Project	AtlantOS – 633211
Deliverable number	7.7
Deliverable title	Atlantic Ocean Data Integration
Description	Integration of EU and North American data for the Atlantic Ocean
Work Package number	7
Work Package title	Data flow and data integration
Lead beneficiary	UniHB
Lead authors	Ketil Koop-Jakobsen
Contributors	WP7
Submission data	31.03.2018
Due date	31.03.2018
Comments	N/A

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 633211.

Stakeholder engagement relating to this task*

WHO are your most important stakeholders?	☐ Private company If yes, is it an SME ☐ or a large company ☐? x National governmental body x International organization x NGO ☐ others EU NSF National Science foundations around the Atlantic …
WHERE is/are the company(ies) or organization(s) from?	☐ Your own country x Another country in the EU x Another country outside the EU Please name the country(ies): Countries surrounding the Atlantic Ocean
Is this deliverable a success story? If yes, why? **If not, why?**	x Yes, because we identified impediments for the cross continental collaboration on Atlantic Ocean data and initiated future action for improvement of data exchange. ☐ No, because …..
Will this deliverable be used? **If yes, who will use it?** **If not, why will it not be used?**	x Yes, a reference for future activities on international sharing for Atlantic Ocean Data. Initiation of new actions planned for Ocean Obs conference 2019. ☐ No, because …..

Contents

Executive summary

In order to explore ways for better Integration of EU and North American data for the Atlantic Ocean, AtlantOS WP7 organized a 2-day international workshop in Brussels on June 7-8, 2017. As AtlantOS has established collaborations with organizations around the entire Atlantic Ocean, we expanded the international outlook to also include South America and Africa in order to better cover the South Atlantic Ocean. The workshop had representation from all bordering continents, and provided an overview of the Atlantic Ocean data landscape identifying key-organizations working with Atlantic Ocean data. Current impediments of transcontinental data exchange and readiness to share data on each continent were discussed from a technical as well as a political viewpoint. Most importantly the workshop established personal contacts among people from different continents, which made the grounds for new collaborations, some which are initiated under AtlantOS.

Recommendations

The data land-scape exploration exercise performed at the workshop revealed that Europe and North America possess advanced infrastructures for the collection, archiving and dissemination of Atlantic Ocean data. In Africa, Ocean data collection, archiving and dissemination are established for the Southern part of Africa but is lacking in Northern Africa. In South America, there are multiple organizations concerned with collecting, archiving and disseminating data. However, open data sharing has not gained a wider acceptance.

The workshop discussions identified that improved transcontinental interoperability can be expanded by including best practices and metadata standards for QA/QC procedures, as well as by improving communication on existing standards and best practices for data and information exchange. Transcontinental harmonization of QA/QC–best practices is an obvious next step for a better interoperability on Atlantic Ocean data. This shall include standardization of metadata for selected variables improving data traceability (starting with platforms, ships and people). An initial step for introducing best practices could be to standardize the concept of uncertainty among data providers. This shall ensure that uncertainty estimates provided by data providers are in compliance with ratified standards. In regard to QC procedures; there is an obvious need to seek interoperable solutions for QC flagging. A future task shall be to identify mapping schemes, which can integrate commonly used flagging systems. This shall include analysis of existing flagging schemes, developing and managing mappings between them, and making this data available in a machine readable form.

Further international activities

Three topics were identified during the workshop as key areas where improved transatlantic collaboration could have a significant impact on data integration:

1. Ocean data - Quality assurance and quality control
2. Data standardization
3. Interoperability, semantics and machine learning from a user community perspective including new services that could emerge from big data technologies

Key persons were assigned with the task of establishing working groups to explore these topics and to produce inputs for AtlantOS blueprint and for white papers to be presented at the OceanObs'19-conference. Altogether 3 abstracts have been submitted to OceanObs'19-conference, which has ties to the discussion at the harmonization workshop were produced in truly international collaboration and with large number of the workshop participants as co-authors. In this way, AtlantOS WP7 has initiated discussions and collaborations on Atlantic Ocean Data Integration, which may extend long into the future.

Introduction

Between mid-2015 and mid-2017, AtlantOS WP7 “Data flow and data integration”, worked with data networks and data integrators in Europe to improve harmonization of Atlantic Ocean data, and facilitate improved interoperability. This resulted in agreement and implementation of key-metadata standards, for example for the use of identifiers for platforms and institutions, as well as common vocabularies and key quality control standards for data shared in near real time (within days or weeks).
In order to include the pool of data covering the whole Atlantic Ocean, it was important to include data collected from all continents bordering the Atlantic Ocean in our efforts to harmonize Atlantic Ocean data. In order to initiate this work, WP7 organized a 2-day workshop in June 2017 with key organizations for Atlantic Ocean data management from all continents bordering the Atlantic Ocean was organized (See agenda appendix 2). The workshop gave an overview of the data landscape, and identified key actions to improve transatlantic ocean data exchange. The workshop was intended as a stepping stone for the strategic planning of future actions to improve transatlantic data exchange across continents, and it succeeded to give an overview of the current status regarding Atlantic Ocean data handling (curation, handling and dissemination), and to identify and prioritize key actions for long-term and lasting improvement of transatlantic data exchange.

In order to assure the transcontinental aspect of the workshop, the scientific content was prepared by a panel with key data management organizations from all continents bordering the Atlantic Ocean. This panel selected two topics for improvement of trans-continental ocean data-exchange covering:
1) **Quality assurance and control:** Improvement of procedures and best practices; from sensor calibration at sea to sensor qualification at Data centers across continents, and analysis of harmonization perspectives.
2) **Data and metadata standardization:** Exploration of opportunities for improvement of transcontinental data-connectivity (cf. interoperability, standardization, accessibility) on a community level, emphasizing metadata assignment.

Following, the report give an overview of the conclusions and recommendations from the workshop as well as post workshop activities preparing new initiatives to be presented at the OceanObs’19 conference.

