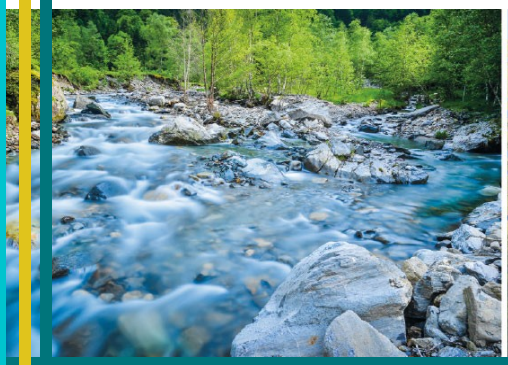
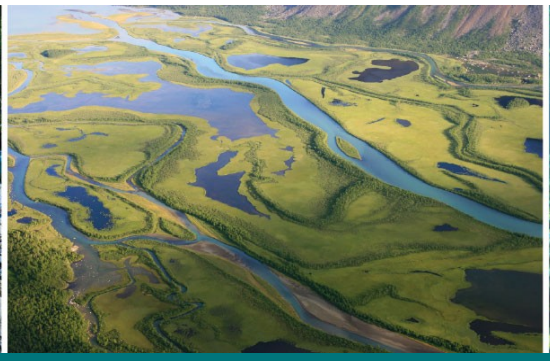
aquacross


D6.1 AQUACROSS data management plan

Deliverable 6.1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 642317.

Authors

Juan Arévalo-Torres, IOC-UNESCO

Ana Barbosa Lopes, IOC-UNESCO

Alejandro Iglesias-Campos, IOC-UNESCO

Julian Barbière, IOC-UNESCO

William Ouellette, IOC-UNESCO

With contributions by: Tim O'Higgins and Declan Dunne UCC, Astrid Schmidt-Kloiber, Florian Pletterbauer BOKU, Aaike De Wever RBINS, Michiel Daam, Antonio Nogueira, Ana Lillebø UAVR

Document reviewed by: Mercedes García-Padilla, Regional Ministry of Environment and Planning of Andalusia (Spain), Environmental Information Network of Andalusia.

Project coordination and editing provided by Ecologic Institute.

Manuscript completed in December 2016

Document title	AQUACROSS data management plan
Work Package	WP6
Document Type	Deliverable (Updated version month 18)
Date	13th December 2016
Document Status	Final Version

Acknowledgments & Disclaimer

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 642317.

Table of Contents

List of Figures

List of abbreviations

API	Application Programming Interface
BISE	Biodiversity Information System for Europe
CKAN	Comprehensive Kerbal Archive Network
CSV	Comma Separated Values
DMP	Data Management Plan
FIP	Freshwater Information Platform
IP	Information Platform
MSFD	Marine Strategy Framework Directive
OGC	Open Geospatial Consortium
WISE	Water Information System for Europe

About AQUACROSS

Knowledge, Assessment, and Management for AQUAtic Biodiversity and Ecosystem Services aCROSS EU policies (AQUACROSS) aims to support EU efforts to protect aquatic biodiversity and ensure the provision of aquatic ecosystem services. Funded by Europe's Horizon 2020 research programme, AQUACROSS seeks to advance knowledge and application of ecosystem-based management for aquatic ecosystems to support the timely achievement of the EU 2020 Biodiversity Strategy targets.

Aquatic ecosystems are rich in biodiversity and home to a diverse array of species and habitats, providing numerous economic and societal benefits to Europe. Many of these valuable ecosystems are at risk of being irreversibly damaged by human activities and pressures, including pollution, contamination, invasive species, overfishing and climate change. These pressures threaten the sustainability of these ecosystems, their provision of ecosystem services and ultimately human well-being.

AQUACROSS responds to pressing societal and economic needs, tackling policy challenges from an integrated perspective and adding value to the use of available knowledge. Through advancing science and knowledge; connecting science, policy and business; and supporting the achievement of EU and international biodiversity targets, AQUACROSS aims to improve ecosystem-based management of aquatic ecosystems across Europe.

The project consortium is made up of sixteen partners from across Europe and led by Ecologic Institute in Berlin, Germany.

Contact	aquacross@ecologic.eu
Coordinator	Dr. Manuel Lago, Ecologic Institute
Duration	1 June 2015 to 30 November 2018
Website	http://aquacross.eu/
Twitter	@AquaBiodiv
LinkedIn	www.linkedin.com/groups/AQUACROSS-8355424/about
ResearchGate	www.researchgate.net/profile/Aquacross_Project2

1 Introduction

This document describes the data management life cycle for all datasets that will be generated by the AQUACROSS project, funded by the European Union's Horizon 2020 Programme for Research (Grant Agreement no. 642317). The documentation of this plan is an outcome to the AQUACROSS general data online survey from Work Package 6 completed by the project partners, as well as other relevant project data providers. The survey was opened from the 15th of September till the 30th of November 2015 with the following objectives (https://ec.europa.eu/eusurvey/runner/aquacrossdatasurvey):

- Make an initial inventory of existing datasets and possible dataset acquisitions;
- Gather information on current data management and metadata practices, the uses and access policies of AQUACROSS partners;
- Data publishing practices, i.e. how partners plan to publish and exchange their datasets across the different WPs and pilot case studies.

The information on current data management practices resulting from this online survey provided useful information for the development of the Data Management Plan (DMP) (Results of the survey available in the Annex section 7 of this document).

