DAFNE

A **D**ecision-**A**nalytic **F**ramework to explore the
water-energy-food **NE**xus in complex and transboundary
water resources systems of fast growing developing countries

DATA MANAGEMENT PLAN

Deliverable D1.2

September 2018

EU H2020 Project Grant No. 690268

Programme Call: Water-5-2014/2015
Project Number:.....690260
Project Title: .....DAFNE

Work-Package: WP1
Deliverable #: D1.2
Deliverable Type: .....Document
Contractual Date of Delivery: 28 February 2017
Actual Date of Delivery: .....28 February 2017
Title of Document: .....Data Management Plan
Author(s): ..... Andrea Castelletti, Marco Micotti, Paolo Burlando, Jos Van Orshoven, Stefaan Dondeyne, Caroline Van Bers, Christian Knieper

Availability:.....This report is public.

Document revisions		
Author	*Revision content*	*Date*
Andrea Castelletti, Marco Micotti	First draft	10.01.2017
Paolo Burlando	Review	20.01.2017
Andrea Castelletti, Marco Micotti	Second draft	20.01.2017
Stefaan Dondeyne, Jos Van Orshoven	Review	13.02.2017
Marco Micotti	Third draft	15.02.2017
Caroline Van Bers, Christian Knieper	Review	17.02.2017
Marco Micotti	Version no. 1	24.02.2017
Paolo Burlando	Final version	28.02.2017

Table of Contents

Abbreviations

CA: Consortium Agreement

GA: Grant Agreement

DoA: Description of Action (Annex I of the Grant Agreement)

GAs: General Assembly

MB: Management Board

PAB: Project Advisory Board

WP: Workpackage

QM: Quality Management

CS: Case Study

EC: European Commission

PO: Project Officer

PR: Project Review

DM: Deliverable Manager

DDP: Deliverable Development Plan

RP: Reporting Period

1. DATA SUMMARY

DAFNE advocates an integrated water resources management approach, which addresses the water-energy-food (WEF) nexus explicitly and from a novel perspective. The project's core is the development of a decision-analytic framework (DAF) that will enable the extensive, quantitative analysis of the anticipated effects of alternative planning options on the broad range of heterogeneous and often competing interests in transboundary river basins of fast developing countries in Africa.

Data in DAFNE will be:

- Collected from multiple sources (e.g. ground stations, remote sensing) to characterize the hydrologic regime, engineering developments, social and economic systems, terrestrial and aquatic ecosystems. A baseline scenario will be defined considering the present situation of the system;
- Generated from models, to build future scenarios and assess effects of changes in the system.

Existing data, available in public repositories or from institutional partners will be re-used whenever possible, while localized field surveys will be carried out to inspect local scale issues or to compensate for eventual data lack.

2. FAIR DATA

DAFNE will adopt three different tools for data management:

- A project **intranet,** composed by shared directories, managed by ETHZ (ETH Zurich) and built using both institutional storage facilities (ETH Zurich *polybox* cloud storage) - to store collected data - and external services (*Dropbox Business Solution*) - for filing generated data in a password-protected way and with access restricted to project partners;
- The **DAFNE Geoportal**, publishing all relevant project data, both public and with restricted access, targeted to be used by project stakeholders during the project. The Geoportal will be powered by a webGIS platform and equipped with customizable user and roles permission tools, providing a subset of common metadata for each dataset, and data viewing and download functionalities. DAFNE Geoportal will be online and working during the lifetime of the project, while the Summer School and the MOOC course will provide the knowledge to ensure its maintenance also after the project completion;
- **Zenodo**[1] repository, to publish and maintain final project outcomes, deliverables and scientific publications, with all the data produced or needed for verification. Project information and related publications and data uploaded to Zenodo will automatically become visible on the OpenAIRE portal.[2]

As for H2020 guidelines on Data Management, this document describes the data management life cycle in DAFNE and will be updated over the course of the project whenever significant changes related to its contents will occur.

1 A data and publication repository, developed by CERN in the OpenAIRE project, freely available to all research programs. (https://zenodo.org/)

2 European Commision Portal for reporting H2020's scientific publications. **(**https://www.openaire.eu/)

2.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

DAFNE will adopt the Zenodo repository to publish project outcomes, including datasets. Using this repository, all the public data of the project will be provided with a Digital Object Identifier (DOI)and a common dataset of metadata (based on *Dublin Core*[3]).

Keywords will be sourced by the standard dictionaries, like the USGS water dictionary[4].

Versions of each dataset will report main and minor changes with dot notation. Main version change will occur after significant changes in the data (e.g. change in the data structure, massive correction or update, changes in the procedure for data collection or generation, …), while minor version changes will occur after data updates or limited correction. Any changes will also be mentioned in the description metadata field.

Naming convention is reported, whenever applicable, in the attached data catalogue (Annex 1[5])

2.2 MAKING DATA OPENLY ACCESSIBLE

Data generated in DAFNE, if relevant for project deliverables or scientific publications, will be published in the Zenodo repository, together with associated metadata.