Atlantic Ocean data Landscape exercise

A prerequisite for improving Atlantic Ocean data integration among different continents is to have a clear map of the data landscape. At the workshop representatives from each continent were assigned the task of presenting the data landscape from their respective continent. A full overview of the Atlantic Ocean data landscape exercise is found in appendix 1.

The data landscape exploration exercise revealed that Atlantic ocean monitoring and research would benefit from an improved transcontinental data management effort including all continents bordering the Atlantic Ocean. Europe and North America possess advanced infrastructures for the collection, archiving and dissemination of Atlantic Ocean data.

In Africa, Ocean data collection, archiving and dissemination are established for the Southern part of Africa. There, Atlantic Ocean data handling is conducted by the Southern African Data Center for Oceanography. However, a centralized transnational data center is missing for Northern Africa. Hence, sustainable archiving and dissemination of Atlantic Ocean data is weak in this area (Gap identified).

In South America, there are multiple organizations concerned with collecting, archiving and disseminating data, with GOOSbrazil in a leading role. However, the movement towards open data sharing, which has gained a foothold in Europe and North America, has not gained a wider acceptance in South America. Hence, there is a cultural barrier (culture restricting access to data), which must be changed in order to improve transcontinental access to data from the South American continent.

Recommendations on harmonization of Quality assurance and control

The following is a selected summary of the recommendations from the discussion at the workshop. Providing transparent and dependable error and uncertainty information for ocean data is a necessity for the generation of reliable data products at the observation network level. Consequently, there is a strong need for harmonization of best practices of quality assurance and quality control procedures in a transcontinental context.

Standardization of QA/QC

Achieving even a basic level of standardization on quality assurance and quality control will be a challenging task, as it will require harmonization of existing protocols and workflows, which are already in use at the instrumentation level. Hence, there is no clear way forward with obvious low hanging fruits. However, the benefits of consolidating best practices to include a level of confidence to be delivered with the data, at least at an EOV level, are obvious. An initial step for introducing best practices could be to standardize the concept of uncertainty among data providers, assuring that the uncertainty estimates provided are in compliance with ratified standards such as ISO, BIPM, and any standardization body of relevance. Additional best practices shall be concerned with use of international vocabulary for metrology. Furthermore, the formulation of a guide to the expression of uncertainty in measurement should be explored, as well as the use of GEO QA4EO, as it defines a quality assurance framework for earth observation.

Traceability of the QA

Best practice guidelines shall also include aspects of traceability of the QA procedures applied. There is a need to know how the data is acquired, therefore, we shall encourage people to know how the data is collected and processed in the QA workflow. The sharing of such information is of primary importance for the platform operators, as well as the data users. Hence, an equivalent responsibility exists for data users to obtain this information, so that data are properly utilized. This is essential for the data stream. Improvement of traceability shall also include going away from free text formatting to a more precise formatting. For example, standard entries for steps taken with regard to quality assurance, such as the frequency of calibrating and preparation of instruments. This kind of information does not necessarily need to be shared on every occasion, but must be traceable and available, when needed. To some extent the improvement of QA traceability is

already ongoing through the development of "smart sensors", which provide this information directly from the sensor along with the data stream.

Even though the sharing of QA information was identified as being essential for a reliable data product, it is also recognized that implementing the necessary procedures for recording, organizing and sharing QA information requests are cumbersome. Consequently, having the capability to share information and then share it with other organizations is often neither prioritized nor used on a regular basis. To the extent that QA information is available for sharing, they most often come in different standards. Hence, integrating QA information from multiple sources has the potential to be time consuming and complex. Considering the current status of QA handling among data providers, as a best practice, we currently recommend that sharing upon request, and upholding preservation of information, has a higher priority, than trying to harmonize all the information.

The data sharing includes sharing of original data (raw data) as well as corrected data, which has been through Quality control procedures. From the QARTOD perspective, it is recommended that data flagged as bad or questionable be retained, in order to be able to redo the QC processing a few years afterwards, if needed, and to be able to improve the QC procedures in hindsight over time.

Recommending critical metadata and data

In order to facilitate the implementation of better and standardized procedures for QA/QC information, it is important that standardized metadata for the measuring system is also provided. We recommend, as a minimum requirement, standardization of the following metadata; unit ID for the sensing unit, and standard vocabulary for platforms, PID and standard vocabulary for responsible institutions, and PID and standard vocabulary for people involved. AtlantOS WP7 has worked on implementing metadata standards (PID and Vocabulary) for platforms and institutions, and future work can be built on the experience from AtlantOS WP7.

Regarding QC metadata from platforms, it should be taken into account that requirements are different for different kinds of platforms: For example; autonomous platforms with continuous automated data transmission, and autonomous platforms, which must be recovered in order to retrieve data. In particular, metadata on operator and host of platforms, instrumentation selection, and maintenance procedures, are parameters that are currently challenging in an international perspective. For these parameters metadata should be standardized. QC metadata from research vessels is in a different category, and can contain significantly more metadata.

There has been a revision of standards conducted by the EuroGOOS DATAMEQ working group, (http://eurogoos.eu/data-management-exchange-quality-working-group-data-meq), but opportunities shall be explored to harmonize this concept with similar guidelines on both sides of the Atlantic Ocean. For the most part, standards are implemented within a country, or in some cases within a continent, this is the reason there is a need for letting cross-continental agreements come into focus. In this regard, it is important to emphasize that the activities shall also include data from South African or South American data providers and repositories.
Communication and direction for data providers as well as users is a critical point; users need a focal point for a central resource to assist in guiding them through the available data providing systems. It is important to break the learning barriers through increased training for providers and users. One way by which we can facilitate an easier approach to metadata on QC is through

mapping and implementation of a centralized external vocabulary. A repository for documentation of standards at IODE or BODC could be a starting point, not just for one region, but for the whole world. Regarding funding such an initiative; typically, the observation networks or individual nations that have interest in maintaining a minimum quality of observational data, already provide funds for these activities. Hence, the expansion to an internationally recognized system should not require major investments on changing infrastructures. Furthermore, it is also necessary to educate the community of data-providers on standards, how to apply them, and the value thereof.

