

**Project Title:** AQUACOSM: Network of Leading European  
AQUatic MesoCOSM Facilities  
Connecting Mountains to Oceans from the Arctic  
to the Mediterranean

**Project number:** 732065

**Project Acronym:** AQUACOSM

**Proposal full title:** Network of Leading European AQUatic MesoCOSM Facilities  
Connecting Mountains to Oceans from the Arctic to the  
Mediterranean

**Type:** Research and innovation actions

**Work program topics  
addressed:** H2020-INFRAIA-2016-2017: Integrating and opening research  
infrastructures of European interest

## Deliverable No 4.2: Database Management Plan adhering to the H2020 Open Research Data Pilot

**Due date of  
deliverable:** 30 June 2017

**Actual submission  
date:** 13 June 2017

**Version:** v1.5

**Main Authors:** Lisette N. de Senerpont Domis (NIOO-KNAW), Deniz Başıoğlu  
(METU), Meryem Beklioglu (METU), Antica Culina (NIOO),  
Simon Keeble (BLIT), Francesca Vidussi-Mostajir (CNRS-  
MARBEC), Dora Neto (CIBIO-UE), Anders Nielsen (AU), Robert  
Ptacnik (WCL), Ioulia Santi (HCMR), Sven Teurlincx (NIOO),  
Daphne Buijert-de Gelder (NIOO)



This project has received funding from the *European Union's Horizon 2020 research  
and innovation programme* under grant agreement No 732065





This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 732065



<b>Project ref. number</b>	<b>731065</b>
<b>Project title</b>	<b>AQUACOSM: NETWORK OF LEADING EUROPEAN AQUATIC MESOCOSM FACILITIES</b>

<b>Deliverable title</b>	Database Management Plan adhering to the H2020 Open Research Data Pilot
<b>Deliverable number</b>	D4.2
<b>Deliverable version</b>	Version 1.4
<b>Contractual date of delivery</b>	30 June 2017
<b>Actual date of delivery</b>	13 June 2017
<b>Document status</b>	FINAL
<b>Document version</b>	
<b>Online access</b>	Yes
<b>Diffusion</b>	Public
<b>Nature of deliverable</b>	ORDP: Open Research Data Pilot
<b>Workpackage</b>	WP4
<b>Partner responsible</b>	NIOO-KNAW
<b>Author(s)</b>	Lisette N. de Senerpont Domis and WP4 workshop on “present status data collection and processing” participants.  Deniz Başıoğlu (METU), Meryem Beklioglu (METU), Antica Culina (NIOO), Simon Keeble (BLIT), Francesca Vidussi-Mostajir (CNRS-MARBEC), Dora Neto (CIBIO-UE), Anders Nielsen (AU), Robert Ptacnik (WCL), Ioulia Santi (HCMR), Sven Teurlincx (NIOO), Daphne Buijert-de Gelder (NIOO)
<b>Editor</b>	
<b>Approved by</b>	Carla Pinho and Jens Nejstgaard (IGB)
<b>EC Project Officer</b>	Agnes Robin

<b>Abstract</b>	AQUACOSM will collect aquatic mesocosm data from 37 facilities throughout Europe. The AQUACOSM objective is to advance mesocosm science through more standardized and synchronized experimentation for better understanding of cause and effect relationships aquatic ecosystem. As part of the H2020 Open Research Data Pilot a Data Management Plan is developed. This Data management Plan sets the guidelines for how data will be generated in a standardized manner, and how data and associated metadata will be made accessible. This Data Management Plan is living document
-----------------	---



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 732065



	and will be updated through the lifecycle of the project.
<b>Keywords</b>	Open data, Metadata, Data collection, Data standardization, Standard Operating Procedures



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 732065





## Table of Contents

1. Executive summary .....	6
2. Data Management Plan .....	7
1. Definitions, acronyms and abbreviations.....	10
2. Data Summary.....	11
3. FAIR data .....	13
3. 1. Making data findable, including provisions for metadata .....	13
3.2. Making data openly accessible .....	14
3.3. Making data interoperable.....	15
3.4. Increase data re-use (through clarifying licences) .....	15
4. Allocation of resources .....	15
5. Data security.....	15
6. Ethical aspects .....	16
7. Other issues .....	16
3. References .....	17



## 1. Executive summary

This deliverable provides the first version of the Data Management Plan of AQUACOSM. AQUACOSM will collect aquatic mesocosm data from 37 facilities throughout Europe. The AQUACOSM objective is to advance mesocosm science through more standardized and synchronized experimentation for better understanding of cause and effect relationships aquatic ecosystem. As part of the H2020 Open Research Data Pilot a Data Management Plan is developed. The development of the Data Management Plan have followed a stepped approach.

