Project	AtlantOS – 633211
Deliverable number	7.15
Deliverable title	AtlantOS Data Services Report
Description	Report on AtlantOS data services for AtlantOS observations
Work Package number	7
Work Package title	Data flow and data integration
Lead beneficiary	
Lead authors	V. Harscoat, S. Pouliquen
Contributors	WP7 Partners
Submission data	20 December 2018
Due date	31 July 2018
Comments	The report delivery was postponed in order to include a valuable view of the traceability dashboard service under development and aiming to provide statistics on data usage to data providers.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n° 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 ☐? ☒ National governmental body ☒ International organization ☒ NGO ☐ others Please give the name(s) of the stakeholder(s): …
WHERE is/are the company(ies) or organization(s) from?	☒ Your own country ☒ Another country in the EU ☒ 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?**	☒ Yes, because the report shows the effective results achieved within AtlantOS WP7 from the collaborative work of the networks and integrators partners in Europe to setup new data services or enhance existing data services to better serve the observation data and the users ☐ No, because …..
Will this deliverable be used? **If yes, who will use it?** **If not, why will it not be used?**	☒ Yes, by in-situ observation data system managers, by users of in-situ Atlantic Ocean observation data, by data providers ☐ No, because …..

NOTE: This information is being collected for the following purposes:

1. To make a list of all companies/organizations with which AtlantOS partners have had contact. This is important to demonstrate the extent of industry and public-sector collaboration in the obs community. Please note that we will only publish one aggregated list of companies and not mention specific partnerships.
2. To better report success stories from the AtlantOS community on how observing delivers concrete value to society.

*For ideas about relations with stakeholders you are invited to consult D10.5 Best Practices in Stakeholder Engagement, Data Dissemination and Exploitation.

Contents

1 References 5
2 Executive summary/ Abstract 6
3 Introduction 7
3.1 Scope 7
3.2 The integrated data system 7
4 The AtlantOS catalogue 9
4.1 Presentation 9
4.2 Features 9
4.2.1 Discovering by browsing through the catalogue entries 9
4.2.2 Facilitating the access to services 10
4.3 GEOSS integration 12
5 The traceability service 13
5.1 Principles for sharing information on data usage 13
5.2 'Pivot' format for shared log information 14
5.3 The AtlantOS dashboard 15
6 Data services at Integrator level 18
6.1 SeaDataNet 18
6.2 EMODnet physics 21
6.3 Copernicus INS TAC 23
6.4 ICES 24
6.5 EMODnet biology and EurOBIS 26
7 DOI services at Network level 28
8 List of acronyms 29

Illustrations

Figure 1: The integrated European data system in yellow the observing network systems and in blue the integrators in green and red the harmonization and integration elements. 7
Figure 2: discovering the in-situ observing networks managing Chlorophyll-a and fluorescence data 9
Figure 3: discovering the catalogue entries related to "argo" 10
Figure 4: viewing the sea_water_salinity parameter from one Copernicus INS TAC product 10
Figure 5: accessing to "Gliders" data services from the AtlantOS catalogue 11
Figure 6: metadata from the AtlantOS catalogue integrated in GEOSS 12
Figure 7: global schema of the traceability service designed with AtlantOS 13
Figure 8: processing schema in Semaphore from raw usage data to exposed data (statistics) via consolidated usage data (pivot format) 14
Figure 9 : Example of a log information in a pivot file 14
Figure 10: The AtlantOS dashboard main menu 15
Figure 11: example of Argo network in the AtlantOS dashboard on data usage for December 2018 17
Figure 12: SeaDataNet/Copernicus INS TAC/EMODnet physics data integration and services 18
Figure 13: Map with Euro-Argo coverage, and summary of contributors 21
Figure 14: EMODnet Physics data portal dedicated to AtlantOS observations 22
Figure 15. ARGO platform page 22
Figure 16: Glider platform page 23
Figure 17: Copernicus INS TAC monitoring tools 24
Figure 18: ICES Vocab web interface, showing a DATRAS shipcode mapped to a SeaDataNet Platform 25
Figure 19: DOI architecture at ICES, September 2018 26
Figure 20: Mapping of species observations in the North Atlantic Ocean from the EurOBIS database, using the EMODnet biology Portal 27

1 References

Ref	Title	Type	Version / Date
D7.4	Data Management Handbook http://doi.org/10.13155/48139	Report	V1.0 November 2016
D7.5	Service specifications https://www.atlantos-h2020.eu/download/deliverables/7.5%20Service%20specification.pdf	DEM	V1.0 29 March 2017
D7.8	Integration in GEOSS https://www.atlantos-h2020.eu/download/deliverables/AtlantOS_D7.8.pdf	Report	V1.0 22 March 2018

2 Executive summary/ Abstract

Within the H2020 AtlantOS project, WP7 has designed and started to set up an integrated data system that harmonizes work flows, data processing and distribution across in-situ observing network systems, and integrates in-situ observations into existing European and international data infrastructures, named Integrators (e.g. Copernicus INS TAC, SeaDataNet NODCs, EMODnet, EurOBIS, GEOSS).

The targeted European data system is not a new system but integrates existing data systems. These systems are being enhanced to ingest and deliver more in-situ observation data on Atlantic Ocean and to better serve the users, in a harmonized way across the systems.

The scope of data services for the AtlantOS observations is discovery, download, viewing and monitoring/traceability services for both users and data providers. This document reports the data services set up or enhanced within WP7 for AtlantOS observations: AtlantOS catalogue, AtlantOS traceability of use service and new or enhanced data services at Integrator level.

3 Introduction

3.1 Scope

This report describes the data services set up or enhanced within WP7 for AtlantOS observations.

- AtlantOS catalogue that provides users with a discovery service of the networks, integrators and products related to the Atlantic Ocean observation data, and it facilitates the access to services (viewing, downloading and monitoring) of the existing data systems.
- Traceability services that aim to give visibility of data usage to data providers
- Data services at integrator level

3.2 The integrated data system

The targeted European data system is not a new system but integrates existing data systems. These systems are being enhanced to ingest and deliver more in-situ observation data on Atlantic Ocean and to better serve the users, in a harmonized way across the systems. The existing data systems participating in "data harmonization and integration" activities in AtlantOS are diverse in-situ observing network systems operating in the Atlantic Ocean and existing European and international data infrastructures and portals, termed integrators (e.g. Copernicus INS TAC, SeaDataNet NODCs, EMODnet, ICES, EurOBIS, GEOSS).

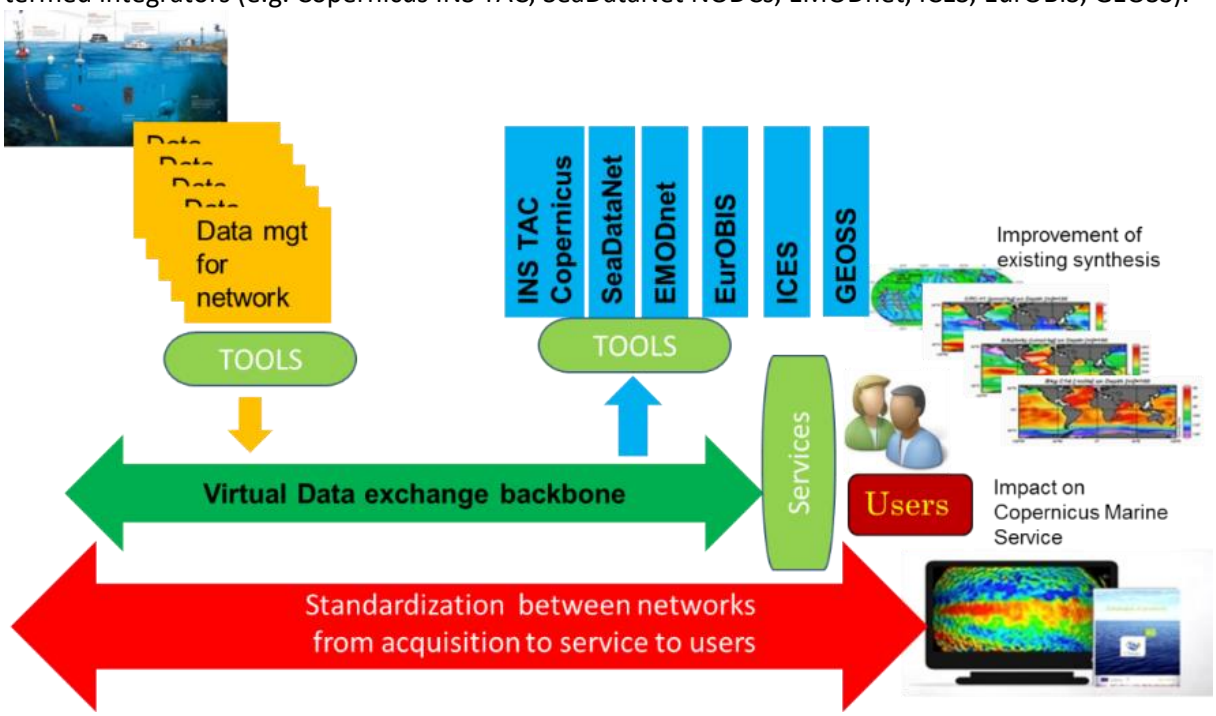


Figure 1: The integrated European data system in yellow the observing network systems and in blue the integrators in green and red the harmonization and integration elements.

An integrated data portal (https://www.atlantos-h2020.eu/project-information/integrated-data-portal/) proposes the access via web links to high level data services set up in well-established infrastructures:

- The "**AtlantOS catalogue discovery service"** set up within WP7
- The two **monitoring services** designed under the JCOMMOPS and EuroGOOS umbrellas: web based tools developed within AtlantOS WP9.1 (see D9.1 and D9.2 for more details)
  - the international (IOC/Jcommops), to track the implementation of networks (http://www.jcommops.org/board?t=atlantos)
  - the European (including coastal/regional regions) (embedded into EuroGOOS webpage http://eurogoos.eu/atlantos/atlantos-dashboard/) to track from the user point of view

- The **traceability service** is the most recent service, still under development but well underway, that aim to give visibility of data usage, through integrators with whom in situ data were shared, to data providers. It is a joint development from Ifremer, ETT and JCOMMOPS within WP7.

The system designed is based on existing and sustained infrastructures that will continue to operate even after the end of the AtlantOS project. The implementation has started within the project and it will go on within the framework of other projects.

4 The AtlantOS catalogue

4.1 Presentation

The **AtlantOS catalogue tool** is an element of the **data exchange backbone** of the integrated AtlantOS data system and the entry point of the integrated data portal. It intends to describe the elements of the integrated data system (the networks, the integrators and the relevant products related to Atlantic Ocean observation data), it provides users with **a discovery service** of these elements, and it **facilitates the access to services** (viewing, downloading and monitoring) of the existing data systems. It is also the front end window for the WP7 AtlantOS efforts to aggregate and federate observations.

