www.aquacosm.eu

TRAINING VIDEO AND GUIDANCE DOCUMANTATION FOR TA USERS

Deliverable No: 4.3

Deliverable title	**Training video and guidance documentation for TA users**
Deliverable number	D4.3
Deliverable version	1
Contractual date of delivery	31 March 2020 (M12)
Actual date of delivery	30 March 2020 (M12)
Dissemination level	
• **PU** Public	
Nature of deliverable	
• **OTHER**	
Work package	WP4
Lead Beneficiary	HCMR
Author(s)	Katerina Symiakaki (HCMR), Paraskevi Pitta (HCMR), Lisette de Senerpont Domis (NIOO)
Editor	Katharina Makower (IGB), Jens Nejstgaard (IGB)
EC Project Officer	Blagovesta CHOLOVA

Table of Contents

1.   Abstract

A central goal of the AQUACOSM-plus project is that all data produced during the project shall be findable, accessible, interoperable and re-usable (FAIR). When this is achieved, we propose that the data from AQUACOSM-plus will not only be useful for the current and future mesocosm scientists, but also for environmental assessment agencies, water quality managers, and companies with interest in water quality. To support the development towards open mesocosm data, a centralized metadata portal for mesocosm data was created in the previous H2020 project AQUACOSM that will be updated in AQUACOSM-plus. This deliverable describes the guidance video and documentation that were created to facilitate entering metadata in the Mesocosm   Metadata   Catalogue (https://www.aquacosm.eu/geonetwork/srv/eng/catalog.search#/home).

2.   Introduction

It is the goal of AQUACOSM-plus to have all data produced during the lifetime of the project to be findable, accessible, interoperable and re-usable (FAIR). AQUACOSM-plus 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. As part of the approach towards open mesocosm data, a Meta-database was created. The Mesocosm Metadata Catalogue (https://www.aquacosm.eu/geonetwork/srv/eng/catalog.search#/home) is a Meta-database tailored 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. The AQUACOSM-plus metadata interface is publicly exposed via an API and we will contact relevant data gateways (Pangaea, EMODnet etc.) to encourage them to harvest the metadata.

Upon completion of experiments carried out by Transnational Access Users under AQUACOSM and AQUACOSM-plus, metadata should be entered in the metadata database, as part of the Transnational Access (TA) requirements. Apart from details on the mesocosm experimental design and methodology, the metadata also contains the contact details of the data owner as well as the Digital Object Identifier (DOI), if available. As an important step on the way to

remove potential barriers for using the metadata portal, we here report on instructional videos and a manual that have been created to provide step-by-step guidance to entering metadata.

3.   Videos

Three training videos were recorded with the goal to be shared with TA users, helping them throughout the process of providing the metadata related to their projects. Videos were recorded using screen capturing software. We recorded an introductory video to illustrate the need and the importance of metadata and why users have to provide them (***Figure** 1*).
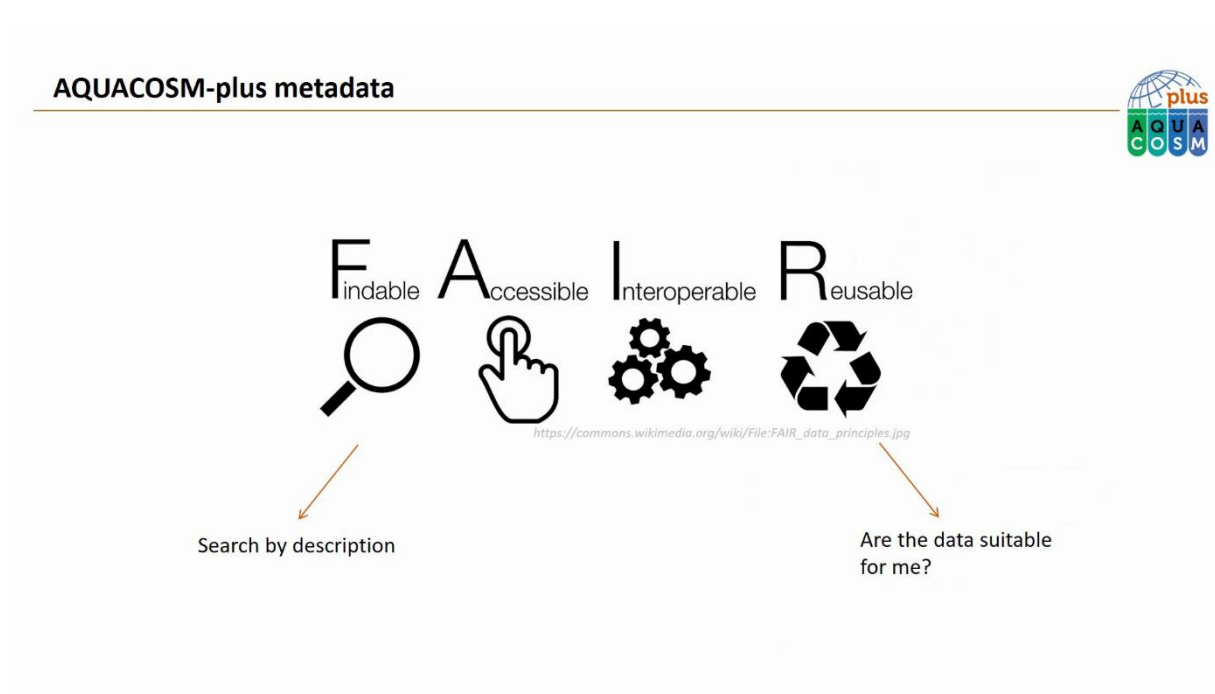
Figure 1: *Screenshot of the first video, giving an introduction to metadata and their importance.* https://youtu.be/-VrlZVd1KSo

