Joint European Research Infrastructure network for Coastal Observatory – Novel European eXpertise for coastal observaTories - **JERICO-NEXT**	
Deliverable title	Report on data management best practice and Generic Data and Metadata models
Work Package Title	WP 5 – Data Management
Deliverable number	D5.9
Description	QA7QC best practices and ISO quality metadata elements
Lead beneficiary	
Lead Authors	CNR-ISMAR
Contributors	G. Manzella, A. Griffa, L.P.de la Villéon
Submitted by	G. Manzella
Revision number	V2.1
Revision Date	28/07/2017
Security	Public

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

History			
Revision	**Date**	**Modification**	**Author**
1.0	28/12/2015	Draft	G. Manzella
2.0	12/06/2017	finalised	G. Manzella
2.1	28/07/2017	Addition of the Annex on the near real time data flow	L.P.de la Villéon

Approvals				
	Name	**Organisation**	**Date**	**Visa**
Coordinator	P. Farcy	Ifremer	19.06.2017	PF
WP Leaders	L. Perivoliotis	HCMR	08.09.2017	LP
	P. Gorringe	EuroGOOS	13.09.2017	PG

Table of contents

1. Executive Summary

WP5 of JERICO NEXT has the main objective of enabling free and open access to data having high quality accompanied by information that allow their easy use. This requires well defined architectures, policies, practices and procedures that properly manage the full data lifecycle needs of data producers and data users. In other words, in JERICO NEXT it is essential to assure the quality of data, control, protect, deliver and enhance the value of data and information assets.

Quality control of data is an essential component of oceanographic data management. Data quality control information tells users of the data how it was gathered, how it was checked, processed, what algorithms have been used, what errors were found, and how the errors have been corrected or flagged. Without it data from different sources cannot be combined or re-used to gain the advantages of integration, synthesis, and the development of long time series.

Information quality means consistently meeting the information customer's expectations. Information quality means the degree to which information has content, form, and time dimensions which give it value to specific end users. Based on user requirements, ISO standards are providing the quality elements that define the extent to which data sets or data set series can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

2. Introduction

JERICO NEXT objectives

The JERICO-NEXT project aims at extending the EU network of coastal observations developed in JERICO (FP7) by adding new innovative infrastructures while integrating biogeochemical and biological observations. The main target of JERICO-NEXT is to provide the researchers with continuous and more valuable coastal data coupling physical and biological information.

JERICO NEXT Data management objectives

In particular, WP5 has the main objective of enabling free and open access to data, by integrating all relevant coastal data and by facilitating their management through the JERICO Portal, EMODnet data systems (physical, chemical and biogeochemical) as well as other data management infrastructures such as SeaDataNet, Copernicus Marine, OBIS.

Marine Data Quality Management definitions

Data management can be defined as the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of data producers and data users. Data management has also the scope to assure the quality of data, control, protect, deliver and enhance the value of data and information assets.[1] Ocean Teacher defines Marine Data Quality Managements as 'the procedures, exercises and systems employed to insure the reliability and accuracy of marine data'.[2] Quality Management is usually considered to consist of the two phases Quality Assurance (QA) and Quality Control (QC).

In the data management definitions there are some concepts that need to be clarified for a correct implementation and application of procedures and protocols.

Data producers and data users

In marine science, data producers and users are (in most of the cases) the same group of communities. However, to better understand their needs it is necessary suppose that they are different communities with different requirements. The concept at the base of this report is that data sets or data set series are products created by data producers and consumed by data users. This allows the implementation in the JERICO Portal of many ISO standards that should be adapted to the JERICO NEXT case.

Data producers

On the base of ISO19115-1:2014(E) and ISO19157:2013(E) data producers are creating data sets or data set series to satisfy the purposes for their intended uses.

Data users

Data set or data set series could be used with intentions different from the one for which they were originally created. Based on user requirements, ISO19157:2013(E) is providing the quality element that is called usability, i.e. the extent to which data sets or data set series can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

[1] DAMA-DMBOK Guide (Data Management Body of Knowledge) Introduction & Project Status (https://www.dama.org)

[2] http://library.oceanteacher.org/OTMediawiki/index.php/Marine_Data_Quality_Management

The difference between data producers and data users was considered in the first Data Adequacy Report of EMODnet Mediterranean Checkpoint, where the concept was elaborated in terms of 'Universe of Discourse'[3].

Universe of Discourse

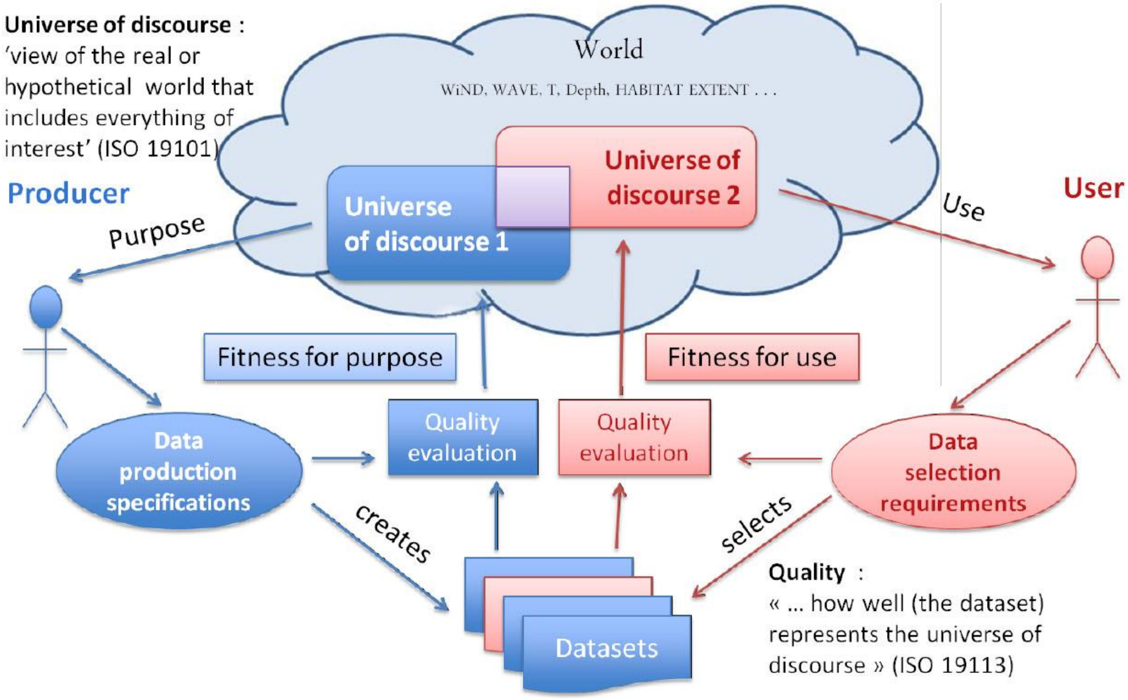


Figure 1: Universe of Discourse (from EMODnet Mediterranean Checkpoint [3])

Fitness for purpose means that data sets or data set series should be suitable for the intended purposes and is a principle of quality assurance, that is the 'set of planned and systematic actions necessary to provide appropriate confidence that a product or service will satisfy the requirements for quality[4]. QA includes management of the quality of materials, assemblies, products and components, services related to production, and management, production and inspection processes.

Quality assurance and fitness for purpose

Quality control of data is an essential component of oceanographic data management. Data quality control information tells users of the data how it was gathered, how it was checked, processed, what algorithms have been used, what errors were found, and how the errors have been corrected or flagged. Without it data from different sources cannot be combined or re-used to gain the advantages of integration, synthesis, and the development of long time series.[5]

Quality control

Information on quality and fitness for use require the use of meaningful metadata and the involvement of technical arrangements that ensure interoperability.

[3] http://www.emodnet-mediterranean.eu/wp-content/uploads/2015/06/D11.2-revised-V11.pdf
[4] http://library.oceanteacher.org/OTMediawiki/index.php/Marine_Data_Quality_Assurance
[5] http://library.oceanteacher.org/OTMediawiki/index.php/Marine_Data_Quality_Control

Best practices on Quality Assurance and Quality control have been listed in the EMODnet Physics portal (http://www.emodnet-physics.eu/portal/bibliography). In particular for the Quality Assurance methodologies it is recommended to consult the British Columbia Field Sampling Manual[6] and in general the QARTOD Manuals on QA / QC[7].

A Wiki on quality control procedures for real time, delayed mode and reprocessed data has been recently developed: https://150.145.136.231:8081/WordPress/.

To reach the objectives of JERICO two elements are important:

- assure that all components are adopting common best practices
- provide access to information on those quality elements that allow the best use of data

This report is discussing ISO elements that should be introduced in a Data Quality Management System from the production to the consume of data sets or data set series.

3. Objective of the report

Objective of the report

The objective of this report is not to provide an overview of best practices on data management and on metadata models. All the necessary information can be found in many documents accessible via internet. Particular bibliographic indications will be provided here.

The objective is the clarification of links existing between QA/QC protocols and ISO Quality Elements, and suggest how all these elements could be include in a metadatabase.

An effort done in this report is the inclusion of both quantitative and non quantitative elements, including users' feedbacks or expert evaluations in metadata.

4. Data quality

Data Quality

Data quality concepts provide an important framework for data producers, as well as, for data users. A data producer is given the means for validating how well a data set reflects its universe of discourse as defined in the data product specification. Data users can assess the quality of a data set to ascertain if it is able to satisfy the requirements of the data user's application (Figure 1).

QA/QC

Quality evaluation procedures may be used in different phases of a product's life cycle. In general two phases are defined in data quality management: Quality Assurance and Quality Control. However, an additional element is added in this report, taking into consideration that

[6] http://www.env.gov.bc.ca/epd/wamr/labsys/field_man_pdfs/fld_man_03.pdf
[7] http://www.ioos.noaa.gov/qartod/welcome.html

data can be re-used many times. By using the nomenclature of data Quality in INSPIRE, this third phase is called Quality Control during data set update.