The DMP presented herein has been prepared by taking into account the template of the Version 1.0 of the "Guidelines on Data Management in Horizon 2020" (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf) and has the following structure:

1 *General information:* description of the main project datasets, which are organized according to the main topics of the AQUACROSS project:
    - Dataset 1on drivers and pressures affecting aquatic ecosystems;
    - Dataset 2 on biodiversity, ecosystem functions and services;
    - Dataset 3 on projected changes in drivers and pressures, and model scenarios outputs for the case studies;
    - Dataset 4 on ecosystem-based management towards policy objectives;
2 *Standards and metadata:* reference to existing suitable standards of metadata that will be used by the project.
3 *Data Publishing*: description on how data will be published, including access procedures, tools for dissemination as well as software and other tools necessary to enable re-use, and definition of access rights for specific groups. This section

also identifies the location and type of repository (institutional, standard repository for the discipline, etc.) where data will be stored, if already existing and identified.

4 *Data preservation and archiving:* Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its estimated final volume, what the associated costs are and how these are planned to be covered.

The AQUACROSS Data Management Plan is an evolving document that will be edited and updated throughout the project. This deliverable was officially submitted in month 9. This version (Month 18) is an updated version of the official deliverable submitted including the following updates:

- Clarification on the submission of the Data Management Plan updates
- Clarification on data repositories that will be used in the project
- Clarification on the http://data.freshwaterbiodiversity.eu/ as a component of the Freshwater Information Platform (FIP)
- Clarification the sustainability of FIP
- Clarification on how data is organized and the area of coverage
- Clarification on how the data products will be linked to any relevant publications

This deliverable will be updated and re-submitted if required as part of the AQUACROSS periodic reports.

2 General Information

The objective of this section is to provide a detailed description of the different topics of information according to which all the datasets managed by the AQUACROSS project will be categorized. These categories have been defined based on the different topics that will be addressed by the project.

AQUACROSS will establish an Information Platform (IP)that will gather all types of information generated by the project on aquatic ecosystems and biodiversity, including: spatial data in raster and vector formats, multimedia files (videos, images, etc.), tables, graphs, web services and any other relevant information sources available in a digital format. Open data formats will be preferred in an easy re-usable format; e.g.

- geospatial data: shapefiles for vector files orgeo-referenced TIFF for raster files;
- imaging: JPEG2000, PNG, GIF;
- digital videos: MPEG-4 High Profile;
- documentation and scripts: Open Document Text (.odt), Rich Text Format (.rtf) Text file (.txt);
- tabular data: comma-separated values (CSV) file (.csv);

However, other popular data format such as Microsoft Office Formats, PDF will be also accepted.

In addition to this, a version control system will be put in place as part of the Information Platform in order to enable AQUACROSS partners to keep track of their contributions and submission dates.

The AQUACROSS Information Platform will be based on CKAN software, an open source framework for developing open data portals, which provides a reliable entry point for all data and metadata collected, processed or generated by the project. CKAN is based on the following technologies:

- **Apache** as a web server;
- **Postgres** to store metadata and data files into a relational database;
- **Python** scripts for the server-side;

The application's front-end uses a collection of JavaScript libraries such as recline.js and leaflet.js. Other libraries can also be integrated, e.g. for visualisation purposes.

CKAN is a modular software and therefore functionalities can be extended or further developed by using different extensions (Figure 1).

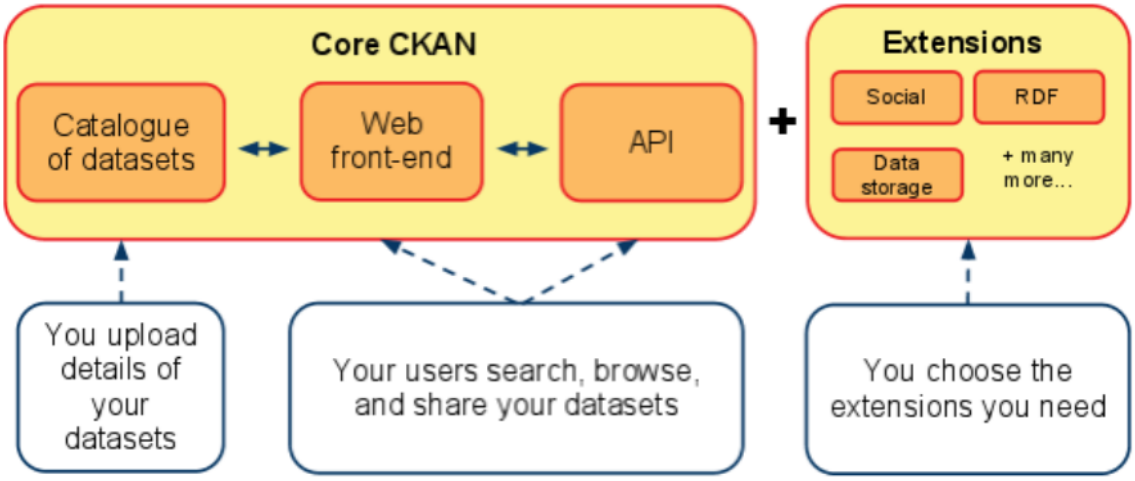


Figure 1: Overview of CKAN IT architecture highlighting the main functionalities provided.

Source: Open Knowledge Foundation https://goo.gl/L1nYzo

CKAN is currently the technical solution implemented by the European Commission to publish pan-European open datasets across the European Union (http://open-data.europa.eu/en/data/). Moreover, data publishing and interoperability of AQUACROSS outputs into various established EU platforms such as BISE, WISE, FIP and other relevant visualization tools will be ensured.

2.1 AQUACROSS thematic categories

The AQUACROSS datasets will be divided into thematic categories following the main topics of the project:

- Drivers of Change and Pressures on Aquatic Ecosystems
- Biodiversity, Ecosystem Functions and Services
- Forecasting Biodiversity and Ecosystem Service Provision
- Ecosystem-based Management Towards Policy Objectives

The AQUACROSS partners should report the data according to the categories, whether the data is generated in the context of the work package or a case study. For instance, data produced under the work package 5 on Causalities between Biodiversity, Ecosystem Functions and Services will go under the category 2 on “Data on biodiversity, ecosystem functions and services”; and data produced under the

case study 1, which refers to the Causalities between Biodiversity, Ecosystem Functions and Services also goes under the Category 2.