Interoperability – a term with many definitions

Improvement of interoperability is an essential task often discussed among data providers, data integrators, and data users. However, interoperability can be discussed at many different levels and in many different contexts; like syntactic and semantic interoperability. Consequently, there is not only one definition for interoperability; it is context dependent. Information on context level, and scope is needed in order to define the meaning of interoperability in each case. This provides a significant challenge for the overall discussion of "how to improve interoperability". Hence, working towards a common internationally accepted definition would be preferable, but most likely not doable. A good starting point could be the definition provided in the FAIR principles, which have been adapted in Europe under the H2020 program. The FAIR principles define interoperability, as the ability of a system or a product to work with other systems or products without special effort on the part of the user. Furthermore, the FAIR principles set up a thorough list of requirements, which shall ensure interoperability. Other international organizations have gone through the same exercise of defining interoperability; within ODIP, IOOS & QARTOD the definition of interoperability follow the same mind-set, where the current definition is "The ability of two or more systems to exchange and mutually use data, metadata, information, or system parameters using established protocols or standards." Hence, overall definitions of interoperability is being discussed at a higher level, however, it does not prevent the challenges of different interpretations, when interoperability is discussed for a specific task. Consequently, interoperability must always be clarified by the context of the task at hand, and in regard to quality control, it may be beneficial to define interoperability in a QA/QC–context.

Consolidating Real-Time QC recommendations among Atlantic partners

Many basic physicochemical ocean data are recorded in real-time in the Atlantic Ocean by institutions and governments from all continents bordering the Atlantic Ocean. The basic technical approaches to these measurements are very similar across continents. Hence, the adaption of a more broadly accepted set of minimum requirements for RTQC for specific variables should be an achievable task. The first step shall be the exploration of opportunities on how to harmonize existing recommendations. This includes identifying commonalities as well as differences in the QC protocols, and their implementation across the Atlantic. For this purpose, Dissolved Oxygen (DO) is a good starting point as there are groups on both sides of the Atlantic working on updating DO QC manuals including QARTOD QC, which manuals are updated every 2-3 years.
For future work on RTQC, a priority list of EOVs shall be produced, where the prioritization is based on importance of the variable as well as the ease by which a transatlantic common standardization can be achieved. In this regard, it would be useful, prior to a prioritization of EOVs, to identify how views and prioritization of EOVs differs between various scientific communities.

Common Flagging system

Flagging and mapping of flags is a key instrument to make life easier and simpler for the users of ocean data, and inform them on the quality of the data they are using. Many differing flagging standards are already entrenched. Hence, finding a compromise with one standardized flagging system that accommodates all the flagging system already in use is challenging. This challenge includes both data providers as well as data users. Making users who are used to one flagging system, convert to another one, will almost never work, because they will use something else. Mapping between flagging systems is a solution, which can accommodate multiple flagging systems. However, in this case another aspect of the flagging system to be identified is then the level of information that should be provided by a flagging system, whether it only flags obviously erroneous data, or whether questionable data should also be flagged. In conclusion, there is an obvious need to seek interoperable solutions for QC flagging. A future task shall be to identify mapping schemes, which can integrate commonly used flagging systems

Cross network harmonization

As different observing networks in many cases have developed their own recommendations on various QA/QC procedures, implementing new recommendations across networks is challenging. The first step is to identify a sustainable effort for the future of the harmonization process. AtlantOS can be used as a test-bed for ideas and some demonstrators, but it does not solve the need for a long-term solution. Same conditions apply for ODIP.
Consequently, we should aim to work in an Atlantic context rather than in an AtlantOS context. Other frameworks for continuing the work should be explored. This can include Blue Cloud, Blue Planet, OES and RDA and may involve operational programs such as GOOS.

In general, we should aim for recommending a comprehensive description of the workflows on how observational data are generated and processed. This includes all instrumental and deployment aspects, the calibration procedures and history, and how corrections have been applied.

Recommendations on harmonization of transcontinental data-connectivity

The following is a selected summary of the conclusions and recommendations from the discussion at the workshop on the improvement of data interoperability, standardization and accessibility.

There is a big push from policy makers to move to open data policies for data observation acquired with public funds. Many funding schemes currently facilitate open access by making open data access a requirement for funding projects. This poses a big change in some communities that still consider that they own the data and decide when and how they will share them. It is also an issue with the private sector even if there is some experience like the SIMORC project at EU-level that share data with some constraints.

Reasons for implementing restricted access to data

In general, there is still a need for future discussions on data ownership in regard to accessibility. There are many reasons for not sharing the data:

1) Sharing is restricted due to Quality concerns: Quality concerns is an important aspect, where the data provider want to reach a reasonable level of certainty in regards to the data quality, before the data is made openly accessible. This is particularly important in regards to sharing of real time data, where there is limited time for intelligent quality control measures. Flagging algorithms can be applied, leaving it up to the user to apply a layer of intelligent quality control. Hence, it is of great importance that 1) the data carries a sufficient metadata on the level of processing and 2) training and informing of users on the interpretation of metadata, and further processing of data for real time and delayed mode.

2) Sharing can be restricted in order to secure first priority to publish the data. This is a concern, which may decline in the future, as requirements for dissemination of research projects move towards showing not only publications, but also evidence of re-use of the data in a broader context.

3) Incentive to share is lost due to lack of knowledge on data publication. The incentive to share data can easily be lost, if the data provider lacks knowledge of appropriate data publishers and/or data networks to collaborate with. In this regard, the investment needed to transform their data to fit the requirements is a concern. There have been some success stories on initiatives to improve data publication services in both USA and Europe. Such services facilitate the contact between the data provider and the appropriate data archiving network/publisher. Furthermore, these services assist data providers with their obligations to make their data available for use through appreciation documentation. These services can become an important means to improve the open accessibility to Atlantic Ocean data from South America or Africa. In this regard, there is a need for certified repositories to assure standardization. A lot of funding agencies in Europe are looking at certified repositories, where they are encouraging people to submit their data. Certified repositories have to go through an approval process to demonstrate they have a certain level of best practices.

