



SWAMP

SMART WATER MANAGEMENT PLATFORM

Project n^o: 777112

WP7

D7.2 Data Management Plan

Editor: Juha-Pekka Soininen (VTT)

Author(s): Markus Taumberger (VTT), Marcos Visoli (EMBR)

Status – Version: 1.0

Date: 14 May 2018

Distribution – Confidentiality: Public

Code: 777112-SWAMP - D7.2 Data Management Plan



Disclaimer

This document contains material, which is the copyright of certain *SWAMP* contractors, and may not be reproduced or copied without permission. All *SWAMP* consortium partners have agreed to the full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information. The *SWAMP* Consortium consists of the following companies:

The information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Participant no.	Participant organisation name	Part. short name	Country
1 (European Coord.)	Teknologian tutkimuskeskus VTT	VTT	FI
2	Intercrop	ICRO	ES
3	University of Bologna	UBO	IT
4	Consorzio di Bonifica dell'Emilia Centrale	CBEC	IT
5	Quaternium	QUAT	ES
6 (Brazilian Coord.)	Federal University of ABC	ABC	BR
7	Centro Universitário da FEI	FEI	BR
8	Federal University of Pernambuco	UFPE	BR
9	LeverTech Tecnologia Sustentável	LEV	BR
10	Brazilian Agricultural Research Corporation	EMBR	BR

Document revision history

Date	Issue	Author/Editor/Contributor	Summary of main changes
14 March 2018	0.1	Juha-Pekka Soininen	Preliminary document with standard template layout.
25 March 2018	0.2	Marcos Visoli	Open access section added
14 April 2018	0.3	Juha-Pekka Soininen	DMP principles added
14 May 2018	1.0	Juha-Pekka Soininen	Final formatting

Internal review history

Date	Reviewer	Summary of comments
21 April 2018	Alessio Domeneghetti	Accepted with minor changes
23 April 2018	Rodrigo Filev Maia	Accepted with minor changes

Table of contents

Abbreviations	5
Executive Summary	6
1 Open access publication policy	7
1.1 SWAMP project repository	8
1.2 Allocation of resources to open access publishing	8
2 Research Data Summary	9
3 FAIR data	10
3.1 Making data findable, including provisions for metadata	10
3.2 Making data openly accessible.....	10
3.3 Making data interoperable.....	10
3.4 Increase data re-use (through clarifying licenses)	10
4 Allocation of resources	11
5 Data security	12
6 Ethical aspects	13

Abbreviations

ICT	Information and Communication Technologies
DMP	Data Management plan
PSG	Project Steering Group
FAIR	Findable, Accessible, Interoperable and Reusable

Executive Summary

This deliverable D7.2 Data Management Plan results from task in WP7 and creates a practical guide for members of the SWAMP consortium.

It covers issues related to open access of publications and handling and opening of the research data to be collected in SWAMP project.

As the details of pilot implementations and the data to be collected are not known at the beginning of the project, this report gives only principles and generic ideas how the data will be managed.

