

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

Project number: 731065

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 4.4: Updated H2020 Open Research Data Pilot Database Management Plan

Due date of deliverable: 30 June 2018

Actual submission date: 29 June 2018

Version: v2

Main Authors: Lisette N. de Senerpont Domis



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 731065





This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 731065



Project ref. number	731065
Project title	AQUACOSM: NETWORK OF LEADING EUROPEAN AQUATIC MESOCOSM FACILITIES CONNECTING MOUNTAINS TO OCEANS FROM THE ARCTIC TO THE MEDITERRANEAN

Deliverable title	Database Management Plan adhering to the H2020 Open Research Data Pilot
Deliverable number	D4.4
Deliverable version	Version 2.2
Contractual date of delivery	30 June 2018
Actual date of delivery	29 June 2018
Document status	FINAL
Document version	
Online access	Yes
Diffusion	Public
Nature of deliverable	ORDP: Open Research Data Pilot
Workpackage	WP4
Partner responsible	NIOO-KNAW
Author(s)	Lisette N. de Senerpont Domis, Kathryn Keeble, and WP4 workshop on “present status data collection and processing” participants.
Editor	Jens C Nejstgaard, Katharina Makower
Approved by	Jens C Nejstgaard
EC Project Officer	Agnès Robin

Abstract	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 in aquatic ecosystems. 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 a living document and will be updated through the lifecycle of the project.
Keywords	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 731065





Table of Contents

1.	Executive summary	5
2.	Data Management Plan	6
2.1	Definitions, acronyms and abbreviations	8
2.2	Data Summary	9
2.3	FAIR data	12
2.3.1	Making data findable, including provisions for metadata	12
2.3.2	Making data openly accessible	12
2.3.3	Making data interoperable	13
2.3.4	Increase data re-use (through clarifying licences)	13
2.4	Allocation of resources	14
2.5	Data security	14
2.6	Ethical aspects	14
2.7	Other issues	14
2.8	References	15
2.9	Appendix 1	17
2.9.1	Summary of the Data Management Plan	17
2.10	Appendix 2:	18
2.10.1	Guidance notes on Uploading Metadata to the Mesocosm Metadata Catalogue	18



1. Executive summary

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

This data management plan builds on the first version of the data management plan as uploaded to the EC portal (version 1.4) and has been amended for developments that have taken place in the period June 2017 to June 2018. In short, it includes developments of standard operating procedures (SOPS) on data collection, data quality and assurances and processing (D4.1), Guidelines for database management, including controlled vocabulary (D4.3), and development of a metadatabase webfiling system, embedded in the AQUACOSM webportal (www.aquacosm.eu).



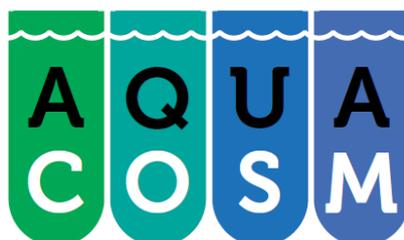
2. Data Management Plan

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:*

2.1	Definitions, acronyms and abbreviations	8
2.2	Data Summary	9
2.3	FAIR data	12
2.3.1	Making data findable, including provisions for metadata	12
2.3.2	Making data openly accessible	12
2.3.3	Making data interoperable	13
2.3.4	Increase data re-use (through clarifying licences)	13
2.4	Allocation of resources	14
2.5	Data security	14
2.6	Ethical aspects	14
2.7	Other issues	14
2.8	References	15
2.9	Appendix 1	17
2.9.1	Summary of the Data Management Plan	17
2.10	Appendix 2:	18
2.10.1	Guidance notes on Uploading Metadata to the Mesocosm Metadata Catalogue.....	18



2.1 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

2.2 Data Summary

AQUACOSM will collect aquatic mesocosm data from 37 facilities throughout Europe (see 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 Transnational 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 in aquatic

ecosystems. 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 (*Figure 1*).