Quality Assurance

The Quality Assurance is the set of procedures and protocols applied before and during the data set creation.

Quality Control

The Quality Control is an inspection for conformance to a data product specifications. Data product specification is defined as the "detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party". If the data set passes inspection (composed of a set of quality evaluation procedures), the data set is considered to be ready for use.

Quality control during data set update is composed by those procedures used both for update and for benchmarking the quality of the data set after an update has occurred.

Quality control during data set update

5. Data production monitoring and diagnostic

One of the preliminary actions identified to achieve the objectives of the Data Quality Management System is the provision of an overview of the sensors used on platforms and methodologies used for data collection. The quality of data and products can be assured by:

1. knowledge of the sensors, accuracy;
2. calibration and inter-comparison of sensors;
3. Quality of the sampling strategy;
4. Quality assurance of field work;
5. Quality assurance of collected data;

Quality assurance protocols and best practices have been developed on the base of the international agreed policies and standards (e.g. IOC, ICES, WMO, ISO). The application of the quality assurance guidelines into the specific fieldwork is guaranteed by a list of actions to be checked every time.

ISO17359 standard adapted for marine data quality assurance

The standard that is better fitting this purpose is ISO17359:2002(E). Figure 2 is providing the block diagram of procedures to be adopted for the assurance of quality of a product.

At the beginning of each data collection there is the need to have the most general descriptions of the instruments working principle and the expected results. A criticality assessment of all instruments/platforms is recommended in order to create a prioritized list of sensors/platforms to be included (or not) in data collection systems. This may be a simple rating system based on factors such as:

- cost of sensor/platform down-time or lost production cost
- failure rates, mean time to repair
- consequential or secondary damage

- replacement cost of the instruments/platforms
- cost of maintenance or spares
- life cycle costs
- cost of the monitoring system
- safety and environmental impact.

Failure modes and effects analysis (FMEA) or failure mode effect and criticality analysis (FMECA) are adopted to identify expected faults, symptoms, and potential parameters to be measured which indicate the presence or occurrence of the faults. The FMEA and FMECA audits will produce information on the range of parameters to be measured for particular failure modes.

Consideration should be given to the feasibility of acquiring the measurement including ease of access, complexity of required data acquisition system, level of required data processing, safety requirements, cost, and whether surveillance or control systems exist which are already measuring parameters of interest.

Records of monitored parameters should include as a minimum the following information:

- essential data describing the sensors and platforms,
- the measurement position,
- the measured quantity units and processing, and date and time information.

Other information useful to allow comparison include details of the measuring systems used, and the accuracy of each measuring system. Details of instruments/platforms configuration and any component changes should be also included.

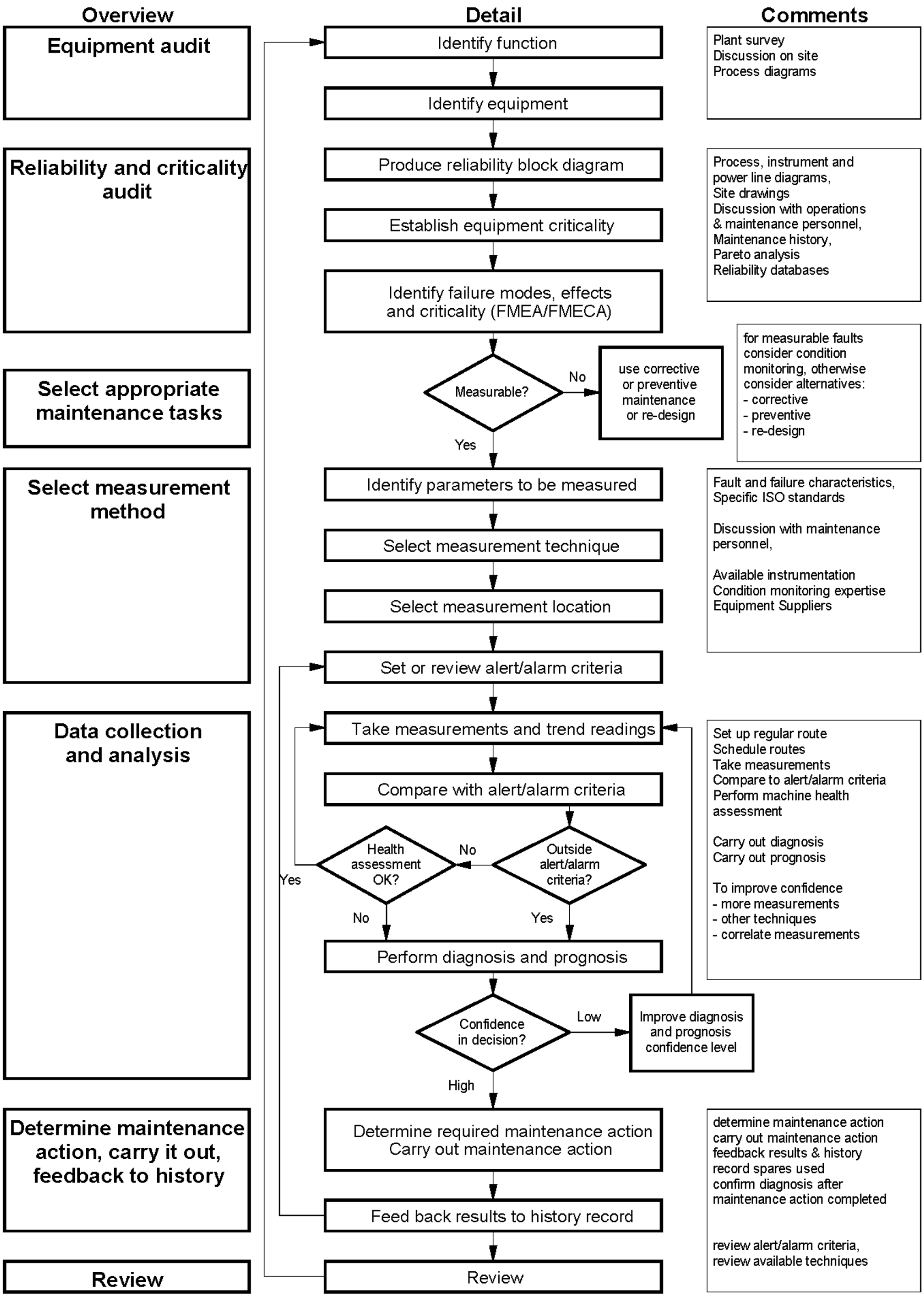


Figure 2. ISO19359 block diagram to be adapted for marine data quality assurance.

The application of ISO19359 concepts would allow making the right choice in case of problems encountered during data collection. As matter of fact, this requires the presence of trained and qualified personnel. The presence of trained/qualified personnel, the knowledge of instruments and the environment to be investigated and results expected allow making critical decisions in case of any material malfunction, deviation or deficiency relative to: ISO19359

- description of procedures
- changes in the ship track with respect to the monitoring program
- malfunctions of the instrumentation
- failures in the data acquisition
- any kind of problem that can arise during the monitoring.

A pictorial scheme of the steps that should be applied to assure the quality of the measurement program strategy is shown in Figure 3.

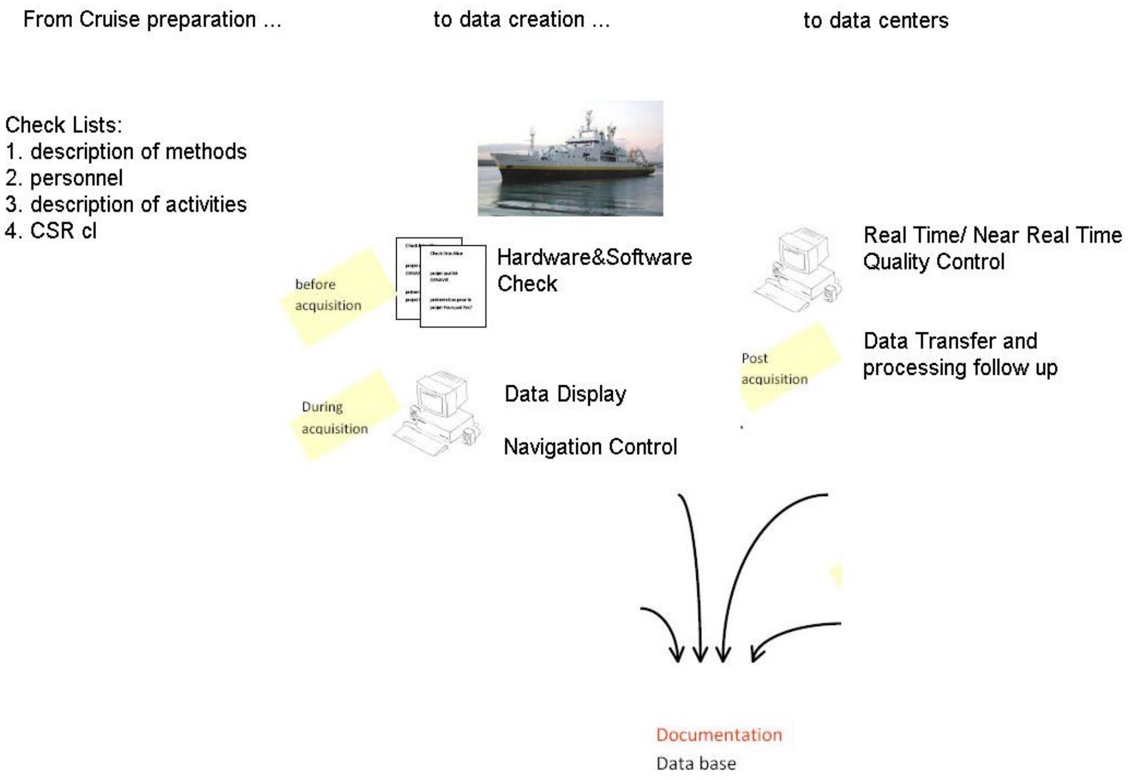


Figure 3. The best practices for data acquisition.