For the first version of the document, the DMP focusses on describing the dataset categories. As data become available we plan to update these descriptions as required and include specific examples. In any case, information on individual datasets will be documented through the metadata editor of the Information Platform (IP), and will include details such as dataset name, reference, description, origin and data type (raw, processed, quantitative, qualitative).

The following sections describe the typical datasets under each dataset category.

2.1.1 Category 1: Data on drivers and pressures affecting aquatic ecosystems

Dataset

This category will include the main outcomes from WP4: all drivers or decisions taken by social and economic agents individually, as well as the pressures on how the socio-economic system affects and directly transforms the environment. The resulting data and information produced herein will be for example: population density, built-up density, sectorial Gross Value Added (GVA) (fisheries, agriculture, industries and services),in the land use (i.e. land take[1], deforestation/afforestation) maritime and inland transport performance of goods, invasive alien species, number of tourists (daily visitors), production and consumption of ozone-depleting substances, water consumption (public, livestock and irrigation, industrial), loss of accessibility for migratory fish due to dams in river basins, among others.

Economic and demographic drivers may have effects on the aquatic ecosystems and biodiversity through the pressures that they might exert. Driver-pressure-related data reflect for example: changes, water stress index, pollutant concentrations inland, in freshwaters and coastal and marine waters, decline of habitat quality, overexploitation and overfishing, pH, eutrophication, tourism intensity, urban development along shore and natural reserves, among others.

This data will typically be gathered from publicly available sources such as the European Environment Agency (EEA). Project partners will be requested to post links to these data source on the IP. The majority of these data is expected to be available

[1] http://glossary.eea.europa.eu/terminology/concept_html?term=land%20take

through both OGC web services and available for download in raster of shapefile format.

An example dataset for this category is:

Name: Corine Land Cover 2000 – 2006 changes

Reference/link:http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-3

Description/metadata:

Temporal coverage:2000-2006

Last upload:08 Mar 2014

Geographic coverage: Albania, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Europe, Finland, France, Germany, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia (FYR), Malta, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom

Tags: corine land cover changes | raster data | land cover | vector data | eea owned data sets | geospatial data | corine land cover

Rights:EEA standard re-use policy: unless otherwise indicated, re-use of content on the EEA website for commercial or non-commercial purposes is permitted free of charge, provided that the source is acknowledged (http://www.eea.europa.eu/legal/copyright). Copyright holder: European Environment Agency (EEA).

Coordinate reference system: EPSG:3035

Data sources: Corine Land Cover 2000, Corine Land Cover 2006: December 2013

Owners: European Environment Agency (EEA)

Processors: European Topic Centre for Spatial information and Analysis (ETC/SIA)

Permalink to this version: f497a90b18dc496b823e3b71137eff7a

Permalink to latest version: Q4T9TYUK84

In addition to the metadata available at the source, we plan to document for which analyses the data was used, and provide links to the resulting products and linked to relevant publications on the IP.

2.1.2 Category 2: Data on biodiversity, ecosystem functions and services

This category will include the main outcomes from WP5: The assessment of status and conditions of ecosystems is based on the reported data under the Water Framework Directive (WFD), Habitats Directive (HD) and the Marine Strategy Framework Directive (MSFD). These datasets include the ecological status (WFD), environmental status (MSFD), species and habitat conservation status and red list index (HD art. 17).

Under the umbrella of Category 2, AQUACROSS will deliver a collection of maps to assess the status of habitats and other important areas for biodiversity within the case studies. Examples of this category include indicators such as: abundance and distribution of selected species, species of European interest, species diversity, environmental status (MSFD), Red List Index for selected European species. In addition to this, the dataset will also include indicators for the current assessment of ecosystem services, divided into:

- Freshwater: e.g. freshwater aquaculture production, water abstracted, area occupied by riparian forests, chemical status and ecological status, morphological status, National Parks and Natura 2000 sites, proximity to urban areas of scenic rivers or lakes.

-Coastal and Marine: e.g. nutrient load to coast, Heavy Metal and Persistent Organic Pollutants deposition, Oxygen depletion Risk, composite indices based on extent of selected emerged, submerged and intertidal habitats, coastline slope and coastal geomorphology, wave regime, tidal range, relative sea level, storm surge, species distribution, C stock , C sequestration ,,presence of iconic/endangered species .

An example dataset for this category is:

Name: Joint Danube Survey 1-3 data (mentioned as one of the dataset in Annex 7)

Reference/link: No link available at this stage, metadata link at the Freshwater Information Platform:

http://data.freshwaterbiodiversity.eu/metadb/bf_mdb_view.php?entryID=BF21

Description/metadata:

JDS was an initiative by the ICPDR to investigate the whole Danube river both with biological and chemical parameters. Goal: chemical status / biological status assessment; www.icpdr.org

2001:

94 sites along the Danube river incl. some larger tributaries at confluence

benthic invertebrate data, qualitative sampling

2007:

96 sites

benthic invertebrate data, airlift sampling

For this dataset, metadata and data links will be included in the AQUACROSS IP. In parallel, occurrence data will be made available on-line in Darwin Core-Archive format through the Freshwater Information Platform (http://www.freshwaterplatform.eu). Links to the occurrence data will also be included on the AQUACROSS IP. As mentioned for the example for the dataset Category 1, the use of this data in analyses will be recorded.

