



AGROECOLOGY PARTNERSHIP



DELIVERABLE 5.1

Data Management Plan

Reproduction is authorised provided the source is acknowledged.

© AGROECOLOGY Partnership, 2024



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Technical References

| | |
|----------------------------------|----------------|
| Deliverable No. | 5.1 |
| Dissemination level ¹ | Public |
| Work package | 5 |
| Lead beneficiary | LifeWatch ERIC |
| Version | 1.0 |
| Due date of deliverable | June 2024 |
| Actual submission date | June 2024 |

Versions

| Version | Person | Partner | Date |
|---------|---------------|---------------|------------|
| 1.0 | José M. Avila | Final version | 27.06.2024 |

Approved by the Coordinator on: 28.06.2024

Disclaimer. Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.



TABLE OF CONTENTS

| | |
|---|----|
| Summary..... | 4 |
| Definitions..... | 5 |
| Acronyms..... | 6 |
| Data overview table..... | 6 |
| 1 Purpose of the data management plan..... | 7 |
| 1.1 Supporting policy..... | 9 |
| 1.1.1 European Commission - open science policy..... | 9 |
| 1.1.2 Horizon Europe - open science..... | 10 |
| 1.1.3 GDPR..... | 12 |
| 1.1.4 Open Data Directive..... | 13 |
| 1.2 DMP change management..... | 14 |
| 2 Data Summary..... | 16 |
| 2.1 Data overview..... | 16 |
| 2.2 Data types and formats..... | 17 |
| 2.3 Data reuse..... | 18 |
| 2.4 Data utility..... | 18 |
| 2.5 Co-funded research projects..... | 19 |
| 3 FAIR data..... | 20 |
| 3.1 Making data findable..... | 20 |
| 3.2 Making data accessible..... | 21 |
| 3.3 Making data interoperable..... | 22 |
| 3.4 Making data reusable..... | 22 |
| 4 Other research outputs..... | 23 |
| 5 AGROECOLOGY Data Infrastructure..... | 23 |
| 6 Allocation of resources..... | 24 |
| 7 Data security..... | 25 |
| 8 Data ethics..... | 26 |
| 9 Conclusions..... | 29 |
| 10 References..... | 30 |
| 11 Annex A – Metadata file template for data assets and other research outputs..... | 31 |

Summary

This deliverable defines the data management plan (DMP) elaborated for the European partnership AGROECOLOGY. The document outlines strategies and procedures for managing all types of research data assets generated and reused throughout the partnership's lifecycle, ensuring adherence to Horizon Europe guidelines. The document clarifies all administrative and technical concerns related to data involved in the whole project lifecycle. It emphasizes data quality, accessibility, and long-term preservation, covering data collection, storage, sharing, and preservation practices while addressing legal, ethical, and security considerations. This initial DMP aims to provide comprehensive guidelines for preserving data and other research outputs collected or generated during the partnership's implementation by consortium members and provide a strategy to promote the adoption of these general guidelines for the transnational research projects funded by the partnership. Effective data management practices, including accurate data recording, metadata documentation, and secure storage and dissemination of results, will be performed to ensure safe and reliable data preservation.

Given the diverse range of data expected over the partnership duration, the DMP incorporates various strategies for responsible data management and is regularly revised to accommodate newly generated data throughout the partnership's lifecycle. This first version provides information about the data management strategy and includes a first identification of the research data items and other research outputs. This process will last the partnership's duration in order to ensure that all types of research data are sufficiently addressed. After the initial gathering of information, all data items will be analysed, and the initial specific procedure to ensure data are properly managed will be improved. New versions of the DMP document will be available periodically.

Definitions

| Terms | Definition |
|-----------------------------------|---|
| Co-funded research project | AGROECOLOGY will conduct calls for transnational research projects in the agroecology thematic remit. The calls will address relevant research performers in agroecology in countries providing funding and others under certain conditions, taking into account national regulations. |
| Data Management Plan | The document outlines from the start of a project the main aspects of the lifecycle of research outputs, notably including data. This includes their provenance, organization and curation, as well as adequate provisions for their access, preservation, sharing, and eventual deletion, both during and after a project. |
| Digital data repository | A digital data repository is an established and reliable online platform or infrastructure that stores, manages, and provides access to research data in a secure, long term and sustainable manner. Ideally, such repositories adhere to recognized standards and best practices for data management, ensuring the stored data's integrity, authenticity, and long-term preservation. Trusted repositories typically provide features such as persistent identifiers, version control, metadata standards, access control, and data curation services. They play a crucial role in facilitating data sharing, reuse, and reproducibility, thereby advancing research, innovation, and collaboration within the scientific community. |
| FAIRness | When data meet principles of findability, accessibility, interoperability, and reusability. |
| Metadata | Descriptive information accompanying the data assets, including details on data collection methods, equipment used, sampling protocols, and any associated contextual information. |
| Open Science | An approach based on open cooperative work and systematic sharing of knowledge and tools as early and widely as possible in the research process. |
| Personal data | Information relating to an identified or identifiable natural person. An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (Article 2(a) EU General Data Protection Regulation 2016/679 (GDPR)). |
| Research data asset | Research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion or calculation. |
| Research data management | The process within the research lifecycle includes the organisation, storage, preservation, security, quality assurance, allocation of persistent identifiers (PIDs) and rules and procedures for sharing of data, including licensing. |
| Research output | Results to which access can be given in the form of scientific publications, data or other engineered results and processes such as software, algorithms, protocols, models, workflows |

Acronyms

| Acronyms | Description |
|-------------|---|
| DMP | Data Management Plan |
| DPO | Data Protection Officer |
| DOI | Digital Object Identifier. This is a unique identifier managed by the International DOI Foundation (IDF), a non-profit membership organization responsible for governing and managing the federation of Registration Agencies that offer DOI services and registration. The IDF serves as the registration authority for the ISO standard (ISO 26324) governing the DOI system. |
| EC | European Commission |
| EOSC | European Open Science Cloud |
| FAIR | Findable, Accessible, Interoperable, Reusable (FAIR). This acronym delineates a collection of characteristics that scientific data assets should possess to facilitate reproducible research and maintain relevance within the research community. |
| GDPR | General Data Protection Regulation. This is the EU law that establishes the legal framework for personal data management |
| IPR | Intellectual Property Right |
| IT | Information Technology |
| PIDs | Persistent Identifiers |
| RDM | Research Data Management |
| RDF | Resource Description Framework |
| SRIA | Strategic Research and Innovation Agenda |
| WP | Work Package |

Data overview table

| | |
|-------------------|-------------------|
| Data set 1 | Short description |
| Data set 2 | Short description |
| Data set 3 | Short description |

No data sets were shown as per M6. The above table serves as a placeholder for future versions.

