



The PRIMA programme is supported and funded under Horizon 2020, the Framework European Union's Programme for Research and Innovation



Deliverable 6.1

Data Management Plan

VALMEDALM: Valorization of Mediterranean Almond orchards through the use of intercropping integrated strategies



Index

Document information	3
Acknowledgements	4
Executive Summary	5
1. Introduction	6
2. Data Summary	7
2.1. Purpose of data collection and relation to the objectives of the project	7
2.2. Types of data	7
2.3. Formats of the data	7
2.4. Origin and size of the data	8
2.5. Organization of data	8
2.6. Re-use existing data	8
2.7. Potential users of published research data	9
2.8. Availability, verification, curation and preservation of data	9
3. FAIR Data Management	14
3.1. Making data findable	14
3.2. Making data openly accessible	14
3.3. Making data interoperable	14
3.4. Increase data re-use (through clarifying licences)	15
4. Allocation of resources	15
4.1. Estimation of costs	15
4.2. Responsibilities for data management	15
5. Data security, storage and backup	16
5.1. Data security as specified for SharePoint	16
5.2. Research	16
6. Ethical aspects	16

Document information

Deliverable Number:	6.1
Deliverable name:	Data Management Plan
Contributing WP:	WP6: Dissemination, communication and exploitation plan
Contractual delivery date:	M6, November 2022
Dissemination level	Public
Responsible partner:	NRC
Reviewers	All partners
Version:	1

Acknowledgements

VALMEDALM project has financial support provided by PRIMA programme, Call 2021, Section 2 2021 Multi-Topic. The PRIMA programme is supported by the European Union.

Executive Summary

This report constitutes the deliverable D6.1 – Data Management Plan of the VALMEDALM project. This internal document includes information on how research data will be collected, processed, or generated and organized during the lifetime of the project; what methodology and standards will be adopted; whether and how these data will be shared and/or made open and how it will be curated and preserved during and after the project.

During the 36 months of the project, a SharePoint site will be used as the online working and collaborative platform. SharePoint is only accessible to project participants and can provide further access control through establishing folders and sub-sites with stricter access granted than to the main site. During the project all anonymised (public) datasets will be uploaded to this site and stored in accordance with the ethics and privacy strategy of VALMEDALM.

This document is the first version of the Data Management Plan it will be updated at opportune intervals, whenever important changes to the project occur due to inclusion of new datasets, changes in consortium policies or external factors.

This Data Management Plan follows the H2020 Template for the Data Management Plan v1.0, 13/10/2016 and draws on the Guidelines on FAIR Data Management in Horizon 2020 (v.3, 26/07/2016).

1. Introduction

Deliverable 6.1 details the initial version of the Data Management Plan (DMP) of project VALMEDALM. This document intends to provide guidance to the consortium members on how to manage the data that will be collected and produced within the activities conducted in the VALMEDALM project.

Therefore, the Data Management Plan includes:

- Data Summary: Overview of which data will be gathered and processed during the project.
- Determine how data will be stored and processed according to the H2020 FAIR Data Management principles, in order to make data findable, accessible and interoperable, as well as reusable.
- Resource allocation: Consider the costs of making data FAIR in this project.
- Principles on Data Security, Storage and Backup: How is intended to keep the project data secure.
- Ethical aspects: A summary on the ethical and privacy strategies in this project.

The members of the VALMEDALM consortium must follow this Data Management Plan while managing any VALMEDALM related data. The data management must be in accordance with the European Commission (EC) guidelines as well as with open access, FAIR and IPR data principles. In such sense, the VALMEDALM consortium might define certain datasets to remain closed for potential commercial exploitation according to the principle “as open as possible, as closed as necessary”. The Data Management Plan reflects the current state of consortium agreements on data management which is consistent with the requirements for exploitation and Intellectual Property Rights (IPR).