First, to gain a broad understanding about perceived challenges and opportunities of mesocosm data collection and management online surveys were distributed among the AQUACOSM partners. This first survey pertaining to data management were filled in for 21 out of 37 facilities.

Second, the results of this survey were discussed during a two-day workshop June 1 and 2, 2017 in Wageningen, The Netherlands.

A first outline of this Data management Plan was drafted during the workshop and reflects the ideas of the workshop participants. This Data Management Plans sets the initial guidelines for how data will be generated in a standardized manner, and how data and associated metadata will be made accessible. This Data Management Plan is a living document and will be updated through the lifecycle of the project.



## 2. Data Management Plan

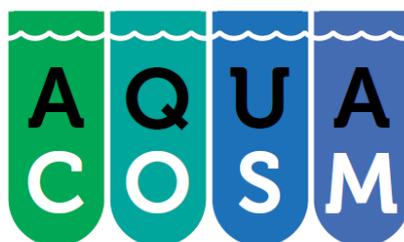


# DATA MANAGEMENT PLAN

**Project Number:** 731065

**Project Acronym:** AQUACOSM

**Project title:** Network of Leading European AQUATIC MesoCOSM Facilities Connecting  
Mountains to Oceans from the Arctic to the Mediterranean





## Table of Contents Data Management Plan

Data Management Plan.....	7
1. Definitions, acronyms and abbreviations.....	10
2. Data Summary.....	11
3. FAIR data .....	13
3.1. Making data findable, including provisions for metadata .....	13
3.2. Making data openly accessible .....	14
3.3. Making data interoperable.....	15
3.4. Increase data re-use (through clarifying licences) .....	15
4. Allocation of resources .....	15
5. Data security.....	15
6. Ethical aspects .....	16
7. Other issues .....	16



# 1. Definitions, acronyms and abbreviations

**BODC:** British Oceanographic Data Centre

**Copernicus:** a European Union Programme aimed at developing European information services based on satellite Earth Observation and in situ (non-space) data.

**DMP:** Data Management Plan

**DOI:** Digital Object Identifier is a persistent identifier used to uniquely identify objects, standardized by the ISO

**EML:** Ecological Metadata Language

**FAIR:** Research data that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

**GB:** Gigabytes

**GitHub:** is a web-based Git or version control repository and Internet hosting service, mostly used for code. A Git is a version control system (VCS) for tracking changes in computer files and coordinating work on those files among multiple people.

**ISO:** International Organization for Standardization, is an international standard-setting body composed of representatives from various national standards organizations.

**Metadata:** data that provides information about other data. Three types of metadata can be distinguished, including descriptive metadata, structural metadata and administrative metadata.

**NERC:** Natural Environment Research Council, the United Kingdom's leading public funder of environmental science

**Open data:** Research data that that can be freely used, re-used and redistributed by anyone for any purpose. Open data is free of restrictions from copyright, patents or other mechanisms of control.

**PhD:** doctoral degree awarded by universities.

**Processed data:** also known as secondary data. This data that has been part of a processing routine, "cleaning" by researchers to remove outliers, obvious instrument reading errors or data entry errors, or any analysis (e.g., determining central tendency aspects such as the average or median result). In addition this data may me been subjected to more statistical forms of analysis

**QA:** Quality Assurance

**QC:** Quality Control

**R:** an open source programming language and software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing

**Raw data:** also known as primary data, is data (e.g., numbers, instrument readings, figures, etc.) collected from a source. Raw data has not been subjected to processing or any other manipulation by a software program or a human researcher, analyst or technician.