1 Purpose of the data management plan

AGROECOLOGY, committed to open science principles, will generate important research findings, materials and data of different natures that will be jointly managed to ensure their openness, integrity, and reproducibility, storage, availability, display, and efficiency for queries. Some of the open science practices that AGROECOLOGY will implement are making study materials freely accessible (e.g., data, measures, experimental protocols, and analysis files), pre-registering study designs (e.g., registering a study and analysis plan prior to data collection, which will be integrated into AGROECOLOGY website), and offering open access to journal content. WP5 will implement an AGROECOLOGY data management plan (DMP) including open science principles, which is considered a key for consistent research data handling that complies with the EC guidelines for FAIR (Findable, Accessible, Interoperable and Reusable) data management to improve the openness, integrity, and reproducibility of research by preventing research misconduct or reducing questionable research and/or reporting practices [1-3]. While it should be noted that co-funded research projects resulting from the transnational calls in WP6 are not property of the partnership *per se*, the majority of funders joining AGROECOLOGY and subject to funding such projects actively promote open science practices, e.g. making eligible publication costs in open access journals. The DMP will ensure research data are FAIR to ensure that knowledge is integrated and available for re-use in future research.

The AGROECOLOGY DMP plays a crucial role in the research process, aiming to establish best practices for managing research outputs, notably including data, throughout the partnership lifecycle. The DMP covers various aspects such as data provenance, organization, curation, access, preservation, sharing, and eventual deletion. It is not just a part of the research methodology but a cornerstone that contributes to efficiency, time savings, information security, and data impact. Research data management involves multiple stakeholders, including data creators, database designers, data managers, project managers, and information technology (IT) staff. This interdisciplinary field ensures effective handling of research data, maximizing its value and impact. The document has been prepared based on previous experiences of LifeWatch ERIC in projects of related initiatives [4-9].

The DMP is an adaptable document that evolves alongside the partnership. This initial version of the DMP will be continually refined and updated throughout the partnership's duration; it is foreseen to be updated at least by Month 22, 40, 64 and 82 to ensure its relevance over time. Revisions will be driven by an iterative process that will consider the identification of research data, its analysis, and the revision of the research data management strategy. This iterative process will be driven by partnership milestones, feedback, consortium reviews, external and internal insights (including the Ethics Advisory Board), and advancements in technology and best practices. To facilitate these updates, we have

implemented a comprehensive change management framework outlined in the next section of the present document.

The research data generated in the context of the AGROECOLOGY partnership will be published in domain-specific or domain-generic trusted digital data repositories, employing standardized data formats and vocabularies, depending on the data type. A data infrastructure is envisaged for specific datasets needed by some research-intensive work packages (WP), such as WP5. This infrastructure will be implemented following open-source principles and a cost-effective approach. This database will also be aligned with the FAIR principles for the preservation of information and will allow the identification of information owners as well as its traceability. It will utilize a standardized metadata schema, guarantee persistent identifiers, offer open access to metadata (either through public or restricted access policies) and enable machine-readable metadata.

While the AGROECOLOGY partnership will not directly manage all research data, this document sets forth guidelines and best practices on how research projects, funded by the partnership should generate, acquire, handle, share and curate data during all activities carried out within and beyond the AGROECOLOGY partnership.

In general terms, the AGROECOLOGY data will first be kept in the local repository of the consortium partners or co-funded project beneficiary, or shared in the AGROECOLOGY Microsoft Sharepoint enabled for this purpose for the partnership members (See section 7 Data security). The AGROECOLOGY policy for data publication follows the H2020 Guidelines to Open Access [3]. It is the will and commitment of the partners to share non-commercially sensitive knowledge and experience with the research community to ensure learning is transferred and errors are not repeated. Research data will be deposited and published on trusted repositories, such as Zenodo and linked to the AGROECOLOGY website¹ during the lifetime of the partnership and then classified as external. Research outputs will also be linked to the AGROECOLOGY platform components developed in WPs 5, 7 and 8. Figure 1 depicts the data workflow in AGROECOLOGY, which will be subsequently described in detail.

The DMP will be publicly accessible to ensure that all consortium members and partners funded by the open calls are informed about the practices to be adhered to during these activities. Research project applicants, together with the funders, are required to review this document and adhere to the policies outlined herein when planning their data management activities.

¹ <https://www.agroecologypartnership.eu/>

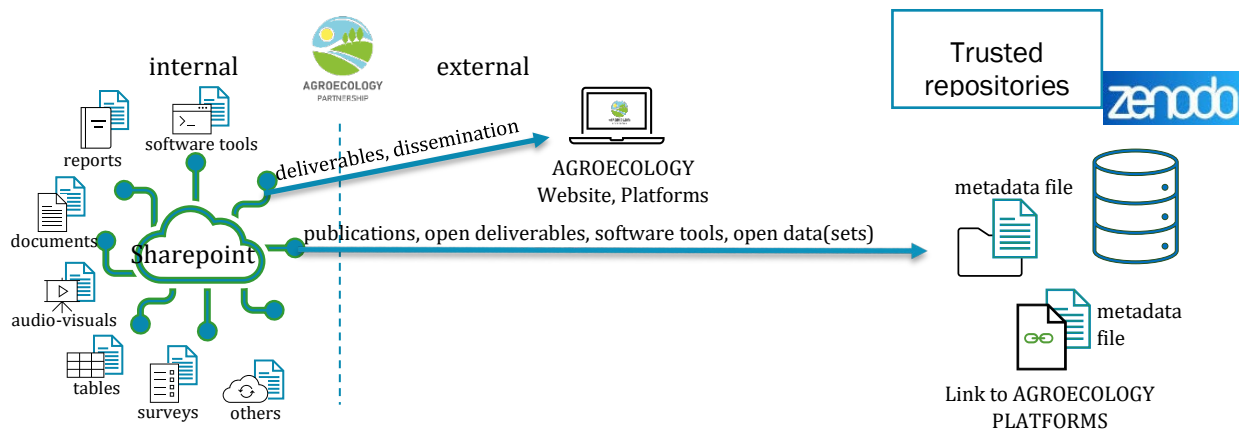


Figure 1: Data workflow for AGROECOLOGY.

1.1 Supporting policy

The present deliverable, D5.1 – Data Management Plan, is a comprehensive guideline adhering to European and national regulations on the acquisition, handling, processing, and archiving data generated throughout the AGROECOLOGY partnership and beyond its completion. This chapter reviews relevant European policy concerning open access and data management. This helps to better appreciate the context regarding the development of the AGROECOLOGY DMP.

