ATLAS Deliverable 8.4:
Delivery of Research outputs

Project acronym:	ATLAS
Grant Agreement:	678760
Deliverable number:	**D8.4**
Deliverable title:	**Delivery of research outputs:** Report that assesses tools and methods for improved delivery of research outputs to EMODnet and other major dissemination and engagement platforms for re-use and re-assembly of data and information products.
Work Package:	**WP8**
Date of completion:	**15 May 2020**. An extension from 30 April until maximum 15 May was agreed with the project officer, to accommodate for Covid-19 affecting the working schedule for some key authors with young children at home. On 28 April SBE (lead author) sent a very mature draft for review to WP8 lead and BGS (as key author). Comments were received by 30 April and the D8.4 was updated with review comments by 6 May 2020, sent for Project Coordinator review and then submitted.
Authors:	**Authors:** Tim Collart (SBE), Kate Larkin (SBE), Stéphane Pesant (UniHB), Joana Gafeira (BGS) **Acknowledgments:** Antonio Novellino (ETT, EMODnet Physics), Patrick Gorringe (SMHI, EMODnet Physics); Joana Beja (VLIZ, EMODnet Biology), Leen Vandepitte (VLIZ, WoRMS, EMODnet Biology); Dick Schaap (MARIS, EMODnet Data Ingestion), Ioana Athanasia (Sissy) (HCMR, EMODnet Data Ingestion); Helen Lillis, Natalie Askew, Eleonora Manca (JNCC, EMODnet Seabed Habitats), Mickael Vasquez (Ifremer, EMODnet Seabed Habitats); Jan-Bart Calewaert, Pascal Derycke*, Conor Delaney (SBE, EMODnet Secretariat) *until December 2019

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 678760 (ATLAS). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Contents

1 Executive Summary

Marine data are needed for many purposes: for acquiring a better scientific understanding of the marine environment, but also, increasingly, to provide information and knowledge to support ocean and coastal economic developments and underpin evidence-based ocean and wider environmental management decision making. Data must be of sufficient quality and at the right resolution to meet the specific users' needs. They must also be accessible in a timely manner and in appropriate formats – not only in raw data but as integrated datasets, data products, etc. – for use by marine and maritime professionals. Such expert users span scientific research, policy and industry. In addition, providing engaging and user-friendly interfaces and tools for wider society to explore marine data and information e.g. through visualisations, is vital to promote a knowledge-driven, ocean literate society. In addition, the blue economy, policy makers, researchers and wider society increasingly require data that are Findable, Accessible, Interoperable and Reusable (FAIR) across multiple parameters, spatial scales and resolutions. Many data services and initiatives already exist in Europe and there is a drive towards collaboration and interoperability of these to ensure data can be discovered through web services by human queries and through machine-to-machine communication.

This ATLAS deliverable (D8.4) is driven by the philosophy of Open Data and Open Science, adding value to the diverse datasets produced by ATLAS, making them more FAIR and so, ultimately, increasing their long-term use and impact. To this end, project partner Seascape Belgium (SBE) provided and customised a web-GIS Platform for the ATLAS project. Using an open source geospatial content management system – GeoNode – the ATLAS GeoNode was developed as a tool to share, visualise and download geospatial data with the ATLAS consortium and wider stakeholders. In addition, ATLAS data and data products are being ingested into the European Marine Observation Data Network (EMODnet)[1] as a long-term solution to data availability, discovery and use. This report summarises the work conducted by SBE, in collaboration with University of Bremen (UniHB) and the PANGAEA[2] information and data publisher for earth and environmental data, British Geological Survey (BGS) and others partners, to valorise the marine data being produced by ATLAS, namely building on existing methods and tools to add value, use and impact of marine data along the pipeline from data production to end-user. This contributes in particular to the 3rd key objective of ATLAS, to transform new data, tools and understanding and make it accessible to wider stakeholders for effective ocean governance. To achieve this, SBE has worked together with UniHB (as data management and WP8

[1] www.emodnet/eu
[2] https://www.pangaea.de/

lead) and BGS to assess, optimise and – where possible – innovate the data flows in place. A key focus has been at the mid-point of the "data pipeline", where curated data can be 'valorised' through methods including data visualisation and data integration, to make them more accessible to multi- and inter-disciplinary research communities and to wider stakeholders including policy and industry. SBE administers the EMODnet Secretariat, and so has been able to facilitate direct dialogues between EMODnet Data Ingestion and the seven thematic areas of EMODnet (Bathymetry, Biology, Chemistry, Geology, Human Activities, Physics and Seabed Habitats) with ATLAS data providers to ensure a longer-term ingestion of data into EMODnet.

As a North Atlantic basin scale project with strong industry partnerships ATLAS has offered an opportunity to assess data flows and pipelines from major research activities and projects via existing data publishers and assembly centres to EMODnet, and to recommend further ways to optimise these in the future. This report also looks at the relevance of ATLAS data and outputs to policy and industry, including recommendations from meetings and consultations conducted by ATLAS WP6 and WP7. These include recommendations from ATLAS D6.4 that a desire from offshore maritime industry to see greater connectivity and interoperability between marine data to increase their impact and use and to streamline the process of marine data discovery, uptake and exploitation.

Particular focus has also been dedicated to investigate the flow of data from PANGAEA data publisher to EMODnet. This has resulted in stronger collaborations between the two initiatives, leading to more systemic and operational exchanges in data flows, including a move towards automated data harvesting. The project has also offered an opportunity to develop an innovative online GIS platform as a community tool for sharing and integrating geospatial data. This was developed as a pilot and the positive user feedback shows its potential for making data 'come alive', connecting it to wider stakeholders and offering useful maps and products which marine and maritime professionals can use for their professional needs e.g. marine spatial planning.

Recommendations from this report in terms of data stewardship and data flows can be taken forward by marine data initiatives and by the marine research community in the future. The advances that have been taken in ATLAS towards FAIR data are important steps towards streamlining the ingestion of data into EMODnet. In EMODnet, data are discoverable through data and web services, contributing to the European Union's policy on marine knowledge, the "Marine Knowledge 2020" initiative. Here, EMODnet has a key mandate to transform Europe's fragmented data landscape into an interoperable sharing framework, in addition to supporting coordinated European observation activities. This will increases the information available, and therefore the efficiency, for marine and maritime professionals from industry, public authorities and academia to discover and use marine data,

information and knowledge. This encourages innovation that reduces our present uncertainty as to what is happening beneath the sea surface. Beyond 2020, EMODnet is working with key data initiatives to federate existing infrastructure and contribute to a Blue-Cloud cyber platform[3] that will offer enhanced capabilities for marine research including a virtual research laboratories, computational power and storage and the latest data discovery and interoperability to access data from a large diversity of data initiatives and data providers.

2 Introduction: Valorising marine data and information

2.1 Overview of D8.4

SBE lead this task, with inputs from UniHB as Task contributors and overall lead of WP8 on open science and research outputs, and BGS for data and metadata management and support in the ATLAS GeoNode. The ultimate goal of this report is to demonstrate how Work Package 8 has worked with existing data repositories, data publishers and longer-term EU data services to ensure a legacy for the wide diversity of ATLAS data and research outputs that promotes Findable, Accessible, Interoperable and Discoverable (FAIR) data, metadata and information for users in Europe, and beyond (see Figure 1 below and D8.3 for more information).

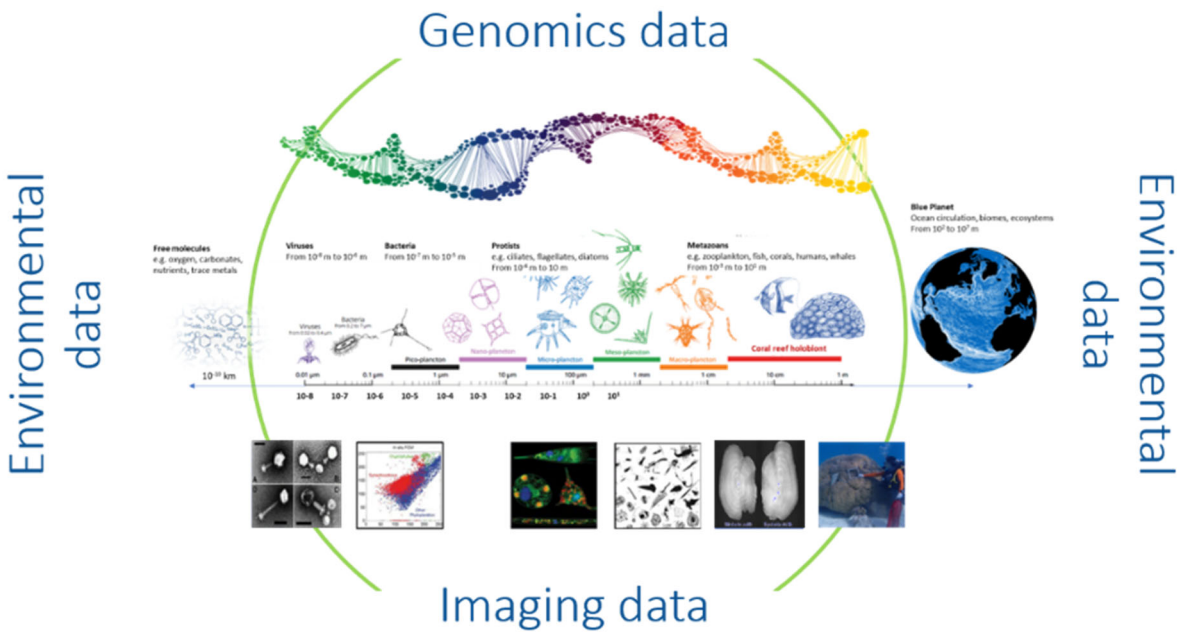

Figure 1: ATLAS research data (adapted from Stéphane Pesant – presentation to the WP6 Industry workshop 2019)

[3] EU H2020 project Blue-Cloud: https://www.blue-cloud.org/

D8.4 assesses the tools and methods used in the ATLAS project for improved delivery of research outputs for long-term discovery to the European Marine Observation and Data Network (EMODnet). Examples of the existing processes for data discovery between open data services e.g. EMODnet and PANGAEA are also presented, together with recommendations for next steps e.g. automated harvesting.

It also looks at novel tools that were developed in WP8 Task 8.5 and led by SBE. For example, the ATLAS GeoNode is a concrete example of an existing open source software that was customised and adapted to the needs of the ATLAS research community, as a novel method to share geospatial data during the course of a marine research project. Implemented in 2019 onwards, the ATLAS GeoNode demonstrates the potential for enhancing the exchange of (geospatial) data and information between project partners during the project life-time, but also improves the promotion, communication and ultimately the transfer of ATLAS research outputs to science, policy and industry stakeholders. In addition, the transfer of research outputs to other major dissemination and engagement platforms e.g. the European Atlas of the Seas[4], has been investigated with concrete developments for making ATLAS data accessible and available for re-use and re-assembly of data and information products, in combination with other open data from multiple sources e.g. EMODnet, Copernicus Marine Service[5] (CMEMS) and the International Council for the Exploration of the Sea[6] (ICES).

This report is set in the wider context of Task 8.6 (Assess the suitability of WP8 methodology for transfer of Atlantic ecosystem-based research outputs on the long-term), with UniHB contributing a section on the OpenAIRE infrastructure (that includes Zenodo), with updated statistics on ATLAS research outputs compared to other marine and non-marine projects, and an assessment of the use of OpenAIRE as an open science tool. This is being further elaborated by SBE, with input from UniHB and dialogue with ATLAS data producers and wider experts e.g. EMODnet Data Ingestion. This deliverable is therefore related to D8.3, a synthesis of ATLAS research outputs available in open access (See Table 2 therein for a list of datasets and wider outputs, linked to Task 8.4 Monitor and report on ATLAS research outputs using OpenAIRE). This is led by UniHB with the same delivery time (M48). SBE and UniHB have exchanged information for both D8.3 and D8.4 to ensure the full list of research outputs (data and other outputs) is comprehensive and identifies further long-term discovery

[4] http://www.european-atlas-of-the-seas.eu/
[5] https://marine.copernicus.eu/
[6] https://www.ices.dk/Pages/default.aspx

possibilities e.g. data relevant to the 7 thematic areas of EMODnet (see further information in this report and Annex I).

2.2 Increasing the long-term availability, discovery and impact of ATLAS data for users in Europe, and beyond

ATLAS project Work Package 8 adds value to the diverse datasets produced by the project, making them more FAIR and, ultimately, increase their long-term use and impact. A key issue with large-scale integrated research projects can be the limited sharing of data during a project lifetime, due to data restrictions by data providers and the lack of a suitable platform to share and visualise data. In addition, much of the ATLAS project data are geospatial and there was a need to share data and visualise it to meet research needs. To address these issues, SBE developed a customised application of the GeoNode, an open source geospatial content management system, to provide geospatial data visualization for the project community (section 5). An additional benefit of GeoNode is that it brings together mature and stable open-source software under a consistent and easy-to-use interface allowing users, with little training, to quickly and easily share data and create interactive maps. This was also an invaluable tool to showcase the spectrum of geospatial data produced by ATLAS to wider stakeholders, and to enable other open source data to be discovered and visualised through remote services.

Figure 2 shows the FAIR data principles and also the multiple best practices and components that are required from the data production to the data and metadata curation and management stage that enable the provenance of the data to be traceable and fully described, acknowledging the data provider together and including a persistent identifier such as a Digital Object Identifier (DOI).

Figure 2: Best practices to ensure data is FAIR – Findable, Accessible, Interoperable and Reusable

Scientific marine research data are often archived in national data repositories and from there, the data are connected and further standardised, assembled and integrated into more comprehensive datasets for particular regions (e.g. regional seas) and at wider (e.g. pan-European) level. There are a large number of data management experts working at national, regional and European levels for institutions, initiatives and networks that are the foundation for the "data pipe-line" that curates the data and ensures the data are coupled with metadata that describe the data, its provenance and quality, format etc.

The European Marine Observation and Data Network (EMODnet) is a long-term key marine data initiative of the European Union, funded by the European Commission's Directorate for Maritime and Fisheries (DG MARE) since 2009 (see Figure 3 and Figure 4). EMODnet already collaborates with many key data repositories e.g. national marine data centres, data publishers such as PANGAEA and wider European initiatives such as SeaDataNet and the Copernicus Marine Service (CMES) and the International Council for the Exploration of the Seas (ICES). The EMODnet central portal offers a single gateway to Europe's diverse *in situ* marine data, related metadata and data products, underpinned by seven thematic portals: bathymetry, geology, physics, chemistry, biology, seabed habitats and human activities. EMODnet is driven by the philosophy to collect data once and use it many times. This adds value to individual datasets through standardisation and integration, data visualisation and data products and improving the user experience with data and web services and machine-machine readable metadata. This relies on a large network and community of more than 150 organisations, including existing marine data management centres at national level (e.g. national data management centres such as the British Oceanographic Data Centre, BODC), regional level (e.g. Regional Sea Convention data collation efforts), pan-European marine data infrastructures (e.g. SeaDataNet), and

the expert networks that underpin each EMODnet thematic area. Crucially, EMODnet not only offers integrated datasets but the expert networks of the thematic assembly groups produce added value data products that are driven by user-demand (see Figure 3) such as vessel density maps (EMODnet Human Activities, including collaboration with the European Maritime Security Agency) or marine litter maps (EMODnet Chemistry, in collaboration with the European Environment Agency, Joint Research Centre of the EC and Regional Sea Conventions for use in Member State reporting for the Marine Strategy Framework Directive). Data are always paired with metadata set to minimum requirements to adhere to European standards (e.g. those set by the INSPIRE Directive for geospatial data). The data and web services also ensure more discoverability of data by people and machine-machine through remote services. EMODnet marine data and data products are indispensable to an increasingly diverse stakeholder community, as a key component in blue economy industry operations and enabling new knowledge about the marine environment to underpin evidence-based policy making e.g. Marine Spatial Planning.

