<table>
<thead>
<tr>
<th><strong>Project</strong></th>
<th>AtlantOS – 633211</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deliverable number</strong></td>
<td>D3.13</td>
</tr>
<tr>
<td><strong>Deliverable title</strong></td>
<td>OceanSITES Data</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Harmonization and supervision of existing and new data streams across the OceanSITES biogeochemistry network.</td>
</tr>
<tr>
<td><strong>Work Package number</strong></td>
<td>WP3</td>
</tr>
<tr>
<td><strong>Work Package title</strong></td>
<td>Enhancement of autonomous observing networks</td>
</tr>
<tr>
<td><strong>Lead beneficiary</strong></td>
<td>NERC</td>
</tr>
<tr>
<td><strong>Lead authors</strong></td>
<td>Richard Lampitt (NERC)</td>
</tr>
<tr>
<td><strong>Contributors</strong></td>
<td>[list further contributors]</td>
</tr>
<tr>
<td><strong>Submission data</strong></td>
<td>14 September 2019</td>
</tr>
<tr>
<td><strong>Due date</strong></td>
<td>PM45</td>
</tr>
<tr>
<td><strong>Comments</strong></td>
<td>Additional work was required in order to present a clear description of the way in which this deliverable has been achieved.</td>
</tr>
</tbody>
</table>
### 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  
☐ NGO  
☐ others  
Please give the name(s) of the stakeholder(s):  
... |
| WHERE is/are the company(ies) or organization(s) from? | X Your own country  
X Another country in the EU  
X Another country outside the EU  
Please name the country(ies):  
These data are of global significance and not restricted to any particular country. |
| Is this deliverable a success story? If yes, why?  
If not, why? | X Yes, because without such coherent data management, its value would be completely lost.  
☐ No, because ..... |
| Will this deliverable be used?  
If yes, who will use it?  
If not, why will it not be used? | X Yes, by all those concerned with temporal and spatial trends in the properties of the ocean, particularly the so called Essential Ocean Variables.  
☐ 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.*
OceanSITES Data: Harmonisation and supervision of existing and new data streams across the OceanSITES biogeochemistry network

Background
The OceanSITES program is the global network of open-ocean sustained time series sites, being implemented by an international partnership of researchers. OceanSITES provides fixed-point time series of various physical, biogeochemical, and atmospheric variables at different locations around the globe, from the atmosphere and sea surface to the seafloor. The program’s objective is to build and maintain a multidisciplinary global network for a broad range of research and operational applications including climate, carbon, and ecosystem variability and forecasting and ocean state validation, and to provide data from all those stations.

All OceanSITES data are publicly available. More information about the project is available at: http://www.oceansites.org, with the OceanSITES mission statement, governance documents, and general contact information.

A major objective of AtlantOS has been to ensure that the streams of biogeochemical (and other) data from AtlantOS Eulerian Observatories are harmonised so that they are easily accessible by the public and members of the scientific community, particularly those who need to incorporate the data into computational models. As an international network, OceanSITES has been working to achieve this over the past two decades for the entire international network and under the fundamental data policy and principles outlined by CLIVAR. AtlantOS WP3 has made a significant and very positive contribution to this both with existing and new data streams.

The principles of data management are:
1. Free and unrestricted exchange
2. Timely exchange
3. Quality control
4. Metadata
5. Preservation of data
6. Plan for reuse in reanalysis
7. Easy access
8. Use of existing national and international mechanisms and centres

Data Flow
This is facilitated through three organizational units:
1. Principal Investigators (PI),
2. Data Assembly Centres (DAC), and
3. Global Data Assembly Centres (GDAC).