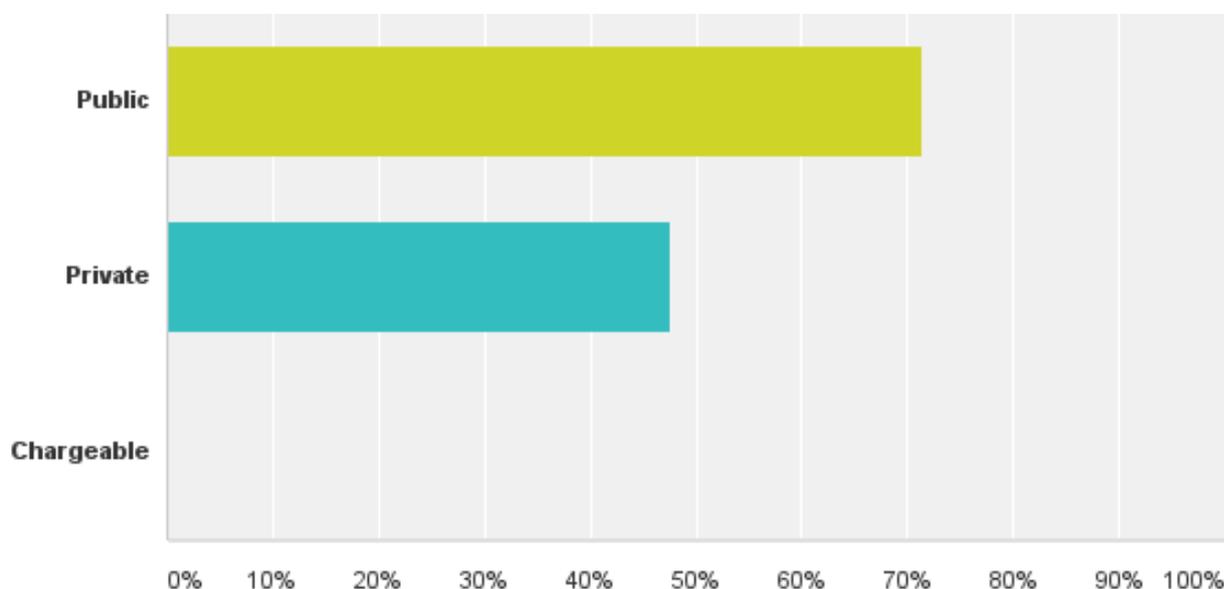


Figure 1: Answer to the question whether experimental data is public, private and/or chargeable (with multiple answers possible). The question was answered for 21 out of 37 facilities.

Standardisation of data collection will be achieved through the development of SOPs and best practices guidelines as developed under WP3 and WP4. These SOPs are developed under a standardized format and include information on best practices in data collection and processing, health and safety and further reading. At present (June 2018) the SOPs displayed in Table 2 are in different stages of revision by the AQUACOSM community. These revisions are necessary to ensure that the best practices as described in the SOPs reflect the diversity of the mesocosm facilities within AQUACOSM, including marine, brackish, stagnant, and running waters.

Table 2: Overview of SOPs developed under AQUACOSM with comments on revision stage,

SOP name	Stage
Phytoplankton	Final
Zooplankton	Revised by core team
Water chemistry	Revised by core team
High Frequency data	Revised by core team
Periphyton	Revised by core team
Quality assurance and Quality control	Revised by core team

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 (Figure 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.

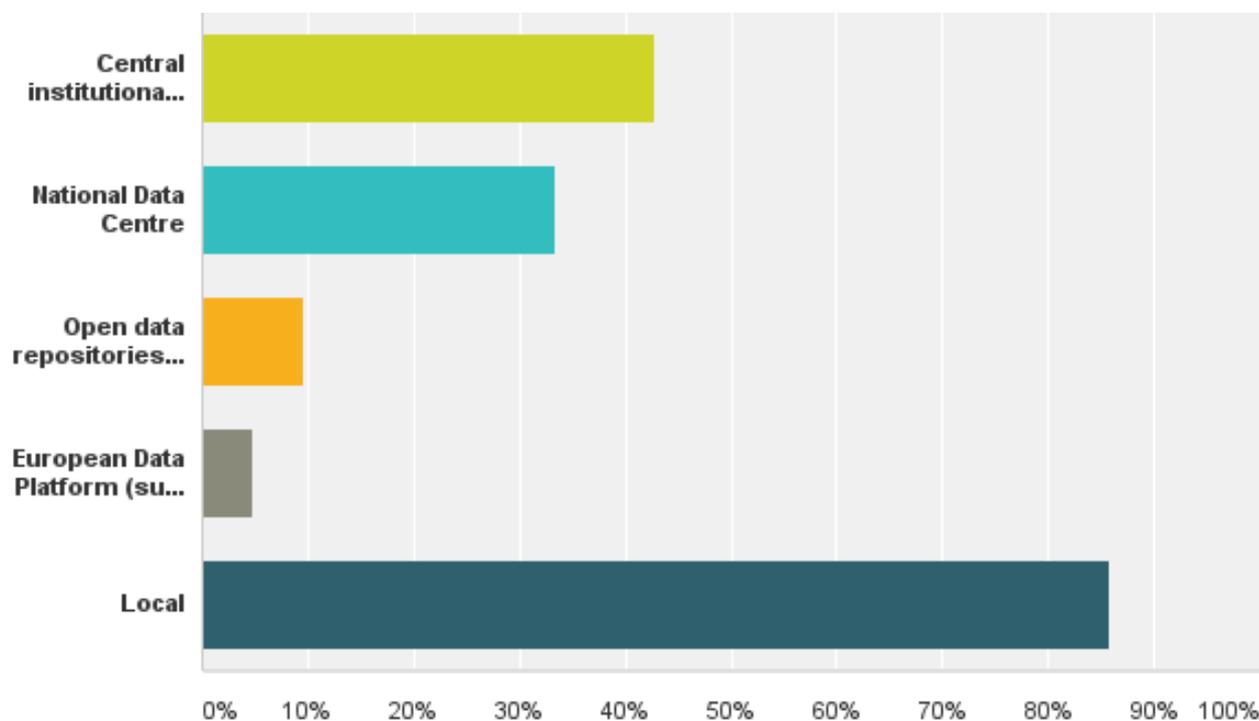


Figure 2: Answer to the question where experimental data is currently stored (with multiple answers possible). The question was answered for 21 out of 37 facilities.