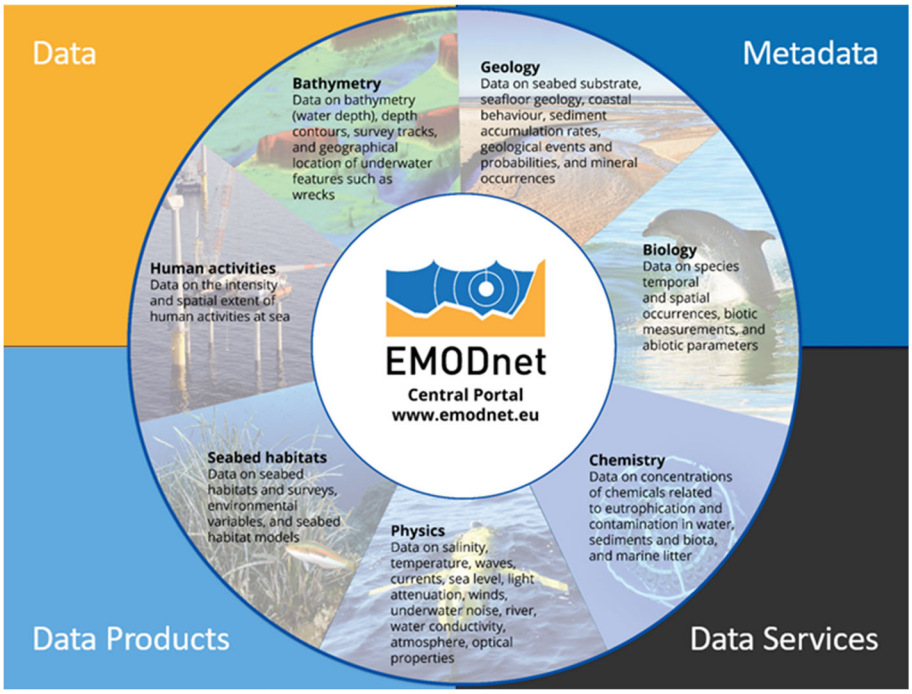

Figure 3: EMODnet services in 7 thematic domains

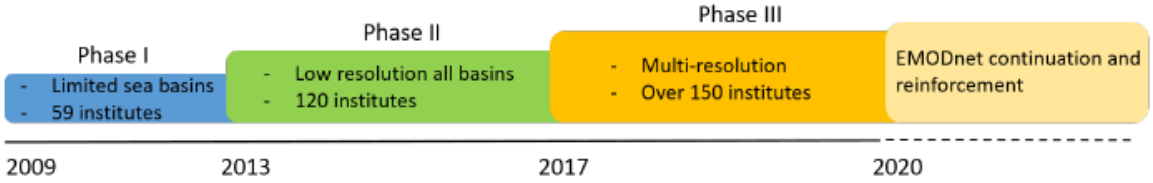

Figure 4: EMODnet evolution. More information on next phase EMODnet at: https://www.emodnet.eu/next-phase-emodnet

This report is an internal assessment of the data flow within ATLAS, with a particular focus on the flow between PANGAEA and EMODnet (section 4). This includes an evaluation of methods and bottlenecks, conducted in dialogue with project partners UniHB and BGS, and recommendations for optimising operational data flows between these marine data initiatives that can be taken forward beyond the life-time of the ATLAS project. In addition, the longer-term ingestion of ATLAS data into EMODnet is being discussed with EMODnet Data Ingestion, a service that was set up during the life-time of the ATLAS project and provides a dedicated service to ingest datasets from all data producers, including research projects, industry etc.

3 ATLAS Open Science Resources

OpenAIRE[7] is the overarching infrastructure funded by the European Commission to explore, monitor and connect the research outputs of its research communities. OpenAIRE harvests essential metadata that allow simple, but informative statistics and metrics about research outcomes, such as publication types, publication routes, access modes, co-funding and collaborations across research projects. Other open science resources funded by the European Commission, such as the European Open Science Cloud, are currently under development and will offer additional benefits to research projects.

As outlined in its Grant Agreement and in its Data Management Plan, project ATLAS had selected OpenAIRE as one of the building blocks to facilitate open science. Through the OpenAIRE-Connect initiative, project ATLAS acted as a pilot community to help evaluate and develop such services. Best practices were adopted from the very beginning of project ATLAS to ensure that all research outputs were properly acknowledged and made available in open access, including institutional repositories and ZENODO, the European open access repository. Here we outline some statistics and metrics made available on the OpenAIRE portal about project ATLAS.

[7] https://www.openaire.eu/

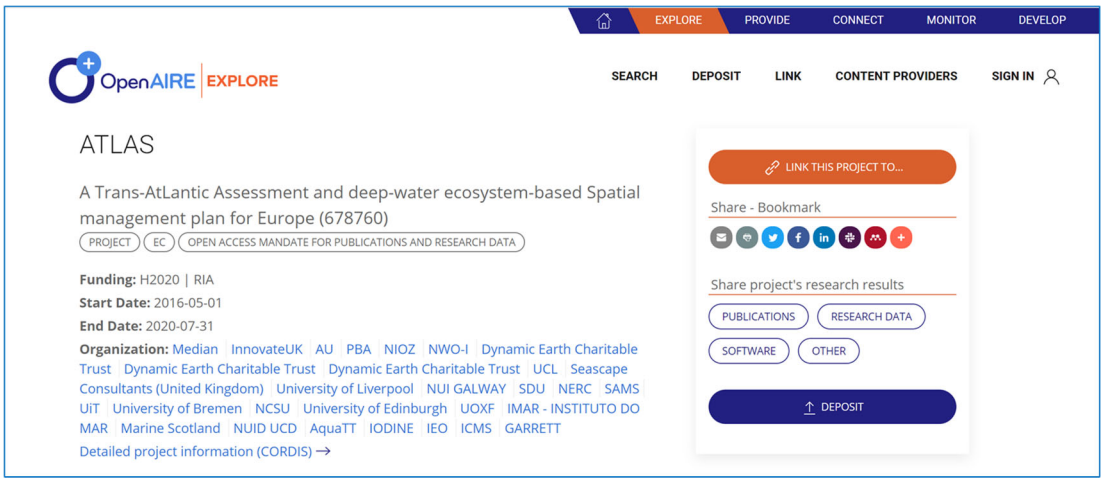

Figure 5: Landing page of project ATLAS on the OpenAIRE infrastructure portal.

At the time of submitting this report (2020.05.15), a total of 469 research outputs acknowledging project ATLAS had been harvested by OpenAIRE from eleven content providers (Figure 6). They consist of various types of research output, including 61 (13%) peer-reviewed articles[8], 89 (19%) data sets, four software, one video, and 315 (67%) other types of research outputs such as deliverables, presentations, posters, cruise reports and newsletters. These research outputs are available in open access (93%), restricted access (4%), under embargo (2%) or in closed access (1%).

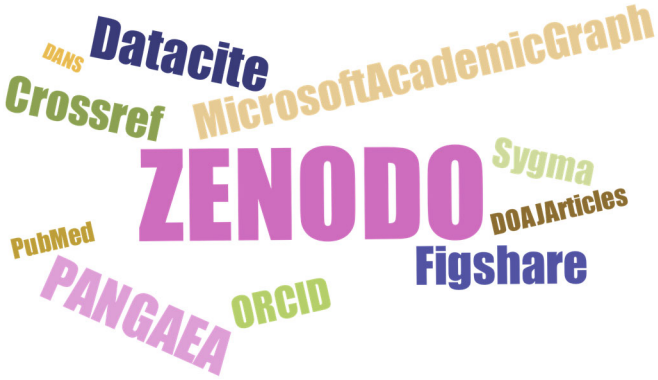

Figure 6: ATLAS Research outputs were collected by OpenAIRE from 11 content providers: ZENODO (273), PANGAEA (62), Datacite (50), Figshare (43), Microsoft Academic Graph (38), Crossref (38), ORCID (29), Sygma (25), DOAJ-Articles (11), PubMed Central (10), DANS (5)

Metrics harvested from the different providers indicate that ATLAS research outputs had 6,674 views and 4,430 downloads in local repository. ATLAS peer-reviewed articles provide evidence of co-funding by the Research Council UK (10 articles), Academy of Finland (3 articles), Science Foundation Ireland

[8] At the time of writing the ATLAS project consortium had published 85 peer-reviewed papers with an updated list available on the project website (https://www.eu-atlas.org/resources/atlas-library), of which 61 are available on OpenAIRE

(3 articles) and the Netherlands Organisation for Scientific Research (2 articles). Furthermore, they show synergies between ATLAS and 15 other projects funded by the European Commission (Figure 7).

Figure 7: A total of 47 research outputs from project ATLAS show synergy with 15 other projects funded by the European Commission: HADES (10), SponGES (9), CODEMAP (4), Blue-Action (4), ASSEMBLE (3), MERCES (3), MIDAS (3), NACLIM (3), OpenAIRE-Connect (2), AtlantOS (1), BLUE MINING (1), CALMARO (1), CORALFISH (1), EMSODEV (1), FIXO3 (1).

4 Data flows from PANGAEA to EMODnet

PANGAEA is a long-term data publisher for Earth and Environmental Science hosted by Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI) and the Centre for Marine Environmental Sciences, University of Bremen (UniHB). As lead of WP8, UniHB has produced a data policy and submission guidelines to ensure the majority of ATLAS project data are archived in PANGAEA. In PANGAEA, datasets were labelled as belonging to the ATLAS project and were assigned a Digital Object Identifier (DOI) in order to allow for long-term discovery of the produced resources.

In order for ATLAS data archived in PANGAEA to be maximally FAIR (Findable, Accessible, Interoperable and Reusable), relevant data sets are also making their way to the European Marine Observation and data Network (EMODnet). This enables ATLAS data to reach the wider community of marine scientists, industry stakeholders and policy makers.

The flow of marine data from PANGAEA to EMODnet is not only relevant for ATLAS data but also has a much wider scope. Therefore, we have evaluated the existing data flows to see what processes were already in place for transferring data from research projects and existing open source data repositories to EMODnet and recommended next steps. Figure 8 shows the 5 key data repositories involved in the

ATLAS data curation, namely PANGAEA, SEANOE, SeaDataNet, Zenodo and ENA (see D8.3 for more information).

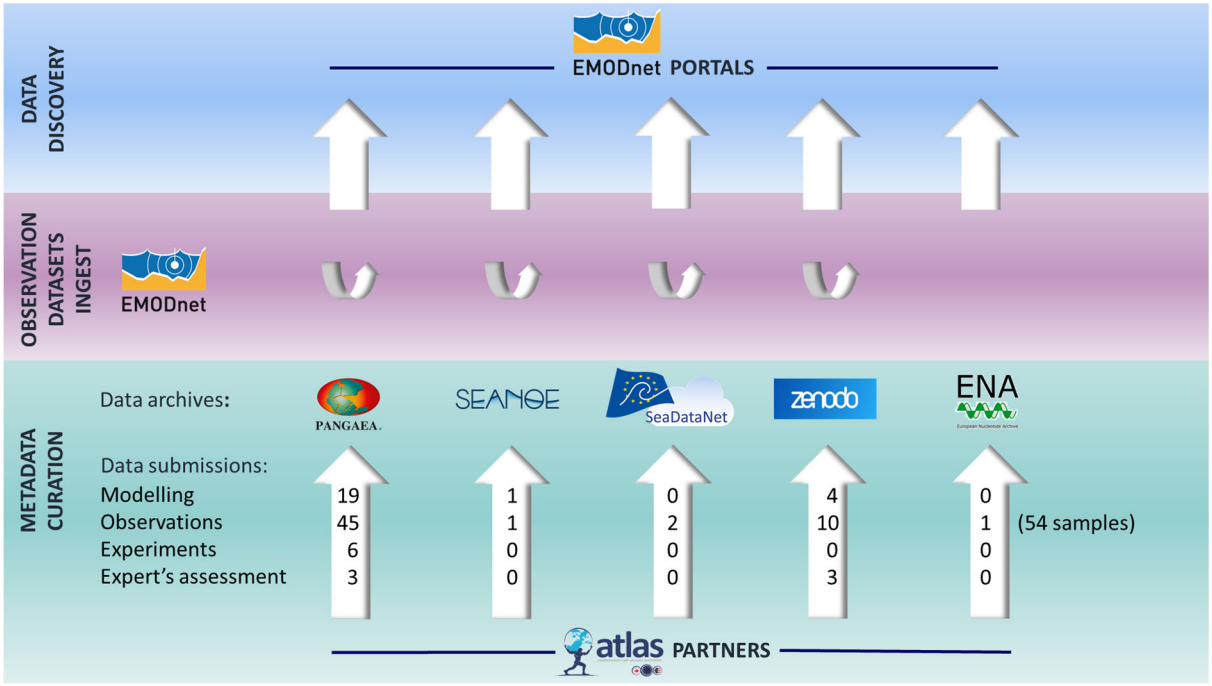

Figure 8: Flow of data sets from their deposition by ATLAS partners in selected data repositories, metadata curation by the archives, to their upload in EMODnet. Note that the latter step of the workflow is currently in progress (see Section 6.2).

4.1 Physical oceanography data

A semi-automated data flow of physical oceanography data (predominantly CTD data) from PANGAEA to EMODnet Physics was realised within the framework of the Southern Ocean Observing System (SOOS)[9] and later expanded to EMODnet Physics[10] in general. This flow has so far managed to make 481,119 CTD records archived in PANGAEA available through the SOOS and the EMODnet Physics data portals.

The process for the semi-automated dataflow is displayed graphically in Figure 9 and described below. The data flow is constrained to a limited set of PANGAEA datasets which were labelled as "EMODnet" to indicate they can be harvested. A list of metadata records for the datasets within this set[11] are generated by the Open Archives Initiative - Protocol for Metadata Harvesting (OAI:PMH) service offered by PANGAEA. For each returned metadata record and each event[12] within that record, the

[9] http://www.soos.aq/
[10] https://www.emodnet-physics.eu/
[11] https://ws.pangaea.de/oai/provider?verb=ListRecords&metadataPrefix=pan_md&set=EMODnet
[12] In Pangaea, an event holds the metadata for a site or station where a measurement or sampling has been performed. For additional information see https://wiki.pangaea.de/wiki/Event

following flow is performed. The dataset is downloaded and the DateTime information and Geographical location (longitude & latitude) is determined, either from the data table itself or from the event metadata. Per DataTime, a depth value is determined from the data table.

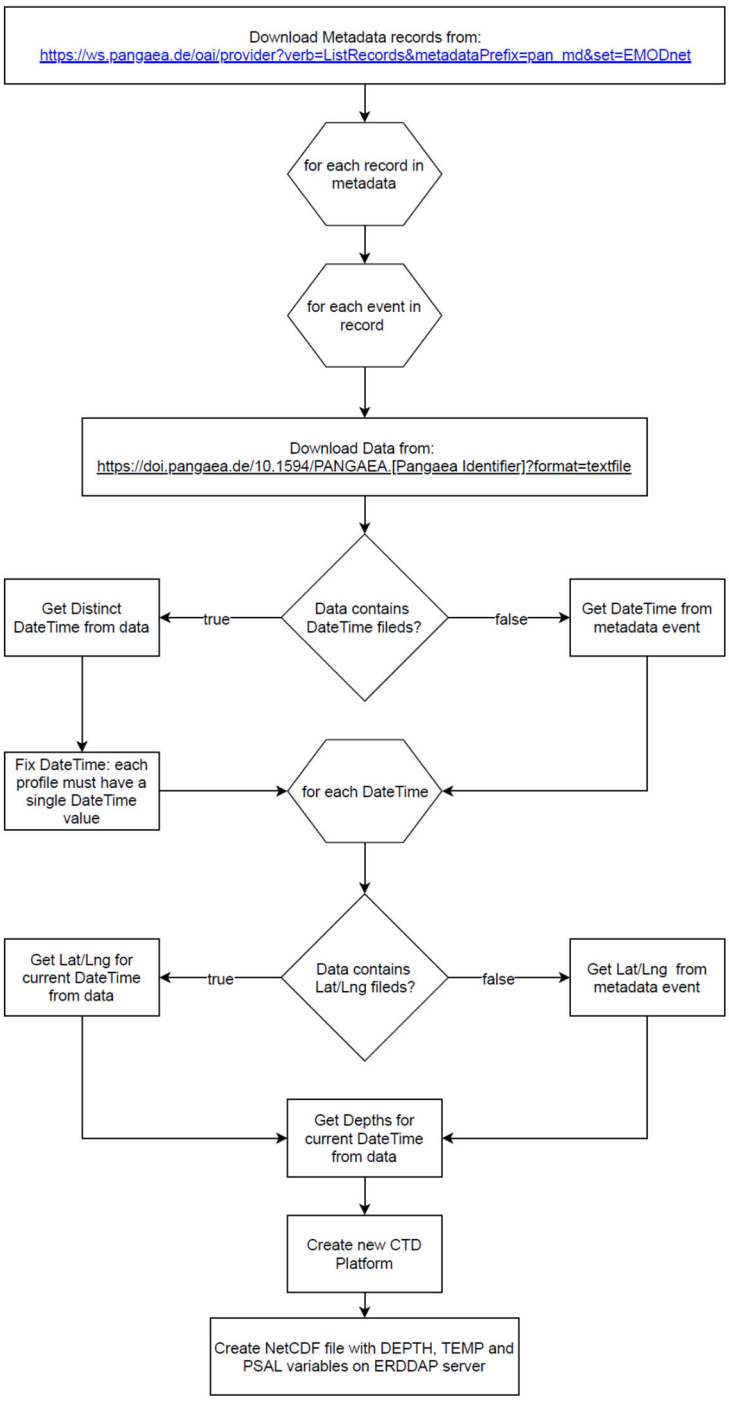

Figure 9: Semi-automated flow of CTD data from PANGAEA to EMODnet Physics

A new EMODnet Physics CTD platform is created and the DateTime, Location and Depth, together with the temperature and practical salinity (PSAL) information are then saved within the SeaDataNet

netCDF format[13]. The netCDF files can then be discovered and accessed through the EMODnet Physics ERDDAP server[14] in a variety of common file formats. Important to note is that the original datasets remain archived in PANGAEA but the processing flow is able to create a regularly updated cache of the data that is available through EMODnet Physics in a format that follows the standards set out by SeaDataNet[15]. This has the advantage that, in case updates are done to any of the datasets, the same version is offered through both PANGAEA and EMODnet.