At this early stage of the project, there are several relevant issues regarding the data that cannot be fully answered because they are still under debate by the consortium, however, they will be detailed in the next version of this document. Moreover, the Data Management Plan will be modified and updated accordingly in accordance when major changes might occur, such as the production of new data, potential changes in consortium policies or changes in the composition of the consortium, as well as other external factors.

2. Data Summary

This section of the Data Management Plan has the purpose of presenting the numerous data that is going to be created and collected throughout the execution of the VALMEDALM project as well as its formats, origin and size (Table 1). The data collection will be complying with all the existing national and EU ethical and legal requirements.

2.1. Purpose of data collection and relation to the objectives of the project

The purpose of the VALMEDALM project is to empower local almond production of the Mediterranean through the implementation of intercropping practices as an integrated strategy aligned with economic and social aspects, as well as sustainable principles towards an adaptation to climate change. Given the concept of the project, whereby each partner will generate its own data, from the inventory of intercropping systems used in almond orchards in the Mediterranean (WP1) to the assessment of the impact of such intercropping systems using environmental, economic and social indicators (WP4), structured databases will be built and shared by all partners in order to optimise exploitation.

So, VALMEDALM will produce high quality integrated datasets in the areas of agronomy, pest and weed management, food production, including access to scientific papers and studies based the implementation of intercropping strategies. For this reason, VALMEDALM data management approach is described in this document, which applies to all datasets that the project will generate. However, potential deviations from the default approach will be needed for some datasets, due to reasons such as confidentiality of data or embargo periods related to IPR or scientific publication.

2.2. Types of data

The data generated through the VALMEDALM project will cover a wide range of research fields. Such deviations are possible, as long as they are well explained and documented in a future update of this Data Management Plan.

Therefore, the data generated by VALMEDALM will include:

- Data of success and failure factors of intercropping system in almond orchards (WP1)
- Strategies to encourage farmers to adopt sustainable agriculture (WP1)
- Data of intercropping effect on crop production and environmental impact (WP1)
- Data of practices for pest and weed management (WP2)
- Data of nutritional and biochemical composition of almonds and intercrops (WP3)
- Economic, social, environmental data (WP4)
- Treated data on main results of the demo sites (WP1-WP4)
- Literature data, textbooks, webpages, newspaper articles, social network data, and blog posts (WP5)

2.3. Formats of the data

A dataset might include different types of formats. The VALMEDALM project will be using only widely accepted and standardized formats for data generation, such as:

- Documents/Reports/Publications/Presentations: .pdf, .txt, .doc/.docx, .ppt, .xlsx;
- Databases: .csv;
- Audio files: .mp3, .wav, .wma, .ra;
- Pictures: .jpg, .png, .tiff, .gif;
- Video: .avi, .flv, .mov, .mp4, .wmv, .mkv;
- For statistical purposes: sas7bdat (SAS), RData (R), .SAV (SPSS), .mat (MATLAB).

Applicable data formats may be migrated when new technologies become available and proven to be robust enough to ensure digital continuity and continued availability of data.

2.4. Origin and size of the data

The origin of the generated/collected data will be:

- Feedback from participants at stakeholder workshops/seminars;
- Survey responses obtained in experimental field trials;
- Market survey;
- Literature study/review and open data (re-use of existing data);
- Project meetings, field visits, conference calls, net meetings, emails, shared data platforms, projects, original synthesis and review articles;
- The dataset originates from the project's social media accounts (Facebook, Instagram, Twitter) and websites.

The exact size of the data cannot be anticipated nor determined at the current stage. Such size of the data will be evaluated during the execution of the project and regularly registered and reported in the updated version of the Data Management Plan.