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), 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.



2.3 FAIR data

2.3.1 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, is 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 is currently built upon the Ecological Metadata Language (EML, see Fegeaus et al. 2005), with options to extend if required to accommodate future 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. To facilitate that the AQUACOSM DMP is widely embraced in the TA community, a summary of this DMP is provided to each TA user and provider (see 2.9, Appendix 1). In addition, a guideline on how to use the metadata webfilling system, embedded in the AQUACOSM webportal (www.aquacosm.eu) is provided to both TA user and provider (see 2.10, Appendix 2:).

Apart from metadata, primary 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. For guidance, AQUACOSM has published guidelines for database management, including controlled vocabulary (D4.3).

These reasons for not making the publishable dataset openly accessible 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/) 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.

2.3.2 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. For guidance, AQUACOSM has published guidelines for database management, including controlled vocabulary (D4.3).

The associated metadata will be available in centralized portal embedded in the AQUACOSM website (<http://aquacosc.eu/data>). Data and associated metadata will be linked. The exact location of primary data (depending on institutional policy) will be made available through the AQUACOSM website and later through mesocosc.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 such as, e.g. 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.

2.3.3 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.

2.3.4 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, in 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.



2.4 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 metadata database 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.

2.5 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 storage lies with the institutions.

2.6 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 include, e.g. legislation on animal experimentation, legislation on privacy of human trial subjects, legislation on experimentation on exotic and/or invasive species.

2.7 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.



2.8 References

EuropeanCommission. 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.

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.



HISTORY OF CHANGES		
Version	Publication date	Change
2.1	22.06.2018	§ First full draft, send out to coordinator
2.2	29-06-2018	§ Second full draft, coordinator comments implemented



2.9 Appendix 1

2.9.1 Summary of the Data Management Plan

This is a summary of the Database Management Plan adhering to the H2020 Open Research Data Pilot, version 1.5

Data obtained during AQUACOSM needs to be FAIR. FAIR stands for findable, accessible, interoperable and reusable. Transnational access (TA) requirements to make data FAIR:

- During the AQUACOSM project, we recommend to make use of the AQUACOSM SOPs, as far as this is allowed by institutional policy. Otherwise, partners need to refer to their used standards. You have to describe any deviation from the AQUACOSM SOP in the metadata.
 - Data should be openly accessible within 6 months after completion of the publishable dataset. A publishable dataset is a dataset that has been subject to processing routines aimed at e.g. quality assurance (QA) and quality Control (QC).
 - Store data in a text based format.
 - All data should have a Digital Object Identifier, raw data and processed data (QA QC).
 - Use SOPs and best practice guidelines for file naming conventions and version numbering.
 - The data should be accessible without password protection and without additional software.
 - Data should be deposited in a distributed network. The AQUACOSM program has the freedom to advise their preferred data repository.
 - In case, data is not accessible within 6 months after completion of a TA project, TA users should provide the coordinator and reason why they are not able to meet the deadline. These reasons may include competitive advantages. For example in case of a PhD thesis, an embargo of three years will be upheld.
 - Data produced and made openly available under AQUACOSM will be available for third parties.
- Software developed exclusively under AQUACOSM will be deposited in open code repositories (e.g. GitHub)
- **Metadata** of the experiment, carried out under AQUACOSM, must be available on the Mesocosm Metadata Catalogue, which is available on the AQUACOSM website (<http://aquacosm.eu/data>).
 - Metadata should be provided after completion of the TA
 - The Mesocosm Metadata Catalogue uses a standard metadata vocabulary, built on the Ecological Metadata Language (EML).
 - Each metadata record held on the Mesocosm Metadata Catalogue will provide detailed information on the data collected, including the spatial and temporal coverage.
 - Provision of the metadata is required, even when the 6 month data deadline cannot be reached or when there are restrictions on the use of data, software or codes. The later allows contacting the data owner, who can still grant full access to the data.

Costs for supporting people/institutions in making data open access are not covered.

The responsibility for data security, data sensitivity and long-term storage lies with the institute. The institutes need to adopt a suitable tested back up strategy.

