-SITU TAC archiving strategy and folder structure, • CMEMS IN-SITU TAC naming convention. **Data var names**: • SeaDataNet P09*	***Mandatory Attr.*** • to comply with CF-1.6 and OceanSITES conventions. ***Recommended Attr.*** • to comply with INSPIRE and Unidata Dataset Discovery conventions. ***Suggested Attr.*** • relevant in describing the data, whether it is part of the standard or not.	CMEMS IN-SITU TAC – OceanSITES: **0**. unknown, no QC **1**. good, all QC passed **2**. probably good, **3**. Probably correctable, data used without scientific correction/calibration **4**. Bad data, one or more QC failed **7**. Nominal value, data not observed but reported **8**. interpolated value **9**. Missing value	• chosen among the ones listed in the **QARTOD** manual. • **are manufacturer-independent**, i.e. they do not rely on particular variables or information provided only by a specific device. • defined **for both radial and total velocity data** • **required for labelling** the data as Level 2B (for radial velocity) and Level 3B (for total velocity) data.

*) to be updated with HFR related variables

Currently applied protocols (including metadata requirements) for handling data relating to the specific parameter under consideration within JERICO-NEXT.

QC tests defined for radial data

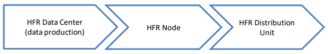
QC test	Meaning	QC variable type
Syntax	Test ensuring the proper formatting and the existence of fields within the radial netCDF file.	scalar
Over-water	Test labeling vectors that lie on land.	gridded
Velocity Threshold	Test labeling velocity vectors beyond a maximum velocity threshold.	gridded
Variance Threshold	Test labeling velocity vectors beyond a maximum variance threshold.	gridded
Median Filter	For each source vector, the median of all velocities within radius of <RCLim> and whose vector bearing (angle of arrival at site) is also within <AngLim> degrees from the source vector's bearing is evaluated. If the difference between the vector's velocity and the median velocity is greater than a threshold, then the median velocity is used	gridded
Average Radial Bearing	Test determining that the average radial bearing is close to shore normal and does not fluctuate from each hourly measurement.	scalar

QC tests defined for total data

QC test	Meaning	QC variable type
Data Density Threshold	Test checking if the minimum number of radial velocity is present for the combination into the total velocity vector.	gridded
Balance of contributing radials	Test checking if the number of radials coming from the different contributing sites are balanced for the combination into the total velocity vector.	gridded
Velocity Threshold	Test labeling velocity vectors beyond a maximum velocity threshold.	gridded
Variance Threshold	Test labeling velocity vectors beyond a maximum variance threshold.	gridded
GDOP Threshold	Test labeling velocity vectors beyond a maximum GDOP threshold.	gridded

---

data dissemination (e.g. scales, units and conversions, processed vs. unprocessed data, data reduction practices, further metadata needs …).


HFR Data Center (data production)
HFR Node
HFR Distribution Unit

HFR Data Production:

- Data Production: run HFR site or assemble HFR data
- Quality control: apply automatic quality controls that have been agreed

HFR Node:

- Acquire Data: gather available HF Radar data though collaboration with regional and national partners.
  - If the data provider can set up the data flow according the defined standards, the regional coordinator only has to link and include the new catalogue and data stream
  - If the data provider cannot setup the data flow (because of lack of experience, technical capacity etc), the regional coordinator has to work on harvesting the data from the provider, harmonize and format these data and make them available from the regional catalogue
- Data format and naming harmonization
- 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 neighbors but could not be detected by automatic QC.

HFR Distribution Unit:

- Distribution Unit: assemble data into an integrated dataset and uniform catalogue, make the data available in NRT within the European infrastructures and to the external users.

---

data dissemination (e.g. scales, units and conversions, processed vs. unprocessed data, data reduction practices, further metadata needs …).

Data infrastructure:

- THREDDS data server + INSTAC naming conv
  - Last day - RR_LATEST_ZZ_XX_CODE_YYYYMMDD.nc
  - Latest - RR_LATEST_ZZ_XX_CODE.nc
  - Monthly - RR_YYYYMM_ZZ_XX_CODE.nc
  - History - RR_YYYY_ZZ_XX_CODE.nc (e.g. IR_2016_TL_HR_BasqueHFR.nc)

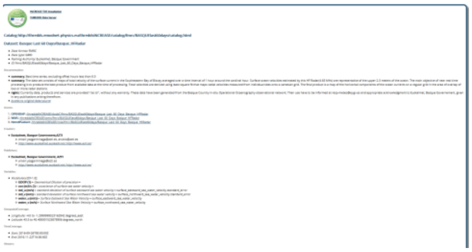
Data infrastructure:

RR = regional bigram
ZZ = type of prod. (TL/RD)
XX = HF
CODE = system name


Catalog http://thredds.cmems-increase.eu/threddsINCREASE/catalog.html

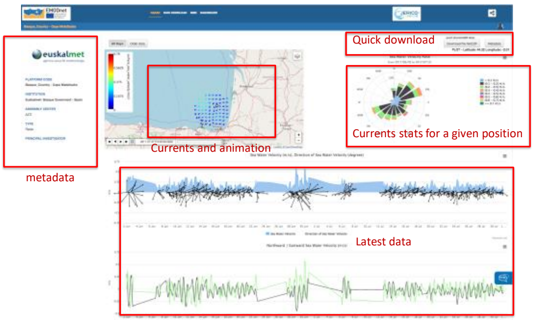
Dataset	Size	Last Modified
INCREASE		--
Last day/		--
Last 30 days/		--
Monthly/		--
History/		--

data dissemination (e.g. scales, units and conversions, processed vs. unprocessed data, data reduction practices, further metadata needs …).