2.1.3 Category 3: Related data on projected changes in drivers and pressures, and outputs model scenarios for the case studies

This category will include the main outcomes from WP7: The assessment of impacts refers to changes in drivers and the implementation of alternative policies, such as setting biodiversity strategy targets. This includes data on projected drivers and pressures, as well as data on their impact on ecosystem services, functions and biodiversity.

Indicators under category 3 will include, for instance, projected drivers and pressures, important areas of biodiversity and ecosystem services delivery areas for selected case studies according to different scenarios. In other words, the same list of indicators as described in Dataset 1 and 2 will be reproduced accordingly to the scenarios defined by the project.

The datasets we envisage under this category are largely ones that will be generated in the course of the AQUACROSS project. As methodology for these analyses is being discussed, we are not able to include examples at this stage. Nevertheless, we mostly expect the outcomes to be available as map products. As outlined in the

AQUACROSS GIS guidelines, we are primarily expecting map layers in raster or shapefile format along with standardised map images for reporting purposes. When making these datasets available, project partners will include documentation on the dataset used, and provide links to data in the IP or on external websites where available.

2.1.4 Category 4: Ecosystem-based management towards policy objectives

This dataset will include the main outcomes from WP8: All type of information and indicators that will help monitoring and informing policy makers on the effects of the responses or actions taken by society, individually or collectively. More specifically, these indicators will provide information on designated areas for policy intervention such as multi-zoning planning according to the spatial prioritization of the different scenarios. In addition to this, distance-to-targets indicators could also be used to quantify the efforts required to reach policy targets and compare the differences of the scenarios in terms of important areas of biodiversity and the provision of ecosystem services.

3 Standards and metadata

Metadata is "data about data". Metadata helps a user to find or discover the data that they need and, thereafter, evaluate whether this resource satisfies the user's requirements. It contains information that documents the basic characteristics of a specific dataset. Metadata will be key to organise all the digital information generated and processed by AQUACROSS, providing a digital identification and helping us to support archiving and preservation of the information resources. Three different kind of metadata are envisaged in the AQUACROSS project:

1 Discovery metadata for each dataset including information related to the ownership, use limitations, creation of data, distribution, scale, extent, contributors and reference publications. It chiefly concerns discovery metadata with some additional exploration/evaluation metadata elements. For spatial datasets the AQUACROSS discovery metadata profile will adopt the INSPIRE Directive metadata model using the ISO 19115/19139 standards. For non-spatial datasets, the Dublin Core standard will be adopted for discovery.

2 Structural or usage metadata describes each attribute of the data (column), for example value, numbers, units used in a dataset. It concerns specific metadata about the actual data format itself. Its overall model is based on the data specification level. In essence, it contains information about how to use the data. This structural information can be linked to the discovery metadata profile as a URL reference.

3 Indicator factsheets describe the purpose, policy context and methods used for producing the indicator. These factsheets provide essential project context for the data and will be useful in improving the impact and dissemination of information generated in the project frame. The Indicator factsheets can be linked to the discovery metadata profile as a URL reference, providing supplemental information.

The information collected during the project will be appropriately documented following relevant metadata standards and will be INSPIRE/Open Geospatial Consortium (OGC) compliant. The purpose of the discovery metadata records stored in a database will be to facilitate the discovery of information on aquatic ecosystems and biodiversity.

The standards used to fulfil metadata discovery will be based on the ISO 19139 which provides the XML implementation scheme for ISO 19115 specifying the metadata record format. These standards will be used to describe, validate, and exchange geospatial and non-spatial metadata prepared in XML.

In addition to this, other relevant standards such as the Catalogue Service for the Web (CSW) from the OGC will be implemented in order to facilitate metadata information exchange with third party platforms.

The discovery metadata files will be generated using a graphical user interface as part of the Information Platform in which authorized users will be able to enter and edit the metadata. The metadata files will be stored in a relational database following INSPIRE metadata standards for spatial datasets and Dublin Core for non-spatial datasets. The final catalogue will be open to all the AQUACROSS project partners and the general public. Before releasing the Information Platform it is expected that AQUACROSS project partners generate and hold metadata files in a transcript form and accessible file formats such as .xls and .csv (Excel). A template will be provided as part of the GIS guidelines for that purpose until the Information Platform is ready to be used.

The AQUACROSS information platform will take advantage of the latest developments and interoperability standards for harvesting and collecting relevant data and metadata files from other European reference data portals. The aim of this functionality will be to create an access point to the different existing data portals without replicating the information by developing a decentralized information system.

4 Data publishing – data access and intellectual property

The AQUACROSS Information Platform (IP) will be available online for users to access all digital data and information of the project via a CKAN data catalogue, without having to install any specific software or plug-in, in their favourite internet browsers. The IP will be compatible with most standard internet browsers.

After registration, project partners and associated organisations of the AQUACROSS consortium will be able to upload and publish the metadata and data. The first beta release of the IP is expected in June 2016. In order to protect privacy, security, confidentiality and intellectual property rights, a mechanism will be put in place allowing users to specify the recommended data license before publishing data or metadata into the IP (see below, Figure 3). All outputs for which AQUACROSS and/or its partners holds all intellectual property rights, will be free for general use and will be promoted within the IP to increase the impact of AQUACROSS (Figure 2).