1 Open access publication policy

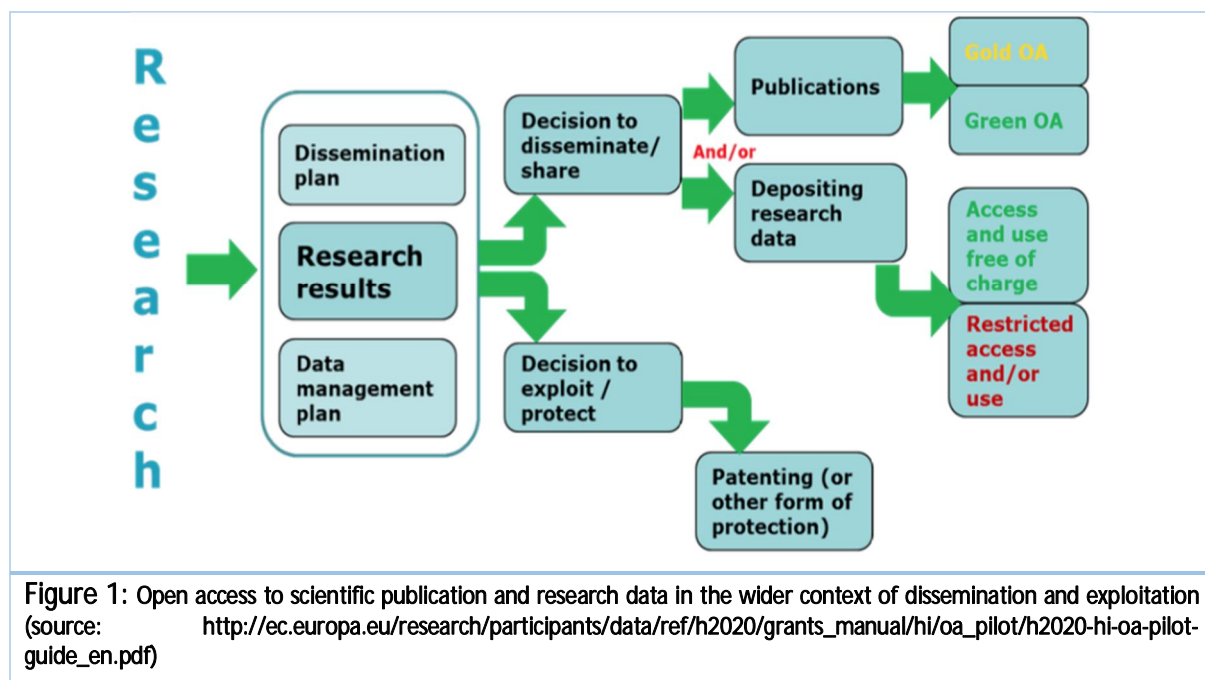
Open access (OA) refers to the practice of providing online and free of charge access to end-users to scientific information.

1. Peer-reviewed scientific research articles (published in scholarly journals or conferences) or
2. research data (data underlying publications, curated data and/or raw data).

This document provides a plan for managing the generated and collected data during the execution of the SWAMP project. We provide guidelines for the data management life cycle, including:

- the handling of research data during and after the project;
- what data will be collected, processed or generated;
- what methodology and standards can be applied;
- whether data will be shared/made open and how;
- how data will be curated and preserved after the project conclusion.

The open access policy for research data is aligned to research dissemination plan and follow the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, summarized in the flowchart depict in Figure 1.



The flowchart distinguishes two main routes for open access publications:

1. Self-archiving, or "green" open access, where the research result is deposited in an online repository (personal webpage, institutional, disciplinary, project or funder repository or open archive portals). Self-archiving should not violate copyright constraints of the publisher, which may restrict the archival accepted (no final) version or request an embargo period.
2. Open-access publishing, or "gold" open access, where the publisher charges a fee to cover the costs of the publication to the authors, and free access to the content is granted to the readers immediately after the publication.

Research data associated to publications may accompany the research article (as supplementary materials) or made available by the same means used for self-archiving, or dedicated data repositories. Data publication may be associated to data papers, which consists of a scholarly publication of searchable metadata describing

the structure of the dataset (or a group of datasets). Data papers can be published in venues, which accept both research or data papers, or specialized journals, which only publishes data papers.

SWAMP project policy

The SWAMP partners will publish project results and software outcomes as follows:

- Scientific papers will be published in fora that implement open access policies, unless there are reasons that either prevent open access publishing or there is significant benefit for publishing result in other fora. In both cases, there must be an explicit decision from PSG to do this.
- SWAMP aims at achieving maximum possible impact with its results, therefore public deliverables will be made publicly available. This default procedure will not be complied with only in cases of particular interests of consortium partners. Results in form of software will be made available according to the deliverable plan that defines which prototype deliverables will be public at what point in time.
- The SWAMP platform will be highly modular, allowing independent deployment of the modules. This makes it possible to publish them under different licenses to account for different use case scenarios and consortium partner interests. If open source is used as a license, allowing commercial usage will typically be required. For modules developed by single partners, those partners will decide about their desired licensing models. In case of joint developments or dependencies between modules, all involved parties will agree on a suitable license for each component. Too restrictive licenses such as the GPL will typically be forbidden since they prevent commercial parties to uptake such open source components. The partners will agree on a detailed process concerning the use of open source in the Consortium Agreement.
- Despite the consortium's efforts to publish open source there are limitations by the background owned by partners. That is why some deliverables are marked restricted at the moment. This might be changed during the project by PSG. The consortium targets making all results public where possible.

Knowledge management

The owner of knowledge will provide adequate and effective protection for knowledge that is capable of industrial or commercial application. The consortium participants may publish information on knowledge arising from the project provided this does not affect the protection of that knowledge. So before any knowledge dissemination that may impact on the exploitation potential of one or more partners takes place, the matter should be agreed with the Project Steering Group.