Further, we created videos where a user demonstrates how to register on the Mesocosm Metadata Catalogue (**Figure 2**) and create a new metadata entry based on an example TA project (**Figure 3**).
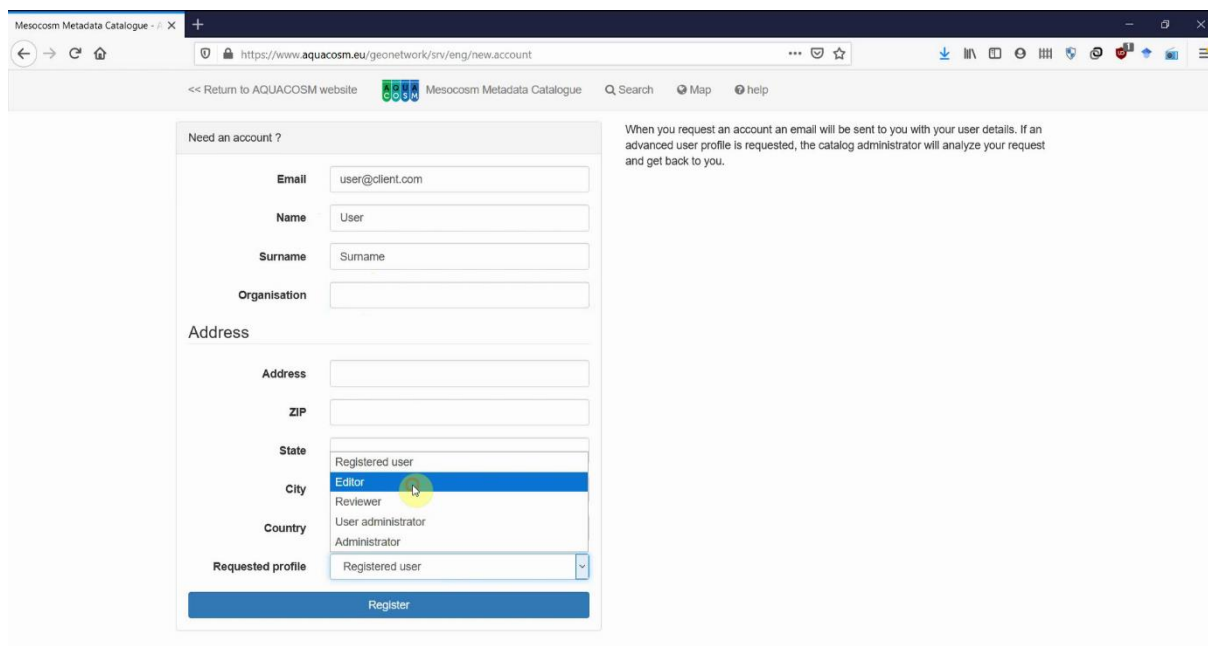
Figure 2: *Screenshot of the second video demonstrating how to create an account for the Mesocosm Metadata Catalogue. https://youtu.be/7qCk3ZSxNBI*
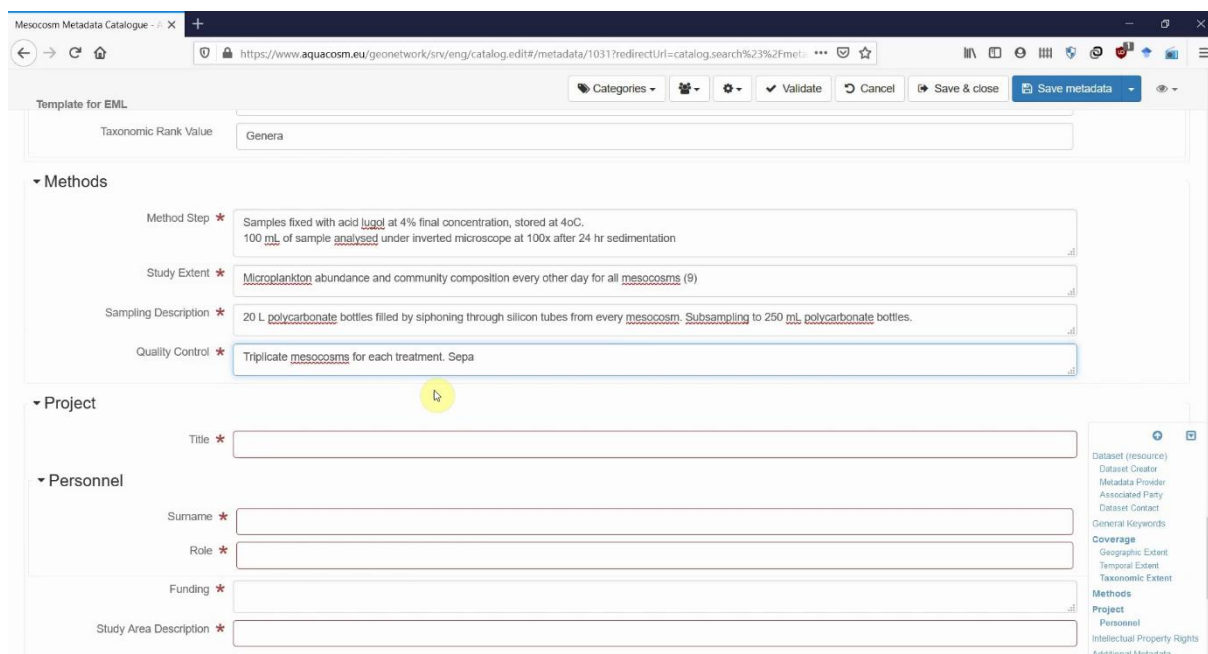
Figure 3: *Screenshot of the third video where a user goes through each field of the EML Template for a new metadata entry. The user explains what information need to be given and provides an example. https://youtu.be/8Qa7Uhjl3zc*