The setup started end of 2016 and, since March 2017, it is on-line on the AtlantOS web portal (https://www.atlantos-h2020.eu/atlantos-catalogue): all in-situ observing networks contributing to the integrated system and all the integrators are described, together with product descriptions relevant on the Atlantic Ocean (from Copernicus INS TAC, SeaDataNet, SOCAT, GLODAP,…). The catalogue content, services, data and products available in the integrated system are progressing as existing systems move forward in achieving the goals of data harmonization and integration.

After the lifetime of the AtlantOS project, the sustainability of the catalogue at European level will be achieved through GEOSS and EuroGOOS.

4.2 Features

It is implemented with the GeoNetwork component of the **Sextant** Spatial Data Infrastructure (Ifremer) and is sustained as one of the thematic catalogues managed in the Sextant tool. All entries in the catalogue comply with ISO 19139 standard for metadata description.
Thanks to an API facility it is possible to integrate the catalogue interface in any web site and that is what was done on the AtlantOS web site.

4.2.1 Discovering by browsing through the catalogue entries

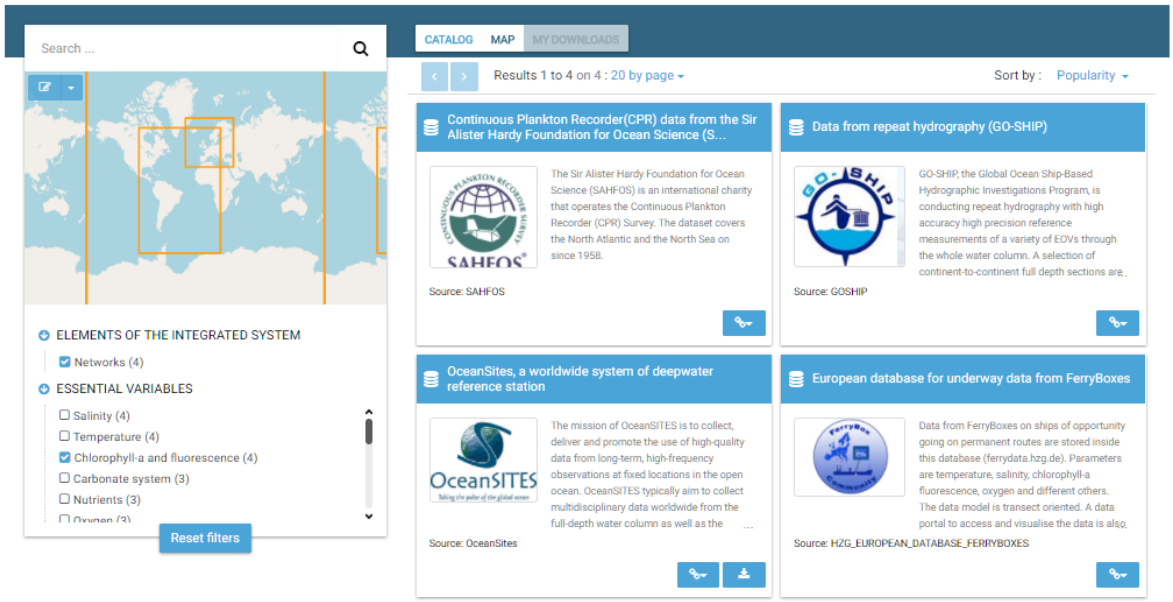


Figure 2: discovering the in-situ observing networks managing Chlorophyll-a and fluorescence data

On the left side of the catalogue interface (see the figure), two filters are available to browse the catalogue entries: **ELEMENTS OF THE INTEGRATED SYSTEM** that applies on ('Networks', 'Integrators', 'Products', 'Monitoring') and **ESSENTIAL VARIABLES** that applies on the terms from the vocabulary 'AtlantOS EVs' setup within WP7.1 (https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/A05/).

Furthermore, a user can enter a text in the general search field and this selects the catalogue entries that contain the text in any of the metadata of the description. In the example below, 4 entries contain the text "argo" and the user can see what elements (products, monitoring, network) and essential variables (Salinity, Temperature…) are concerned by these entries. Then the user can also use the two specific filters to apply a further selection on the general search.

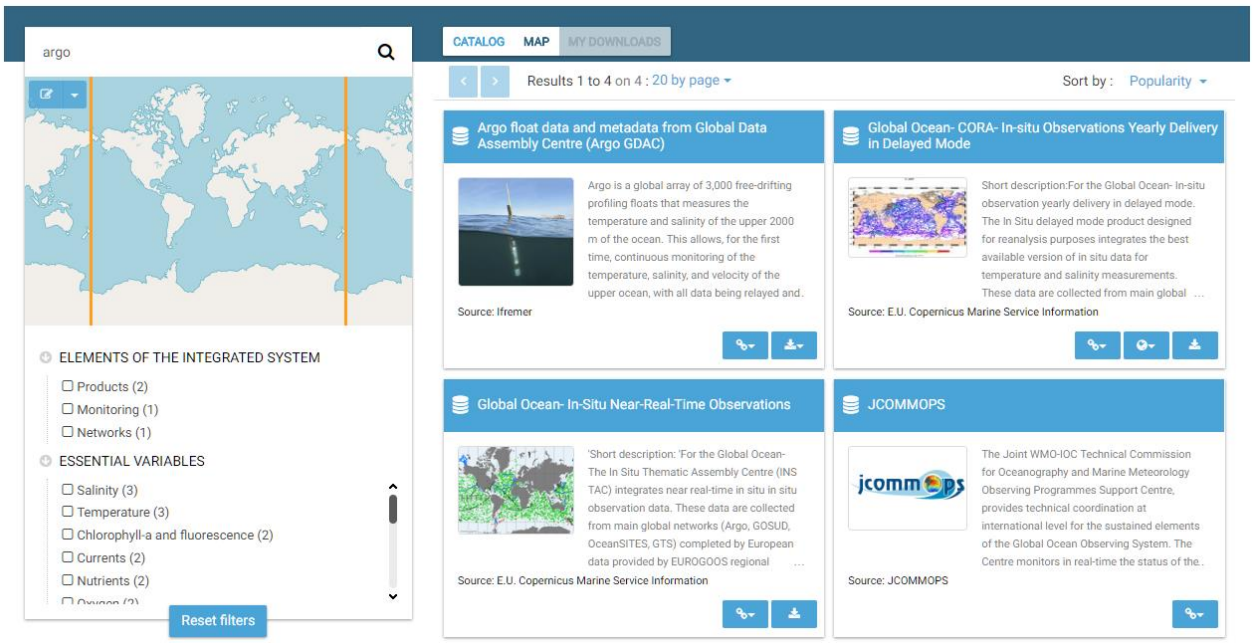


Figure 3: discovering the catalogue entries related to "argo"

4.2.2 Facilitating the access to services

Each catalogue entry proposes the access to the services available for this entry.

- **Viewing**

If a WMS service is available for one catalogue entry, the different layers can be an be displayed in the map facility of the web tool using the viewing facility , as illustrated in this figure

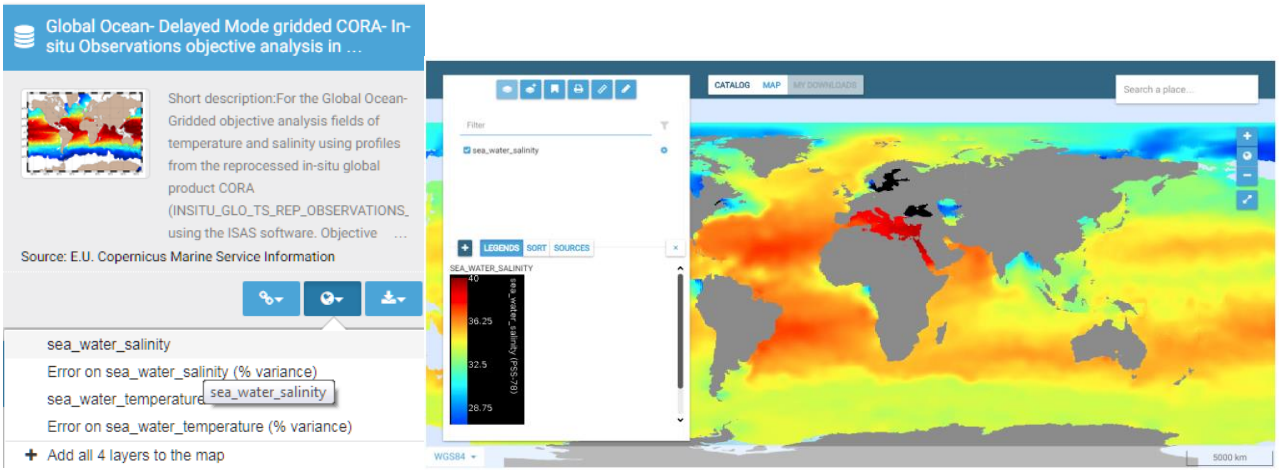


Figure 4: viewing the sea_water_salinity parameter from one Copernicus INS TAC product

- **Downloading**

If a downloading service is available for one catalogue entry, like FTP or other specific tool, it can be access using the downloading facility that will open a new tab in web browser, as illustrated in this figure

- **Accessing to external web links**

The external web links facility allows to access to any URL, the accessed link being open in a new tab of the web browser. With this facility networks or integrators can exposes their different and relevant data services from the AtlantOS catalogue.

For example, the entry for "Gliders" proposes three main data services for users of this in-situ network: EGO data portal, format reference manual and data processing chain