1.1.1 European Commission - open science policy

The 2018 Recommendation of the European Commission on access to and preservation of scientific information (2018/790/EU) [2], which updates and replaces the 2012 Recommendation (2012/417/EU) [1], sets forth several policy recommendations to support open access to research outcomes. These policies aim to provide researchers and the public with free, open, and non-discriminatory access to peer-reviewed scientific publications, research data, and other research outputs as early as possible in the dissemination process. They also facilitate the use and re-use of scientific research results. Open access contributes to enhancing research quality, reducing unnecessary duplication, accelerating scientific progress, combating scientific fraud, and fostering economic growth and innovation. Licensing decisions should be made to promote the dissemination and re-use of scientific publications. Helping researchers manage research outputs effectively is becoming a standard practice in the scientific community [2]. The 2018 recommendations are divided into several measures, as outlined in Table 1 [2], which provides context and guidance for AGROECOLOGY in developing and implementing the DMP.

Table 1 EUROPEAN COMMISSION'S OPEN SCIENCE POLICY MEASURES

| Measure | Summary |
|---|---|
| Open access to scientific publications | Policies for the dissemination of and open access to scientific publications resulting from publicly funded research. |
| Management of research data, including open-access | Policies for the management of research data resulting from publicly funded research, including open access. |
| Preservation and re-use of scientific information | Policies for reinforcing the preservation and re-use of scientific information (publications, datasets and other research outputs). |
| Infrastructures for open science | Policies for further developing infrastructures underpinning the system for access to, preservation, sharing and re-use of scientific information and for promoting their federation within the European Open Science Cloud (EOSC). |
| Skills and competences | Policies for the necessary skills and competencies of researchers and personnel of academic institutions regarding scientific information. |
| Incentives and rewards | Policies for adjusting, with regards to scientific information, the recruitment and career evaluation system for researchers, the evaluation system for awarding research grants to researchers, and the evaluation systems for research performing institutions. |
| Multi-stakeholder dialogue on open science at national, European and international level | Multi-stakeholder dialogues on the transition towards open science at national, European and international level. |
| Structured coordination | National point of reference expert group to support coordination among EU members. |

1.1.2 Horizon Europe - open science

The European Commission's open science policy is embedded within the Horizon Europe 2021–2027 funding program for research and innovation. According to the Horizon Europe Programme Guide, open science involves the collaborative and systematic sharing of knowledge and tools at the earliest possible stages to reach the widest audience. This approach aims to improve research quality and efficiency, accelerate the advancement of knowledge and innovation through shared results, enhance reusability, and improve reproducibility. It necessitates the involvement of all relevant stakeholders. Horizon Europe stipulates both mandatory and recommended open science practices, which may differ based on specific work programs or call conditions. The typical mandatory open science practices are outlined in the following table:

Table 2 Mandatory open science practices (Horizon Europe)

| Open science practices | |
|------------------------|---|
| 1 | Open access to scientific publications under the conditions required by the grant agreement. |
| 2 | Responsible management of research data in line with the FAIR principles of findability, accessibility, interoperability and reusability, notably through the generalized use of data management plans, and open access to research data under the principle “as open as possible, as closed as necessary”, under the conditions required by the grant agreement. |
| 3 | Information about the research outputs/tools/instruments needed to validate the conclusions of scientific publications or to validate/re-use research data. |
| 4 | Digital or physical access to the results is needed to validate the conclusions of scientific publications, unless exceptions apply. |
| 5 | In cases of public emergency, if requested by the granting authority, immediate open access to all research outputs under open licenses or, if exceptions apply, access under fair and reasonable conditions to legal entities that need the research outputs to address the public emergency. |

Recommended open science practices go beyond the mandatory ones and include activities such as engaging all relevant knowledge actors, including citizens, early and open sharing of research, managing outputs beyond research data, and implementing open peer review. The Horizon Europe Programme Guide [3] defines research data management (RDM) as a series of activities within the research lifecycle. These activities encompass data collection or acquisition, organization, curation, storage, long-term preservation, security, quality assurance, assignment of persistent identifiers (PIDs), provision of metadata to meet disciplinary standards, licensing, and rules and procedures for data sharing. RDM is crucial for any project that generates, collects, or reuses data, requiring measures to ensure responsible management, such as selecting appropriate repositories, ensuring adequate access provisions, and complying with legal regulations like the General Data Protection Regulation (GDPR). Additionally, data management must adhere to the FAIR principles, enabling researchers to easily find, access, interoperate, and reuse each other’s data, thereby improving research effectiveness and reproducibility.

Data management plans (DMPs) are essential for responsibly managing research outputs. They help researchers manage a wide range of research outputs in line with the FAIR principles. These outputs may be physical or digital, including original software created during the partnership, workflows, protocols, and new materials such as samples and cell lines, among others. A DMP should be a living document that is updated as the partnership evolves. A comprehensive DMP should include:

Table 3 Data management plan aspects (Horizon Europe)

| DMP aspect | Summary |
|--|--|
| Dataset description | Detailed description of the data generated or re-used. |
| Standards and metadata | The protocols and standards used to structure the data. |
| Name and persistent identifier for the datasets | A unique and persistent identification (an identifier) of the datasets and a stable, resolvable link to where the datasets can be directly accessed. |
| Infrastructures for open science | Policies for further developing infrastructures underpinning the system for access to, preservation, sharing and re-use of scientific information and for promoting their federation within the European Open Science Cloud (EOSC). |
| Curation and preservation methodology | Information on the standards that will be used to ensure the integrity of the datasets and the period during which they will be maintained, as well as how they will be preserved and kept accessible in the longer term. |
| Data sharing methodology | Information on how the datasets can be accessed, including the terms-of-use or the license under which they can be accessed and re-used, and information on any restrictions that may apply or relevant security and privacy considerations. |

These Horizon Europe open access and data management policies will be implemented in the AGROECOLOGY partnership.

1.1.3 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) is a European Union law requiring organizations to safeguard personal data and uphold the privacy rights of anyone in the EU. Personal data is defined as any information that relates to an individual who can be directly or indirectly identified (e.g. names, email addresses, location, ethnicity, gender, etc.). The regulation includes seven principles of protection that must be implemented and a series of privacy rights that must be facilitated. The seven data protection principles are [10]:

Table 4 GDPR - data protection principles

| Data protection principle | Summary |
|--|---|
| Lawfulness, fairness and transparency | Processing must be lawful, fair, and transparent to the data subject. |
| Purpose limitation | You must process data for the legitimate purposes specified explicitly by the data subject when you collect it. |
| Data minimization | You should collect and process only as much data as absolutely necessary for the purposes specified. |
| Accuracy | You must keep personal data accurate and up-to-date |

| Data protection principle | Summary |
|--------------------------------------|---|
| Storage limitation | You may only store personally identifying data for as long as necessary for the specified purpose. |
| Integrity and confidentiality | Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption). |
| Accountability | The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles. |

GDPR also recognizes a series of privacy rights, including:

- The right to be informed
- The right to access
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights in relation to automated decision-making and profiling

The GDPR regulation must apply to any personal data managed by AGROECOLOGY (See section 8 Data ethics).