As the AQUACOSM-plus project progresses with the completion of other data and metadata management tasks (4.1, 4.3, 4.4), the videos will be updated following changes in the Mesocosm Metadata Catalogue and will be enriched with information and guidance for future data collection portals.

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

4.1   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 4**)
2. Next, click on "Create an account (**Figure 5**)
3. Fill in your details to create your account and select "Editor" from the "Requested profile" dropdown box. Then click register (**Figure 6**). 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.
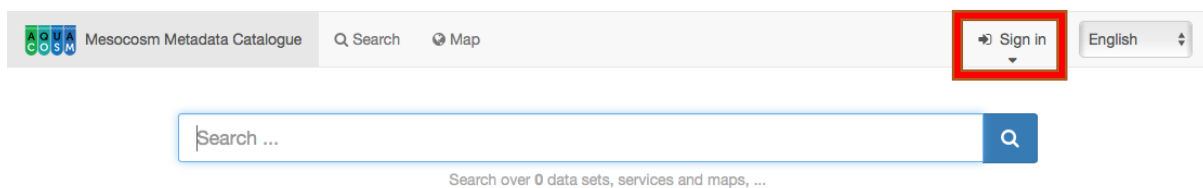

Figure 4: *Go to http://aquacosm.eu/data and click on the "Sign in" icon shown in the red box.*
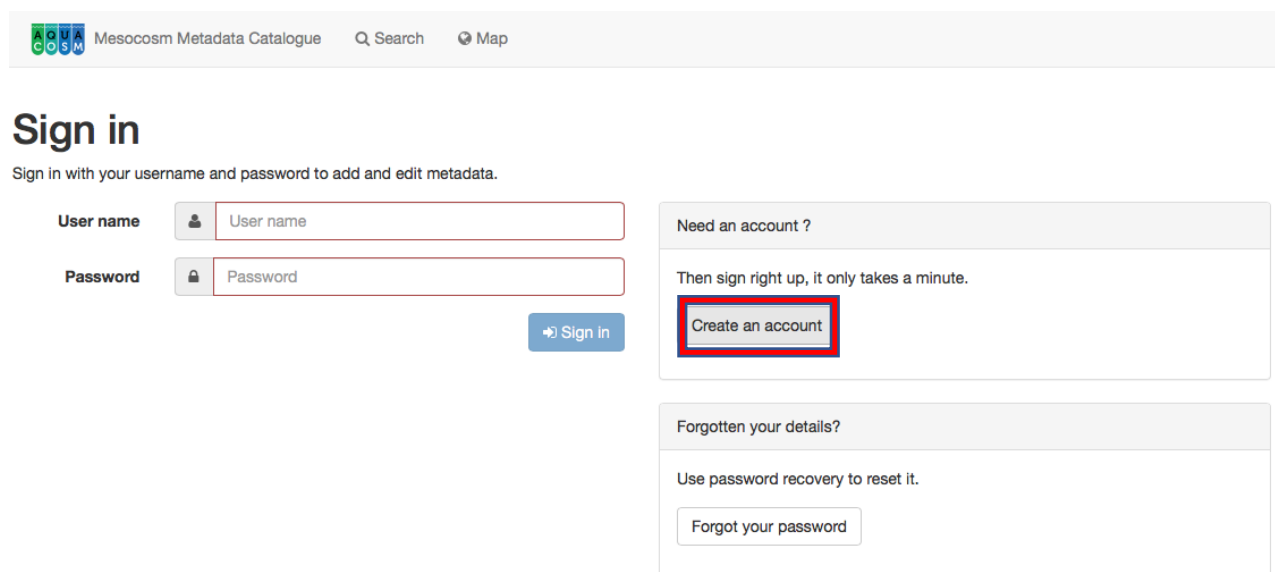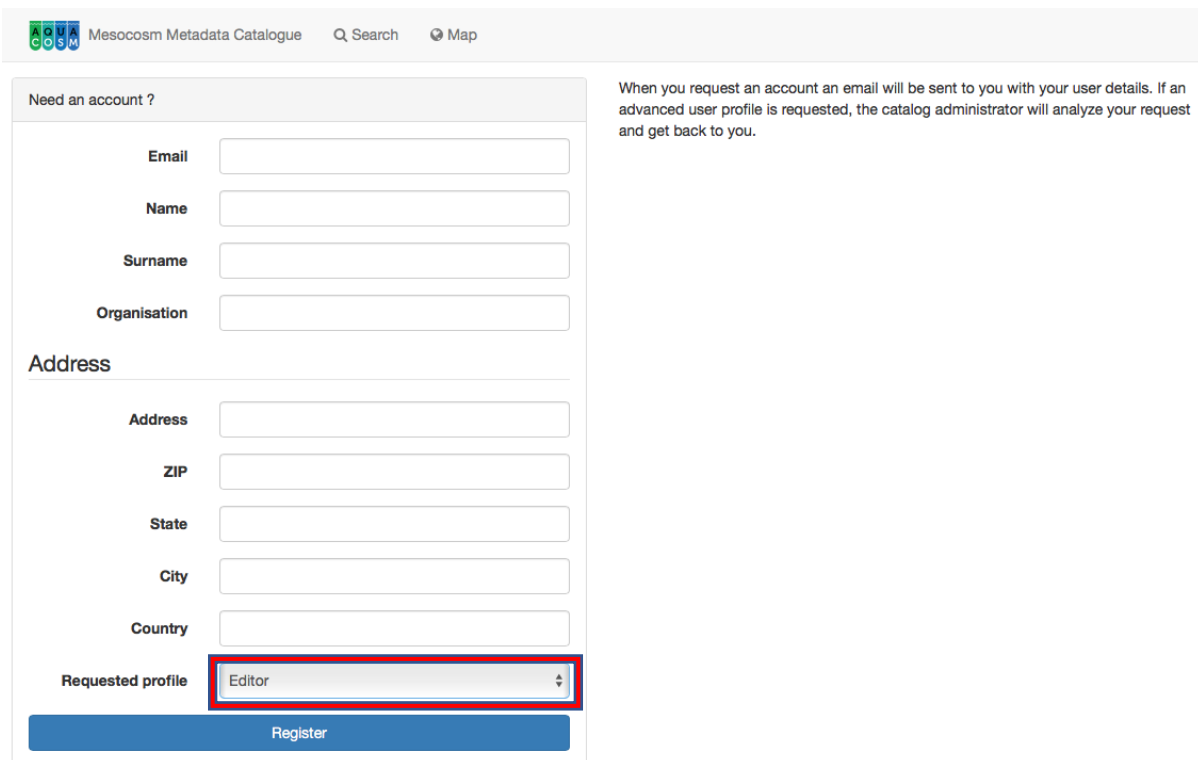
Figure 5: *Select "Create an account" by clicking on the "Create an account" button shown in the red box.*

