VALERECO

Valorization Legumes Related Ecosystem Services

D7.2: Data Management plan (Version 1)

Responsible Authors: Ilias Travlos, Metaxia Kokkini [AUA]

Funded by
the European Union

Document Information

Grant Agreement No.	**101135472**		
Project Acronym	**VALERECO**		
Project Title	**Valorization Legumes Related Ecosystem Services**		
Type of action	HORIZON Innovation Actions		
Call	HORIZON-CL6-2023-BIODIV-01		
Start – ending date	01/06/2024 – 31/05/2028	**Duration**	48 months
Project Website	https://www.valereco.eu/		
Work Package	WP7: Project management & coordination		
WP Lead Beneficiary	Agricultural University of Athens (AUA)		
Relevant Task(s)	T7.3 Data Management		
Deliverable type [1]	DMP — Data Management Plan	**Dissemination level** [2]	PU-Public
Due Date of Deliverable	30 November 2024		
Submission Date	29 November 2024		
Responsible Author	Ilias Travlos [AUA]		
Contributors	Metaxia Kokkini [AUA]		
Reviewer(s)	Alexandros Tataridas [UC]		

[1] *Please consult the Grant Agreement: R: Document, report; DEM: Demonstrator, pilot, prototype, plan designs; DEC: Websites, patents filing, press & media actions, videos, etc.; DATA: Data sets, microdata, etc; DMP: Data management plan; ETHICS: Deliverables related to ethics issues; SECURITY: Deliverables related to security issues; OTHER: Software, technical diagram, algorithms, models, etc.*

[2] *Please consult the Grant Agreement: PU – Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page); SEN – Sensitive, limited under the conditions of the Grant Agreement; Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444; Classified C-UE/EU-C - EU CONFIDENTIAL under the Commission Decision No2015/444; Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444*

Document History

Version	Changes	Date	Contributor
0.1	Initial Version	15/10/2024	Ilias Travlos, Metaxia Kokkini (AUA)
0.2	Several changes	27/10/2024	Nikos Antonopoulos, Metaxia Kokkini (AUA)
0.3	Several changes	11/11/2024	Ilias Travlos (AUA)
0.4	All partners completed the tables in Section 3 (Data Summary)	20/11/2024	All partners
0.5	Project Partner Review	25/11/2024	Alexandros Tataridas (UC)
1.0	Final Version submission	28/11/2024	Metaxia Kokkini (AUA)

VALERECO Consortium

No.	Participant organization name	Short name	Country
1	GEOPONIKO PANEPISTIMIO ATHINON	**AUA**	GR
2	ESC DIJON BOURGOGNE	**BSB**	FR
3	DELPHY BV	**DELPHY**	NL
4	UNIVERSIDADE DE COIMBRA	**UC**	PT
5	REFRAME FOOD ASTIKI MI KERDOSKOPIKI ETAIRIA	**RFF**	GR
6	INSTITUT ZA RATARSTVO I POVRTARSTVO INSTITUT OD NACIONALNOG ZNACAJA ZA REPUBLIKU SRBIJU	**IFVCNS**	RS
7	GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER	**LUH**	DE
8	STICHTING WAGENINGEN RESEARCH	**WR**	NL
9	INSTITUTO NAVARRO DE TECNOLOGIAS E INFRAESTRUCTURAS AGROALIMENTARIAS SA	**INTIA**	ES
10	AgriFood Lithuania DIH	**AFL**	LT
11	UNIVERSITA DEGLI STUDI DI FIRENZE	**UNIFI**	IT
12	UNIVERSITA DI PISA	**UNIPI**	IT
13	AG FUTURA TECHNOLOGII DOOEL SKOPJE	**AGFT**	MK
14	SCUOLA SUPERIORE DI STUDI UNIVERSITARI E DI PERFEZIONAMENTO S ANNA	**SSSA**	IT
15	HELVETAS Swiss Intercooperation	**HELVETAS**	CH

Executive Summary

VALERECO Data Management Plan (DMP) is part of Work Package (WP) 7 and will be based on FAIR (Findable, Accessible, Interoperable and Reusable) principles and the Open Science concept of the Horizon Europe (HE) programme. This DMP has been prepared in accordance with the: (i) Articles of the Grant Agreement (GA) of the VALERECO project (101135472), (ii) Article 16 for Open Science of the Horizon Europe programme guide[1], (iii) Data Management Plan Template of Horizon Europe[2]. This document is the first version of the Data Management Plan, delivered in Month 6 of the project, and will be updated twice in Month 24 (Deliverable D7.3) and Month 48 (Deliverable D7.4).