**RDA:** Research Data Alliance, is a research community organization started in 2013 by the European Commission, the American National Science Foundation and National Institute of Standards and Technology, and the Australian Department of Innovation. Its goal is to build social and technical infrastructure to enable the open sharing of data

**SOP:** Standard Operating Procedure

**TA:** Transnational Access. Transnational Access means free of charge, trans-national access to research infrastructures or installations for selected user groups. The access includes the logistical, technological and scientific support and the specific training that is usually provided to external researchers using the infrastructure.

**WP:** Work package

## 2. Data Summary

AQUACOSM will collect aquatic mesocosm data from 37 facilities throughout Europe (see [www.aquacosm.eu](http://www.aquacosm.eu)). AQUACOSM is organized in nine work packages (Table 1).

*Table 1: Workpackages in AQUACOSM, acronyms of beneficiaries as detailed in the grant agreement of AQUACOSM*

WP number	WP title	Lead beneficiary
WP1	Consortium Management	1 - FVB-IGB
WP2	Integrated science strategy and governance from local to European scales	2 - UNI
WP3	Transnational network of best practices, harmonization and international training	3 - UiB
WP4	Data collection, standardisation and sharing	4 - NIOO-KNAW
WP5	Outreach: Dissemination and Stakeholder engagement	7 - HCMR
WP6	Provision of trans-national access to all AQUACOSM facilities	1 - FVB-IGB
WP7	Transforming leading freshwater and marine technologies to enable pan-European experimental ecosystem studies in all climates	11 - UMU
WP8	Autonomous measurements	12 - WCL
WP9	Lasting Science Integration	13 - LMU



The AQUACOSM objective is to advance mesocosm science through more standardized and synchronized experimentation for better understanding of cause and effect relationships aquatic ecosystem. To this end a centralized metadata repository will be built, following current standards in metadata vocabulary. Data will be collected within the lifecycle of the project through transnational access (TA) to the different mesocosm facilities, in various freshwater and marine habitats. A survey on data collection, standardization and management carried out within the AQUACOSM consortium showed that for more than 70% of partners the experimental data is currently public, and none is chargeable (Fig. 1).

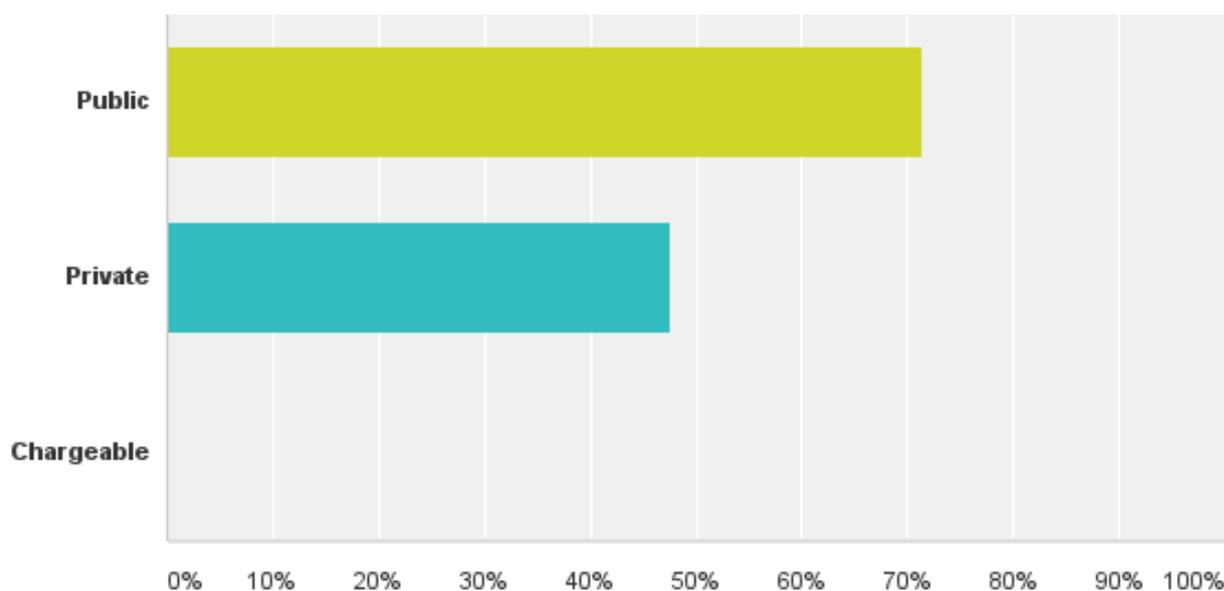


Figure 1: Answer to the question whether experimental data is public, private and/or chargeable (with multiple answers possible). The question was answered for 21 out of 37 facilities.