4) Incentive to share openly is lost as it complicates traceability of data use. Many research platforms are often judged based on the use of the data produced with users/publications, therefore the ability to tightly trace the data to assure platform accreditation is essential. Consequently, there is a need for improving traceability of data use. This can be achieved by standardization of metadata for the data providers, regardless of whether it is a permanent stationary autonomous platform, or a temporarily deployed moving platform. Activities on setting such capabilities should be addressed at a trans-continental level.

5) Concerns regarding misinterpretation and misuse. Although higher level authorities want the publicly funded data to be released, the open data licenses allow users to do what they want with the data. Consequently, some researchers producing the data are concerned that open and unrestricted access may result in misinterpretations and/or misuse of the data.

Improvement of data citation

There has been a lot of discussion on data citation, and how to define a strategy that satisfies the need of the various data network and data integrator. DOI assignment to data-set is widespread now-a-days, but the strategies for DOI assignment varies among data networks and integrators. In particular, the assignment of DOIs to open-ended data time-series made accessible in near real-time pose a significant challenge.

In general, the necessary technological tools are available, and DOIs are manageable on both the data provider and the data user level. However, there is no commonly accepted best practice on the temporal assignment of DOIs to continuous data series. Furthermore, the same situation is present in regard to accrediting the people behind the data. From a technological side, ORCIDs are becoming more and more widespread as it is a PID for people. However, there is no strategy for how to connect Peoples IDs to the data-series. This is particularly noticeable with autonomous platforms operated by governmental monitoring programs, where it is difficult to determine who the accreditable personnel are. The data is usually handled at various levels, by technical personal, responsible for the daily operations, as well as scientific personnel, making decisions on data validity, and application of QC procedures, and the funding is provided by an overarching governmental monitory agency. This poses a challenge for the accreditation of people to data-series, and there is currently no best practise for this task.

Getting the users interested in the improved technological solutions available for data discovery

Ocean Data is available to serve academics as well as commercial interests. A better knowledge of the user-community is needed in order to accommodate their needs. A strategy must be developed for how to accommodate both academic and scientific users. It is of great importance that the data providers and/or data integrators are able to reach their widespread user community and inform them on the latest developments. Webinars and catchy web videos is one solution.

Commercial search engines, such as Google, will probably begin to include scientific data such as ocean data with in a foreseeable future. In order to guarantee the best possible use of the scientific data available, it is important to assure that advice provided to commercial companies direct them to the best certified services/products/search engines. This is best done in a collaborative effort.

Minimizing the risk of duplication of data

Duplication of data sets via the internet, as they are picked up by various search engines, scientific as well as commercial, is a general challenge prompted by increase open data accessibility. As big data technologies improve interoperability between systems, and assure the linkage of products to the original data, the challenge regarding duplicates will probably decrease. This highlights the importance of assigning persistent unique IDs for data, right from first version provided by the data-provider. This shall further be improved in the near future as the development of smart sensors will facilitate the automated assignment of PIDs to the data-sets produced.

Further international activities

White papers for OceanObs'19 conference and inputs for AtlantOS blueprint

Three topics were identified as key areas where improved transatlantic collaboration could have a significant impact on data integration:

1. Ocean data - Quality assurance and quality control
2. Data standardization
3. Interoperability, semantics and machine learning from a user community perspective including new services that could emerge from big data technologies

Key persons were assigned with the task of establishing working groups to explore these topics and to produce inputs for the AtlantOS blueprint and for white papers to be presented at the OceanObs'19 conference.

The working groups succeeded and altogether 3 abstracts were submitted mid-March 2018 to OceanObs19 conference, which has ties to the discussion at the harmonization workshop. These are abstracts on 1) Quality Assurance and Real-Time Quality Control of Ocean Data 2) Data standards from sensors to the web 3) Development of integrated observation data services in Europe based on enhanced data system of systems implementing FAIR principles. The Abstracts were produced by workshop participants as lead-PIs in a truly international collaboration and with large number of the workshop participants as co-authors.

Working group on for Quality assurance and quality control

The working group on Quality assurance and quality control has established itself as an active international forum and is currently working and very active at a global level with representation from many of the key international organizations working on ocean data quality control.
The working group on Ocean data – Quality assurance and quality control was established under the leadership of Mark Bushnell, IOOS. The Scope is to establish a globally harmonized QA/QC framework linking to ongoing initiatives and making use of existing reference documents developed in relevant projects (FIXO3, QARTOD, Neptune Canada, OTN, ,ENVRI+, EMSODEV, etc.) as well as the GOOS best practices repository and the QARTOD program of IOOS GOOS Bio/Eco panel ISO initiative.

The goal of this working group is to establish guidelines for providing transparent and dependable error and uncertainty information for ocean data, which is a necessity for the generation of reliable data products. In general, there is a strong need for harmonization of best practices of quality assurance and quality control procedures in a transcontinental context. These goals shall be reached through Standardization of QA/QC procedures. Achieving even a basic level of standardization on quality assurance and quality control requires harmonization of existing protocols and workflows, which are already in use at the instrumentation level. The benefits of consolidating best practices to include a level of confidence to be delivered with the data, at least at an EOV level, are obvious.

The starting point of the working group will be the introduction of best practices by standardizing the concept of uncertainty among data providers. This will assure that the uncertainty estimates provided are in compliance with ratified standards such as ISO, BIPM, and any standardization body of relevance. The best practices shall be concerned with use of international vocabulary for metrology, and will include the formulation of a QA manual to the expression of uncertainty in measurement, as well as the use of GEO QA4EO, as it defines a quality assurance framework for earth observation. The work started on October 1st 2017 and presented and discussed ideas for QA manual at the Fall AGU meeting 2017.

Appendix 1: Overview of the data landscape on the 4 continents bordering the Atlantic Ocean

Europe

The European data landscape section is based primarily on a talk given by Sylvie Pouliquen (Ifremer) and information available from webpages of the data centers/programmes/projects mentioned.