Partners of VALERECO will comply with all Articles of the VALERECO GA and as related to the Data Management Plan, with the following Articles:

- ARTICLE 14 — Ethics and values
- ARTICLE 15 – Data Protection
- ARTICLE 16 – Intellectual Property Rights (IPR) – Background and Results – Access Rights and Rights of Use
- ARTICLE 17 – Communication, Dissemination and Visibility
- Specific rules in Annex

[1]https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

[2] https://www.openaire.eu/images/Guides/HORIZON_EUROPE_Data-Management-Plan-Template.pdf

Contents

List of Tables

List of abbreviations

DMP	Data Management Plan
WP	Work Package
FAIR	Findable, Accessible, Interoperable and Reusable
HE	Horizon Europe
GA	Grant Agreement
IPR	Intellectual Property Rights
EC	European Commission
LL	Living Lab
GDPR	General Data Protection Regulation
DLIH	Digital Legume Information Hub
R&I	Research and Innovation
EOSC	European Open Science Cloud
DOI	Digital Object Identifiers
CAP	Common Agricultural Policy
EU	European Union
LCA	Life Cycle Assessment

1. Introduction

The Data Management Plan (DMP) serves as a comprehensive guide for how research data will be handled throughout the course of the project. It provides a detailed overview of how data will be collected, processed, or generated, ensuring clarity on the methodologies and standards to be followed. Additionally, the DMP outlines the plans for data sharing, specifying whether the data will be made open to the public or remain restricted, and the mechanisms by which this will occur. It also includes a strategy for the curation and long-term preservation of the data, both during the project's duration and beyond its completion. Importantly, the DMP is designed as a living document, evolving alongside the project to reflect new datasets, emerging methodologies, and updated policies, ensuring that data management remains adaptable and effective throughout.

2. Open access & open science

The European Commission (EC) is dedicated to Open Science[3]. Open science consists in the sharing of knowledge, data and tools as early as possible in the Research and Innovation (R&I) process, in open collaboration with all relevant knowledge actors, including academia, industry, public authorities, end users, citizens and society at large. Open science has the potential to increase the quality, efficiency and impact of R&I, lead to greater responsiveness to societal challenges, and increase trust of society in the science system. The Open Science policy of HE5 aims among others to:

• ensure that beneficiaries retain the intellectual property rights they need to comply with their open access obligations;

• require research data to be FAIR and open by default following the principle "as open as possible, as closed as necessary" (with exceptions notably for commercial purposes and to safeguard the privacy of the participants);

• promote the adoption of open science practices, from sharing research outputs as early and widely as possibly, to citizen science, and developing new indicators for evaluation research and rewarding researchers;

• engage and involve citizens, civil society organizations and end-users in co-design and co creation processes and promote responsible research and innovation.

[3] https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en#future-ofopen-science-under-horizon-europe

Obligation to disseminate results

According to Article 17 of the VALERECO GA, unless otherwise agreed with the granting authority, the beneficiaries must promote the action and its results by providing targeted information to multiple audiences (including the media and the public), in accordance with Annex 1 and in a strategic, coherent and effective manner. The beneficiaries must disseminate their results as soon as feasible, in a publicly available format, subject to any restrictions due to the protection of intellectual property, security rules or legitimate interests. A beneficiary that intends to disseminate its results must give at least 15 days advance notice to the other beneficiaries (unless agreed otherwise), together with sufficient information on the results it will disseminate. Any other beneficiary may object within (unless agreed otherwise) 15 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the results may not be disseminated unless appropriate steps are taken to safeguard those interests.

Obligation and right to use the EU emblem

According to Article 7.2 of VALERECO GA, unless otherwise agreed with the granting authority, communication activities of the beneficiaries related to the action (including media relations, conferences, seminars, information material, such as brochures, leaflets, posters, presentations, etc., in electronic form, via traditional or social media, etc.), dissemination activities and any infrastructure, equipment, vehicles, supplies or major result funded by the grant must acknowledge EU support and display the European flag (emblem) and funding statement (translated into local languages, where appropriate).

Moreover, it must indicate the following disclaimer (translated into local languages where appropriate): "*Funded by the European Union under Grant Agreement No. 101135472. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them.*"

Open access to scientific publications

According to the Specific rules in Annex 5 of the VALERECO GA and related to Article 17, The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:

- At the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications

- Immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and

- Information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

Metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication. Only publication fees in full open access venues for peer-reviewed scientific publications are eligible for reimbursement.

Open science: research data management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- Establish a data management plan ('DMP') (and regularly update it)
- As soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions, this repository must be federated in the EOSC in compliance with EOSC requirements
- As soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:
    - be against the beneficiary's legitimate interests, including regarding commercial exploitation,
    - be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP
- Provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data

Metadata of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs.

3. Data Summary

3.1. Types of Data

The majority of the data and research outputs of the VALERECO project will be generated in the context of WP1-WP5. The data majority will involve a knowledge database of societal/collective decision-making biases of farmers, drivers, challenges, barriers and opportunities to the adoption of legumes and the use of questionnaires and interviews (WP2). These results will be published as csv/tsv and pdf/txt files as sets of qualitative and quantitative data. Typically, experimental and technical data will be generated by WP3, WP4 and WP5, regarding the legumes tested in the LLs and will be accessible through the VALERECO platform to download in the form of csv/tsv and json files.
The following Tables highlights the: (1) **re-use of existing data** and **for what it will be re-used**, (2) **types and formats of data** that will be generated in the VALERECO project, (3) **purpose of data generation or re-use** and its **relation to the Objectives** of the project, (4) **expected size** of the data, (5) **origin/provenance** of the data, (6) **data utility** outside the project.