1.1.4 Open Data Directive

The European Union Directive on open data and the re-use of public sector information provides common rules for government-held data at national, regional and local levels. It aims to overcome barriers to fully exploit the potential of public sector information for the European economy and society. The implementation of the Open Data Directive is ongoing and requires the adoption by the European Commission via a future implementing act, with a list of high-value datasets taking into account guidelines on recommended standard licenses and datasets and charging for the reuse of information resources. To achieve maximum impact and to facilitate re-use, the high-value datasets should be made available for reuse with minimal legal restrictions, free of charge, and published via Application Programming Interfaces (APIs). However, the directive does not prevent public sector bodies from charging no more than marginal costs (e.g. reproduction costs) for the reuse of their data. However, particular charging exceptions are allowed for public sector bodies that are required to generate revenue to cover a substantial part of their costs; libraries, including university libraries, museums and archives; and public undertakings. Focus areas include [10]:

Table 5 Focus areas of the Open Data Directive

| Open data focus areas |
|---|
| Provision of real-time access to dynamic data |
| Increasing the supply of public data for re-use, including public undertakings, research performing organizations and research funding organizations |
| Tackling the emergence of new forms of exclusive arrangements and the use of exceptions to the principle of charging at marginal cost |
| Outline the relationship between this and other directives, such as: GDPR: align with GDPR personal data protection measures Database: public bodies should not be exercising data ownership protection measures to prevent data re-use |

In line with the European Commission’s open science policy, the Open Data Directive describes obligations regarding the re-use of publicly funded research data:

“The Commission Recommendation of 25 April 2018 on access to and preservation of scientific information describes, among other things, relevant elements of open access policies. Additionally, the conditions, under which certain research data can be reused, should be improved. For that reason, certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidised by public funding or co-funded by public and private-sector entities. Under the national open access policies, publicly funded research data should be made open as the default option. However, in this context, concerns in relation to privacy, protection of personal data, confidentiality, national security, legitimate commercial interests, such as trade secrets, and to intellectual property rights of third parties should be duly taken into account, according to the principle ‘as open as possible, as closed as necessary’.”

1.2 DMP change management

AGROECOLOGY recognizes the potential need to adjust its data management policies and IT requirements over the duration of the partnership. This is due to the dynamic nature of regulations and best practices, as well as the rapidly evolving IT service environments. To facilitate effective change management, Task 5.1 Develop and Implement a data management and monitoring plan will oversee revision processes of the DMP over the course of the partnership at least in M22, M40, M64 and M82. The revision of the DMP is discussed within the Executive Team, with the representation of all WP Leaders. However, based on the iterative process in the DMP development, other revisions can be included during the partnership implementation.

In general, the process for DMP revisions involves several steps:

- Identification of change requests: The DMP committee will identify new data management requirements imposed externally or emerging internally, evaluating their urgency and feasibility for implementation within the scheduled DMP update.
- DMP change specification: For each change request, the committee will design appropriate procedures and draft a candidate revision to the DMP.
- Pilot phase: The revised DMP procedures will be tested on a voluntary basis with a data item impacted by the change. WP5 will oversee and coordinate this pilot phase.
- Change implementation: Upon successful completion of the pilot, the DMP will be updated, and the new management procedures will be implemented.

This version includes a preliminary identification of research data by the WP leaders with the support of the partners. After a preliminary analysis, a data management strategy has been specified in this document. The next version of M22 will include a revised analysis of the research data identified, including new items, and a revision of the specific procedures to ensure FAIRness.

The Ethics Advisory Board will revise all ethic-related issues. The DMP will be changed based on their inputs. External requirement providers such as the EU, the ESFRI community, the EOSC association, and projects funded under the INFRA-EOSC initiative may influence new data management requirements. The foreseen DMP revisions are anticipated to take effect by months 22, 40, 64 and 82, respectively.

In the rare event of disruptive external change requests requiring an immediate response, T5.1 reserves the right to conduct extraordinary DMP revisions, which may bypass the pilot phase and become immediately effective. In such cases, the consortium will be promptly informed of the ongoing revision activities.

2 Data Summary

The objective of this chapter is to provide a summary overview of the different types of data and information being used in AGROECOLOGY, including existing data sources and new data being generated by the partnership.

In the partnership, a protocol has been established to support the implementation of the DMP at the WP level, described in the next section. Each WP identifies research data and outputs for open access, fills out a supervised metadata file, and records it in a central repository. This process, based on the EC's guidelines, ensures that research outputs are properly deposited and published in trusted repositories, following FAIR data principles.

During its operational phase, AGROECOLOGY will deal with a variety of data types. Some of these data will be directly generated and managed by AGROECOLOGY WPs, while others will be independently managed by funded research projects. In either scenario, for the sake of uniform data management, each distinct data asset should be categorized according to the descriptions outlined in this chapter and will be carefully managed to ensure its quality, integrity, accessibility, and (long-term) preservation. The DMP, precisely this and the following sections, will be updated and enriched as the partnership implementation progresses.

2.1 Data overview

A protocol is established to support the implementation of the DMP at the Work Package (WP) level. Thus, a template was distributed among WP leaders (Data Overview table) in order to collect essential information and check compliance with the DMP and European Commission (EC) recommendations. This template is based on the document “EU Grants: Data management plan [12]”, provided by the EC. Each WP should identify the research data and other research outputs, including their utility and their selection for open access based on their usefulness for the partnership consortium and other parties. After this decision is made, for data to be opened, the data owner must fill in the metadata file (Annex A – Metadata file template), which should be supervised by the leader of the WP and/or Task the data is related to before informing the DMP leader (LifeWatch ERIC). The metadata file will be included in the deliverable for which the data was created or used, and it will be recorded in the T5.1 repository (sharepoint). The metadata file comprehensively outlines descriptive information about the corresponding research data asset, ensuring a detailed data report aligned with the guidelines outlined in the Horizon Europe DMP template (HE, 2022) [12]. This information is essential to deposit and publish the research output in the trusted repository (see 2 FAIR data).

The Data Overview and Metadata file template table will be completed, considering the principles and guidelines provided in the initial DMP. The information collected from each WP following this procedure will contribute to providing updated versions of the DMP.

2.2 Data types and formats

The AGROECOLOGY partnership will gather a great variety of research data, including surveys, questionnaires and interviews, a list of tools and indicators, mappings of concepts, institutions and organisations, including living labs and research infrastructures, and data for monitoring agroecology transition, including environmental, social and economic data. AGROECOLOGY will also generate other types of research outputs, such as algorithms, workflows, capacity-building material, conceptual frameworks, strategies and systems, repositories, platforms and data infrastructures, among others. Table 6 outlines various types of resources anticipated to be generated as part of the AGROECOLOGY partnership.