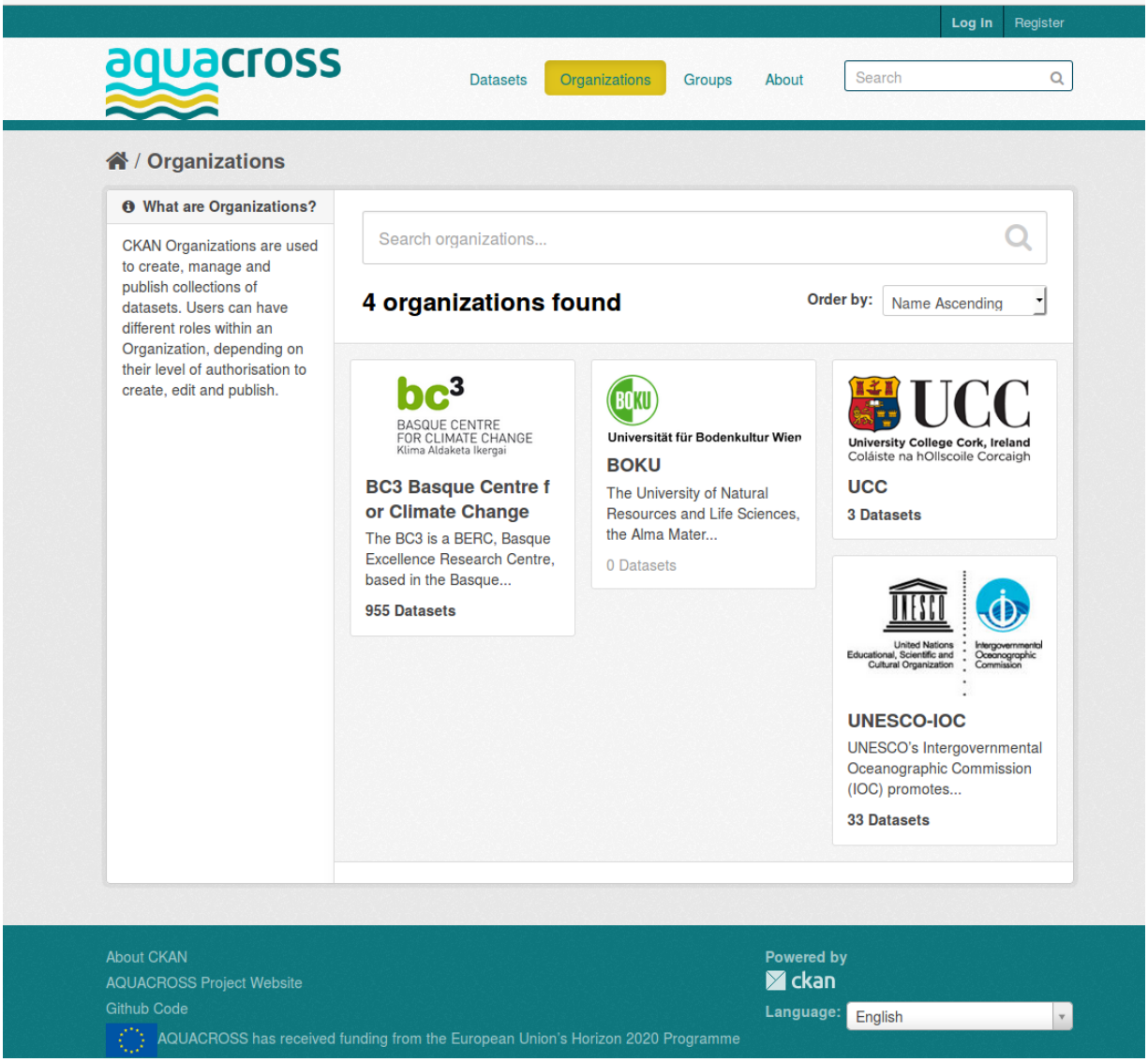


Figure 2: Organizations registered within the Information Platform as initial configuration test in a local environment.

According to Article 29.3 of the Model Grant Agreement (article applied to all projects participating in the Pilot on Open Research Data in Horizon 2020), "*the digital research data generated in the action ('data'), the beneficiaries must: (a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user*" (Guidelines on Data Management in Horizon 2020).

Following these requirements, all the information produced within the AQUACROSS project will be released under an Open Data License (Creative Commons CC Zero License or Creative Common Attribution License-CC-BY v4.0), unless there are restrictions imposed by third parties. Furthermore, the data acquired by AQUACROSS

partners from third parties should, if not published elsewhere, be considered for publication under an Open License. However, depending on their origin and ownership, data may have different requirements in terms of access and re-use as outlined below.

In the case that AQUACROSS partners have full intellectual ownership of the data and no reasons for non-disclosure of data have been specified, the data will be released under an Open Data License (Creative Commons CC Zero License or Creative Common Attribution License-CC-BY v4.0). The default rule is that data produced in AQUACROSS shall be available without restrictions on access and re-use. This means that the data shall be made available in accordance with the standard EC Reuse and Copyright notice (see http://ec.europa.eu/geninfo/legal_notices_en.htm).

In the case that AQUACROSS partners are fully owner of the intellectual property rights of the data, but reasons for non-disclosure of data have been specified (e.g. privacy and/or integrity of individuals, commercial interests of a natural or legal person, including intellectual property, court proceedings and legal advice, the purpose of inspections, investigations and audits), the reasons specified will be reviewed by the Data Management Team. All ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country, and as described in the Ethics Requirements section in the Project Description of Action[2].

In the case of the data holders who do not have full intellectual ownership of the data, they are requested to verify who can access the data and which restrictions apply. Following the recommendations of the Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020 (EC, 2013)[3], the Figure 3 shows the scheme on data origin and licensing scheme to be used.

In case the data produced involves third party data or the AQUACROSS project partner / data holder does not hold full ownership of the data, they will be requested to verify the use conditions (licensing) of the data. If the third party data has been previously published under a specific license, the data holders should verify the license used (see below how to proceed in this situation). If the third party data has not yet been published, the AQUACROSS partner/ data holder will be requested to make reasonable effort to obtain permissions from the third party to make the data

[2]http://ec.europa.eu/research/participants/portal/doc/call/h2020/h2020-msca-itn-2015/1620147-h2020_-_guidance_ethics_self_assess_en.pdf

[3]https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

available. The information regarding the dataset contributors will be recorded in the metadata file.

