Data Management Plan Framework

Version No.: 1

Date: 30. Sep 2015

Introduction

This data management plan (DMP) describes the data that will be authored and how the data will be managed and made 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

A novelty in Horizon 2020 is the **Open Research Data Pilot** which aims to **improve and maximize access to and re-use** of research data generated by projects. AtlantOS has opted to follow this pilot action and therefore the related regulations set up by the EC apply. It is required to make available the data needed to validate the results presented in scientific publications. Participating projects are receiving dedicated support. Applicants must provide a short, general outline of their policy for data management, including the following issues:

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved?

The 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 of that is that WP7 mainly deals with harmonizing data management procedures and policies. 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. On a technical level these issues are strongly linked to the assignment of DOIs to data ensembles which is an ongoing discussion in some Research Data Alliance (RDA) working groups. In any case this information will be added in the revised version of the data management plan, which is essential to fulfill the integration task that AtlantOS has as one of their priorities.
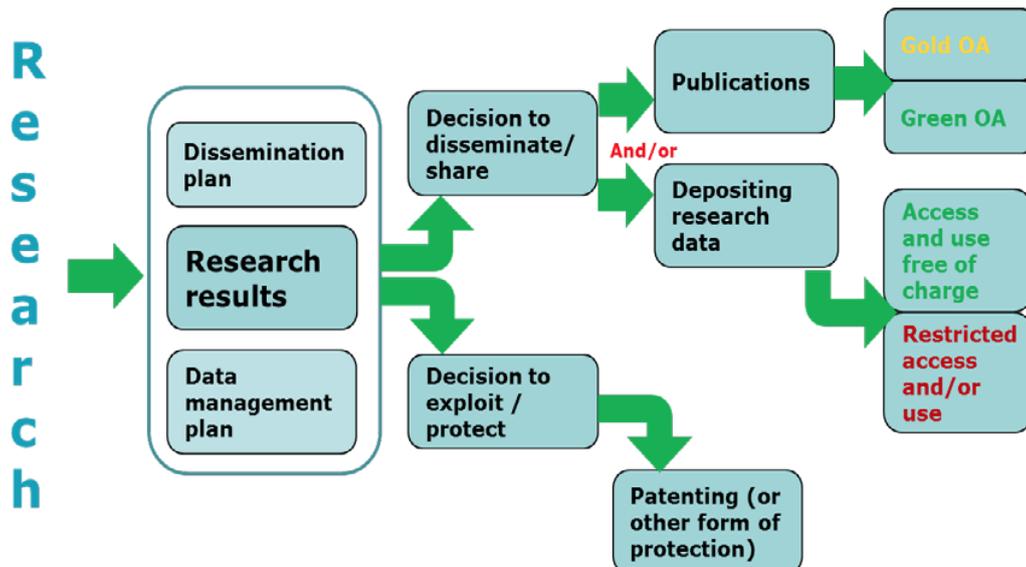
Background of the AtlandtOS project

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 (IAOOS). 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.

Sources and types of data produced by the observing nodes under the umbrella of AtlantOS

The following tables summarize the sources and types of data produced and made available by data-providers participating in AtlantOS. The data providers are divided in 3 categories according to the structure of the Atlantos DOA; ship based observing networks (WP2), autonomous observing networks (WP3) and coastal observing systems (WP4). The data description gives an initial characterization of data available of selected essential ocean variables. AtlantOS is closely following the continuous debate and recommendations on Essential Ocean Variables (EOVs) and Essential Climate Variables (ECVs), and 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: Data contributions from each data providing network in AtlantOS WP2; M indicates a major EOV for the network), A indicates an additional/complementary EOV for the network)

			WP2 - ship based observing networks									
			GOSHIP		SOOP		CPR		Fish + plankton survey		Seafloor mapping	
EOV	Physics	Temperature	M		A	sea surface	A		A		A	All of water column
		Salinity	M		A	sea surface	A		A		A	All of water column
		Current	M									
		Sea Level										
		bottom depth									M	
	Biogeochemistry	Oxygen	M						A			
		Chla/Fluo	M				A	Fluoro				
		Nutrients (**nitrate NO3**,…)	M	**NO$_3$**/NO$_2$, PO$_4$, SiO$_3$								
		Carbonate system (inorganic carbon)	M	**at least two of** pCO$_2$, DIC, TAlk, pH	M	sea surface pCO$_2$						
		Dissolved Organic Matter	A	DOC,DON								
		Transient Tracers	M	CFC-11,CFC-12,SF6,^{3}H/^{3}HE								
		Nitrous Oxide	A	N$_2$O								
		Turbidity										

							M	biomass, abundance	M	biomass, abundance		
Biology/ecosystems		Zooplankton										
		Phytoplankton					M	biomass, abundance				
		Species					M	Species counts (phytoplank. and zooplank.)	M	size, age and maturity		
		Eggs and larvae					M	abundance	M	stages, quantity		

Table 2 Data contributions from each data providing network in AtlantOS WP3. (M indicates a major EOV for the network, A indicates an additional/complementary EOV for the network)

EOV			Argo		Glider		Drifter		Oceansites		EATN	
			WP3 - autonomous observing networks									
	Physics	Temperature	M		100%		100%	sea surface	M			
		Salinity	M		100%		4%	sea surface	M			
		Current	A		100%		100%	sea surface	A			
		Sea Level										
		bottom depth										
	Biogeochemistry	Oxygen	M		80%				M			
		Chla/Fluo	M		50%	Chla			M			
		Nutrients (**nitrate NO$_3$,…**)	M	NO$_3$	2%	NO$_3$			M	NO$_3$		
									A	PO$_4$		
		Carbonate system (inorganic carbon)							A	pH, CO$_2$		
		Dissolved Organic Matter			25%	CDOM						
		Transient Tracers										
		Nitrous Oxide										
	Biol./ecosys	Turbidity			50%							
		zooplankton										
		phytoplankton										
		species									M	acoustic telemet.
		Eggs and larvae										

Table 3 Data contributions from each data providing network in AtlantOS WP3. (M indicates a major EOV for the network, A indicates an additional/complementary EOV for the network)

EOV			WP4 - coastal observing systems			
			Tide Gages		Ferrybox	
	Physics	Temperature			M	sea surface
		Salinity			M	sea surface
		Current				
		Sea Level	M			
		bottom depth				
	Biogeochemistry	Oxygen			M	
		Chla/Fluo			M	
		Nutrients (**nitrate NO$_3$,...**)			M	NO$_3$
		Carbonate system (inorganic carbon)			M	
		Dissolved Organic Matter				
		Transient Tracers				
		Nitrous Oxide				
		Turbidity				
	Biol./ecosys.	zooplankton				
		phytoplankton				
		species				
		Eggs and larvae				

Standardization approach

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 such as SEADATANET or Copernicus. AtlantOS will closely follow these specifications and facilitate the implementation within the involved data providing networks. International integration is a key issue for AtlantOS, thus GEOSS compatibility shall be ensured, which implies that a close interaction with those people who are in charge for the GEOSS Common Infrastructure will be established.

In the European legislative context, metadata encoding has to follow 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 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: 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.

1

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

IV. <u>Quality assurance:</u> 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. <u>Time compliance:</u> 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 in close cooperation with the WP7 leads, supported by the Steering Committee of AtlantOS, is to keep track of the implementation of these principals.

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 - the AtlantOS joint data policy
Description of how data will be shared, including access procedures, outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use.

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.