The operators of the Observing Systems

Data management is designed by the networks, organized within a European (EuroGOOS http://eurogoos.eu/) or International context (JCOMM http://www.jcomm.info/). Data Systems are targeted to the specific platforms and fit with the Network community needs in particular for Quality control. Nearly 100 institutes coordinate their activities through EuroGOOS, and the Regional coordination (ROOS), which operates the observing systems and process their data for their national needs with a regional coordination.

European Integrated data Systems

SeaDataNet (https://www.seadatanet.org/) is a pan-European infrastructure set up and operated for managing marine and ocean data in cooperation with the NODCs and data focal points of 34 countries bordering the European seas. SeaDataNet provides various metadata directories essential for international standardization of ocean data. An upgraded architecture is planned for the near future and will be concerned with i.a. data replication and Virtual research environments (VREs) in the cloud.

Copernicus, the European earth observation programme, is aimed at developing European information services based on **satellite Earth Observation and in situ ocean data.** The Copernicus environmental monitoring service (CMEMS) (http://www.copernicus.eu/main/marine-monitoring) is a world-leading marine environment and monitoring service focused on global and European regional marine products. The CMEMS provides regular and systematic core reference information on the state of the physical oceans and regional seas. The observations and forecasts produced by the service support all marine applications, this also includes core services for maritime safety, effective use of marine resources, healthy waters, informing coastal and marine hazard services, and supporting climate services. The products delivered by the Copernicus marine environment monitoring service are provided free of charge to registered users through an Interactive Catalogue available on the marine.copernicus.eu web portal.

EMODnet (http://www.emodnet.eu/) is the European Marine Observation and Data network. It is a long term marine data initiative from the European Commission Directorate- General for Maritime Affairs and Fisheries (DG MARE). More than 100 organizations assembling marine data, products and metadata contribute to EMODnet. The goal is to facilitate availability of marine resources to public and private users. EMODnet is concerned with quality-assurance, standardization and harmonization of marine data, which shall assure full interoperability and accessibility to European Marine data.

Connectivity between operators of the Observing Systems and Integrated data Systems

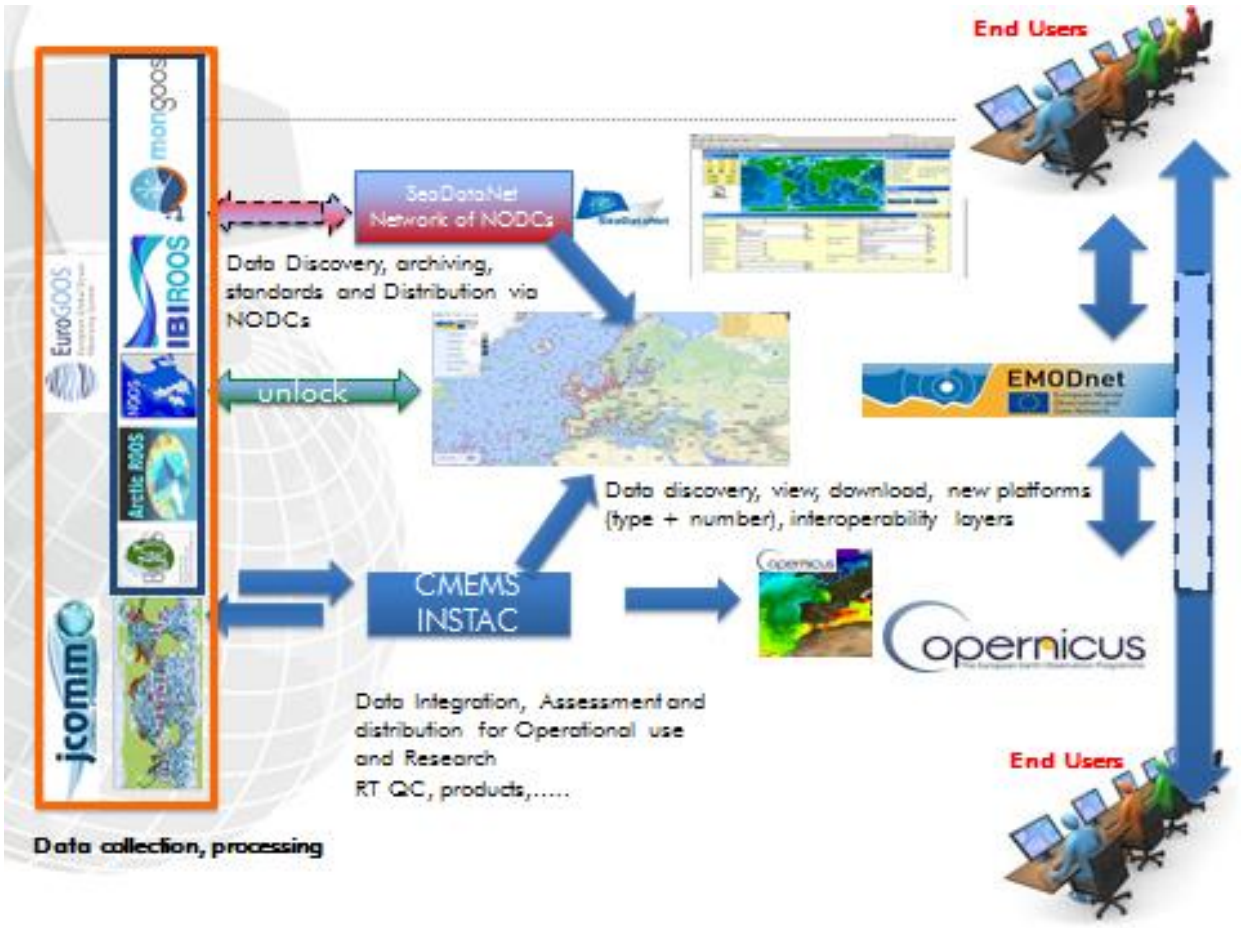

Figure 1: Overview of essential pathways for data management in Europe

Within the past 15 years, Interoperability between the different components of the European Data system has been coordinated through the DATAMEQ (DATA Management Exchange and Quality) working group set up by EuroGOOS. It gathers networks representatives from JCOMM and EuroGOOS ROOSes, main data integrator infrastructure representatives (CMEMS, SeaDataNet, EMODNet) and representatives of EU projects that have a significant data integration activity (ODIP, AtlantOS, Jerico-next, ENVRI+..) and meet every 2 years . A strategy has been defined to connect and facilitate data exchanges between the different components and promote Open Data principles and International Standards when available, taking into account existing best practises whenever possible . Based on the networks/integrators/projects progress, deliverable from projects are synthesised and endorsed as recommendations at the annual assembly and made available at http://eurogoos.eu/data-management-exchange-quality-working-group-data-meq/ .