If anybody can access the data obtained from the third party, the following open licenses are recommended:

- CC0 1.0 Universal (Creative Commons Public Domain Dedication License) https://creativecommons.org/publicdomain/zero/1.0/
- CC-BY v4.0 (Creative Common Attribution License) https://creativecommons.org/licenses/by/4.0/

If permission from the third party has only been granted under specific conditions and data use is permitted for authorised users only, AQUACROSS partners should use one of the Non-open licenses for data not fully owned by the AQUACROSS project:

License Name	You are free to:	Under the following terms:	Additional restrictions
CC-BY-SA v4.0 (Creative Common Attribution Share-Alike Licence) http://creativecommons.org/licenses/by-sa/4.0/	Share copy and redistribute the material in any medium or format Adapt remix, transform, and build upon the material for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the license terms.	Attribution You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. Share Alike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.	No additional restrictions You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.
CC-BY-NC v4.0 (Creative Common Attribution Non-Commercial Licence)	Share copy and redistribute the material in any medium or format Adapt remix, transform, and build upon the	Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the	No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the

https://creativecommons.org/licenses/by-nc/3.0/	material The licensor cannot revoke these freedoms as long as you follow the license terms.	licensor endorses you or your use. Non Commercial — You may not use the material for commercial purposes.	license permits.
CC-BY-NC-SA v4.0 (Creative Common Attribution Non-Commercial Share-Alike Licence) https://creativecommons.org/licenses/by-nc-sa/2.5/	Share — copy and redistribute the material in any medium or format Adapt — remix, transform, and build upon the material The licensor cannot revoke these freedoms as long as you follow the license terms.	Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. Non Commercial — You may not use the material for commercial purposes. Share Alike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.	No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

The end-user community will be able to query and interrogate the data and metadata of all public datasets. Datasets will be published following all the relevant standards for publishing information on the web and ensuring maximum interoperability with other existing environmental information platforms such as FIP, WISE and BISE. The information will be available using open data file formats and machine-readable formats such as Comma Separated Values (CSV), JSON, TIFF, Shapefiles and XML. Moreover, the CKAN API (Application Programming Interface) will be used to expose the data and functionalities developed as part of the Information Platform, allowing end-users to connect and interact with the AQUACROSS datasets.

Following the license scheme described above AQUACROSS partners will be able to decide whether the access will be widely open (this should be the case of all foreground project outputs), available after user registration, or restricted to specific groups:

- Free access to the Information Platform: any user without registration can download and view the datasets;
- Authenticated users: only users who have previously registered to the Information Platform can have access to the datasets; authenticated users subscribed to a specific group, will have access to certain datasets, such as, for instance, unpublished datasets of the project case studies.