Regarding the selection of file formats, it is crucial to publish data in open and interoperable formats whenever feasible. This promotes enhanced data reusability and eliminates dependency on proprietary software licenses for data reuse. Typically, the original 'raw' file format or a lossless format derived directly from it is most suitable for long-term storage. Table 6 presents a variety of recommended file formats. Other formats can be used while processing/analysing the data, but it is encouraged to convert to a recommended file format before archiving/publishing the data. Ideally, file formats should be non-proprietary (open), unencrypted, uncompressed, and commonly used by the research community. If it is not possible to use one of these file formats, the software name, version, license, and company should be documented in the metadata. AGROECOLOGY anticipates generating or reusing data ranging from megabytes to gigabytes in size. AGROECOLOGY will revise and update the data types and recommended formats listed in Table 6 throughout the duration of the partnership.

Table 6 Data types and formats

| Data Type | Data format |
|----------------------|---|
| Documents | <ul style="list-style-type: none"> - PDF (.pdf) - OpenDocument Format for text documents (.odt and .fodt) - Text file (.txt) - Microsoft Word (.doc) |
| Presentations | <ul style="list-style-type: none"> - OpenDocument Format for presentations (.odp and .fops) - Microsoft PowerPoint (.pptx) |
| Tabular | <ul style="list-style-type: none"> - Comma-separated values (CSV) (.csv) - OpenDocument Format for spreadsheets (.ods and .fods) - Microsoft Excel (i.e., .xls, .xlsx) |

| Data Type | Data format |
|----------------|---|
| | - Relational data (e.g., sql) |
| Images | - PNG (.png) - JPEG (.jpg) - GIF (.gif) - TIFF (.tiff) - Hyperspectral images (e.g., BIP, BSQ, BIL) |
| Videos | - MPEG-4 (.mp4) |
| Audio | - MP3 (.mp3) - FLAC (.flac) - WAV (.wav) |
| Spatial | - Shapefiles for vector files (.shp, etc.) - GeoTIFF for raster files (.tif, etc.) |
| Code | - Plain text (usually with an extension that represents the source language) |
| Other | - Compressed files (e.g., .zip, .gz) - Hierarchical data such as those based on XML (e.g., RSML to store 2D or 3D image metadata) - Resource Description Framework (RDF) for metadata |

2.3 Data reuse

An extensive review of archive data and data available in trusted repositories will be conducted periodically. AGROECOLOGY expects to reuse data from a variety of sources, including long-term monitoring, biodiversity databases, statistics databases, etc. Future versions of the DMP will provide a list of potential sources of public data that will be used in the AGROECOLOGY partnership.

2.4 Data utility

Data used and produced in the AGROECOLOGY partnership will serve to achieve the shared vision to “team-up and unlock the transition to agroecology so that farming systems are resilient, productive and prosperous, place-sensitive, as well as climate, environment, ecosystem, biodiversity and people-friendly by 2050”. To achieve this general vision, the partnership will support research and related activities that contribute to achieving objectives of key strategies under the Green Deal, notably the Farm to Fork and the EU Biodiversity strategies and specific Sustainable Development Goals (SDGs), enabling transformative change in the agricultural sector towards agroecology.

By providing robust, research-based insights, the data will enhance knowledge of the benefits and challenges of agroecology, particularly its potential to mitigate and adapt to climate change and improve farming practices, ecosystem services, and societal impacts. The data will support the development and co-creation of innovative practices, reducing transition risks for both individual

farmers and collectives through real-life evidence and shared experiences. Improved knowledge sharing and access will be facilitated by leveraging collected data and reinforcing agricultural knowledge and innovation systems across Europe. This will involve the Network of Living Labs and Research Infrastructures, as well as targeted communication with diverse stakeholders, to help remove barriers to agroecological transition. A comprehensive monitoring and data framework will be established, using harmonized methods and common indicators to measure progress in agroecological transition. This will integrate fragmented data repositories and make them accessible, thus improving data valorisation and sharing. The data will also facilitate exchanges with policymakers and stakeholders, providing evidence-based insights to improve governance, policies, and institutions supportive of agroecological practices. This will ensure the mainstreaming of agroecological practices and contribute to the development of supportive mechanisms necessary for effective agroecological transition.

2.5 Co-funded research projects

By means of co-funded joint calls for transnational research projects, AGROECOLOGY will pool the resources of the EC and the states involved to fund high-level research generating appropriate knowledge and technologies aligned with the core themes described in the SRIA [13]. However, the co-funded projects resulted from transnational calls are not property of the partnership and as such, any ethical issues are primarily in responsibility of the partners funding the co-funded projects (Grant Agreement) [14].

For all proposals of the co-funded transnational calls, an ethical-self assessment will be a mandatory part of the proposal. The Ethics Advisory Board will scrutinise proposals selected for funding accordingly. Compliance with the ethical principles and the applicable EU, international, and national laws for the ethics issues identified in the Ethics Summary Report and any additional ethics issues that may emerge in the course of the grant will be ensured. For any applicable ethics issue, the guidance provided in the European Commission ethics self-assessment guidelines will be rigorously followed.

The AGROECOLOGY DMP will be the basic strategy to ensure proper data management for co-funded projects to facilitate the fulfilment of the FAIR principles and data-related ethical issues. Co-funded project applicants are expected to read the present document and to conform to the herein-described policies upon planning their data management activities, also preparing a data management plan as a deliverable during the project implementation. Apart from the DMP, specific activities or guidelines will be provided in order to encourage the adoption of the AGROECOLOGY DMP by co-funded research projects.

3 FAIR data

The FAIR principles of findability, accessibility, interoperability and reusability are at the core of the AGROECOLOGY data management plan, and are discussed in this chapter.

3.1 Making data findable

AGROECOLOGY adopts the open science principle, so data or metadata should be deposited in a trusted repository as soon as possible, no later than six months after their creation, and always before the end of the partnership.

All data assets generated within AGROECOLOGY must be associated upon publication with a persistent identifier (DOI or another globally unique and persistent identifier) assigned by a trusted digital data repository (e.g Zenodo, GBIF, etc.) for effective and persistent citation. This persistent identifier can be used in any relevant publication to direct readers to the underlying data.