While the periodic caching of datasets described above is an automated process that is currently ran annually, the addition of new PANGAEA datasets to the set that can be harvested by EMODnet requires an often manual remapping of the variety of different representation for physical parameters (e.g. temperature, salinity) in PANGAEA[16] to the common vocabularies used by SeaDataNet and EMODnet Physics (commonly known as BODC vocabularies)[17]. As such, adoption of common standards or robust translation tables at the data level could greatly benefit the data flow between the two open data publishers and result in more frequent periodic caching of updated and new datasets. Nevertheless, the established dataflow for physical oceanography data can serve as an example which can be extended to implement data flows from PANGAEA to EMODnet for other environmental data types (e.g. chemical oceanography data, geological data, etc.).

4.2 Biological occurrence data

EMODnet Biology and PANGAEA have been collaborating on data flow for more than a decade. From 2006 until 2011, a successful data flow of marine species occurrence data archived in PANGAEA to EurOBIS[18], the database infrastructure underpinning EMODnet Biology, was established. This accomplished the ingestion of data archived under approximately 7088 DOI's in PANGAEA resulting in approximately 1.4 million occurrence records in EurOBIS[19] from which approximately 1.1 million records passed the EurOBIS quality control procedures[20]. Since then, this data flow has not operated at the same level, with one key reason being that PANGAEA is not a partner in the current EMODnet Biology partnership. However, some data flow has continued, with a bulk ingestion of 70 thousand occurrence records (of which 69 thousand passed the quality control checks) in 2016[19]. While efforts

[13] https://www.seadatanet.org/Standards/Data-Transport-Formats
[14] https://erddap.emodnet-physics.eu/
[15] https://www.seadatanet.org/Standards
[16] https://wiki.pangaea.de/wiki/Parameter
[17] https://www.seadatanet.org/Standards/Common-Vocabularies
[18] http://www.eurobis.org/
[19] Datasets tagged "PANGAEA" in http://www.eurobis.org/dataset_list
[20] http://www.eurobis.org/#qc-nfo

to re-establish this data flow were undertaken in 2014, the full potential of this has not yet been realised. This leaves a potentially large number of archived occurrence records in PANGAEA which are not yet available through EurOBIS and EMODnet and could be ingested into these long-term data services. In this section, we first describe the original process of ingestion of PANGAEA occurrence records into EurOBIS and outline the main challenges that were encountered. Secondly, opportunities for a renewed collaboration and data flow are suggested.

The ingestion of PANGAEA data into EurOBIS in 2006-2011 was a manual to semi-automated process. PANGAEA datasets were first filtered for the parameter group "Taxa" with abundance units to retrieve several thousands of PANGAEA datasets with occurrence records. A challenge for publishing these PANGAEA datasets in EurOBIS was the discrepancy in dataset granularity between the two initiatives. The EurOBIS definition of a dataset is a relatively large collection of occurrence records, often aggregated by the project which collects them[19]. In contrast, PANGAEA has a different definition of a dataset. For instance, a dataset in PANGAEA often contains a smaller number of occurrence records collected by a single cruise or published in a single publication or publication table. In order for the data to be published in EurOBIS, the PANGAEA datasets were first grouped by research project and/or institution into 38 aggregated datasets[19]. As a result of this, dataset metadata records could not be extracted directly from PANGAEA but aggregated metadata had to be manually created. Links to the original metadata records were maintained by including the PANGAEA DOI in the citation at the individual occurrence record level. The taxa in the datasets were semi-automatically mapped to the taxonomy used by EurOBIS, based on the World Register of Marine Species (WoRMS)[21], and checked with the LifeWatch Taxon match services[22]. The datasets were then published in Darwin Core Standard (DwC) format[23] in a Distributed Generic Information Retrieval (DiGIR) web service[24], used as the database integration system by EurOBIS at the time. As EurOBIS has since moved to the open-source GBIF Integrated Publishing Tookit (IPT)[25] web service, the datasets were later republished in an IPT server [26].

As both EMODnet Biology/EurOBIS and PANGAEA have continuously improved the technology and quality of their services, new opportunities exist to re-establish a successful dataflow between them.

[21] https://www.marinespecies.org/
[22] http://www.lifewatch.be/data-services/
[23] https://www.gbif.org/darwin-core
[24] http://digir.sourceforge.net
[25] https://www.gbif.org/ipt
[26] http://ipt.vliz.be/eurobis/

First, the WoRMS database now has support for fossil species which make up a significant portion of occurrence record data submitted to PANGAEA. Second, the web services now provided by WoRMS[27] may facilitate the taxon matching with the taxonomy used by PANGAEA. Moreover, PANGAEA has also expanded their web services[28] which may further facilitate the data flow. Potential improvements could also be made in the way PANGAEA datasets can be aggregated in EurOBIS. For example, the taxon match could be used to group the PANGAEA datasets into 9 dataset collections by functional domain (i.e. benthos, zooplankton, phytoplankton, fish, birds, mammals, reptiles, algae, angiosperms and paleo/fossil) which can be handled by EurOBIS ingestion. Given these opportunities, the re-establishment of a collaboration between PANGAEA and EMODnet Biology has much potential to further improve the open access to biological data.

5 Geospatial data "hub": The ATLAS GeoNode

5.1 The ATLAS GeoNode

The ATLAS GeoNode[29] was set up as a novel approach to add value to the wealth of geospatial data produced by the project. This geospatial data hub was built using the state-of-the-art, open source, geospatial content management system - GeoNode[30]. The ATLAS GeoNode was developed by partner SBE, in collaboration with partner BGS, to raise awareness among the marine and maritime community of the wealth of data and products produced by ATLAS and to provide a community tool for project partners to visualise geospatial data required for research and marine management e.g. Marine Spatial Planning purposes. The ATLAS GeoNode user interface was customised to reflect the style of ATLAS and further customisations (e.g. menu structure, metadata fields, help pages) were adapted to fit the needs of the ATLAS consortium.

The ATLAS GeoNode provides a platform with an intuitive user interface where both the ATLAS consortium and stakeholders can search for, visualise, download and share ATLAS geospatial data, metadata and map products (Figure 10). Furthermore, it provides functionality to visualise ATLAS data together with data from external open data providers in interactive maps (e.g. EMODnet, Marine Copernicus, Marine Scotland, National Oceanic and Atmospheric Administration (NOAA) and many others; Figure 10). Guidelines on how users can interact with the ATLAS GeoNode User Interface were

[27] http://www.marinespecies.org/aphia.php?p=webservice
[28] https://www.pangaea.de/about/services.php
[29] http://www.atlas-horizon2020.eu/
[30] http://geonode.org/

documented in a help page[31] that covers actions like making an account, creating and publishing interactive maps and accessing static maps. In addition to a user interface, the ATLAS GeoNode provides access to ATLAS geospatial data and metadata through Open Geospatial Consortium (OGC)[32] and INSPIRE[33] compliant web services. These enable ATLAS consortium members and stakeholders to access ATLAS data and metadata directly within widely used GIS applications (QGIS, ArcGIS) and allow web applications to be built on top of the ATLAS GeoNode infrastructure. Examples of such applications are the European Atlas of the Seas (see section **Error! Reference source not found.**) and the Marine Spatial Planning application developed by University of Galway, as part of ATLAS WP 6. Finally, gathering of ATLAS geospatial data and metadata in the ATLAS GeoNode has the advantage of facilitating the flow to EMODnet portals, which make use of similar spatial data infrastructures by collecting rich metadata and troubleshooting potential problems with the provided data formats.

5.2 Designing the data and metadata flow to the ATLAS GeoNode

The main challenge in making the ATLAS GeoNode reach its full potential was designing an efficient flow for the ingestion of geospatial data and metadata. In order to accomplish this, a GeoNode Help team, administered by SBE and BGS, was established. The flow of data from external data providers that support industry standard geospatial web services (WMS[34], ArcGIS REST MapServer[35]) was straightforward as it made use of the GeoNode's functionality to connect with these remote services and add the relevant data layers with a minimum of metadata to the GeoNode catalogue (Figure 10). If the provider updates the data, these changes are automatically reflected in the ATLAS GeoNode. Upon request of users, remote services were added by the help team and from these services, users can then select the data layers of interest as documented in the ATLAS GeoNode guidelines[31]. For the data flow from providers within the ATLAS consortium and relevant external providers without web services, two different procedures were adopted (Figure 10).

As the majority of ATLAS data is archived in PANGAEA (Figure 8), a semi-automated data flow was established (Figure 10). This includes a python script that automatically fetches the metadata from the PANGAEA OAI-PMH service[36], and ensures it can be read by the ATLAS GeoNode metadata ingestion. The standard metadata set of GeoNode and its ingestion service was also expanded to

[31] http://www.atlas-horizon2020.eu/help/
[32] https://www.ogc.org/
[33] https://inspire.ec.europa.eu/
[34] https://www.ogc.org/standards/wms
[35] https://developers.arcgis.com/rest/services-reference/hosted-map-service.htm
[36] https://wiki.pangaea.de/wiki/OAI-PMH

include additional fields (DOI and data authors) which were collected by PANGAEA. Furthermore, several python scripts were developed to automatically download raster data, transform it to a standard coordinate reference system (registered within the EPSG registry[37] or spatialreference.org) and format (Geotiff) required by GeoNode and generate a file describing how the layer should be styled (according to the Style Layer Descriptor standard[38]). This workflow has the advantage of making use of the metadata curation performed by PANGAEA and transferring multi layered raster datasets produced by ATLAS quickly and efficiently. However, several challenges were also encountered and may be improved upon. First, there was no protocol for informing the GeoNode help team when new ATLAS datasets were made available in PANGAEA. For this reason, it was not possible to consistently detect and upload new datasets the moment they were made available. This could be optimised with consistent and comprehensive labelling of data. Second, by uploading the layer directly from PANGAEA into the ATLAS GeoNode, the help team is not directly informed of how the data should be displayed (i.e. colour scale, classes thresholds). In some cases, that information can be inferred from figures presented in the paper associated with the dataset but that isn't always available. Therefore, datasets should first be uploaded to the GeoNode with restricted viewing permissions and then both display and metadata should be reviewed by the authors before making the layer publicly available. A third challenge was encountered: when datasets in PANGAEA are under embargo, only the metadata is available to be transferred to the ATLAS GeoNode. The authors have to additionally provide the datasets to the GeoNode help team that will upload it (with restricted permission) and combine it to the metadata extracted from PANGAEA. That way, while the dataset is still under embargo, it can become available to (all or just some) project partners within ATLAS GeoNode. Once the embargo period is finished, permissions will have to be manually changed, by the help team or by the authors. As this may only occur after the end of the project, this can raise several issues. A fourth challenge was related to the projected coordinate reference system (Albers, an equal-area conical projection centred in the middle of the study area, used by ATLAS basin-scale outputs) used in the production of ATLAS raster data output. While this projection is optimal to minimise distortion within the study area, it is not a standard projection defined within the EPSG registry or spatialreference.org, traditionally more focussed on land area. As a standard projection is required for transmission of geospatial data using OGC protocols, the data had to be transformed ahead of ingestion in the ATLAS GeoNode, which results in a minor loss of fidelity in case of raster data. An opportunity to resolve this problem is to register the projection used by ATLAS in a community-driven registry such as spatialreference.org.

[37] https://epsg.io/
[38] https://www.ogc.org/standards/sld

Not all data relevant for the ATLAS GeoNode were archived in PANGAEA, with some datasets being archived in other open data repositories (see Figure 8) and some not required to be archived (e.g. working files only relevant within the context of the ATLAS project). Therefore, a separate manual data flow to the ATLAS GeoNode was established and managed by BGS, with support of SBE (Figure 10 and Figure 11). The data and metadata were manually collected and uploaded into the ATLAS GeoNode through the user-friendly data upload and metadata wizard services. The metadata wizard service has the advantage that it informs the metadata provider on the minimal metadata fields required by INSPIRE. The majority of datasets which followed this flow were provided by the 12 ATLAS case study (CS) leads via the BGS Sharefile folder, which was created at an early stage of the project when the CS leads were requested to send all the geospatial data used in delivery 6.1 to the BGS. However, as the creation of an ATLAS GeoNode was not part of the project at this stage, the documentation provided with the geospatial data was inconsistent and mostly scarce or non-existent. As such, the main challenges in this workflow were the collection of metadata and identifying the source of the datasets (many of which were pre-existing datasets describing each case study area). While trying to harvest missing information, it became evident that many of the datasets uploaded to the BGS ShareFile folders had been obtained from several portals that also allow access to the data via remote services. In these cases, datasets were linked in the ATLAS GeoNode via the remote services instead of using the layers archived in the BGS Sharefile folder. These issues can be avoided by always providing metadata together with the datasets, to allow the original data source to be determined.