The normal procedure is that a PI provides the data and metadata information to a DAC, which formats this information into the OceanSITES file format and passes it on to the GDAC.
Accessing OceanSITES Data

There are two data Access Pathways

1. **Anonymous FTP:** Two GDACs provide access points for OceanSITES data via ftp servers. One is in France at the Coriolis centre (http://www.coriolis.eu.org), the other is in the US at the NOAA National Data Buoy Centre (NDBC, http://www.ndbc.noaa.gov). The two servers at the GDACs are synchronized at least daily to provide the same OceanSITES data redundantly. The user can access the data at either GDAC’s ftp site:
   - ftp://data.ndbc.noaa.gov/data/oceansites

2. **OPeNDAP/THREDDS:** The GDACs also provide OceanSITES data via a THREDDS:
   - http://dods.ndbc.noaa.gov/thredds/dodsC/data/oceansites/DATA/
   - http://tds0.ifremer.fr/thredds/CORIOLIS-OCEANSITES-GDAC-OBS/CORIOLIS-OCEANSITES-GDAC-OBS.html

Data Users Guide

The OceanSITES data users guide was developed by Dr Matthias Lankhorst at Scripps Institution of Oceanography, University of California San Diego (see: http://www.oceansites.org/docs/OceanSITES_DataUsersGuide.pdf)

Example of AtlantOS-supported existing data stream

It is not useful to provide a listing of OceanSITES biogeochemistry data streams which were supported by AtlantOS and to contrast these with data streams which were supported by other programmes or by the data service providers of member states which do, in fact, provide the majority of data. It is however useful to give some examples of data streams and systems which were massively improved as a result of AtlantOS support. The example chosen is from the sustained observatory over the Porcupine Abyssal Plain, PAP-SO http://projects.noc.ac.uk/pap/

The PAP observatory is situated in the Northeast Atlantic away from the continental slope and Mid Atlantic Ridge. The site is an open ocean time-series representing processes in the North Atlantic Drift Region and accessible from many EU ports and has had an observatory of increasing significance and sophistication for the past 30 years with only a few gaps in data supply.

Latitude and Longitude: 49.0°N 016.5°W

Depth: 4850m

Data are obtained on meteorology from the deck of a surface buoy managed by the UK Meteorological Office (UKMO), from sensors and samplers attached to the keel of the buoy, from sensors and samplers on an instrument frame at 30m depth, from sensors and samplers at 3000 and 4700m depth and from sensors above the seabed at 4850m depth.

http://tds0.ifremer.fr/thredds/catalog/CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/catalog.html

Here we present an example of one particular 2019 data stream which is for CO2, a variable which does, of course, have massive environmental significance:
Dataset: PAP/OS_PAP-1_201806_R_PCO2.nc

- **Data size:** 28.69 Kbytes
- **ID:** CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/OS_PAP-1_201806_R_PCO2.nc

**Documentation:**


**Access:**

1. **OPENDAP:** [/thredds/dodsC/CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/OS_PAP-1_201806_R_PCO2.nc](/thredds/dodsC/CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/OS_PAP-1_201806_R_PCO2.nc)
2. **HTTPServer:** [/thredds/fileServer/CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/OS_PAP-1_201806_R_PCO2.nc](/thredds/fileServer/CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/OS_PAP-1_201806_R_PCO2.nc)
3. **SOS:** [/thredds/sos/CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/OS_PAP-1_201806_R_PCO2.nc](/thredds/sos/CORIOLIS-OCEANSITES-GDAC-OBS/deployment_data/PAP/OS_PAP-1_201806_R_PCO2.nc)

**Dates:**

- **2019-06-24T02:10:04Z (modified)**

**Variables:**

- **Vocabulary [CF-1.0]:**
  - PSAL (psu) = sea_surface_salinity
  - TEMP (Celsius) = sea_surface_temperature
  - DOXY (percent) =

**GeospatialCoverage:**

- **Longitude:** -180.0 to 179.5 Resolution=0.5 degrees_east
- **Latitude:** -77.01048 to 89.89626000000001 Resolution=0.5 degrees_north
- **Altitude:** 0.0 to 2000.0 meters (positive is down)

**Properties:**

- aggregation-level = "product"

**Viewers:**

- [NetCDF-Java ToolsUI (webstart)](netcdf.java/toolsui)
New Data streams supported by AtlantOS

In addition to significant progress in harmonising existing data streams, AtlantOS has also stimulated some new data streams which will, in future, be of major significance although are not currently listed as Essential Ocean Variables (EOVs) by the Global Ocean Observation System (GOOS). The most significant of these has been in the area of so called ‘Omics’, which has been reported in D3.17. This neologism “omics” informally refers to a field of study in biology ending in -omics, such as genomics, proteomics or metabolomics. Omics aims at the collective characterization and quantification of pools of biological molecules that translate into the structure, function, and dynamics of an organism or organisms.

