



AG-WaMED | Advancing non conventional water management for innovative climate-resilient water governance in the Mediterranean Area

Grant Agreement Number: UNIFI: 391 del 20/10/2022

Deliverable 1.3.2

Data Management Plan v1



Partnership for Research and Innovation in the Mediterranean Area Programme (PRIMA)

The AG-WaMED project has received funding from the PRIMA Programme, an Art.185 initiative supported and funded under Horizon 2020, the European Union's Framework Programme for Research and Innovation

Deliverable Identification

Deliverable No and Title	D1.3.2 - Data Management Plan v1		
Grant Agreement No	UNIFI: 391 del 20/10/2022	Acronym	AG-WaMED
Project Full title	Advancing non conventional water management for innovative climate-resilient water governance in the Mediterranean Area		
Funding Instrument	Partnership for Research and Innovation in the Mediterranean Area Programme (PRIMA)		
Call	PRIMA CALL SECTION 2 2021 – MULTI-TOPIC		
Work-Package No and Title	Work Package 1: Project Coordination and Management		
WP- Main Beneficiary	UNIFI		
WP-Leader	Elena Bresci, elena.bresci@unifi.it		
Task No and Title	Task 1.3 - Data Management Plan production		
Task Leader	Elena Bresci, elena.bresci@unifi.it		
Main Author	Elena Bresci, elena.bresci@unifi.it		
Contributors	Eleonora Forzini eleonora.forzini@unifi.it		
Status	Draft <input checked="" type="checkbox"/> Final <input type="checkbox"/>		
Dissemination Level	Internal <input type="checkbox"/> Public <input checked="" type="checkbox"/>		
Reviewed by	Giulio Castelli giulio.castelli@unifi.it , Lorenzo Villani lorenzo.villani@unifi.it		
Abstract	<p>The Data Management Plan (DMP) addresses all the issues related to the data gathered and processed during AG-WaMED project and the outputs derived from them. In particular, the DMP defines:</p> <ol style="list-style-type: none"> the roles and responsibilities for data collection and management data sources and formats framework for data reuse data storage, findability and accessibility data security 		
Key words	NONE		
DOCUMENT HISTORY			
Planned Release Date	30 June 2023	Actual Release Date	31 July 2023
Version	V1	Released Version No	1

Table of Contents:

List of acronyms and definitions	4
1. Introduction	5
2. Data Summary	8
2.1. Roles and responsibilities	9
2.2. Data sources and formats	9
2.3 Data verification or reuse	11
2.4 Data utility: potential users of the AG-WaMED research data	11
3. FAIR Data	12
3.1 Data storage and metadata: making data findable and openly accessible	12
3.1.1 Data storage: making data openly accessible	12
3.1.2 Metadata: making data findable	13
3.2 Making data interoperable	14
3.3 Increase data re-use (through clarifying licenses)	14
4. Allocation of Resources	16
5. Data Security	17
4.1. Security settings for AG-WaMED Google Drive shared folder	17
5.2. Security settings for Zenodo	17
5.3. Measures to prevent data loss during the research work phases	18
5.4. Personal data	18
6. Ethical Aspects	19
7. Refences	23

List of acronyms and definitions

Anonymization: Include methods that can be used to convert personal data into anonymized data.

Anonymized data: Anonymized data are no longer considered personal data. Items of information that do not allow the identification of individuals in the data material, neither directly through names or personal ID numbers, nor indirectly through background variables, a list of names, connection keys, encryption formula, or codes. However, a combination of background variables such as gender, age, location, etc. increases the risk of identifying individuals in the data material.

CC: Creative Commons

CC license: Creative Commons licenses are tools to grant copyright permissions to creative work

Data controller: The institution/company/other legally responsible person who 'alone, or jointly with others determines the purposes, conditions, and means of the processing of personal data'. The Data Controller is a formal position and involves requirements for compliance with a number of duties in the Personal Data Act (General Data Protection Regulation, GDPR¹).

DOI: Document Object Identifier.

EC: European Commission

FAIR data: Findable, Accessible, Interoperable, Re-usable data

GA: General Assembly

“Gold” Open Access: Authors make a one-off payment to the publisher so that the scientific publication is immediately published in open access mode.

“Green” Open Access: Due to the contractual conditions of the publisher, the scientific publication can undergo an embargo period up to six months since publication date before the author can deposit the published article or the final peer-reviewed manuscript in open access mode.

Informed consent: Principle according to which human participants can enter research freely (voluntarily) with full information about what it means for them to take part, and that they give consent before they enter the research.

Open Access: Open access is defined as the practice of providing on-line access to scientific information that is free of charge to the reader and that is reusable.

WP: work package

Zenodo²: Zenodo is an open access repository for publications and data by researchers. It is managed by the CERN (Conseil européen pour la recherche nucléaire) for OpenAIRE network and makes it possible to self-archive even for researchers whose institution does not have an institutional repository or does not allow the archiving of certain formats.

¹ <https://gdpr-info.eu/#:~:text=General%20Data%20Protection%20Regulation%20GDPR>

² <https://zenodo.org/record/51902>

Disclaimer

This Deliverable reflects only the author's views, and the PRIMA program is not responsible for any use that may be made of the information contained therein. All intellectual property rights are owned by AG-WaMED consortium members and are protected by the applicable laws. Reproduction is not authorized without prior written agreement. The commercial use of any information contained in this document may require a license from the owner of that information.

1. Introduction

Studies available at Mediterranean level show the evidence on the applicability of technologies to use Non Conventional Water (NCW) for multiple purposes. What is still missing is a “good governance” for the equitable use of such water resources and this needs an advancement of current management practices to be fostered by integrating innovative approaches and evidence-based, co-produced decision making. In this framework, a multi-actor approach that can guarantee the representation of multiple stakeholders in the decision-making process will be promoted to improve water governance schemes.