Therefore an efficient data flow and interfaces has been set up between these elements and we are working together to reduce the identified weaknesses and propose solutions to the different actors. Deliverables from AtlantOS (Data Harmonisation, NRT QC recommendations, Data Management Handbook and Full Life Cycle) will be within a year endorsed by the EuroGOOS Assembly. And serve as starting documents for other initiatives and projects.

Current status and future initiatives:

Europe has well-developed infrastructures to accommodate the ever growing amount of data from the Atlantic Ocean within an Open Data Policy pushed both at National and European level in many domains. This is generated through a framework of various Research Infrastructures including Observation RIs and Integrating initiatives Copernicus, EMODNet, SeaDataNet in link with GOOS, IODE, GEO and GODAE Ocean View. It provides a sustainable framework for these RIs of utmost importance in order to assure the best possible use of the Atlantic Ocean data available. This shall also include the framework for continuous development and addition of new technologies and services; which includes Smart Sensors (SWE) and remote Cloud, VRE and HPC services.

North America

The North American data landscape section is based primarily on the talk given by Mark Bushnell (NOAA) and information available from webpages of the data centers/programmes/projects mentioned. The North American Ocean data landscape consists of multiple data programmes, centers and projects concerned with ocean data. There are 5 main players, IOOS, NOAA,NSF and CDIP in the US, and OSD in Canada. Each is responsible for the operations of various ocean data collecting programs. Selected essential programs are described below

IOOS

The Integrated Ocean Observing System (U.S. IOOS®) lead the integration of ocean observing capabilities in collaboration with Federal and non-Federal partners in order to maximize access to data and generation of information products.

The IOOS data catalog is an important tool for accessing data from key research and monitoring organizations in North America. The IOOS has data from various sources including *industry, academia, government and non-governmental organizations. The data and services provided are available to a wide range of users.* IOOS coordinates and provide the architecture for data integration from various sources in partnership with IOOS regions. The IOOS consist of 11 regional associations, 4 covering the Atlantic Ocean.

NOAA

The National Ocean and Atmospheric Administration (NOAA) administers several monitoring systems with connected data centers:

NOAA - Observing System Monitoring Center (OSMC) (http://www.osmc.noaa.gov/)

The Observing System Monitoring Center provides integrated access to ocean data from various platforms.

The OSMC is furthermore concerned with data and metadata standardization and best practices as well as the development of tools that promote increased discoverability, accessibility, usability and interoperability of diverse ocean/climate scientific data streams.

NOAA - The National data buoy center http://www.ndbc.noaa.gov/

The National Oceanic and Atmospheric Administration (NOAA) - National Data Buoy Center (NDBC) is a part of the National Weather Service (NWS). NDBC designs, develops, operates, and maintains a network of data collecting buoys and coastal stations.

NOAA - The Center for Operational Oceanographic Products and Services (CO-OPS) https://tidesandcurrents.noaa.gov

The CO-OPS provides data on water-level and currents from along the US coastline. The vision is to provide open and ready access to tide, water level, current and other coastal oceanographic information needed for informed decision-making. Data contributions go to GLOSS, an international programme concerned with collection of sea level monitoring globally.

NOAA - The Global Drifter Center (http://www.aoml.noaa.gov/phod/dac)

The Global drifter center provides Satellite-tracked surface drifting buoy observations from all Oceans

NOAA - National Centers for Environmental Information (https://www.ncei.noaa.gov/)

NOAA's National Centers for Environmental Information (NCEI) hosts and provides public access to one of the most comprehensive archives for environmental data, which includes atmospheric, coastal, oceanic and geophysical data.

NSF

The National Science Foundation has funded the development of EarthCube Integrative Activities.

NSF - OOI http://oceanobservatories.org/data-portal/

Ocean Observatories Initiative (OOI) is an integrated infrastructure project composed of science-driven platforms and sensor systems that measure physical, chemical, geological and biological

properties in the ocean. The OOI network operates on a global scale and has arrays in the North and South Atlantic Ocean. The OOI network was designed to address critical science-driven questions that will lead to a better understanding and management of our oceans. The data is available online without restrictions via the OOI-portal.

NSF - Seaview

The SeaView project is working on making better and centralized access to data from existing ocean data repositories. It is creating a federated collection of data sets that are organized around common scientific themes and published online by building connections between the data resources in five major oceanographic facilities: Biological and Chemical Oceanography Data Management Office (BCO-DMO), CLIVAR and Carbon Hydrographic Data Office (CCHDO), International Ocean Biogeographic Information System (OBIS) (external partner), Ocean Observatories Initiative (OOI) (external partner) and Rolling Deck to Repository (R2R) Program.

The Coastal Data Information Program (CDIP) (http://cdip.ucsd.edu)

The Coastal Data Information Program (CDIP) is an extensive network for monitoring waves and beaches along the coastlines of the United States. CDIP has a publically available database with data from 1975 and after. CDIP is operated by Scripps and contributes to IOOS

The Oceanography and Scientific Data (OSD)

OSD is the National Oceanographic Data Centre for Canada within the IODE, providing support for programs including Argo Canada (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/index-eng.html), Global Drifting Buoys (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/drib-bder/index-eng.htm), and the Atlantic Zone Monitoring Program (http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/azmp-pmza/index-eng.html) among other activities.

