Data Management Plan Framework

Date: 30. September 2015: The AtlantOS Data management plan DMP

Date: November 30th 2016. Version No.: 1.1. -Implementation of minimum requirements and guidelines for best practice provided by AtlantOS WP7.

Date: January 30th 2018. Version No.:1.2. -Implementation of AtlantOS Catalogue and GEOSS requirements.

The DMP was produced by AtlantOS WP7 on behalf the AtlantOS management.

Introduction

This Data Management Plan (DMP) is produced as part of the Open Research Data Pilot that AtlantOS comply with. The DMP sets the framework for the handling of data produced in AtlantOS from acquisition over curation to dissemination, and shall thereby assure the implementation of best practice procedures for lifecycle management of AtlantOS data during and beyond the lifetime of the project. This data management plan (DMP) describes the data that will be authored and how the data will be managed and made openly accessible throughout the lifetime of AtlantOS. The content of the data management plan includes:

- the types of data to be managed;

- the standards that would be applied, for example format and metadata content;

- provisions for archiving and long-term preservation;

- access policies and provisions; and

- quality assurance

This DMP is not a fixed document; it evolves and gains more precision and substance during the lifespan of the AtlantOS project. AtlantOS has a specific approach in regard to data management as the project builds up on existing observing networks. The consequence thereof is that WP7 mainly deals with harmonizing data management procedures. Therefore, we foresee that more specific information will be provided in regard to the provenience, discoverability and accessibility to data coming from individual networks that are involved in AtlantOS during the course of the project.

AtlantOS aims to move towards an integrated data system that harmonizes work-flows, data processing, quality assurance procedures, and distribution across in-situ observing network systems in AtlantOS WP2, WP3 and WP4, and integrates in-situ observations into existing European and international data infrastructures, in AtlantOS termed data Integrators.

The overarching objective of AtlantOS is to achieve a transition from a loosely-coordinated set of existing ocean observing activities producing fragmented, often mono-disciplinary data, to a sustainable, efficient, and fit-for-purpose Integrated Atlantic Ocean Observing System. This will be achieved through research and innovation activities focused on: defining requirements and systems design, improving the readiness of observing networks and data systems, engaging stakeholders around the Atlantic, as well as strengthening Europe's contribution to the Global Ocean Observing System (GOOS), a major component of the Group on Earth Observations' (GEO), its Global Earth Observation System of Systems (GEOSS), and specifically on the emerging "Oceans and Society: Blue Planet" initiative. AtlantOS contributes to blue growth by merging new information needs relevant to key sectors such as transport, tourism, fisheries, marine biotech, resource extraction and energy with existing requirements.

The Networks and Integrators participating in the AtlantOS data integration

The Networks

The primary data providers in AtlantOS are existing observing Networks as described in WP2 to WP4 that have been established before the start of AtlantOS:

- **Ship-based observation Networks (WP2)** : GO-SHIP (Global Ocean Ship-based Hydrographic Investigations Program), VOS (Voluntary Observing Ship)/SOOP (Ship of Opportunity Program), CPR (Continuous Plankton Recorder), fish and plankton surveys, seafloor mapping
- **Autonomous observing Networks (WP3)**: Argo, Gliders, Drifters, OceanSITES, EATN (European Animal Tracking Network)
- **Coastal observing systems (WP4)**: Ferrybox, FOS (Fishery Observing System), coastal profilers, fixed moorings

To facilitate the harmonization across the observing networks AtlantOS rely on existing data integrators, i.e. no separate data portal or data archiving system will be build.

The Integrators

The integrators in AtlantOS are:

- For marine environmental data: SeaDataNet for <u>validated and archived</u> data; and the In-Situ Thematic Assembling Centre (INS TAC) component of *Copernicus* Marine Environment Monitoring Service (CMEMS) for <u>NRT (near real-time)</u> data and for the past 60 years of historical data assembled for reanalysis needs
- for marine biodiversity data: the ICES system, and EurOBIS

The Portals involved as integrators in AtlantOS are:

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

Sources and types of data produced as part of AtlantOS

The data landscape of AtlantOS has been analyzed and ocean variables essential to AtlantOS has been selected. These variables are categorized by the data networks as "major" variables or "complementary" variables depending on their relative importance. The following table summarizes the sources and types of data produced and made available by primary data-providers participating in AtlantOS focusing on selected essential variables.