The aim of the project “Advancing non conventional water management for innovative climate-resilient water governance in the Mediterranean Area (AG-WaMED)” is to provide innovative, evidence-based participatory management solutions to water scarcity governance that can be scaled at the Mediterranean level. The project will propose a transdisciplinary approach, integrating the state of the art on land, water and agronomic modeling to support evidence-based water management in four Living Labs (LLs) located in Mediterranean watersheds in Italy, Tunisia-Algeria, Spain, Egypt.

A participatory modeling suite will allow the co-definition and co-evaluation of different water allocation scenarios for improving water governance in each LL, taking into account the decisive impact of NCW use to ensure climate-resilient strategies under a climate change vision. A flexible yet comprehensive socio-economic approach procedure will enable the evaluation of such scenarios and will trigger a feedback mechanism to generate an adaptive learning environment in the AG-WaMED LLs. The full process will inform new water allocation models and will produce an innovative framework for improving water governance and for reducing conflicts in water-stressed contexts of the Mediterranean.

This report has been elaborated in the framework of Project coordination work package one, task 1.3 (D1.3.1). This deliverable foresees the development of a Data Management Plan (DMP) for AG-WaMED project produced by the coordinating beneficiary (UNIFI), which is the leader of this action and submitted to the General Assembly (formed during the AG-WaMED kick-off meeting) for evaluation. The DMP describes the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. It also reflects the current state of the Consortium agreement on data management. AG-WaMED activities will use and generate data involving and benefiting from the expertise of eight institutions in six Mediterranean basin countries. The DMP provides an analysis of

the main elements of the data management policy that will be used by the Consortium regarding the project research data. This deliverable details the rules concerning the management of data. Its purpose is to establish the processes and the standards that will be followed throughout the project life cycle with regards to the AG-WaMED used and produced data. AG-WaMED DMP will ensure that data management and protection is compliant with EU principles and standards (e.g., General Data Protection Regulation [GDPR]) and with relevant national data protection laws and institutional data management policies. Produced data will be treated according to the management guidelines.

AG-WaMED will use primary data collected during the project and will analyze secondary data that are publicly available or for which agreements will be established during the first months of the project. In the framework of the AG-WaMED project, archival data from partner institutions will be exploited with novel modeling tools, to produce reliable sets of project related indicators. In parallel, a database of the relevant inputs will be obtained from diverse experimental setups in the project's LLs. Data will be stored in structured databases for the activities to be carried out in the project. They will be used for achieving project objectives and related scientific dissemination activities, including follow up scientific activities. Project data will be used both by the partners collecting and generating them and by other partners. Specific agreements to use data jointly involving partners that have not participated in data collection will be restricted during the project to specific authorship objectives without involving monetary transactions. For all the data collected in structured databases, each partner can freely use the data that was collected, whilst the use of pooled data will be granted based on authorship agreements (for scientific publications) or project roles (for deliverables). Where secondary data are used, they will be requested and managed by the partners in charge of the related task and used according to the rules of the providers, including non-delivery to other project partners if this is not part of the agreement made. Data may be used in collaboration with individuals affiliated to non-partner institutions when this is useful to maximize impact. In this case, data use will be granted under a specific non-disclosure agreement. Due to the nature of the project, some personal data will be collected and elaborated in pseudonymized form. This will be done in compliance with EU and partners' countries ethical standards and regulations, including a competent institutional data protection officer, i.e. the UNIFI Ethical Commission. Additional cases that may arise will be regulated according to the Grant Agreement and the Consortium Agreement (CA).

This DMP used the H2020 Template for the Data Management Plan (v1.0, 13th October 2016³) and the Guidelines on FAIR Data Management in Horizon 2020 (v.3, 26th July 2016)⁴. This document is a dynamic document that can change and update if relevant changes occur during the course of the project. The document consists of 6 sections. Each section contains a complete description of a set of operations or processes or links to additional internal documents where the topic is developed in more detail. The sections of the DMP are organized as follows:

3

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

4

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- Section 2 Data Summary (Aim, origin, types, and formats of data generated/collected in the project)
- Section 3 FAIR Data (Data and metadata storage; making data findable, openly accessible, interoperable, and re-usable)
- Section 4 Allocation of Resources (Costs and responsibilities of data management activities in AG-WaMED)
- Section 5 Data Security (Security settings for AG-WaMED and Zenodo platforms, measures to prevent data loss and personal data protection)
- Section 6 Ethical Aspects (Some identified ethical issues)

2. Data Summary

The overall purpose for data collection in AG-WaMED is to establish and fulfill the tasks and subtasks of the 6 project's work packages (WP) figure 1, including support to the project coordination (WP1).