By way of background, the GOOS Biology and Ecology Panel has recognised the need to account for microbial life when reporting on the ocean’s essential ecology. Investigations and assessments of the biomass and diversity of marine microbes have shown great promise in the creation of bio-indicators for a wide range of phenomena, and also in developing our basic understanding of how marine ecosystems function and respond to stressors. Global efforts to provide long-term solutions for monitoring microbes are rapidly increasing in their readiness, however, these lack an overarching framework and coordination body to align activity, methods, and information products to address basin-scale and global needs. Further, large expanses of the surface ocean and almost all of the deep ocean are not observed over time, and no baselines are available to detect key environmental change.

In addition to an underdeveloped framework, the following key gaps exist:

1) the lack of field-tested, autonomous technologies to extend the reach of omics sampling to environments such as the Arctic and deep sea,
2) a lack of FAIR omics data and metadata products and standards that are tuned to the needs of omics observatories and their scientific and societal missions,
3) a lack of cyberinfrastructure dedicated to observatory-grade activities and associated dissemination points for distributed and linked data,
4) a scientific evaluation of the significance of omics-based observations of microbial communities for time-series observations of ecosystems, and
5) a lack of intercalibration between observatories using rapidly evolving omics technology.

Activities in task 3.2 to build capacities for omics-based observations were mainly carried out by partners AWI, Riboc, and NOC and focused on approaches for a monitoring of microbial communities in the Atlantic Ocean and beyond. Activities include the assessment, optimization, and development of technologies for sampling and sample preservation, the demonstration of the feasibility of omics approaches for time-series observations of microbial communities and their use to address connections to biogeochemical processes and element cycling, a comparison of existing methodologies for lab- and bioinformatics analyses, as well as the implementation of data management and analysis procedures to improve data accessibility and integration. Further, these partners have created a Global Omics Observatory Network (GLOMICON) to fill the need of a global coordination framework to integrate activities in this emerging field. This activity will be enhanced by the AWI partner’s new role as the lead of the emerging Microbial EOV (which heavily relies on omics) on behalf of GOOS.

From the perspective of AtlantOS D3.13, and the issue of data harmonisation the most significant development has been in GLOMICON which has a mission to federate omically enabled observatories and create an integrated, global system of multi-omic monitoring. This will enhance our capacity to understand, investigate, and monitor the biosphere in a fundamental manner and has been massively supported by AtlantOS: https://sites.google.com/view/glomicon/home
Apart from the obvious extreme technical demands of this objective, the data obtained are large and complex and substantial effort has been made to ensure that from the outset there are appropriate systems established to ensure the data are harmonised and can therefore be widely compared subsequently.

As reported in D3.17, omics-based approaches produce large amounts of raw data (e.g. DNA and/or RNA sequences) which can be analyzed with a wide variety of bioinformatics tools and methods. The results derived from these tools can vary considerably, and a lack of coordination of methods across omics-observatories hinders inter-comparability, interoperability, and consolidated reporting. Mainstreaming omics in global assessment and monitoring requires greatly improved harmonization of omics data products in what is still largely an innovation- and research-driven field. This is especially true as more marine observatories begin to deploy omics technologies in operational settings, often creating local and ad hoc data standards which make data integration across observatories difficult. Indeed, without a core set of common tools and reference resources (reference gene catalogues, taxonomies, etc.) operational usage of omics technologies loses practicality.