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



2.10 Appendix 2:

2.10.1 Guidance notes on Uploading Metadata to the Mesocosm Metadata Catalogue.

CONTENTS

<i>Guidance notes on Uploading Metadata to the Mesocosm Metadata Catalogue.</i>	18
Introduction	19
Accessing the Mesocosm Metadata Catalogue	19
Creating a new record	20
Notes for filling out the Template	21
Getting Help	22
Glossary of terms:	23
Dataset (resource)	23
General Keywords.....	26
Coverage	26
Methods.....	27
Project.....	27
Intellectual property rights	28
Additional metadata	28

1. INTRODUCTION

The Mesocosm Metadata Catalogue is a bespoke Meta-database for recording all types of metadata describing data collected using Mesocosms.

The catalogue is built using the GeoNetwork application (<https://geonetwork-opensource.org>). The Metadata template uses standard metadata vocabulary defined by the Ecological Metadata Language (EML) standards. The standard conforms to the ISO standards.

2. ACCESSING THE MESOCOSM METADATA CATALOGUE

The first step is to register and request editor access to the Mesocosm Metadata Catalogue:

- 1) Go to <http://aquacosm.eu/data> and click on “Sign in” located top-right (Figure 3)
- 2) Next, click on “Create an account” (Figure 4)
- 3) Fill in your details to create your account and select “Editor” from the “Requested profile” dropdown box. Then click register (Figure 5). Users with Editor privileges are permitted to enter and edit data.
- 4) You will receive an email containing your login details and you will be granted “Registered User” status on the site. An email will also be sent to the administrator notifying them that you have requested an Editor User profile. The administrator will be required to upgrade your account to grant you Editor User privileges before you are able to input any data. The administrator will email you once your request has been granted.
- 5) Once you have been approved as an Editor, you will have permission to enter and edit datasets.

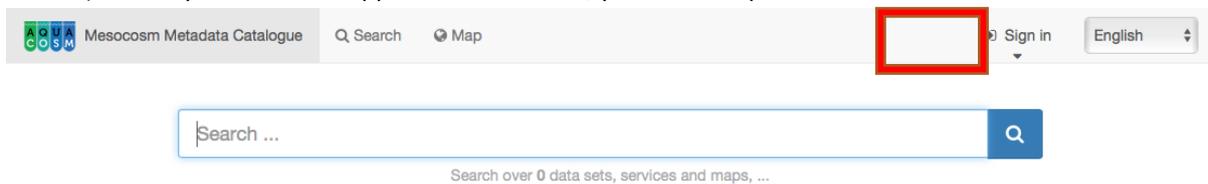


Figure 3 Go to <http://aquacosm.eu/data> and click on the “Sign in” icon shown in the red box

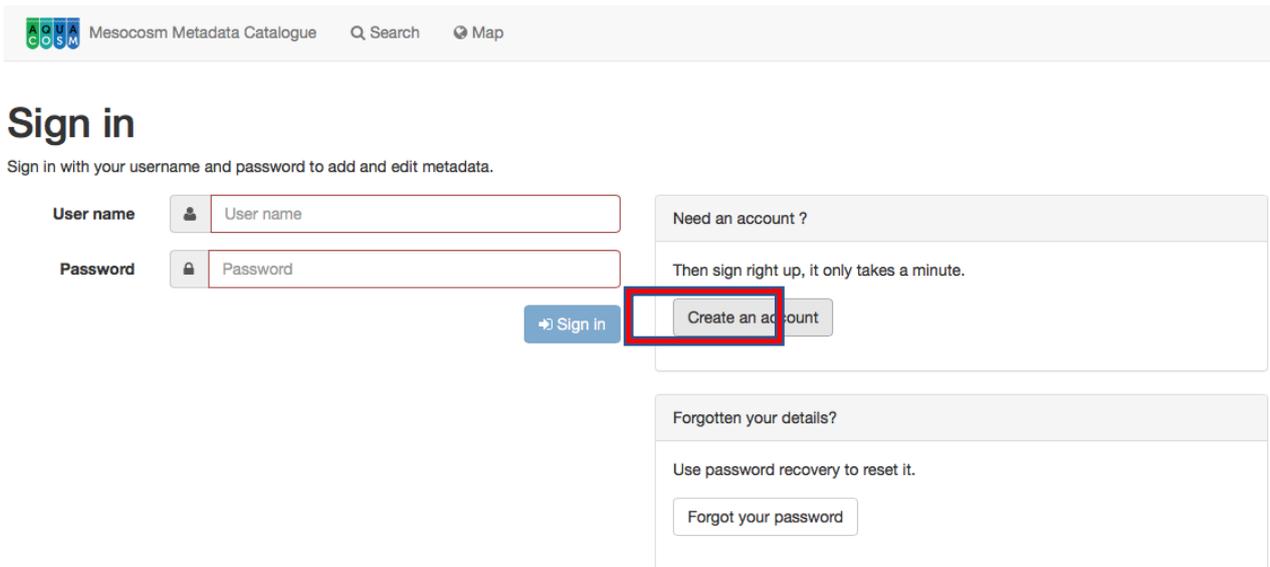


Figure 4: Select “Create an account” by clicking on the “Create an account” button shown in the red box.