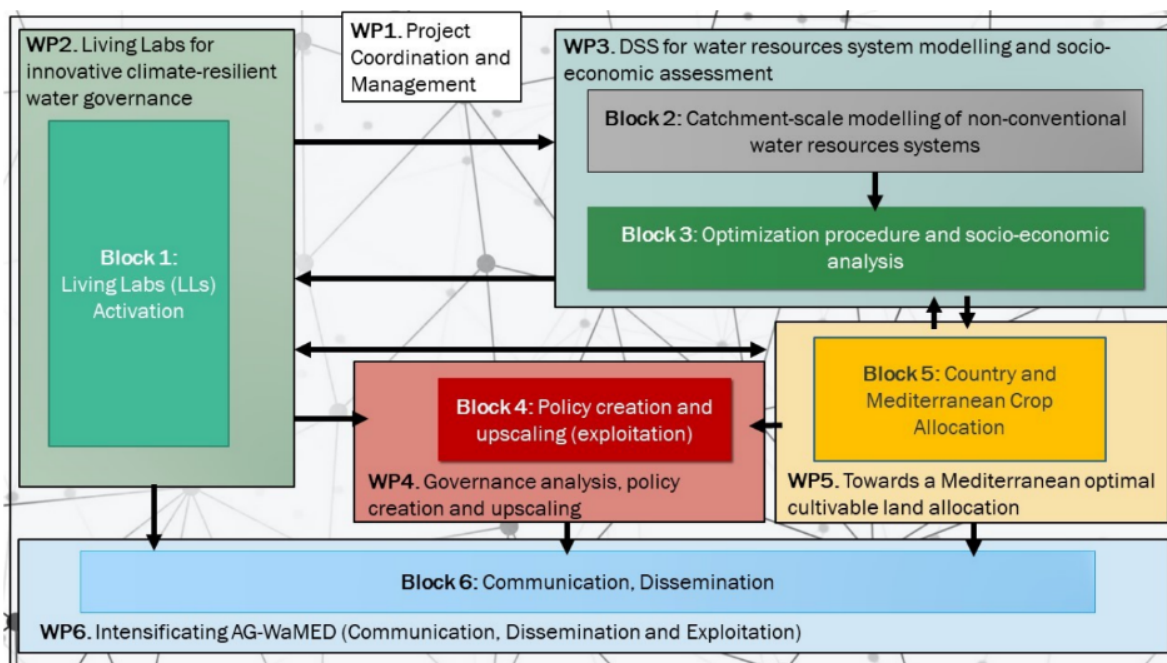


Figure 1. AG-WaMED's project WPs and their inter-relation

The aim of each WP and its respective tasks are as follows:

WP2: Living Labs for innovative climate-resilient water governance

- to define a methodology to set up living labs (LLs) (Task 2.1)
- to carry out a stakeholder analysis and create the LLs (Task 2.2)
- to set the adequate data ground for participatory modeling and scenario analysis (Task 2.3)

WP3: Decision Support System (DSS) for water resources system modeling and socio-economic assessment

- to model NCW at watershed scale (Task 3.1)

- to assess T3.1 modeling results for alternative pathways of socio-economic and environmental sustainability (Task 3.2)
- to analyse the social perceptions and evaluation of benefits from NCW use (Task 3.3)

WP4: Governance analysis, policy creation and upscaling

- To analyse governance and policies (Task 4.1)
- To outscale Integrated Watershed Management Plans and NCW (Task 4.2)
- To out-scale policies at Mediterranean scale (Task 4.3)

WP5: Towards a Mediterranean optimal cultivable land allocation

- To characterize the water availability and water use in the Mediterranean region (Task 5.1)
- To generate scenario and optimal crop allocation alternatives (Task 5.2)
- To co-design and generate policy-relevant conclusions (Task 5.3)

WP6: Intensifying AG-WaMED (Communication, Dissemination and Exploitation)

- To develop and implement an Integrated Strategy for Dissemination, Exploitation and Communication (IS-DEC) (Task 6.1)
- To communicate the project's development and outcomes to the general public (Task 6.2)
- To disseminate the project's outcomes to the scientific community (Task 6.3)
- To use dissemination and exploitation for capacity building of relevant stakeholders (Task 6.4)

In the context of research and innovation and the mentioned AG-WaMED tasks, scientific information can refer to peer-reviewed scientific research articles or research data. Research data refers to information, facts or numbers collected to be examined and considered, and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. In addition to research data multimedia data (text, pictures, audio, video, animation etc.) in concerns to communication, dissemination and exploitation actions will also be collected, elaborated, or generated.

2.1. Roles and responsibilities

Each AG-WaMED partner must respect the policies set out in this DMP. Datasets must be created, managed, and stored appropriately by each partner. Each partner should also ensure that backups are performed, and that proprietary data are secured. The Project Coordinator has a particular responsibility to ensure that data shared are easily available. Registration of datasets and metadata is the responsibility of the partner that generates the data in each WP. Metadata constitutes an underlying definition or description of the datasets, which facilitates finding and working with particular instances of data. Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data. Quality control of these data is the responsibility of the relevant WP Leader supported by the Project Coordinator. If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is

available in the case of publicly available data. Finally, all Consortium members must consult the concerned partner(s) before publishing data that can be associated with an exploitable result in the open domain.

2.2. Data sources and formats

The data types that will be produced during the project are based on the project's WPs and tasks and their results. Datasets can be both collected, unprocessed data as well as analyzed, generated data. Specific datasets may be associated with scientific publications (i.e. underlying data), public project reports and other raw data or curated data not directly attributable to a publication. According to such consideration, Table 1.1 reports a list of categories of research data that AG-WaMED will produce. These research data types have been defined, including data structures, sampling and processing requirements, as well as relevant standards. This list may be adapted with the addition or removal of datasets in the final version of the DMP to take into consideration the project developments and scientific publications. A detailed description of each dataset is given in the following sections of this document.

Table 1: Data origin, type and relationship to AG-WaMED activities

Set no	Data Type	Origin	Data format	Relation to WP
1	Project coordination and management data	Data generated for project management and coordination purposes (agendas, minutes, reports, timesheets, etc.)	Texts, PDFs, Spreadsheets, Presentations, images, and videos	WP1
2	Literature data	Data gathered from desk studies based on in-depth analysis of literature, grey literature shared by partners, reports, etc	Texts, PDFs, Spreadsheets, Bibliographic information files	WP1-6
3	Survey data	Data collected through surveys, questionnaires, and interviews	Texts, PDFs, Spreadsheets, Audio, Video, Pictures	WP2-3-4
4	Generated experimental data	Primary data collected in the field experiments realized in the project LLs (Italy, Spain, Tunisia, Egypt, Algeria) or from participatory and laboratory studies	Texts, PDFs, Spreadsheets, GIS files, Codes, Pictures	WP 2-3-4-5
5	Collected experimental data	Field experimental data collected in the LLs (Italy, Spain, Tunisia, Egypt, and Algeria) by the AG-WaMED consortium or external parties.	Texts, PDFs, Spreadsheets, GIS files, Codes, Pictures	WP 2-3-4-5
6	Generated or collected remote sensing data	Remotes sensing data from aerial or satellite platforms and data acquisitions by drones	GIS files	WP 2-3-5
7	Generated or collected models' data	Data generated by modeling approaches.	Spreadsheets, raster outputs	WP 2-3-5