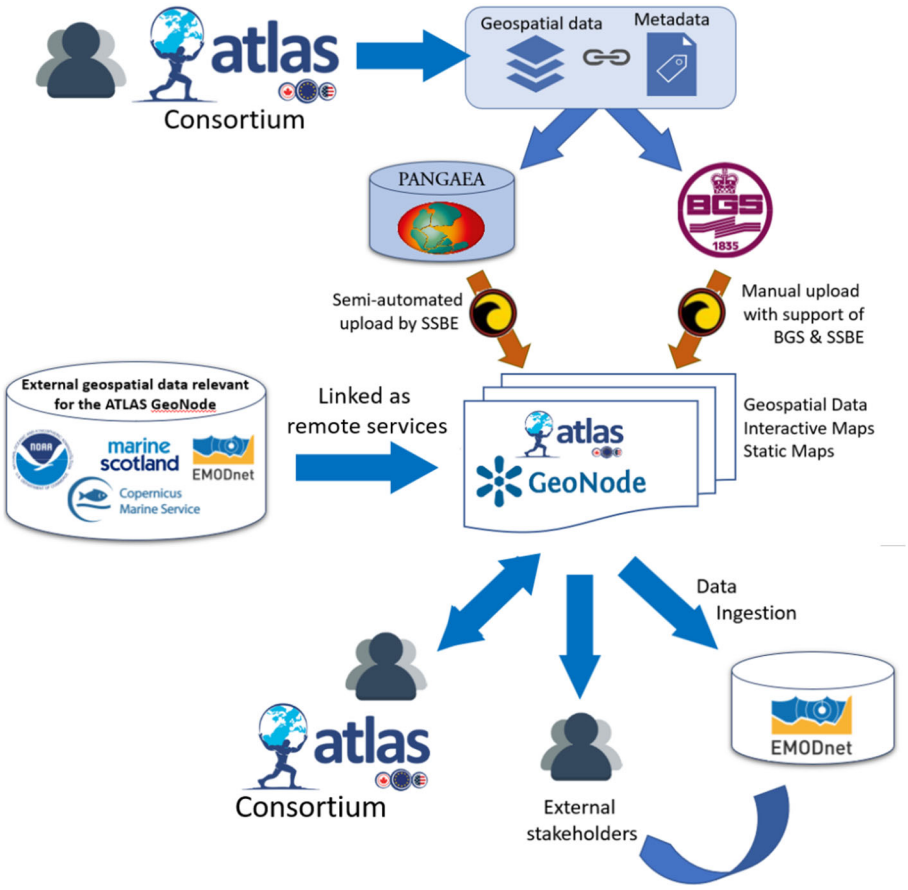

Figure 10: Flow of geospatial data and metadata to and from the ATLAS GeoNode

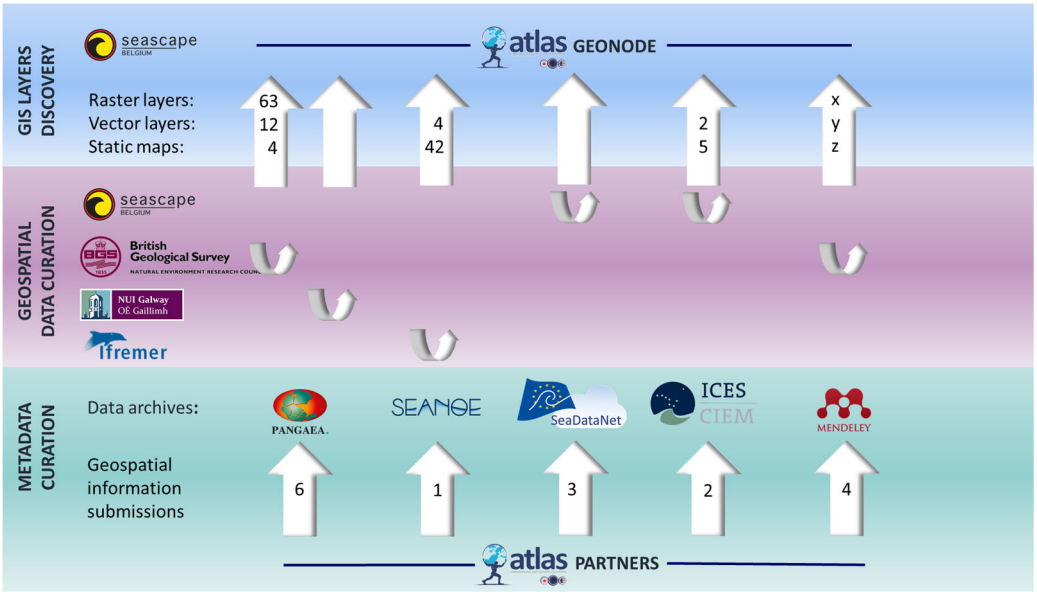

Figure 11: Flow of data from curation, including metadata curation, to the upload of data onto the GeoNode (geospatial data curation) and access on the GeoNode (GIS Layers discovery). Note that one dataset in an archive may contain multiple GIS Layers in the GeoNode.

5.3 ATLAS data in the ATLAS GeoNode

SBE and BGS have worked closely together to operationalise the data flow and pipeline for geospatial data from the data curation e.g. from PANGAEA and other repositories to the ATLAS GeoNode. Annex

I lists all the datasets produced in the project, with an indication of the data layers that are relevant, already uploaded (by earlyMay 2020) and soon to be uploaded to the ATLAS GeoNode as well as the ones identified for longer-term ingestion to EMODnet (see section 6).

In addition, there are a number of further layers expected to be made available and uploaded to the GeoNode – once the correct metadata have been supplied. These includekey outputs from ATLAS Work Package 6, in particular D6.2 "Ecosystem goods and services valuation and environmental risk assessment" led by National University of Ireland Galway (NUI Galway). This has generated several datasets and numerous static maps (produced by BGS in collaboration with WP6). From these, the first 2 maps of the EUNIS Habitat classification system for Europe (http://www.atlas-horizon2020.eu/documents/499 and http://www.atlas-horizon2020.eu/documents/502) have been uploaded to the ATLAS GeoNode. Further static maps, including for case study specific areas, will be uploaded in the coming weeks before the end of the project (including the no-cost extension to 31 July 2020). The list of additional ATLAS layers from D6.2 that can be uploaded to the GeoNode in the coming weeks includes:

- 12 layers that score each cell as High/Low/Negligible according to their ***Ecosystem Service Scores*** for Air Quality, Biodiversity, Cognitive, Disturbance, Feelgood, Food, Leisure, Nutrient, Photosynthesis, Raw material, Reproduction and Waste, respectively.
- 12 model outputs, one for each coral/deep-sea fish habitat using predicted presence/absence, for each case study.
- 3 layers with composite risk analyses were produced taking account of (1) **all predicted VME**[39] (based on predicted presence of all 6 VME indicator taxa whose distributions were modelled), (2) **all predicted deep-sea fish habitat** (based on predicted presence of 6 deep-sea fish species whose distributions were modelled), and (3) **the risk to all of these VME and deep-sea fish habitats considered collectively**.

Other layers from existing data services were also added to complement the new ATLAS data, including 1) EUNIS classification 2) the fisheries landings data (tonnes) and 3) landing value per ICES rectangle.

These layers require further metadata description by the data providers then, once on the ATLAS GeoNode, will further complete the ATLAS GeoNode as a tool for transforming new data, tools and

[39] VME = Vulnerable Marine Ecosystem

understanding into robust ocean governance in line with an adaptive ecosystem-based maritime spatial planning (MSP) approach to achieve ecosystem preservation, sustainable exploitation, and Blue Growth.

5.4 User feedback from science, policy, marine spatial planners and industry

During the course of the project the WP8 developments on open source data and resources were presented at a number of events. The ATLAS GeoNode was initially presented at the General Assembly 2019 in Mallorca, followed by the 3rd ATLAS science-policy panel meeting in Brussels in May 2019. This included discussions with policy makers and industry representatives about the optimum ways to enhance policy and industry relevance. Participants welcomed the ATLAS GeoNode as a potential way to improve the sharing of data and scientific knowledge in support of policy developments and marine management needs in a timely and relevant way.

The industry relevance of this work was further explored on 11 December 2019 when the ATLAS GeoNode was also presented at a workshop in Dublin that addressed aspects of maritime spatial planning that could benefit industry and support Blue Growth. This was co-organised by the ATLAS project, the Marine and Renewable Energy Research, Development and Innovation Centre (MaREI)/RPS Ireland and the Irish Offshore Operators Association (IOOA). The workshop was attended by representatives of the oil and gas industry from Ireland and Norway, engineering and environmental consultancies, Irish Government officials, and academics, Marine Scotland Science and ATLAS project partners. Discussions included demonstrating enhanced planning decision support tools that better represent the complexity of the marine environment, enabling improved access to data for more cost-effective environmental impact assessments, and the options available to industry to mitigate impacts. ATLAS-supported open-source resources and services were promoted as a means by which industry could engage with, and benefit from, data sharing. The ATLAS GeoNode was presented as a pragmatic clearing house solution to enable sharing of geospatial data, with minimum data and metadata quality checks preparing the data for ingestion and publication through EMODnet.

At the meeting, industry representatives expressed a desire to see greater connectivity and interoperability between marine data to increase their impact and use and to streamline the process of marine data discovery, uptake and exploitation. While some financial, work culture and data ownership barriers still exist, there is a new recognition in industry of the benefits of sharing (non-

commercial) data in open-access portals such as EMODnet (Murray *et al.,* 2018[40])**.** Read the full news article in the 7th edition of the ATLAS Newsletter, published in February 2020[41].

Further surveys of industry stakeholders were conducted by ATLAS WP7 and WP6 with interesting outcomes. Of particular relevance here, ATLAS Deliverable D6.4[42] produced six key recommendations on data-sharing opportunities and barriers to unlock Blue Growth potential. Whilst noting that EMODnet as a key data-sharing mechanism, D6.4 highlighted the need raise more awareness around EMODnet and its potential to greatly assist the wider offshore community in best practices and reducing costs of EIAs, e.g., as in the German windfarm use case study identified during the very first Blue Growth Data workshop in Edinburgh in 2017 (Murray et al., 2018). It was also recommended that EMODnet continue and strengthen proactive engagement with deep and open ocean industries to ensure awareness is raised. This is continuing in a dialogue with the International Oil and Gas Producers, including Beyond Petroleum (BP) and Equinor, as a result of ATLAS project developments.

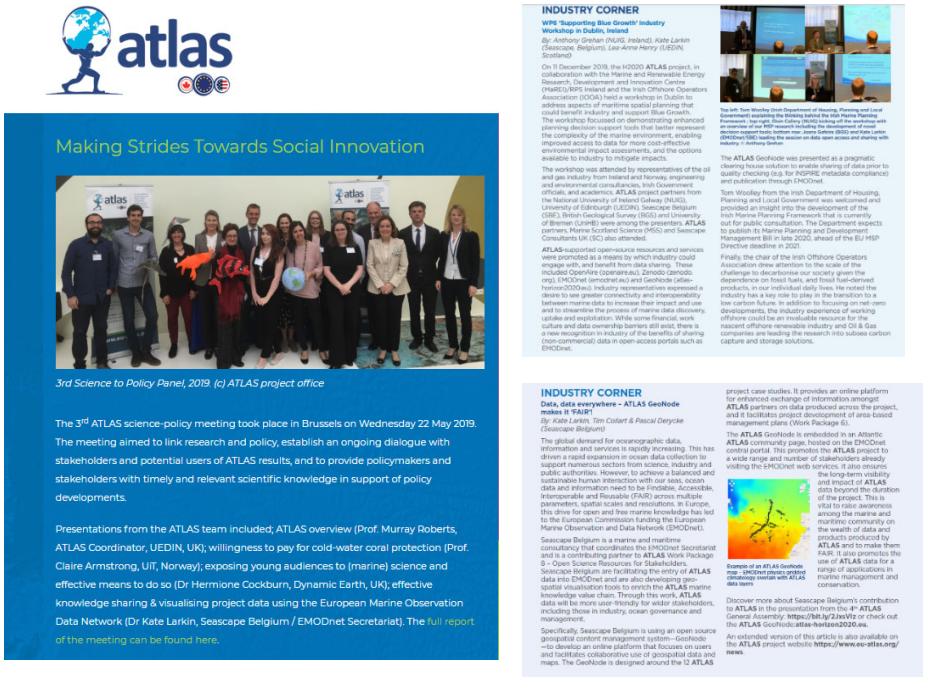

Figure 12: Examples of SBE presentations on the ATLAS GeoNode and valorization of project outputs to the 3rd ATLAS science-policy panel (May 2019, Brussels), a "Supporting Blue Growth" industry workshop, December 2019, Dublin, Ireland (7th Newsletter in January 2020) and an article on the GeoNode written for the 6th Newsletter in August 2019

[40] Murray F, Needham K, Gormley K, Rouse S, Coolen JWP, Billet D, Dannheim D, Birchenough SNR, Hyder K, Heard R, Ferris JS, Holstein JM, Henry L-A, McMeel O, Calewaert J-B, Roberts JM (2018) Data challenges and opportunities for environmental management of North Sea oil and gas decommissioning in an era of blue growth. Marine Policy 97: 130-138

[41] https://www.eu-atlas.org/news/browse-previous-newsletters/498-atlas-newsletter-7-february-2020/file

[42] ATLAS D6.4: Improving business practice and costs through data-sharing and the identification of potential mitigation measures for adaptive marine spatial planning (Henry et al., 2020)

SBE also wrote an article dedicated to the ATLAS GeoNode for the ATLAS Newsletter, 6th edition, published August 2019 [43] (see Figure 12 for examples).

The user uptake and visits to the ATLAS GeoNode are also an important indicator of the user requirements, interest and use. In August 2019, a Matomo instance was installed to track web traffic to the ATLAS GeoNode (Figure 13). This showed an upwards trend in the visitors with spikes of interest and downloads seen during ATLAS-related meetings (e.g. General Assembly) and rising towards the end of the project. A final analysis of this will be included in the final reporting of ATLAS, to take place in July 2020 due to the no-cost extension.

In related activities, EMODnet also promoted the involvement of ATLAS researchers in the second OpenSeaLab Hackathon which took place in Ghent, Belgium, from the 4th to the 6th of September 2019, offering marine data enthusiasts the opportunity to work in groups to develop new tools and products related to three marine themes: sustainable blue economy, blue society and ocean literacy and marine environment protection and management. Such events are an important opportunity to increase the visibility and value of marine data. Johanne Vad, a postdoctoral researcher based in the Changing Oceans Group at the University of Edinburgh working within the EU-ATLAS project, was amongst the participants (see her article on the ATLAS website[44]). As part of the team Overlap (with team members from Germany, Norway, Belgium and UK) a machine learning based tool was developed to predict the impact of new human activities on benthic community composition. To achieve this, the team used data from multiple EMODnet portals, the ICES database as well as the Copernicus repository. This was an important way for ATLAS early career researchers to be engaged in a training, networking and innovation event, inspiring a next generation of marine data scientists to tackle societal challenges and develop applications. It has also sparked new collaborations between the marine research community and marine data initiatives that will continue with future hackathons, not only organised by EMODnet but also the upcoming Hack4Oceans (rescheduled to October 2020) and the EUDatathon (October 2020) to name a few.

[43] https://www.eu-atlas.org/news/browse-previous-newsletters

[44] https://www.eu-atlas.org/news/project-news/having-fun-with-marine-data-opensealab-2

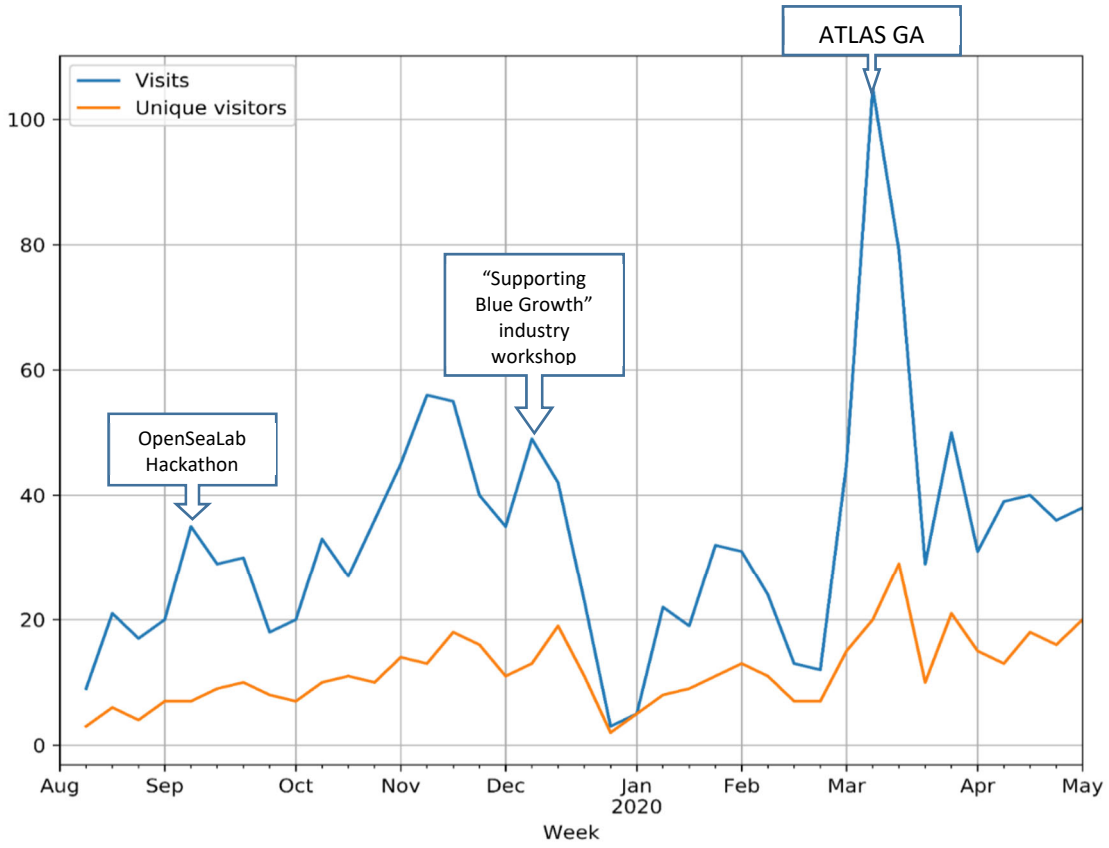

Figure 14: Evolution of the number of weekly visits and unique visitors (unique IP addresses) to the ATLAS GeoNode with indication of relevant outreach events.

5.5 ATLAS GeoNode legacy

The ATLAS GeoNode has been shown to be a useful resource and the decision was made by SBE, in dialogue with ATLAS Coordinators University of Edinburgh, to fund a 5-year legacy for the ATLAS GeoNode. This means that after the project end the ATLAS GeoNode will remain operational and open for data upload, viewing and map creation/download. This section explains a little more about the ATLAS GeoNode legacy requirements.