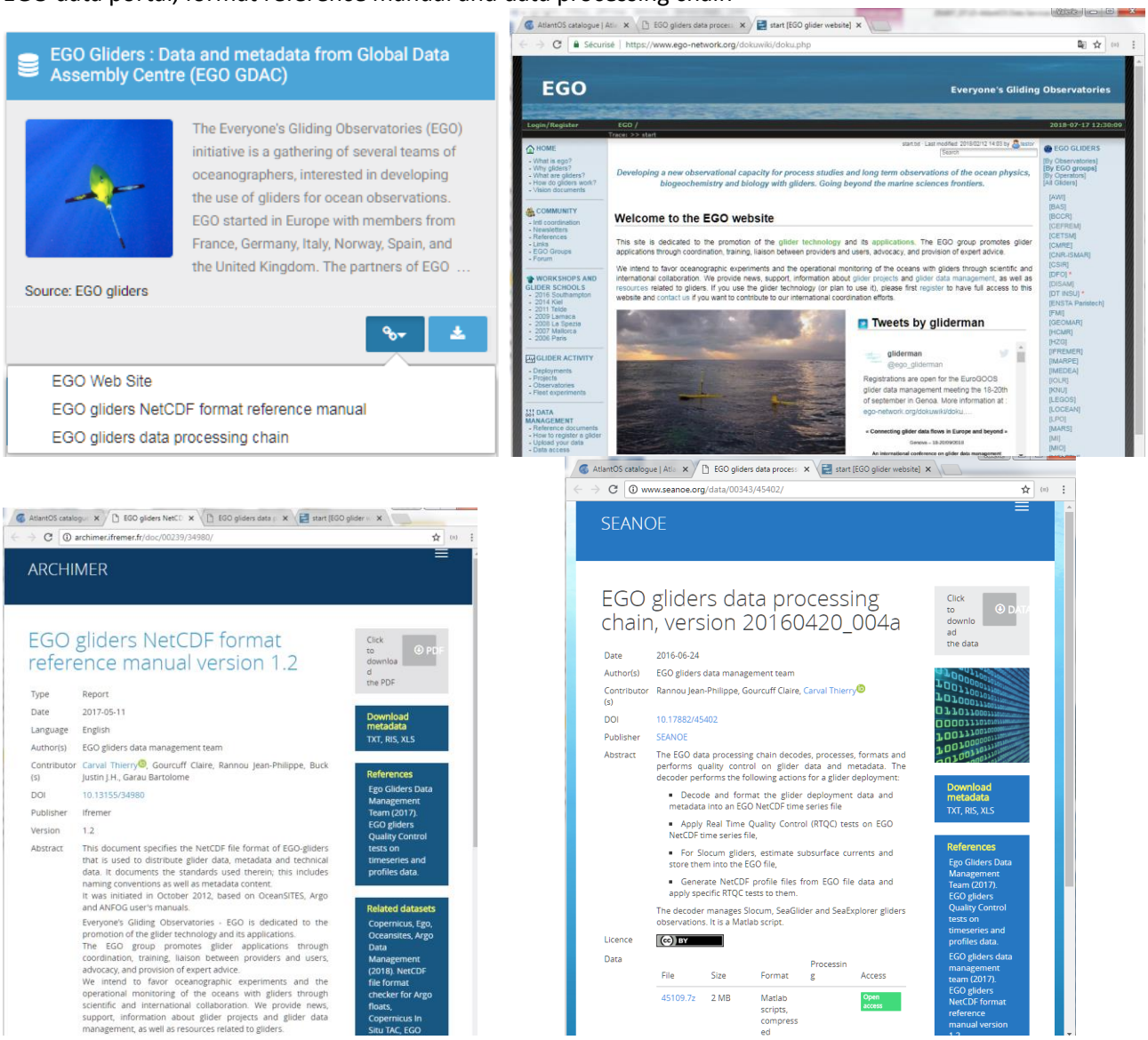


Figure 5: accessing to "Gliders" data services from the AtlantOS catalogue

4.3 GEOSS integration

The integration of the AtlantOS catalogue in GEOSS portal is achieved through an interoperability mechanism activated on GEOSS side to broker all the descriptions available in the AtlantOS catalogue, and feeding and updating automatically a mirror site in the GEOSS portal (see D7.8 Integration in GEOSS for more details).

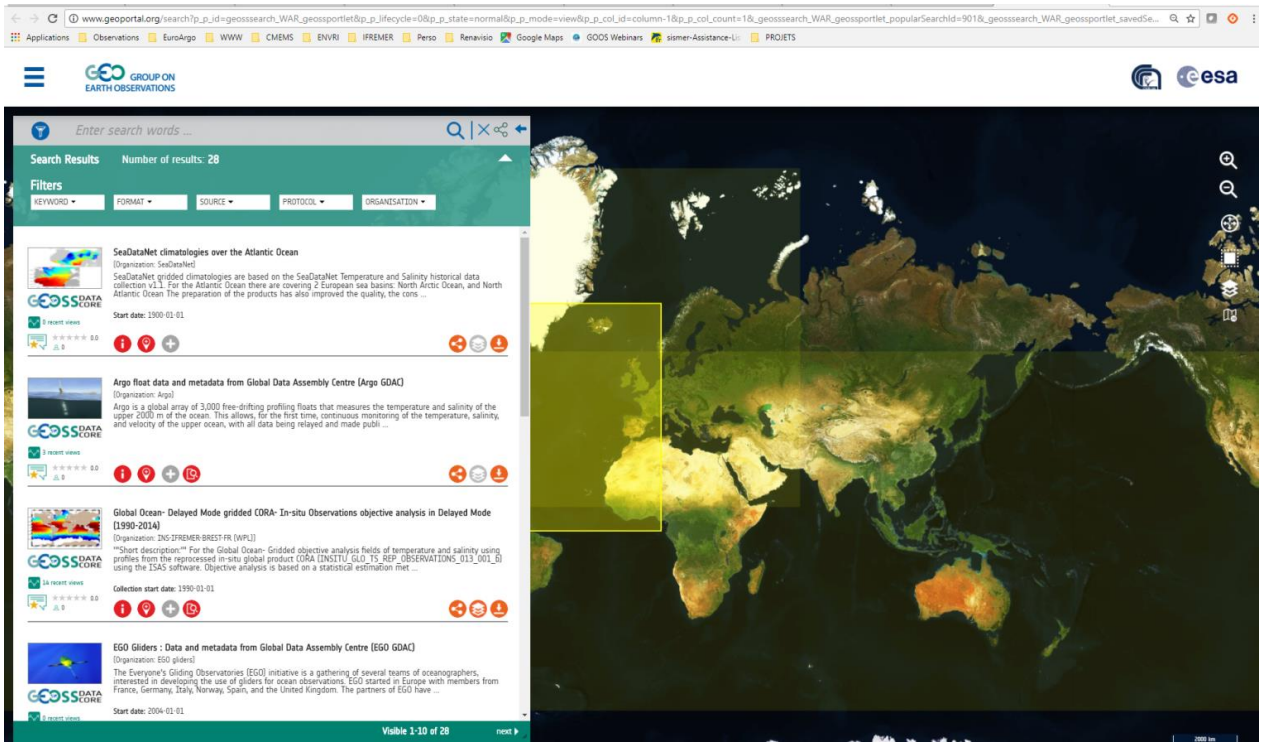


Figure 6: metadata from the AtlantOS catalogue integrated in GEOSS

Furthermore GEOSS is using AtlantOS as a demonstrator for the community approach, which is also shown in the news about the AtlantOS published by GEOSS on their webpage.

5 The traceability service

There is a diversity among the data services (discovery, download, viewing, DOI,…) proposed by the existing systems of networks and integrators: data plots without any restriction, services with different rules and different service level agreements, privacy rules,… But it is still possible to maintain a common set of minimal tracking information that can be shared and used to compute core statistics on data usage through a central integrated traceability service.

This service proposes a web interactive dashboard presenting core statistics that give visibility of data usage to both data providers and to data managers.

It is setup thanks to the joint developments from Ifremer, ETT and JCOMMOPS. Within the AtlantOS WP7 the objective was set up the initial version of the AtlantOS traceability dashboard taking into account the logs for data downloads from FTP data services managed at (1) Ifremer for Copernicus INS TAC till May 2018, SeaDataNet products, Argo, Glider and OceanSITES GDACs and (2) at EMODnet physics. This goal was achieved at the end of 2018.

The further upgrades of this traceability service (logs from more data services, both from more type of services and from more networks/integrators) are going to be achieved and sustained over the coming months and years within other projects at European level.

5.1 Principles for sharing information on data usage

The traceability service is based on a common format, named 'Pivot format' (see https://docs.google.com/document/d/1QnMrpmsPYvCi1e7Su-RrYvrripnUznCjvRazG8fQdtA/edit?usp=sharing), defined to share the tracking information (logs) across the existing data services.

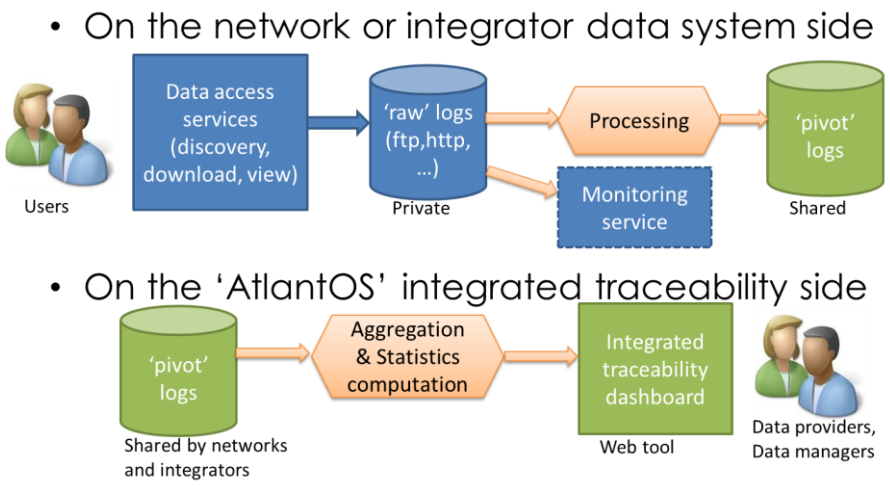


Figure 7: global schema of the traceability service designed with AtlantOS

The global raw logs generated by the data services are 'processed' to generate the pivot logs that can be shared at a central level: the data access from search engines are filtered, the information is enriched with metadata on data providers and platforms when available, and the user information is anonymized to comply with GPDR (setup in May 2018 by the EU to protect the personal data).

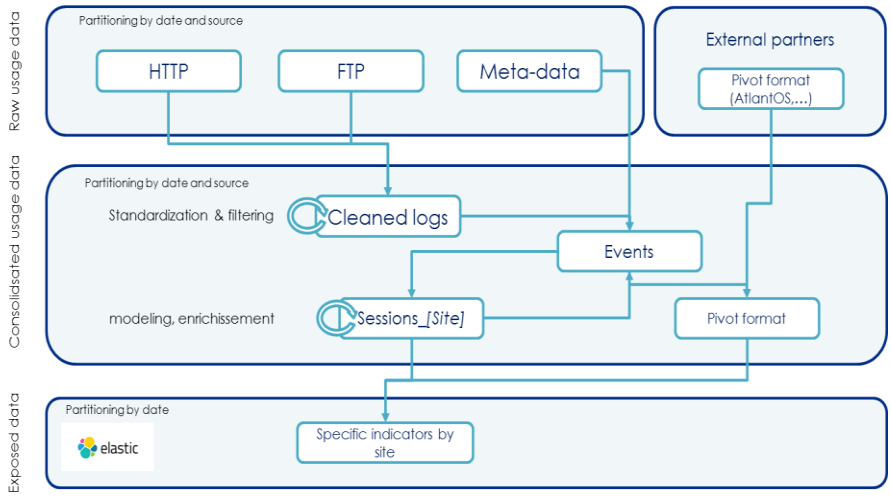


Figure 8: processing schema in Semaphore from raw usage data to exposed data (statistics) via consolidated usage data (pivot format)

In this framework, a joint development has gone on between three partners involved in WP7:

- JCOMMOPS manages and provides the metadata on the platform deployments for the networks involved in the AtlantOS projects,
- Ifremer develops tools (Semaphore framework) that perform the log 'processing' from raw information to pivot information.
- ETT develops the AtlantOS dashboard by uploading the pivot logs and computing core statistics from those logs.

The development implements big data technologies (Spark, Elasticsearch, Kibana) for the log processing, the uploading and computation of core statistics.

5.2 'Pivot' format for shared log information

The metadata to enrich the information available in raw logs are (1) metadata on platforms (JCOMMOPS) including the EDMO code of the data provider (organization responsible for the platform), (2) the information from EDMO catalogue (SeaDataNet) and (3) the metadata from Sextant catalogue (IFREMER) for description of datasets.