8	Multimedia data	Data collected or generated to implement effective communication, dissemination and exploitation actions	Texts, PDFs, Pictures, Audio, Video, Animations, Presentations	WP 6
---	-----------------	--	--	------

Concerning file formats, AG-WaMED will use the most common and widely accepted formats for data generation, such as:

- Documents: .PDF/A, .txt, .doc/.docx, .odt, .csv
- Spreadsheets: .xls/.xlsx, .ods, .csv
- Audio: .mp3, .wav, .wma, .ra, .m4a –
- Pictures: .bmp, .gif, .jpg/jpeg, .png, .tiff
- Presentations: .ppt/.pptx, .odp - Video: .avi, .flv, .mov, .mp4, .wmv
- Bibliographic information files: .enl, .ris, .enw, .bib, .xml, .txt, .PDF
- Internet media: typetext/x-fasta
- Geographic Information System (GIS) files: .shp, .dbf, .gdb, .qgz, .tif, .kmz
- Remote sensing images: .tiff, .img, .Geotiff.
- Codes: .R, .IDL, *java*, *py*

The expected size of the project data will be manageable in terms of storage capacity of the Consortium.

Documents and table datasets will be characterized by limited size (few MB maximum), while geospatial datasets, such as the outcomes of the WATNEEDS model⁵, expressed as rasters and vector data cover wide areas and are characterized by tens or hundreds of MB each.

2.3 Data verification or reuse

For the activities foreseen in Subtask 2.3.1, 4.1.2 and Task 5.1, two main data types will be re-used:

- (1) bibliographic research in major academic search engines and expert advice. This includes in-depth literature reviews, scientific articles, other scientific and nonscientific reports, grey literature shared by partners and other communications with experts.
- (2) Numerical, geo-spatial and meteorological data. All the data sources will be full-text or numbered references, and most types of data of this type that will be gathered by the researchers are publicly available, either with or without payment. Regardless of the source of the data, the consortium will ensure compliance with all relevant licenses and maintain clear source referencing.

In case secondary data are used, they will be requested and managed by the partners in charge of the related task and used according to the rules of the providers, including non-delivery to other project partners if this is not part of the agreement taken. Data may be

⁵

https://springernature.figshare.com/collections/Global_Gridded_Dataset_of_Crop-specific_Green_and_Blue_Water_Requirements/4893084

used in collaboration with individuals affiliated to non-partner institutions when this is useful to maximize impact. In this case, data use will be granted under a specific non-disclosure agreement.

In addition, the background⁶ reported in the attachment 2 of AG-WaMED's Consortium Agreement will be used, including the relevant datasets of the thematic model WATNEEDS. WATNEEDS model provides physically based water, crops and ecosystem resource management and scenario analytics with built-in features to simulate effects of climatic, hydrologic and environmental change and analyse water-agriculture-ecosystem interdependencies.

Table 2. AG-WaMED background

Consortium partner	Background	Specific limitations and/or conditions for <u>implementation</u> (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for <u>exploitation</u> (Article 25.3 Grant Agreement)
POLIMI	WATNEEDS dataset and algorithm (version available at https://doi.org/10.6084/m9.figshare.c.4893084)	WATNEEDS dataset and algorithm will be applied in forming AG-WaMED as described in the workplan. WATNEEDS data/output/authors will have to be referenced in all actions and products of AG-WaMED (including reports, software, scientific papers and so on).	WATNEEDS dataset and algorithm will be applied in forming AG-WaMED as described in the workplan. WATNEEDS data/output/authors will have to be referenced in all actions and products of AG-WaMED (including reports, software, scientific papers and so on).

2.4 Data utility: potential users of the AG-WaMED research data

The research data underlying project deliverables requires a large investment of resources, the main goal of sharing data and making data public is for others to use and build on this data, in order to avoid repetition of work and improving efficacy, quality and speediness of

⁶ In accordance with the Consortium agreement, background is defined as “*data, know-how or information (...) that is needed to implement the action or exploit the results*”.

research. Datasets generated within the AG-WaMED project can be of significant interest for:

- i. members of the scientific community as researchers, professors and students;
- ii. professionals operating in the agricultural sector interested in the sustainability aspects of the land and water resources;
- iii. other research projects that can keep developing or combine their work with AG-WaMED datasets;
- iv. decision makers, regional authorities, and governments at local, national, Mediterranean and EU levels.
- v. stakeholders such as research institutes, non-governmental organizations, civil protection, policymakers, environmental agencies, service providers, public authorities, and industrial/agricultural associations; practitioners/end-users/industry and commercial actors
- vi. EU and National advocacy organizations/NGOs, press, civil society

3. FAIR Data

AG-WaMED data management will follow the FAIR guiding principles for scientific data management (Findability, Accessibility, Interoperability, and Reusability of data) aiming to maximize access and re-use of research data after the publication process. Importantly, the principles discussed here are applied not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. At the same time, some parts of the datasets produced within this project cannot be made public in order to protect the privacy of volunteer participants. Data will be stored and made available in common easily accessible formats, with metadata, documentation, and licenses (if relevant), in order to facilitate discoverability, quality assessment and further re-use. All datasets will be identified using persistent identifiers (e.g., DOI) that will be linked to publications, as the underlying data of the research.