In Canada, the mandate of the Oceanography and Scientific Data (OSD) is to manage and archive ocean data collected by department of fisheries and Ocean or acquired through national and international programmes conducted in ocean areas adjacent to Canada, and to disseminate data, data products, and services to the marine community in accordance with the policies of the Department.

Ocean Networks Canada monitors the wet and east coast of Canada and the Arctic to continuously gather data in real-time for scientific research that helps communities, governments and industry make informed decisions about our future.e

Current status and future initiatives:

In North America, there are multiple organizations concerned with collecting archiving and disseminating ocean data, following international standards for data and metadata. In this way, the technical capability to curate and disseminate all available kinds of ocean data is present, and there is movement towards more open access to ocean data and in general where possible international standards and best practices are sought implemented. With the open data execute order and open data policy signed by the Obama administration in 2013, the US took a big step providing access to data generated by federal funds. The Government of Canada has undertaken a similar approach through directives intended to provide more easy and transparent access to data and information from federal government departments and agencies. The Government of Canada has established an Open Government portal (http://open.canada.ca/en/), Open Data portal (http://open.canada.ca/en/open-data), and an Open Government license (http://open.canada.ca/en/open-government-licence-canada).

In North America, many of the initiatives mentioned also work in an international setting to improve the current status of ocean data interoperability and data quality. Canada and the US collaborate very closely internally as well as with international communities, and in many cases contribute data to systems/programs already led by international partners.

A key initiative is Quality Assurance / Quality Control of Real Time Oceanographic Data (QARTOD), which strives to sustain a process for Ocean data. QARTOD shall develop authoritative QA/QC procedures for 26 of the IOOS variables, through the production of written manuals for these QA/QC procedures. Through this work, QARTOD shall facilitate QA/QC integration with the Global Ocean Observing System (GOOS) and other international ocean observation efforts including procedures for calibration and quality flagging.

South America

The South American data landscape section is based primarily on a talk given by Rafael Sperb, (FURG) and information available from webpages of the data centers/programmes/projects mentioned. In 1971 Brazil began operation of the Oceanographic National Database (BNDO) under the Directorate of Hydrography and Navigation (DHN). From 1988, it is a requirement in accordance with the Coastal shelf and exclusion zone that every research or activity that produces ocean related data must submit it to BNDO. The BNDO collect data from various sources. Data is kept in its original form, and there is currently no standardization of metadata or validation process installed. Data is available upon request and there is no public access.
Furthermore, DHN is responsible for facilitating the implementation of ocean observation programs in Brazil with institutions in Argentina and Uruguay, and provides support activities for GOOS Brazil. DHN also operates some ocean data collecting stations, some of which provide real-time data on various ocean variables.

GOOS-Brasil and included programs

GOOS-Brasil is an essential organization uniting 13 Key institutions involved with ocean data from the South Atlantic Ocean. *GOOS-Brasil* is a regional observing system under the Global Ocean Observing System (GOOS). *GOOS-Brasil* is an overarching organization coordinating ocean data collection from the following programs

The PIRATA program (*Prediction and Research Moored Array in the Tropical Atlantic* (PIRATA) is an essential project concerned with collecting ocean data in the tropical Atlantic Ocean. PIRATA is operated in an international context and is primarily involved with the collection of data concerning ocean-atmosphere interactions.

PNBOIA is the national buoy-program of Brazil operating along the Brazilian coast and in the South Atlantic Ocean.

GLOSS-Brazil is a regional entity of GLOSS, an international programme concerned with collection of sea level monitoring globally.

ARGO Brazil is a regional program of ARGO.

ANTARES is a program concerned with collection of data on chlorophyll concentration and sea surface temperature obtained by satellites for the oceanic regions adjacent to the South and Southeast coasts of Brazil. Other regional programs cover Argentina and Chile.

Rede Ondas is a network of buoys based in shallow waters along the Brazilian coast with the objective of providing real-time monitoring of wave-activity.

Movar is a program concerned with monitoring the Regional variability of heat transport in the surface layers of the South Atlantic Ocean between Rio de Janeiro and Ilha de Trindade. MOVAR is a pilot project in collaboration with NOAA.

Other programs

SIMCOSTA (http://www.simcosta.furg.br/) is the Brazilian Coastal Monitoring System project. SimCOSTA operates several coastal buoys providing real-time data measurement along the Brazilian coastline.

Gas and Oil Industry: The gas and oil industry possess a lot of oceanic data from along the Brazilian coastline. The Brazilian Data Center is the unit inside ANP responsible for storing all technical data acquired on petroleum exploration and production activities throughout Brazilian sedimentary basins. Primarily Seismic data and well data.

BAMPETRO is an Environmental Database for the Oil Industry. The database is specialized in storing, exchanging and disseminating georeferenced environmental information. It is composed of a Portal and a central database.
PREPS is a National Program for the Tracking of Fishing Vessels by Satellite.
SIMMAM - Support System for the Monitoring of Marine Mammals.

Biodiversity Data networks
For Data on South American ocean biodiversity and biogeographic data OBIS is the primary system used for data curation and dissemination. OBIS is the Ocean Biogeography Information System, which globally connects 500 institutions from 56 countries. Furthermore, some Ocean biodiversity data from Brazil can also be found through, the ***species*Link** network, which is used for data curation and dissemination of Brazilian Species data. *species*Link promotes free and open access to data, information, and tools available to any individual or group. However, quality control is not available and there may be restrictions to data usage.
Current status and future initiatives:
In South America, there are multiple organizations concerned with collecting archiving and disseminating data, following international standards for data and metadata, with GOOSBrazil in a leading role in terms of Observing System management. However, the movement towards open data sharing, which has gained a foothold in Europe and North America, has not gained a wider acceptance in South America. Hence, there is a cultural barrier (culture restricting access to data), which must be changed in order to improve trans-continental access to data from the South American continent. Overall, although the technical capability to handle ocean data is present, and knowledge on international standards and best practices to a wide extend is implemented, there is a lot of ocean data from South America, which does not become accessible to the larger international community as it is only stored locally often with restricted access and no discovery means. In general, there is limited willingness to make ocean data available from a scientific and economical point-of-view, and the accessibility to ocean data in South America has not been a matter of major concern.