The 'Pivot' format for shared logs includes a minimum set of mandatory and shared fields to allow core statistics computation and optional fields (only if available and sharable by a system) to allow more detailed statistics.

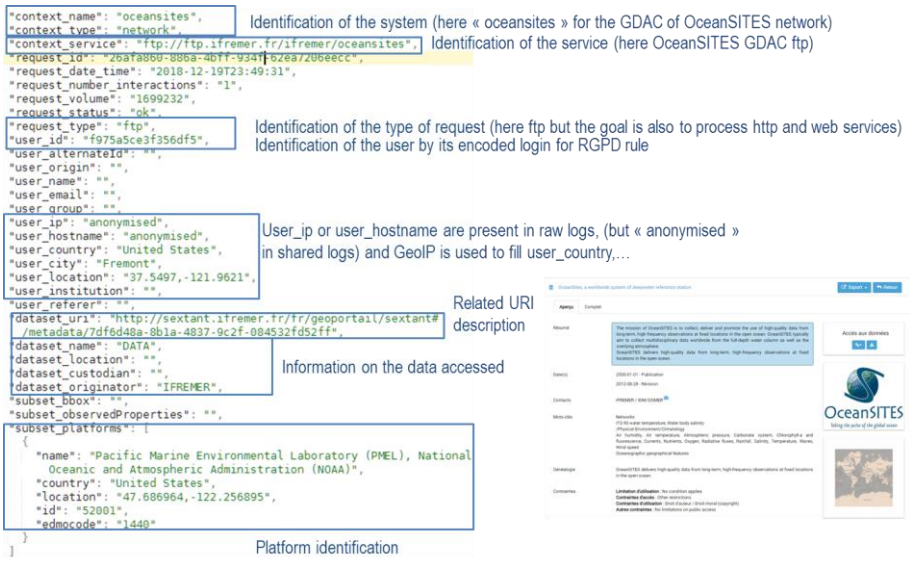


Figure 9 : Example of a log information in a pivot file

5.3 The AtlantOS dashboard

The traceability service proposes a web interactive dashboard (http://www.emodnet-physics.eu/atlantos/Dashboard/KibanaDashboard.aspx) of core statistics presented by network or integrator of the integrated AtlantOS system, for the last 7 days or last 30 days or last year or a user period selection.

The initial list of metrics/map/graphics included in the dashboard is:

- Number of downloads
- Number of unique identified users
- Map of downloads per geographical location of users
- Top 50 of data providers (Institutions responsible of platforms, EDMO codes and names)
- Number of downloads per day
- Top 50 of user countries
- Top 50 of platforms ids

Thanks to interactivity features, the metrics/map/graphics are automatically updated when the user changes the period to visualize or when the user selects an element on each map/graphics, allowing getting global or detailed statistics (for example getting the statistics of a specific platform in the top 50).

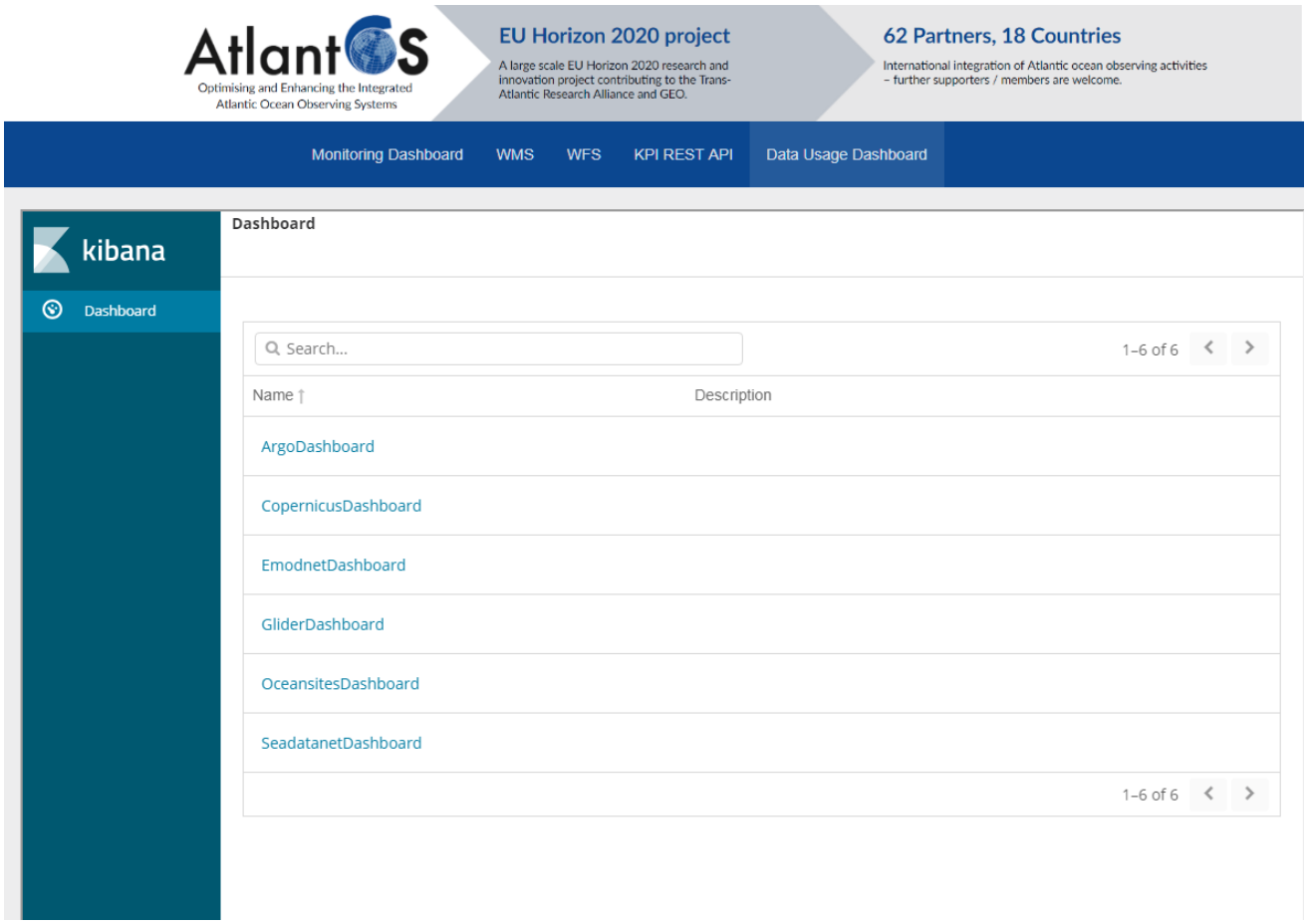
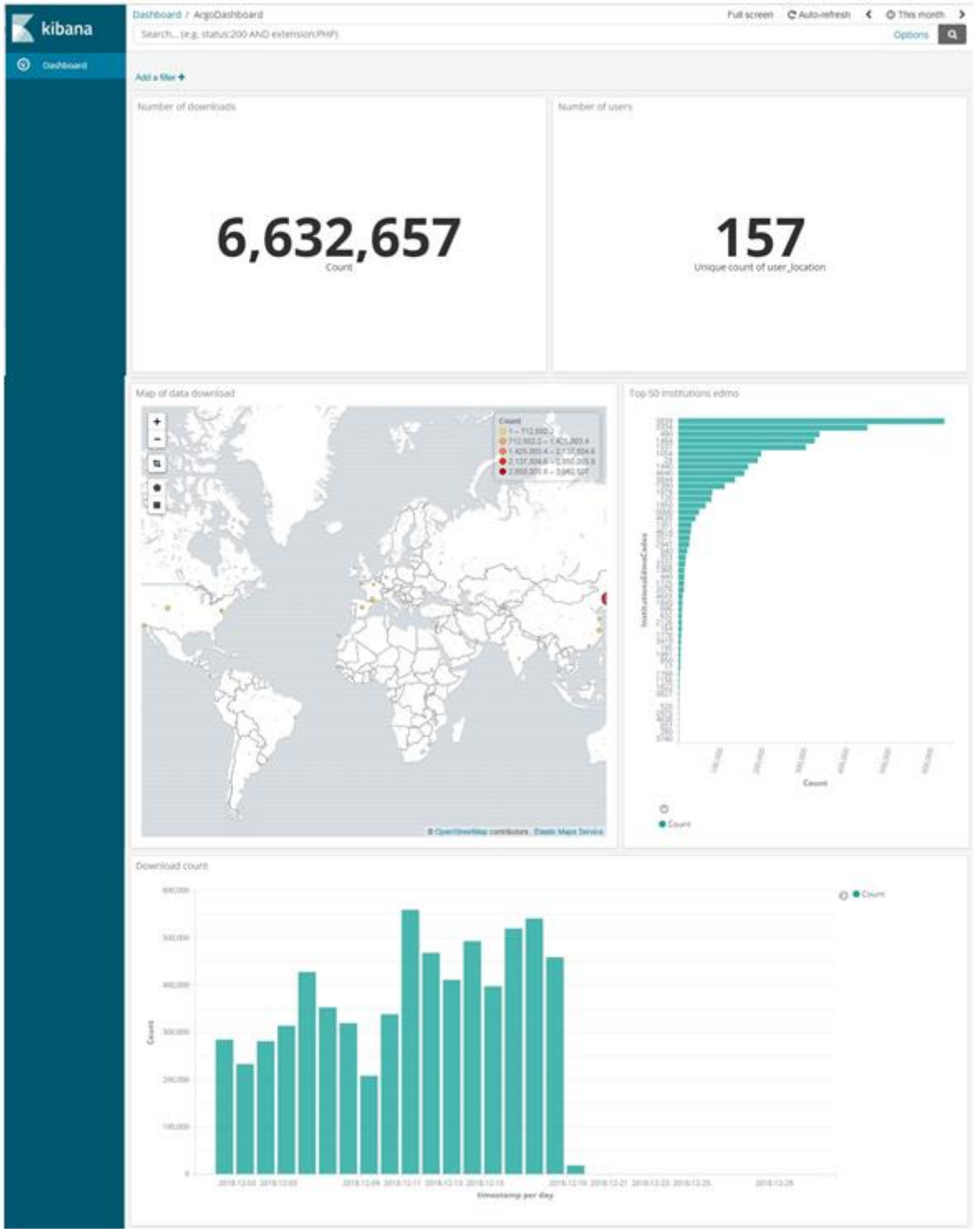


Figure 10: The AtlantOS dashboard main menu

The initial version made available at the end of 2018 and presented in this report includes the core statistics for FTP download data services. The later versions shall include the statistics gathered from other data services (such as http, wms…).

Below the example of Argo network in the AtlantOS dashboard on data usage for December 2018, the starting month of the automatic daily processing of FTP data downloads.