6. Quality assurance

To assure good quality of data it is necessary to follow precise steps from cruise preparation to data delivery to shore. The surest way to operate a robust quality assurance program is to adopt standard procedures through the use of methods and manuals agreed at international level.

All major programs have issued data management manuals for the specific methods and procedures they adopt for their own use, however, these manuals are not comprehensive for all purposes. Here the general concepts for quality assurance are presented.

Check Lists

These check lists are intended as reminder for the action to be done. Behind any check list there are the best practices that are defined in reports and protocols that must be well known by the technicians and scientists dedicated to data collection. Also the formal education of the personnel in data collection is the best assurance of data quality. The check list is a useful reminder that all necessary actions have been done before, during and after the cruise. It can be composed by many components that are herewith presented.

Description of Methods Check List

This part is related to the a priori knowledge on methods and instruments to be used during the measurement program, the results that can be expected and working instructions to personnel. It is required that the personnel is knowing the instruments, their working principles, what are the limitations in the use of such instruments and what is presumed to be the data to be collected. This means that manuals and protocols specifying how to use the instruments are available to personnel and have been assimilated in the working practices. The availability of historical data can greatly improve the quality of the work. The check includes:

Methods check list

- Principle of method described,
- What will be investigated,
- Representativeness of method stated,
- Drawing up working instructions
- Equipment list drawn up,
- Supplementary observations and measurements specified

Personnel Check List

Personnel check list

This makes sure that the personnel is able to do the required work, and if the organization of the work is well defined and clear to everybody. It includes:

- Education of field personnel specified
- Personnel plan drawn up
- Field of responsibility for task defined

Description of Activities Check List

This is related to the planning of the cruises. Areas of investigation and ownership of data is included in this list:

Activities check list

- Station points clearly identified
- Description of measuring site prepared
- Other field activity described

Sensors Check List

It is also recommended to have a check list containing:

- a list of sensors
- information on validation test performed before the use
- information on behavior during the use
- state audit sensors and position

Sensors check list

Software Check List

Tests have to be done also to the software used for data acquisition. In some cases, the complete data acquisition system composed by sensors, sensors' connections, cable to computer and computer could have a response that is varying on the base of sensor and computer type, as well as the hardware included in the computer for the acquisition of data. A check of the response is particular important when data are collected with different systems.

Software check list

Cruise Summary Report Check List

This check list is important for the production of the CSR at the end of each cruise. Also in this case the responsibilities have to be well defined:

Report check list

- Institute responsible for the cruise stated
- Person responsible for the cruise stated
- Record the personnel
- Note special aspects during the measurements
- Note deficiencies
- Ownership status clearly identified

Data acquisition hardware/software

This part is containing the necessary checks to be done on the instrumentation in order to assure:

Monitoring sensors and software

- consistency checks on the sensors before the cruise
- monitoring control of the electronic
- state audit of sensors

This is done by means of tests on sensors and connection between sensors and a control of the sensors and connection functioning during their use.

Data display

This is part of the quality control process. Visualization of sensors' data will help the personnel providing alert and identifying malfunctions. This includes:

Data visualisation

- visualization of all data (raw and engineering from sensors)
- visualization of wrong data
- identification of malfunctions
- reporting of sensors conditions

7. Quality control

Quality Assurance is a way of preventing mistakes in data and avoiding problems when delivering data derived products to users. This 'defect' prevention in quality assurance differs

Quality control focus

subtly from defect detection and rejection in quality control that is essentially focused on 'process output'.

Quality control is a process by which data are validated and the quality of all factors involved in production reviewed. Scientific, analytical and statistical evaluation of data must determine if data are of the right quality to support their intended use. Quality control ensures the data consistency within a single data set and within a collection of data sets, and ensures also that the quality and errors of the data are apparent to the user, who has sufficient information to assess its suitability for a task.

The quality of data is the degree to which it satisfies the stated and/or implied needs of its various users, and thus provides value. Those users' needs (functionality, performance, security, maintainability, etc.) are precisely what are represented in the quality model, which categorizes the product quality into characteristics and sub-characteristics.

Some ISO/IEC 25010 elements that can be included in Quality Control procedures are 'Usability' (ISO/IEC 25010 Product Quality) and 'Effectiveness and Context Coverage' (ISO/IEC 25010 Quality in Use):

QC ISO elements

- ISO defines usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use."
- Effectiveness is the extent to which data fulfils the intended use.
- Context Coverage is the degree to which data can be used with effectiveness and satisfaction in both specified contexts of use and in contexts beyond those initially explicitly identified.

The measurement method and the data quality control procedure for a parameter are dependent on each other, because each measurement method and each parameter type need some special data quality control procedures in addition to the generic checks on timing, position etc. Data quality control procedures can be divided into procedures which are:

a) applied by the owner or originator of data to improve the data consistency within the data set, and

b) applied by a data manager to improve the data consistency within a data bank, or in a multi source data set.

Regarding the data quality control measures, the originator is responsible for the following:

Data originator responsibility

- use of documented or international recommended standard measurement methods and equipment;
- national and international calibration of measurement methods and instruments;
- data validation according to results of calibration and intercalibration as well as in comparison with standard methods;

- information on temporal and spatial sampling;
- tests of fixed and computed limits, gaps and constant values;
- detection, correction, and flagging of spikes;
- detection, correction, and flagging of errors in position and time;
- documentation of the process of data sampling and validation, including any algorithm applied;
- documentation of QC checks carried out and their results.

Data quality flags provide the user of the data with clear information about actions taken to change the original data.

The data quality procedures ensure the data consistency within a data bank. They include procedures for:

Data manager responsibility

- test of format coding;
- check of incoming data set against location and identification errors;
- tests of fixed and computed limits;
- tests according to climatological standard;
- visual inspection;
- duplicates check;
- parameter screening;
- oceanographic and meteorological assessment.

8. Information quality

Information quality metadata elements

According to Arndt and Langbein ([8]), information quality means consistently meeting the information customer's expectations. Information quality means the degree to which information has content, form, and time dimensions which give it value to specific end users ([9]).

ISO standards for cataloguing the information includes:

- 8601 Representation of date and time
- 19108 Temporal characteristics of geographic information
- 19113 revised by 19157 standard for geographic information
- 19115 Geographical information metadata
- 19119 Taxonomy of services
- 19139 Geographical information metadata implementation specification

ISO 25010 Quality

[8] Arndt, D. and N. Langbein (2002). Data Quality in the Context of Customer Segmentation. International Conference on Information Quality.

[9] O'Brien, J. A. (2003). Introduction to Information Systems (Twelfth Edition), Mc Graw-Hill/Irwin.

The product quality model defined in ISO/IEC 25010 comprises quality characteristics shown in figure 4:

Product Quality - ISO/IEC 25010

Characteristics	Sub-Characteristics	Definition
Functional Suitability	Functional Completeness	degree to which the set of functions covers all the specified tasks and user objectives.
	Functional Correctness	degree to which the functions provides the correct results with the needed degree of precision.
	Functional Appropriateness	degree to which the functions facilitate the accomplishment of specified tasks and objectives.
Performance Efficiency	Time-behavior	degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements.
	Resource Utilization	degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements.
	Capacity	degree to which the maximum limits of the product or system, parameter meet requirements.
Compatibility	Co-existence	degree to which a product can perform its required functions efficiently while sharing a common environment and resources with other products, without detrimental impact on any other product.
	Interoperability	degree to which two or more systems, products or components can exchange information and use the information that has been exchanged.
Usability	Appropriateness recognisability	degree to which users can recognize whether a product or system is appropriate for their needs.
	Learnability	degree to which a product or system enables the user to learn how to use it with effectiveness, efficiency in emergency situations.
	Operability	degree to which a product or system is easy to operate, control and appropriate to use.
	User error protection	degree to which a product or system protects users against making errors.
	User interface aesthetics	degree to which a user interface enables pleasing and satisfying interaction for the user.
	Accessibility	degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use.
Reliability	Maturity	degree to which a system, product or component meets needs for reliability under normal operation.
	Availability	degree to which a product or system is operational and accessible when required for use.
	Fault tolerance	degree to which a system, product or component operates as intended despite the presence of hardware or software faults.
	Recoverability	degree to which, in the event of an interruption or a failure, a product or system can recover the data directly affected and re-establish the desired state of the system.
Security	Confidentiality	degree to which the prototype ensures that data are accessible only to those authorized to have access.
	Integrity	degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data.
	Non-repudiation	degree to which actions or events can be proven to have taken place, so that the events or actions cannot be repudiated later.
	Accountability	degree to which the actions of an entity can be traced uniquely to the entity.
	Authenticity	degree to which the identity of a subject or resource can be proved to be the one claimed.
Maintainability	Modularity	degree to which a system or computer program is composed of discrete components such that a change to one component has minimal impact on other components.
	Reusability	degree to which an asset can be used in more than one system, or in building other assets.
	Analyzability	degree of effectiveness and efficiency with which it is possible to assess the impact on a product or system of an intended change to one or more of its parts, or to diagnose a product for deficiencies or causes of failures, or to identify parts to be modified.
	Modifiability	degree to which a product or system can be effectively and efficiently modified without introducing defects or degrading existing product quality.
	Testability	degree of effectiveness and efficiency with which test criteria can be established for a system, product or component and tests can be performed to determine whether those criteria have been met.
Portability	Adaptability	degree to which a product or system can effectively and efficiently be adapted for different or evolving hardware, software or other operational or usage environments.
	Installability	degree of effectiveness and efficiency in which a product or system can be successfully installed and/or uninstalled in a specified environment.
	Replaceability	degree to which a product can replace another specified software product for the same purpose in the same environment.