Standardisation of data collection will be achieved through the development of SOPs and best practices guidelines as developed under WP3 and WP4. Data will be openly accessible through a distributed network as developed in AQUACOSM WP5. The survey showed that more than 85% of the partners have their data stored locally, and less than 10% of the partners have their data stored in open data repositories (Fig 2). The survey clearly showed the current practices in open data management within the consortium and illustrates the challenges in migrating towards open data within the consortium.

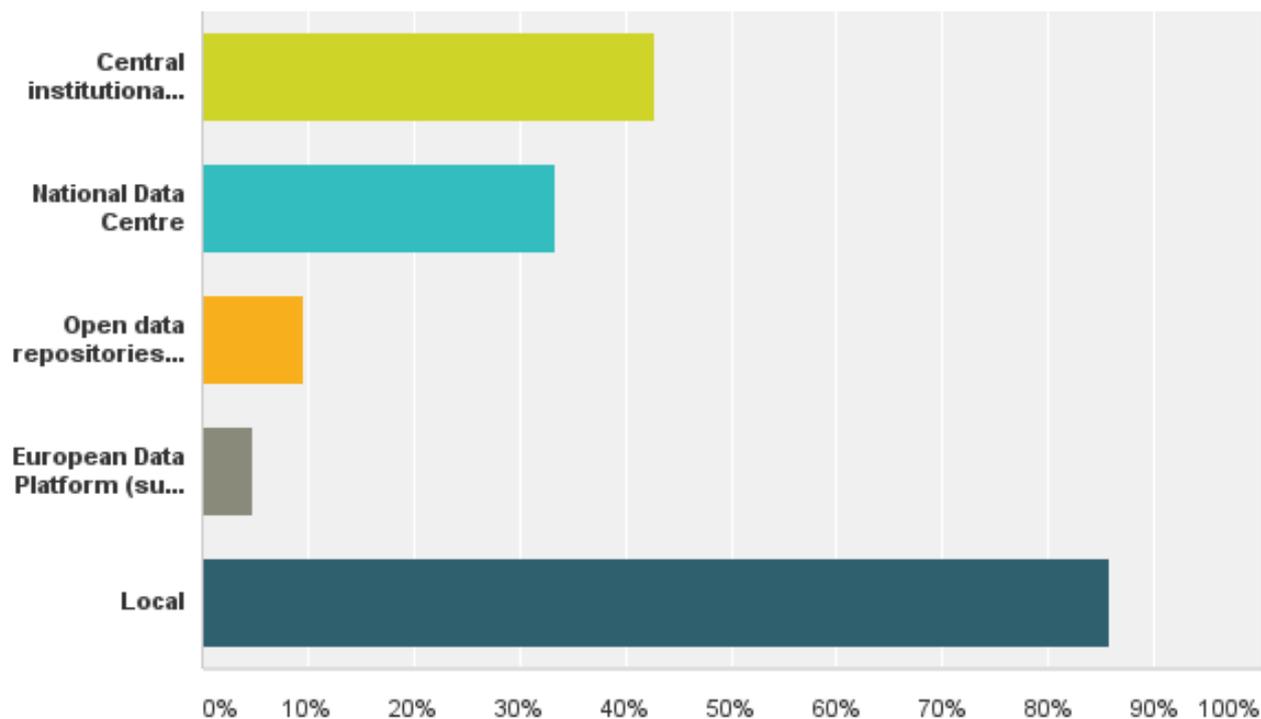


Figure 2: Answer to the question where experimental data is currently stored (with multiple answers possible). The question was answered for 21 out of 37 facilities.