WP1 – A knowledge base on the ecosystem services provided by legumes		
Task 1.1: Legacy of the legumes and ecosystem services **Task Leader: UC**	**Re-use of existing data**	Yes
	Types and formats of data	Excel spreadsheet .xlsx and/or Word document .docx., and/or PDF files
	Purpose of data generation/re-use	provide a comprehensive assessment of the contribution of legume crops to ecosystem services
	Expected data size	<10 MB
	Origin/Provenance	Existing scientific literature, scientific and technical findings from previous international and national projects, existing databases
	Data utility outside the project	Academia, policy makers and farmers/advisors
Task 1.2: Legacy of the legumes and ecosystem services **Task Leader: IFVCNS**	**Re-use of existing data**	Yes
	Types and formats of data	Word document .docx., and or/PDF files
	Purpose of data generation/re-use	Preparation of deliverable
	Expected data size	<20 MB
	Origin/Provenance	Open access literature review
	Data utility outside the project	Academia
Task 1.3: Legumes in alignment with the new	**Re-use of existing data**	Yes
	Types and formats of data	Excel spreadsheet .xlsx and/or

CAP **Task Leader: WR**		Word document .docx., and/or PDF files
	Purpose of data generation/re-use	Preparation of deliverable
	Expected data size	<20 MB
	Origin/Provenance	Open access literature
	Data utility outside the project	Academia, policy makers and farmers/advisors
Task 1.4: Establishment of the Living Labs **Task Leader: AUA**	**Re-use of existing data**	N/A
	Types and formats of data	Excel spreadsheet .xlsx and/or Word document .docx., and/or PDF files
	Purpose of data generation/re-use	Implementation of LL and LL Boards
	Expected data size	<10 MB
	Origin/Provenance	Experimental layouts and measured parameters of each LL
	Data utility outside the project	N/A
Task 1.5: Synergies with other projects **Task Leader: DELPHY**	**Re-use of existing data**	Yes
	Types and formats of data	Excel spreadsheet .xlsx and/or Word document .docx., and/or PDF files
	Purpose of data generation/re-use	To connect the VALERECO ecosystem with other relevant EU-funded projects, thematic networks and upcoming initiatives
	Expected data size	<20 MB
	Origin/Provenance	public
	Data utility outside the project	yes

Table 1. WP1 Data Summary

WP2 – Behavioural changes in the value chain towards legumes		
Task 2.1: Status quo analysis: Perceptions and knowledge systems of supply chain actors and consumers. **Task Leader: LUH**	**Re-use of existing data**	N/A
	Types and formats of data	Excel spreadsheet .xlsx or .csv and/or Word document .docx., and/or PDF files .pdf
	Purpose of data generation/re-use	Understanding current perceptions, knowledge, attitudes and actions of supply chain actors
	Expected data size	unknown
	Origin/Provenance	Surveys and Interviews
	Data utility outside the project	N/A
Task 2.2: Market penetration and acceptance of leguminous crops **Task Leader: BSB**	**Re-use of existing data**	Yes
	Types and formats of data	Online database
	Purpose of data generation/re-use	Analysis of market-wide information on demand and industry dynamics to understand the importance of legume in the alimentary basket across EU countries. The data about legume consumption in the EU countries will be analysed in the light of the competition with other foods. At the same time, data on leguminous crop production in the EU countries will be analysed. Input will also feed the cost-benefit analysis in Task 4.3
	Expected data size	unknown
	Origin/Provenance	Secondary data from 1. Euromonitor, Statista, Eurostat, FAO, national statistical institutes to have a complete picture of demand and supply of legume in the EU; 2. Eurostat and UNComtrade to depict a detailed and updated global picture of trades in legumes
	Data utility outside the project	Academia; statistics and economics agencies
Task 2.3: Behavioural design strategies to study consumers attitudes and promote healthier and more sustainable diets. **Task Leader: BSB**	**Re-use of existing data**	N/A
	Types and formats of data	Excel spreadsheet .xlsx or .csv and/or Word document .docx., and/or SPSS files (.sav) and/or PDF files .pdf and/or STATA files (.dta)
	Purpose of data generation/re-use	Determine the ways of implementing successful nudges and the best domains for their application in order to propose the

		best regulatory and market environment for the promotion and enhancement of legume consumption.
	Expected data size	unknown
	Origin/Provenance	Online surveys via panel provider and laboratory experimentation
	Data utility outside the project	Academia
Task 2.4: Behavioural design strategies to leverage the adoption of legumes by farmers. **Task Leader: LUH**	**Re-use of existing data**	N/A
	Types and formats of data	Excel spreadsheet .xlsx or .csv and/or Word document .docx., and/or PDF files .pdf
	Purpose of data generation/re-use	Understanding farmers' decision-making, identifying communication strategies
	Expected data size	unknown
	Origin/Provenance	(Online) experiments and surveys
	Data utility outside the project	Anonymized data shared as open access resource

Table 2. WP2 Data Summary

WP3-On-station participatory trials to quantify multiple ecosystem services from legume crops		
Task 3.1: Set up of on-station participatory trials and establishment of the experimental protocols following 3D diversification strategies Task Leader: UNIPI	**Re-use of existing data**	N/A
	Types and formats of data	MS-Word and MS-Excel files documents with experimental/assessment protocols
	Purpose of data generation/re-use	Establish homogeneous protocols for the implementation of and data collection within field experiments
	Expected data size	Few Mb
	Origin/Provenance	Own contribution from task partners
	Data utility outside the project	Only as metadata accompanying bio-physical data that will be published
Task 3.2: Optimization of legume crops yield, stability and grain quality through agroecological approaches. Task Leader: UNIPI	**Re-use of existing data**	Maybe (e.g., existing soil data analyses).
	Types and formats of data	MS-Excel datasets containing all the yield-related data generated in each of the field experiments conducted by each partner of the task
	Purpose of data generation/re-use	Inputs for cross-partners data analyses and scientific publications/dissemination material, feeding the DSS calibration and validation in WP5, activities with LL actors in WP4, training and education activities in WP6
	Expected data size	<20 Mb
	Origin/Provenance	Template produced by UNIPI and periodically filled by each partner responsible person.
	Data utility outside the project	The dataset will be published on public repository once the results will be published on scientific journals to allow also other users analysing for different purposes the same dataset (e.g., for meta-analyses or modelling activities)
Task 3.3: Cascading effects of legume N2-fixation to soil fertility and	**Re-use of existing data**	Maybe (e.g., existing soil data analyses).