3.1 Data storage and metadata: making data findable and openly accessible

3.1.1 Data storage: making data openly accessible

The AG-WaMED project aims to make the research data generated by the project accessible with as little restriction as possible while also guaranteeing the protection of personal data for privacy and/or commercial or security reasons. Data generated in AG-WaMED will be documented and be made discoverable and accessible through a dedicated webpage on the project's website. For scientific publications, a DOI will be assigned to datasets for effective and persistent citation when it is uploaded to a repository. This DOI can be used in any relevant publications to direct readers to the underlying dataset. Each dataset generated during the project will be allocated a dataset identifier. This dataset identifier will be, in

combination with dataset information, included in a metadata file at the beginning of the documentation, and updated with each version.

AG-WaMED will ensure the open accessibility and findability of research data. The project database will be designed to remain operational for 5 years after the project. The following solution for open access and long-term preservation of all public datasets, project deliverables and scientific publications will be considered:

b. The Zenodo repository, which is an open research data repository, for non-military purposes only, that collects research data across all disciplinary areas. Zenodo is hosted and managed by CERN that securely stores all data in the CERN Data Centre's cloud infrastructure.

By the end of the project, the final dataset will be transferred to the Zenodo repository, which ensures sustainable archiving of the final research data. Items deposited in Zenodo will be retained for the lifetime of the repository, which is currently the lifetime of the host laboratory CERN. The final decision on the choice of the repository will be thoroughly discussed by the General Assembly and formalized in the next version of the DMP. In order to ensure maximum availability, all uploads will be linked to the European Commission Funded Research OpenAIRE. In some cases, the use of a particular repository, specific to a research discipline, will also be used for data archiving to make it possible to improve the findability of the data. To determine the most suitable repository for each dataset the Registry of Research data Repositories will be consulted. Partners that generate published data will be responsible for their deposit and compliance with the PRIMA and H2020 Open Data requirements.

Until a dataset is fully finalized and ready for publication, it will be stored privately in a dedicated AG-WaMED cloud for data exchange. Other forms of data exchange, such as email or other cloud services, will also be allowed, as long as all contributors agree, and any restrictions on personal data are respected. At the start of the AG-WaMED project, a Google Drive shared folder, hereafter referred to as the PRIMA AG-WaMED Cloud, will be created to manage the data and prepare the datasets to be uploaded on Zenodo. This is reserved exclusively for users in possession of access details.

Access to the AG-WaMED Cloud has been shared with the Data Controllers of each of the project partners who will be responsible for uploading their public datasets. All anonymized datasets will be stored in the AG-WaMED Cloud that will be the project's online collaboration platform throughout the project life span and until datasets will be useful for the research purposes. The datasets which are not anonymous will be stored: i) locally by the Data Controllers; ii) in a folder with restricted access to the AG-WaMED Cloud. The access to the data of the AG-WaMED folder will be regulated by a specific roles and rights concept according to the WPs logic of the AG-WaMED Project, further, to be defined by the data owners and Project Coordinator. This roles and rights concept will be documented as soon as defined.

3.1.2 Metadata: making data findable

For each dataset uploaded in the AG-WaMED Cloud, the following metadata will be provided: File name, Date, Version, File type, Description, WP (Work Package) number, Responsible person, Lead partner / Data Controller, Dissemination level. Where the dataset is made public on the Zenodo repository, standard Zenodo metadata, together with the Grant Number and Project Acronym, will complement all uploads. Version control and Digital Object Identifiers (DOIs) will be provided by Zenodo for every item. Default Zenodo metadata will be: DOI, Version numbers, Bibliographic information, Keywords, Abstract/description, Associated project and community, Associated publications and reports, Grant information, Access and licensing information, and Language. A Data Controller for each partner will upload the datasets and will be in charge of assigning specific appropriate keywords to these datasets. The dataset-specific keywords must be evocative of the dataset content. For example, a dataset on basin analyses should be tagged with consistent keywords such as "watershed". The following naming conventions will be used to name all the datasets:

Type_StudyCode_Task(orSubTask)Nr_DataController_Description_PRIMA_AG-WaMED_Versioning

Table 2 shows the abbreviations to be used in each component of the naming convention described above.

Table 3 Abbreviations to be used with the AG-WaMED naming convention

Type	Dataset: DS; Report: RP; Deliverable: DL; Bibliographic citations Archive: BA; Patent: PT; Licence: LC
StudyCode	For field studies and surveys the identification codes will refer to the Study Areas as following: Italy: IT, Egypt: EG, Tunisia: TN, Algeria: AL, Spain: ES. For Desk Studies the identification code will be DESK. For Laboratory Studies the identification code will be LAB
WP(or Task or Subtask)Nr	Refers to the Work Package or Task or Sub Task within which the data collection has been carried out. For example, for SubTasks 2.1.1. the code will be ST211. If a dataset refers to more than one Task, then the reference WP will be indicated directly Data Controller Refers to the short name of the partner
DataController	Controller who is responsible for the data set: UNIFI (Università degli studi di Firenze); POLIMI (Politecnico di Milano); UPM (Universidad Politécnica de Madrid); IRA (Institut des Régions Arides); ALEXU (Alexandria University); AGRERI (Hellenic Agricultural Organization ELGO-DIMITRA, Agricultural Economics Research Institute); UTEBESSA (Université Larbi Tebessi de Tébessa); VUA (Vrije Universiteit Amsterdam).
Description	Refers to a brief description of the content of the dataset [max 40 characters including spaces]
PRIMA Acronym	PRIMA AG-WaMED
Version	Refers to the number of the version uploaded

Example: DS_GH_ST221_UPV_watershed_PRIMA_AG-WaMED_001

Once the datasets are published, the Zenodo repository enables further editing and updating of the already uploaded versions through DOI versioning of all uploaded items. In this way it is also possible to cite specific versions of an upload or to cite all versions of an upload. This does not imply that a new DOI will be assigned every time the metadata related to an item is edited. A new DOI is only generated for updates where a new file is uploaded.