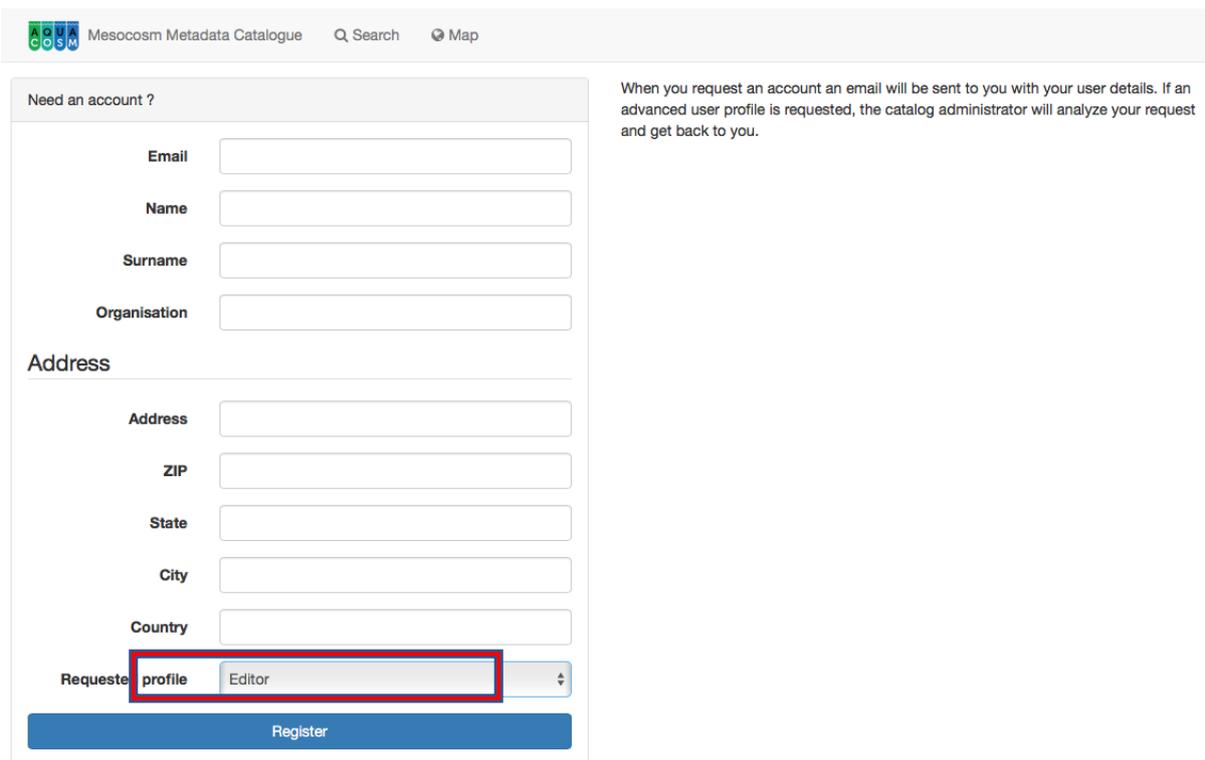


Figure 5: Create an account form. Complete the details, select “Editor” from the “Requested profile” dropdown list (highlighted in red) and click “Register”. You will shortly receive an email with your login details. The Administrator will also receive an email notifying them that you have requested an “Editor” user profile.

3. CREATING A NEW RECORD

This section guides you through the process of adding new metadata records with associated data and/or services into the catalogue.

From the home page, go to the *contribute* section or directly click on add new record menu (Figure 6).

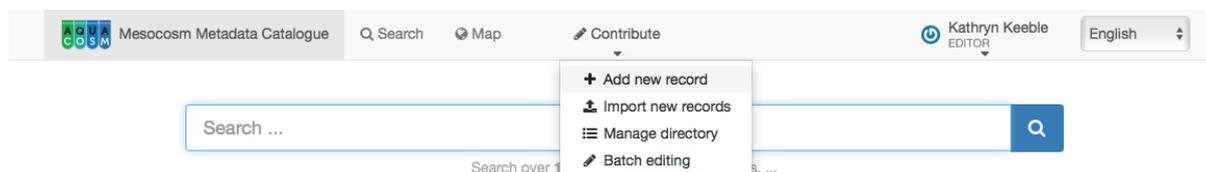


Figure 6: Shows how to select “Add new record” from the “Contribute” menu

Then choose “Create a Dataset” and the “Template for EML” (both will be selected by default) and click “Create” (Figure 7).