In order to make data findable, we need metadata (data about data) to help discover that data. Metadata contains information that describes the basic characteristics of each data asset, helping users to evaluate it and ultimately use it for their needs. Metadata must adhere to standardized formats consistent with community standards. It must be machine-readable, which enables computer processing techniques such as metadata harvesting, indexing, and the ability to cross-link between different research outputs. The structure of the minimum meta-information that will be recorded for each data asset generated within AGROECOLOGY is shown in Annex A. This aligns with the guidelines outlined in the Horizon Europe DMP template (HE, 2022) [12], and encompasses the minimum set of metadata elements as per the DataCite Metadata Schema standards [15]. It includes search keywords aimed at maximizing opportunities for reuse. Other standard metadata formats and vocabularies will be explored, once the preliminary set of datasets and research outputs are analysed and will be updated in the next and future versions of the DMP. Moreover, metadata about the different data assets in the partnership will be described and structured using vocabularies based on the Resource Description Framework (RDF) such as schema.org and [Data Catalog Vocabulary \(DCAT\)](https://www.datacatalog.org/).

For all data assets and research output, a metadata record will be created in the trusted repository, linking to the other repositories where the data can be found. The research outputs generated within AGROECOLOGY are owned by the partner/co-funded research project beneficiary that produces it and should be acknowledged when others use the data, along with acknowledging AGROECOLOGY. Both AGROECOLOGY partners and co-funded research project applicants must be aware that their published data assets may be included in a number of catalogues, such as the EOSC marketplace, ENVRI-Hub or Research Infrastructures data portals.

3.2 Making data accessible

The research output created in AGROECOLOGY is owned by the partner or co-funded project beneficiary that generates it, who is responsible for data dissemination unless there is a legitimate interest to protect it. The data will be made accessible following the accessibility criteria:

- Restricted data: accessible by the owner only, deposited in the repository of the owner partner or co-funded project beneficiary. Data must be kept for at least 5 years after the partnership ends.
- Consortium data: accessible by the consortium members only, deposited in the AGROECOLOGY Sharepoint. Data must be kept for at least 5 years after the partnership ends.
- Open data: accessible by the public, deposited in trusted data repositories. Data is kept for the lifetime of the selected trusted data repository.

AGROECOLOGY adopts open science principles, therefore, as soon as possible, no later than six months after their creation, and always before the end of the partnership, the research output authors should ensure open access, via the repository, to the deposited data, following the principle “as open as possible as closed as necessary”, unless providing open access would in particular:

- be against the data owners’ legitimate interests, including regarding commercial exploitation, or
- be contrary to any other constraints, in particular, the EU competitive interests or the authors’ obligations under the Grant Agreement.

In cases where it is considered necessary, research output owners may request embargo periods of up to two years for specific data assets. This allows time for publication or seeking intellectual property protection. During this period, the research output will be deposited in the trusted repository with an embargo date specified, after which the research output will be made accessible. For research output that cannot be shared, or need to be shared under restricted access conditions, data owners have to explain why, clearly separating legal and contractual reasons from intentional restrictions. Information about the restricted data will be collected in the Data Overview table. Where a restriction on open access to data is necessary, attempts will be made to make data available under controlled conditions to other individual researchers.

In general, AGROECOLOGY will consider data assets accessible if they are available through one or more trusted repositories. Co-funded research project beneficiaries who choose alternative approaches are required to outline in their co-funded research project data management plan how they will ensure accessibility for their assets, how they intend to secure adequate storage and backup

facilities to maintain all produced data assets after the partnership ends, and take full responsibility for long-term data preservation.

3.3 Making data interoperable

Data generated in the AGROECOLOGY partnership will adhere to standard formats upon publication and be as much as possible compliant with available (open) software applications, facilitating interoperability. All data providers are responsible for making sure their data adhere to international standards before publication to ensure interoperability. Which standards apply, depends on the data type.

While the AGROECOLOGY partnership does not mandate specific formats and vocabularies, data assets must be hosted in a trusted repository and appropriately annotated and formatted according to the repository's best practices. If any partner or co-funded project applicant opts for alternative hosting solutions, they must outline the metadata annotations and formats they intend to provide for each of their published data assets or research output in their data management plan.

Acceptable domain-specific metadata formats and vocabularies are listed in [FAIRsharing](#), and their use is further described in [RDMkit](#) and the [FAIR Cookbook](#).

3.4 Making data reusable

The data generated in the AGROECOLOGY partnership will be made available as open data to permit the widest possible reuse. AGROECOLOGY partners and co-funded research project beneficiaries may request data publishing embargoes of up to two years, e.g., to maximize the scientific impacts of the project or prepare exploitation measures. After this period, all data will be made open. For data that can be publicly released open licenses such as CC BY or CCO will be used.

The open data will be made available in trusted data repositories, as described earlier. When necessary, to facilitate data reuse and data analysis validation, documentation about data sources, used methodologies, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc., will be provided for each data asset.

If data assets are updated, the owner of 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. The quality control of the data is the responsibility of the owner of the data.

4 Other research outputs

Other research outputs, such as software, workflows, protocols, models, reports, presentations, and publications will be managed in the same way as research data assets. That is, a record will be collected from the WP Leaders and once selected for publication a metadata template will be filled at least with the compulsory fields for open research outputs (see Annex A). Then the output will be published in a trusted repository.

AGROECOLOGY conforms with Horizon Europe requirements to make all peer-reviewed scientific publications generated by AGROECOLOGY, as any other specific research output, open access immediately after publication. Publications will either (1) be published in an open access journal, or (2) be deposited in a repository for scientific publication (e.g. Zenodo) and be made open access.

5 AGROECOLOGY Data Infrastructure

In the AGROECOLOGY partnership, following the Open Data, Software, and Code Guidelines suggested by the European Commission, we intend to implement a modular and scalable data infrastructure for the application of the harmonised methods and tools for long-term data collection and analysis developed in WP5 Data and Monitoring Agroecology Transition. This infrastructure will allow the integration and standardisation of existing databases, following a scalability strategy. This aggregated database will also be aligned with the FAIR principles for the preservation of information, will allow the identification of information owners, and will ensure information traceability. By depositing such data/metadata into the infrastructure, we aim to facilitate monitoring of agroecology transition within the research community and beyond, promote reproducibility, and comply with funding agency requirements regarding data management and sharing policies.

In the next and following versions of the DMP, more detailed information will be provided because this infrastructure will be designed after M13, following the work developed in WP5 Data and Monitoring Agroecology Transition during the first year. This section will include specific information about data deposit and access control, data sharing and collaboration and a data gathering strategy that will help to understand how this repository will be populated and exploited by WP5.

6 Allocation of resources

The recommended trusted data repositories are free of user charge. Therefore, no immediate costs are anticipated in opening the datasets following the FAIR principles. Partnership members have a specific budget for open access publication.

The costs generated by hosting non-public partnership results will be carried by the partners owning them and the partners owning the non-public results commit to keep and manage them for a minimum of 5 years after the partnership finalization.