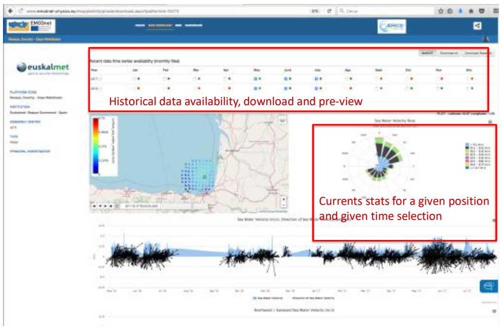


Details of the THREDDS catalogue for a given platform – last 60 days

data dissemination (e.g. scales, units and conversions, processed vs. unprocessed data, data reduction practices, further metadata needs …).



data dissemination (e.g. scales, units and conversions, processed vs. unprocessed data, data reduction practices, further metadata needs …).



Proposals/suggestions for improving the JERICO-NEXT terms of reference for handling data relating to the specific parameter under consideration.

- Make Radials available
- Start working on waves

Suggestions for better and more efficient connection with major European Data infrastructures

- Use of EDMO
- Mapping of HFR metadata vs. SDN CDIs
- Need for ragistry of systems
- Unique identifier for HFR → easier integration into GOOS

Suggestions for better and more efficient connection with major European Data infrastructures

Data gap filling and refined grid products	Short term prediction	Lagrangian products
• **Needed for key applications** (using a Lagrangian Particle-Tracking Model (LPTM) • **Gap filling** by Open Mode Analysis (OMA)[1] from radials or Variational Analysis[2] • to **provide accuracy estimations** also on the gap-filled products • Product to come together with total current field without data gap-filling	• Simple approaches e.g. **empirical models** to forecast future currents based on a short time history of past observations • **Products of interest for Marine Services - SAR and oil spill** apps • methods to be tuned up on the geographical areas of application («predictable» patterns).	• trajectory predictions using currents derived from HFR • Lagrangian particle transport model • … Not for CMEMS catalogue but as downstream application

1) Lekien et al., 2004, Kaplan and Lekien, 2007

2) Yaremchuk and Sentchev, 2009

Marine carbonate system observations: the data producer's perspective

Andrew King and Kai Sørensen
Marine Biogeochemistry and Oceanography
Norwegian Institute for Water Research (NIVA)

What type of data do we produce?

Sensor observations
pCO_2
pH

Discrete samples
Total alkalinity
Total dissolved inorganic carbon
pH (Total scale and NBS)

Sensor observations: pCO_2

Basic principle

- Equilibration of CO_2 with air space
- Either showerhead or membrane
- CO_2 in air is typically dried
- CO_2 measured by detector (mostly NDIR, e.g., LICOR)

Calibration

- Most systems only calibrate the detector (i.e., equilibration system assumed to be constant)
- Air:CO_2 mixtures are humidified and sent to drying system + detector
- Best calibration gases are from NOAA Earth System Research Lab (<0.1 ppm uncertainty)

What you are measuring and how to get to *in situ*

- xCO_2 (dry) is being measured at chamber T and P, need P to convert to pCO_2
- Measured pCO_2 is dry, need P to convert to 100% humidity
- Measured pCO_2 of seawater is at the chamber T during equilibration; need *in situ* T to correct for warming
- Most labs will also correct for non-ideality of CO_2 and calculate fCO_2 using salinity, in situ T and chamber T

What is reported

- fCO_2 (μatm) at *in situ* temperature
- Could useful to report fCO_2 at chamber T, chamber T, and *in situ* T?

Sensor observations: pH

Basic principle

- Seawater is pumped into a cell that has light sources and light detectors (spectrophotometry)
- pH sensitive indicator dye is added (e.g., thymol blue)
- Absorption at different wavelengths are measured and the ratio of these wavelengths are used to calculate pH

Calibration

- Indicator dyes need to be fully characterized to determine extinction coefficients
- Indicator dye addition can change sample pH – standard addition of dye should be carried out
- Can be calibrated using CO_2 CRMs or Tris buffer

What you are measuring and how to get to *in situ*

- pH (total scale) of seawater sample at cuvette T and perturbated by dye addition
- Need to correct for dye addition perturbation by making standard additions
- Need to correct from cuvette T to *in situ* T using empirical pH-T relationship

What is reported

- pH (total scale) at *in situ* temperature
- Again, could useful to report pH at cuvette T, cuvette T, and *in situ* T?

Discrete samples: CT (total dissolved inorganic C)

Basic principle

- Seawater is warmed up in a closed bottle to 25 deg C and pumped into stripping cell
- Phosphoric acid is added to convert all DIC into CO_2
- CO_2 is dried in a Peltier cooler and carried to coulometric detector by an N_2 gas stream

Calibration

- Pipette (~20 ml) for measuring out seawater sample needs to be checked for volume
- Sample S and T are important to calculate sample density which is combined with volume to calculate mass
- Na_2CO_3 standards can be used
- Also CO_2 CRMs should be used on a regular basis

What you are measuring and how to get to *in situ*

- Total DIC (μmol kg^{-1})
- Total DIC in a sample is independent of T as long as the bottle is gas tight
- The fraction of total DIC that is CO_3^{2-}, HCO_3^-, and CO_2 is dependent on *in situ* T
- DIC speciation can be calculated using CO2SYS as long as you have CT, AT *or* pH *or* fCO_2, S, and T; phosphate and silicate are needed if you are using AT and nutrients are high

What is reported

- Total DIC (μmol kg^{-1}), and *in situ* S and T
- Normally it is also reported with total alkalinity – next slide
- Data analyzers are free to choose which constants to use in CO2SYS for further work

---

Discrete samples: AT (total alkalinity)

Basic principle

- Seawater is warmed up in a closed bottle to 25 deg C and pumped into stripping cell
- 0.1 M HCl is added in small aliquots to titrate total alkalinity components
- Sample pH during the titration and the volume of titrant added are used to compute total alkalinity

Calibration

- Pipette (~100 ml) for measuring out seawater sample needs to be checked for volume
- Sample S and T are important to calculate sample density which is combined with volume to calculate mass
- 0.1 M HCl titrant and dosimat need to be carefully checked
- Na_2CO_3 standards can be used
- Also CO_2 CRMs should be used on a regular basis