3.2 Making data interoperable

In order to make the published datasets interpretable by third parties without any connection to the project, each dataset will be accompanied by a description of the methodology, sources, definitions and scope of the data contained therein. Whenever possible datasets should be structured in such a way that they can be combined, in whole or in part, with other datasets, from the project or from any other data source. In order to ensure interoperability of the datasets it is necessary that standards and methods used in the same research discipline are used. Therefore, for each dataset it will be necessary to check if there are specifications, metadata and/or vocabulary by consulting the Research Data Alliance (RDA's) Metadata Standards (website). Fairsharing.org is a valuable resource where relevant standards as well as databases and repositories can be identified. In the event that the use of uncommon or project-specific ontologies or vocabularies is unavoidable, mapping to the most popular ontologies will be provided. As an internal representation of the metadata, the Zenodo repository uses the JSON schema and allows export to other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and Mendeley. The metadata of the data records will use the vocabulary provided by Zenodo.

3.3 Increase data re-use (through clarifying licenses)

Most AG-WaMED activities will produce scientific information suitable for open-access publication and possibly transferable to stakeholders such as policymakers, public authorities, and industrial/agricultural associations, as well as to the general public. For some technologies, patents or commercially exploitable results could be envisaged. As the dissemination of the project research should not jeopardize the potential protection of generated IP (e.g., patent, product design) and further industrial applications, before any dissemination activity (publication, presentation) strict rules of prior notice to all partners will be applied, according to EC guidelines, to the General Assembly (GA) and to the details included in the next DMP versions. Given the strategic importance of broadening access to publicly funded research by shifting publication costs from readers to researchers, the AG-WaMED consortium will ensure that most of the publications stemming from the project will be published in "gold open access". For those publications not published in "gold open access", a "green open access" option will be used, and maximum embargo terms negotiated with publishers. The underlying data of the scientific publications will be shared after publication in the journal. The data will be associated with the publication. The data related to public deliverables will be made available only once the deliverable has been approved and accepted by the EC. For other public datasets not directly associated with a scientific publication or deliverable, datasets will be shared after evaluation by the Data

Controllers when they are ready for publication. Both for the data directly and not directly associated with a scientific publication or deliverable, the dataset will be made publicly available at the latest in the final month of the project.

If applicable, the AG-WaMED project will regulate the access, extraction, exploitation, reproduction, and dissemination of all public datasets by third parties through the use of Creative Commons (CC) Licenses, which are tools for granting copyright permissions to creative work. The license application will be assessed on a case-by-case basis in close cooperation with the coordinator and interested partners. The license chosen must be registered in the register of characteristics of the attached dataset.

Data classified as confidential will not be reusable by default for privacy reasons. Zenodo has a default mechanism for selecting the appropriate license. All non-anonymous data will be deleted at the end of the project. In case of authorization by the party that provides and owns the data, some non-anonymous data will be kept for a maximum of 4 months from the date of the contractual conclusion of the project.

The additional 4 months are used to keep the underlying datasets available to allow the completion of any scientific publications in preparation towards the end of the project. Exceptions are photos and videos, made with the consent of the volunteer pilot participants, used for communication purposes. If the consent is not revoked before, this data will be kept until it is needed for communication purposes in order to fulfill the contractual obligation of the EC to continue the dissemination and exploitation activities after the end of the project. If a party withdraws consent to the use of this material (images, videos), it will be deleted without delay. Storing data on the Zenodo repository would ensure that it will remain reusable for at least 20 years, as declared by the CERN host laboratory. CERN has also ensured that in the case Zenodo closes its operations, all content (including metadata) will migrate to other suitable repositories. The lead beneficiary responsible for a dataset should strive to make the dataset public as soon as possible, respecting the consent restrictions of all contributing beneficiaries and sufficient time for quality control before publication. In case of publication on Zenodo, a DOI is obtained, to be used in scientific publications and in the deliverables of the project. Beneficiaries should also ensure that they provide the repository address and basic access instructions as part of any dissemination related to the datasets. In order to ensure that the individual datasets contain clear instructions (directly in the relevant dataset files or in a readme.txt file). At least one senior reviewer who was not involved in preparing the dataset will review the dataset. The Data Manager can assume this role. Partners will have the possibility to refuse public dissemination of their own knowledge (background or foreground) when it could potentially harm their own interests. To protect commercially significant innovations by patent and/or copyright, partners will agree on intellectual property rights (IPR) in the Consortium Agreement. Relevant background for AG-WaMED implementation has been listed by each partner in the Consortium Agreement, as well as terms for accessing it, ensuring transparency and availability of the needed Background. Each partner will comply with the PRIMA open access requirements, under the terms and conditions laid down in the Consortium Agreement. In case parts of the research data cannot be openly shared, the updated version of the DMP will provide proper motivations and data availability statements. Not publishing, or embargo or other restrictions is allowed, but only in cases where there are adequate and necessary reasons to do so. An

example of this may be confidential data received from case studies or data collected through interviews.

4. Allocation of Resources

The costs of data management activities in AG-WaMED will be covered by the resources allocated by the EU in the project budget. Part of the time of staff recruited for the project will contribute to the development of the FAIR management through collection and description of the data. The costs relate to the implementation of the AG-WaMED website (domain name registration and web hosting fees etc.) and dissemination material. No additional cost will be necessary for publication on the site because this is included in the cost of managing the site as a whole, including the public and private access areas. If the long-term conservation of the public data is entrusted also to Zenodo, this repository is free for PRIMA projects. Other resources needed to support reuse of data after project finishes will be solved on a case-by-case basis. The overall responsibility for data management lies with the AG-WaMED project coordinator. A Data Management Team, composed of the Data Controllers for each partner and the WP Leaders will support the Coordinator in this task. Those responsible for the activities that generate each dataset have the ultimate responsibility for the correct implementation of the provisions of the DMP.