Given these challenges, we attempt to put as little barriers for open access data as possible, and the expected size of the data is relatively small (<20 GB per experiment), we will allow for different data management formats, whilst keeping the metadata format standardized. Terms of use for re-using of the data will follow the Copernicus guidelines ([https://spacedata.copernicus.eu/web/cscda/document-library/-/asset\\_publisher/Hng0LG1u1OHJ/content/data-discovery-and-download-guidelines](https://spacedata.copernicus.eu/web/cscda/document-library/-/asset_publisher/Hng0LG1u1OHJ/content/data-discovery-and-download-guidelines)), but institutional restrictions might apply. If specific restrictions apply, they will be part of the metadata. AQUACOSM data will not only be useful for the current and future generation of mesocosm scientists, but also environmental assessment agencies, water quality managers, and companies with a vested interest in water quality.

### 3. FAIR data

#### 3. 1. Making data findable, including provisions for metadata

The data produced through TA during the lifecycle of the project will be discoverable via a centralised metadata database. This database, will be built as part of the platform of the AQUACOSM web portal, and later transferred to the mesocosm.eu website, the one stop portal for mesocosm science. Metadata can be filled in online and will build upon the Ecological Metadata Language (EML, (Fegraus et al. 2005)), and extended if required, with existing ISO standards. We will maintain an active connection with the Research Data Alliance (RDA) to follow up on standards used there.

The AQUACOSM interface will send periodic updates to existing data portals, including DataOne. Updates will be initiated manually according to an agreed schedule (annually, or more frequently depending on the volume of updates). Upon completion of experiments carried out under AQUACOSM, metadata should be entered in the metadatabase, as part of the transnational access requirements. In



addition, data should be openly accessible within 6 months after completion of the publishable dataset, with reasons given if this is not done. The publishable data set is defined as a dataset that has been subject to processing routines aimed at e.g. QA and QC. These reasons may include competitive advantages such as the completion of a PhD thesis, in which case an embargo of three years will be upheld. All data should have a Digital Object Identifier, with raw and processed data (QA and QC'ed data) deposited separately. File naming conventions and version numbering will be as outlined in the SOPs and best practices guidelines produced within WP4. We will adopt the standards in keywords vocabulary (e.g. BODC common vocabulary of the NERC, see [https://www.bodc.ac.uk/resources/products/web\\_services/vocab/](https://www.bodc.ac.uk/resources/products/web_services/vocab/)) and tailor it to mesocosm science if required. If these existing vocabularies are not complete, keywords will be added, and we will feed this back to dictionary manager, allowing for a full integration of mesocosm science in the scientific realm.

### 3.2. Making data openly accessible

It is our intention that all data produced through TA in the lifecycle of the AQUACOSM project is openly accessible. As part of the TA requirements, data should be made openly accessible after completion of the experiment. As pointed out above data should be openly accessible 6 months after completion of the publishable dataset. For reasons of competitive advantages a data embargo may apply, including the completion of a PhD thesis, in which case an embargo of three years will be upheld. Data will be deposited in a distributed network, where AQUACOSM institutions have freedom to choose their preferred data repository. The associated metadata will be available in centralized portal embedded in the AQUACOSM website (<http://aquacosm.eu/>). Data and associated metadata will be linked. The exact location of data (depending on institutional policy) will be made available through the AQUACOSM website and later through mesocosm.eu. The data is accessible through the AQUACOSM website via one link without password protection. As data is stored in a text based format, no additional software is needed for accessing the data. Software developed under AQUACOSM, e.g. software tools for processing data or automatically QA&QC 'ing of data will be deposited in code repositories (such as GitHub). Where possible existing software will also be made accessible, e.g. existing R scripts. Restriction on use of data, software and code are documented in the AQUACOSM grant agreement, and may vary according to institutional and national policies and legislations. In case of restrictions on use, metadata is still provided, which allow for contacting of the data owner. The request will then be up for consideration of the data owner, and depending on the data owner's decisions full access to the data may be granted.