2.5. Organization of data

The research data files and folders will be labelled and organised under a system that ensures that they will both perfectly identifiable and accessible for current and future users. For that purpose, some useful criteria (of organisation, context and consistency) will be taken into to provide a naming convention or protocol to follow recommended rules (example: 20221121_NAME_15):

- Do not use generic names to avoid conflicts
- Use short and relevant file names (under 25 characters)
- Use underscores instead of full stops and spaces
- Distinguish the different versions of your data consistently

2.6. Re-use existing data

Project VALMEDALM will capitalize on already existing data and will integrate the newly generated data during the project. Some examples are:

- In WP1, existing databases on intercropping practices used in the Mediterranean, internet sources (websites, Youtube, patent databases, etc.), review of scientific and technical literature to create an inventory on success and failure factors for the implementation of intercropping practices;
- In WP2, an inventory of the main pests affecting almond orchards will be created to identify and characterize the main pests in almond orchards across the Mediterranean region;
- In WP3, to quantify some compounds (e.g. fatty acids, tocopherols, phenolic compounds) software databases will be used to assist on the identification;
- In WP4, existing databases, internet sources and scientific and technical literature will be used for setting up the most suitable environmental, technical and social indicators to be used in order to evaluate the impact of intercropping practices in almond orchards.

Furthermore, in order to support such data re-use, will be made available lists of datasets collected and generated during the project on the VALMEDALM website and some partners' repositories, and also access procedures drafted for those data. For all the available data will be given appropriate references to authors and institutions.

2.7. Potential users of published research data

Datasets generated by project VALMEDALM might present interest for:

- Members of the scientific community
- Agri-food sectors, including primary production
- Professionals linked to plant-based products
- General public
- Other research projects that can continue to build on or combine their work with the VALMEDALM datasets
- Policy makers in relevant topics such as agriculture, climate change, sustainability and pest management

2.8. Availability, verification, curation and preservation of data

Each WP leader is responsible for data recording and validity checking according to the consortium's procedures and uploading (intra-consortium portal). Through the upload of results in the defined portal, confidentiality and data protection will be carefully taken into consideration. The project Coordinator and the Project Management Team will be developing a strategy for the identification of the most suitable data sharing and protection methods, using clear rules on data accessibility and quality verification. Post-project maintenance, management, and exploitation of the VALMEDALM results will be undertaken by coherently articulating the needs and opportunities that will emerge from the Exploitation Plan. The consortium will then ensure long term access to data through website maintenance after the conclusion of the project (in a period of two years). The use of public datasets will be assigned to the VALMEDALM website (<https://valmedalm.eu/>) and open dataset will be made available in open access repositories openly accessible and free of charge.

Table 1. Overview of expected datasets in VALMEDALM

Work packages	Task	Partner Responsible	Datatype & Dataset Description	Description/Purpose Data	Data Format	Access type	Utility
WP1	1.1	CNCFS	Experimental data: Inventory of failure and success factors of intercropping systems	This dataset will contain an inventory on the use of intercropping systems and strategies in order to present failure and success factors to be taken into account.	Tabular data (.csv, .xls), text data (.txt, .doc), Graphs (e.g. .tiff, .jpeg, .gig), email (.eml), electronic docs (.egt)	Public	Researchers, general public, nut and/or legumes producers, agri-food sector
	1.2	CNCFS	Experimental data: establishment of almond orchards with intercropping practices	The data will provide information about the implemented demo sites		Confidential results used for confidential deliverable	
	1.3	CNCFS	Experimental data: evaluation of the effect of intercropping practices on almond orchards	Data of the productivity, plant growth, use of resources, input use, soil fertility, soil and plant microbial population composition			
WP2	2.1	ARO	Experimental data: Inventory of main pests affecting almond orchards in the Mediterranean region	Data on the main pests in almond orchards in the Mediterranean region to characterize exposures to pests, environmental and agroecological conditions that contribute to sensitivity or tolerance to different pests	Tabular data (.csv, .xls), text data (.txt, .doc), Graphs (e.g. .tiff, .jpeg, .gig) email (.eml), electronic docs (.egt)	Public	Researchers, general public, agri-food sector, ecological sector
	2.2	ARO	Experimental data: Inventory of practices for pest and weed management associated with intercrops	Data on the main practices for pest and weed management, associated with intercrops, already applied by local farmers in the Mediterranean		Confidential	