What you are measuring and how to get to *in situ*

- Total alkalinity (μmol kg^{-1})
- Total alkalinity is independent of T as long as the bottle is gas tight
- AT is used in CO2SYS along with CT *or* pH *or* fCO_2, S, T, and phosphate and silicate if nutrients are high to calculate the rest of the carbonate system
- If you are in coastal waters (high humic content) or in very productive waters, there can be substantial non-carbonate alkalinity component – this is not characterized in CO2SYS

What is reported

- Total alkalinty (μmol kg^{-1}), and *in situ* S and T
- Normally it is also reported with CT
- Data analyzers are free to choose which constants to use in CO2SYS for further work

Discrete samples: pH

Basic principle

- Spectrophotometric method is the same as the pH sensor, except samples should be warmed to a fixed temperature (typically 25 deg C)
- Potentiometric pH measurements are made using an electrode that measures electromotive force

Calibration

- Spectrophotometric method: same as for pH sensor
- Potentiometric pH electrode needs to be calibrated using NBS buffers, but better to calibrate using seawater Tris buffers to reduce shock of going between low/high ionic strength

What you are measuring and how to get to *in situ*

- Spectrophotometric method: pH (total scale) of seawater sample at cuvette T and perturbated by dye addition
- Potentiometric method: If using seawater Tris buffers, pH (total scale); if using NBS buffers, pH (NBS scale)
- Both require knowledge of measurement T and *in situ* T
- Any sample warming/cooling must be done with closed gas tight bottle

What is reported

- pH (total scale or NBS scale) at *in situ* temperature
- Could useful to report pH at measurement T, measurement T, and *in situ* T?

What type of data do we produce?

Sensor observations

pCO_2 (ppm pCO_2 or µatm fCO_2)
pH (total scale)

- Both need to be corrected for *in situ* T and measurement T
- pCO_2 is calibrated using calibration gases
- pH is based on standard characterization of pH indicator dye

Discrete samples

Total alkalinity (µmol kg^{-1})
Total DIC (µmol kg^{-1})
pH (total or NBS scale)

- AT, CT, pH all are measured at a fixed T (e.g. 25 deg C) and need CO2SYS software to calculate other carbonate system variables at *in situ* T – dependent on with constants you choose
- AT and CT both use carbonate standards AND CRMs
- Electrode pH can be on total or NBS scale depending on the kind of calibration solutions used

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654410.

JERICO-NEXT
JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

HORIZON 2020

Marine carbonate system observations: the data manager's perspective

Benjamin Pfeil
Bjerknes Climate Data Centre
RI ICOS Ocean Thematic Centre
University of Bergen
benjamin.pfeil@uib.no

JERICO-NEXT
JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

HORIZON 2020

Marine inorganic carbon observations: marine biogeochemistry community's perspective

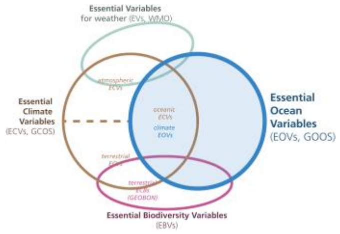
Benjamin Pfeil
Bjerknes Climate Data Centre
RI ICOS Ocean Thematic Centre
University of Bergen
benjamin.pfeil@uib.no

OUTLINE

- Purpose - why do we measure: climate vs environmental monitoring
- Quality vs coverage
- Global vs regional
- Research infrastructures vs networks, projects
- Data availability
- NRT data
- Collaboration

Driven by requirements, negotiated with feasibility

Essential Ocean Variables



- **We cannot measure everything, nor do we need to**
- Basis for including new elements of the system, for expressing requirements at a high level
- Driven by requirements, negotiated with feasibility
- Allows for innovation in the observing system over time

Essential Ocean Variables according to GOOS

EOV Information	Inorganic Carbon
Sub-Variables	**Dissolved Inorganic Carbon** (DIC), **Total Alkalinity** (TA), **Partial pressure of carbon dioxide** (pCO2) and **pH**. **[At least two of the four Sub-Variables are needed.**]
Derived Products	Saturation state (aragonite, calcite), Dissolved carbonate ion concentration, Air-sea flux of CO2, Anthropogenic carbon, Change in total carbon
Supporting Variables	**Temperature**, **Salinity**, wind speed, Atmospheric CO2 (xCO2), Barometric pressure, Oxygen, Calcium concentration, Transient tracers, Oxygen to argon ratio (O2/Ar)

Landscape for EOV Inorganic Carbon Data Management

Data availability

- Data repositories, observing and community networks
- Data products (e.g. SOCAT, GLODAP)
- Data brokers (e.g. IODE, GCMD, SeaDataCloud, GEO)
- Integrated networks (EMODnet, JERICO-Next)
- Research infrastructures (IOOS, ICOS; ARGO, EMSO)
- Research projects

ICOS INTEGRATED CARBON OBSERVATION SYSTEM

RI ICOS: Ocean Thematic Centre

Director: Truls Johannessen (University of Bergen)
Adminstrative Director: Erik Sandquist (Uni Research)
Deputy director: Benjamin Pfeil (University of Bergen)
Data management: Benjamin Pfeil, Steve Jones, Camilla Stegen Landa

---

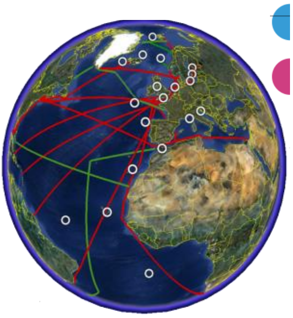
ICOS Ocean Thematic Centre

Mission: highest possible quality!

