European Research Area for Climate Services

CLISWELN

Climate Services for the Water-Energy-Land-Food Nexus

European Research Area for Climate Services
Joint Call for Transnational Collaborative Research
Topic A – Researching and Advancing Climate Service Development by Advanced Co-development with users

Start date of project: 1 October 2017
Duration of project: 3 years

Deliverable 3.1: Data Management Plan

Due date of deliverable: 03-2018
Actual submission date: 03-2018
Organization name of lead contractor for this deliverable: CREAF
Dissemination level: Public

Funders

Project CLISWELN is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by BMBF (DE), UEFISCDI (RO), BMBWF (AT), and MINECO (ES) with co-funding by the European Union (Grant 690462).

• Ministerio de Economía y Competitividad (MINECO, Spain).

• Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF, Austria).

• Executive Agency for Higher Education, Research, Development and Innovation Funding (UEFISCDI, Romania)

• Bundesministerium für Bildung und Forschung (BMBF, Germany).

Document history

Version	*Date*	*Reason of change*
1	2018/01/23	First draft
2	2018/02/06	Incorporate partner's contributions to the previous draft
3	2018/03/07	Incorporate partner's contributions to the previous draft
4	2018/03/31	Final version

License

Authors

Diana Pascual, Anabel Sanchez, Roger Cremades, Nicu Constantin Tudose, Serban Octavian Davidescu, Hermine Mitter, Katrin Karner

Cite as

Pascual D, Sanchez A, Cremades R, Tudose NC, Davidescu SO, Mitter H, Karner K (2018). Data Management Plan. Deliverable 3.1. CLISWELN project.

Executive summary

This deliverable contains the Data Management Plan (DMP) of CLISWELN project. The DMP defines the data sets that will be used and generated, the mechanisms to preserve and store this data and their main standards and metadata, the degree of accessibility of the produced data and materials, as well as the mechanisms to release them and the licenses that will be used.

Contents

1. Introduction

CLISWELN is an **ERA-NET for Climate Services project**. The aim of CLISWELN is to advance the provision of Climate Services (CS) for drought-related decision making, by using the water-energy-land nexus (WELN) to integrate the cross-sectoral links of drought and climate change challenges risk management with further synergistic co-benefits between the provision of climate services and long-term societal objectives like sustainable land planning, mitigation of CO2 emissions and other locally relevant policy targets connected with the Sustainable Development Goals. CLISWELN has **three case studies** with conflicting water uses dealing with cities, regions and river basins that have specific drought-related vulnerabilities: agricultural irrigation affecting the groundwater level with considerable effects on a natural reserve, suboptimal land use management in mountainous areas (mostly forested) compromising water availability downstream, and a vulnerable touristic sector operating in the dry season. The cases are substantially conditioned by the links captured by the WELN, the stakeholders lack tools to understand all its implications, and considering the WELN is necessary for achieving policy coherence with long-term societal objectives.

The Data Management Plan (DMP) of CLISWELN describes the life cycle of the data collected and processed in the project. The DMP will include:

- Raw data, initially compiled to initiate project activities
- Processed data, raw data previously modified from the source and used as input for the project models
- Generated data, selected outputs from the project models used for deliverables, reports and publications.
- Data obtained during the workshops, focus group and interviews of the project
- Publications

This DMP defines the data sets that will be used and generated, the mechanisms to preserve and store this data and their main standards and metadata, the degree of accessibility of the produced

data and materials, as well as the mechanisms to release them and the licenses that will be used. This DMP has followed the JPI Climate Guidelines on Open Knowledge[1] and the FAIR principles[2] (Fair, Accessible, Interoperable and Reusable) for data management and sharing of the EC.