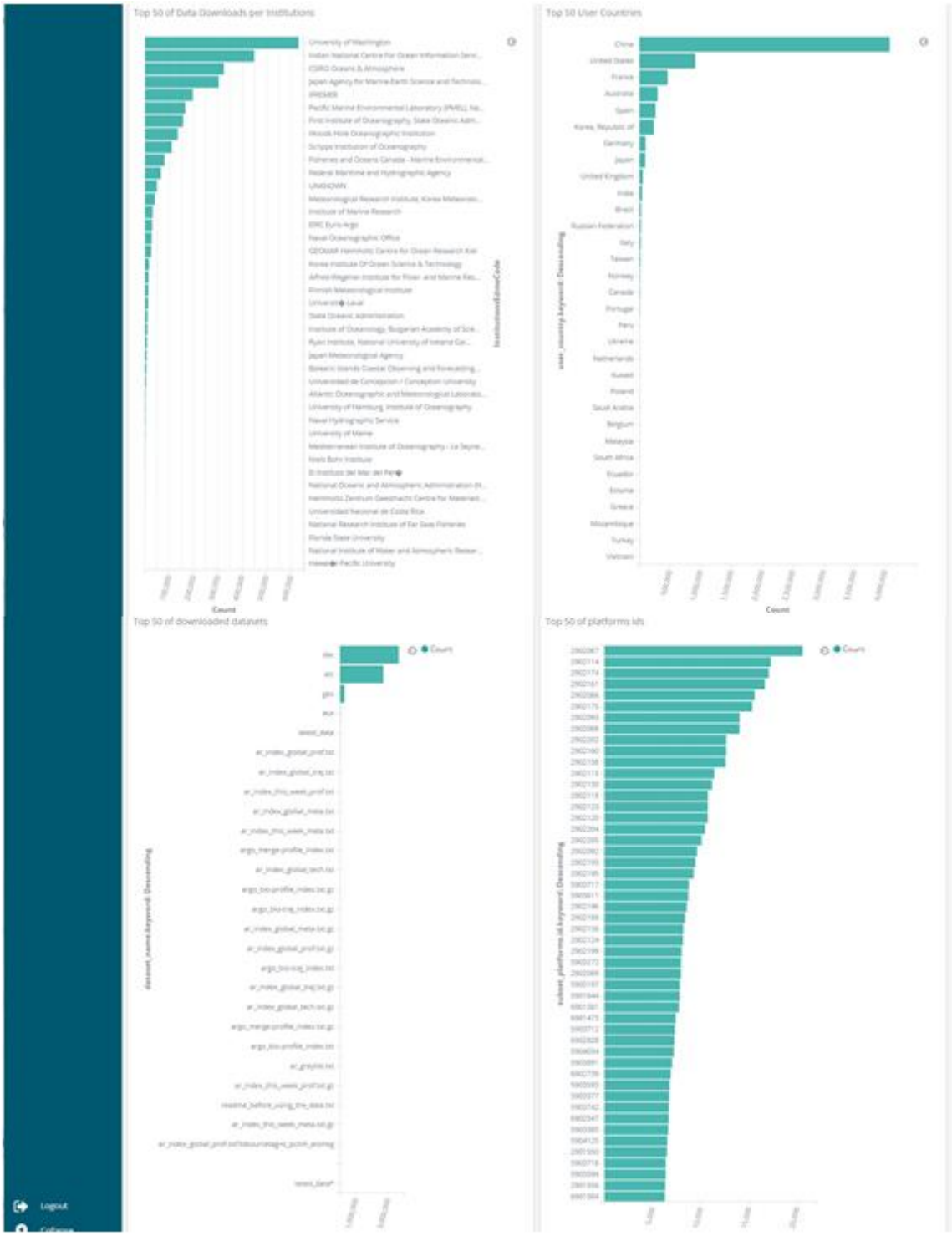


Figure 11: example of Argo network in the AtlantOS dashboard on data usage for December 2018

6 Data services at Integrator level

This part of the report highlights the improvements in data services at Integrator level in order to better serve Atlantic Ocean observations from in-situ networks involved in AtlantOS, among:

- Serving more data
- Implementation of AtlantOS recommendations for metadata (platform ID and EDMO codes for Institution) and parameters (agreed vocabularies), in tools and data format
- Enhanced data services, like plots and views customized according to the platform/network

The European infrastructures or global assembly centres involved as Integrators in AtlantOS are:

- For marine environmental data: SeaDataNet for validated and archived data; and the Copernicus Marine Environment Monitoring (CMEMS) INS TAC (In-Situ Thematic Assembling Centre) for NRT data and for the past 60 years of historical data assembled for reanalysis needs
- for marine biodiversity data: the ICES system and EurOBIS

The Portals involved as integrators in AtlantOS are:

- EMODnet lots (physics, chemistry, bathymetry, biology) fed by Copernicus INS TAC, SeaDataNet and EurOBIS
- GEOSS

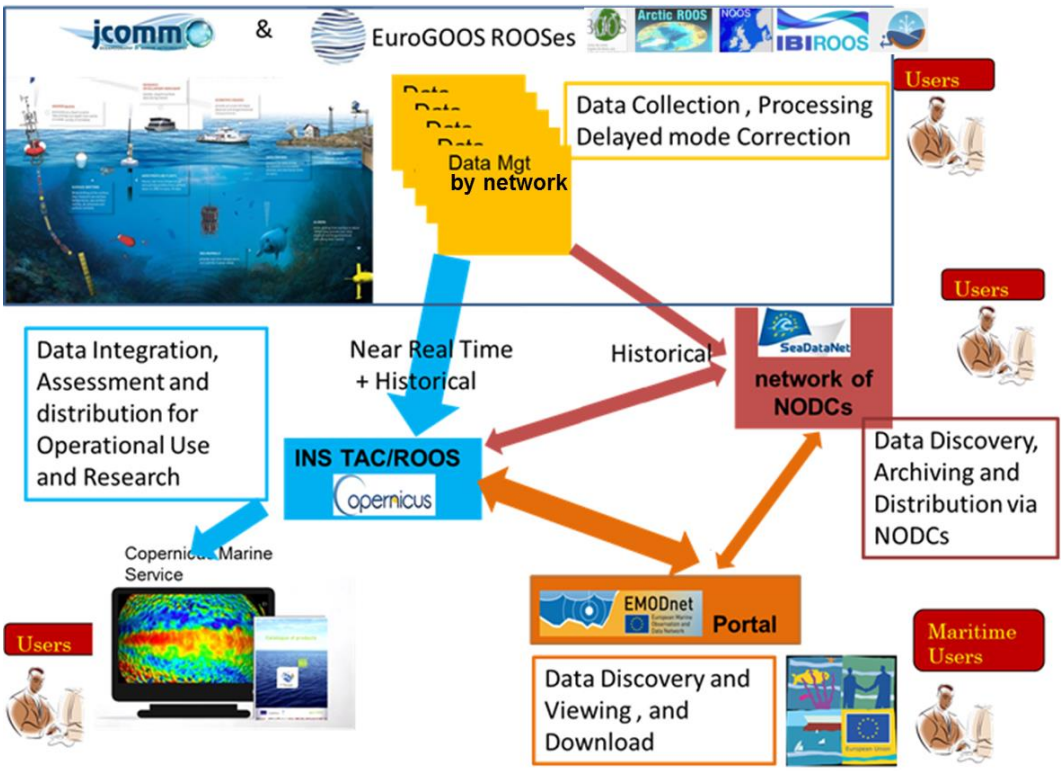


Figure 12: SeaDataNet/Copernicus INS TAC/EMODnet physics data integration and services

6.1 SeaDataNet

SeaDataNet (SDN) is one of the integrators of validated historical marine datasets, but at the same time also an important contributor to marine data standards. When looking at the improvements in SeaDataNet data services, and improvements coming forward from SeaDataNet, as a result of AtlantOS related actions we find:

- Adoption of SeaDataNet standards in the data management of Atlantic Near Real Time monitoring networks to enhance traceability of data
- Expansion of SeaDataNet vocabularies for platforms and stations
- Expansion of available data archives via the CDI data access system, in the end resulting in more available data for research and EMODNet data products.

Adoption of SeaDataNet standards

SeaDataNet is an important player in setting metadata and data standards for validated marine datasets. Important part of these standards are the vocabularies to describe important attributes of the data: Organisations (EDMO) https://www.seadatanet.org/Metadata/EDMO-Organisations, parameter collected, instrument, data format, etc (see https://www.seadatanet.org/Standards/Common-Vocabularies). Adoption of such standards as early as possible in the lifecycle of data (so in the networks involved in AtlantOS) contributes highly to the provenance of the data, and insight in the quality, when the data is later distributed into data products, as well as directly to researchers via system like the SeaDataNet CDI system https://www.seadatanet.org/Data-Access. In close communication with the network representatives big steps have been made in adoption of EDMO and the vocabularies (station and platform codes, parameter codes) in e.g. Euro-Argo, CPR, JCOMMOBS, and Copernicus INS TAC.

The progress of adoption of standards in networks involved in AtlantOS WP7 is presented in the following table:

	Network	C17 platforms and ICES station code	EDMO mapping
1	VOS/SOOP, GOSHIP/GOSUD	Part via SOCAT ongoing, via NODCs done	Part via SOCAT ongoing, via NODCs done
2	Ferryboxes	Via SDN, Copernicus INS TAC , EMODnet Physics done	Via SDN, Copernicus INS TAC , EMODnet Physics done
3	Drifters	Via WMO codes	Included in work plan, Ongoing (via JCOMMOPS and Copernicus INS TAC)
4	Argo	Via WMO platform code	Ongoing (via JCOMMOPS and Copernicus INS TAC)
5	OceanSITES	Done as part of NODC work.	Done as part of NODC work
6	Tide gauges	Via bulk upload to ICES station vocab	Done (via NODC's and ROOSes)
7	Gliders	Done, via WMO codes	Done
8	Seafloor Mapping	N.A.	N.A.
9	Fish and plankton survey (ICES)	Partly done	Ongoing
10	ETN	Investigated with ICES	Via action in EMODNet Biology
11	CPR/MBA	Done during SDN connection	Done during SDN connection, and via EMODNet biology
12	EMODNet biology	Will be checked in datamodel, ongoing	Started
13	JCOMMOPS	Vessels included in C17	Done and ongoing

Expansion of SeaDataNet vocabularies

The adoption of standards like EDMO and the SDN Vocabularies is a two-way process. During the adoption of EDMO it turned out several organisations involved in the monitoring (as operator or data centre) where not yet part of EDMO. They have been added in cooperation between MARIS, the network and the national data centre contact. The same approach was undertaken for platform (vocabulary C17) and station (ICES directory) codes. They have been added upon request and contributed to the quality and completeness of these directories.

Expansion of available data archives in SeaDataNet CDI

In AtlantOS there are several Atlantic monitoring networks involved, but the data flow from the actual observation (Real-Time/Near-Real-Time to validated archives) is very different for each network. In some cases (e.g. Euro-Argo, Gliders, CPR) and in a different way in SOCAT (data is collected from other centres, validated and aggregated into a product)) there is clear stream to a central DAC and GDAC, but for other data networks (e.g. tide gauges, VOS/SOOP, OceanSITES) the data flows into several national centres and find their way to SeaDataNet, Copernicus INS TAC and EMODNet via these centres.
In the last case however this data stream is often not complete and many archives are not yet published. In cooperation with the network representatives actions are ongoing to improve the situation, but results are not (yet) clearly measurable.