Quality in Use - ISO/IEC 25010

Characteristics	Sub-Characteristics	Definition
Effectiveness		accuracy and completeness with which users achieve specified goals
Efficiency		resources expended in relation to the accuracy and completeness with which users achieve goals
Context Coverage	Context Completeness	degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in all the specified contexts of use
	Flexibility	degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in contexts beyond those initially specified in the requirements

Figure 4. ISO 25010 products elements and quality in use of products.

ISO 25012 elements

In 2008 the standard ISO/IEC 25012 was introduced defining a general data quality model for data retained in a structured format within a computer system. The standard can be used to establish data quality requirements, define data quality measures, or plan and perform data quality evaluations. ISO/IEC 25012:2008 categorizes quality attributes into fifteen characteristics:

- Accuracy,
- Completeness,
- Consistency,
- Credibility,
- Correctness,
- Accessibility,
- Compliance,
- Confidentiality,
- Efficiency,
- Precision,
- Traceability,
- Understandability,
- Availability,
- Portability,
- Recoverability

Considered by two points of view: inherent and system dependent. It also underline that data quality characteristics will be of varying importance and priority to different stakeholders.

Information quality is a wider concept than data quality and is referred to a wide type of data and characteristics represented according to different heterogeneous models, such as texts, maps, images, etc. However, the many attempts done to define the corresponding quality elements have been applied in the field of marketing ([10],[11],[12]).

9. ISO 19xxx Core Elements

ISO19115

The ISO 19115 Standard requires a basic minimum number of metadata elements that are essential for the checkpoint:

- Dataset or dataset series on specific challenges ('what'),
- Geographic bounding box ('where')",

[10] A. Parasuraman, V.A. Zeithaml, and L.L. Berry (1985), "A conceptual model of service quality and its implications for future research", Journal of Marketing, Vol. 49, autumn, pp. 41-50.

[11] Nitin Set, S.G. Deshmukh, Prem Vrat (2005), "Service quality models : a review", International Journal of Quality & Reliability Management, vol. 22, No 9, pp.913-949

[12] Hongxiu Li, Reima Suomi, (2009), " A proposed scale for measuring E-service Quality", Interntional Journal of u- and e-Service, Science and technology , vol. 2n No. 1, 2009.

- Temporal extent (‘when’),
- Contact point to learn more about or order the dataset (‘who’).

Additional elements increase interoperability and usability.

ISO 19113 defines quality principles, which are applied in ISO 19115 (geographic metadata). There are also related works in the Guide to the Uncertainty in Measurement series13. The metadata records in the current GEOSS use the ISO 19115 data model and its companion XML encoding (ISO 19139). ISO 19138 defines quality measures, but it is superseded by the standard, ISO 19157, which also supersedes ISO 19113.

ISO19157

INSPIRE is requiring the following Data Quality elements:

INSPIRE Data Quality Elements

1. Consistency - Logical, Conceptual, Domain, Temporal, Format;
2. Completeness – Commission, Omission;
3. Accuracy – Positional, Temporal, Thematic, Classification correctness;
4. Lineage;
5. Usability.

All these elements are part of the quality metadata set.

The first four quality elements might be characterized as *producer data quality elements* because they are known by the data producer and are the ‘classical’ quality indicators for the evaluation of spatial data quality.

10. User needs for data quality

Metadata Scope Elements

When using the data (e.g. for research, modelling or decision-making) quantified accuracy judgements are required in order to make best use of the quality information. For example, in data discovery/search, users might be most interested in expert statements about the utility of the dataset for a specific purpose. They could not use all information on quality, but they may need to know that detailed information exists, and is reliable. This introduces the concept of the granularity, or scope, of the data quality descriptors.

From a study conducted by Yang et al.[14] It was found that users are interested in:

- ‘Soft’ knowledge about data quality—i.e. data providers’ comments on the overall quality of a dataset, any data errors, potential data use and any other information

[13] Bich W, Cox MG, Harris PM. (2006). Evolution of the ‘guide to the expression of uncertainty in measurement’. Metrologia 43, 161. (doi:10.1088/0026-1394/43/4/S01)10.1088/0026-1394/43/4/S01

1. [14] X. Yang, J. D. Blower,L. Bastin, V. Lush, A. Zabala, J. Masó, D. Cornford, P. Díaz, and J. Lumsden, 1983. An integrated view of data quality in Earth observation. Phylosophical Transaction A A Math Phys Eng Sci. 2013 Jan 28; 371: 20120072. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3538291/

that can help to assess fitness-for-use of the data (these are contained in the ISO 'purpose', 'lineage', 'usage').

- Knowledge on expert selection of most suitable datasets.
- Dataset provenance (this is contained in ISO 'lineage') as well as citation and licensing information when assessing whether data are fit for purpose (this is in policy metadata – Identification and Constraint Information of ISO19115). The reputation of data providers was identified as a key factor in dataset selection. A problem can arise on the fact that users typically rely on data from producers that they already know or those who have a good reputation in the community. In checkpoints the only way approach this problem is based on expert knowledge.
- Quality information visualization for complex data sets would allow geospatial datasets to be compared more effectively, especially when datasets are very similar and differences are hard to distinguish. Such functionality would support and simplify data searches, decision-making and data quality evaluation, particularly for less knowledgeable and non-expert users who find it hard to manually inspect data to assess their fitness-for-use. In ISO 19157 coverage and grid of data uncertainty can be managed.

The interest on this study by Yang et al. Rely on the fact that a case study was 'Ocean Data Reanalysis' generated from consultation with a scientist on the MyOcean project (now Marine Copernicus), who was concerned about *validation*, in which the reanalysis results are compared with known and trusted datasets in order to characterize the quality of the reanalysis. This case study showed that users sometimes require extremely detailed information about the heritage of a dataset, encompassing producer-specified data (e.g. provenance and error estimates) and user-generated metadata (citations and comments).

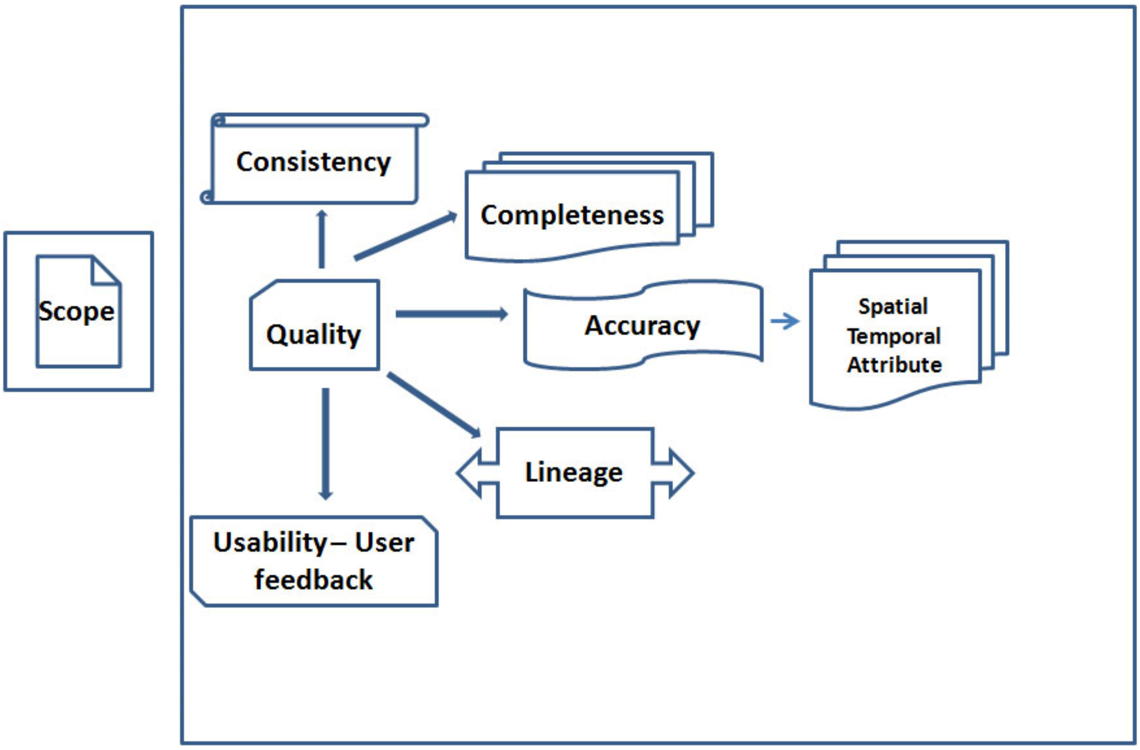


Figure 5. The metadata Scope elements

11. General quality requirements

Other general requirements from the user side are the access to best data (in terms of minimal uncertainty with respect to reality) subject to some constraints on spatial and temporal coverage of the data, and their relevance to user tasks. Users need to have appropriate quality information, which can include:

- per-product quality;
- availability and reliability of per-item quality information; granularity of per-item quality information (dataset, field or pixel);
- nature of per-item quality information (quantified values);
- extent, completeness, legend, lineage and reputation of data (based on expert knowledge);
- community assessments of data relevance and usability within this application domain.

The JERICO Next datasets have spatial, temporal, and thematic elements that must be considered for the definition of quality components. For each of the spatial, temporal and thematic dimensions, several components of quality (including accuracy, precision, consistency, and completeness) can be identified.

