BioValue

Deliverable 11.12

Data Management Plan

Lead Beneficiary:	AUTH
Delivery Date:	31.03.2022
Dissemination Level:	Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000499

Project Acronym	BIOVALUE
Project Title	Fork-to-farm agent-based simulation tool augmenting BIOdiversity in the agri-food VALUE chain
Activity	RIA Action
Call	H2020-SFS-2020-2
Funding Scheme	GA No: 101000499

Deliverable No.	11.12
Dissemination Level[1]	Public
Work Package	11
Task(s)	Task
Lead Beneficiary	AUTH
Contributing beneficiary(ies)	All partners
Due date of deliverable	31.03.2022

Document History

Version	Date	Beneficiary	Author
1.0	31.03.2022	AUTH	ANASTASIOS MICHAILIDIS ASTERIOS THEOFILOU

Executive Summary

This deliverable provides the BIOVALUE data management plan (DMP) version 1. The deliverable outlines how the research data collected and/or generated will be handled during the project. It describes which standards and methodology for data collection and generation will be followed, and whether and how the data will be shared. It will serve primarily as a guide for the BIOVALUE consortium and at the same time, being a public document, it will enable the easy identification of the datasets generated within the project by other researchers that want to reuse them. The DMP will define the ways and methods by which datasets will be generated and used within the project, how they will be categorized and technically detailed in terms of data collection, processing, and generation.

This DMP V1, is a first draft of the document. It will be a "living document", that will be continuously updated with the information provided by all partners in the project. A final version of the DMP, will be submitted to the European Commission by the end of the project.

This document follows the template provided by the European Commission in the Participant Portal (Data Management Template (HE): V1.0 – 05.05.2021). In line with the EU's guidelines regarding the DMP, this document addresses for all datasets (collected, generated, and processed), the following aspects: Dataset description, reference, name, standards, metadata, sharing, archiving and preservation. This document covers the methods and strategies developed by the consortium in order to address the aspects concerning the datasets.

List of Tables

List of Figures

Abbreviations and Acronyms

Abbreviations	Explanation
D	Deliverable
WP	Work Package
M	Month

Table of Contents

1. Data summary

This section contains information regarding the the re-use of data, its types and formats, its purpose, its expected size, its origin, and its utility.

1.1. Data re-use and generation of datasets

This section contains information regarding the the re-use of data, its types and formats, its purpose, its expected size, its origin, and its utility.

1.2. Types and formats of data

In this current version of the DMP (V1), a first identification of standards has been compiled:

For textual data, the project will use Microsoft Word 2010 (and any other compatible version), .doc, .docx, .xls, .xlsx, .ppt, .pptx, .csv, and .txt file formats.

All finished and approved documents will additionally be made available as .pdf files.

For images, illustrations, and graphic designs, the .jpg, .png, .psd, .ai, .tiff, and .vsd file formats will be used.

For audio files, the mp3 and wav file formats will be used.

For video files, the .mp4, .mov, and .wmv file formats will be used.

The file formats mentioned above have been chosen for their wide accepted standards, their widespread use, and the ease of access to the software needed to handle them.

The files will be converted to open file formats where possible for long-term storage.

Metadata will be comprised of two formats: contextual information about the data in a text-based document and ISO 19115 standard metadata in an xml file. These two formats for metadata are chosen to provide a full explanation of the data (text format) and to ensure compatibility with international standards (xml format).

1.3. Data purpose

The BIOVALUE tool aspires to model the interaction of agents in the agri-food value chain and the impact those interactions have in Biodiversity, in novel and more complete ways. To this aim, the partners will re-use existing and create new datasets, that help “paint” a better and clearer picture of those interactions. Regarding the generated datasets, the details for the datasets that will be generated have not been defined yet. Currently, a template has been drafted to guide all partners that will generate datasets with the information needed to help characterize and identify those datasets.

1.4. Data size

At this point of the project’s lifetime, an estimation of the size of the datasets cannot be provided.