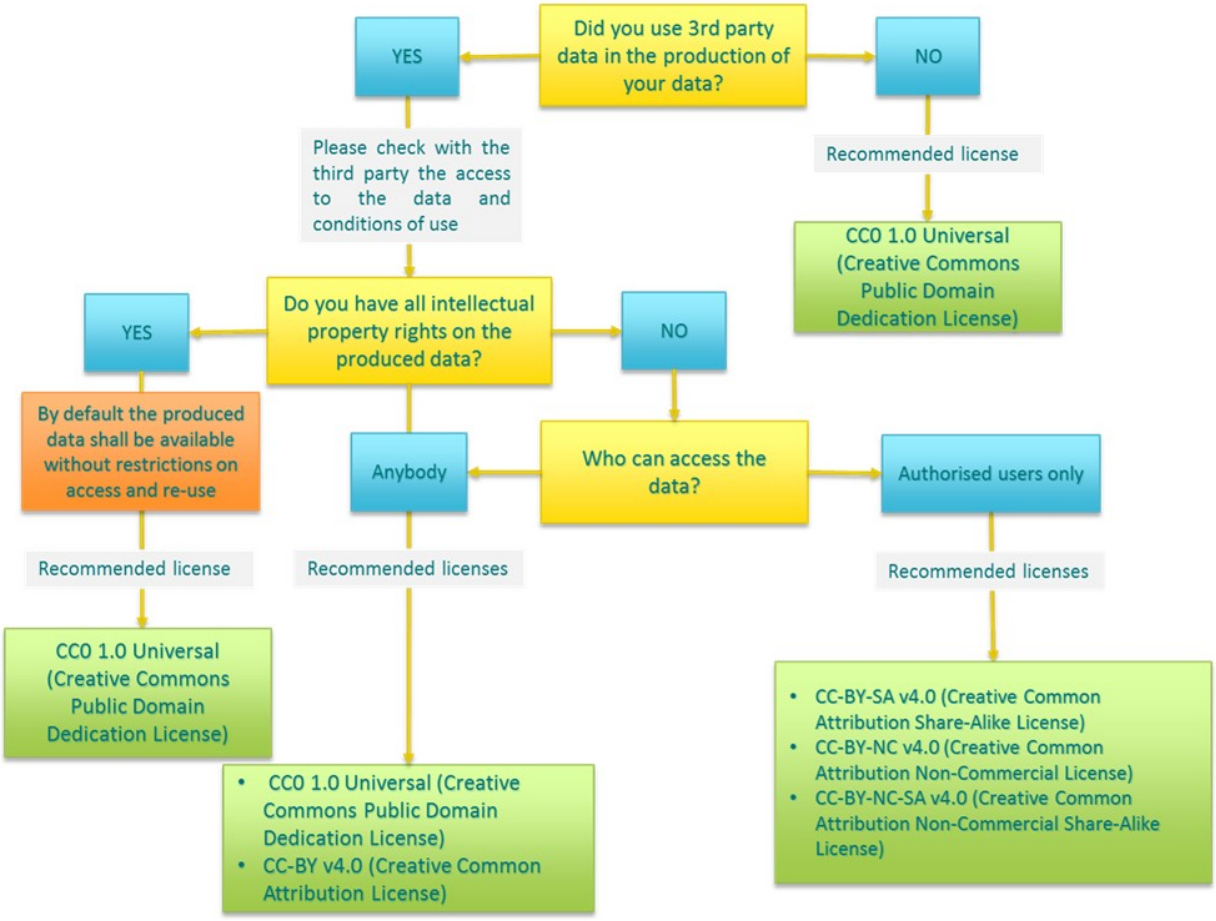


Figure 3: Data origin and respective licensing scheme (based on the Guidelines on Open Access (EC, 2013)).

5 Data preservation and archiving

All data and information collected in the AQUACROSS project database will be accessible through the IP together with the project website **for three years after the end of the project**, as stated in the Project Description of Action. During this period, unless otherwise decided by the AQUACROSS consortium members, the IP functionality will remain the same as during the project duration. Preservation of the datasets will be guaranteed by archival on a dedicated archive server. Backups will be performed weekly or daily depending on the amount of data transferred and the use level of the Information Platform. All backups will be stored on a dedicated server and, if possible, also on a local machine. A backup strategy will be implemented based on the following:

- Automatic: An automatic backup will be scheduled every night on both the production and development servers. This will include a backup of uploaded AQUACROSS datasets and metadata, the AQUACROSS Data Management Portal and any underlying databases and logs. Depending on the backup size, this can be a daily full backup or else a daily incremental backup coupled with a weekly full backup. Email notification of any backup issues will be implemented.
- Location: The backup will be stored offsite and not on the same server as the portal. Ideally at least 2 offsite copies should be stored.
- Test: The backup will be tested to ensure the backup routine will restore the data and portal.
- Code: In addition, the portal source code will be maintained in GitHub.

Following this 3-year period, the consortium members will guarantee the continuity of the Information Platform (IP) via a Memorandum of Understanding attaching the information gathered by the project into other existing information platforms with long-term support. Freshwater related data and results, for example, will be incorporated into different components of the Freshwater Information Platform (FIP; www.freshwaterplatform.eu), partly already during the project's lifespan. This will involve storing metadata into the Freshwater Metadatabase and releasing the information as publication in the Freshwater Metadata Journal, including spatial results into the Global Freshwater Biodiversity Atlas

(http://atlas.freshwaterbiodiversity.eu), storing occurrence data into the Freshwater Biodiversity Data Portal (http://data.freshwaterbiodiversity.eu/) and its data repository, as well as research and educational means into the Freshwater Resources section (http://www.freshwaterplatform.eu/index.php/resources.html). The institutions involved in the FIP have established an agreement to ensure the sustainability of the initiative and long-term availability of the published data, but are seeking for longer-term, sustainable funding to ensure further development and sustained support for data and information contributors.

Other options for long-term data archiving and preservation of data generated or mobilised during the project will be carefully considered throughout the project based on the best research data management practices coming from the OpenAIRE (https://www.openaire.eu) and Zenodo data repository (http://www.zenodo.org/) projects. For data from other realms (in each of the thematic categories) we will consider other suitable community repositories (as listed on re3data.org) or turn to more generic archiving repositories including Zenodo and Dryad (http://datadryad.org). OpenAIRE2020 assists in monitoring H2020 research outputs and is a key infrastructure for reporting H2020's scientific publications, as it is loosely coupled to the EC's IT backend systems. The EC's Research Data Pilot (https://www.openaire.eu/opendatapilot) is supported through European-wide outreach for best research data management practices and Zenodo, which will provide long-term data storage, enabling researchers to share and preserve any research outputs in any size, any format and from any discipline.

The data uploaded in Zenodo will be also stored in CERN Data Centre. Both data files and metadata are kept in multiple online replicas and are backed up to tape every night. In the highly unlikely event that Zenodo will have to close operations, there is a guarantee that they will migrate all content to other suitable repositories. Since all uploads will be registered with Digital Objective Identifiers (DOIs), all citations and links to Zenodo resources (such as AQUACROSS data) the availability of AQUACROSS data would not be affected by the closure of the repository.

6 References

European Commission (2015) Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Online version: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

European Commission (2015) Guidelines on data management in Horizon 2020.Online version: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

AQUACROSS General Data Survey: https://ec.europa.eu/eusurvey/runner/aquacrossdatasurvey

AQUACROSS GIS Guidelines: (Link to be included once the final version is ready and uploaded to the intranet)

7 Annex

AQUACROSS online survey results

Data providers

Case Study 1 North Sea	Case Study 2 IBR Mediterranean	Case Study 3 Danube River	Case Study 4 Lough Erne	Case Study 5 Vouga River	Case Study 6 Ronne	Case Study 7 Swiss Plateau	Case Study 8 Azores
BC3	ACTEON	BOKU	ACTEON	BC3	SRC-SU	ACTEON	ACTEON
IMARES	BC3		UCC	UAVR			
RBINS	IOC- UNESCO						

List of datasets (preliminary)

Description of the existing datasets

National Park Donau-Auen (Austria)-dataset comprises: fish, amphibians, macro invertebrates, water birds, hydromorphological classifications, human alterations classification, hydromorphology, chemical data, digital maps.

FLOX- Floodplain Index Database- dataset contains: fish, amphibians, Mollusca, Odonata, Trichoptera from the Danube floodplain.