### 3.3. Making data interoperable

Interoperability of data collected within the AQUACOSM life cycle is promoted through the development of AQUACOSM SOPs and best practices guidelines. Where possible we will follow these AQUACOSM SOPs, if allowed for by institutional policy. In these cases, each partner needs to refer to the standards applied (e.g., institutional standards, international standards/ISO standards), and describe deviations from the AQUACOSM SOPs in the metadata. . We will be using standard metadata vocabulary, which builds on the Ecological Metadata Language (EML) and will be appended if required with other existing ISO90155 conform metadata libraries.

### 3.4. Increase data re-use (through clarifying licences)

Within the lifecycle of AQUACOSM the data made openly available will be licenced following the service and licence commitment of Copernicus (<http://marine.copernicus.eu/services-portfolio/service-commitments-and-licence/>). Data collected under AQUACOSM will be made available for re-use upon completion of the experiment. As pointed out above, for reasons of competitive advantages a data embargo may apply, including the completion of a PhD thesis, which case an embargo of three years will be upheld. Data produced and made openly available under AQUACOSM will be available for third parties. Restriction on use of data, software and code are documented in the grant agreement, and may vary according to institutional and national policies and legislations. In case of restrictions on use, metadata is freely still provided, which allow for contacting of the data owner. The request will then be up for consideration of the data owner, and depending on the data owner's decisions full access to the data may be granted. It is the intention to keep the data available indefinitely but additional costs for keeping the web links alive might be applicable. Data quality assurances will be outlined in SOPs produced within WP4.

## 4. Allocation of resources

Costs of the start-up of making the data FAIR (EuropeanCommission 2016) are covered under AQUACOSM, and pertain to e.g. the centralized metadatabase as well as the steps towards more standardized data collection and processing with SOPS and guidelines on best practices. AQUACOSM does not cover data management resources other than through AQUACOSM website links to the distributed institutional repositories. Costs for supporting people/institutions in making data open access are currently not covered. We intend to provide access to the open data indefinitely; however, curation of the aquacosm.eu web portal after the project end is not included under the current AQUACOSM grant provisions. We roughly expect these additional costs to open access to research data will be an additional 37 days for a data manager, and an additional 5 days for each of the 37 facilities.

Lisette de Senerpont Domis (NIOO-KNAW) will be responsible for data management in the AQUACOSM project, with Simon Keeble (Blue Lobster Ltd) as second in command of data management.

## 5. Data security

None of the data generated in this project are considered sensitive data, thus data security regulations are deemed not necessary. Locally stored data may be at risk for data recovery depending on routines in operation under institutional policies. Best practices will be addressed through provision of SOPs and best practices guidelines in WP4. Partners are expected to adopt suitable tested backup strategy that



allow for full recovery of the data in case of a catastrophic event in which the responsible person or location of the data storage is compromised. The responsibility for data security and long term stories lies with the institutions.

## 6. Ethical aspects

Partners within AQUACOSM need to comply with the Ethics on research integrity as described in the Description of Action. In addition, AQUACOSM partners have to comply with national or international legislation related to data collection. These includes, e.g. legislation on animal experimentation, legislation on privacy of human trial subjects, legislation on experimentation on exotic and/or invasive species.

## 7. Other issues

Data management procedures are open for AQUACOSM partners, although standardization of data management is encouraged through the provision of SOPs and guidelines on best practices as developed under WP4.



### 3. References

European Commission. 2016. Guidelines on FAIR Data Management in Horizon 2020.

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

Fegraus, E. H., S. Andelman, M. B. Jones, and M. Schildhauer. 2005. Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. *Bulletin of the Ecological Society of America* **86**:158-168.



<b>HISTORY OF CHANGES</b>		
<b>Version</b>	<b>Publication date</b>	<b>Change</b>
1.1	5.06.2017	§ Draft outline
1.2	6.06.2017	§ First full draft, send out for internal review under WP4 participants
1.3	13-06-2017	§ Second full draft, send out for internal review under AQUACOSM participants
1.4	23-06-2017	§ Final first version of the DMP, with implemented changes brought up by the AQUACOSM participants
1.5	28-06-2017	§ Final first version of the DMP, with implemented changes brought up by coordinator