



## Data Management Plan

The AQUACLEW project will work with the Open Science philosophy. The data management plan will identify which data the project will collect and produce, which formats, which will be used and how the data will be made accessible beyond the project partners and how the data will be maintained after the project. In general, the data will be open under the license Creative Commons Attribution 4.0 International (CC BY 4.0) and made available through downloading functions from the project website, together with metadata. The ambition of AQUACLEW is to provide data, metadata, guidance and tools to already existing and sustained Climate Services. The data management plan will not be a fixed document, but will evolve during the lifespan of the project. It will address each single dataset and reflect the current status within the consortium about the data that will be produced.

The datasets generated can be stored and distributed on pan-European climate services as well on national services. The access of new data requires to be communicated to (a) the Scientific Community (b) Engineering and Consulting offices and (c) Authorities. Informing these users about the availability of data sets can be done by means of publications in national journals. Especially engineering and consulting offices as well as public authorities use primarily national journals as the primary source for information on new developments. The authorities and clients in AQUACLEW play a key role, as they are the users who define design parameters to be used in projects and will judge if the AQUACLEW data is reliable/relevant also in future projects.

For the scientific community, the project partners will chose open access journals for publication and make sure to upload data and metadata in open repositories for easy inspection and access. To the extent possible, some scientific experiments will be performed by using the recently developed Virtual Water-Science Laboratory of the EU FP7-project Switch-on <http://www.water-switch-on.eu/> to ensure transparency and reproducibility.

FAIR data management: In general terms, research data should be 'FAIR', 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. (FAIR data principles, <https://www.force11.org/group/fairgroup/fairprinciples> and FAIR principles, <http://www.nature.com/articles/sdata201618> )



## 1 Data Summary

### 1.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?

In relation to the objectives of the AQUACLEW project, the purposes of data collection and data generation are as follows:

Type of Data	Purpose/relation to project objectives
Essential Climate Variables (ECV) and/or Climate Impact Indicators (CII)	Tailored climate information for use cases.



Protocols	Store guidelines for best practice when using climate services in case studies. Three protocols are developed here: <ul style="list-style-type: none"> <li>• Expert Elicitation</li> <li>• Differential Split Sample Test for Bias Adjustment</li> <li>• Differential Split Sample Test for Hydrological models</li> </ul>
Climate friendliness	Climate impact of work within the project, e.g. large computer resources and disk storage facilities, commuting and travels to events.
code repositories	Repositories for sharing open or closed access scientific code, e.g. on github.

## 1.2 What types and formats of data will the project generate/collect?

Type of Data	Format
CII	netcdf, ascii
ECV	netcdf, ascii
Protocols	HTML, PDF
Reports	PDF

## 1.3 Will you re-use any existing data and how?

Yes. All efforts are based on already available data in climate service portals, and from data generated in coordinated experiments such as CMIP5 and CORDEX.

## 1.4 What is the origin of the data?

The origins of data (re-)used in indicator calculations are; regional/global climate models (CORDEX, IMPACT2C) and local datasets.

Dataset	Origin of data
CORDEX	ESGF nodes or in some cases from local storage at contributing institutes
CMIP5	ESGF
Local datasets, e.g. elevation, building data,	Mainly open, but in some cases closed data sources. This is case specific.



infrastructure	

### 1.5 What is the expected size of the data?

The expected size of the datasets will be ..... Mb, divided as follows: See table below

Dataset	Data size
ECVs	~10GB
CIIs	~100MB
Other	~10MB

### 1.6 To whom might it be useful ('data utility')?

The data produced is mostly interesting for the scientific community and water authorities in the regional areas to which the case studies apply to. Below a table is displayed with an overview of the case studies and the (potential) stakeholders who might be interested in the outcomes.



Table 1: Data Utility (experiments)