biodiversity **Task Leader: UNIPI**	**Types and formats of data**	MS-Excel datasets containing all the N2-fixation and soil fertility data generated in each of the field experiments conducted by each partner of the task
	Purpose of data generation/re-use	Inputs for cross-partners data analyses and scientific publications/dissemination material, feeding the DSS calibration and validation in WP5, activities with LL actors in WP4, training and education activities in WP6
	Expected data size	<20 Mb
	Origin/Provenance	Template produced by UNIPI and periodically filled by each partner responsible person.
	Data utility outside the project	The dataset will be published on public repository once the results will be published on scientific journals to allow also other users analysing for different purposes the same dataset (e.g., for meta-analyses or modelling activities)
Task 3.4: Legume contribution to the reduction of disservices from biodiversity aiming at pesticide-free agricultural systems. **Task Leader: SSSA**	**Re-use of existing data**	Review of pest weed and disease control by legume crop and collect unpublished data from project partners
	Types and formats of data	Excel files and word documents.
	Purpose of data generation/re-use	Collect evidence and measure if legumes reduce pest, weed and disease pressure in european cropping systems
	Expected data size	unknown
	Origin/Provenance	Scopus, project partners and project experiments
	Data utility outside the project	Accademia, policy makers and farmers/advisors
Task 3.5: Assessment of legume varieties for their climate change adaptation capability.	**Re-use of existing data**	Meta-analysis, review of literature data and comparison with experimental ones produced in the current project
	Types and formats of data	Fasta files, excel files with

Task Leader: UNIFI		normalized omics data
	Purpose of data generation/re-use	Sheds lights into gene and molecular regulatory networks underlying legume responses to drought stress
	Expected data size	> 1 Gb
	Origin/Provenance	Project experiments and published data in public repositories
	Data utility outside the project	Academia, breeders, farmers, advisors

Table 3. WP3 Data Summary

WP4-Living labs for the economic, environmental and technical assessment of ecosystem services		
Task 4.1: Living-Labs for demonstrating innovative uses of legumes. Task Leader: WR	Re-use of existing data	Yes
	Types and formats of data	Excel spreadsheet .xlsx and/or Word document .docx., and/or PDF files
	Purpose of data generation/re-use	To guide LL in demo's and learning with expertise from former projects
	Expected data size	<10MB
	Origin/Provenance	Open access literature
	Data utility outside the project	No
Task 4.2: Determination of the technical feasibility of legumes inclusion in crop rotations. Task Leader: SSSA	Re-use of existing data	No
	Types and formats of data	Excel files and word documents
	Purpose of data generation/re-use	To assess the technical viability of agroecological diversification strategies by exploring the advantages of incorporating legumes into crop rotations. The outcome of this task will be a set of recommendations for integrating legume crops into crop rotation systems.
	Expected data size	unknown
	Origin/Provenance	The data will be gathered through Task 1.2, which involves analyzing current knowledge on legumes to identify technical barriers and gaps for their inclusion in agroecological systems. Task 4.2 will also assess existing crop rotations across regions, with the selection of legume crops aligned with CAP eco-schemes, as outlined in Task 1.2.
	Data utility outside the project	Accademia, policy makers and farmers/advisors
Task 4.3: Cost-benefit analyses. Task Leader: DELPHY	Re-use of existing data	yes
	Types and formats of data	Excel spreadsheet .xlsx or .csv and/or Word document .docx., and/or PDF files .pdf
	Purpose of data generation/re-use	Focus on the assessment of the economic viability of LL

		solutions
	Expected data size	<100 MB
	Origin/Provenance	Project partners and (scientific) literature
	Data utility outside the project	Yes
Task 4.4: Life Cycle Assessments **Task Leader: WR**	**Re-use of existing data**	Yes
	Types and formats of data	Excel spreadsheet .xlsx and/or Word document .docx., and/or PDF files
	Purpose of data generation/re-use	Data collecection to run LCA, reuse of data collection formats
	Expected data size	<100MB
	Origin/Provenance	LL partners, project partners
	Data utility outside the project	Via deliverable