AtlantOS is closely following the continuous debate and recommendations on Essential Ocean Variables (EOVs) and Essential Climate Variables (ECVs). AtlantOS can contribute to that development by providing statistics on what type of parameters are collected on a routinely basis in the involved observing networks and the assessment of the readiness level for that type of observation. Furthermore a comprehensive description of the mission goals and the observational infrastructure can be provided. A link to the GOOS Bio-Eco Panel and the GEO MBON (Marine Biodiversity Observation Network) has been established to support the process of defining ecosystem and biodiversity Essential Ocean Variables.

The data description for AtlantOS will be under continuous development in order to assure a standardized data characterization with better precision and substance during the lifespan of the AtlantOS project.

Table 1: Overview of data resources expected produced by main data networks in AtlantOS WP2-4. The relative importance of each variable for the individual data network is indicated by "M" for major EOV for the network and "A" indicating an additional/complementary EOV for the network. Gliders and Drifters have assigned a percentage of total data acquisitions.

EOV group	EOV	WP2 - ship based observing networks					WP3 - autonomous observing networks					WP4 - coastal observing systems				
		GOSHIP	SOOP	CPR	Fish + plankton survey	Seafloor mapping	Argo	Glider	Drifter	Oceansites	EATN	Tide Gauges	Ferrybox	FOS (RECOPESCA)	Coastal profilers	Fixed moorings
Physics	Temperature	M	A	A	A	A	M	100%	100%	M	A	A	M	M	M	M
	Salinity	M	A	A	A	A	M	100%	49%	M		A	M	M	A	M
	Current	M	A	A	A	A	M	100%	100%	M		A	A	M	A	A
	Sea Level						A			A		M				A
	bottom depth					M	A									
	Radiative fluxes									A						
	Waves									A						A
	Rainfall									A						
	Wind direction									M		A	A			A
	Wind speed									M		A	A			A
	Atmospheric pressure								100%	M		A	A			A
	Air humidity									M		A	A			A
	Air temperature									M		A	A			A
Biogeochemistry	Oxygen	M					M	80%		M	A		M		A	M
	Chla/Fluo	M	M	A	A		M	50%		M			M		A	M
	Nutrients (nitrate NO3,...)	M					M	2%		M			M			
	Carbonate system (inorganic carbon)	M					A	not yet		A			A			
	Dissolved Organic Matter	A								A						
	Transient Tracers	A					A	25%								
	Nitrous Oxide													A	A	A
Biology/ecosystem	Turbidity			M				50%					M			M
	Zooplankton			M	M											
	Phytoplankton	A		M	M											
	Species			M	M						M					
	Eggs and larvae				M											

Standardization and harmonization

During the last decade, a series of standards for data and metadata formats as well as exchange protocols have been established within the marine community where projects and organizations like SeaDataNet and Copernicus played a significant role. AtlantOS will analyze the individual data processing and distribution steps to identify possible gaps and impediments in the use of standards, build on these specifications and facilitate the implementation within the involved data networks. AtlantOS is concerned with improving the interoperability among Networks and Integrators through standardization of basic data features within the AtlantOS community.

The AtlantOS project consists of many mature data providers and data integrators with well-established workflows and data policies, which administers a heterogeneous pool of data resources. Consequently, a prescriptive one-model-fits-all rulebook is neither feasible nor desirable. Hence, AtlantOS shall take a pragmatic approach to improve standardization among the involved Networks and Integrators by defining an essential set of minimum requirements, and set guidelines for best practice. Established infrastructures like Euro-Argo are providing templates for that approach.

In the first period of AtlantOS, Networks and Integrators, through the work of WP7- "Data flow and data integration", have agreed on a set of minimum requirements and provided guidelines for best practice on essential data management aspects that will ensure cross platform coherence and facilitate better data discovery and integration. Following, the charted minimum requirements and guidelines for AtlantOS are listed.

- **Distribution:**
  As the minimum requirement for delivery service, an FTP service at the level of Network data management shall be provided. Additional services such as Web services can also be provided but are not mandatory.