Case Study Name	Scope	Lead	Area of application	Stakeholders
Flooding from pluvial flash floods in pre-alpine areas		BOKU, UIBK		
Drought and water resource allocation for tourism, agriculture, energy sectors	Assessing the prevision of water allocation in different sectors involved in the watershed management limiting the uncertainty associated to the data provided by pan-European Climate Services.	UCO	Hydropower energy production Reservoir management Tourism development	Hydropower companies Reservoir Manager City councils
Hydropower production	Investigating the use of large-scale climate services to improve the performance of water-energy systems, based on the development of indicators of local impact on reservoir operations in the hydropower sector	Irstea	Hydropower reservoir management	EDF R&D climate services group is following the developments. Other energy companies dealing with hydropower in Europe might also be potentially interested in the outcomes
Irrigation and drainage requirements for agricultural production	Evaluate the application of climate services to assess changes in the drainage and irrigation needs for agriculture focusing on variations of the soil wetness and groundwater table	GEUS	Agriculture management	Water manager - Central Denmark region, agricultural consultant company, Aarhus University - department of Nature, Environment and Farming
Biodiversity decline	To look for future changes in specific indicators that can be used in biodiversity assessments	SMHI	Biodiversity management	Regional municipality in Sweden
Fluvial and coastal interactions under Mediterranean climate	Analyze changes in physical processes such as sea waves, fluvial discharges and sediment transport, that interact and control the dynamics of coastal and fluvial	UGR	Coastal management Urban/tourism planning and development	Local government and coastal managers Port authorities



conditions	environments.			
Urban flash floods in the city of Hagen, Germany	Conducting hydrological simulations based on rainfall observation and projection data in order to determine flood extent, depth and velocity maps that can be used for planning future urban expansion and emergency response planning	TUDO	Urban planning; civil protection/emergency response planning	City of Hagen: Urban planning department, Fire and disaster prevention department



## 2 Fair Data

### 2.1 Making data findable, including provisions for metadata

#### Discoverability

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

The data produced by AQUACLEW has been:

- Provided with the corresponding metadata (see metadata topic below) using the SWITCH-ON tool for *registering open datasets* [4]
- Uploaded using the SWITCH-ON tool for *uploading open datasets* [4]
- Provided with a persistent and unique DOI

#### Metadata

What naming conventions do you follow? Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers? What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Result datasets have been provided with corresponding metadata which complies with the following:

- Based on ISO-19115
- Based on NetCDF Climate and Forecast (CF) metadata conventions
- Keywords lists:
  - started with CUAHSI [5] + extended
  - INSPIRE Topic Categories

### 2.2 Making data openly accessible

#### Open Data by default

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

All data will be open and accessible by default

#### Tools to access the data

How will the data be made accessible (e.g. by deposition in a repository)? What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

All data will be uploaded to the appropriate (OpenDAP, Geoserver and Subversion) repositories by using the SWITCH-ON tool for *Open Data Registration and Upload*.



Access to the datasets is provided by the means of the Open Data Search tool called BYOD. This tool enables to user to select a dataset using a clickable URL so that the dataset itself can be downloaded for further inspection use and use. Within the tools documentation is available.

Both tools are open source and available in GITHUB.





**(physical) Location of the data**

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible. Have you explored appropriate arrangements with the identified repository?

Collected data in	Location
Climate service SWICCA	SMHI
National services	?
Protocols	<a href="http://aquaclew.eu/">http://aquaclew.eu/</a>

**Data access (restrictions and security)**

If there are restrictions on use, how will access be provided? Is there a need for a data access committee? Are there well described conditions for access (i.e. a machine readable license)? How will the identity of the person accessing the data be ascertained?

Collected data in	Restriction(s)	Access provided through
Protocols	No restrictions on use	web
Code repositories	CC BY 4.0	github?
SWICCA	CC BY 4.0 ?	<a href="http://www.swicca.eu">www.swicca.eu</a>

## 2.3 Making data interoperable

**Allowed data exchange**

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers?



institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Yes. All data produced is intended for re-use between researchers, institutions, organisations etc. By examining the corresponding metadata and downloading the datasets (re-)users get an impression on the data quality, formats etc.

#### Metadata vocabularies

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability? In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

As mentioned, the AQUACLEW vocabulary is based on the CUAHSI keywords list which has been extended. This keywords list is well-known in the hydrological community and for this reason this was taken as a basis.

## 2.4 Increase data re-use (through clarifying licences)

#### Data license and timing

How will the data be licensed to permit the widest re-use possible? When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Dataset	License	Available
Protocols	CC BY 4.0	
Climate service data SWICCA	CC BY 4.0	
Code repositories	CC BY 4.0	

Data are under no embargo.

#### After project re-use

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why. How long is it intended that the data remains re-usable?

Yes. The data becomes available for reuse by third parties and will stay available for.....?



---

Metadata quality assurance check list

<https://drive.google.com/open?id=1y5F7XhkO1TBt5vQsVVUYnsrBXQWWMb2zJbEZI1Jc6Dc>

### 3 Allocation of resources

**TO BE COMPLETED**

#### 4 Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

#### 5 Ethical aspects

No ethical aspects are anticipated.

#### 6 Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management?  
If yes, which ones?