



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. 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.

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
CII	Tailored climate information for use cases



Protocols	Store guidelines for best practice when using climate services in case studies.
Climate friendliness	Climate impact of work within the project, e.g. large computer resources and disk storage facilities.
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; CORDEX, IMPACT2C,

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 beMb, 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 by	Area of application	Potential Stakeholders
Flooding from pluvial flash floods in pre-alpine areas		OKU, UIBK (with sub-consultant: HUMER)		
Drought and water resource allocation for tourism, agriculture, energy sectors		CO		
Hydropower production		RSTEA		
Agricultural production		EUS		
Biodiversity sdecline		MHI		
Fluvial and coastal interactions under Mediterranean climate conditions		GR		



Urban flooding		UDO		

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
- 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
SWITCH-ON repos	Deltares?
Protocols	Deltares?

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 ?	www.swicca.eu

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.....?



Data quality assurance

Are data quality assurance processes described?



Metadata quality assurance check list

<https://drive.google.com/open?id=1y5F7XhkO1Tbt5vQsVVUYnsrBXQWWMb2zJbEZI1Jc6Dc>

3 DMP Component: Allocation of resources

TO BE COMPLETED

4 DMP Component: Data security

Virus scans

etc.

firewalls

5 DMP Component: Ethical aspects

TO BE COMPLETED

6 DMP Component: Other issues

TO BE COMPLETED

7 DMP Summary Table

The table below describes the summary of the SWITCH-ON Data Management Plan.

Table 1: DMP Summary Table

DMP Component	Issues (to be)	addressed
Data Summary	INSERT SUMMARY	HERE
Fair Data	INSERT SUMMARY	HERE
Allocation of Resources	INSERT SUMMARY	HERE
Data security	INSERT SUMMARY HERE	
Ethical aspects	INSERT SUMMARY HERE	
Other issues	INSERT SUMMARY HERE	

[Redacted]

[Redacted]

[Redacted]