1.5. Data origin

The datasets that will be used and generated cover the fields of climate research, land and soil quality and characteristics, farm characteristics, biodiversity, Agriculture, farmers/consumers/all agri-food value chain stakeholders’ behavior. They will come from EU, national and regional existing data sources as well as from participatory research and data analysis and research within the BIOVALUE project.

1.6. Data utility

The data that will be collected and generated for the project, can be used by Policy makers, researchers, other projects, to validate/reproduce the project's outcomes, to expand the project's outcomes, to complement various other research activities.

2. FAIR data

This section presents the ways and methods by which the BIOVALUE consortium will handle the data in order to make it Findable, Accessible, Interoperable and Reusable (FAIR).

2.1. Making data findable, including provisions for metadata

All datasets will be assigned a unique name that will serve both as a unique identifier and to be distinguished among various other datasets.

For that reason all datasets will follow a consistent naming convention: a) all dataset names will start with the name "BIOVALUE", b) next a brief description of the content, c) a version number, or the word "FINAL" will follow, d) the short name of the organization that provided the dataset.

All the parts of the dataset names will be joined with underscores.

Example "BIOVALUE_ResultsFromSurveyOfFarmersinGreece_V1_AUTH"

Each dataset to be used by the project will be accompanied by a generated form (utilizing information derived from the Annex 1 template), that will contain useful details describing it.

If any of the dataset description fields is not determined at the time of the dataset creation, the label TBD (to be determined), will be included in the description. If a description field is not applicable, the NAP label will be included in the description.

The BIOVALUE consortium will utilize the ARDIT tool, created by the AGRICORE project to characterize, and render the datasets metadata available to the public. To do so, all the ARDIT recommendations and guidelines will be followed and respected.

Doing so, the datasets metadata will be indexed and discoverable by keywords search.

2.2. Making data accessible

Following the guidelines presented by the EC, the BIOVALUE consortium, will comply with the H2020 requirements for Open Access publishing. The projects data will be self-archived in publicly accessed repositories, ensuring "green" and "gold" publishing. The datasets of results, will be made publicly available through the project's website, public software repositories such as GitHub and. In the case that certain datasets cannot be shared, the partner that decides so, will provide an explanation, clearly separating legal and contractual reasons from intentional restrictions. The non-open research data will be archived and stored long-term in the BIOVALUE REDMINE platform, administered by AUTH. The BIOVALUE REDMINE platform is currently being employed to coordinate the project's activities and to store all the digital material connected to BIOVALUE. If certain datasets cannot be shared (or need restrictions), legal and contractual reasons will be explained.

2.3. Making data interoperable

In this current version of the DMP (V1), a first identification of standards has been compiled:

For textual data, the project will use Microsoft Word 2010 (and any other compatible version), .doc, .docx, .xls, .xlsx, .ppt, .pptx, .csv, and .txt file formats.

All finished and approved documents will additionally be made available as .pdf files.

For images, illustrations, and graphic designs, the .jpg, .png, .psd, .ai, .tiff, and .vsd file formats will be used.

For audio files, the mp3 and wav file formats will be used.

For video files, the .mp4, .mov, and .wmv file formats will be used.

The file formats mentioned above have been chosen for their wide accepted standards, their widespread use, and the ease of access to the software needed to handle them.

The files will be converted to open file formats where possible for long-term storage.

Metadata will be comprised of two formats: contextual information about the data in a text-based document and ISO 19115 standard metadata in an xml file. These two formats for metadata are chosen to provide a full explanation of the data (text format) and to ensure compatibility with international standards (xml format).

2.4. Increase data re-use

All data created and used by the project will be curated based on the sharing policies attached to them. The aim is to preserve the data and render it easily accessible to all interested parties, for the whole duration of the project and beyond. To this aim, a public API will be provided to registered users allowing them to access the data storage platform. This will ensure the correct implementation of the security policies on the databases, the analysis of the entire database, the implementation of user rights access, the application of strong/useful passwords. With the database implementation, the following measures will be considered:

- Dataset minimization. The minimum amount of data needed will be stored in order to prevent potential risks.

- Access control list for user and data authentication. Depending on the dissemination level of the information an Access Control list will be implemented reflecting there for the data sets that can be accessed by each user.