Joint Danube Survey 1-3 data- dataset (directly located at BOKU) contains: macro invertebrates; additional data on all measurements that were collected during the Joint Danube Surveys (fish, macrophytes, phytoplankton, microbiology, hydromorphology, chemistry, toxic substances etc.) can be requested from the ICPDR.

EFI+ database- dataset contains: fish data (relevant Danube data can be extracted), stressor/pressure data, land use, temperature, etc.

FLOODmi – macro invertebrate data from Austrian floodplain areas.

X-GIG – inter-calibration of large rivers dataset.

Database of Trawl Surveys – All international fisheries surveys data executed under auspices of ICES is available via the ICES website.

Aries – DEMs, Watersheds, drainage networks, population density data, roads, etc.

SIOSE – Land use/cover map of Andalusia.

Water information system Andalusia – The Water Information Dataset considers all those involved in water management: physical and biotic environment, human environment and infrastructure.

Coastal and Marine information system Andalusia – This dataset contains information on coastal physiography, characterization of the coastline, bathymetry, land, sea level, waves, winds, currents, fields, water quality, fisheries, artificial reefs, etc.

Landscape Information System Andalusia– The Integrated Information System Landscape covers land use, topography, vegetation, climate, habitats fauna, cultural resources, historic architecture, colour, texture, structure of ownership, etc.).

BPNS Benthos– Habitat Suitability Map for the Belgian Part of the North Sea.

Description of the datasets planned acquisitions

Meteorological data obtained by a meteorological station from UAVR team.

Danube Park data – biological, chemical, hydromorphological data from all Danube Parks along the Danube (current status: discussions about what is available and could be requested).

Trans-National Monitoring Network data – all data of the TNMN database can be requested from the ICPDR.

Metadata available and published

Are the metadata available and published for all the existing datasets?

Options		Answers	Ratio
Yes		3	33.33%
No		6	66.67%
No Answer		0	0%

Metadata standards used

Options		Answers	Ratio
INSPIRE (ISO 19115 profile)		0	0%
ISO 19115 (Geographical metadata)		2	22.22%
SDMX (Statistical Data and Metadata Exchange)		0	0%
No standard (i.e., a custom metadata schema is used)		5	55.56%
Other		1	11.11%
No Answer		2	22.22%

Please add the links to the webpages describing the metadata.

http://geo.ices.dk/geonetwork/srv/en/main.home?search=ices

http://www.juntadeandalucia.es/medioambiente/site/rediam/menuitem.aedc2250f6db83cf8ca78ca731525ea0/?vgnextoid=a3aba721

Use limitations

Are there any reasons why the data should not be made openly available?

Options		Answers	Ratio
Yes		3	33.33%
No		6	66.67%
No Answer		0	0%

If yes, then describe which datasets are restricted to the AQUACROSS partners.

Need to verify this as some datasets are public and others were provided under a bilateral protocol.

For all datasets individual IPRs apply and usage needs to be discussed separately.

Who should be able to access the data?

Options		Answers	Ratio
Everybody: the datasets can be downloaded anonymously by anybody (under an open access license).		3	33.33%
Authorised users: datasets can be accessed only by accredited users (e.g. restricted to the partners of AQUACROSS, to a specific group under request)		2	22.22%
No Answer		4	44.44%

Data Policies

Will any permission restrictions need to be placed on the data?

Options		Answers	Ratio
Non-commercial: data are available only for non-commercial use		1	11.11%
Non derivatives: derived datasets (resulting from modifying the original dataset) cannot be distributed		0	0%
Share-alike: derived datasets must be shared according to the same terms and conditions of the source datasets		0	0%
Specific agreement		1	11.11%
No Answer		7	77.78%

Data sharing

How will you make the data available?

Options		Answers	Ratio
Website: the data is/will be available on our website		2	22.22%
FTP: The data will be uploaded on the AQUACROSS FTP server.		2	22.22%
FTP: The data files are available on our institute's FTP server and can be accessed by AQUACROSS.		0	0%
API: The data can be accessed by an Application Programming Interface (API).		1	11.11%
EMAIL: We will send the raw data files via email.		3	33.33%
HTTP: The data files are available on our HTTP server, and can be collected by HTTP requests.		2	22.22%
Other (please, specify which ones):		2	22.22%
No Answer		0	0%

AQUACROSS PARTNERS

Ecologic Institute (ECOLOGIC) | Germany
Leibniz Institute of Freshwater Ecology and Inland Fisheries (FVB-IGB) | Germany
Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization (IOC-UNESCO) | France
Stichting Dienst Landbouwkundig Onderzoek (IMARES) | Netherlands
Fundación IMDEA Agua (IMDEA) | Spain
University of Natural Resources & Life Sciences, Institute of Hydrobiology and Aquatic Ecosystem Management (BOKU) | Austria
Universidade de Aveiro (UAVR) | Portugal
ACTeon – Innovation, Policy, Environment (ACTeon) | France
University of Liverpool (ULIV) | United Kingdom
Royal Belgian Institute of Natural Sciences (RBINS) | Belgium
University College Cork, National University of Ireland (UCC) | Ireland
Stockholm University, Stockholm Resilience Centre (SU-SRC) | Sweden
Danube Delta National Institute for Research & Development (INCDDD) | Romania
Eawag – Swiss Federal Institute of Aquatic Science and Technology (EAWAG) | Switzerland
International Union for Conservation of Nature (IUCN) | Belgium
BC3 Basque Centre for Climate Change (BC3) | Spain

Contact aquacross@ecologic.eu
Coordinator Dr. Manuel Lago, Ecologic Institute
Duration 1 June 2015 to 30 November 2018

Website http://aquacross.eu/
Twitter @AquaBiodiv
LinkedIn www.linkedin.com/groups/AQUACROSS-8355424/about
ResearchGate www.researchgate.net/profile/Aquacross_Project2