1.1 SWAMP project repository

The SWAMP project will set up and maintain two repositories for publications.

The VTT OneDrive will be used for storing and sharing publications between project partners.

The SWAMP website (www.swamp-project.org) will have a page that provides open access to publications that are stored at the server hosting the website.

1.2 Allocation of resources to open access publishing

Each academic project partner has allocated 2-3000€ funding to open access publishing costs. Dissemination plan outlines the publishing process and project has a task T6.1 for dissemination that is meant to coordinate also publishing process.

2 Research Data Summary

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

The objective of SWAMP is to develop a high-precision smart irrigation system concept for agriculture. The main idea is to provide tools for the optimisations of irrigation, water distribution and consumption based on a holistic analysis that collects data and information from several aspects of the crop system including the natural water cycle, irrigation infrastructure and the agricultural knowledge related to the crop.

To achieve this objective, the SWAMP platform will provide mechanisms for data acquisition from heterogeneous sensors, including infield soil water absorption at different depths, climatological data, aerial images collected by unmanned aerial vehicles (UAVs), as well as irrigation map plans and crop development status. This data collection is paramount to the success of the SWAMP project, as it will provide input information to optimize models for water need estimation as well as validate the findings.

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

The project will generate data from heterogeneous sources and sensor, including data from soil sensors, climatological sensors, aerial images, amount of irrigation water, irrigation efficiency, crop yield, crop quality, etc. The actual data formats will be defined during the project.

Will you re-use any existing data and how?

The re-use of data is an open issue. During the project, data will be used in analytics of water need and in calculating the correlations between plant growth and crop yield with amount of irrigation and weather conditions.

What is the origin of the data?

Data will be collected from the SWAMP pilot farms and other open public data sources (weather data).

What is the expected size of the data?

The expected size cannot be estimated very accurately at this stage. The size of data will be dominated by amount of pictures taken by possible hyperspectral cameras or other imaging devices. Other data to be collected are relatively simple sensor values that are not even recorded very often. The time resolution will be minutes or hours rather than sub-seconds. The number of sensors that are recorded (spatial resolution) will be less than 1000. The duration of recording will roughly be 16-20 months.

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

The data might be useful as comparison data for farming related research. We do not expect very big demand for the data, but think that it is still feasible to store it for possible use.

3 FAIR data

The project has not yet done decision whether the research data will be put into open access, because we are not yet sure if there is enough value for the expenses of this, and the availability of open access depends on the agreements with data owners. The project pilots are so few that normal anonymization process cannot guarantee privacy, for example.

Because of these reasons, only initial ideas on how possible open sharing of data will be managed in SWAMP.

3.1 Making data findable, including provisions for metadata

The metadata to be created will include what, where, how, and when data was collected. The main intention is to make metadata that allows the replication of experiments possible as we see that the value of data is probably biggest as a comparison data.

Metadata will be described as semantic data and with subset of FAO ontology models. Metadata will be added as part of data collection process.

3.2 Making data openly accessible

This will be decided during the project when we better understand the value of data series. SWAMP data repositories are the first options of sharing platform.

3.3 Making data interoperable

We expect the data to be very simple data series collected from pilot sites. Data will be stored using .csv format.

Interoperability will be tackled using ontology models derived from FAO ontologies.

3.4 Increase data re-use (through clarifying licenses)

These topics will be decided when open access is decided. The intention of the project is to make data as available and usable as possible.

4 Allocation of resources

What are the costs for making data FAIR in your project? How will these be covered?

The project has not reserved additional resources for making data FAIR in the project. The intension is to streamline the data management process so that additional costs are kept as low as possible. The intension is to use otherwise existing servers and resources.

Who will be responsible for data management in your project?

Data management is under task 7.4 and the knowledge and innovation manager is responsible for data management.

Long-term management of data (if considered necessary) will be done by coordinator, i.e. VTT.

5 Data security

The data is not considered confidential and it does not have privacy related aspects. Security measures are adapted to this. We also intend to ensure the anonymization of data.

During the data collection phase the focus in data security is to prevent unauthorised access to the servers that are handling the raw data collection.

The stored collected data will be secured with normal data security processes that are applied at coordinator (VTT).

6 Ethical aspects

The collection of data will happen under informed consent with the field owners. A written agreement on the use and sharing of data will be done with each pilot site owner.