- ISO 19113 non-quantitative elements
    - "purpose" the rationale for creating a dataset, containing information about its intended use. Purpose
    - "usage" the application(s) for which a dataset has been used by the data producer (if the producer use is different from creation purpose) or by other, distinct, data users. Usage
    - "lineage" the history of a dataset and the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form. Lineage
- Quantitative elements
    - "completeness" (commission: excess data present in a dataset; omission: data absent from a dataset) Completeness
    - "logical consistency" (conceptual consistency: adherence to rules of the conceptual schema; domain consistency: adherence of values to the value domains; format consistency: degree to which data is stored in accordance with the physical structure of the dataset; topological consistency: correctness of the explicitly encoded topological characteristics of a dataset) Logical Consistency
    - "positional accuracy" (absolute or external accuracy: closeness of reported coordinate values to values accepted as or being true; relative or internal accuracy: closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true; gridded data Positional Accuracy

position accuracy: closeness of gridded data position values to values accepted as or being true)

- "temporal accuracy" (accuracy of a time measurement: correctness of the temporal references of an item; temporal consistency: correctness of ordered events or sequences, if reported; temporal validity: validity of data with respect to time)
- "thematic accuracy" (classification correctness: comparison of the classes assigned to features or their attributes to a universe of discourse; non-quantitative attribute correctness: correctness of non-quantitative attributes; quantitative attribute accuracy: accuracy of quantitative attributes)

Temporal Accuracy

Thematic Accuracy

12. General quantitative and non-quantitative information quality elements

4 dimensional quality elements

In the scheme presented in Figure 6, it can be better specified the 'Usability' and 'User Feedback'. Usability is an element that can be based on expert knowledge, and the above scheme can be completed in a four dimensional scheme containing both ISO standard quality information and knowledge base quality information divided by: intrinsic information quality, contextual information quality, representational information quality and accessibility information quality.

Part of these four dimensional quality elements are represented in ISO 'scope' standard elements (e.g. accessibility, timelines), other are part of an evaluation process based on expert judgement or community knowledge (e.g. believability, reputation, value-added).

To update the existing records the following elements needs a better specification of element included in the checkpoint assessment criteria: Reliability, Purpose, Usage, Logical Consistency

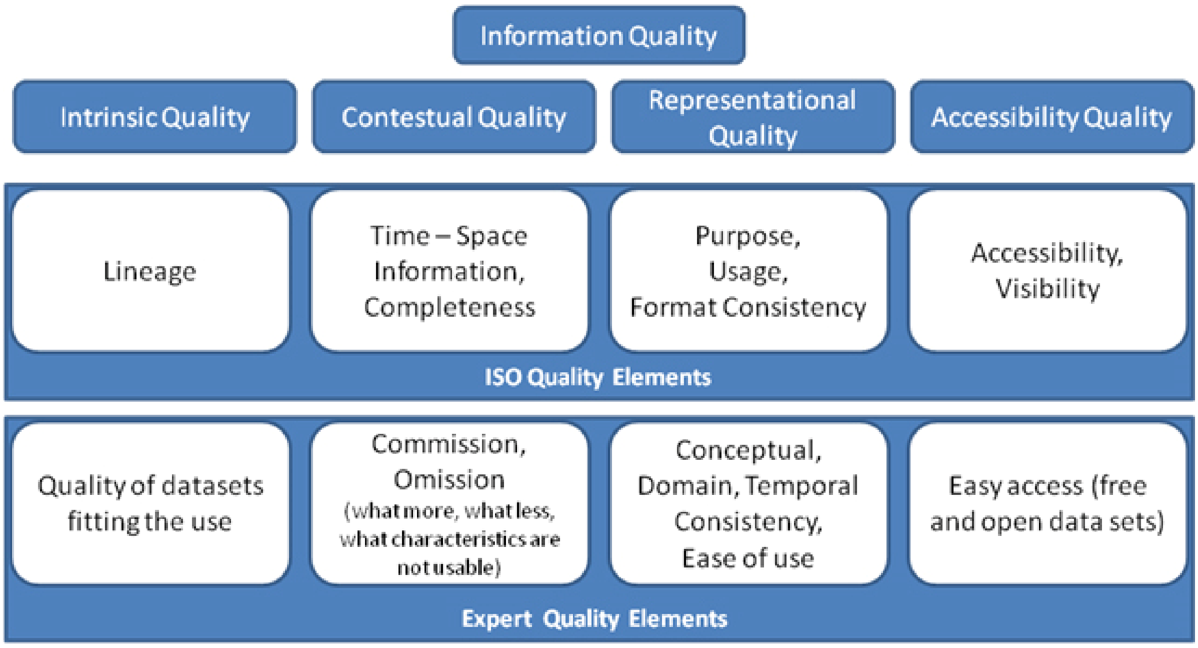


Figure 6. Information quality elements

Purpose

Purpose describes the rationale for creating a dataset and contain information about its intended use. It is containing the story / general description of the characteristics.

Usage

Usage describes the application(s) for which a dataset is used. Usage describes uses of the dataset by the data users. This element should contain also information about the heritage of a dataset.

Logical consistency

The main interpretation problems are related to the conceptual consistency, that must be simplified in some way. The conceptual scheme at the base of the logical consistency could be the 'reference model' for each data collection system, i.e. the ideal set of characteristics and their qualities fitting for use. Being the representation of an 'ideal world' the conceptual scheme can provide the metric to assess the quality of data sets and derived products.

User Feedback

Users are belonging to communities of practice and are practically doing an assessment of all procedure done to assure the quality of data sets and information associated to them.

13. Aknowledgement

The development of the concepts on ISO Quality Elements has benefitted of the many inputs provided by Eric Moussat from Ifremer, France.

Annex to D5.9 – Near Real Time Data Flow Description for InSitu Data

IN SITU OCEAN
NEAR REAL TIME DATA FLOW DESCRIPTION
Annex of D5.9
"Report on data management best practice and generic Data and metadata models"

Version 1.0

Abstract:
This "In situ ocean real-time data flow description" is an annex of D5.9 "Report on data management best practice and generic Data and metadata models"
Its describes how the ocean data flows in near-time in the frame of European Projects or initiatives and in particular in the Copernicus Marine Service –CMEMS- in situ TAC component.

Change Record

Issue	Date	§	Description of Change	Author	Validated By
V1.0	28/07/2017	All		L Petit de la Villéon, Ifremer	Leonidas Perivoliotis, HCMR, WP5 leader Patrick Gorringe, EuroGOOS, WP5 co-leader

Table of contents

14. Introduction and context

In the continuity of the preparation phase performed during the European Projects Mersea , MyOcean and MyOcean2, the Copernicus[15] Marine Environment Monitoring Service (CMEMS) measures, models and forecasts the state of the global oceans and regional seas, providing in a pre-operational mode more than 150 specific products comprising data from satellite images, ocean forecast models and ocean observations (measurements taken in the sea).
CMES is the major European program for ocean observations and forecasts.
In situ ocean observed data is a key element for ocean forecast (data assimilation in near-real time, model reanalysis and model validation in delayed mode)

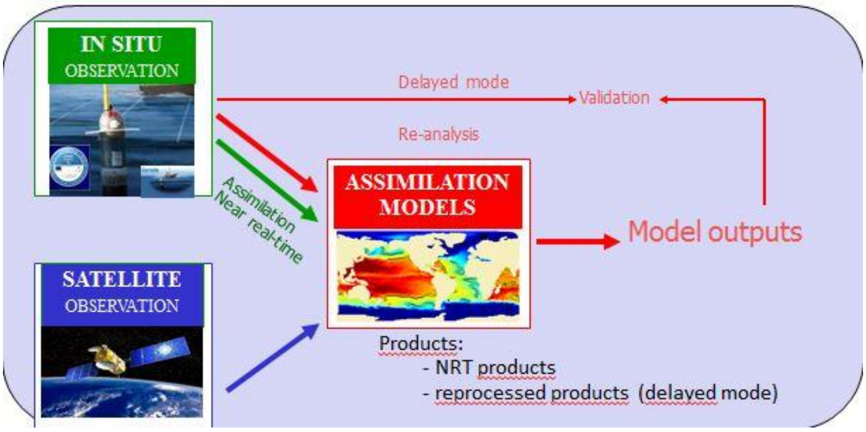


Figure 1: The 3 components of an operational oceanography system

Within CMEMS, the In Situ Thematic Assembly Centre (INSTAC) ensures that a steady supply of these in situ ocean measurements is made available to the other service components.

The In Situ TAC is a distributed system built on the existing activities and services developed previously within the EC supported MyOcean FP7 project and EuroGOOS Regional Operational Oceanographic Systems (ROOSes). The In Situ TAC provides the interface between centres, distributing In Situ measurements from national and international observing systems.

It is integrated into a larger framework at European and International level and is developed with interoperability requirements with:

- JCOMM networks (Argo, OceanSITES, GOSUD, GTSPP, DBCP, EGO) and EuroGOOS ROOSes who operate the networks and provide access to the observations needed in Copernicus Marine Environment Monitoring Services
- Thematic centres which aggregate In Situ observation data for specific purposes. The main one for Copernicus In Situ TAC is SeaDataNet who integrates the networks of NODCs that manage historical scientific European data
- Downstream services that provide additional services on INSTAC data jointly with other datasets. One important downstream service for the In Situ TAC is EMODnet-Physics.