The suggested network of stations for the ocean-network:

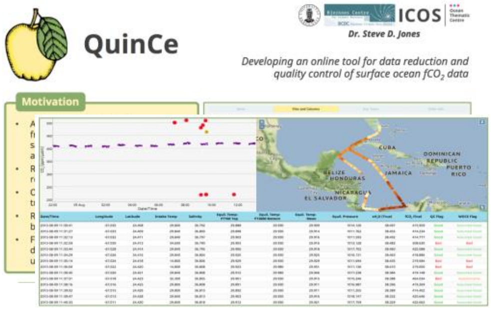
- 18 SOOP/VOS lines
- 22 fixed time series stations
- 7 repeat hydrographic sections

Currently the official OTC network is around 50%



Software for automated data reduction and QC

- Enables integration, interoperability and consistency of data streams incl.



QuinCe

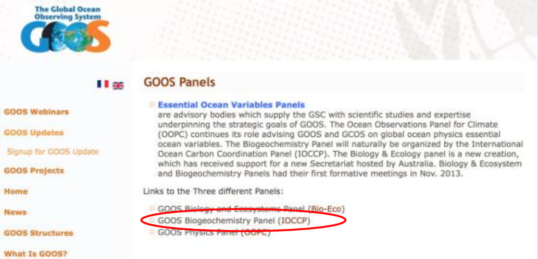
Dr. Steve D. Jones

Developing an online tool for data reduction and quality control of surface ocean fCO_2 data

Motivation

Framework for Ocean Observing
Approved governance structure

GOOS Steering Committee



The Global Ocean Observing System GOOS

GOOS Panels

Essential Ocean Variables Panels are advisory bodies which supply the GSC with scientific studies and expertise underpinning the strategic goals of GOOS. The Ocean Observations Panel for Climate (OOPC) continues its role advising GOOS and GCOS on global ocean physics essential ocean variables. The Biogeochemistry Panel will naturally be organized by the International Ocean Carbon Coordination Panel (IOCCP). The Biology & Ecology panel is a new creation, which has received support for a new Secretariat hosted by Australia. Biology & Ecosystem and Biogeochemistry Panels had their first formative meetings in Nov. 2013.

Links to the Three different Panels:

- ~~GOOS Biology and Ecosystems Panel (Bio-Eco)~~
- GOOS Biogeochemistry Panel (IOCCP)
- ~~GOOS Physics Panel (OOPC)~~

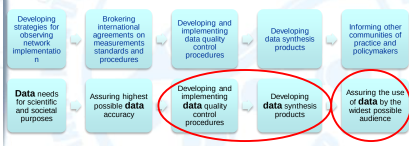
(Observing technologies and networks,
Variable focus: data and products, synthesis, link to models)

IOCCP SSG

Chair
Toste Tanhua (Germany)
Underway pCO_2:
Rik Wanninkhof (USA)
Surface CO_2 Data:
Kim Currie (New Zealand)
Repeat Hydrography:
Masao Ishii (Japan)
Ocean Interior Data:
Are Olsen (Norway)
Time Series Networks:
Laura Lorenzoni (US)
Instruments and Sensors:
Todd Martz (US)
Data Management:
Benjamin Pfeil (Norway)
Nutrients
Michio Aoyama (Japan)
Ocean Acidification
Richard Feely (USA)
SOLAS/IMBER:
Andrew Lenton (Australia)
Niki Gruber (Switzerland)
Project Director:
Maciej Telszewski (Poland)

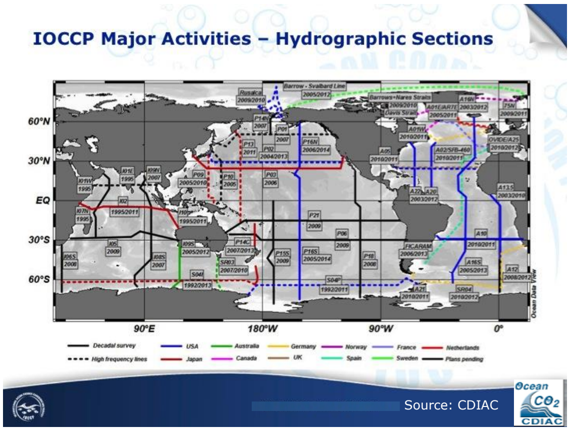
IOCCP Mission and Field of Expertize

The IOCCP promotes the development of a global network of observations for marine biogeochemistry through technical coordination and communication services, international agreements on standards and methods, and advocacy and links to the global ocean observing system. In each of the fields of our interest (left) IOCCP follows the following scheme:



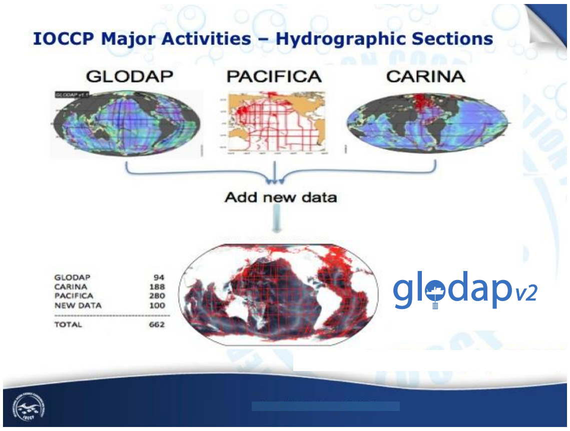
Developing strategies for observing network implementation → Brokering international agreements on measurements standards and procedures → Developing and implementing data quality control procedures → Developing data synthesis products → Informing other communities of practice and policymakers

Data needs for scientific and societal purposes → Assuring highest possible **data** accuracy → Developing and implementing **data** quality control procedures → Developing **data** synthesis products → Assuring the use of **data** by the widest possible audience