Table 4. WP4 Data Summary

WP5 – Development of tools and networks for the exploitation of legumes in agroecologically-based farming systems		
Task 5.1: Decision support system for farmers and communication tool for biodiversity-friendly legume cropping Systems **Task Leader: LUH**	**Re-use of existing data**	data from previous LUH research projects LEGU-MED and MANUELA, data from other Valereco WPs (most likely WP 1, WP2 and WP3)
	Types and formats of data	Excel spreadsheet .xlsx or .csv and/or Word document .docx., and/or PDF files .pdf
	Purpose of data generation/re-use	development of decision support system
	Expected data size	unknown
	Origin/Provenance	algorithms from Manuela, literature and field data (experimental layouts and measured parameters) from LEGU-MED, literature data from WP 1, behavioural strategies from WP2, field data (experimental layouts and measured parameters) from WP3
	Data utility outside the project	N/A
Task 5.2: Implementation of a digital legume information hub. **Task Leader: AUA**	**Re-use of existing data**	The task is expected to reuse existing data from various sources, including scientific findings from previous projects, existing databases like the European Grain Legume Dataset, and technical solutions available from EU-funded and national projects
	Types and formats of data	The types of data expected to be used in Task 5.2 include a knowledge database of societal collective decision-making biases of farmers, drivers, challenges, barriers, and opportunities related to legume adoption, as well as questionnaires and interviews data in formats such as csv, tsv, and pdf
	Purpose of data generation/re-use	The purpose of data generation/reuse is to develop a user-friendly digital hub that hosts information about the impact of legumes on ecosystem services and biodiversity, facilitating informed decision-making among stakeholders in sustainable agriculture and food production

	Expected data size	The expected data size may vary depending on the volume of information collected from literature reviews, primary data from field trials, and data from existing databases, but it is anticipated to be substantial to support the comprehensive digital legume information hub
	Origin/Provenance	The data originates from a combination of literature reviews, primary data collection from field trials conducted as part of the project, and existing databases developed in previous projects, ensuring a diverse and comprehensive dataset for the digital hub
	Data utility outside the project	The utility of this data outside the project extends to scientific and research organisations for further research on legume ecosystem services, the agri-food industry for market strategies based on research insights, farmers and agricultural advisers for informed decision-making, and policymakers for evidence-based policy formulation regarding nitrogen fertiliser use practices
Task 5.3: Formulation of policy recommendations at EU, national and regional level **Task Leader: WR**	**Re-use of existing data**	Yes
	Types and formats of data	Excel spreadsheet .xlsx and/or Word document .docx., and/or PDF files
	Purpose of data generation/re-use	Preparation of deliverable
	Expected data size	<20 MB
	Origin/Provenance	Open access literature, partners
	Data utility outside the project	Academia, policy makers and farmers/advisors

Table 5. WP5 Data Summary

WP6 – Ecosystem engagement, capacity building & sustainability		
Task 6.1: Dissemination, exploitation & communication plan **Task Leader: RFF**	**Re-use of existing data**	Multi-Actor Approach tools (adopted from the LIAISON H2020 project) **Data types:** Images and/or documents **Data formats:** .jpg, .jpeg, .png, .doc, .pdf
	Types and formats of data	**Data types:** Images and/or Videos and/or Presentations and/or Documents and/or Podcasts **Data formats:** *.jpg and/or *.jpeg and/or *.png and/or *.svg and/or *.mp4 and/or *.mov and/or *.ppt and/or *.pptx and/or *.doc and/or *.docx and/or *.rtf and/or *.pdf
	Purpose of data generation/re-use	Dissemination, exploitation & communication of the project, Stakeholders engagement, Project's sustainability
	Expected data size	0.5 – 25 MB for images, presentations and documents 500 MB – 2 GB for videos and podcasts files
	Origin/Provenance	Dissemination and communication materials prepared by Partners and RFF
	Data utility outside the project	-
Task 6.2: Boosting stakeholder & ecosystem engagement **Task Leader: RFF**	**Re-use of existing data**	Multi-Actor Approach tools (adopted from the LIAISON H2020 project) **Data types:** Images and/or documents **Data formats:** *.jpg and/or *.jpeg and/or *.png and/or *.doc and/or *.pdf
	Types and formats of data	**Data types:** Images and/or Videos and/or Presentations and/or Documents and/or Podcasts **Data formats:** *.jpg and/or *.jpeg and/or *.png and/or *.svg and/or *.mp4 and/or *.mov and/or *.ppt and/or *.pptx and/or *.doc and/or *.docx and/or *.rtf and/or *.pdf
	Purpose of data generation/re-use	Dissemination, exploitation & communication of the project, Stakeholders engagement, Project's sustainability
	Expected data size	0.5 – 25 MB for images/ presentations/documents

		500 MB – 2 GB for videos/podcasts files
	Origin/Provenance	Dissemination and communication material prepare by Partners and RFF
	Data utility outside the project	-
Task 6.3: Training & capacity building **Task Leader: AfL**	**Re-use of existing data**	Data from other Valereco WP's (WP1-WP5)
	Types and formats of data	Excel spreadsheet .xlsx or .csv and/or Word document .docx., and/or PDF files .pdf
	Purpose of data generation/re-use	Foundation for creating a comprehensive training program that will develop as reference material for capacity-building workshops and activities
	Expected data size	N/A
	Origin/Provenance	A knowledge base on legumes ecosystem services from WP1; value chain behaviours from WP2; multiple ecosystem services from WP3; economic, environmental and technical assessments from WP4; exploitation of legumes in agroecologically based farming systems from WP5.
	Data utility outside the project	Data sharred trough training platform for capacity building
Task 6.4: IPR management & sustainability **Task Leader: RFF**	**Re-use of existing data**	N/A
	Types and formats of data	**Data types:** Presentations and/or Documents and/or Spreadsheets **Data formats:** *.ppt and/or *.pptx and/or *.doc and/or *.docx and/or *.rtf and/or *.pdf and/or *.xls and/or *.xlsx
	Purpose of data generation/re-use	Exploitation of Project's results and outcomes, Project's sustainability
	Expected data size	0.5 – 25 MB (presentations/ documents/spreadsheets)
	Origin/Provenance	Materials prepared by RFF to manage IPR and ensure project's sustainability
	Data utility outside the project	-