[15] Copernicus was previously known as GMES (Global Monitoring for Environment and Security)

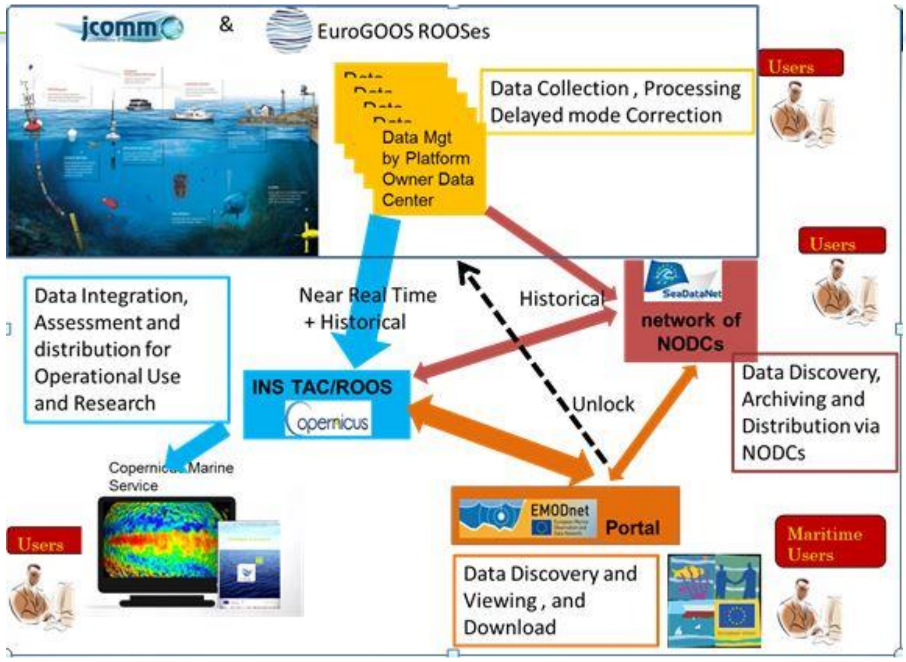


Figure 2: links between INSTAC
and other European initiatives dedicated to ocean observation data management

Copernicus In Situ TAC (INSTAC) consolidates the global and regional components, based on expertise developed within Copernicus and the ROOSes, and to develop the setup of bio-geochemical and waves in the In Situ TAC. In addition, considerable benefits to the Global Monitoring for Environment and Security (GMES/Copernicus) In Situ users will be gained in terms of In Situ product choice, service, timeliness, quality, robustness and accuracy.

As an operational infrastructure, the In Situ TAC sets the necessary production capacities and quality control procedures in answer to Europe’s request for service level agreements with the external users as defined in Copernicus. In addition, In Situ TAC has to provide In Situ data to internal users, the Global and regional MFCs.[16]

This document is describing the near-time data flow of ocean observation data.

[16] Source CMEMS INSTAC System Requirements Document DOI : 10.13155/40846

15. Near Real Time data organisation

Copernicus INSTAC is a decentralized data structure based on national or institutional data centres, productions and distribution units[17]

National or Institutional Data centres: a national or institutional data centre is responsible for assembling data performed by a set of observing systems. The centre collects, controls and distributes data according to its own rules. It is an outside interface for Copernicus. There are either international programs (Argo, GOSUD, OceanSITES, GTS…) or ROOSes data centres or SeaDataNet National Data Centres (NODC) for historical data.

Production Units: a Production Unit (PU) is responsible of assembling data provided by national or institutional data centres into an integrated dataset. The PU collects, controls data according to Copernicus In Situ TAC agreed rules and validates the dataset consistency in its area of responsibility. The validation activity can be managed by a different institute according to the regional INS TAC organization

Distribution Units: a Distribution Unit (DU) is responsible for assembling data provided its region Production Units into an integrated dataset. The Distribution Unit collects, distributes data according to Copernicus In Situ TAC agreed rules in its area of responsibility. The Global PUs assemble data from the international networks (JCOMM) observation for the global ocean.

The Global Distribution Unit: a Global Distribution Unit collects data distributed by Regional Distribution Units for the European seas complemented with Global PUs. The Global DU acts as a backup for the Regional DUs These aggregated products are distributed to the Global MFC, to Copernicus users working at global scale and as a backup to regional MFCs.

The External Users: an external user access to the INS TAC products using the MFTP and WMS services set up by the INSTAC through the Copernicus CIS. For example the ROOSes and EMODnet-Physics have developed customized viewing service on INS TAC DUs.

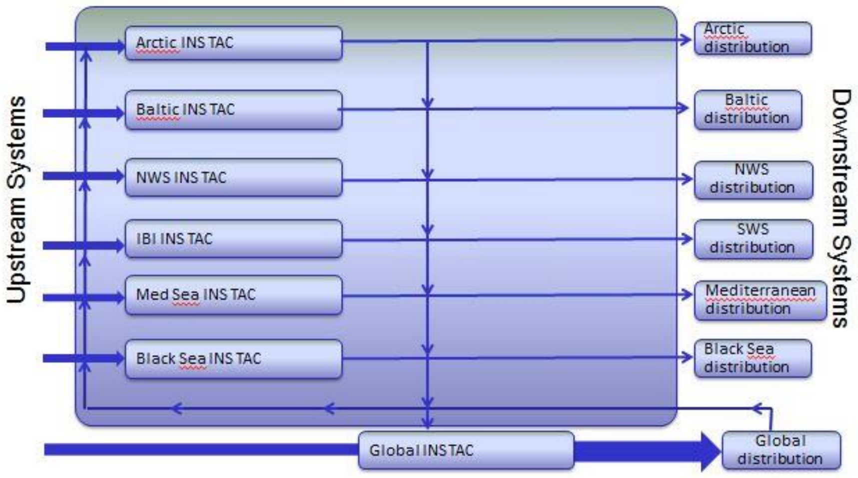


Figure 3: INSTAC as a decentralized structure

Where the different components are split according to the EuroGOOS / ROOS region areas

[17] Source CMEMS INSTAC System Requirements Document DOI: 10.13155/40846

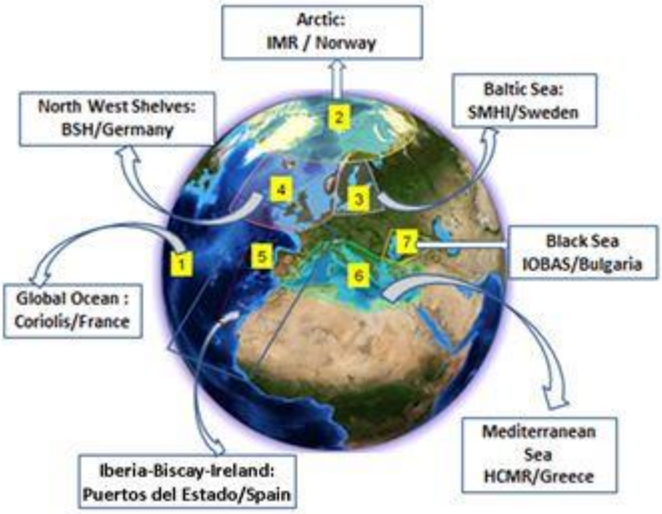


Figure 4: ROOSes region distribution

The region split is detailed below

INSTAC areas limits				
Distribution Unit	**Eastward**	**Westward**	**Northward**	**Southward**
Arctic	-180	180	90	60
Baltic 1	8	15	62	53
Baltic 2	15	31	66	53
Black sea	26	42	47.5	40
Global	-180	180	90	-90
IBI - Iberia Biscay Ireland	-50	9	60	19
Mediterranean 1	-5.61	37	41	28
Mediterranean 2	0	20	45.8	41
North West Shelf	-45	10	71.5	48

Also available on http://www.seanoe.org/data/00333/44395/ through a kml (Google map format) file

16. Near Real Time data definition and standards

Near real-time data may have several meanings:

- data that circulate from the originator to the data centre from a few hours to no later than 30 days after data collection (definition of the WMO –World Meteorological Organization)
- data acquired by continuous, automatic and permanent observation networks
- data that have been passed through an initial quality control check. Their quality can be later enhanced by using more accurate quality checks and and/or calibrations the data may be re-submitted as delayed mode data

16.1. Format

Copernicus INSTAC distributes data in NetCDF following the EuroGOOS recommendations and OceanSITES version 1.2 NetCDF conventions.

- *OceanSITES data management team (2010). OceanSITES User's Manual. NetCDF Conventions and Reference Tables.* *http://doi.org/10.13155/36148*

The specific Copernicus INSTAC adaptation of the NetCDF OceanSITES format is listed in chapter 7 of CMEMS INSTAC System Requirements Document DOI : 10.13155/40846

It is notable that for better recognition of the data sources or data providers, the SeaDataNet[18] station identifier has been added in the INSTAC NetCDF format as a global attribute. SeaDataNet is a data provider for Copernicus.