	2.3	ARO	Experimental data: Impact of intercropping practices on pest and weed management	Data on the impact of intercropping practices on the pest and weed management, namely biodiversity, almond orchards characterization, soil analysis, etc.		Public	
WP3	3.1	IPB	Experimental data: Nutritional and physicochemical characterization	This dataset will contain results from the assessment of nutritional and physicochemical composition on almonds and intercrops produced under intercropping practices in order to determine the most promising samples.	Tabular data (.csv, .xls), text data (.txt, .doc), Graphs (e.g. .tiff, .jpeg, .gif) email (.eml), electronic docs (.egt)	Public	Researchers, general public, food sector
	3.2	IPB	Experimental data: Chemical composition	This dataset will contain results from the characterization of almonds and intercrops samples concerning its composition in fatty acids, organic acids, vitamin C, carotenoids, tocopherols, folic acid and phenolic compounds.		Public	
	3.3	IPB	Experimental data: Bioactive compounds characterization	This dataset will contain results from the characterization of the samples in terms of bioactive compounds, in order to select the most promising		Confidential	
WP4	4.1	UNIPA	Experimental data: Economic analysis of almond production	This dataset will contain study costs (labour, energy, fuel, seeds, etc.) from almond production using intercropping practices	Tabular data (.csv, .xls), text data (.txt, .doc),	Public	Researchers, general public, food sector

	4.2	UNIPA	Experimental data: Determination of the most suitable environmental, technical and social indicators	This dataset will contain results covering the three pillars of sustainability (environmental performance, economic development, and social growth)	Graphs (e.g. .tiff, .jpeg, .gig) email (.eml), electronic docs (.egt)	Public	
	4.3	UNIPA	Experimental data: Assessment of the integration of the different indicators	This dataset will contain results on the joint assessment conducted using MCDA tools to integrate environmental, economic, and social sustainability results		Public	
WP5	5.1	CNCFS	Interaction channels (workshops, webinars and conferences)	This dataset will contain results from the implementation of interaction channels to promote knowledge transfer	Tabular data (.csv, .xls), text data (.txt, .doc), Graphs (e.g. .tiff, .jpeg, .gig) email (.eml), electronic docs (.egt)	Public	Farmers, associations, farmers cooperatives, companies, the scientific community, consumers
	5.2	CNCFS	Demo site visits	This data will contain the knowledge transfer derived from visits and field trips along the demo sites of the project		Public	
	5.3	MORE	Data collected through the implementation of network strategies	Data will result from the creation of a knowledge networking framework to promote collaborative work, knowledge transfer and women's engagement in productive activities		Public	
WP6	6.1	NRC	Data management plan	Information on how research data will be collected, processed or generated	Tabular data (.csv, .xls), text data (.txt, .doc), Graphs (e.g.	Public	Consortium project, Researchers, general
	6.2	MORE	Multimedia collected to promote intercropping practices in almond orchards	Data used in communication and dissemination activities (e.g. website, newsletter, social media, etc.)		Public	

			and shared through different means		.tiff, .jpeg, .gig) email (.eml), electronic docs (.egt)		public, agri-food sector
	6.3	NRC	Data on technical, legal, and economic aspects to exploit the results of the project	Data to identify of strengthens, weaknesses, opportunities and threatens, Marketing Strategy and financial projections.		Public	
WP7	7.1	MORE	Reporting communication data and	This dataset contains documentation to the European Commission (deliverables, cost statements, audit certificates, etc.); project intranet and documents server; partners contact list (contacts of consortium and Management Board).	Text data (.txt, .doc), Tabular data (.csv, .xls); pdf; email (.eml), electronic docs (.egt)	Confidential	Consortium project
	7.2	MORE	Administrative data	This dataset contains documentation related with administrative activities (agreements, certificates, reports...).			
	7.3	MORE	Technical and contractual data	This dataset contains documentation related with technical and contractual activities (financial statements, relevant certificates, contracts, reports...).			
	7.4	MORE	Financial data	Bank details for the distribution of the financial contribution of the Financing Authority			

3. FAIR Data Management

VALMEDALM will manage data in accordance with the principles of FAIR data management (Findable, Accessible, Interoperable and Re-usable data). The project aims to maximise the access to, and to re-use-of research data generated by the project. At the same time, there are datasets, or parts of datasets, that will be generated within this project which cannot be shared in order to protect the privacy of data.