IOCCP Major Activities – Hydrographic Sections

Source: CDIAC

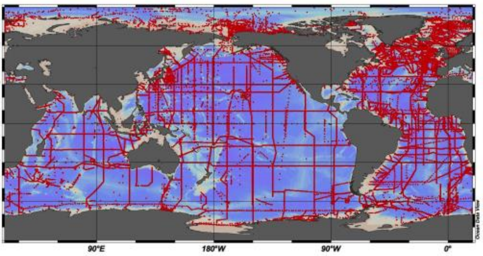


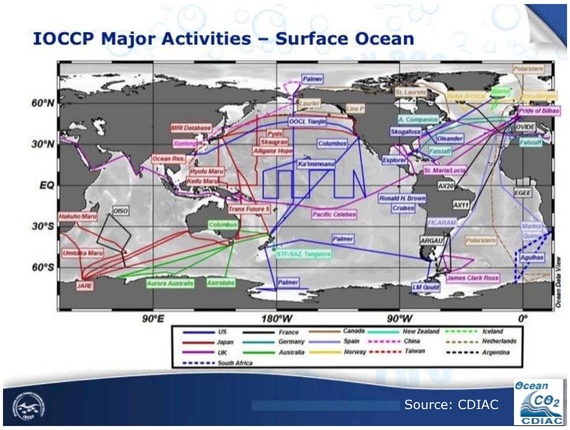
IOCCP Major Activities – Hydrographic Sections

GLODAP	94
CARINA	188
PACIFICA	280
NEW DATA	100
TOTAL	662

glodap v2

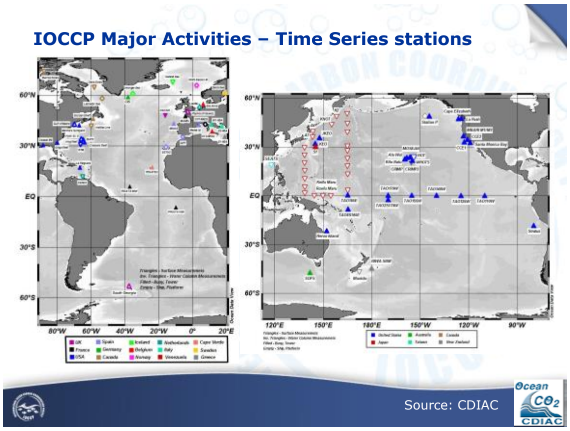
GLODAPv2 Map (45,475 stations)





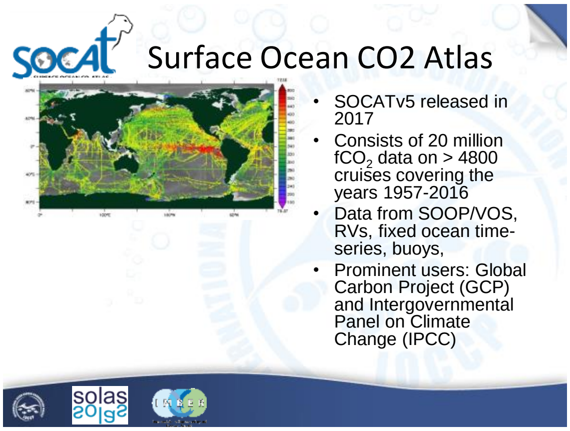
IOCCP Major Activities – Surface Ocean

Source: CDIAC



IOCCP Major Activities – Time Series stations

Source: CDIAC



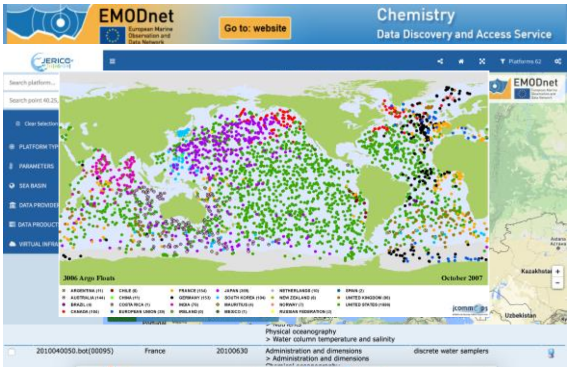
Surface Ocean CO2 Atlas

- SOCATv5 released in 2017
- Consists of 20 million fCO_2 data on > 4800 cruises covering the years 1957-2016
- Data from SOOP/VOS, RVs, fixed ocean time-series, buoys,
- Prominent users: Global Carbon Project (GCP) and Intergovernmental Panel on Climate Change (IPCC)

Is the global community (under GOOS) in perfect shape?

No

- Severe lacks in data availability (time series stations, coastal data)
- EOV Inorganic Carbon data often not covered entirely (e.g. SOCAT)
- ’New’ sensors (e.g. AUVs)
- Certain systems are not included (FerryBox)
- NRT data availability



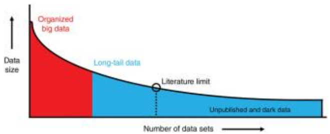
Physical oceanography
> Water column temperature and salinity

2010040050.bot(00095) France 20100630 Administration and dimensions > Administration and dimensions discrete water samplers

---

Issues - Data availability

- Duplicates
- Lack of documentation for metadata and quality control
- Varying quality control
- Movement towards *Big Data* but limited integration and access to long-tail data from individual researchers

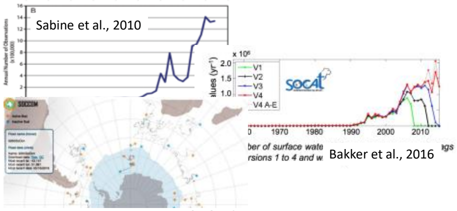


(Ferguson et al., 2014, Nature)

Data access and user friendliness

- Difficult to access data across data sources
- Reproducibility and citation
- Uncertainties (lack of QC, metadata)
- NRT data availability
- Lack of integration, interoperability and consistency (vocabulary, flags, data, metadata)
- Data products exist but just for certain data
- EOV Inorganic Carbon not covered entirely

Increasing data volumes

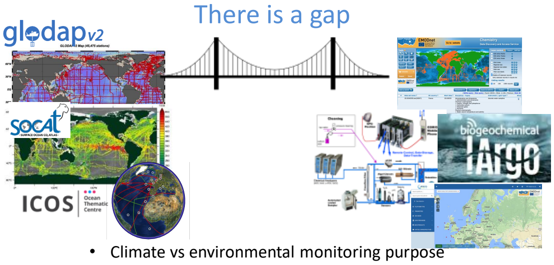


Sabine et al., 2010

Bakker et al., 2016

- New sensors and platforms generate more and more data (BGC ARGO, glider)
- Need for persistent QC procedures incl. documentation

There is a gap



- Climate vs environmental monitoring purpose
- Agency vs university sector
- Quality vs coverage
- Global vs regional
- Constant limitation: $$$, we can not do everything!