16.2. Parameters

The valid list of parameters is published in an Excel spreadsheet "In Situ TAC parameters list" available on the landing page of this document: http://doi.org/10.13155/40846

[18] SeaDataNet is the European project that federates the network of EU national oceanographic data centres. https://www.seadatanet.org/

17. Catalogue, Documentation and information

17.1. Catalogue of in-situ products

Mercator-Ocean maintains a global catalogue of CMEMS products. This includes the description of the prediction products elaborated by the Marine Forecasting Centres (MFCs) and the Ocean Observing Products elaborated by the Thematic Assembling Centres (TACs).
Consequently the in situ ocean observing products elaborated by INSTAC are described in this global catalogue:

- http://marine.copernicus.eu/services-portfolio/access-to-products/
- Select "product with depth level"

Figure 5: CMEMS catalogue with focus on situ products

The catalogue part dedicated to the in situ products is fed by the INSTAC

A paper document also describes all the CMEMS products:

- http://marine.copernicus.eu/wp-content/uploads/catalogue-cmems.pdf

The INSTAC range of products in continuously evolving:

- **Version 1:** temperature and salinity data
- **Version 2:** temperature, salinity and current data
- **Version 3 (present/ 2017):** Version 2 + parameters related to waves
- **Version 4 (2018)**: Version 3 + biogeochemical parameters

Product reference	Product Name	Area	Parameters
INSITU_GLO_NRT_OBSERVATIONS_013_030	GLOBAL OCEAN- IN-SITU NEAR-REAL-TIME OBSERVATIONS	Global Ocean	Physical
INSITU_GLO_TS_OA_NRT_OBSERVATIONS_013_002_A	GLOBAL OCEAN- REAL TIME IN-SITU OBSERVATIONS OBJECTIVE ANALYSIS	Global Ocean	Physical/ BGC/Waves
INSITU_ARC_NRT_OBSERVATIONS_013_031	ARCTIC OCEAN- IN SITU NEAR REAL TIME OBSERVATIONS	Arctic	Physical/ BGC/Waves
INSITU_BAL_NRT_OBSERVATIONS_013_032	BALTIC SEA- IN SITU NEAR REAL TIME OBSERVATIONS	Baltic	Physical/ BGC/Waves
INSITU_NWS_NRT_OBSERVATIONS_013_036	ATLANTIC- EUROPEAN NORTH WEST SHELF- OCEAN IN-SITU NEAR REAL TIME OBSERVATIONS	NWS	Physical/ BGC/Waves
NSITU_IBI_NRT_OBSERVATIONS_013_033	ATLANTIC IBERIAN BISCAY IRISH OCEAN- IN-SITU NEAR REAL TIME OBSERVATIONS	IBI	Physical/ BGC/Waves
INSITU_MED_NRT_OBSERVATIONS_013_035	MEDITERRANEAN SEA- IN-SITU NEAR REAL TIME OBSERVATIONS	Mediterranean Sea	Physical/ BGC/Waves
NSITU_BS_NRT_OBSERVATIONS_013_034	Black Sea- In-Situ Near Real Time Observations	Black Sea	Physical/ BGC/Waves

List of the INSTAC NRT V3 products (information updated July 2017)

17.2. Documentation

Each INSTAC product is documented with 2 types of documents:

- PUMs: Product User Manuals
- QUIDs: Quality Information Documents

Both types of documents have been drafted to fulfil the final user needs of information (User point of view).

PUMs describe the way to use the products:

- How to register?
- Where to reach the products?
- Format description
- Product specification
-

An example is given through the following links:

- http://cmems-resources.cls.fr/documents/PUM/CMEMS-INS-PUM-013.pdf
- http://dx.doi.org/10.13155/43494

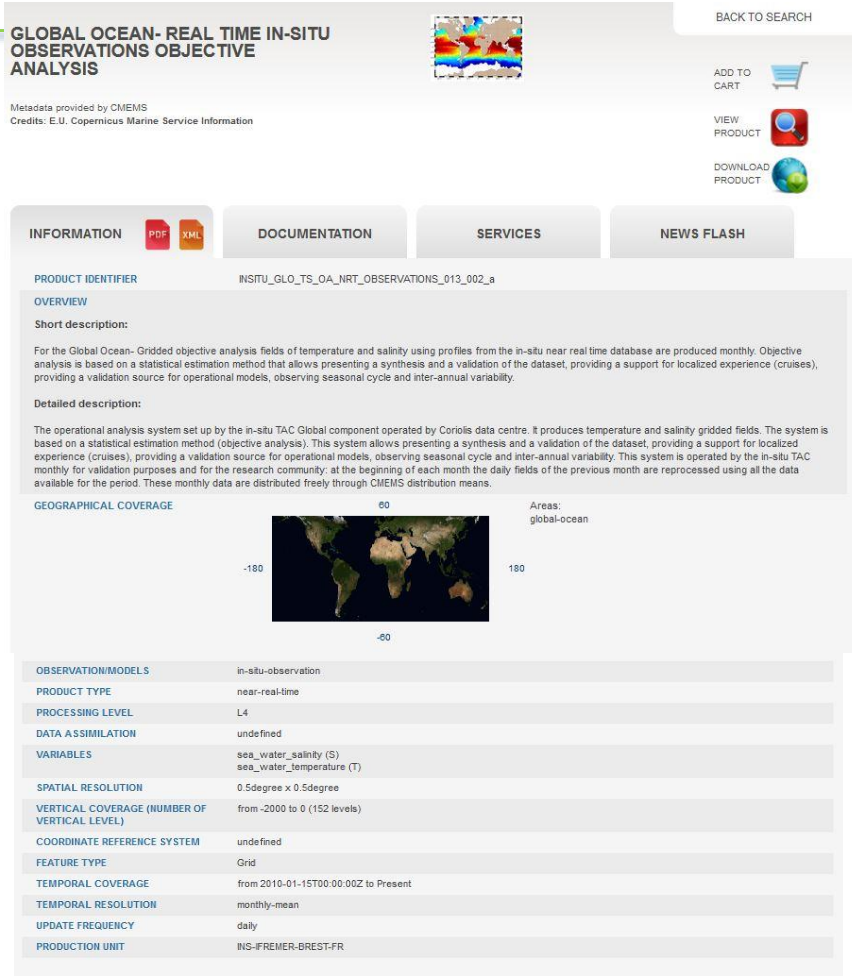
QUIDS describe the way the product has been elaborated:

- Elaboration method
- Quality control procedure
- Validation framework
- Validation results

An example is given through the following links:

- http://cmems-resources.cls.fr/documents/QUID/CMEMS-INS-QUID-013-030-036.pdf

All those documents are centralised through the CMES catalogue


GLOBAL OCEAN- REAL TIME IN-SITU OBSERVATIONS OBJECTIVE ANALYSIS

Metadata provided by CMEMS
Credits: E.U. Copernicus Marine Service Information

BACK TO SEARCH | ADD TO CART | VIEW PRODUCT | DOWNLOAD PRODUCT

INFORMATION | DOCUMENTATION | SERVICES | NEWS FLASH

PRODUCT IDENTIFIER	INSITU_GLO_TS_OA_NRT_OBSERVATIONS_013_002_a

OVERVIEW

Short description:

For the Global Ocean- Gridded objective analysis fields of temperature and salinity using profiles from the in-situ near real time database are produced monthly. Objective analysis is based on a statistical estimation method that allows presenting a synthesis and a validation of the dataset, providing a support for localized experience (cruises), providing a validation source for operational models, observing seasonal cycle and inter-annual variability.

Detailed description:

The operational analysis system set up by the in-situ TAC Global component operated by Coriolis data centre. It produces temperature and salinity gridded fields. The system is based on a statistical estimation method (objective analysis). This system allows presenting a synthesis and a validation of the dataset, providing a support for localized experience (cruises), providing a validation source for operational models, observing seasonal cycle and inter-annual variability. This system is operated by the in-situ TAC monthly for validation purposes and for the research community: at the beginning of each month the daily fields of the previous month are reprocessed using all the data available for the period. These monthly data are distributed freely through CMEMS distribution means.

GEOGRAPHICAL COVERAGE — 60, -180, 180, -60 — Areas: global-ocean

OBSERVATION/MODELS	in-situ-observation
PRODUCT TYPE	near-real-time
PROCESSING LEVEL	L4
DATA ASSIMILATION	undefined
VARIABLES	sea_water_salinity (S) sea_water_temperature (T)
SPATIAL RESOLUTION	0.5degree x 0.5degree
VERTICAL COVERAGE (NUMBER OF VERTICAL LEVEL)	from -2000 to 0 (152 levels)
COORDINATE REFERENCE SYSTEM	undefined
FEATURE TYPE	Grid
TEMPORAL COVERAGE	from 2010-01-15T00:00:00Z to Present
TEMPORAL RESOLUTION	monthly-mean
UPDATE FREQUENCY	daily
PRODUCTION UNIT	INS-IFREMER-BREST-FR

Figure 6: example of a INSTAC detailed product description

17.3. Quality control

The NRT quality control procedures rely on the "Recommendations for in-situ data Near Real Time Quality Control"[19] drafted by the EuroGOOS DATA-MEQ working group.
For each product the related Product Quality information document (QUID) the detailed quality controls steps that have been applied

- [19] EuroGOOS DATA-MEQ working group (2010). Recommendations for in-situ data Near Real Time Quality Control. http://dx.doi.org/10.13155/36230

18. Data Distribution

The main distribution channels for in situ data are ftp sites that have the same directory organisation for the global NRT product and the Regional products
This information is available on each NRT product Product User manual (PUM) and is recalled below

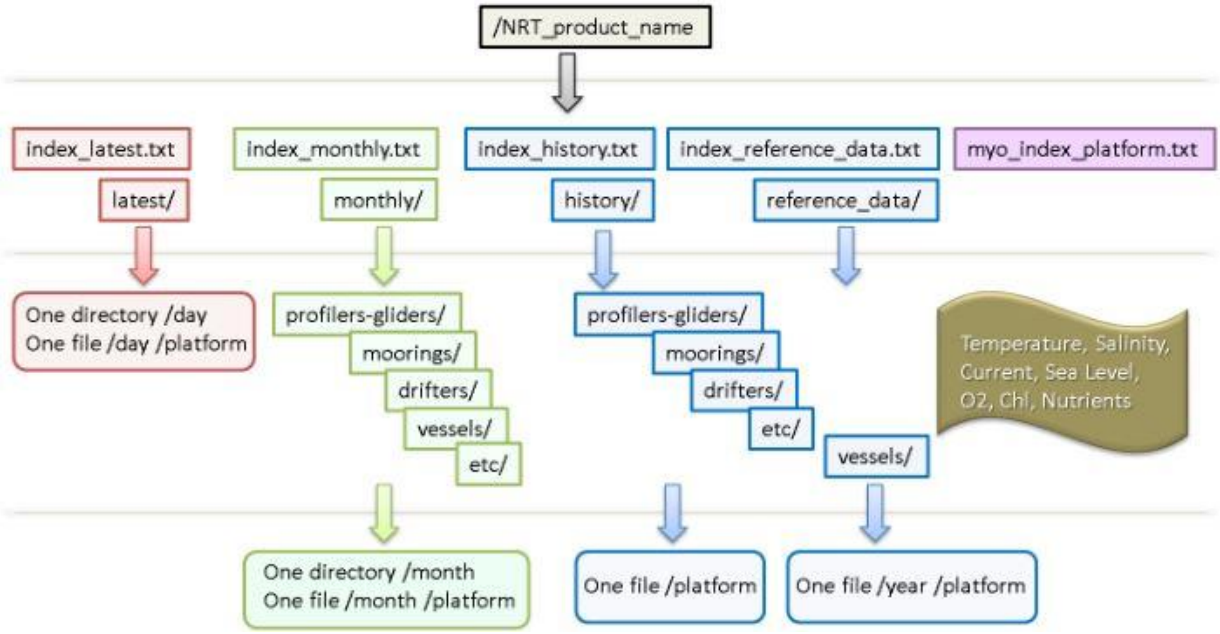


Figure 7: INSTAC NRT data distribution on ftp sites

For both regional and global ftp distributions,

- The latest_data directory is updated at least once a day
- The monthly directory is updated at least once a month
- The history and reference_data directories are updated at least once a year

In addition to the ftp data distribution, additional data distribution services are also provided
These services rely on tools provided by Copernicus Central Information Service (CIS).