LifeWatch ERIC will be the responsible consortium partner for the implementation of the DMP. All the consortium partners will also undertake to support the data management protocol by providing the types of generated data that have been described in the methodology. In more detail, LifeWatch ERIC will:

- prepare the DMP,
- ensure its update during the partnership implementation,
- ensure the recording of the produced datasets,
- have the overall responsibility for the implementation of the DMP.

Each AGROECOLOGY partner should respect the policies set out in this DMP. Data sets must be created, managed, and stored appropriately and in accordance with local legislation and following the indications included in this DMP. The partner that generates the data is responsible for following the instruction once new data is generated but also for data set validation, metadata registration, and data backup for sharing through repositories.

7 Data security

The AGROECOLOGY consortium does not enforce specific provisions on data security, but nevertheless expects all its partners and co-funded research project beneficiaries to comply with their parent institution's policies and best practices, therefore the AGROECOLOGY consortium declines any responsibility on data breaches and/or data losses.

All consortium partners fully comply with current security standards in terms of data security, backups, anti-virus scans, etc. The servers, on which the data files will be stored are password protected. The passwords are unique, contain more than eight symbols and do not contain any personal information. Only staff working directly on this partnership will have access to the files containing the personal data of the research subjects. The computers and files are password-protected. In case more strict data policies in specific institutions are in place, they will be followed by affected partners.

The AGROECOLOGY partnership has a Microsoft Teams sharepoint that is shared with all partners and is used to store and share common datasets (experimental data), results, contact lists, templates and other data that is required for the implementation and progress of the partnership. The MS-Teams channel is managed by ILVO, providing rights to partners following a strict protocol. The sharepoint is regularly backed up, and it serves as a tool for document sharing, with an attached and dedicated data repository for each WP. After the partnership finalization, ILVO will keep all partnership data for at least 5 years. MS Teams have a security layer implemented under an Authentication, Authorization, and Accounting Infrastructure (AAAI) that requires a username and password.

In case of sensitive data, it should be managed by its respective owners and should never be uploaded to the shared partnership data space.

Open research data will be stored in trusted repositories, which have provisions in place for data security.

8 Data ethics

As AGROECOLOGY will operate in an international context and include multiple stakeholders, ensuring fair and legal data processing to all assets is of vital importance. To cope with the different nature of the data assets, methodologies to address ethical requirements as well as data privacy and security are of utmost importance to establish a general framework in this heterogeneous environment. This section applies to all the research data created and/or curated within the AGROECOLOGY partnership and its cascading co-funded research projects, but in particular to research data assets containing personal information.

As stated in the AGROECOLOGY Consortium Agreement, results, thereby including all types of data assets, are owned by the party that generates them. Where it is impossible to establish each participant's respective contribution to the result or separate the result, then to apply for, obtain or maintain their protection, the provisions for Joint Ownership (detailed in Grant Agreement Article 8.2 and Article 16.4 and its Annex 5) apply. AGROECOLOGY is therefore committed to a fair and transparent management of **intellectual property** and condemns all predatory behaviour aimed at superseding, circumventing, or working around the above-described data ownership principles. The AGROECOLOGY consortium acknowledges that FAIR data and open data practices should be instruments to promote reproducible research and knowledge transfer, not means to transfer intellectual property and/or prevent data owners from exercising their rights.

AGROECOLOGY and its participants are committed to guaranteeing their users lawful, fair, and **transparent personal data processing**, as established by Article 5 of the GDPR. In this section, we address the general principles according to which personal data is collected and gathered within AGROECOLOGY. Personal data is information relating to an identified or identifiable natural person. An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (Article 2(a) EU General Data Protection Regulation 2016/679 (GDPR)). Individuals are not considered 'identifiable' if identifying them requires excessive effort.

The collection, processing and transmission of personal data will be analysed under the principles of

- the GDPR (Regulation (EU) 2016/679);
- the Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data;
- the national laws applying their provisions.

For countries to which the GDPR does not apply (e.g., Serbia, Switzerland, and Türkiye in the consortium and others in the co-funded research project), we will observe that the applicable laws and regulations are compliant with the GDPR. If this is not the case, the GDPR applies to all personal data transferred from the EU to the respective non-EU country.

In a preliminary analysis at M6, AGROECOLOGY will collect personal data only for internal purposes, i.e., for co-funded project applications, organizing meetings and workshops, sending email, etc. and does not handle sensitive data. Personal data includes, but is not limited to:

- a. AGROECOLOGY co-funded project applications that will require registration from applicants and a list of experts to compile the international expert panel to evaluate proposals.
- b. Registration forms for representatives' external organizations, participating in an AGROECOLOGY organized workshops.
- c. Seminars organized by AGROECOLOGY, that require participant registration.
- d. Surveys (either technological or user surveys) that are not anonymous for specific reasons.
- e. Stakeholder, project and initiative mappings that can support the implementation of the actions and can contribute to the liaison strategy.

These data can be considered research data as they might be examined and considered and serve as a basis for reasoning and discussion. Other personal data that is not considered research data is not part of the DMP and is being considered in the WP9 Ethics, and always managed under the GDPR regulation.

AGROECOLOGY promotes the implementation of several different co-funded research projects. Hence, the project stakeholders (internal and external) and the people and organisations affected by the data collection will vary from project to project. The collection, processing and transmission of personal data should follow the principles of this DMP and following national and international regulations. Any additional regulations at national level that do not fall under the GDPR and apply to data protection or any other sensitive information will also be taken into account. The responsibility of lawful and ethical personal data management lies upon the co-funded project beneficiary collecting and managing personal data. Before implementation, each co-funded research project is expected to provide the consortium with an ethical self-assessment, including the planned personal data collection, storage, and processing activities. The ethical self-assessment checklist and executive form are made available through the platform and its fulfilment is a mandatory step in the research project application process.

Personal data will be managed with care within AGROECOLOGY. They will be stored and managed in a secured IT environment by the partners or by using the Microsoft Teams sharepoint. Access to the data requires a member of guest status in the AGROECOLOGY MS Teams, which is given to partnership members by ILVO following a specific procedure. Privacy of data subjects will be secured by fully

complying with the General Data Protection Regulation (Regulation (EU) 2016/679 of the European Parliament and of the Council). The partnership consortium has appropriate technical and organizational measures in place to carry out data protection during the partnership.

In accordance with the GDPR's provisions, data subjects must be able to express consent to data collection and processing explicitly, as their consent cannot be taken for granted under any circumstance. Similarly, data subjects must be provided with the means to exercise their rights on the data they provided the consortium with. The AGROECOLOGY consortium expects each partner and each co-funded research project beneficiary to manage user consent and to administer access to the fundamental data subject rights in accordance with their institution's guidelines and best practices. The AGROECOLOGY consortium will not enforce additional policies and/or constraints on its members and/or related organisations as long as significant incompatibilities with the consortium's General Ethical Guidelines described in D9.1 emerge, in such a case the ethical conflict resolution procedures described in the same document (D9.1) will be applied. In the event a partner or research project applicant does not have a consent management and data subject rights administration policy to refer to, the AGROECOLOGY consortium expects the partner or co-funded research project beneficiary to produce one during project implementation. Such custom policy must include an example of a consent form to be administered to data subjects (a model template is proposed in D9.1) and a clear explanation of how they intend to allow their data subjects to exercise their fundamental rights (namely: the right to be informed, the right of access, the right to rectification, the right to erasure/be forgotten, the right to restrict processing, the right to data portability, and the right to object) as established by the GDPR.