We can not do everything!

- Insufficient funding for data management activities
- Need for complimentary efforts
- More interaction between RIs, networks -> room for each other activities
- Need for modernisation
- Focus on expertise (let the experts do the QC for the various EOVs)
- Interoperable systems (QC feedback)

Collaboration instead of competition!

Demands from funding agencies and society

- Data availability across networks (GEO, SDC)
- FAIR (Findable, Accessible, Interoperable, Re-usable)
- UN Sustair

Need for change

THE INTERNATIONAL OCEAN CARBON COORDINATION PROJECT

A joint project of Scientific Committee on Oceanic Research and Intergovernmental Oceanographic Commission of UNESCO and an affiliate program of the Global Carbon Project.

Project Office:
Ul. Powstańców Warszawy 55,
81-712 Sopot, Poland
Tel.: +48 (0)58 731 16 10
Fax: +48 (0)58 551 21 30
Web: www.ioccp.org

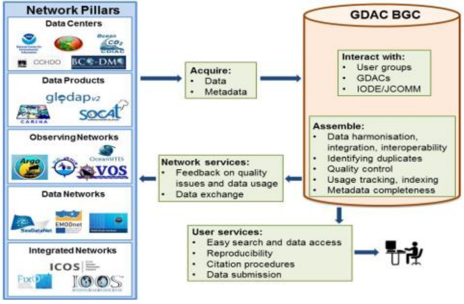
Global Ocean Biogeochemistry Data Management
IOCCP Position Paper

The international ocean biogeochemistry community is mainly using and depending upon one global data center, the Carbon Dioxide Information Analysis Center ocean trace gases section (CDIAC-Oceans) at the U.S. Department of Energy's Oak Ridge National Laboratory, USA. CDIAC-Oceans provides data management support for ocean carbon measurements from Repeat Section cruises, VOS/SOOP lines, time series and moorings data, has accommodated most community requests for data archival and data access and has also actively engaged with the science community, supporting large synthesis projects like SOCAT, the LDEO Database, GLODAP, CARINA, PACIFICA and GLODAPv2. The withdrawal of funding support for the ocean trace gases section of CDIAC puts in jeopardy

---

Community effort ensures stability

Global Data Assembly Centre for Marine Biogeochemistry



Network Pillars

- Data Centers
- Data Products
- Observing Networks
- Data Networks
- Integrated Networks

Acquire:
- Data
- Metadata

GDAC BGC

Interact with:
- User groups
- GDACs
- IODE/JCOMM

Assemble:
- Data harmonisation, integration, interoperability
- Identifying duplicates
- Quality control
- Usage tracking, indexing
- Metadata completeness

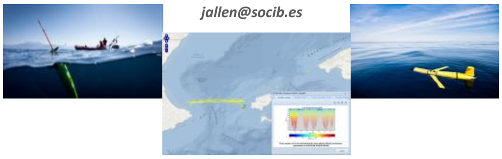
Network services:
- Feedback on quality issues and data usage
- Data exchange

User services:
- Easy search and data access
- Reproducibility
- Citation procedures
- Data submission

JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

Observations using gliders: the data producer's perspective

John Allen
SOCIB
jallen@socib.es



- Gliders are now a mature robotic ocean observing platform.

- Typically, gliders carry an instrument payload that includes a pumped CTD, an oxygen sensor and one or more biogeochemical sensors such as a fluorometer, an optical backscatter sensor and/or a PAR sensor.

- Additional payloads have included passive acoustic hydrophones, high frequency ADCPs, a nitrate sensor and a micro-structure/turbulence sensor .

- Data are typically available in three modes,
  - Real Time – a reduced dataset sent each time the glider surfaces for communications
  - Near Real Time – a complete dataset downloaded after each glider mission recovery
  - Delayed Mode – a complete dataset that has been 'field' corrected to historic datasets

- Automatic and semi-automatic data QC criterion continue to be developed and refined for the first two of these modes. Data from the RT and NRT modes are then suitable for rapid environmental assessment, and model constraint and assimilation

- Delayed mode, 'field' correction is being adopted and adapted for glider data from historical experiences with towed and lowered observational platforms. At SOCIB we are beginning to look at how this can be applied in a more semi-automatic manner – this is the subject of deliverable D 5.15, and, to a large extent, this presentation.

OBJECTIVES

- Respond to scientific and societal challenges by maintaining and enabling world class quality control of glider data at high temporal and spatial resolutions.

- Develop methods and tools to apply well-established procedures before, during, and after every mission.

- Incorporate routine multi-platform calibration and inter-calibration procedures in the validation and correction process.

- Monitor and record information concerning the calibration, validation and correction in the metadata file.

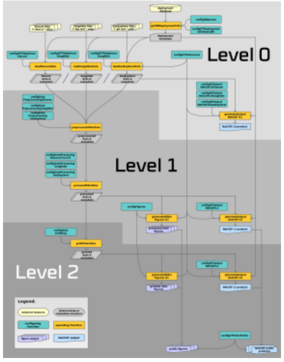
- Quantify and clearly describe the achieved accuracy and therefore residual error in the final delayed mode product.

- Guarantee traceability in the data calibration, validation and correction chain.