Measurable results can be found for Euro-Argo, CPR and SOCAT. For Euro-Argo, actions under AtlantOS have led to an improved dataflow by automatic processes on Euro-Argo side that aggregate monthly records, create CDI files and share this upon request of the NODC in a certain country. When the NODC does not hold the Euro-Argo data themselves, the data archives are shared directly from the GDAC. First results are visible for France and Italy, see screenshot below.

The methodology developed for Euro-Argo pave the way for other networks (Gliders, OceanSITES, drifters, Ferrybox, HF Radar…)

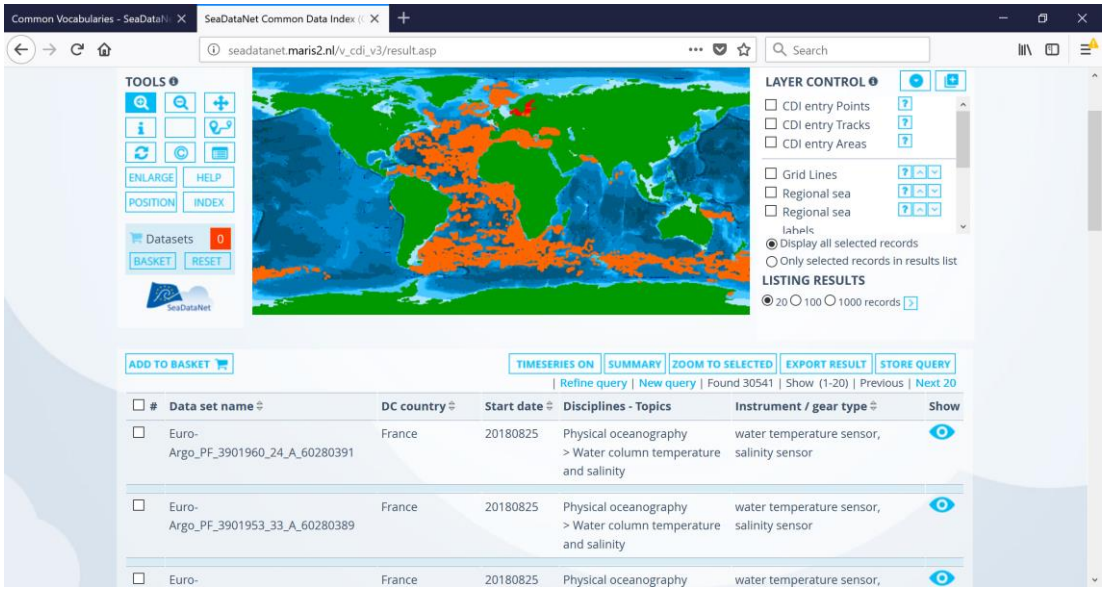


Figure 13: Map with Euro-Argo coverage, and summary of contributors

CPR CDI connection (via MBA), and SOCAT connection to the SeaDataNet CDI system are underway but currently still in test modus.

6.2 EMODnet physics

EMODnet Physics is a portal with GIS features designed to provide the user with discover, access, plot, and download of near real time and historical data and data products on physical parameters of the oceans. It is built in collaboration with and integrates data from the EuroGOOS ROOSs, Copernicus INS TAC and SeaDataNet infrastructures.

Implementing AtlantOS agreed standards

- **Metadata – EDMO codes for data providers**: EMODnet Physics is promoting the adoption of EDMO, in collaboration with Copernicus INS TAC and SeaDataNet does period checks (next check planned by end of 2018)
- **Metadata – Unique platform/station IDs**: EMODnet Physics follows and promotes the recommendations of DATAMEQ, AtlantOS and Copernicus INS TAC. Each platform in EMODnet Physics has a unique URL.
- **Common vocabularies for parameters**: EMODnet Physics follows and promotes the recommendations of DATAMEQ and SeaDataNet. When new parameters are popping up – it opens discussion with BODC (e.g. Under Water Noise is undergoing).

Serving more data

EMODnet Physics benefits from all improvements in data services from the related infrastructures to better serve Atlantic Ocean observations.
For near real time data, it is directly connected with the Copernicus INS TAC to make available data from about 2000 platforms and provides the users with access to about 6000 platforms in the AtlantOS area since the start of the project.
EMODnet Physics (and EMODnet lots in general) benefits from the EMODnet Data Ingestion project facility (www.emodnet-ingestion.eu). Data ingestion are first submitted and then processed and made available via the European infrastructures (NRT data flow into ROOSs, Copernicus INS TAC and EMODnet Physics; historical data flow into NODCs, SeaDataNet and EMODnet Physics).

Dedicated data services for AtlantOS

EMODnet Physics has set up a data portal http://www.emodnet-physics.eu/atlantos/ dedicated to AtlantOS observations (based on AtlantOS boundary file developed by JCOMMOPS) to facilitate data use, integration and interoperability, to attract/unlock new data providers and better quality data, to facilitate the link between users and providers (visibility, acknowledgment …).

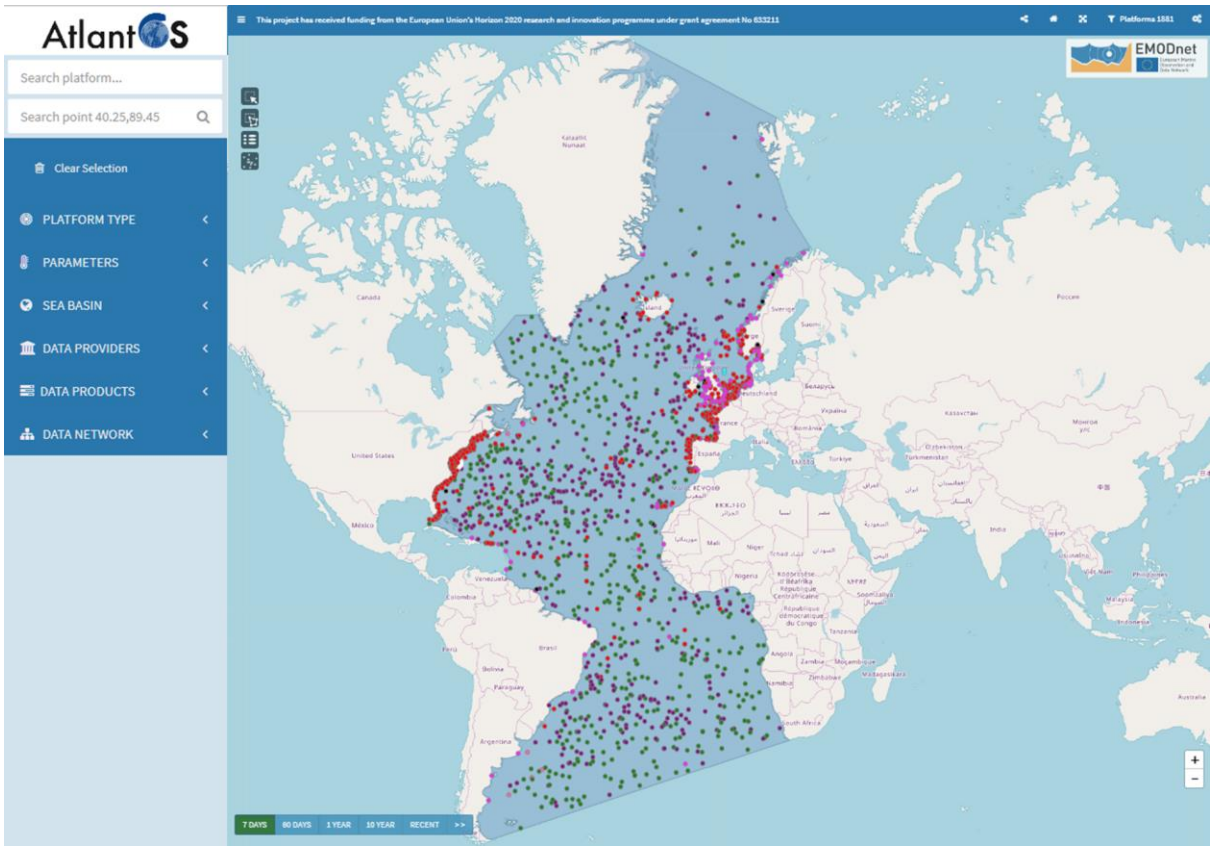


Figure 14: EMODnet Physics data portal dedicated to AtlantOS observations

For each specific platform type related to an in-situ observing network involved in AtlantOS, the dedicated EMODnet-Physics/AtlantOS web portal proposes a customized page with metadata (data provider, data curator, data assembly center), data plots (timeseries, profiles) and data download features. Users can download data without any restriction; authentication may be requested for data older than two months.