- Oceanotron: The Copernicus server for In Situ data which provides web services such as:
  - OPeNDAP
  - OGC WMS mapping services
- MOTU: The Copernicus web server for data distribution
- THREDDS: The Unidata NetCDF web server for data distribution

19. Monitoring the system

This paragraph is not very detailed because, for the moment, most of the tools for monitoring the activity are not publically available. They are a key component for a (pre) operational service.

KPIs: Key Performance Indicators allow to monitor the INSTAC activity in terms of data flowing, data

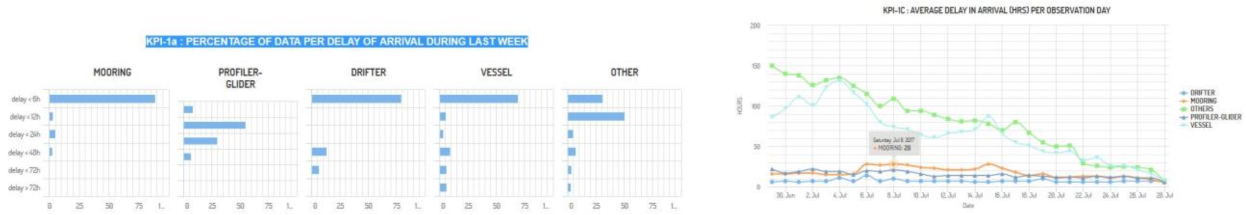


Figure: Examples of KPIs monitoring the INSTAC activity

With the same view of setting up a pre-operational service, it is important also to carefully monitor the availability of the different servers contributing to the whole system.

20. Service desk

As the INSTAC is run as a pre-operational system, a service desk has been set up. The service desk make the links:

- Between the global CMEMS service desk (hosted by Mercator)
- Between INSTAC external data providers and the INSTAC
- Between the INSTAC and the final data users

The address of the INSTAC service desk is cmems-service@ifremer.fr

21. Data submission

As described in paragraph 2, it is the responsibility of the regional Distribution Units to aggregate the data at the region level

Consequently the first contact must be taken at region levels

Service Desk	Email address	Operated by
Arctic sea	cmems-service@imr.no	IMR (Norway)
Baltic sea	cmems-service@smhi.se	SMHI (Sweden)
North West Shelf	cmems-service@bsh.de	BSH (Germany)
South West Shelf / Iberia-Biscay-Ireland	cmems-service@puertos.es	Puertos del Estado (Spain)
Mediterranean sea	cmems-service@hcmr.gr	HCMR (Greece)
Black sea	cmems-service@io-bas.bg	IO-BAS (Bulgaria)

And for global ocean observation network at

Service Desk	Email address	Operated by
Global	cmems-service@ifremer.fr	Ifremer (France)

It depends on what kind of platform the data is originated.
Depending on the platform type, the data are managed differently, therefore the methods depends on the kind of platform from which the data originates:

- Argo floats: it is the responsibility of the global in situ TAC (Coriolis)
- Ferryboxes: it is the responsibility of one of the Regional Centres
- Most other platforms: it is the responsibility of one of the Regional Centres
- Sea level and tide gauges: there are national coordinated actions but as yet no European approach
- Sea mammals data: there was no coordination at European level, a first effort for such a coordination has been recently implemented through the relevant Task Team of EuroGOOS

Therefore for most platforms, a new data provider will contact the regional data managers or the INSTAC service desk. Generally a new data provider will have to provide access to his data via an FTP server without changing his in house format as long as it contains enough metadata information to generate the standard NetCDF files. The basic information in this metadata is generally: what (platform name, institution, WMO number when available, type of platform, etc.), where (space coverage), when (time coverage), who (data provider, contact, PI, Data Centre, etc.), how (update interval, QC information), in addition some per platform specifics may be recommended (see details below).

21.1. Summary of data submission, by platform, for near Real-Time data

Near real time data				
Platform type	**Submit to**	**Data format**	**Contact**	
Argo Floats	INSTAC Global Distribution Unit which will make the contact with the Argo DACs	Argo NetCDF (note metadata must be submitted prior to data)	codac@ifremer.fr or cmems-service@ifremer.fr	
Research Vessels CTD	INSTAC Regional Distribution Units	.cnv	See table above for regional contacts	
XBT Data	INSTAC Regional Distribution Units	.edf or Devil NetCDF	See table above for regional contacts	
Gliders	INSTAC Global Distribution Unit	data in EGO NetCDF v1.1 (possibly also .json metadata only)	codac@ifremer.fr or cmems-service@ifremer.fr	
Near real time data				
Platform type	**Submit to**	**Data format**	**Contact**	
Moorings	INSTAC Regional Distribution Units	NetCDF Oceansites format (also Medatlas format in some cases)	See table above for regional contacts	Oceansites format (prefered solution) http://www.oceansites.org/docs/ oceansites_user_manual_version1.2.pdf
Surface Drifters	Most data are transmitted through Argos to CLS (France) and then to GTS. Some data are transmitted with Iridium	Improved GDP standards	codac@ifremer.fr or cmems-service@ifremer.fr	Two Surface Drifter GDACS are on the way to be established

	and then to GTS			
Tide Gauges / Sea Level	Not yet implemented in an operational mode	Not yet implemented in an operational mode	Not yet implemented in an operational mode	The tide gauges data acquisition and dissemination activities are coordinated within theGlobal Sea Level Observing System (GLOSS) –project (see http://www.gloss-sealevel.org/ The data repository is maintained at the PSMSL (Permanent Service for Mean Sea Level- http://www.psmsl.org/
Thermosalinograph	GOSUD	Colcor ASCII format or GOSUD V3.0 NetCDF format	gosudcontact@listes.ifremer.fr or codac@ifremer.fr	
Ferry Box	NIVA Norway (to mid 2015, MyOcean project period)		Pierre Jacquard (NIVA)	
Fishery Observing Systems				Not yet implemented in an operational mode
Near real time data				
Platform type	**Submit to**	**Data format**	**Contact**	
HF Radar				Not yet implemented in an operational mode
Sea Mammals				Not yet implemented in an operational mode
Other platforms			cmems-service@ifremer.fr	
Tide Gauge	National Oceanographic Data Centre (NODC) or direct to SeaDataNet	TBC	TBC	

22. Conclusion

The area of operational oceanography is evolving quickly and continuously. Consequently, the data management dedicated to in situ near real time should be adapted accordingly.
The next steps will have to take into account should involve new observation platforms such as tide gauges, HF radar, VM-ADC, etc., as well as to make available in operational mode new parameters (enhanced BioGeochemical data management).

Until now, the system is run in a pre-operational mode. The next step will be also to move to a fully operational system.

This document presented the state of the art in mid 2017

ACRONYMS

Acronym	Meaning
BAL	Baltic Sea
BGC	Biogeochemical
BS	Black Sea
CIS	Central Information System
CMEMS	Copernicus Marine Environment Monitoring Service
CTD	Conductivity temperature and depth
DATA-MEQ	EuroGOOS DATA Management, Exchange and Quality Working Group
DOI	Digital Object Identifier
DUs	Dissemination Units
EGO	Everyone's Gliding Observatories
FTP	File Transfer Protocol
GDAC	Global Data Assembly Centre
GIS	Geographic Information System
GLO	Global Ocean
GOSUD	Global Ocean Surface Underway Data
GTS	Global telecommunication system
IBI	Iberian-Biscay-Irish Seas
ICINGA	Server Monitoring Server
INSTAC	In Situ Thematic Assembly Centre
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
Med	Mediterranean Sea
MFC	Marine Forecasting Centre
MOTU	Motu is a high efficient and robust Web Server which fills the gap between heterogeneous Data Providers to End Users.
NAGIOS	Server monitoring Software
NetCDF	Network Common Data Form
NRT	Near real-time
NWS	Atlantic North West Shelf
OCEANOTRON	IFREMER tool in response to the problem of the multiplication of data formats. This server generates plugins that read different data formats, i.e. netCDF / OceanSites, SGBDR diagram and ODV binary format.
OceanSITES	Network of reference sites that measure a set of physical, biogeochemical and atmospheric parameters.
OpenDAP	Open-source Project for a Network Data Access Protocol
PU	Production Unit
PUM	Product User Manual
QC	Quality control
QUID	Quality Information Document
ROOSs	Regional Ocean Observing Systems
SD	Service Desk
SDN	SeaDataNet
TAC	Thematic Assembly Centre
Threds	
XBT	eXpendable BathyThermograph