SOCIB Glider Toolbox

- A set of MATLAB/Octave scripts and functions – the toolbox
- Supports Slocum, Seaglider and recently added SeaExplorer platforms
- Provides:
  - Tools to generate standard netCDF files and figures from raw glider data
  - Advanced processing features, e.g. thermal lag correction
  - Standard RT and NRT quality control tests (range, spikes)
  - Covers main initial steps of the glider data management process
- Modular structure and user configurable output, for addition of new sensors, etc.
- Built for an operational facility, also useful for scientist users as standalone

User-friendly, real-time/delayed mode, processing toolkit for glider data
Delivers data ready for science and operations
- available at https://github.com/socib/glider_toolbox



- Flow diagram the best guide - information processing levels, scripts, inputs and outputs, user configurable components

- Essentially 4 processing steps and 3 levels of output: L0, L1 and L2

Contents lists available at ScienceDirect

Methods in Oceanography

journal homepage: www.elsevier.com/locate/mio

Full length article

A toolbox for glider data processing and management

C. Troupin[a,*], J.P. Beltran[a], E. Heslop[a], M. Torner[a], B. Garau[b], J. Allen[a], S. Ruiz[c], J. Tintoré[a,c]

Delayed Mode – Field Correction

- Focusing on CTD data, develop methods and software tools to make routine inter-calibrations between gliders and other platforms. Following international leading procedures and standards where they exist and taking an international lead in promoting new standards where they do not.

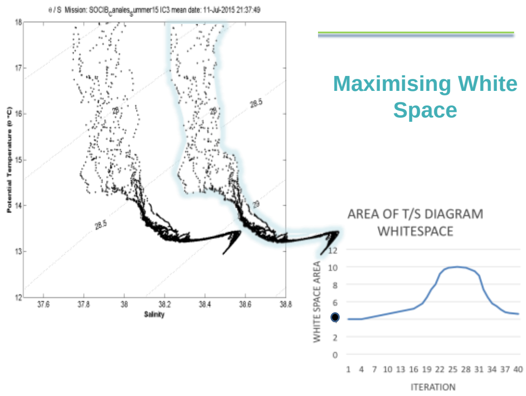
  e.g. "virtual bottle stops" points of T/S comparison with CTDs or maximising "white space" in T/S diagrams (*Allen, Gardiner and Heslop, in prep.*)

$$conductivity = A * (measured\ conductivity)$$

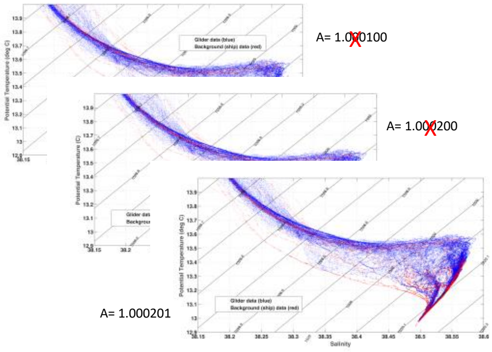
$$conductivity = A + B * (measured\ conductivity)$$

$$conductivity = A + B * (measured conductivity) + C * (measured conductivity)^2$$

Maximising White Space



Jerico-Next – WP5 meeting – Helsinki 2017



Metadata ….

To guarantee the traceability, the netCDF file should have an updated header, containing all the information necessary to understand how and when the calibration was performed.

We propose the following global attributes to be added:

```
correction_date = "2015-05-26T10:33:33+00:00" ;
correction_author = "J. Allen"
correction_email = " ... @socib.es"
correction_method = "Least-square linear regression"
correction_equation = "0.9987654
```

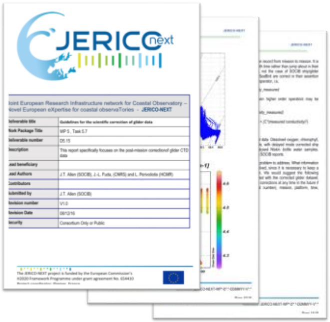
If all the variables are calibrated using the same instrument (e.g., the R/V CTD), then two more lines can be added:

```
correction_instrument = "SeaBirdCTD ..."
correction_dataurl =
"http://thredds.socib.es/thredds/dodsC/research_vessel/ctd/socib_rv-scb_sbe9002/L1/2015/dep0011_socib-rv_scb-sbe9002_L1_2015-05-19.nc"
or
correction_campaign = "SCB-POS001_SOCIB_ENL_Canales_Apr2015"
```

If some variables are calibrated against different instruments or data set, then this information should be reflected in the variable attributes, not the global attributes.

```
double salinity_corrected(time) ;
          salinity_corrected:long_name = "water salinity"
          salinity_corrected:standard_name = "sea_water_salinity"
          salinity_corrected:units = "PSU" ;
          salinity_corrected:coordinates = "time depth latitude longitude"
          salinity_corrected:_FillValue = 1.79769313486232e+308
          salinity_corrected:author = " operator1" ;
          salinity_corrected:email = " " ;
          salinity_corrected:date = " " ;
          salinity_corrected:equation = " " ;
```

Task 5.7: Deliverable D5.15 - Draft



Joint European Research Infrastructure network for Coastal Observatory – Novel European eXpertise for coastal observaTories - JERICO-NEXT

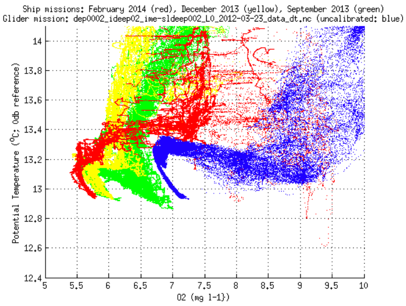
Deliverable title	Guidelines for the scientific correction of glider data
Work Package Title	WP 5 , Task 5.7
Deliverable number	D5.15
Description	This report specifically focuses on the post-mission correctionof glider CTD data
Lead beneficiary	
Lead Authors	J.T. Allen (SOCIB), J.-L. Fuda, (CNRS) and L. Perivoliotis (HCMR)
Contributors	
Submitted by	J.T. Allen (SOCIB)
Revision number	V1.0
Revision Date	06/12/16
Security	Consortium Only or Public

The JERICO-NEXT project is funded by the European Commission's H2020 Framework Programme under grant agreement No. 654410

Potential and collaborations with other WPs

Oxygen Calibration

Winkler titration to Seabird 43 to Aandera Optode



Challenging Topics

- Delayed Mode scientific correction Metadata close to full implementation at SOCIB now for physical parameters – Biogeochemical data will be a bigger challenge as parameters such as community structure will have a large part to play, as will multi-instrument diagnoses, e.g. using a combination of optical backscatter and fluorescence to determine DCMs, surface quenching, and auto versus mixo – trophic communities.