Figure 6: *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.*

4.2   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 7**).
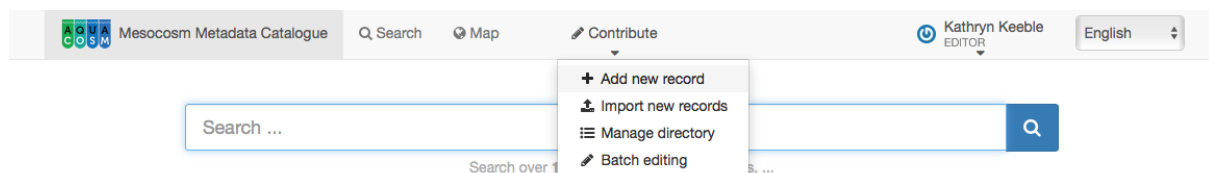
Figure 7: *Shows how to select "Add new record" from the "Contribute" menu.*

Then choose "Create a Dataset" and the "Template for EML" (both will be selected by default) and click "Create" (**Figure 8**).
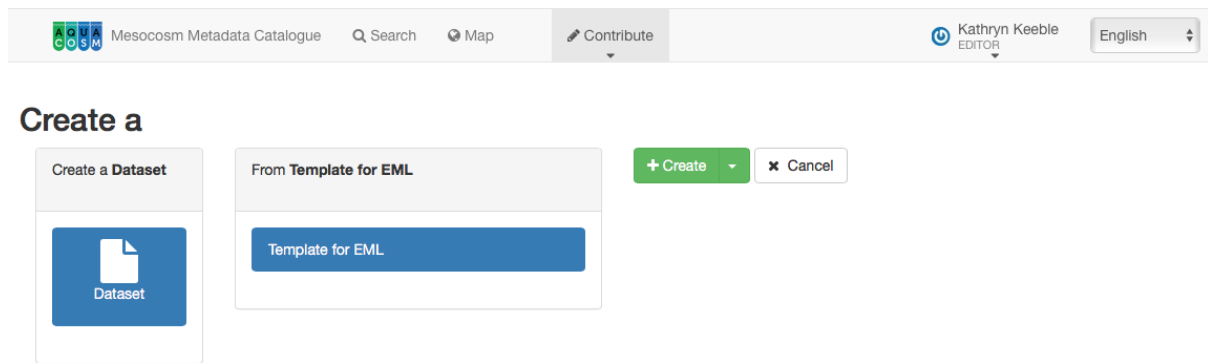
Figure 8: *Shows the next screen with "Create a Dataset" and "Template from EML" selected. Click the green "Create" button to the right of the screen to open the form.*

This will open the template.

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

4.3   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!

4.4   Getting Help

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

4.5   Glossary of terms:

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

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, the person that created the dataset.

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 they are 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

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 website. 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.

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, the person that creates the entry in the metadata catalogue.

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 they are referred to in citations, e.g. Carson

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 website. 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 website or other related online information about the party, e.g., http://www.yourdomain.edu/~doe.

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 they are 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.

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 they are referred to in citations, e.g. Carson.

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 website. 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 website or other related online information about the party, e.g., http://www.yourdomain.edu/~doe.

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 keyword Set, e.g., IRIS keyword thesaurus.

Coverage

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

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".

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

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.

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.

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.

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

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 or AQUACOSM-plus 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.

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. Read more about common license types for datasets at https://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets

Purpose: A description of the purpose of this dataset.

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

5.   Dissemination Activities Related to this Deliverable

Videos and guidance documentation will be shared with TA users prior to each experiment to provide the necessary training on reducing data heterogeneity. The videos and guidance documentation will be available on the TA information page under the "Data management and reporting related to TA?" section. They will be communicated to the AQUACOSM-plus consortium through the newsletter. The videos will be uploaded on the YouTube channel of AQUACOSM-plus and disseminated through social media, accompanied by links to the AQUACOSM-plus webpage and AQUACOSM-plus TA information page.