Data collected will be published on the same repository, unless limitations specified in the attached data catalogue (Annex 1).

Data with restricted access will be published in the DAFNE Geoportal, which will be made accessible through user identification to all project partner and stakeholders.

Data will be stored using standard formats specified in Annex 1 for each dataset and any open source software and tool developed within DAFNE needed to access data will be published on Zenodo or on other public software code repositories, like Github[6].

2.3 MAKING DATA INTEROPERABLE

DAFNE strives to integrate data and information from different disciplines and domains: in order to provide a common understanding on data within the project itself, the use of the *Dublin Core Metadata Element set* vocabulary will be adopted.

Whenever possible and useful, more discipline-specific metadata will be also adopted, as those defined by OGC[7] for geospatial data.

2.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENCES)

Generated data will be licenced with CC-BY-SA licence[8].

Collected data will be generally subject to the same licence, published on Zenodo repository to be accessible to third parties without any time restriction, unless with the limitations reported in Annex 1.

Quality assurance process will be performed through the following steps:

- Collected data

[3] A set of “core metadata" for simple and generic resource descriptions. (http://dublincore.org)

[4] A Water Science Glossary of Terms, compiled by USGS (https://water.usgs.gov/edu/dictionary.html)

[5] Annex 1 table will evolve with the progress of the project, as the data availability and their format will become clear from the data collection activities. Therefore, Annex 1 to this document has to be considered a preliminary version of the consolidated document that will be available at the end of the data collection activity.

[6] Online project hosting platform (http://github.com/)

[7] Open Geospatial Consortium (http://www.opengeospatial.org/)

[8] Creative Commons Attribution-ShareAlike 4.0 International Public License (https://creativecommons.org/licenses/by-sa/4.0/legalcode)

1. Storage of raw datasets, without any further processing, in a dedicated folder in the intranet;
2. Data check and editing for assuring positional, attribute and temporal quality, completeness and consistency, under the responsibility of the project partner listed as reference for each dataset in Annex 1;
3. Compilation of metadata (both generic and domain specific where applicable) reporting a brief summary with the editing done;
4. Storage of the final version of the datasets in a dedicated folder in the intranet;
5. Uploading of the datasets on DAFNE Geoportal, if relevant for sharing with project stakeholders, and/or on Zenodo repository, if compliant with limitation detailed in Annex 1 and if relevant for maintaining also after the project lifetime.

- Generated data
  1. Compiling metadata (both generic and domain specific where applicable) reporting a brief summary specifying the generation process (lineage);
  2. Storing the final version of the generated datasets in a dedicated folder in the intranet;
  3. Uploading of the datasets on DAFNE Geoportal, if relevant for sharing with project stakeholders, and/or on Zenodo repository, if relevant for maintaining also after the project lifetime.

3. ALLOCATION OF RESOURCES

The data catalogue reported in Annex 1 identifies, for each dataset, the responsible project partner. Costs are included in the tasks related to data collection and generation and cannot be listed separately.

The DAFNE Geoportal will be hosted on the Politecnico di Milano (POLIMI) servers using internal resources.

Costs of the project intranet on *polybox* are covered by internal ETHZ resources. As already mentioned in Section 2, *Dropbox Business* accounts will also be activated in order to have more space where to store and share raw datasets and simulation outputs. Costs related to these accounts will be changing according to the numbers of accounts and disk space needed for each of them and they will be covered by project budget.

4. DATA SECURITY

The tools mentioned in Section 2 of the present document are hosted partially on external services (e.g. Zenodo) and partially on ETHZ and POLIMI servers protected by firewall and institutional security policies.

More precisely:

- The **intranet** is relying both on ETHZ storage facilities and on external storage services; accessibility is anyway reserved only to registered users, via *HTTP Secure protocol* (https), for both upload and download functionalities;
- **DAFNE Geoportal** will be hosted at POLIMI**:** data upload will be performed from registered users through *Secure Shell Protocol* (SSH), while download functionalities will be available for registered project partners and stakeholders, unless if not differently specified in Annex 1;

- **Zenodo** repository is hosted at CERN and it is subject to its rules for data security as reported at https://zenodo.org/policies

All datasets maintained on the ETH Zurich *polybox* Intranet and DAFNE Geoportal will be periodically subject to incremental backup in order to avoid data loss.

5. ETHICAL ASPECTS

Ethical aspects related to data management have been already addressed in the Deliverable D8.2.

6. OTHER ISSUES

The Zambezi Watercourse Commission (ZAMCOM) has implemented ZAMWIS[9], a system to support Riparian States with an efficient and timely means of sharing data and information on water resources in the Zambezi River basin.

ZAMCOM is involved in DAFNE as a key stakeholder for the Zambezi: data collected by ZAMWIS will be used as primary source of data for DAFNE, while data produced in the DAFNE project will be made available for integration into ZAMWIS.

[9] Zambezi Water Resources Information System (http://zamwis.wris.info/)

7. ANNEX 1 – DATA CATALOGUE