Table 6. WP6 Data Summary

WP7 – Project management & coordination		
All Tasks of WP7 [7.1-7.2-7.3-7.4] Task Leader: AUA	**Re-use of existing data**	N/A
	Types and formats of data	Excel (xlsx, .csv), Word documents (docx.), and/or PFD files (.pdf.), Audio-visual material
	Purpose of data generation/re-use	To coordinate both the technical and financial aspects of VALERECO project
	Expected data size	Unknown
	Origin/Provenance	Minutes of meetings. Activities of VALERECO partners. Periodic reporting. Data related to ethics and IPR issues.
	Data utility outside the project	-

Table 7. WP7 Data Summary

4. FAIR Data

In compliance with Horizon Europe's Open Science policy, the data generated/re-used by VALERECO will be 'FAIR', that is Findable, Accessible, Interoperable and Re-usable. According to the consortium's policy, curated datasets and respective peer-reviewed scientific publications will be available both for the scientific community and the general public through the project's website and an online repository which will be created on the Zenodo platform. The project results will be published on the Open Research Europe platform and other open-access data repositories (e.g., Zenodo developed by OpenAIRE, EU-FarmBook), enabling the relevant knowledge actors to be aware of the latest scientific discoveries. These platforms include citations to all supporting data & materials, so the information can be reached, reanalyzed, replicated and reused. Dissemination materials will also be shared through the Organic E-prints repository.

Research output management: It is foreseen that project outcomes will be centralized and stored in common formats, committed to homogeneity, in order to facilitate the easiest access and exploitation possible. Mandatory technical standards will be adhered to ensure that scientific information, publications, data and other outputs, as well as the relevant metadata, are available for re-use in the long term. This will involve the use of persistent and unique identifiers, of certified repositories that are compliant with the standards of the European Open Science Cloud (EOSC) and in compliance with the FAIR principles for the management of research data produced by the project. Apart from the outputs that will be available through the EOSC, every practical and useful outcome of the VALERECO project will be stored in easy-to-read material in the project's on-line portal.

Each beneficiary will be responsible for the curation of their data, while the upload of this data to the VALERECO website will be the responsibility of the Communication Manager (RFF). The Data Management will be the responsibility of AUA.

4.1. Making data findable, including provisions for metadata

The VALERECO website (basic information site – www.valereco.eu; DLIH—to be launched later in the project lifetime) will be the main portals for finding the project's data. The project's curated data, respective metadata and peer-reviewed open-access scientific publications will be available on the

VALERECO website, to maximize findability of the project's results. Furthermore, the scientific publications will receive Digital Object Identifiers (DOI) unique identifiers upon publication and will be accompanied by keywords to optimize possibilities for re-use. Curated datasets will be also deposited on the Zenodo repository and thus receive unique identifiers. VALERECO will also publish its research outputs and data in FAIR and open-access data repositories (i.e., EOSC, Open Research Europe) ensuring that the results will be findable via the assignment of unique, persistent identifies to them and annotated with rich metadata. The VALERECO knowledge database will be made available in the EU-Farmbook.

Early and open sharing of research: All peer-reviewed scientific publications will be made in Open Access Journals and be available free of charge.

4.2. Making data accessible

VALERECO will conform to the FAIR data principles making research outputs and data open and freely accessible through the open-access repositories that have been considered for the publication of the VALERECO results, and by using more tools such as: printed and digital dissemination material, publications in open-access peer-reviewed journals, online repositories and platforms. Additionally, VALERECO will support the Open Access initiative of the EC by sharing datasets with interested parties on-demand and respecting privacy security-commercial-ethical related rules, intellectual property, and the General Data Protection Regulation (GDPR).
VALERECO will primarily share all data through its website. The website will contain curated datasets, scientific publications and links to the online repository/databases. Furthermore, the project's website will present the project's major achievements in a language easily understandable by agricultural practitioners and the general public, highlighting the main benefits for the farmers, consumers, and other relevant stakeholders. What is more, the consortium will make sure to upload on the project website the produced data in formats that will be easily downloaded and read by the majority of the relevant stakeholders (e.g., .pdf, .docx, .xlsx). Most of the primary and curated datasets will be made openly available after their publication on the open-access online repository, while scientific publications will also be accessible through open-access platforms (in accordance with the rules set out in the Open Access and Open Science section of the present DMP). The project will also promote technology transfer through workshops, conferences, webinars, agricultural fairs, etc. As the project progresses and more data is generated, further information will be outlined in subsequent versions of the DMP.

If certain datasets cannot be shared during and after the end of the project, it will be described in the updated versions of this deliverable (M24, M48).