- Monitoring and Log of Activity. For each user in the platform, there will be kept a track log with their activity (datasets accessed, time spent, downloaded material), in order to track and detect possible harmful behavior of users.

- Implementation of alert system that informs in real time of the violation of procedures or hacking attempts.

- Liability. Identification of a person responsible for safe keeping of the stored information.

- When possible, the information will be also made available in the initiative that the EC has launched for open data sharing from research (ZENODO.ORG).

All the aforementioned actions, aim to reduce the risks related to data storage. However, due to the activities that are going to be carried out in the project, the amount of time for which the data will be stored in the platform is still not defined.

3. Other research outputs

In case that the project will generate or re-use other research outputs, either digital of physical, due consideration will be given for its management, in regard to application of the FAIR principles.

4. Allocation of resources

Within the BIOVALUE project, data management will be performed as part of WP11 and AUTH as project coordinator, will be responsible for data management in BIOVALUE project. Therefore, AUTH will be in charge of continuously updating the DMP, and to deliver its final version by the end of the project. All partners are responsible for updating in the BIOVALUE-REDMINE site, any kind of data used or generated by them (i.e., publications, dissertations, presentations, datasets that have been generated/used in the frame of the project).

A part of the overall WP11 budget and person months has been allocated by AUTH for these activities. For the time being, the project coordinator is responsible for FAIR data management. Costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). Resources for long term preservation, associated costs, and potential value, as well as how data will be kept beyond the project and how long, will be discussed by the whole consortium during General Assembly meetings.

5. Data security

For the duration of the project, datasets will be stored on the responsible partner's storage system. Every partner is responsible to ensure that the data are stored safely and securely and in full compliance with European Union data protection laws. After the completion of the project, all the responsibilities concerning data recovery and secure storage will go to the repository storing the dataset.

All data files will be transferred via secure connections and in encrypted and password-protected form (for example with the open source 7-zip tool providing full AES-256 encryption: http://www.7-zip.org/ or the encryption options implemented in MS Windows or MS Excel). In regards to passwords, precaution measures will be taken to restrict unwanted/unnecessary access to platforms, repositories, DWH's (data ware house), that contain datasets.

6. Ethics

This section deals with ethical and legal compliance issues, like the consent for data preservation and sharing, protection of the identity of individuals and companies and how sensitive data will be handled to ensure it is stored and transferred securely. Data protection and good research ethics are major topics for the consortium of this project. Good research ethics meet all actions to take great care and prevent any situation where sensitive information could get misused. This is what the consortium wants to guarantee for this project. Research data which contains personal data will just be disseminated for the purpose for which it was specified by the consortium. Furthermore, all processes of data generation and data sharing must be documented and approved by the consortium to guarantee the highest standards of data protection.

BIOVALUE partners must comply with the ethical principles as set out in Article 34 of the Grant Agreement, which states that all activities must be carried out in compliance with:

Ethical principles (including the highest standards of research integrity) and

Applicable international, EU and national law.

Data collection, storage, protection, retention, and destruction will be carried out through the BIOVALUE-REDMINE platform. Interviewees/beneficiaries/recipients will be informed about data security, anonymity and use of data as well as asked for accordance. Participation happens on a voluntary basis.

Additionally, all ethical issues will be addressed in the context of WP1.

7. Other issues

At this phase of the DMP, no other issues have been identified by the BIOVALUE project, to be addressed.

8. Conclusions

This deliverable provides the first version of the BIOVALUE data management plan. The DMP will be a "living" document, continuously updated based on the project's needs. The final version of the DMP will be submitted by the end of the project. At this early stage of the project, a first identification of standards for types and formats of data has been compiled, and initial considerations have been documented regarding the purpose, origin, and utility of data. Additionally, the ways and methods by which the BIOVALUE consortium will handle the data in order to make it Findable, Accessible, Interoperable and Reusable (FAIR), have been presented. In the Allocation of resources section of the DMP, the responsibilities for data management have been defined, and finally the initial steps for data security and ethics have been addressed in the respective parts.