- **Metadata:**
  The minimum metadata requirements are platform or station identifiers and the data provider codes. These are mandatory in the data files and the associated platform catalogues from AtlantOS, thus guaranteeing a continuum between data-platform-institution in an unambiguous way across the Networks. For unique IDs assignments shall be used either (1) C17 controlled vocabulary of SeaDataNet listing the codes for all platforms or (2) ICES station directory. Institutions shall be identified by a unique code from the EDMO (European Directory of Marine Organizations) catalogue.

- **Permanent identifiers for data:**
  A guideline for the *best practice for DOI* assignment to AtlantOS data was formulated and is available online: http://dx.doi.org/10.13155/44515

- **Vocabulary:**
  AtlantOS has a recommended vocabulary for AtlantOS essential variables mapped with existing EU (SeaDataNet vocabularies) or international (CF or WoRMS for Taxa) vocabulary standards. This vocabulary for AtlantOS essential variables and the associated mappings with existing vocabularies is available online.
  https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/A05/

Each Network has to define the mapping between the metadata for the parameters in their data and the standards recommended. By doing this, a Network allows mapping on the fly without having to change its datasets.

- **Platform catalogue:**
  A catalogue technique was formulated to be used at network global data assembling (GDAC) or portal level to facilitate the discovery of platforms and data files. This technique is already implemented at the Copernicus INS TAC. The document describing this recommendation is available online: http://dx.doi.org/10.13155/45063

- **AtlantOS catalogue: All AtlantOS networks, integrators and products relevant for Atlantic Ocean must be described in the AtlantOS catalog**
  The AtlantOS catalogue is the entry point to the integrated data system of AtlantOS. It provides a discovery service to users and it facilitates the access to existing services (viewing, downloading and monitoring) customized to show the Atlantic Ocean as defined within AtlantOS. Entry to the AtlantOS catalog is available here: https://www.atlantos-h2020.eu/atlantos-catalogue/. The sustainability at European level of the catalogue will be studied with EuroGOOS.

- **AtlantOS and GEOSS**. AtlantOS registered end of November 2017 directly with the GEOSS Yellow-Pages that will be fed from AtlantOS catalogue entries. AtlantOS encourages all partners to make their resources available through GEOSS. An AtlantOS community site is under development in the GEOSS portal which will be able to display the AtlantOS community resources. Guidelines for registering resources are available here: https://www.earthobservations.org/documents/gci/201711_gci_manual_01.pdf

An elaborated description of these initiatives can be found in the AtlantOS Data Management Handbook D7.4. This includes an overview of the individual networks data management procedures in regards to data-flow, data-integration, scientific disciplines covered, quality control features, meta-data, data formats and data distribution means.

Also, a core of seven AtlantOS essential variables (T, S, Current, Sea level, Oxygen, Chla, Nitrate and Carbon) are selected for implementation of common Quality Control procedures because they are acquired and controlled in NRT (24h to several days) by more than one Network among the Networks involved in AtlantOS integration activity. The recommendations have been compiled in the AtlantOS QC Report D7.2.

Additionally, in the European legislative context, AtlantOS encourages data providers to comply with metadata encoding following the requirements of the INSPIRE directive as well as the Directive 2003/4/EC[1]. Therefore, INSPIRE compliant XML formats such as ISO 19115 or accordingly enriched simpler formats such as extended Dublin Core are the preferred metadata profile. If metadata is integrated within a data file, the file needs to be in an agreed format such as OceanSites NetCDF or the ASCII MEDAR format. The usage of observing networks specific standards such as the OGC SWE standards will be used to improve e.g. real time data accessibility (OGC SOS) as well as unified documentation of instrumentation (OGC

---

[1]

DIRECTIVE 2003/4/EC „on public access to environmental information and repealing Council Directive 90/313/EEC"

SensorML). For the exchange of metadata, protocols such as OGS CSW or OAI-PMH will be used, alternatively simpler REST formats or sharing of metadata via HTTP or FTP. Used data formats depend on the scientific target user group, common formats are ASCII or NetCDF or for real-time data OGC O&M via SOS as mentioned before.