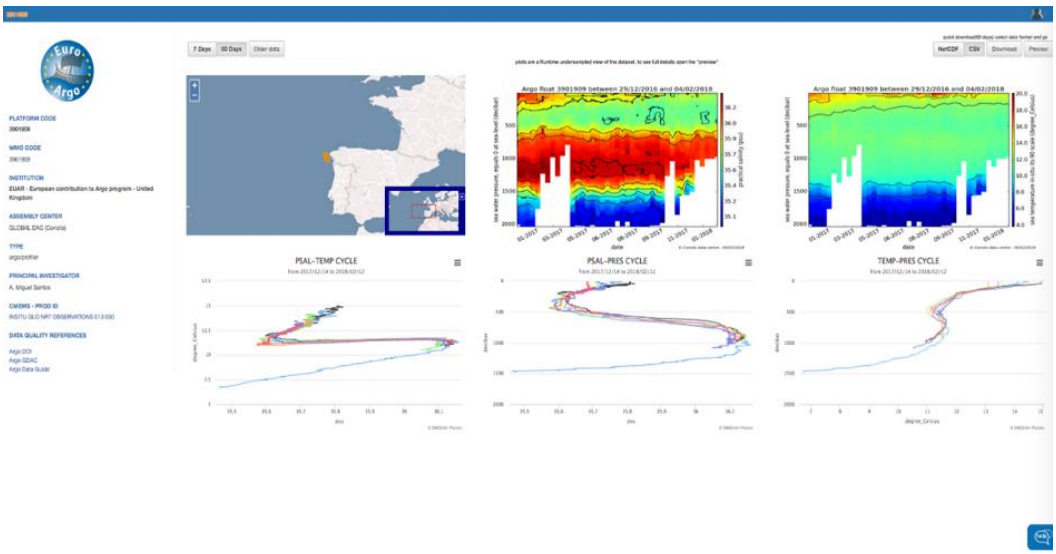


Figure 15. ARGO platform page

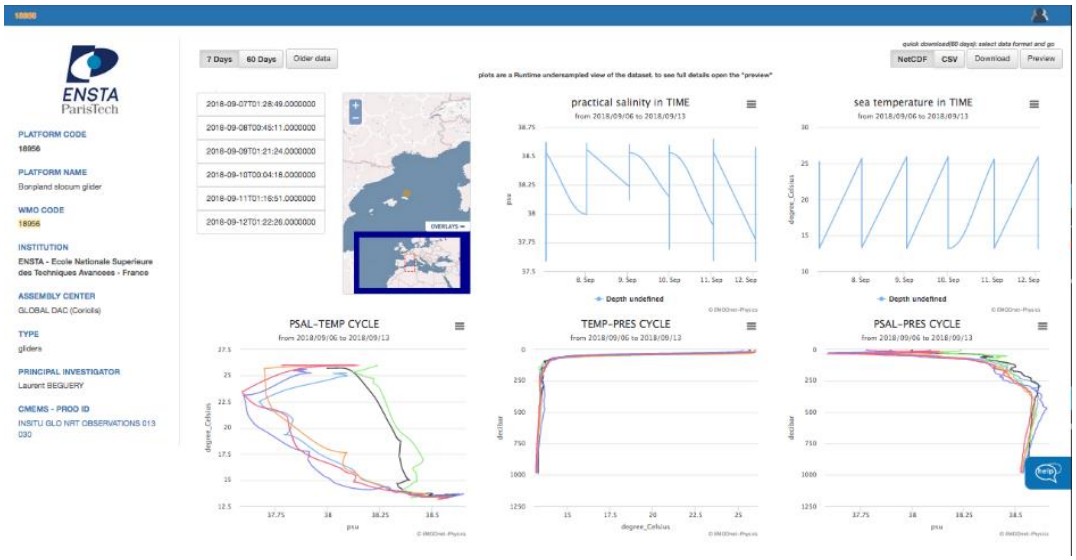


Figure 16: Glider platform page

The web portal is also implementing monitoring services that have been used:

- to monitor the adoption of AtlantOS recommendations for metadata (platform ID and EDMO codes for Institution) and parameters (agreed vocabularies),
- to monitor and report on quantitative and graphical information on the accessibility of data in terms of temporal distribution,
- to let the data providers to check which and how many of their platform are fully integrated and accessible at integrator level,
- to let the provider to retrieve information on data use and data users (from the dedicated EMODnet Physics/AtlantOS portal)

Furthermore it is possible to have information on the spatial distribution of the data collection for a given time frame: http://www.emodnet-physics.eu/atlantos/dashboard/Section26.aspx?shape=atlantos

6.3 Copernicus INS TAC

The Copernicus Marine Environment Monitoring (CMEMS) INS TAC aggregates the in situ data for operational oceanography (assimilation in models for forecast and reanalysis as well as model validation). These products are also used by downstream services using In situ products jointly with Model and Satellite products and by the research community.

For Near Real Time products it is connected to JCOMM Networks and EuroGOOS ROOS regional networks. For Delayed Mode products it is connected to SeaDataNet Network of NODCs, JCOMM Network sand EuroGOOS ROOSes.

Serving more data

Within AtlantOS Copernicus INS TAC has upgraded its data ingestion procedures to integrate the new data that were provided by the networks involved in AtlantOS (more data are available for Gliders, strong activity for Argo-BGC) and also has corrected anomalies in existing products thanks to a more streamlined data stream from providers to integrators. Copernicus INS TAC will connect to GDAC drifter as soon it is up and running to ingest more data.

Furthermore, Copernicus INS TAC has also work jointly with EMODnet, SeaDataNet and EuroGOOS to unlock access to existing data. Within the phase2 of Copernicus INS TAC started in January 2018, routine integration of HF Radars in link with the EuroGOOS HF radar task team has started and plans to be operation early 2019, as well as integration of carbon product into Copernicus INS TAC catalogue in link with ICOS-ERIC.

Implementing AtlantOS agreed standards

Copernicus INS TAC has enhanced its data system to implement the AtlantOS metadata recommendations: EDMO codes added to the files for all platforms when the provider is well identified, and the recommendations for platform IDs and common vocabularies for parameters are implemented. It has also ensured that the provider metadata are kept all along the data flow from provider to users.

Enhanced and new services

For viewing, a dashboard has been setup (http://www.marineinsitu.eu/dashboard/) and has been upgraded using existing services either from Coriolis or EMODnet Physics for viewing service. For downloading, enhanced subsetting facilities will be addressed in near future and already easy to download data from a specific platform.
Also monitoring tools are now available (http://www.marineinsitu.eu/monitoring/).

Figure 17: Copernicus INS TAC monitoring tools

Furthermore, jointly with EMODnet-physics and SeaDataNet partners Copernicus INS TAC has contributed to the designing and prototyping of the AtlantOS traceability service on data usage to prove the feasibility of such reporting service. To be able to provide feedback on data usage to Copernicus INS TAC providers, INS TAC will work with Mercator to turn such a service into operations.

6.4 ICES

ICES has a well-established Data Centre which manages a number of large dataset collections related to the marine environment. Data are served through the http://ocean.ices.dk portal. ICES is part of the SeaDataNet network of NODC's, it is a thematic integrator and it is linked to EMODnet Biology and EurOBIS for biological data from trawl surveys and acoustic.

Serving more data

The data integrated in ICES come from the EU CFP but also non-EU countries. ICES is improving data availability for non-EU by working further on this with specific workshops outside of AtlantOS scope.

Implementing AtlantOS agreed standards

For acoustic data, the data transport format and meta data standard were developed at the start of the AtlantOS project, so platform IDs and EDMO codes recommended for metadata are used from the outset as a core attribute.

For biological trawl surveys in DATRAS, the management of platforms has been historically separate from the platform management standards, or at least have diverged over time. The DATRAS platform field [TS_Ship] is now linked in the ICES semantic vocabulary to ICES SHIPC code, this being the list of platforms that are provided to the SeaDataNet C17 platform vocabulary (see below). There are however still a handful of older ships that are yet to be resolved to the respective SHIPC code. When this exercise is completed with the data providing institutes, the SeaDataNet platform code will also be added to the output formats/web services.

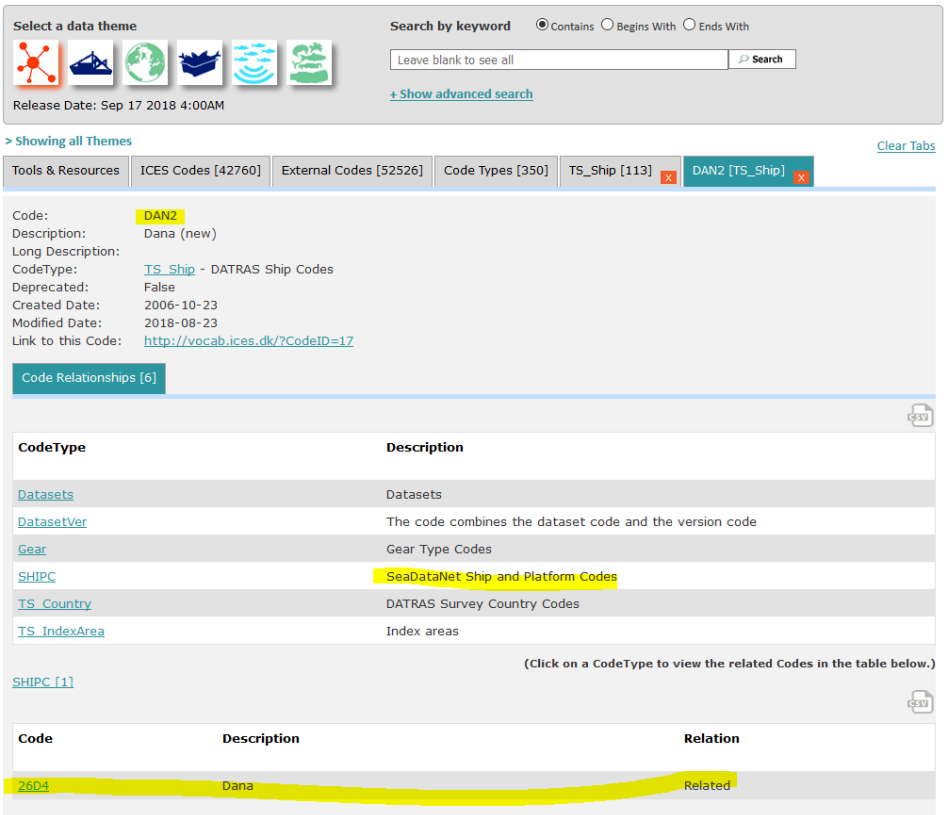


Figure 18: ICES Vocab web interface, showing a DATRAS shipcode mapped to a SeaDataNet Platform

Concerning EDMO codes for organizations, it was not as straightforward for DATRAS data as there has not been an attribute for 'institute' in the reporting formats that are used for biological trawl survey data. Work has now completed on mapping the combination of [TS_Country] and [Datasets] to the EDMO codes. Although this in the portal (vocab.ices.dk), this is not visualised correctly and either the interface needs to be modified, or a bespoke web service made from DATRAS for this. ICES is underway to simplify this, so it can be shown in the interface more clearly.

Enhanced and new services

ICES is a DOIs publisher and has redesigned the architecture of minting DOI's from the ground up to work completely from web services that can be called from any of the data systems, such as DATRAS or the

Acoustic portal (see below). The best approach for minting incoming data to the portals is still under discussion as the ability to mint a DOI for each data submission or resubmission is not necessarily helpful to the data provider, or to the end user. Focus will first be on the data outputs (survey indices and standard calculated products) as these are already versioned and so the decision about what to mint and when is more clear-cut.

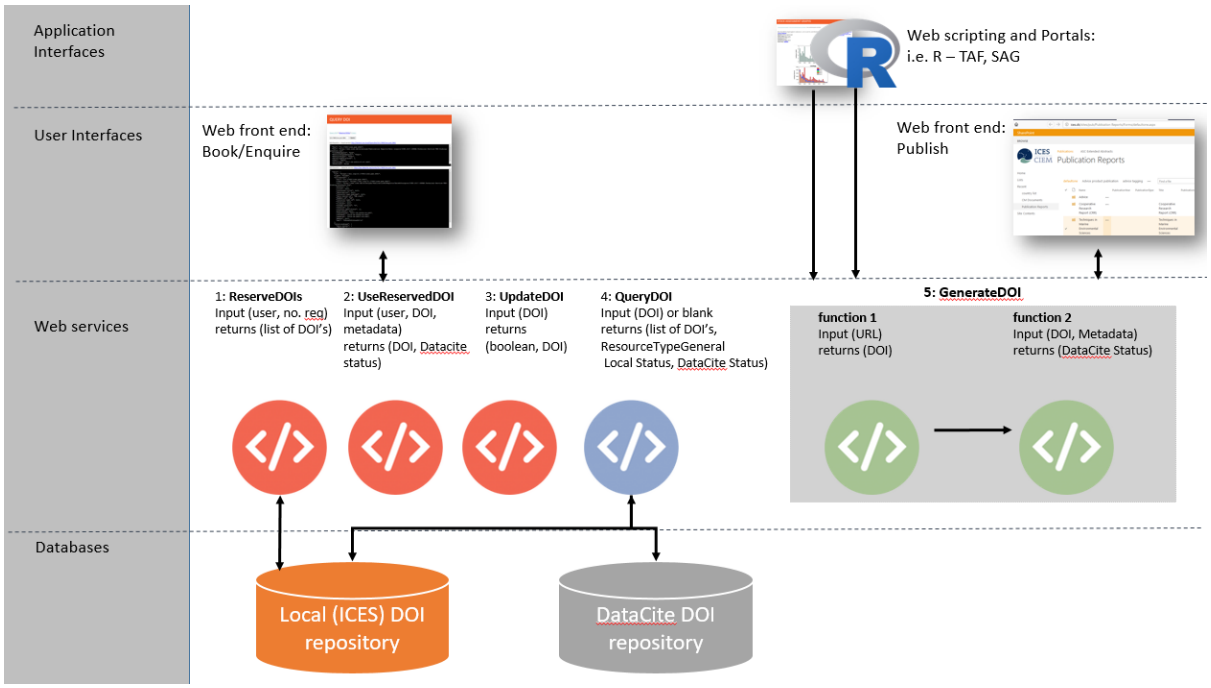


Figure 19: DOI architecture at ICES, September 2018

Also ICES is working on getting survey areas defined as map resources in Geoserver/network.