- Automating the QC and correction, where possible, of bad data resulting from bio-fouling of the conductivity sensor is still problematic and more work needs to be done. Nonetheless significant progress has been made in this direction.

- Standardise and semi-automate Depth Averaged Velocity calculations, combining navigation data, flight models, and compass correction data files. This is now well understood throughout the community, but generally follows a very manual and non-routine process.

JOINT EUROPEAN RESEARCH INFRASTRUCTURE NETWORK FOR COASTAL OBSERVATORIES

Observations using AUVs
The glider data manager's perspective

Thierry Carval
Ifremer
Thierry.Carval@Ifremer.fr

The data manager's perspective

- Currently applied protocols (including metadata requirements) for handling data relating to the specific parameter under consideration within JERICO-NEXT.
- Known limitations of the described protocols, and their level of compatibility with other EU and global data management initiatives.
- Issues affecting data dissemination (e.g. scales, units and conversions, processed vs. unprocessed data, data reduction practices, further metadata needs …).
- Proposals/suggestions for improving the JERICO-NEXT terms of reference for handling data relating to the specific parameter under consideration.
- Suggestions for better and more efficient connection with major European Data infrastructures

---

EGO data management

EGO contributors agree on a common NetCDF-CF file format to manage glider metadata and observations, organized by deployment

- *EGO gliders NetCDF format reference manual version 1.2.*
  http://doi.org/10.13155/34980
- *EGO gliders Quality Control on time series and profiles data*
  http://doi.org/10.13155/51485
- *EGO gliders data processing chain*
  http://doi.org/10.17882/45402
- *EGO gliders NetCDF file format checker*
  http://doi.org/10.17882/45538

EGO NetCDF-CF implementation for gliders

EGO gliders NetCDF format reference manual
http://doi.org/10.13155/34980

- One file per glider deployment
  - Metadata : a list of NetCDF global attributes
  - Observations : a timeseries of parameters (each observation has a time stamp)
- EGO & Argo share the same list of CF parameters
  http://www.argodatamgt.org/Documentation Core and BGC parameters
  - Endorsed by AtlantOS, Copernicus Marine, SeaDataNet

Quality control procedures

EGO gliders QC on time series and profiles data
http://doi.org/10.13155/51485

- Real-time quality control
  - 14 quality controls derived from Argo real-time QC
- Delayed mode quality control
  - A long work underway
    - Physical parameters
    - BGC parameters

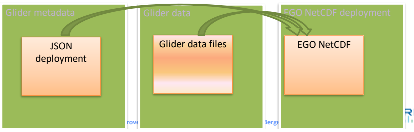
EGO data processing chain

EGO gliders data processing chain
http://doi.org/10.17882/45402

- A matlab data processing chain, freely available under CC-BY-4.0 license
- The EGO data processing chain decodes, processes, formats glider data
  Slocum, SeaGlider, SeaExplorer
- The decoder also performs the additional actions
  - Apply Real Time Quality Control (RTQC) tests on EGO file time series,
  - Estimate Slocum subsurface currents and store them into the EGO file,
  - Generate NetCDF profile files from EGO file data and apply specific RTQC tests to them.

EGO data processing chain

- The data processing chain works with
  - The deployment data files sent from the glider
  - A JSON collection of metadata
    - One deployment metadata JSON file
    - One JSON file for each sensor
    - Example
      ftp://ftp.ifremer.fr/ifremer/glider/v2/ifm12/ifm12_20170403/



EGO ftp data server

- The EGO GDAC (Global Data Assembly Centre) is available at:
  - http://www.ifremer.fr/co/ego/ego/v2
  - ftp://ftp.ifremer.fr/ifremer/glider/v2

- A directory per glider, a sub-directory per deployment
- Each deployment contains
  - The EGO NetCDF data and metadata file
  - The deployment JSON file (used for data processing)
  - A directory of all vertical profiles
    - One NetCDF file per profile (Argo NetCDF format)
    - Extracted from EGO data file (descending and ascending phases)

EGO file format checker

NetCDF file format checker for EGO gliders
http://doi.org/10.17882/45538

- Useful to check glider NetCDF files format from various providers
  - EGO data processing chain is not mandatory
    IMR and SOCIB use their own tools

Data citation

- Work underway with AtlantOS funding
  - Assign a DOI for each deployment
  - Manage DOIs of DOIs to group a series of deployment
    - Network level
    - Science process
  - Use ORCID to give credit to PIs and contributors

- Examples
  - Tintin in Greenland http://doi.org/10.17882/51473
  - Tintin & Moose http://doi.org/10.17882/51472

EGO online community glider services

http://gfcp.ego-network.org/private/login.php?ref=/private//missions/php/index.php

With « Landstation » - IP address

Back up / operational piloting service

Piloting tools
Alarms
Real time Data Visualisation
(geographical, scientific and technical data)
Automatic Mission Report

Data management System

EGO NetCDF formating
RTQC
DOI on demand
Distribution to Data integrators
✓GTS
✓EGO website
✓Copernicus Marine
✓EMODNET
✓SeaDataNet
✓OceanGliders

The last 30 days of observations