1.1. JPI Climate Guidelines of Open Knowledge

JPI Climate establishes the following set of recommendations for funders, researchers and stakeholders in the climate research system[1]:

1. Open licensing. In order to reconcile the proper management of intellectual property and the broad distribution and use of information and knowledge, the use of the Creative Commons *Attribution license* (CC-BY) and the *public domain license* (CC0) is recommended when dealing with research results and data (including meta-data), respectively.
2. Open formats. The use of open formats i.e. ODF (e.g. *.odt, *.ods, *.odp) in any working document, at least for internal purposes is to be born in mind.
3. Open Access publishing. JPI Climate encourages the publication of research results in Open Access journals, books or proceedings (i.e. gold Open Access) and/or self-archiving of subscription-based formats incl. embargos (i.e. green Open Access). Other possibilities offering restrictive Open Access rights through e.g. national contracts with given publishers are explicitly discouraged.
4. Publishing costs. JPI Climate encourages policy makers and funding agencies to assume in their budgets those costs related to open access data managing and research results publication, including the so-called Author Processing Charges (APCs), if any.
5. Open Data. Establishing Data Management Plans (DMP) as a required criterion for any publicly funded research activity is explicitly recommended. The DMP should be submitted in the proposal and can be evaluated as a part of it.

[1] Sancho Reinoso A, Helgenberger S (2015). JPI Climate - Guidelines on Open Knowledge. Improving the societal

[2] Guidelines on FAIR Data Management in Horizon 2020, version 3.0. 26 July 2016 (URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

6. Open Access and Open Knowledge in the Joint Programming community. JPI Climate will promote the use of these guidelines in the European Research Area (ERA) by cooperating with those initiatives promoting common framework conditions in the public-to-public (P2P) community.
7. Open Knowledge in the climate research community. By means of this set of Guidelines (and particularly chapter 5), JPI Climate invites those actors involved in the climate research community to actively promote, design and implement comprehensive open knowledge policies in order to enhance research activities' societal benefit.

1.2. FAIR principles

The FAIR principles are a proposal of the EC to ensure a good data management. The FAIR principles have been designed by a complete set of stakeholders (academia, industry, funding agencies and publishers) with the aim to act as a guideline to favour the reusability of the data. Following, the FAIR Guiding Principles are described as proposed by Wilkinson et al. (2016)[3]:

To be Findable:

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data it describes

F4. (Meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (Meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 The protocol is open, free, and universally implementable

[3] Wilkinson et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, Nature. doi:10.1038/sdata.2016.18. https://www.nature.com/articles/sdata201618

A1.2 The protocol allows for an authentication and authorization procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

To be Reusable:

R1. Meta (data) are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

2. Raw data

Raw data is the original data needed to feed the models, methodologies and approaches of CLISWELN project. This data is produced neither by the project nor by the partners, but it is provided by data providers, such as climate services, water agencies, public authorities …

2.1. Application of JPI Climate recommendations and FAIR principles to raw data

Raw data will be **findable** on the websites and official portals or by direct contact with the data providers. Some of the raw data will be provided to the project under confidential clauses, under payment or under research conditions limiting its dissemination. For this reason, raw data will not be findable nor openly accessible to the open public, only to project partners. Any public interested should be able to easily turn to data providers to get these same raw data sets.

Raw data used in the project will be **clearly identified** in each deliverable using those data. The following template will be used as standard in all deliverables using raw data:

Raw data	Data source	Type of data	Scale / Resolution	Temporal timeframe	Description
e.g. Meteorological data from the Spanish case-study	*National Meteorological Agency*	*Numerical (mm and ºC)*	*Randomly distributed per meteorological stations in the whole case-study area*	*Daily data, 27 years (1990-2017)*	*Daily precipitation, daily minimum and maximum temperature in 10 meteorological stations*
…					

3. Processed data

Processed data refers to the raw data previously modified from the source and used as input for the project models. Processed data will be produced by project partners, processing raw data to obtain input for project models.