6.5 EMODnet biology and EurOBIS

EMODnet Biology makes available information of marine species including observations from phytoplankton, zooplankton, angiosperms, macro-algae, benthos, marine mammals, marine reptiles and fish and produces digital data products allowing to analyse changes in species abundance and extent over time and space. Using EurOBIS as data backbone, EMODnet biology aims to harmonize and centralize biogeographic data on marine species collected by European institutions. The data can either be collected within or outside (e.a. Atlantic Ocean) European waters. As long as the data providing institute is within Europe, EurOBIS acts as the responsible node to make these data available to the OBIS community. The Ocean Biogeographic Information System (OBIS) is the world's largest database on the diversity, distribution and abundance of marine life.

Serving more data

At this moment, EMODnet Biology contains 1.265 marine biological dataset descriptions (metadata) of which 875 datasets are integrated and available through EurOBIS, representing 23.860.954 occurrence records, of which 20.671.981 quality controlled records (87%). This corresponds with 77.724 species names – of which 60.868 are accepted species names – linked to the WoRMS. If we perform a geographic query on the database that corresponds to the IHO Region - North Atlantic Ocean, we obtain 4,077,653 occurrence records from different taxa (from phytoplankton, zooplankton, angiosperms, macro-algae, benthos, marine mammals, marine reptiles and fish). The data can be retrieved using the EMODnet biology download portal: (http://www.emodnet-biology.eu/toolbox/en/download/selection/15b9aaa5829edb).
Also integration of data the European animal Tracking network (ETN) is underway. Furthermore EMODnet Biology benefits from the portal set up by EMODnet Ingestion that helps providers to submit data.

Implementing AtlantOS agreed standards

The taxonomic standards and data formats used within this service to process and integrate the scattered marine biological datasets are based on the World Register of Marine Species (WoRMS), the authoritative and comprehensive list of names of marine organisms worldwide, and recommended standard by AtlantOS. Through the implementation of EurOBIS as marine biological data infrastructure of EMODnet Biology, data is processed following the Darwin Core Archive, an internationally recognised biodiversity informatics data standard that simplifies the publication of biodiversity data. Much effort went into the parameter standardization using the recommended BODC vocabularies. The common data structure and standards used consist of a DarwinCore (Dwc) Event Core in combination with a DwC Occurrence Extension and a proposed enhancement to the DwC MeasurementOrFact Extension. This new structure enables the linkage of measurements or facts - quantitative and qualitative properties - to both sampling events and species occurrences, and includes additional fields for property standardization (using the BODC Vocabularies). The data management uses ISO 19115-compliant metadata record for all the datasets that contribute to EurOBIS and EMODnet Biology.
Concerning EDMO codes to identify institutes in datasets, an action is ongoing to implement a mapping between the institutes linked to EurOBIS datasets and the EDMO codes.

Enhanced and new data services

The EMODnet Biology portal has an improved functionality to plot and view species observations, also for large marine areas. For example, the figure below shows the number of species observation records per grid cell (1° by 1°) for the North Atlantic Ocean. The number corresponds to the number of observations that were found within a grid cell. When zooming on the map, the size of the grid cells will automatically decrease to 30', 6' and finally the exact location of the actual observation (platform).

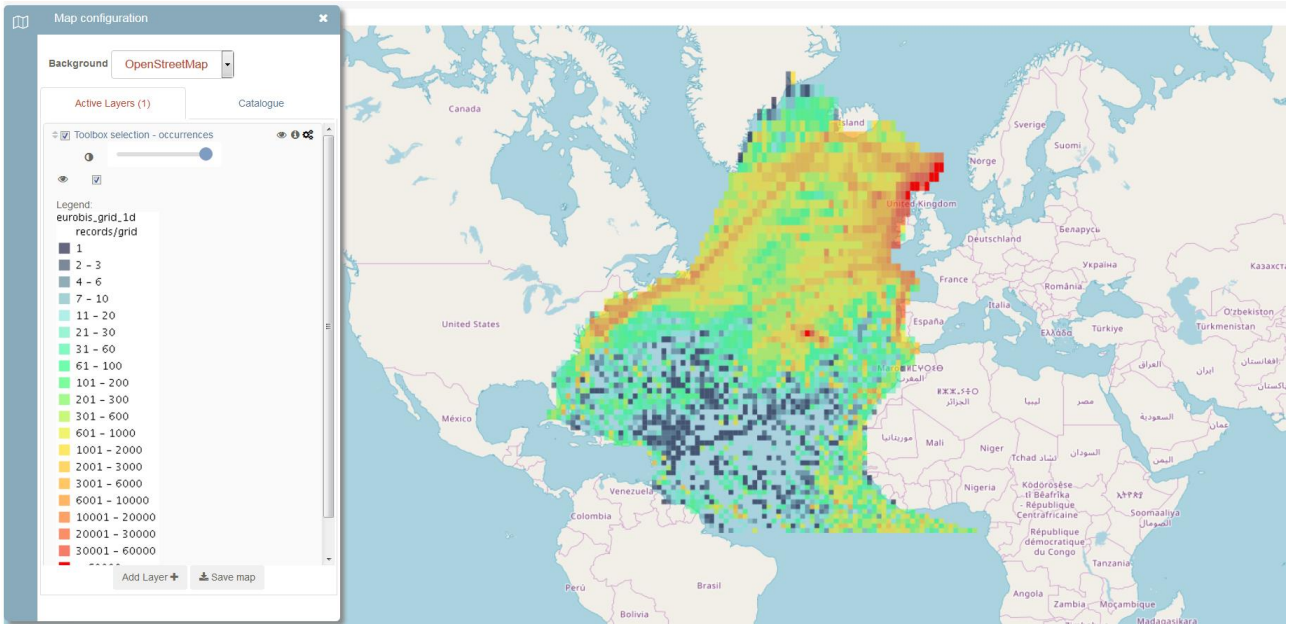


Figure 20: Mapping of species observations in the North Atlantic Ocean from the EurOBIS database, using the EMODnet biology Portal

The EMODnet Biology data download tool provides also the possibility to retrieve the species data as a WFS service. For example, a user can retrieve directly from the download portal of EMODnet Biology a web service that retrieves all the occurrence data of all marine species occurring within the IHO area North Atlantic Ocean.
The exact selection can also be stored as a JSON file. In this specific example, we only select by marine region (North Atlantic Ocean) where the polygon is described in the marine regions database at http://www.marineregions.org/gazetteer.php?p=details&id=1912.
We can possibly further select by many other fields as occurring in the Darwin Core data scheme (taxon, lat, long, depth, year, month, taxonomic group…).

7 DOI services at Network level

In-situ Ocean observing networks can assign DOIs to data for three main objectives. The first objective is to rationalize data citation, as a DOI is a direct and permanent link to the data used in a publication. The second one is to provide traceability on data usage, as the publications that cite a dataset with a DOI are easier to track in bibliographic surveys. The last one is to propose a simple data access service, as the landing page of a DOI can offer a direct access to a dataset through ftp or http links to files.

Recently one significant break to management of data DOIs for in-situ Ocean observing networks was reached with a methodology defined to handle data continuously updated in time. The Research Data Alliance (RDA) group dedicated to best practices on scientific data dissemination proposed a way to better manage DOIs on dynamic data. Although programs like Argo or Ocean Gliders succeeded to manage DOIs in an appropriate and efficient way, many of the In-situ Ocean observing networks are still in the phase to define what should be their strategy to manage DOIs for their network data.

Ifremer is the supporting WP7 partner for In-situ Ocean observing networks to define and set their data DOI strategy, and will add the definition of the strategies in the guidelines initialized within the AtlantOS project (see http://dx.doi.org/10.13155/44515) and promote them at the beginning of 2019 as best practices through oceanbestpractices.net/ initiated by IOC.

8 List of acronyms

API	Application Programming Interface
BGC	Biogeochemistry
BODC	British Oceanographic Data Centre
CDI	Common Data Index
CMEMS	Copernicus Marine Environment Monitoring Service
CPR	Continuous Plankton Recorder
DAC	Data Assembly Centre
DATAMEQ	DATA Management Exchange and Quality
DATRAS	DATabase of TRAwl Surveys
DBCP	Data Buoy Cooperation Panel
DM	Delayed Mode
DOI	Digital Object Identifier
ETN	European animal Tracking Network
EDMO	European Directory of Marine Organizations
EMODnet	European Marine Observation and Data Network
EV	Essential Variable
EurOBIS	European Ocean Biogeographic Information System
EuroGOOS	European Global Ocean Observing System
FTP	File Transfer Protocol
GDAC	Global Data Assembly Centre
GEOSS	Global Earth Observation System of Systems
GIS	Geographic Information System
GLODAP	Global Ocean Data Analysis Project
GOOS	Global Ocean Observing System
GO-SHIP	Global Ocean Ship-based Hydrographic Investigations Program
GPDR	General Data Protection Regulation
ICES	International Council for the Exploration of the Sea
ICOS	Integrated Carbon Ocean System
INS TAC	In-Situ Thematic Assembly Centre
IHO	International Hydrographic Organization
IOC	Intergovernmental Oceanographic Commission
ISO	International Organization for Standardization
JCOMM	Joint Technical Commission for Oceanography and Marine Meteorology
JCOMMOPS	JCOMM in-situ Observing Programmes Support Centre
JSON	JavaScript Object Notation
MBA	Marine Biological Association
NODC	National Oceanographic Data Centre
NRT	Near Real Time
OBIS	Ocean Biogeographic Information System
OGC	Open Geospatial Consortium
QC	Quality Control
ROOS	Regional Ocean Observing System
RT	Real Time
SOCAT	Surface Ocean CO_2 Atlas
SDN	SeaDataNet

SOOP	Ship of Opportunity Program
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VOS	Voluntary Observing Ship
WMO	World Meteorological Organization
WMS	Web Map Service
WoRMS	WOrld Register of Marine Species