The management of the data collected and produced by VALMEDALM will be in compliance with the article 29.3 of the H2020 Annotated Model Grant Agreement (AMGA).

3.1. Making data findable

Data and metadata should be found through open data repositories (institutional repositories if possible, otherwise ZENODO can be an alternative that is normally suitable for European Institutions and others), that are using unique identifiers (DOIs) as well as bibliographic metadata standards (e.g. Dublin Core) for the dataset description. (Meta)data will be registered and indexed in standardized data repositories that make searchable and findable where is included bibliographic metadata (as recommended in article 29.2 of the AMGA), such as:

- Terms “PRIMA”, “European Union (EU)” and “Horizon 2020”;
- The name of action, acronym and grant number;
- The publication data and length of embargo period if applicable;
- A persistent identifier (DOI).

3.2. Making data openly accessible

The Grant Agreement (Art 29.3) aims to make research data which is generated in H2020 projects to be accessible with as few restrictions as possible, but also to accept protection for personal and sensitive data due to privacy concerns and/or commercial or security reasons.

All the datasets used in VALMEDALM will be accessible through an open repository thanks to the DOI (persistent identifier). Datasets with dissemination levelled as “confidential” (non-anonymous datasets) will not be shared due to privacy concerns, therefore only be accessible within the project consortium. Potentially, some datasets might be restricted due to protection for commercial exploitation. If such cases would arise during the execution of the project, such information will be reported in the final version of the Data Management Plan.

The list of expected datasets presented in Table 1 constitutes the initial version of dataset description and we acknowledge that it will be developed throughout the evolution of VALMEDALM. In addition, there is information concerning the datasets that remains unknown at the moment, e.g. size of the datasets. An updated version of such list will be produced in the end of the project.

3.3. Making data interoperable

Data produced in the VALMEDALM project must be interoperable to allow data exchange and re-use between researchers, institutions, organizations, countries, regions, etc. In order to be interoperable, research data and metadata must adhere to standard formats, and as much as

possible, be compliant with available (open) software applications, and in particular facilitating recombinations with different datasets from different origins.

- ❖ While selecting a file format, VALMEDALM will prefer formats that are non-proprietary, not encrypted, uncompressed, open and documented by the community, having common character encoding and adapted for the data type;
- ❖ VALMEDALM will prefer using the most common formats, according to the European Data Portal: .csv, .txt, .html, .json, .pdf, .xls and .xml;
- ❖ Research data should include qualified references to other research data;
- ❖ Research data must use formal, accessible, shared, and widely applicable language for knowledge representation;
- ❖ Research should have vocabularies that follow the FAIR data principles and, if necessary, generate project specific ontologies or vocabulary.

3.4. Increase data re-use (through clarifying licences)

Datasets will be made available through open data repositories by the format of Creative Commons license (e.g. CCBY and CCO), which ensures long-term archiving (at least 10 years). In the data re-used principle is included the rest of FAIR principles: fundability, accessibility and interoperability. Datasets must be well-documented (information of the provenance of data: who, what, where and when of the data).

4. Allocation of resources

4.1. Estimation of costs

Public project deliverables and close datasets (bibliographic linked cites) will be uploaded to the VALMEDALM website. There will be no additional cost for publication on the website as it was developed by a partner (MORE) which will also be responsible for its management. On the other hand, VALMEDALM open datasets will use standard formats in a free of charge research data repository (institutional repositories, or ZENODO as an alternative).

However, some project partners institutions have budgeted dissemination costs supporting scientific articles to Open Access Journals. Other costs of data management activities will be covered by allocated resources in the project budget. Anyway, long-term preservation of public data will be ensured through partners repositories. Other resources needed to support the re-use of data prior to the end of the project will be solved on a case-by-case basis.