Processed data will be organized within a database in a spreadsheet format, using when possible an open format, i.e. ODF (e.g.*.odt, *.ods, *.odp). The database will be identified with the name *Database 3.1. Inventory of processed data* where the numbers refer to the Work package (WP3) of origin and the ordinal number of the database within the WP (1). This naming convention is the same one used to identify the deliverables in the project. The information included in the database will be:

- Description of the data: data type, scale, resolution, temporal timeframe, data source, regularity of updates.
- Format: .xls, .odt, .tiff
- Archiving and preservation
- Openness and licensing

3.1. Application of JPI Climate recommendations and FAIR principles to processed data

Together with the database, a set of datasets containing the processed data will be generated. All processed data will be described using a standard metadata such as Dublin Core Schema, as it is a flexible and common used standard and is also the one adopted by the European OpenAIRE project. Each prepared data type (as for example raster files, txt files) will be identified with the name *PD3.1. Description of data* where PD refers to processed data, the numbers refer to the Work package (WP3) of origin and the ordinal number of the database within the WP (1). The datasets will be uploaded in the ZENODO repository for data and documents (https://www.zenodo.org/), directly linked with the OpenAIRE repository (https://www.openaire.eu/).

Open formats for working documents will be encouraged (i.e. ODF as *.odt, *.ods, *.odp).

All the data in the datasets will be made **openly accessible** if there is not private information stored in it. The complete datasets will be accessible from the project website (http://clisweln.info) and from the ZENODO and OpenAIRE repositories.

The use of the Dublin Core standard will ensure the **interoperability** of the data. The data will be prepared in a format and type that allows interoperability.

The datasets will be shared as **reusable** data. For this reason, when deposited in the repository, a Creative Commons (CC) Attribution (BY) license (CC-BY) will be requested. This license allows to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) but conserve attribution (appropriate credit must be given to the licensor, provide a link to the license, and indicate if changes were made). The datasets will be available for re-use from the project website and will also be findable and reusable through the OpenAIRE repository. The data will remain re-usable after the end of the project by anyone interested in it with no access or time restriction. No updates are foreseen after the project ends.

There are **no costs associated** to the described mechanisms to make the datasets FAIR and long term preserved. The value for the long term preservation is low, as the responses will soon become out of date.

All the responsibilities concerning **data recovery and secure storage** will go to the repository storing the datasets and database. The OpenAIRE and ZENODO repositories have properly addressed this issue.

4. Generated data

Generated data refers to the outputs from the project models as part of the results of the project, which may be transformed into tables, maps or graphics and included in deliverables, reports, publications and presentations. Generated data will be produced by project partners, introducing processed data in the project models and analysing their outputs.

Generated data will be organized in a database in a spreadsheet format, using when possible an open format, i.e. ODF (e.g.*.odt, *.ods, *.odp). The database will be identified with the name *Database 3.2. Inventory of generated data* where the numbers refer to the Work package (WP3) of origin and the ordinal number of the database within the WP (1). This naming convention is the same one used to identify the deliverables in the project. The information included in the database will be:

- Description of the data: data type, scale, resolution, temporal timeframe, data source, regularity of updates.
- Format: .xls, .odt, .tiff
- Archiving and preservation
- Openness and licensing

4.1. Application of JPI Climate recommendations and FAIR principles to generated data

Together with the database, a set of datasets containing the generated data will be prepared. All generated data will be described using a standard metadata such as Dublin Core Schema, as it is a flexible and common used standard and is also the one adopted by the European OpenAIRE project. Each generated data type (as for example raster files, txt files) will be identified with the name *GD3.1. Description of data* where GD refers to Generated data, the numbers refer to the Work package (WP3) of origin and the ordinal number of the database within the WP (1).). The datasets will be uploaded in the ZENODO repository for data and documents

(https://www.zenodo.org/), directly linked with the OpenAIRE repository (https://www.openaire.eu/). **Open formats** for working documents will be encouraged (i.e. ODF as *.odt, *.ods, *.odp).

All the data in the datasets will be made **openly accessible** if there is not private information stored in it. The complete datasets will be accessible from the project website (http://clisweln.info) and from the ZENODO and OpenAIRE repositories.

The use of the Dublin Core standard will ensure the **interoperability** of the data. The data will be prepared in a format and type that allows interoperability.

The datasets will be shared as **reusable** data. For this reason, when deposited in the repository, a Creative Commons (CC) Attribution (BY) license (CC-BY) will be requested. This license allows to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) but conserve attribution (appropriate credit must be given to the licensor, provide a link to the license, and indicate if changes were made). The datasets will be available for re-use from the project website and will also be findable and reusable through the OpenAIRE repository. The data will remain re-usable after the end of the project by anyone interested in it with no access or time restriction. No updates are foreseen after the project ends.