Data exploitation and reuse strategy

Data produced by observing networks contributing to AtlantOS will be made available through appropriate e-infrastructures capable to comply with the requirements of the Horizon2020 data pilot in particular:

I. Discoverability: Data shall be described by appropriate metadata. Metadata associated with observing networks contributing to AtlantOS data shall meet accepted international standards and shall contain the necessary information to attribute identification, authorship, geographical and temporal coverage, type of measurements and observations, access constraints as well as the responsible organization and/or PI when possible.

II. Accessibility: Data procured with AtlantOS funds and published by AtlantOS partners needs to be openly available, and easily accessible in accordance with the guidelines of the H2020 Open research data pilot. This will be done by facilitating integration in existing integrators (SeaDataNet, Copernicus INSTA, EMODNet, GEOSS, ICES , OBIS, PANGAEA, CORIOLIS)

III. Reusability: Data archived in AtlantOS partner e-infrastructures must be usable beyond the original purpose or scientific community it was collected by. Published data formats shall meet accepted international standards. Reference to existing suitable standards of the discipline has to be given. If these do not exist, an outline on how and what metadata will be created should be provided.

IV. Quality assurance: Data submitted have to be documented properly; documentation is stored together with each dataset. The scientific quality is always in the responsibility of the PI and all authors. Fields for its documentation like quality flags for single values, adjustable precision or documentation of methods are made available. Technical quality control, i.e. completeness of metadata, consistence of formats, and correctness of download is in the responsibility of the data manager.

V. Time compliance: Observing networks contributing to AtlantOS will make such data and metadata publicly available without undue delay.

In fulfillment of the concepts of the Open Research Data Pilot, data providers have to indicate whether they are considering exploiting or disseminating specific data sets. The role of the coordination team together with the steering committee of AtlantOS is to keep track of the implementation of these principles.

Long term perspective:
E-infrastructures (SeaDataNet, Copernicus INSTA, EMODNet, GEOSS, ICES , OBIS, PANGAEA, CORIOLIS) have installed effective procedures to ensure long-term preservation of the data in accordance with international accepted standards for long term archives. It relies on existing facilities such as the network of National Oceanographic Data Centers (NODCs), World Data System (WDS) or existing JCOMM networks facilities.

Principles of access and sharing

Observing nodes under the umbrella of AtlantOS will follow the principle of free and open access to data produced by their facilities and feel committed to work towards the realization of this principle.

AtlantOS aims to identify unnecessary or obsolete barriers towards open access to their data and will continuously work towards the stepwise elimination of these obstacles.

Free and open access without any restrictions shall be granted to the metadata of the data holdings of each observing networks contributing to AtlantOS in order to enable and ease data discovery and fitness-for-use evaluation of the data holdings of each infrastructure.

Data published by observing networks contributing to AtlantOS shall be made available free of charge. RI specific regulations deviating from this general rule may apply when data is used for e.g. non-scientific or commercial purposes. Separate fees may apply for the reproduction and delivery of data when web-based transfer of data is not possible to cover reproduction costs.

In general, data shall be made available by observing networks contributing to AtlantOS as soon as possible and without undue delay.

Each observing networks contributing to AtlantOS shall nominate 'data stewards' to facilitate such data requests. This 'data stewards' are participating in WP7 workpackage

Observing networks contributing to AtlantOS acknowledge national and international intellectual property rights regulations. Each infrastructure is responsible for the warranty of copyrights and intellectual property rights which may apply for its data holdings.

Observing networks contributing to AtlantOS will clearly indicate licenses and terms of use for each dataset in the corresponding metadata.

If data or information produced by observing networks contributing to AtlantOS is used in published or unpublished work, attribution for the used resources is required. Data citations shall exclusively use the information provided within the metadata of each data set.

Principles for re-use and distribution

Data and products from observing networks contributing to AtlantOS will be integrated in existing data integrators facilitating there re-use and re-distribution. As the data will be distributed with a minimum set of metadata that will trace their originators, feedback on their use through the integrators will be feasible. Re-use of non-open and free data will not be facilitated by AtlantOS as these data won't be integrated in the AtlantOs data system. Only discovery will be facilitated by integration in European and international catalogues.

Data Management Plan production

This DMP was produced in a collaborative effort in AtlantOS WP7 on behalf the AtlantOS management with MARUM and IFREMER in leading roles and with inputs from all partners in WP7. The DMP will be developed continuously throughout the lifespan of AtlantOS.