Research data assets will be published if and only if they do not include non-anonymized personal data. If the personal data therein cannot be anonymized in any way, the data asset will be published only if all the data subjects will give explicit and verifiable consent to publication.

Each consortium partner has an appointed Data Protection Officer (DPO), who is responsible for advising the partnership with regard to compliance with the GDPR. Moreover, a confirmation that all collected and processed personal data will be handled according to the GDPR and in addition the Privacy Policy of all partners will be requested. For countries that GDPR do not apply, a confirmation will be required to ensure that the applicable laws and regulations are compliant to the GDPR. If this is not the case, the GDPR will be respected for all personal data transferred from EU to the respective non-EU country.

The Ethics Advisory Board will revise all ethic-related issues. The DMP will be changed based on their inputs, following the procedure for DMP change management (see section 1.2. DMP change management).

9 Conclusions

The present document describes the DMP of the AGROECOLOGY partnership and emphasizes the importance of effective data management in achieving partnership objectives. By following this DMP, the AGROECOLOGY partnership can ensure the efficient and responsible management of data generated by partnership partners, but also co-funded projects, maximizing the impact and value of research outcomes and promoting collaboration and knowledge exchange within the research community.

10 References

- [1] European Commission, “Commission Recommendation (EU) 2012/417/EU of 17 July 2012 on access to and preservation of scientific information,” Official Journal of the European Union, vol. 55, no. L 194, p. 39–43, 21 July 2012.
- [2] European Commission, “Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information,” Official Journal of the European Union, vol. 61, no. L 134, pp. 12-18, 31 May 2018.
- [3] European Commission, “Horizon Europe Programme Guide (V4.1),” 1 May 2024. [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf [Accessed May 2024].
- [4] MARBEFES, D7.2 Data Management Plan, v2.0. February 2024.
- [5] ALL-Ready D6.2 Data Management Plan. V 3.0. October 2023.
- [6] AGROSERV D9.4 Data Management Plan. V1.0. February 2024.
- [7] PATH2DEA Data Management Plan. V1.0. June 2024.
- [8] ICT-AGRI-FOOD Joint Call Template. December 2016.
- [9] ERA CoBioTech Info DM Joint Call for Proposals Data Management. December 2016.
- [10] Proton AG, “What is GDPR, the EU’s new data protection law?,” 2023. [Online]. Available: <https://gdpr.eu/what-is-gdpr> [Accessed January 2024].
- [11] European Commission, “Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast),” Official Journal of the European Union, vol. 62, no. L 172, pp. 56-83, 26 June 2019.
- [12] EU Grants: Data management plan (HE):V1.1 – 01.04.2022 [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx [Accessed January 2024].
- [13] SCAR, The AGROECOLOGY partnership’s SRIA [Online] Available: https://scar-europe.org/images/Agroecology/SRIA_rev23-02-2023.pdf [Accessed January 2024]
- [14] European Commission, AGROECOLOGY Grant Agreement N°: 101132349.
- [15] DataCite Metadata Working Group. (2024). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.5. DataCite e.V. <https://doi.org/10.14454/g8e5-6293>



11 Annex A – Metadata file template for data assets and other research outputs

(* compulsory fields)

| DATA DENOMINATION |
|---|
| <p>Data set reference name*</p> <p>The reference name will start with the prefix “DA_AGROECOLOGY” indicating it is a data asset <i>E.g.: RO_AGROECOLOGY_stakeholders_survey</i></p> |
| <p>Data set title*</p> <p>The title of the data asset which should be easily searchable and findable <i>E.g.: Stakeholders_Survey_Q2_20232</i></p> |
| <p>Description</p> <p>A brief description of the data asset <i>E.g.: Survey on technological tools used in agroecology.</i></p> |
| <p>Keywords</p> <p>The keywords associated with the data asset <i>E.g.: survey, tools, living labs</i></p> |
| <p>Version Number*</p> <p>To keep track of changes to the data assets <i>E.g.: V 1.0</i></p> |
| DATA ORIGIN |
| <p>Name(s) of data set creator(s)*</p> <p>Lead partners responsible for the creation of the data asset, and their affiliation <i>E.g.: Name Surname, Affiliation, Country</i></p> |
| <p>Data Source</p> <p>How/why was the data asset generated/re-used <i>E.g.: generated data set</i></p> |
| <p>Creation Date*</p> <p>Date of generation of the data asset <i>E.g.: dd.mm.yyyy</i></p> |
| <p>Quality Assurance</p> <p>Brief description of the quality assurance processes the data has gone through <i>E.g.: response rate monitoring, data profiling,</i></p> |
| DATA SPECIFICATIONS |
| <p>Type*</p> |



Type of data contained in the data asset

E.g.: survey data

Format*

This could be CSV, PCAP, HTML, etc.

E.g.: .xlsx

Expected Size*

The approximate size of the data asset

E.g.: 1 MB

DATA ACCESSIBILITY

Data Location*

The repository where the data are stored: partner owned, AGROECOLOGY Sharepoint, other trusted data repository (e.g., Zenodo, GBIF, etc.)

E.g.: GBIF

Date of Repository Submission*

The date of submission to the repository can be added once it has been submitted

E.g.: yyyy.mm.dd

Persistent Digital Identifier*

The PDI can be entered once the data asset has been deposited in the repository

E.g.: <https://doi.org/10.xxxx/dome.xxxx>

Access Status*

Whether a data asset is “open”, “consortium” or “restricted”

E.g.: consortium

Embargo Period*

Duration of embargo foreseen for a data asset

E.g.: no embargo

Funding statement

A disclaimer indicating that the specific data asset was generated within the AGROECOLOGY partnership that received funding from the EU

DATA UTILITY

Significance to the partnership*

The associated work package or co-funded project this data asset originates from

E.g.: WPx – Task y.z

Significance outside the partnership

To whom the data might be useful outside the partnership

E.g.: policy makers



DATA PUBLICATIONS

Related Publications

Bibliographical details of publications based on the data asset

E.g.: deliverable x, scientific article y, ...

Data Citation

A 'ready-to-use' citation reference for the data asset

E.g.: Author1 Surname, Name, Author2 Surname, Name. (yyyy). Publication title. Publisher.