There are **no costs associated** to the described mechanisms to make the datasets FAIR and long term preserved. The value for the long term preservation is low, as the responses will soon become out of date.

All the responsibilities concerning **data recovery and secure storage** will go to the repository storing the datasets and database. The OpenAIRE and ZENODO repositories have properly addressed this issue.

5. Workshops, focus group and interviews data

This data refers to the information compiled during personal interviews, focus group meetings and workshops. This information refers to the challenges, preferences and needs identified by the project stakeholders along the different participatory activities.

5.1. Application of JPI Climate recommendations and FAIR principles to workshops, focus group and interviews data

The information collected in the different participatory activities of the project will be anonymous, so that personal identification will not be possible. The activities may be recorded and transcribed, but not directly published. The information will be processed, analysed and summarized in the project deliverables, respecting the participants confidentially. The information will be accessible through the deliverables, which will be uploaded in the ZENODO repository for data and documents (https://www.zenodo.org/), directly linked with the public and **findable** OpenAIRE repository (https://www.openaire.eu/).

The deliverables will be **openly accessible**, always guarantying the confidentiality of each participant. The **use of open formats** i.e. ODF (e.g. *.odt, *.ods, *.odp) will be favoured when possible. The deliverables will be accessible from the project website (http://clisweln.info) and from the ZENODO and OpenAIRE repositories.

The deliverables will be shared as **reusable** data. For this reason, when deposited in the repository, a Creative Commons (CC) Attribution (BY) license (CC-BY) will be requested. This license allows to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) but conserve attribution (appropriate credit must be given to the licensor, provide a link to the license, and indicate if changes were made). The deliverables will be available for re-use from the project website and will also be findable and reusable through the OpenAIRE repository. The deliverables will remain re-usable after the end of the project by anyone interested in it with no access or time restriction. No updates are foreseen after the project ends.

There are **no costs associated** to the described mechanisms to make the deliverables FAIR and long term preserved. The value for the long term preservation is low, as the responses will soon become out of date.

All the responsibilities concerning **data recovery and secure storage** will go to the repository storing the deliverables. The OpenAIRE and ZENODO repositories have properly addressed this issue.

6. Publications

Publications refers mainly to the academic working papers (WP2 and WP5), Data management plan (WP3), Engagement and Societal Impact Plan (WP5), policy reports (WP5), Communication and Dissemination Plan (WP6) and other deliverables produced in the project.

6.1. Application of JPI Climate recommendations and FAIR principles to publications

All project deliverables (except the deliverables of WP1) will be **public, findable and openly accessible** in the project website (http://clisweln.info) and the ZENODO and OpenAIRE repositories (https://www.zenodo.org/) (https://www.openaire.eu/). The use of open formats i.e. ODF (e.g. *.odt, *.ods, *.odp) will be favoured when possible. Deliverables of WP1 are internal to the project, describing the issues and agreements reached in the different project meetings. The *Deliverable 1.4 Report of final policy workshop* will be stored as intern document but main outputs of the workshop will be included in the deliverable of WP5, which will be **public, findable and openly accessible**.

The deliverables will be shared as **reusable** data. For this reason, when deposited in the repository, a Creative Commons (CC) Attribution (BY) license (CC-BY) will be requested. This license allows to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) but conserve attribution (appropriate credit must be given to the licensor, provide a link to the license, and indicate if changes were made). The deliverables will be available for re-use from the project website and will also be findable and reusable through the OpenAIRE repository. The deliverables will remain re-usable after the end of the project by anyone interested in it with no access or time restriction. No updates are foreseen after the project ends.

There are **no costs associated** to the described mechanisms to make the deliverables FAIR and long term preserved. The value for the long term preservation is low, as the responses will soon become out of date.

All the responsibilities concerning **data recovery and secure storage** will go to the repository storing the deliverables. The ZENODO and OpenAIRE repositories have properly addressed this issue.