The ATLAS GeoNode uses a Virtual Private Server (VPS) service, offered by OVH[45] which has a monthly cost to maintain the processor, memory, storage (of data that needs to be hosted on the ATLAS GeoNode) and an open source operating system (Linux OS family) and backup. As more data (including modelling outputs) are being uploaded to the ATLAS GeoNode the storage will be upgraded to 150GB. Since any storage of data on the ATLAS GeoNode local spatial database has a cost implication, where possible, data are not hosted directly on the ATLAS GeoNode but linked through remote services. For this to be possible, there needs to be availability of web service access to these external data e.g. as

[45] https://www.ovh.com/

provided by EMODnet. Where availability of web service access to the data is not possible the data will need to be hosted on the ATLAS GeoNode local spatial database and later ingested into a longer-term open access data brokerage service. In both cases (whether hosted on the ATLAS GeoNode or connected through remote service), all data entries/maps will link to the original data archive using its digital object identifier (DOI). In addition, the standardised metadata will ensure that both the dataset and the data provider are fully acknowledged for any future use of the data. In addition to storage costs, a minor fee is required to maintain the atlas-horizon2020.eu domain and associated email address, registered at Gandi.net. For further detail, including costs, please refer to the "ATLAS GeoNode Legacy document" (internal document produced by SBE for the ATLAS Steering Committee in March 2020).

Many of the ATLAS datasets are now available in the ATLAS GeoNode. However, some datasets are still being finalised by researchers across other Work Packages and SBE will work with WP8 partners to maintain the dialogue and ensure as many datasets as possible are curated in PANGAEA or other relevant data repositories and data publishers, so these can be uploaded to the ATLAS GeoNode. This will continue into the no-cost extension of ATLAS, now agreed until 31 July 2020.

During the final months of the project, SBE has also increased dialogue with other related projects to see the possibilities of uploading (or connecting via remote services) to other relevant data e.g. from the SPONGES project, which has produced mainly biological data and has started dialogue with EMODnet Biology.

Dialogue has also started, and will continue through SBE (that administers the EMODnet Secretariat), with other Atlantic projects e.g. iAtlantic, TRI-ATLAS, Mission Atlantic, AtlantECO, related projects e.g. SPONGES, and cooperation actions including AAnCHOR. By the end of the 5-year legacy of ATLAS, all ATLAS data will either be ingested into EMODnet and/or ATLAS data layers can be migrated into other related GeoNodes e.g. iAtlantic-GIS. Since data layers are tagged per project, ATLAS data layers would still be discoverable and the provenance/data provider/research funding recognised. It would also encourage the use of ATLAS data in the iAtlantic project, to complement and integrate with H2020 iAtlantic data.

6 Long-term data discovery in EMODnet

6.1 EMODnet Data Ingestion overview

EMODnet offers a dedicated service called the EMODnet Data Ingestion, which exists specifically to guide and assist with data ingestion. The EMODnet Data Ingestion team, in collaboration with the EMODnet Secretariat, facilitate the ingestion process of relevant geospatial observational data. In

ATLAS this focuses on new datasets that are produced by project partners during the project, and includes any additional datasets that become available for open access distribution, even beyond the project lifetime.

Data providers from across Europe can submit their data through the central EMODnet Ingestion Portal. Each data submission is assigned to a data centre to complete the metadata. These completed submissions are then published openly with their data packages 'as is' in the View Submissions service of the portal, where users can search, browse and download the data packages. Since October 2019 the number of published data submissions has increased from by 50 to 600, with almost 250 datasets being ingested at data centres and made available to national, European and EMODnet thematic portals (see Figure 15 for the flow of data to EMODnet Ingestion).

EMODnet data ingestion is now increasingly known in the marine research community (see recent article[46] in the Marine Environmental Data and Information Network (MEDIN), a partnership of UK organisations committed to improving access to marine data).

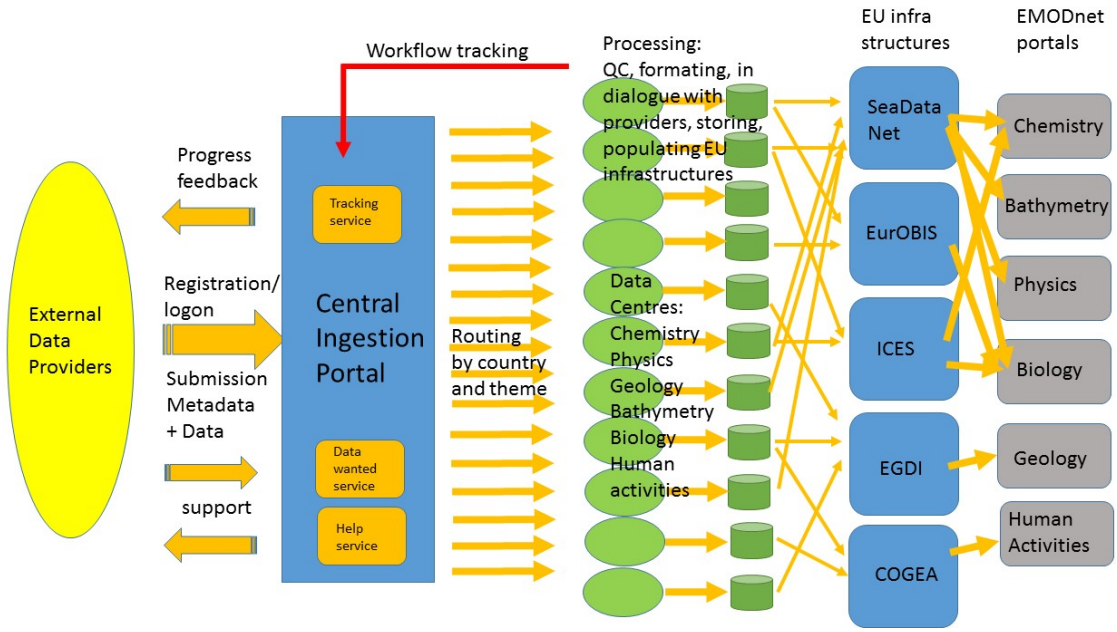

Figure 15: Flow of data from provider to National Databases and EMODnet portals in the EMODnet ingestion service.

EMODnet receives a growing number of data submissions from **private companies/industry.** Dialogue through EMODnet Associated Partners, Industry events and research projects like ATLAS are leveraging more data sharing by industry. Figure 16 shows the data ingestion by organization type from EMODnet Data Ingestion in the last Quarter of 2019.

[46] https://mailchi.mp/3194068c3b51/marine-data-news-april-2020

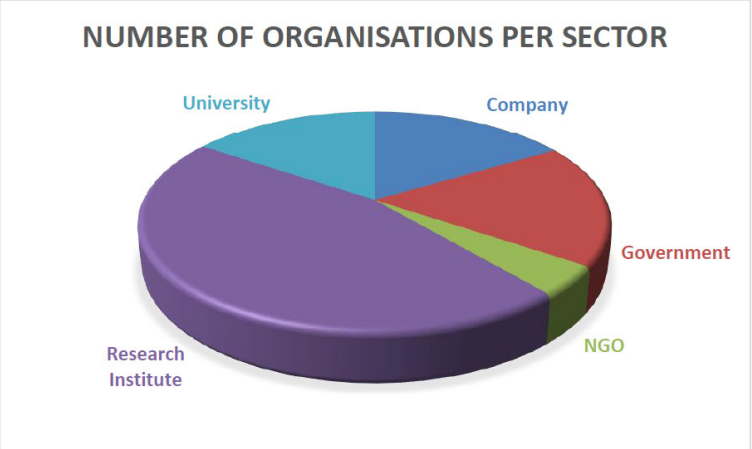

Figure 16: Data ingestion by organization type from EMODnet Data Ingestion in the last Quarter of 2019.

It is noted that, despite the private sector collecting huge quantities of marine environmental data, there remain a number of bottlenecks and barriers to sharing of these data. As outlined by ATLAS D6.4, more can be done to build effective and long-term partnerships with industry. EMODnet already has a growing Associated Partnership Scheme with 21 maritime industry members building win-win solutions to data sharing and data use and uptake. However, the five recommendations from ATLAS D6.4 clearly show that more can be done to communicate the value of open data sharing and the opportunities of visualising data through a common platform e.g. ATLAS GeoNode, which include the ability to visualise whilst setting restrictions on data access, to be in-line with any commercial restrictions that may be in place for datasets from the private sector. Going forward, it is therefore recommended that industry utilises ATLAS' various outputs at regional and basin scales, including maps of marine environmental parameters that are openly shared, to underpin their mitigation options and planning. This includes utilising ATLAS' outputs on the understanding of Vulnerable Marine Ecosystems (VMEs) and the coupled hydrodynamic-biogeochemical models developed by ATLAS. ATLAS partners should also continue, and strengthen, the dialogue with industry, including further communication on EMODnet and the added value, and availability, of open access data and data products that could inform industry operations and planning. Industry is also encouraged to more explicitly consider the information that ATLAS has collected, visit and use the ATLAS GeoNode

6.2 Status of ATLAS data ingestion to EMODnet

Ingestion of data into EMODnet requires a minimum set of metadata (see Annex II). This is why the GeoNode tool is so effective at streamlining the pipeline and ingestion to EMODnet as it also requires a minimum standard metadata very similar to the needs of EMODnet. For this reason, metadata

requirements were compared between PANGAEA, the GeoNode and EMODnet to make these as aligned as possible. Early in the project, EMODnet data ingestion registered ATLAS in the European Directory of Marine Environmental Research Projects (EDMERP; Figure 17).

Figure 17: ATLAS listed in the EDMERP database of SeaDataNet.

Over the course of the project the flow of ATLAS data to EMODnet has been started by making a selection of ATLAS datasets containing relevant data for the different thematic portals of EMODnet. This selection is being made by forwarding a list of ATLAS datasets (see Annex I) and their metadata to the relevant people within each thematic domain of EMODnet. In the table, a preliminary indication to which thematic domain within EMODnet the dataset has relevance has been provided. As EMODnet mainly collects observational data, rather than model outputs, efforts will initially be focused on the former.

It is noted that due to the range of repositories used by ATLAS partners to deposit their data – mainly because of the diversity of the data spanning marine environmental to marine genomic data, the flow of data to EMODnet portals was not generally automated and requires manual interventions. As discussed earlier, dialogue is underway to find ways to optimise data flow between PANGAEA and EMODnet. EMODnet is already collaborating well with SeaDataNet, in particular for some of the thematic areas e.g. Chemistry. For these reasons, it is suggested that data producers in European research projects could deposit data directly to the EMODnet or SeaDataNet ingestion systems, both of which provide automatic deposition in trusted archives and dissemination to EMODnet. In addition, SBE (administering the EMODnet Secretariat) facilitated a discussion with EMODnet portals to identify

which ATLAS datasets could be relevant for their portal so they are aware of the upcoming datasets. First, a comprehensive review of all datasets and their potential contribution to EMODnet thematic data domains was conducted by SBE. These preliminary assessments are indicated for each dataset listed in Annex I. A dialogue was then set up with each EMODnet portal.

For example, EMODnet Seabed Habitats[47] uses the EUSeaMap that extends partially into the Atlantic (see Figure 18) but has (yet) no plans to extend further. Predicted seabed-habitat types are mapped by combining a series of proxy measurements, such as water depth and light levels amongst others, using statistical analysis to identify relationships with biology and GIS modelling. The habitats are classified following the EUNIS hierarchical system (Manca and Vasquez, 2015[48]). EMODnet Seabed Habitats experts have indicated that key input layers to produce the Seabed Habitat maps are the EMODnet Bathymetry Digital Terrain Model (DTM) and EMODnet Geology seabed substrate layers. They also noted the OSPAR Threatened and/or Declining Habitats contains data within the OSPAR area (the north-east Atlantic Ocean) only, with no plans to extend at the current time. In its first phase (2009-2012) over two million km^2 of European seabed were mapped. However, at this stage, EMODnet Seabed Habitats remains focused on European regional seas and the Northeast Atlantic.

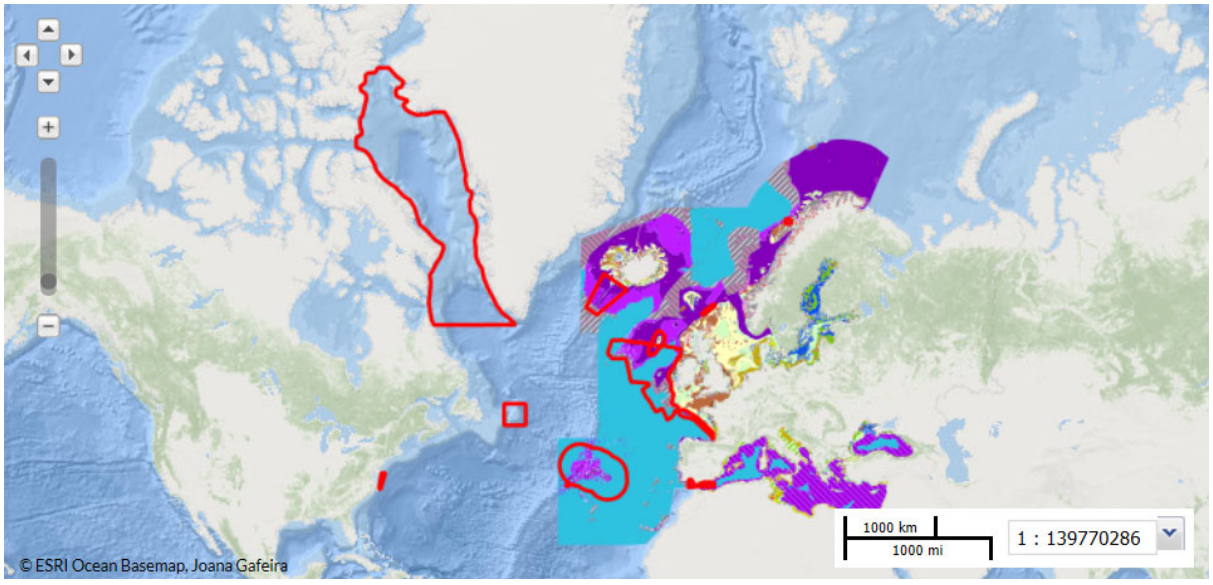

Figure 18: ATLAS Case Study areas & EMODnet seabed habitat map (EUSeaMap) on the ATLAS GeoNode.

On discussing upcoming ATLAS data outputs, EMODnet Seabed Habitats noted some priority datasets, including "Environmental variables that influence habitat type" produced by ATLAS, would be useful

[47] https://www.emodnet-seabedhabitats.eu
[48] https://archimer.ifremer.fr/doc/00318/42910/

to strengthen and extend the dataset available to EMODnet Seabed Habitats. The example shown below in Figure 19 is the set of terrain (static in time) and environmental (dynamic in time) variables used as candidate predictors of present-day (1951-2000) & future (2081-2100) suitable habitat of cold-water corals and deep-sea fishes in the North Atlantic, , recently published by Morato *et al.* (2020)[49].