4.3. Making data interoperable

VALERECO is anticipated to generate a considerable amount of primary data through the interdisciplinary training of experienced researchers. The solid scientific background of the researchers involved in the project will ensure that correct standards and methodologies are followed for the collection of VALERECO results, therefore facilitating their interoperability. More specifically, special emphasis will be given to the usage of standard data and metadata vocabularies, nomenclature, and terminology, to the usage of open software applications where possible, and to the usage of data formats that can be recombined with data of other origin. Furthermore, the consortium will make sure to upload on the project website the produced data in formats that will be easily downloaded and read by the majority of the relevant stakeholders (e.g., .pdf, .docx, .xlsx). The VALERECO project will implement co-creation with society decision makers through a Living Lab

approach to ensure the co-design, testing and validation of solutions, ensuring that they can shape innovation cycles towards the generation of outputs that will have a higher likelihood of being successfully adopted. The methodologies that will be followed during the project's activity will be open, free and available for the public, for the purpose of reproducibility. The data models of the open repositories that have already been considered have been designed based on the re-use of existing, widely adopted ontologies and/or controlled vocabularies. The adoption of commonly employed file formats (e.g., .csv/.tsv and .json files) ensures that any researcher or non-research entity will be able to access and process the VALERECO data by not needing to rely upon the use of proprietary software. If data from other projects or previous surveys are used, then specific reference will be made to them, as well as to the source(s) and owner(s) of the data. Data uploaded by researchers will be restricted to use by the researcher, unless the researcher expressly provides permission for the data to be available to other researchers using DLIH. If data is to be shared, then the user will be required to provide the data in a format that adheres to an appropriate data standard that will be developed in WP5 to ensure consistency of data across the platform.

4.4. Increase data re-use

In addition to ensuring the accessibility of the data, the intention is to facilitate its reuse. Research outputs and datasets with rich provenance details provided in the form of appropriate metadata values, will be published in open repositories and based on the application profiles of them to ensure reusability. Appropriate defined licensing types (e.g., the Creative Commons framework and specific license types available from it, such as CC-BY) will be used in publicly available research outputs and datasets to make explicit concrete (re-)use cases.

The involved partners will provide a description of data provenance and data quality assurance processes and will specify the length of time for which the data will remain re- usable. Readme files with information on methodology, codebooks, data cleaning, analyses, units of measurements will be made available if needed to validate data analysis and facilitate data re-use. There is no limit for how long the DLIH will remain re-usable. Re use of any personal data collected in the project will not be possible, as collection and consent for use of this data will be done on the understanding that it will only be used within the VALERECO project.

5. Allocation of resources

In the interest of ensuring the optimal management, preservation, and dissemination of data, dedicated resources will be allocated for these activities throughout the entirety of the VALERECO project's lifespan. Costs related to open access scientific publications are eligible for reimbursement under the conditions defined in the GA and have been already foreseen in the budget of the relevant VALERECO partners. The handling of the online repository and data management issues fall in the responsibility of the coordinator (AUA), the developer and host of the websites (RFF), the communication manager (RFF) and the responsible for the DLIH of the deliverables D5.2 and D5.3 (AUA). The curation of the VALERECO database of solutions will be delivered as a dedicated database on the VALERECO platform and will be financed by the AUA's budget during the course of the project. In the long-term, the hosting of the database will be made available through the EU-wide interactive knowledge reservoir (EU-FarmBook). The funding of the project will cover any preservation costs for this EU wide reservoir. Open repositories will be exploited to keep publication costs of data and outputs at a minimum.

Curation and hosting of all data assets (images, annotations, metadata): VALERECO's generated results, will be centralised and stored in common formats, committed to homogeneity, facilitating easy access and exploitation in VALERECO platform. At the same time, VALERECO partners aim to make all research publications Open Access. Partners involved in research taking place during the project, are committed to participate in open peer-review papers. Project partners are from the proposal phase of the project committed to the early and open science policy of the EU. This approach to the scientific process focuses on spreading knowledge as soon as it is available using digital and collaborative technology, expert groups, publications, news and events. VALERECO will promote the adoption of open science practices, sharing research outputs as early and widely as possible and to make sure Open Science principles are applied, FAIR metadata will be added to all data collected.

6. Data security

This section outlines the data security protocols adopted by the VALERECO project, which are designed to ensure that all personal data is handled in a secure, legally compliant, and ethical manner. The project follows the General Data Protection Regulation (GDPR) and relevant national and international standards. Robust measures will be implemented throughout the project's lifecycle to safeguard personal data from unauthorised access, breaches and loss. To this end, regular backups will be scheduled to ensure that no critical information is lost. Data sharing protocols will ensure that personal data is transmitted securely within the consortium and to external stakeholders. Confidential data may relate to intellectual property protection of new processes. In the event of a personal data security breach, the responsible partner will notify the project Data Manager (AUA), the data subject and the competent supervisory authority, without undue delay and, where feasible, not later than 72 hours, providing a detailed account of the security breach, its effects, and the remedial action(s) taken, according to article 33 and 34 of the General Data Protection Regulation.

7. Ethics

7.1. Work plan and resources

A limited set of personal data will be collected during the project execution, mainly the personal data of consortium members. Besides the partners, limited personal data of stakeholders and actors (farmers, advisors, researchers, consumers, retailers, organizations, NGOs, EU CAP Network & EU level policy makers, press, industry actors, other actors in the agri-food chain) involved in the project's activities (surveys, interviews, questionnaires, demonstration activities, National workshops and cross-visits) will be gathered (only the data that is relevant and limited to the purposes of the research project in accordance with the data minimization principle).
The requirements concerning Privacy, data protection, data management are part of WP7, Project Management, concerning the establishment of the Data Management Plan (DMP) with the description of the technical and organizational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants will be delivered. These requirements will be taken into account by WP1, WP2, WP3, WP4, WP5, and WP6.