Three key outcomes of AtlantOS WP3 have addressed these needs and have contributed to a firm foundation for future improvements and network building:

1. Through GLOMICON (compare AtlantOS deliverable D6.4), the omics network seeded by AtlantOS, AWI and Ribocon, with support from external partners, have launched strategic efforts to harmonize the data-scape, information resources, and knowledgebase of the global omics observatory community and link them with other key information resources and services:
   a. In cooperation with GFBio (www.gfbio.org; seeded by the FP7 project Micro B3: www.microb3.eu), these partners have established an online registry of omics observatories, harvesting key metadata about the observatories themselves (glomicon.org/registry). This is linked to a web page to improve visibility of GLOMICON’s mission, coordination resources (see below) and activities.
   b. A coordination space for key bioinformatics code and software tools has been launched on GitHub (github.com/GLOMICON) to support collective development and testing in the community.
   c. A coordination space for established and developing omics sampling and field protocols (which are key metadata for downstream in silico analyses) has been established in protocols.io (protocols.io/groups/glomicon). This allows the network to avail of protocol.io’s advanced document management, versioning, and sharing capabilities. From this growing pool of protocols, we will submit a candidate “best practice” workflow to the UNESCO/IOC Ocean Best Practices System (WP6).
   d. An ontology – a human and machine-readable knowledge representation which addresses the “I” in FAIR – for the omics observing domain has been established using the best practices of the Open Biological and Biomedical Ontologies Foundry and Library (github.com/GLOMICON/omicon) and will be linked to the development of ontology resources for the GOOS EOVs (initiated in WP6, with endorsement from GOOS) as the community matures.
e. AWI has initiated a GLOMICON-driven extension to the metadata standards of the Genomic Standards Consortium (GSC) to support omics in the cryosphere/marine realm. “MiXS-cryo” is now gathering input from experts and a manuscript is being refined for publication.

f. A meeting (December, 2018) at the European Bioinformatics Institute between representatives of the Global Biodiversity Information Facility (GBIF), the European Nucleotide Archive (ENA, an INSDC node), SILVA, and GLOMICON secured an agreement on the implementation of data exchange between omics data stores and global biodiversity occurrence aggregators like GBIF. This major step in mainstreaming omics will be pursued in Q1 of 2019, and announced at two workshops pending acceptance of session proposals (submitted).

g. A satellite meeting at the 21st meeting of the GSC (May, 2019) has been organized to discuss the merger GLOMICON with the US-based Genomic Observatory Network (genomicobservatories.org), with particular focus on data and metadata harmonization through resources such as GeOMe (geome-db.org), to further consolidate omics observatories on a global scale.

2. Building on the above efforts (particularly 1b and 1c), Ribocon has established a “baseline” bioinformatics analysis procedure for data exchanged by GLOMICON members. This will create an anchor for intercomparison and coordination between observatories using different methods as the community develops standard practices. This pipeline can be applied to raw sequences from all partners, and adopts the well-established SILVA NGS data analysis pipeline.

3. To systematically demonstrate the utility of 2 within the informatics framework established in 1, an international intercomparison study coordinated by AWI and MBARI was initiated at the AtlantOS and Genome Canada/Agriculture and Agri-Food Canada Workshop on enhancing interoperability & coordination of long-term omics observations (January 2018). Six omically enabled observatories and institutions (inside and outside AtlantOS) joined the experiment and allocated funds for lab and bioinformatics analyses. Sampling is near complete and distribution of samples among participants has been planned. The sequencing and first round of bioinformatics analyses have taken place. The method comparison focuses on analyses of protist communities by means of amplicon sequencing and includes the entire pipeline from DNA extraction to bioinformatics analyses. Environmental samples from different regions and DNA extracts from mock communities are being exchanged between participants for replicate analyses according to the procedures established at partner institutions. The results of this exchange will be needed to estimate the stability of data in our setting, and allow us to test the effectiveness of data exchange and FAIR compliance. The approach is introduced in AtlantOS D6.5. The activity represents a first step towards harmonization across observatory programs, which represents a key requirement by the FOO for any mature contribution to EOV observations as part of the Global Ocean Observation System (GOOS).

In conclusion, through the coordination of data-focused activities in close association with global experimental activities, the AWI and Ribocon partners have leveraged and added significant momentum to existing efforts to enhance FAIR data exchange in omics observatory. There is still work to be done in order secure sustained and robust data coordination. However, with the major
support from AtlantOS, an extensible foundation, supported by growing community endorsement, has been laid and linked to GOOS activities (e.g. see Section 2.9 in GOOS report #232). As AtlantOS concludes, multi-lateral efforts to sustain this work are being pursued by both internal and external partners.