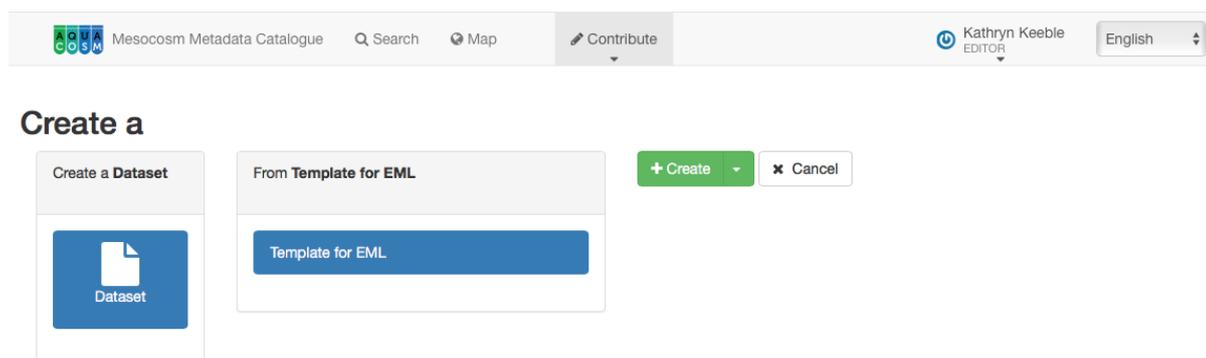


Figure 7: Shows the next screen with “Create a Dataset” and “Template from EML” selected. Click the green “Create” button to the right of the screen to open the form.

This will open the template.

The next step is to fill out as completely as possible the fields provided by default in the template.

4. NOTES FOR FILLING OUT THE TEMPLATE

The elements of the AQUACOSM metadatabase are categorised as follows:

Dataset (resource)

General Keywords

Coverage

Methods



Project

Intellectual property rights

Additional metadata

For each of these elements more detailed information needs to be provided. In the “glossary of terms” section below all elements and subfields are defined. In addition, visit <https://github.com/gbif/ipt/wiki/GMPHowToGuide> for additional information.

Use the Validation button to check that all mandatory fields (indicated by an asterix (*)) have been completed.

Save your data regularly!

5. GETTING HELP

For additional technical queries or queries relating to the metadata template, please contact us at metadata@aquacosm.eu.



6. GLOSSARY OF TERMS:

7. DATASET (RESOURCE)

The dataset field has elements relating to a single dataset (resource). Keep in mind that data (primary data) and metadata require separate details that need to be filled in. Although data and metadata providers may be the same person, this is not necessarily so.

This section contains the following elements:

Alternate identifier: Primary key generated by the web filing system.

Title: A description of the resource that is being documented that is long enough to differentiate it from other similar resources. Multiple titles may be provided, particularly when trying to express the title in more than one language (use the "xml:lang" attribute to indicate the language if not English/en). E.g. Vernal pool amphibian density data, Isla Vista, 1990-1996.

Publication date: The date that the resource was published. The format should be represented as: CCYY, which represents a 4 digit year, or as CCYY-MM-DD, which denotes the full year, month, and day. Note that month and day are optional components. Formats must conform to ISO 8601. E.g. 2010-09-20.

Resource language: The language in which the resource (not the metadata document) is written. This can be a well-known language name, or one of the ISO language codes to be more precise. Recommendation is to use the ISO language code (<http://vocabularies.gbif.org/vocabularies/lang>). E.g., English.

Additional info: Information regarding omissions, instructions or other annotations that resource managers may wish to include with a dataset. Basically, any information that is not characterized by the other resource metadata fields.

Url: The URL of the resource that is available online.

Abstract: A brief overview of the resource that is being documented

8. DATASET CREATOR

The dataset creator field contains subfields detailing information on the organization and/or person that is associated with the primary data of the resource.

Organization name: The full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. E.g., National Center for Ecological Analysis and Synthesis.

Given name: The given name field can be used for the first name of the individual associated with the resource, or for any other names that are not intended to be alphabetized (as appropriate). E.g., Jonny

Surname: The surname field is used for the last name of the individual associated with the resource. This is typically the family name of an individual, for example, the name by which s/he is referred to in citations. E.g. Carson

Position name: The position that the individual associated with the resource holds, i.e. PhD, Postdoc, Professor, Senior Scientist/Faculty

9. Address



This address field contains detailed information on the address that is associated with the dataset creator.

Delivery point: Subfield of the address field that describes the physical or electronic address of the responsible party for a resource. The delivery point field is used for the physical address for postal communication. E.g., GBIF Secretariat, Universitetsparken 15.

City: Subfield of the address field that describes the city of the responsible party for a resource.

Administrative area: Subfield of the address field that describes the administrative area of the responsible party for a resource.

Country: Subfield of the address field that describes the city of the responsible party for a resource.

Phone: Subfield of the address field that describes the phone number of the responsible party for a resource.

Email: Subfield of the address field that describes the email of the responsible party for a resource.

Online URL: A link to associated online information, usually a web site. When the party represents an organization, this is the URL to a website or other online information about the organization. If the party is an individual, it might be their personal web site or other related online information about the party. E.g., <http://www.yourdomain.edu/~doe>.

10. METADATA PROVIDER

The metadata provider field contains subfield detailing information on the organization and/or person that is associated with the metadata of the resource.