The requirements as:

- details on the procedure and criteria that will be used to identify participants
- detailed information on the informed consent procedures that will we implemented for the participation of humans and in regard to data processing

7.2. Compliance with ethical principles and relevant legislations

Involvement of stakeholders and the linked feedback and participatory processes are very important within this proposal. This includes a lot of aspects where ethical and gender issues have to be checked (e.g. the selection of stakeholders, their consent for participation in the workshops, data privacy and protection). In line with the ethics self-assessment, actions will be executed for involvement of humans and for the protection of personal data. VALERECO will produce and process only non-sensitive data [Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (GDPR)], and therefore no significant ethical aspects are raised. The project Data Managers (AUA for the project & RFF for the website) and the Project Coordination Team will have the responsibility of any ethical and gender issues. Furthermore, the VALERECO website will include a Data Policy, Privacy Statement, and a GDPR-compliant cookies bar, informing website visitors about how their personal data will be used, their rights and options to manage preferences. More information on ethical aspects in relation to data management can be found in Articles 14 and 15 of the VALERECO GA. Detailed info on VALERECO's Ethics self-assessment and plan can be found on pages 150-151 of the VALERECO GA.

The consortium will not discriminate against persons of any race, religion or gender. The partners are fully aware of the European Charter for Researchers and Code of Conduct for the Recruitment of Researchers and will pay specific attention to equal opportunities and gender balance in the project team. Gender balance will also be respected in the selection of stakeholders.

Data used in the project will be the subject of a paper analyzing the potential conflict with respect to privacy. When processing personal data, the consortium will comply with the GDPR principles. This means that at the outset, when the data cannot be anonymized fully, in the development of the project, Article 5(1) of the GDPR, enshrining the principles of personal data protection, will be implemented . Before the beginning of an activity raising an ethical issue, the beneficiary will confirm that any ethics committee opinion required under national law has been obtained and is kept on file.

The Consortium confirms that compliance with ethical principles and applicable international, EU and national law in the implementation of research activities not originally envisaged (or not described in detail) in the DoA will be ensured. The Consortium also confirms that any ethical concerns raised by those activities will be handled following rigorously the recommendations provided in the European Commission Ethics Self-Assessment Guidelines.

7.3. Informed consents

A special effort will be made to collect as little restricted and personal data as possible. The personal data involved will be the identification, profession and the opinions and answers given by the participants. All data will be handled only by qualified researchers under strict confidentiality agreements, who will ensure that data access, data protection and privacy standards are in compliance with national and EU regulations. All participants will be volunteers, recruited through in-country VALERECO partners as representative of end and potential end users of the DLIH. All the stakeholders included in the activities of the Living Labs, Living Lab boards, workshops and dissemination-communication-demonstration activities will sign an informed consent (either with digital or handwritten signature, and/or digital or oral consent to be recorded in the case of online questionnaires and recordings, respectively) in which they will be duly informed about: (1) how their personal data will be processed and protected, (2) how the data collected will be used for research purposes in executing the Research and Innovation (R&I) activities of VALERECO and developing the

DLIH, (3) the VALERECO data privacy policy, (4) their rights and responsibilities, including the right to access, rectify erase their data, and to withdraw consent at any time, (5) the period for which the personal data will be stored, and (6) contact details of the controller, data protection officer (DPO) and of the Ethics Mentor. Should any personal data be obtained during the project, the project will see to it that it be made anonymous and rigorously protected for the duration of the action and destroyed at the Conclusion. Data of the attendees and confirmation of informed consent will be kept for a reasonable period after the completion of the project and in no case will exceed five years. To take part in VALERECO activities, attendees must be over 18, have no learning difficulties or an impaired capacity to consent, and will not be 'vulnerable', there will also be no clinical intervention; no further restrictions will be in place. Anonymising the data: All data related to the questionnaires and interviews are anonymized. VALERECO partners are removing personally identifiable information (information that directly or indirectly relates to an identified or identifiable person) from datasets. Personal data will be collected only in the case of: (1) questionnaires, surveys and interviews (WP2), (2) Living Lab boards. For the first category, the stakeholders participating in online questionnaires receive a personal code number that they must enter to the questionnaire which will thereafter be used in handling and typing the data without further possibility of linking the code to the respondent's identity and contact data (name, telephone number or e-mail address). The data collected in this way will be then typed into an electronic format (excel) without any access to identifying information. Only the Living Lab has access to the personal data and the associated personal code number of the respondents to the questionnaires and interviews and is responsible for protecting them. No exchange/sharing of personal and sensitive data is allowed between VALERECO partners, with third parties and any stakeholder.

Informed consent is requested before personal data is collected and stored. Informed consent forms are collected at the start of all VALERECO activities, where appropriate (workshops, surveys, interviews, questionnaires, dissemination-communication-demonstration activities, and cross visits). The consents are made available to all VALERECO partners in the relevant WPs, Tasks and Deliverables, with analytic guidelines for stakeholders' identification and engagement, recruitment criteria for research participants, informed consent, and data collection procedures.

All templates of the informed consent forms and information sheets covering the voluntary participation and data protection issue (in language and terms intelligible to the participants) will be kept on file; will be provided by WP7.

END OF DOCUMENT