Africa

The African data landscape section is based primarily on a talk given by Tammy Morris (SAECON) and information available from webpages of the data centers/programmes/projects mentioned.
Essential project concerned with Atlantic Ocean data
There are a few specific projects generating and disseminating Atlantic ocean data on a broader scale (Beyond regional). Essential data project includes **the EAF Nansen project** - which has collected data on sustainable fisheries, mapping of Oil, gas, pollutants and habitat covering mostly the pelagic ecosystem along the West African coastline. Bottom habitats are to be included in 2019. The South Atlantic MOC Basin-wide Array (**SAMBA**) has the main objective of investigating and monitoring of Indian-Atlantic Ocean exchange. Furthermore, the project possess a series of historical data from 1992 onwards.
The PIRATA array is an essential data provider for African Atlantic Ocean data. Various groups in South Africa (and very possibly the rest of West Africa) make use of the data collected through the PIRATA array.
The Southern African Data Center for Oceanography (SADCO) archives, extracts and manipulates multi-disciplinary oceanographic information in the southern African region, specifically the south-eastern Atlantic Ocean, the south-western Indian Ocean, and the Southern Ocean. Data is obtained from local marine organizations, universities, the South African Weather Service, government agencies and from international data sources. Data available is in the south Atlantic sector from 1900 to 2017. The database in the South Atlantic Ocean from 1970 to 2017 shows significant data coverage specifically in the coastal and equatorial regions. Specific platforms

(e.g. PIRATA, EAF-Nansen cruises) would need further verification as included in the database as large datasets may still be missing.

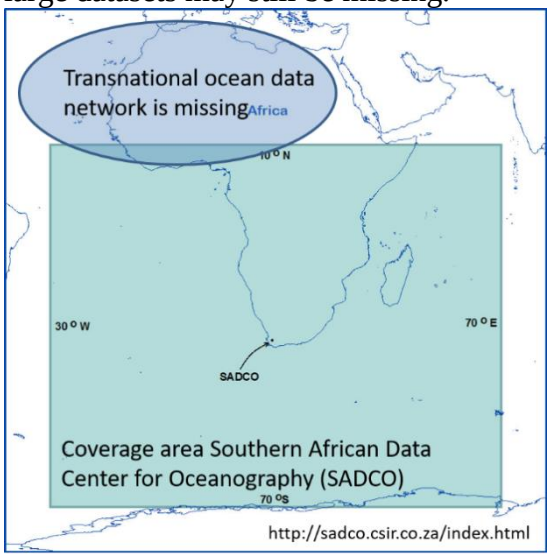

Figure 1: The target area of the SADCO extends from 10°N to 70°S and 30°W to 70°E. Centralized transnational data center is missing for Northern Africa.

ODINAfrica

The Ocean Data and Information Network for Africa (ODINAFRICA http://www.odinafrica.org) is a project of the International Oceanographic Data and Information Exchange programme (IODE). ODINAFRICA involves more than 40 marine institutions from 25 countries in Africa. The scope is to address the challenges faced in accessing data and information for coastal management by making data generated in national, regional and global marine monitoring programmes available to a wide range of users. However, in connection to this workshop, we found it difficult to access data and establish contacts with this program.

Current status and future initiatives:

In Africa, Ocean data curation and dissemination differs significantly between the Northern and Southern part of Africa. SADCO is a well-integrated data-center following international standards for data and metadata. Over the coming time period, SADCO is being transitioned in to an operational oceanography system and will undergo changes.

A centralized transnational data center is missing for Northern Africa, hence sustainable archiving and dissemination of Atlantic Ocean data is weak in this area. ODINAFRICA may a play role in the data collection in this area, but the status is currently unknown.

Appendix 2: Agenda

Day 1 (June 7th) - Atlantic data Landscape 09:15-12:00
09:15-09:45 Data landscape Europe: Sylvie Pouliquen, AtlantOS
09:45-10:15 Data landscape N. America: Mark Bushnell, NOAA
10:15-10:45 Data Landscape Africa: *Tamaryn Morris*, *SAEON* (webex)
10:45-11:00 Break
11:00-11:30 International initiatives: **Ocean Data Interoperability Platform -** Dick M.A. Schaap
11:30-12:00 International initiatives: Biol.-data harmonization- IOC-OBIS - Ward Appeltans
12:00-13:00 Lunch
13:00-13:30 Data Landscape S. America: Rafael Medeiros Sperb, FURG (webex)
13:30- 18:30 Plenary discussions in two Serial sessions
13:30-16:00 Discussion topic 1: Improvement of procedures and best practices; from sensor calibration at sea to sensor qualification at Data centers across continents, and analysis of harmonization perspective. Linking remote sensing and in situ observations in a trans-Atlantic context (Session Chairs: Mark Bushnell-NOAA and Christoph Waldmann-MARUM, Sylvie Pouliquen-Ifremer, *Tamaryn Morris SAEON)*
16:00-18:30 Discussion topic 2: Exploration of opportunities for improvement of transcontinental data-connectivity (cf. interoperability, standardization, accessibility) on a community level Meta data assignment – Finding common grounds for assignment of key permanent identifiers (Sylvie Pouliquen, Ifremer, *Tamaryn Morris SAEON)*
Day 2 (June 8th)
09:00-09:30 The use virtual research environment and cloud services (Thomas Loubrie ifremer)
09:30-10:00: User-perspective on trans-continental Atlantic Ocean data availability - Identifying User-needs in a transatlantic perspective. – an interactive discussion (Ketil Koop-Jakobsen MARUM
10:00-11:30 Synthesis and formulation of action items.
11:30-13:00 Lunch
13:00-13:30 The workshop outcome – ideas for dissemination of workshop outcome and how this workshop fits future initiatives (Sylvie Pouliquen, Ifremer). Strategic planning to realize the outcome of the workshop.
13:30-14:30 Plenary drafting the first outline for the Structural framework for a white paper.
14:30-15:00 Conclusion, Wrap up