Open access publications in peer-reviewed journals will be a great part of AG-WaMED's dissemination activity and the strategy on whether to publish in Gold or Green Open Access will be coordinated by the Project Coordinator. In line with the PRIMA Open Access policies, part of the project funding will be made available for Gold Open Access, in particular in case of any major achievement or progress. This is preferable as it avoids periods of embargo, allows faster diffusion, and subsequently a greater visibility of the research. Along with Gold Open Access, Green Open Access will be used to disseminate project results, which, whilst relevant, are not considered to be major findings or advances.

5. Data Security

Each partner who is owner or co-owner of a dataset is responsible for the storage, preservation, and security of the related data. The period for which, after the conclusion of the project, the collected data will be preserved will be determined on a case-by-case basis by each partner. The project Coordinator (UNIFI) is responsible for storing and preserving the project files (deliverables, manuals, etc.) for a period of at least five years after the end of the project. The maintenance of the website and all the information hosted there will be the responsibility of the project Coordinator. All collected data will be managed securely over the duration of AG-WaMED and thereafter, in order to protect them from loss and unauthorized access.

4.1. Security settings for AG-WaMED Google Drive shared folder

The AG-WaMED Google Drive shared folder will be accessible only by the representatives of the partners in the project's Consortium. In addition, a dedicated folder for research datasets

will be created, enabling a stricter access control, where only anonymous datasets will be uploaded. The private part of the AG-WaMED folder has the following security settings:

- Access level: limited to project members only. Additional access restrictions on specific folders are enabled.
- Threat management: security monitoring and file/data integrity prevent and/or record possible data manipulation. A versioning system will give transparency on data modification.

5.2. Security settings for Zenodo

Data security on the Zenodo platform, which supports Closed, Open and Embargoed Access, is broadly tested. On the front page of the Zenodo website only Open Access uploads are displayed. Closed access uploads are still findable via search queries, through the DOI and any open collections in which they are included. Metadata is licensed under the Creative Commons "No Rights Reserved" license, except for email addresses. Via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), all metadata can be exported and collected. Standard protocols, such as HTTP, provide Access to metadata and data files. Data under a variety of licenses is accepted by Zenodo, however, extra benefits in terms of visibility and credit, and additional services and upload quotas are available for the data stored under the most open licenses. All files uploaded to Zenodo are stored in CERN's EOS service in an 18-petabyte disk cluster. Each file copy has two replicas located on different disk servers with a 12 hour backup cycle with one backup sent to tape storage once a week.

The following list describes the security settings for Zenodo:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is stored as a presentation information package. Derivatives of the data files are generated, but the original content is never changed. Records can be withdrawn from public view; however, the data files and records are preserved.
- Replicas: All data files are stored in CERN's data centres, mainly in Geneva, with replicas in Budapest. The data files are kept in multiple replicas on a distributed file system, which are backed up to tape every night.
- Retention period: items will be retained for the duration of the repository. The Zenodo CERN host laboratory has defined a minimum duration for the repository for the next 20 years.
- Functional conservation: Zenodo makes no promises of usability and comprehensibility of the objects deposited over time.
- File Retention: Data files and metadata are backed up every night and replicated in multiple copies to the online system.
- Firmness and authenticity: All data files are stored together with an MD5 checksum of the file contents.
- Files are regularly checked against their checksums to ensure that the content of the file remains constant.

-Succession plans: In case of closure of the repository, Zenodo has guaranteed the migration of all contents in suitable alternative and / or subject institutional repositories.

The full set of Zenodo's "Terms of Use" can be found online at <https://about.zenodo.org/terms/> while the full set of Zenodo Policies can be found online at <https://about.zenodo.org/policies/>

5.3. Measures to prevent data loss during the research work phases

In order to prevent possible loss of data (due to equipment failure, digital or physical theft, malware, etc.) that may occur while the work is still ongoing, a backup system must be adopted by anyone working on the AG-WaMED project. The existence of a single copy of a research dataset will be allowed for a maximum of 48 hours. As soon as possible within that period, even while traveling or working out of the office, a backup of the data must be created. There are no strict rules on the use of any specific backup or cloud solution, however, an automatic backup system with maximum intervals of 48 hours is compulsory. Information in paper format must be kept in a cabinet or office, where only a limited and known number of people have access to it. Also, in this case to be sure that the data is not lost, a digital backup of the data (by scanning or digitally entering the data) should be created as soon as possible (within at maximum one week). Once the information has been processed and the raw data is no longer needed, appropriate long-term storage or destruction measures must be taken.

5.4. Personal data

Personal data will be stored locally and will only be accessible to those who are authorized (Data Controllers). All partners will be responsible to ensure that data remains protected under all necessary security controls (including backup policies and integrity checks) and access controls (identification, authentication, authorization) within their infrastructure. In the unfortunate event of a personal data breach, the project partners will promptly inform their respective competent national supervisory authorities as well as data subjects who may be affected by the breach. At the same time, they will document any personal data breaches and any related information.

For any personal information received through interviews or case studies, the following restrictions apply:

For information in digital format:

- They should only be stored on encrypted hard drives, such as BitLocker;

For information in paper format:

- It must be kept in a locked cabinet or office, where only a limited and known number of people have access to it;
- Once the information has been processed and the raw data is no longer needed, appropriate long-term storage or destruction measures must be taken.

6. Ethical Aspects

The AG-WaMED work will be fully compliant with the regulations established in the General Data Protection Regulation (GDPR - EU 2016/679) applicable, starting from May 2018, in all Member States of the European Union, as well as in the countries of the European Economic Area (EEA). Furthermore, all data collected by the partners of the project will be managed securely in accordance with the applicable ethical standards and requirements and privacy rules and regulations of the respective data collection countries. Moreover, AG-WaMED will always respect the principles of the European Charter for Researchers, the European Code of Conduct for Research Integrity, including ethical standards and guidelines, regardless of the country in which the research is carried out.