4.2. Responsibilities for data management

Through the execution of VALMEDALM, the Steering Committee (Leader: Alexandre Gonçalves (MORE); Participants: a representative from each organisation) will be the main responsible for data management following the guidelines of this Data Management Plan. In any case, the potential decisions on data management will not conflict with the exploitation of the project's results foreseen by the Plan for Exploitation and Dissemination of Results. Its members will be therefore responsible of informing their respective research teams on the provisions set by this Data Management Plan and thus monitoring its compliance in their own institution.

5. Data security, storage and backup

5.1. Data security as specified for SharePoint

The VALMEDALM project involves the participation of 9 partners and a large amount of data to be generated. Each partner will be responsible for the security of its data, as well as storage and backup, however general recommendations and protocols are given:

- Save three copies of your data on two different storage mediums, as well as one copy off site;
- Backup your files/data on a regular basis and use different media for it (such as external hard drives, computer hard drives, department servers, etc.);
- Do not backup or store sensitive data on commercial clouds (such as Dropbox, Google Drive, etc.);
- Enable computer firewalls and keep antimalware software always up-to-date and operational;
- Users must have access to the computers and/or servers via individual user accounts and not shared accounts;
- On collaborative networks/platforms/Intranets: use permission-controlled files so that users, depending on their status, can “read only”, “write”, or “execute” files. Computers connected in such networks should not store sensitive data, unless data is encrypted, so as to minimize network vulnerabilities;
- Cloud-based storage is useful as a secondary or tertiary storage location for files.

VALMEDALM is working in a virtual workspace (Share Point) to storage and share internal data and documents. As coordinator partner, MORE, will establish a standard procedure by Share Point access to project members only where consortium partners will deposit the data generated by the project. If highly confidential protocol or data are to be restricted, specific private channels only for authorised partners can be created. Each partner has IT departments which ensure security standard protocols mentioned above and each WPs have data storage and backup specific conditions and requirements.

5.2. Research

Data repositories (such as ZENODO, institutional open data repositories...) are recommended in such sense. However, this via is not recommended for sensitive or confidential data that will be encrypted or treated anonymously.

6. Ethical aspects

While no general ethical aspects are foreseen, on the topics of evaluating consumers' interest and acceptance of VALMEDALM outcomes, when it comes to creating surveys and organizing focus groups, partners responsible for case studies that should have a Data Protection Officer (DPO) (or similar figure) as well as an ethics committee in place (or a similar figure). When conducting experiments or surveys, the Principal Researcher and the Project Coordinator require a formal approval from the ethical committee and the DPO. More specifically, the data collection method and the model of the questionnaire as well as the focus group guide must be approved by the ethics committee and the data collection must be conducted according to the ethical principles expressed in the Declaration of Helsinki. The Article 19 will be emphasised which introduces the concept of social justice while extending the scope from individuals to the community as a whole

by stating that “research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research”.

Furthermore, the protection of personal information according to the General Data Protection Regulation (EU) 2016/679 (EC, 2016) (GDPR) must be ensured or in accordance with the national Data protection rules, if required. The development of focus groups, the survey and involvement of personal interviews (surveys and discussion sessions) proposed for extracting data from consumers or stakeholders will be following the European General Data Protection Regulation No 2016/679 amended in May 2018. In particular, consumers will be randomly selected and informed about the purpose of the given study and the funding entity. Consumers will also be informed that all information collected will be exclusively used for research purposes and its confidentiality will be absolutely guaranteed. The surveys respondents will be given the opportunity to omit the questions which they do not want to answer and to withdraw from the survey at any moment of the process.

For the Stakeholders analysis, a face-to-face discussion session will be carried out and a consent form will be provided. Once the participant of the Focus Groups are defined according to the selection criteria, names and contact details will be collected for further participations following the EU rules protection.

The VALMEDALM does not include animals in any of the proposed tasks. Third countries: personal data will be imported from Israel and Morocco to EU countries and treated as described above.