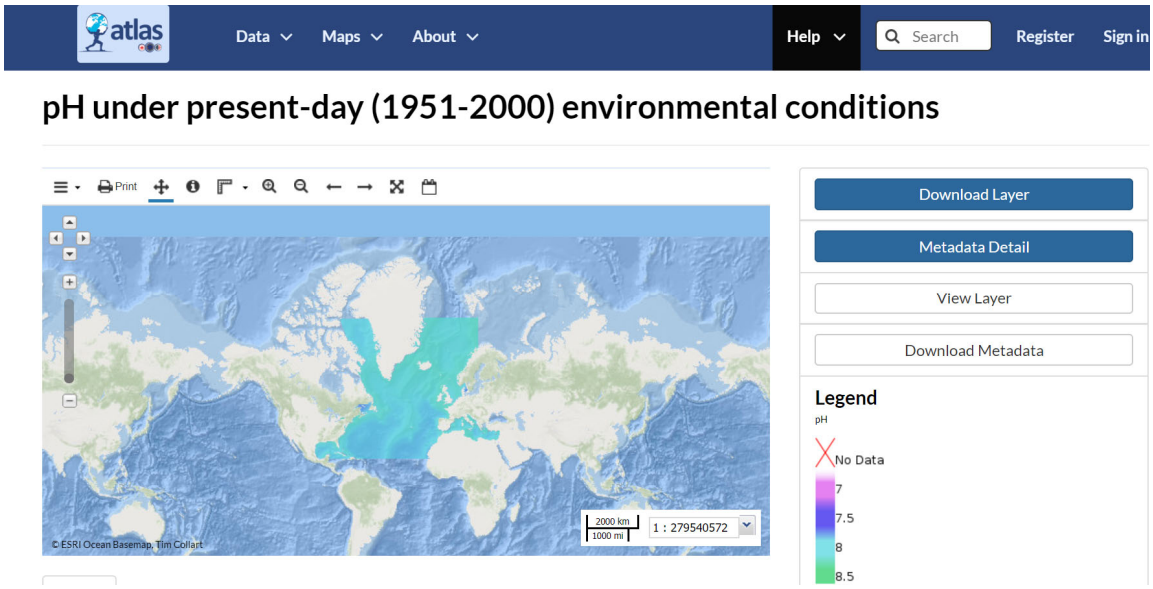

Figure 19: ATLAS data layer ingested into the ATLAS GeoNode: set of terrain (static in time) and environmental (dynamic in time) variables used as candidate predictors of present-day (1951-2000) & future (2081-2100) suitable habitat of cold-water corals and deep-sea fishes in the North Atlantic[50]

EMODnet focuses on ingestion of data collected by European scientists and data providers (both in the European domain and across the world). In terms of other relevant projects, EMODnet is also engaging in a dialogue with other relevant projects including H2020 SponGES. This project has built a SponGIS[51] data portal dedicated for deep-sea sponge data, following international standards of biodiversity data. A dialogue, initiated in Spring 2020, will be taken forward between SBE and the SponGES project to see opportunities for bringing SponGES geospatial data into the ATLAS GeoNode

[49] Morato T, José-Manuel González-Irusta JM, Carlos Dominguez-Carrió C, Chih-Lin W, Andrew Davies A, Sweetman AK, Taranto GH, Beazley L, García-Alegre A, Grehan A, Laffargue P, Murillo-Perez J, Sacau M, Vaz S, Kenchington E, Arnaud-Haond S, Callery O, Chimienti G, Cordes E, Freiwald A, Gasbarro R, Gilkinson K, Hayes VW, Hebbeln D, Hedges K, Henry L-A, Koen-Alonso M, Lirette C, Mastrototaro F, Menot L, Molodtsova T, Orejas C, Pennino MG, Puerta P, Ragnarsson SA, Ramiro-Sánchez B, Rengstorf A, Rice J, Rivera J, Roberts JM, Ross SW, Rueda JL, Siferd T, Stirling D, Treble M, Urra J, Vad J, Walkusz W, Wienberg C, Woillez M, Levin LA, Carreiro-Silva M (2020) Climate-induced changes in the suitable habitat of cold-water corals and commercially important deep-sea fishes in the North Atlantic. Global Change Biology 26: 2181-2202

[50] http://www.atlas-horizon2020.eu/layers/geonode:ph_FromKrige_3km_1951_2000

[51] https://spongis.org/

and to ingest SponGES data into EMODnet. Since most of the data produced by SponGES are biological occurrence data and the procedure would be to have it fed via EurOBIS and then being made available to EMODnet Biology.

There is also a dynamic exchange between the SeaDataNet SEANOE data citing service[52] and EMODnet Ingestion. This has recently been fully deployed (EMODnet Data Ingestion Q4 2019 report) which has resulted in many new submissions starting 2020. The SeaDataNet SEANOE service invites European scientists to publish their scientific papers and associated data collections in return for a DOI which promotes their wider citation. The exchange facilitates the harvesting of (selected) scientific submissions from SEANOE by EMODnet Ingestion for further metadata completion, publishing 'as-is', and elaboration of data sets for inclusion and publishing in national and European portals. SEANOE will undertake further activities for streamlining the flow by adopting controlled vocabularies.

It is noted that EMODnet is not currently harvested by OpenAIRE. This means that data and data products available on that platform are not monitored or accounted for as research outputs by the European Commission. There are a variety of historical and technical reasons for this, however efforts are being made for a dialogue with OpenAIRE to progress this further.

6.3 EMODnet for global users, extending the uptake and impact of ATLAS data in Europe, and beyond

With ever more global coverage of datasets and increasing collaborations worldwide, EMODnet is a trusted resource for marine data and data products in Europe and beyond. An example includes EMODnet Bathymetry, which has a direct cooperation with NOAA as part of Seabed 2030. Through this Memorandum of Understanding, any data submitted to EMODnet Bathymetry will both strengthen European data mapping efforts, and directly contribute to NOAA and the international Seabed 2030 initiative. A further example is EMODnet Biology which supplies more than 50% of the global biological datasets in OBIS, through EurOBIS.

EMODnet has an increasingly global outlook and a growing number of partnerships and collaborations to bring in global datasets from other regions beyond Europe. The ATLAS datasets will significantly enrich the EMODnet coverage in the Atlantic Ocean, offering relevant datasets and outputs for

[52] https://www.seadatanet.org/Software/SEANOE

EMODnet Seabed Habitats, EMODnet Physics, EMODnet Bathymetry, EMODnet Chemistry and EMODnet Geology, and potentially other thematic areas.

During the ATLAS project, EMODnet developed a series of community pages focused on Regional Sea and Basin-Scale level information. An Atlantic community page was set up to promote to users the datasets that EMODnet already makes available in the Atlantic Ocean e.g. through EMODnet Physics and the growing contributions and partnerships that EMODnet has e.g. through the ATLAS project (see Figure 20).

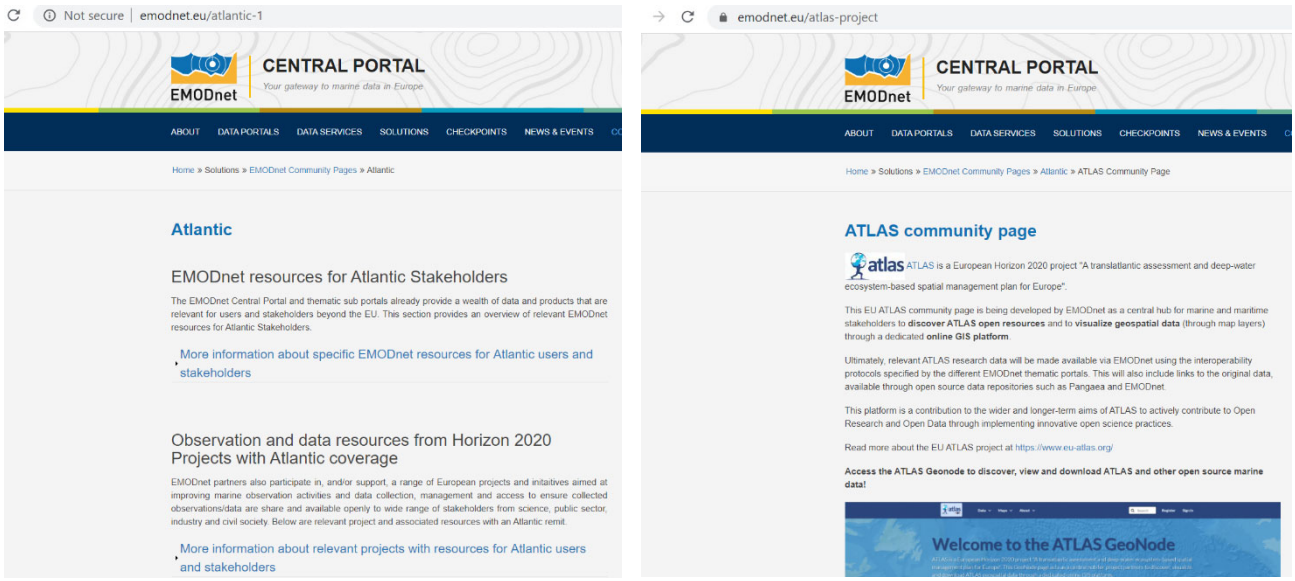

Figure 20: EMODnet Atlantic community page (https://www.emodnet.eu/atlantic-1) referencing EMODnet resources, wider projects, initiatives and partnerships and ATLAS community page (https://www.emodnet.eu/atlas-project).

7 Wider dissemination platforms: The European Atlas of the Seas

To achieve an ocean-literate society and connect people to our blue planet, it is crucial to communicate marine data and information in an attractive, easy to digest and interactive way. To further increase the accessibility and impact of ATLAS data to society, specific geospatial data and map layers from ATLAS may also be included in other open source web platforms designed for wider society, such as the European Atlas of the Seas (EAS)[53] This is a European Commission (DG MARE) web mapping application for the general public which is powered by open marine data e.g. from EMODnet, European Environment Agency, Copernicus Marine (CMEMS) and Eurostat, and provides an interactive online tool on European coasts and seas to promote ocean literacy and a blue society. The European Atlas of the Seas is an educational and interactive web mapping application for the general public, schools and non-expert professionals. It brings at-a-glance data in a comprehensive and fully visual way to a broad public, while at the same time serving as a support tool for marine policy and

the blue economy. The European Atlas of the Seas hosts a wealth of information about Europe's marine environment and related human activities, covering topics such as nature, tourism, security, energy, passenger transport, sea bottom, sea level rise, fish consumption, and much more.

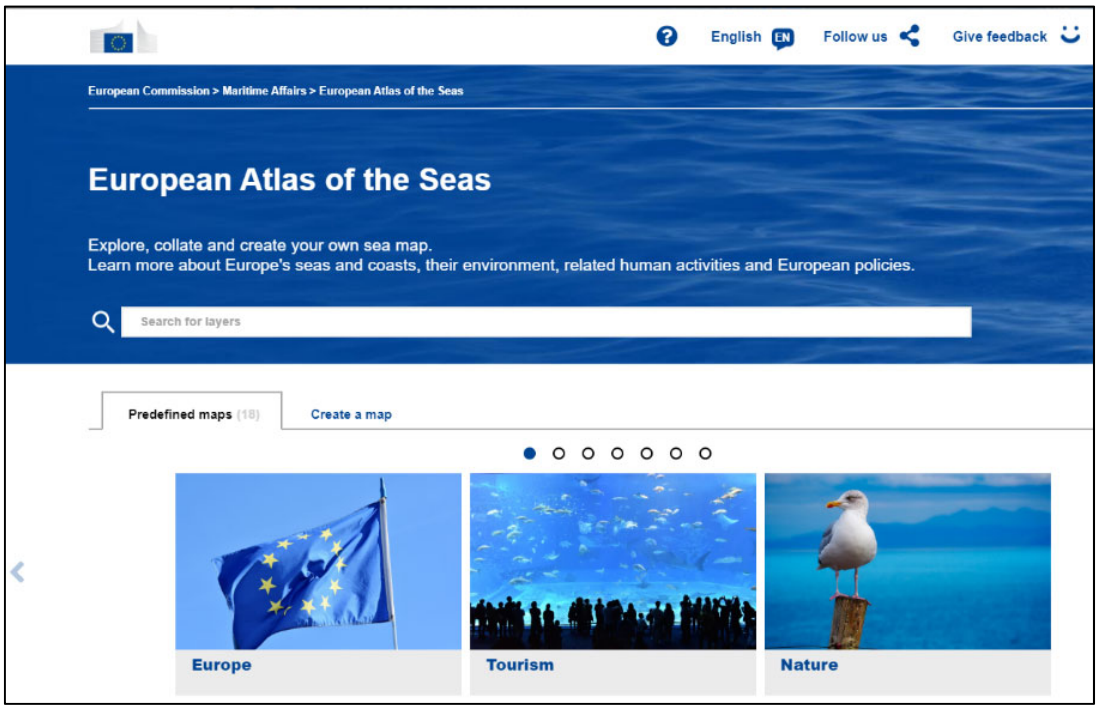

Figure 21: Front page of the European Atlas of the Seas

The mission of the European Atlas of the Seas is to raise awareness about Europe's seas and coasts in the context of the EU's integrated maritime policy, which aims to coordinate across marine policies. To this end, the European Atlas is a go-to tool and one-stop-shop, allowing users an easy interface to explore maps based on diverse source data and learn about Europe's marine environment and how we use it.

The European Atlas of the Seas was identified early on as a potential dissemination platform for ATLAS data, building on the availability of other project research data already ingested in the European Atlas e.g. for Integrated Coastal Zone Management. Since the European Atlas of the Seas is a European Commission tool, there is a request process for new layers to be ingested and this requires discussion with the European Atlas service providers and DG MARE of the European Commission.

In 2019, SBE proposed some potential data layers from ATLAS that could be particularly relevant for wider society and therefore candidates for the European Atlas of the Seas. The two sets of data layers are listed below (with links to the data layers on the ATLAS GeoNode).

- http://www.atlas-horizon2020.eu/layers/geonode:Lophelia_pertusa_GainLossRef_MSS
- http://www.atlas-horizon2020.eu/layers/geonode:all_management_SelFreq0

Both layers convey a clear message about marine ecosystem conservation and management. Following dialogue between SBE (EMODnet Secretariat) and the lead data providers (including Ifremer, France and IMAR, Azores) a formal application to ingest the layers was submitted to the European Commission. These are now currently in the process of being made available (see Figure 22 for an example of ATLAS data layer currently in the staging platform of the European Atlas of the Seas). In the coming months, further 'storytelling' for education and ocean literacy purposes is planned for the layers, together with publicising them through the weekly European Atlas Map of the Week which is publicised on the European Commission DG MARE Maritime Forum webpages, EMODnet Central portal webpages and on social media through @EuropeAtlasSeas, @EMODnet and @EU_MARE.

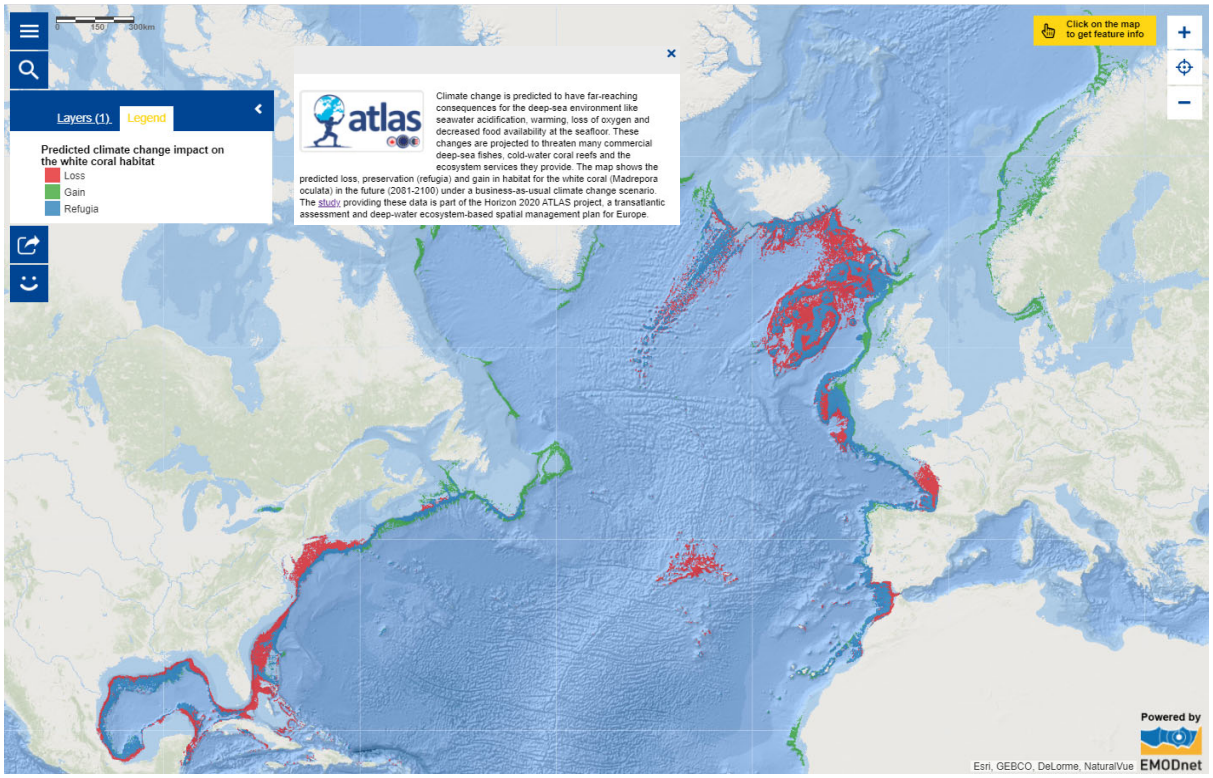

Figure 22: Example of ATLAS data layer available through the European Atlas of the Seas: Projected present-day suitable Lophelia pertusa *habitat loss, gain and acting as climate refugia areas under future scenarios* [53]

8 Conclusions

In the ATLAS project, WP8 has promoted multiple methods and tools for the production, collation and long-term discovery of research data and wider outputs. In addition, SBE has developed a bespoke online GIS-repository – the ATLAS GeoNode – that enables the sharing and visualization of geospatial data from the project, and wider open source data – and has shown good potential and user reviews

[53] http://european-atlas-of-the-seas.eu/

for its use by wider stakeholders including industry and those working at the science-policy interface. Furthermore the European Atlas of the Seas, as an EU online communication tool for society, is also ingesting ATLAS data layers which will further communicate ATLAS data and research outputs beyond the 'traditional' audience of marine and maritime stakeholders to wider professionals and the youth and education sector.