AG-WaMED will collect both farmers' knowledge and opinion, general public opinion as well as information related to ecological and agricultural systems through surveys and questionnaires. During the course of the project, two different types of questionnaires and surveys will be developed. Personal data (e.g. name, surname, age, address, gender, civil status, family composition, email and phone number) of the stakeholders that will be involved in research, communication and dissemination activities will be collected. Other technical data that will be recorded are hydrological information (surface water bodies, reservoir data, aquifers data, water quality data), agricultural data (resources needed in term of fertilizers, water and energy for crop production), energy data (including production sources and consumption) and socio-economic information (including measures and their effects). Also the acceptability of the potential use of NCW resources for agricultural and other uses through small-scale questionnaire surveys with open-ended questions in each LL will be carried out. Every effort will be made to ensure that any discomfort and risk will be caused to participants and that gender equality will be guaranteed. Special categories of personal data will be carefully avoided (data relating to racial or ethnic origin; political opinions; religious or philosophical beliefs; sex lifestyle or sexual orientation; trade union membership; genetic data; biometric data; health data).

The data provided will be used only for the research purposes. The beneficiaries responsible for these data collection are committed to follow the local and national regulations regarding personal data protection and will obtain approval from the local/national authorities in charge of personal data protection. All data will be also treated confidentially and in line with the GDPR⁷.

Only the beneficiary responsible for data collection and dedicated project team members who have direct contact with data subjects will have access to personal information such as Name, Surname, address and email address of the participants and these data will be used only for contacting stakeholders when stakeholder engagement activities will be carried out. No one will have access to the data unless this is necessary to perform the project work. The directly identifiable personal data (Name, Surname, address and email address) will be

⁷ [Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

linked to the data collected through a reference number (code) that refers to a distinct list of names. This list will be stored separately from the data on a secure server in a network belonging to the beneficiaries responsible for the data collection. Access to the separate list of names will be restricted and all data will be protected. Even if not shared openly, personal data such as age, gender, civil status, family composition of the participants can still be used for deliverables and publications, but only in the form of aggregated data analysis, which cannot be linked to individual participants.

The project will finish according to plan on August 31st, 2025. After the end of the project personal data stored on a secured server and identifiable personal data (Name, Surname) and the reference numbers that links them to personal data such as age, gender, civil status, family composition of the participants will be irreversibly deleted after it is no longer needed for the defined research purpose. This is to allow the project to close and delete all data according to the required procedures and to allow the final publications of the project after the end of the project. The anonymous data will be documented and stored on a secured server and in a research data repository as open research data, and then made available to colleagues who wish to replicate the study or process the results.

No AG-WaMED publication, including online publications, either directly or indirectly will violate any agreed confidentiality and anonymity. The research results will be reported without affecting rights to privacy and data protection.

Respondents will be asked to sign an informed consent form. The template for this form, as well as detailed information on (i) the procedures and criteria that will be used to identify/recruit research participants within the context of the AG-WaMED Project; (ii) the informed consent procedures that will be implemented for the participation of humans and in regard to data processing, can be found in deliverable D1.4.2. All ethical aspects related to the project have been detailed in deliverable D1.4.1 that will also be approved by an ethical commission.

Vulnerable groups such as minors and persons unable to freely provide informed consent participation is not envisioned. Participation will be voluntary. Participants will have the option to decline and withdraw their participation at any time.

If participants can be identified in the data material, they have the right to:

- obtain access to their personal data recorded,
- request the rectification of inaccurate or incomplete personal data;
- request the deletion of personal data when they are no longer necessary or if the processing is illegal;
- receive personal data in a machine-readable format (data portability);
- file a complaint with the national Data Protection Authority (DPA).

If participants withdraw, all their personal data, interview transcripts and recordings will be irreversibly anonymized. If anonymization is not possible, the information will be deleted. There will be no negative consequences for participants withdrawing their consent.

National Data Protection Authorities (DPAs) in the countries where the data collection will be performed will be notified of the project's activities and DPA approvals will be submitted to the EC:

AG-WaMED will collect images and videos to be used in communication activities (website, newsletter, social media). Images and videos may contain personal data if an individual is the focus of the image or video. Examples include: 1) photos / videos of individuals stored together with personal data (eg ID cards); 2) photos / videos of people posted on the project website along with biographical details; 3) individual images published in a newsletter. Examples of images and videos that are unlikely to contain personal data are: 1) images / videos where people are accidentally included in an image or are not in the focus (for example in a large conference / seminar); 2) images of people who are no longer alive (the GDPR only applies to living people). Under no case will images containing personal information be shared publicly without the explicit consent of the subject.

For other non-anonymous data, such as images and videos used for project communication activities, these will be retained until it is needed for research and communication purposes. Such data will be shared, only with explicit consent, through the project website, newsletter and social media. If a party withdraws consent to the use of this material (images, videos), it will be deleted without delay. Contact information of other external actors established for the purpose of the project will be managed within the project in accordance with the GDPR. All contact lists generated by the project will be stored on the AG-WaMED project website. Access control will be implemented to ensure that only those who require this information to carry out their activities can access it. Contact information will never be shared with third parties and only essential information that is needed will be kept and stored. At the request of external parties, the project will provide information on the personal data that the project is holding relating to the request and will provide the opportunity to correct or delete the information if consent has been withdrawn. During the lifetime of the project, partners may discover business opportunities based on the project's results that can lead to commercial exploitation. This will be monitored and, if any cases arise, appropriate measures will be taken to protect these results for exploitation purposes. The data underlying these results will not be shared openly.