Organization name: The full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. E.g., National Center for Ecological Analysis and Synthesis.

Given name: The given name field can be used for the first name of the individual associated with the resource, or for any other names that are not intended to be alphabetized (as appropriate). E.g., Jonny

Surname: The surname field is used for the last name of the individual associated with the resource. This is typically the family name of an individual, for example, the name by which s/he is referred to in citations. E.g. Carson

11. Address

This address field contains detailed information on the address that is associated with the dataset creator.

Delivery point: Subfield of the address field that describes the physical or electronic address of the responsible party for a resource. The delivery point field is used for the physical address for postal communication. E.g., GBIF Secretariat, Universitetsparken 15.

City: Subfield of the address field that describes the city of the responsible party for a resource.

Administrative area: Subfield of the address field that describes the administrative area of the responsible party for a resource.

Country: Subfield of the address field that describes the city of the responsible party for a resource.

Phone: Subfield of the address field that describes the phone number of the responsible party for a resource.



Email: Subfield of the address field that describes the email of the responsible party for a resource.

Online URL: A link to associated online information, usually a web site. When the party represents an organization, this is the URL to a website or other online information about the organization. If the party is an individual, it might be their personal web site or other related online information about the party. E.g., <http://www.yourdomain.edu/~doe>.

12. ASSOCIATED PARTY

An associated party is another person or organisation that is associated with the resource. These parties might play various roles in the creation or maintenance of the resource, and these roles should be indicated in the "role" element.

Organization name: The full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. E.g., National Center for Ecological Analysis and Synthesis.

Given name: The given name field can be used for the first name of the individual associated with the resource, or for any other names that are not intended to be alphabetized (as appropriate). E.g., Jonny.

Surname: The surname field is used for the last name of the individual associated with the resource. This is typically the family name of an individual, for example, the name by which s/he is referred to in citations. E.g. Carson.

Phone: Subfield of the address field that describes the phone number of the responsible party for a resource.

Role: Use this field to describe the role the party played with respect to the resource. E.g. technician, reviewer, principal investigator, etc.

13. DATASET CONTACT

The contact field contains contact information for this dataset. This is the person or institution to contact with questions about the use, interpretation of a data set.

Organization name: The full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. E.g., National Center for Ecological Analysis and Synthesis.

Given name: The given name field can be used for the first name of the individual associated with the resource, or for any other names that are not intended to be alphabetized (as appropriate). E.g., Jonny.

Surname: The surname field is used for the last name of the individual associated with the resource. This is typically the family name of an individual, for example, the name by which s/he is referred to in citations. E.g. Carson.

14. Address

This address field contains detailed information on the address that is associated with the dataset creator.

Delivery point: Subfield of the address field that describes the physical or electronic address of the responsible party for a resource. The delivery point field is used for the physical address for postal communication. E.g., GBIF Secretariat, Universitetsparken 15.

City: Subfield of the address field that describes the city of the responsible party for a resource.



Administrative area: Subfield of the address field that describes the administrative area of the responsible party for a resource.

Country: Subfield of the address field that describes the city of the responsible party for a resource.

Phone: Subfield of the address field that describes the phone number of the responsible party for a resource.

Email: Subfield of the address field that describes the email of the responsible party for a resource.

Online URL: A link to associated online information, usually a web site. When the party represents an organization, this is the URL to a website or other online information about the organization. If the party is an individual, it might be their personal web site or other related online information about the party. E.g., <http://www.yourdomain.edu/~doe>.

15. GENERAL KEYWORDS

The keyword fields contain elements for both the keywords and keywords thesaurus elements, both of which are required together.

Keyword: A keyword or key phrase that concisely describes the resource or is related to the resource. Each keyword field should contain one and only one keyword (i.e., keywords should not be separated by commas or other delimiters). The keywords should reflect the experimental treatments (manipulation of independent parameters) being applied in the mesocosm studies, e.g. browning, climate warming, microplastics.

Thesaurus: The name of the official keyword thesaurus from which the keywords were derived. If an official thesaurus name does not exist, please keep a placeholder value such as "N/A" instead of removing this element as it is required together with the keyword element to constitute a keywordSet. E.g., IRIS keyword thesaurus.

16. COVERAGE

Describes the extent of the coverage of the resource in terms of its spatial extent, temporal extent, and taxonomic extent.

17. GEOGRAPHIC EXTENT

A container for spatial information about a resource; allows a bounding box for the overall coverage (in lat long), and also allows the user to draw the spatial extent of the dataset.

Geographic description: A short text description of a dataset's geographic areal domain. A text description is especially important to provide a geographic setting when the extent of the dataset cannot be well described by the "bounding box". E.g., "Manistee River watershed", "extent of 7 1/2 minute quads containing any property belonging to Yellowstone National Park".

18. BOUNDING BOX

Bounding box: Draw the spatial extent or enter the coordinates in the boxes provided. The coordinates will be saved to the XML file.

19. TEMPORAL EXTENT

This field allows coverage to be a single point in time, multiple points in time, or a range of dates.