A challenge for the GeoNode was the diversity of the data being dispersed across multiple data repositories and not knowing the expected volume of data. This was a challenge due to the different parameters and domains from biological and genomics data to geological and other geospatial data. It is recommended that future projects not only list expected data but also their volumes, to give an indication of storage requirements and help plan visualisation e.g. GeoNode applications. In addition, identifying not just one focal point for data management in a project, but also a clear data flow and stewardship is not simple, but could be further optimised to add more value and streamline data valorisation in large integrated research projects.

In the final few months of the project (including to 31 July 2020) but also beyond, there will be a concerted effort to ingest all relevant data from ATLAS and, where possible related projects e.g. SponGES, into long-term data initiatives e.g. EMODnet. This will be done in collaboration and dialogue with the data providers (ATLAS partners) and EMODnet thematic Coordinators and Data Ingestion. The result will mean data will be findable and discoverable through EMODnet and through remote services. EMODnet Data Ingestion provides a service to all data providers wishing to submit data to EMODnet. During the ATLAS project, SBE (who administer the EMODnet Secretariat) are facilitating this process. However, ingestion of relevant ATLAS data into EMODnet will continue, where necessary, beyond the end of the project, to ensure data ingestion into EMODnet (working with marine data infrastructures e.g. national data centres and SeaDataNet and the European Directory of Marine Environmental Research Projects) for longer-term discovery and impact of ATLAS data.

Furthermore, the dialogue with PANGAEA and EMODnet has led to systemic updates to the operational data flows, including plans towards more automated harvesting of PANGAEA data and metadata by EMODnet into the future. The ATLAS project has also facilitated greater exchange and move towards interoperability and optimised data flows between PANGAEA and EMODnet. From assessment of the data flows in place it is also suggested that data producers in European research projects could deposit data directly to the EMODnet or SeaDataNet ingestion systems, both of which provide automatic deposition in trusted archives and dissemination to EMODnet. EMODnet will also explore options to link with OpenAIRE to make EMODnet data and data products even more discoverable through this open science resource.

In addition, the stakeholder contacts e.g. in industry and policy will be continued to further exchange on opportunities for sharing data (in particular those from the private sector), including building on the outcomes of ATLAS D6.4. Industry dialogue also remains vital to receive feedback on the user needs and requirements to inform future datasets and data product development. Highlights of this work were presented at the 12th EMODnet Steering Committee on 21-22 April 2020 and may be further discussed in the context of the European Commission's Marine Knowledge Expert Group (MKEG) which is a network of marine data producers and users, predominantly from the private sector, spanning blue economy areas.

9 Annex I: Table of ATLAS datasets with preliminary indication of relevance to EMODnet

Archive, name	Descriptor, type	Descriptor, domain	Descriptor, parameters	GeoNode, relevance	EMODnet, relevance	EMODnet portal(s)	Title	URL
pangaea	experimental data	biology	exposure	not-relevant	low priority	['Biology']	Results of the cold-water sponge Halichondria panicea exposure experiment to Benzo-A-Pyrene	https://doi.org/10.1594/PANGAEA.911441
pangaea	experimental data	biology	exposure	not-relevant	low priority	['Biology']	Hydrocarbon data of the cold-water sponge Halichondria panicea exposure experiment to Benzo-A-Pyrene	https://doi.org/10.1594/PANGAEA.911443
pangaea	experimental data	biology	feeding	not-relevant	low priority	['Biology']	Feeding biology of two common habitat-forming octocorals in the Azores Archipelago.	https://doi.org/10.1594/PANGAEA.913184
pangaea	experimental data	biology	feeding	not-relevant	yes	['Biology']	Zooplankton capture rates by the octocoral species Dentomuricea aff. meteor and Viminella flagellum in the Azores Archipelago	https://doi.org/10.1594/PANGAEA.913194
pangaea	experimental data	biology	feeding	not-relevant	low priority	['Biology']	Feeding biology of a habitat forming antipatharian in the Azores Archipelago	https://doi.org/10.1594/PANGAEA.913195
pangaea	field observations	biology, chemistry	exposure	not-relevant	low priority	['Biology','Chemistry']	Ingested microplastics identified from macrobenthos collected during the RRS Discovery D340b Research Cruise to the Mingulay Reef Complex (Outer Hebrides, western Scotland, northeast Atlantic)	https://doi.org/10.1594/PANGAEA.897250
pangaea	field observations	biology	inventories	not-relevant	yes	['Biology']	Ground surveys invertebrates sampled in bottom tows from year 2002 to 2013 of the EU NAFO groundfish surveyon boards R/V Vizconde de Eza	https://doi.org/10.1594/PANGAEA.911147
pangaea	field observations	biology	inventories	not-relevant	yes	['Biology']	Presence of Pennatulaceans sampled in bottom tows from year 2002 to 2013 of the EU NAFO groundfish surveyon boards R/V Vizconde de Eza	https://doi.org/10.1594/PANGAEA.911150
pangaea	field observations	biology	inventories	not-relevant	yes	['Biology']	Bottom trawl positions from year 2002 to 2013 of the EU NAFO groundfish surveyon boards R/V Vizconde de Eza	https://doi.org/10.1594/PANGAEA.911151
pangaea	field observations	biology	inventories, traits	not-relevant	yes	['Biology']	Morphological measurements and video annotation of the new octocoral Swiftia phaeton sp. nov. during the cruise MSM 16-3 Phaeton in 2010 off the Mauritanian Slope	https://doi.org/10.1594/PANGAEA.910893
pangaea	field observations	biology	inventory	not-relevant	yes	['Biology']	Species Richness and Taxonomic effort on the study of Octocorallia of the Azores	https://doi.org/10.1594/PANGAEA.889711
pangaea	field observations	biology, environmental	inventories	not-relevant	yes	['Biology','Physics']	Environmental and biological data from deep-sea sponge aggregations in the Faroe-Shetland Channel Nature Conservation Marine Protected Area	https://doi.org/10.1594/PANGAEA.897592
pangaea	field observations	biology, environmental	inventories	not-relevant	yes	['Biology','Physics']	Species and environmental variables per 40m subtransect with coordinates from the Logachev Mound Province	https://doi.org/10.1594/PANGAEA.909118
pangaea	field observations	biology, environmental	inventories	not-relevant	yes	['Biology']	Megafauna presence/absence matrix with Associated Environmental and Spatial Parameters in the Faroe-Shetland Channel	https://doi.org/10.1594/PANGAEA.911023
pangaea	field observations	environmental	bathymetry	uploaded	yes	['Bathymetry']	MBES Bathymetry from Formigas Bank	https://doi.org/10.1594/PANGAEA.890071
pangaea	field observations	environmental	bathymetry	uploaded	yes	['Bathymetry']	MBES bathymetry from Ormonde	https://doi.org/10.1594/PANGAEA.890072
pangaea	field observations	environmental	bathymetry	uploaded	yes	['Bathymetry']	MBES bathymetry from Ormonde to Formigas	https://doi.org/10.1594/PANGAEA.890073
pangaea	field observations	environmental	bathymetry	not-relevant	yes	['Bathymetry']	Two-way vessel track of RV Sarmiento de Gamboa when recording swath bathymetry during MEDWAVES cruise	https://doi.org/10.1594/PANGAEA.890074
pangaea	field observations	environmental	bathymetry, water, atmosphere	not-relevant	already in SeaDataNet?	['Physics']	Full depth ocean properties in the eastern subpolar North Atlantic, Cruise DY052, Extended Ellett Line, 2016, link to raw data in NetCDF format	https://doi.org/10.1594/PANGAEA.881182
pangaea	field observations	environmental	biogeochemistry, fluxes	not-relevant	yes	['Chemsitry']	Benthic oxygen and nitrogen fluxes measured by box core incubation and aquatic eddy covariance technique at a cold-water coral reef in the NE Atlantic	https://doi.org/10.1594/PANGAEA.911412

pangaea	field observations	environmental	physics	not-relevant	yes	['Physics', 'Chemsitry']	Long-term moorings observations in the southwestern Rockall Trough (Logachev cold-water coral mound province)	https://doi.org/10.1594/PANGAEA.911033
pangaea	field observations	environmental	physics	not-relevant	yes	['Physics']	Hydrographic and current velocity data (CTD and LADCP) around the Gazul mud volcano during the MEDWAVES (September 2016)	https://doi.org/10.1594/PANGAEA.911404
pangaea	field observations	environmental	physics	not-relevant	yes	['Physics']	Hydrographic and current velocity data (CTD and LADCP) around the Ormonde seamount during the MEDWAVES (September 2016)	https://doi.org/10.1594/PANGAEA.911405
pangaea	field observations	environmental	physics	not-relevant	yes	['Physics']	Hydrographic and current velocity data (CTD and LADCP) around the Formigas seamount during the MEDWAVES (September 2016)	https://doi.org/10.1594/PANGAEA.911406
pangaea	field observations	environmental	physics	not-relevant	yes	['Physics']	Hydrographic and current velocity data (CTD and LADCP) around the Seco de los Olivos seamount during the MEDWAVES (September 2016)	https://doi.org/10.1594/PANGAEA.911408
pangaea	field observations	environmental	seabed images	not-relevant	low priority as we do not yet have the capacity to manage image data	['Biology']	Seabed images from deep-sea sponge aggregations in the Faroe-Shetland Channel Nature Conservation Marine Protected Area	https://doi.org/10.1594/PANGAEA.897413
pangaea	field observations	environmental	geology	not-relevant	yes	['Geology']	Sediment characteristics of Van Veen samples of MEDWAVES expedition (ATLAS project)	https://doi.org/10.1594/PANGAEA.889083
pangaea	field observations	environmental	geology	not-relevant	yes	['Geology','Biology']	Percentage substrate class per 40 m subtransect with coordinates from the Logachev Mound Province	https://doi.org/10.1594/PANGAEA.909117
pangaea	field observations	environmental	water	not-relevant	yes	['Chemsitry','Physics']	Biogeochemical characterization of water masses over oceanic seamounts off two CTD dives measured during MEDWAVES cruise in 2016	https://doi.org/10.1594/PANGAEA.890197
pangaea	field observations	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera accumulating rates of sediment core GeoB11185-1	https://doi.org/10.1594/PANGAEA.908595
pangaea	model data	environmental	physics, chemistry	uploaded	to check	['Seabed Habitats']	Set of terrain (static in time) and environmental (dynamic in time) variables used as candidate predictors of present-day (1951-2000) and future (2081-2100) suitable habitat of cold-water corals and deep-sea fishes in the North Atlantic	https://doi.org/10.1594/PANGAEA.911117
pangaea	model data	environmental	physics, chemistry	uploaded	to check	['Geology']	GIS layers of seafloor characteristics in the Azores region (North Atlantic), links to files in ArcGIS format	https://doi.org/10.1594/PANGAEA.862152
pangaea	model data	environmental	seabed characteristics	uploaded	to check	['Physics','Chemistyr']	Ocean climatology in the Azores region (North Atlantic) and seabed characteristics, links to GIS layers in ArcGIS format	https://doi.org/10.1594/PANGAEA.872601
pangaea	model data	habitat	indicators -GES	relevant, to upload	to check	['Seabed Habitats']	Shapefiles for ATLAS work in Good Environmental Status in Faroe-Shetland Channel	https://doi.org/10.1594/PANGAEA.911157
pangaea	model data	habitat	indicators -GES	relevant, to upload	to check	['Seabed Habitats']	Shapefiles for ATLAS work on Good Environmental Status in Mingulay Reef Complex	https://doi.org/10.1594/PANGAEA.911163
pangaea	model data	habitat	indicators -GES	not-relevant	to check	['Seabed Habitats']	ATLAS work on Good Environmental Status across 9 study areas in the northeast Atlantic	https://doi.org/10.1594/PANGAEA.911409
pangaea	model data	habitat	inventories, prediction	uploaded	to check	['Seabed Habitats']	Climate-induced changes in the suitable habitat of cold-water corals and commercially important deep-sea fishes in the North Atlantic	https://doi.org/10.1594/PANGAEA.910319
pangaea	model data	habitat	inventories, prediction	relevant, to upload	to check	['Seabed Habitats']	Model output from cold-water corals	https://doi.org/10.1594/PANGAEA.911414
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Silt and age analysis from sediment core KNP178-56JPC	https://doi.org/10.1594/PANGAEA.902449

pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Silt and age analysis from sediment core KNR-178-48JPC	https://doi.org/10.1594/PANGAEA.902454
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Neogloboquadrina pachyderma and age analysis from sediment core KNR158-4-MC10	https://doi.org/10.1594/PANGAEA.902460
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Neogloboquadrina pachyderma and age analysis from sediment core OCE32-MC29D	https://doi.org/10.1594/PANGAEA.902461
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Neogloboquadrina pachyderma and age analysis from sediment core OCE32-MC13A	https://doi.org/10.1594/PANGAEA.902462
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Neogloboquadrina pachyderma and age analysis from sediment core OCE32-MC25A	https://doi.org/10.1594/PANGAEA.902463
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Spheroidal carbonaceous particle from sediment core KNR178-56JPC	https://doi.org/10.1594/PANGAEA.902465
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Spheroidal carbonaceous particle from sediment core KNR178-48JPC	https://doi.org/10.1594/PANGAEA.902468
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Spheroidal carbonaceous particle from sediment core OCE326-MC29D	https://doi.org/10.1594/PANGAEA.902469
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Spheroidal carbonaceous particle from sediment core KNR158-4-MC10	https://doi.org/10.1594/PANGAEA.902471
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Age analysis from sediment core KNR158-4-MC10	https://doi.org/10.1594/PANGAEA.902476
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Age analysis from sediment core OCE326-MC29D	https://doi.org/10.1594/PANGAEA.902481
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Age analysis from sediment core KNR178-48JPC	https://doi.org/10.1594/PANGAEA.902483
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Age analysis from sediment core KNR178-56JPC	https://doi.org/10.1594/PANGAEA.902486
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	210Pb dating from sediment core KNR178-56JPC	https://doi.org/10.1594/PANGAEA.902489
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Lead 210 activity from sediment core KNR178-56JPC	https://doi.org/10.1594/PANGAEA.902490

pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Age and SPG UOHC in 0 - 700 m	https://doi.org/10.1594/PANGAEA.902491
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Alternate binning ages (year CE) are mid-point of bin interval	https://doi.org/10.1594/PANGAEA.902493
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera Mg/Ca, salinity, d18Oseawater, benthic foraminifera accumulating rates, and benthic foraminifera Mn/Ca of sediment core GeoB13731-1, East Melilla	https://doi.org/10.1594/PANGAEA.908461
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera Mg/Ca, salinity, d18Oseawater, benthic foraminifera accumulating rates, grain size, and benthic foraminifera Mn/Ca of sediment core GeoB6718-2	https://doi.org/10.1594/PANGAEA.908469
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera Mg/Ca, salinity, d18Oseawater, benthic foraminifera accumulating rates, grain size, and benthic foraminifera Mn/Ca of sediment core GeoB14885-1, Mauritania Margin	https://doi.org/10.1594/PANGAEA.908470
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera Mg/Ca, salinity, d18Oseawater, benthic foraminifera accumulating rates, grain size, and benthic foraminifera Mn/Ca of sediment core GeoB9064-1, Gulf of Cadiz	https://doi.org/10.1594/PANGAEA.908476
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera Mg/Ca, salinity, d18Oseawater, benthic foraminifera accumulating rates, grain size, and benthic foraminifera Mn/Ca of sediment core GeoB16320-2,Campeche Province	https://doi.org/10.1594/PANGAEA.908477
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera Mg/Ca, salinity, d18Oseawater, and benthic foraminifera Mn/Ca of sediment core GeoB18131-1, West Melilla	https://doi.org/10.1594/PANGAEA.908478
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera Mg/Ca, salinity and d18Oseawater of sediment core GeoB11186-1	https://doi.org/10.1594/PANGAEA.908479
pangaea	model data	environmental	geology	not-relevant	to check	['Geology']	Benthic foraminifera accumulating rates of sediment core TRACOS2010-75, Cape Lookout	https://doi.org/10.1594/PANGAEA.908480
pangaea	model data	paleooceanography	water	not-relevant	to check	['Geology']	Density and age analysis from Central Labrador Sea	https://doi.org/10.1594/PANGAEA.902456
pangaea	model data	paleooceanography	water	not-relevant	to check	['Geology']	Reconstructed water temperature and Salinity fom the north east Labrador Sea	https://doi.org/10.1594/PANGAEA.902459
pangaea	model data	paleooceanography	water	not-relevant	to check	['Geology']	Temperature sub stacks analysis	https://doi.org/10.1594/PANGAEA.902492
pangaea	model data	paleooceanography	water	not-relevant	no	[]	Temperature sub AMOC fingerprints using the EN4 dataset	https://doi.org/10.1594/PANGAEA.902494
pangaea	parent data set			not-relevant	low priority as we do not yet have the capacity to manage image data	['Physics', 'Biology']	Seabed images and corresponding environmental data from deep-sea sponge aggregations in the Faroe-Shetland Channel Nature Conservation Marine Protected Area	https://doi.org/10.1594/PANGAEA.897604
pangaea	parent data set			not-relevant	no	['Geology']	Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years	https://doi.org/10.1594/PANGAEA.902495
pangaea	parent data set			not-relevant	no	['Geology']	Key environmental factors controlling cold-water coral growth over the last 20.000 years	https://doi.org/10.1594/PANGAEA.908558

pangaea	parent data set			not-relevant	yes	['Biology', 'Seabed Habitats']	Species, environmental variables, percentage substrate and coordinates per 40 m video subtransect from the Logachev Mound Province	https://doi.org/10.1594/PANGAEA.909119
zenodo	field observations	biology	map, sample location	not-relevant	yes	['Biology']	Figure1 In Three Species Of Reteporella (Bryozoa: Cheilostomata) In A Diapiric And Mud Volcano Field Of The Gulf Of CÃ¡diz, With The Description Of Reteporella Victori N. Sp	https://doi.org/10.5281/zenodo.1158177
zenodo	lab observation	biology	morphology	not-relevant	yes	['Biology']	Figure 2. Reteporella Mediterranea Hass, 1948.A, Mncn 25.03 In Three Species Of Reteporella (Bryozoa: Cheilostomata) In A Diapiric And Mud Volcano Field Of The Gulf Of CÃ¡diz, With The Description Of Reteporella Victori N. Sp	https://doi.org/10.5281/zenodo.1158178
zenodo	lab observation	biology	morphology	not-relevant	yes	['Biology']	Figure 3 In Three Species Of Reteporella (Bryozoa: Cheilostomata) In A Diapiric And Mud Volcano Field Of The Gulf Of CÃ¡diz, With The Description Of Reteporella Victori N. Sp	https://doi.org/10.5281/zenodo.1158180
zenodo	lab observation	biology	morphology	not-relevant	yes	['Biology']	Figure4 In Three Species Of Reteporella (Bryozoa: Cheilostomata) In A Diapiric And Mud Volcano Field Of The Gulf Of CÃ¡diz, With The Description Of Reteporella Victori N. Sp	https://doi.org/10.5281/zenodo.1158182
zenodo	lab observation	biology	morphology	not-relevant	yes	['Biology']	Figure5. Reteporella Victori N In Three Species Of Reteporella (Bryozoa: Cheilostomata) In A Diapiric And Mud Volcano Field Of The Gulf Of CÃ¡diz, With The Description Of Reteporella Victori N. Sp	https://doi.org/10.5281/zenodo.1158184
zenodo	model data	habitat	inventories	not-relevant	to check	['Seabed Habitats']	Data archive: Niche overlap between a cold-water coral and an associated sponge for isotopically-enriched particulate food sources	https://doi.org/10.5281/zenodo.1198189
zenodo	model data	biology	larvae dispersion	not-relevant	no	[]	Larval dispersal histogram data used for ATLAS deliverable D1.6: Biologically realistic Lagrangian dispersal and connectivity	https://doi.org/10.5281/zenodo.3548344
zenodo	model data	environmental	physics	not-relevant	no	[]	Data set for initializing and forcing of high-resolution local area model implementations in two ATLAS case study areas, Rockall Bank and Condor Seamount	https://doi.org/10.5281/zenodo.3582932
zenodo	field observations	socio-economic	survey	not-relevant	to check	['Human Activities']	Scotland sea and wildlife survey	https://doi.org/10.5281/zenodo.3598266
zenodo	field observations	socio-economic	survey	not-relevant	to check	[]	Norway sea and wildlife survey	https://doi.org/10.5281/zenodo.3598284
zenodo	field observations	socio-economic	survey	not-relevant	to check	[]	North Atlantic high sea survey in Canada	https://doi.org/10.5281/zenodo.3598298
zenodo	field observations	socio-economic	survey	not-relevant	to check	[]	North Atlantic high sea survey in Norway	https://doi.org/10.5281/zenodo.3598301
zenodo	field observations	socio-economic	survey	not-relevant	to check	[]	North Atlantic high sea survey in Scotland	https://doi.org/10.5281/zenodo.3598305
DFO				not-relevant	no	[]	In addition the fishing effort data for the region and elsewhere can be found here	https://open.canada.ca/data/en/dataset/273df20a-47ae-42c0-bc58-01e451d4897a
ENA	field observations	genomic data	metabarcode	not-relevant	no, but see OTU tables at Sextant	[]	metabarcoding of sediment samples	https://www.ebi.ac.uk/ena/browser/view/PRJEB33873
ENA	model data	genomic data	genome	not-relevant	not-relevant, but ultimately, OTU tables	[]	Eunice norvegica (32 Ifremer samples) - end of 2020	end of 2020
ENA	model data	genomic data	genome	not-relevant	no	[]	Deep-sea metazoans (munidopsis, christineconcha, acanella, lophelia, madrepora, eunice, archivesica, calyptogena, dendrophyllia, desmophyllum, paralepetopsis; total 95 Ifremer samples) - end 2020	end of 2020
ENA	model data	genomic data	genome	not-relevant	no	[]	Lophelia pertusa (approx. 60 ifremer samples) - end 2021	end of 2021
ENA	model data	genomic data	genome	not-relevant	no	[]	Madrepora oculata (95 ifremer samples) - end 2021	end of 2021
ENA	model data	genomic data	genome	not-relevant	no	[]	Eunice norvegica (approx. 100 ifremer samples) - end 2021	end of 2021
figshare	field observations	environmental		not-relevant	no	[]	Supplementary material from 'Phylogenetic and functional evidence suggests that deep-ocean	10.6084/m9.figshare.10120151.v1

							ecosystems are highly sensitive to environmental change and direct human disturbance'	
figshare	field observations	environmental		not-relevant	no	[]	Supplementary Material Ashford et al. 'Phylogenetic and functional evidence suggests that deep-ocean ecosystems are highly sensitive to environmental change and direct human disturbance'	10.6084/m9.figshare.8309408.v1
ICES	VME data	habitat	conservation	relevant, to upload	Yes	['Seabed Habitats']	Ormonde VME	https://www.ices.dk/marine-data/data-portals/Pages/vulnerable-marine-ecosystems.aspx
ICES	VME data	habitat	conservation	relevant, to upload	Yes	['Seabed Habitats']	Gazul VME	https://www.ices.dk/marine-data/data-portals/Pages/vulnerable-marine-ecosystems.aspx
ICES	VME data	habitat	conservation	relevant, to upload	Yes	['Seabed Habitats']	Formigas VME	https://www.ices.dk/marine-data/data-portals/Pages/vulnerable-marine-ecosystems.aspx
Ifremer	methods	bioinformatics	code	not-relevant	no	[]	Bioinformatic scripts, config files, and R scripts	https://gitlab.ifremer.fr/abyss-project/
Mendeley	field observations	biology	inventories		low priority (modelled/gridded data?)	['Biology']	Kernel Density Analyses of Coral and Sponge Catches from Research Vessel Survey Data for Use in Identification of Significant Benthic Areas	http://dx.doi.org/10.17632/dtk86rjm86.1
Mendeley	field observations	habitat	seabed images	not-relevant	yes	['Biology']	In Situ Benthic Image Transects from the Disko Fan Conservation Area in the Davis Strait (Eastern Arctic)	http://dx.doi.org/10.17632/cr3xvztwrj.3
Mendeley	model data	biology	inventories	relevant, to upload	yes	['Biology']	Species Distribution Modelling of Corals and Sponges in the Eastern Arctic for Use in the Identification of Significant Benthic Areas	http://dx.doi.org/10.17632/mcb726kcbx.1
Mendeley	model data	biology	inventories	relevant, to upload	yes	['Biology']	Sponge Assemblages and Predicted Archetypes in the Eastern Canadian Arctic	http://dx.doi.org/10.17632/vb4xvxk86v.1
Mendeley	model data	environmental		relevant, to upload	no	[]	Characteristics of Environmental Data Layers for Use in Species Distribution Modelling in the Eastern Canadian Arctic and Sub-Arctic Regions	http://dx.doi.org/10.17632/zmwyjs222s.2
Mendeley	model data	habitat		relevant, to upload	to check	['Seabed Habitats']	Delineation of Coral and Sponge Significant Benthic Areas in Eastern Canada Using Kernel Density Analyses and Species Distribution Models	http://dx.doi.org/10.17632/hnp4xr2sy3.1
seadatanet	field observations	environmental	physics	not-relevant	yes	['Physics']	Time-series physical and chemical transports through the Rockall Trough (2014-2018)	https://cloud.emodnet-ingestion.eu/index.php/s/D8HWQdXjSE98kBJ/
seadatanet	field observations	environmental	physics, chemistry, biology	not-relevant	yes	['Physics']	Full depth ocean properties in the eastern subpolar North Atlantic, Cruise DY052, Extended Ellett Line, 2016, link to raw data in NetCDF format	https://www.bodc.ac.uk/resources/inventories/cruise_inventory/report/16032/
seadatanet	model data	environmental	bathymetry	uploaded	yes	['Bathymetry']	GEBCO_2019	https://doi.org/10.5285/836f016a-33be-6ddc-e053-6c86abc0788e
seadatanet				uploaded	to check	['Human Activities']	OGA Round Indicative Areas (UK Oil&Gas Authority)	https://www.ogauthority.co.uk/data-centre/
sextant/seanoe	field observations	genomic data	metabarcode	not-relevant	no	[]	OTU Tables	https://doi.org/10.12770/0b5d250b-8418-4dda-b39c-960c4481df93
sextant/seanoe	model data	habitat	conservation	uploaded	to check	['Seabed Habitats']	Systematic conservation planning for the North-Atlantic deep sea	https://www.seanoe.org/data/00514/62541/

10 Annex II: Metadata fields required by EMODnet Data Ingestion.

Group	Tag	Name on form	Fields	Single/ Multiple	Data submitter	Automatic	Data repository	remarks
Organisations	Metadata point of contact	**Contact person for this dataset submission form**		Single				
			Name contactperson		X			
			Surname contactperson		X			
			Email contactperson		X			
			Organisation		X			
	Responsible Party	**Organisations responsible for the dataset**		Multiple possible				
			Organisation		X			
			Email organisation		X			
			Role of organisation		X			
Dataset identification	Resource title	**Title of dataset**	**Title of dataset**	Single	X			
	Resource abstract	**Narrative summary of dataset**	**Narrative summary of dataset**	Single	X			
	Resource Type	**Type**	**Type**	Single		X (Dataset)		
	Resource locator	*Download URLs*	*Download URLs*	Multiple possible			X	*NOT on submitter form*
	Unique resource identifier	*Dataset identifiers*	*Dataset identifiers*	Multiple possible			X	*NOT on submitter form*
	Resource language	*Language used in dataset*	*Language used in dataset*	Single		X (Eng(lish))		*Only on publishing form*
	Resource Format	**Data set formats**	**Data set formats**	Multiple possible	X			
	Metadata language	**Language used for completing form**	**Language used for completing form**	Single		X (Eng(lish))		
	Documentation	**Relevant supporting documentation**	**Relevant supporting documentation**	Multiple possible	X			Titles of documentation that should be uploaded as part of the data package
	Lineage	**Summary of data quality processing**	**Summary of data quality processing**	Single	X			
	Metadata date	**Date of metadata creation**	**Date of metadata creation**	Single		X		
		Date of metadata latest revision	*Date of metadata latest revision*	Single			X	*NOT on submitter form*
	Limitations on public access	**Public access**	**Public access**	Single		X(No limitations)		
	Conditions applying to access and use	**License for use**	**License for use**	Single		X(CC-BY-2.0)		
Data types	Mission	**Project / Programme**	**Project / Programme**	Single	X			
	Cruise	*Cruise Summary Reports*	*Cruise Summary Reports*	Multiple possible			X	*NOT on submitter form*
	Platform Name	*Platforms*	*Platforms*	Multiple possible			X	*NOT on submitter form*
	Platform Type	*Platform types*	*Platform types*	Multiple possible			X	*NOT on submitter form*
	Process Methodology	**Summary of processing methodology**	**Summary of processing methodology**	Single	X			
	Parameter/Variables	*Parameters / variables*	*Parameters / variables*	Multiple possible			X	*NOT on submitter form*

	Instruments	*Instruments types*	*Instruments types*	Multiple possible			X	*NOT on submitter form*
	Ocean discovery parameters	**Observation types**	**Observation types**	Multiple possible	X			
	Ocean discovery originating controlled vocabulary	*Ocean discovery originating controlled vocabulary*	*Ocean discovery originating controlled vocabulary*	Single		X (P03)		*Only in XML and database*
	Topic category	*Topic category*	*Topic category*	Single		X(Oceans)		*Only in XML and database*
	Keyword Value -INSPIRE	*Keyword Value -INSPIRE*	*Keyword Value -INSPIRE*	Single		X(Oceanographic Geographical features)		*Only in XML and database*
	Originating controlled vocabulary - INSPIRE	*Originating controlled vocabulary - INSPIRE*	*Originating controlled vocabulary - INSPIRE*	Single		X(GEMET-INSPIRE)		*Only in XML and database*
Locations and Dates	Geographic bounding box	**Geographic Bounding Box**		Single				
			North		X			
			South		X			
			West		X			
			East		X			
	Coordinate reference system	**Coordinate reference system**	**Coordinate reference system**	Single		X(4326=WGS84)		
	Sea areas keyword	**Sea areas**	**Sea areas**	Multiple possible	X			
	Sea areas originating controlled vocabulary	*Sea areas originating controlled vocabulary*	*Sea areas originating controlled vocabulary*	Single		X(C16)		*Only in XML and database*
	Vertical extent, highest level	*Vertical extent of data set*		Single				
			Minimum value				X	*NOT on submitter form*
			Maximum value				X	*NOT on submitter form*
			Unit of Measure				X	*NOT on submitter form*
	Spatial resolution	Spatial resolution	*Spatial resolution value*	Single			X	*NOT on submitter form*
			Spatial resolution unit	Single			X	*NOT on submitter form*
	Temporal extent	**Period of the dataset**	**Start date**	Single	X			
			End date	Single	X			
	Temporal resolution	*Frequency of observations*	*Frequency of observations*	Single			X	*NOT on submitter form*
	Date of publication	*Date of dataset publication*	*Date of dataset publication*	Single			X	*NOT on submitter form*
	Date of last revision	*Date of dataset revision*	*Date of dataset revision*	Single			X	*NOT on submitter form*
	Date of creation	**Date of dataset creation**	**Date of dataset creation**	Single	X			

Document Information

EU Project N°	678760	**Acronym**	ATLAS
Full Title	A trans-Atlantic assessment and deep-water ecosystem-based spatial management plan for Europe		
Project website	www.eu-atlas.org		

Deliverable	**N°**		**Title**	
Work Package	**N°**		**Title**	

Date of delivery	**Contractual**		**Actual**	
Dissemination level		PU Public, fully open, e.g. web		
		CO Confidential restricted under conditions set out in Model Grant Agreement		
		CI Classified, information as referred to in Commission Decision 2001/844/EC		

Authors (Partner)				
Responsible Authors	**Name**		**Email**	

Version log			
Issue Date	**Revision N°**	**Author**	**Change**