Begin date: A single time stamp signifying the beginning of the experimental period. The calendar date field is used to express a date, giving the year, month, and day. The format should be one that complies with the International Standards Organization's standard 8601. The recommended format for EML is YYYY-MM-DD, where Y is the four digit year, M is the two digit month code (01 - 12, where January = 01), and D is the two digit day of the month (01 - 31). This field can also be used to enter just the year portion of a date. E.g. 2010-09-20.

End date: A single time stamp signifying the end of the experimental period. The calendar date field is used to express a date, giving the year, month, and day. The format should be one that complies with the International Standards Organization's standard 8601. The recommended format for EML is YYYY-MM-DD, where Y is the four digit year, M is the two digit month code (01 - 12, where January = 01), and D is the two digit day of the month (01 - 31). This field can also be used to enter just the year portion of a date. E.g. 2010-09-20.

20. TAXONOMIC EXTENT

This field contains taxonomic information about a resource. It includes a list of species names (or higher level ranks) from one or more classification systems. Please note the taxonomic classifications should not be nested, just listed one after the other.

General Taxonomic Coverage: A description of the range of taxa addressed in the data set or collection. Use a simple comma separated list of taxa. E.g., "All vascular plants were identified to family or species, mosses and lichens were identified as moss or lichen."

Taxonomic rank value: The name representing the taxonomic rank of the taxon being described. E.g. *Acer* would be an example of a genus rank value, and *rubrum* would be an example of a species rank value, together indicating the common name of red maple. It is recommended to start with Kingdom and include ranks down to the most detailed level possible.

21. METHODS

This field documents scientific methods used in the collection of the resource. It includes information on items such as tools, instrument calibration and software. Importantly, this element should reflect the parameters measured.

Method step: This field allows for repeated sets of elements that document a series of procedures followed to produce a data object. These include text descriptions of the standard operating procedures followed, relevant literature, software, instrumentation, and source data. E.g. AQUACOSM zooplankton SOP, sedimentation method (Üthermohl, 1958), Determination of phosphorus according to Murphy (1962).

Study extent: This field allows for a textual description of the specific sampling area, the sampling frequency (temporal boundaries, frequency of occurrence), and the response variables measured (e.g. organism groups, such as macrofauna, phytoplankton, and physical-chemical parameter such as pH, Phosphorus). The field "Study extent" represents both the specific parameters measured and the sampling frequency (temporal boundaries, frequency of occurrence).

Sampling description: The "sampling description" field allows for a text-based/human readable description of the sampling procedures used in the research project. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article.

Quality control: The quality control field provides a location for the description of actions taken to either control or assess the quality of data resulting from the associated method step.

22. PROJECT



The project field contains information on the project in which this dataset was collected. It includes information such as project personnel, funding, study area, project design and related projects.

Title: This field should contain both the official TA project title (if applicable as well as a descriptive title for the research project. E.g. Globalrest: Global-change impacts on restored and non-restored ecosystems: Biological responses through different trophic levels

23. PERSONNEL

Surname: The surname field is used for the last name of the individual associated with the project in which this dataset was collected. For TA projects this would be the project leader of the TA proposal. This is typically the family name of an individual, for example, the name by which s/he is referred to in citations. E.g. Carson.

Role: Use this field to describe the role the party played with respect to the resource. E.g. technician, reviewer, principal investigator, etc.

Funding: The funding field is used to provide information about funding sources for the project such as: grant and contract numbers; names and addresses of funding sources. Here, users do not only state AQUACOSM as a funding source, but also any additional funding received to carry out the research

Study area description: The “study area description” field documents the physical area associated with the research project. It can include descriptions of the geographic, temporal, and taxonomic coverage of the research location and descriptions of domains (themes) of interest such as climate, geology, soils or disturbances.

Design description: The field design description contains general textual descriptions of research design. It can include detailed accounts of goals, motivations, theory, hypotheses, strategy, statistical design, and actual work. Literature citations may also be used to describe the research design.

24. INTELLECTUAL PROPERTY RIGHTS

Intellectual property rights: A rights management statement for the resource, or reference a service providing such information. Rights information encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. In the case of a data set, rights might include requirements for use, requirements for attribution, or other requirements the owner would like to impose. E.g., Copyright 2001 Regents of the University of California Santa Barbara. Free for use by all individuals provided that the owners are acknowledged in any use or publication.

Purpose: A description of the purpose of this dataset.

25. ADDITIONAL METADATA

Data stamp: The date time the metadata document was created or modified. E.g., 2002-10-23T18:13:51.235+01:00. This is automatically generated.

Citation: The citation for the work itself. This field reflects how the dataset should be cited. E.g. Peter B. Adler, William R. Tyburczy, and William K. Lauenroth. 2007. Ecological Archives E088-161-D1

URL: URL associated with a resource. E.g., <http://esapubs.org/archive/ecol/E088/161/metadata.htm>