



Boosting functional biodiversity to maximize ecosystem services for Mediterranean crop production

Project acronym: **ECOBOOST**

Deliverable n#	D1.c
Deliverable name	Data Management Plan (DMP)
Version	1.0
Release date	November 25th, 2024
Dissemination level	Public
Status	Draft
Author	UNIPA – Antonino Cusumano

Document History

Version	Date of issue	Content and changes	Edited by
1	16/11/2024	Minor changes to DMP 1.4b	Antonino Cusumano

Table of Contents

1. Data Summary.....	4
2. FAIR data.....	11
2. 1. Making data findable, including provisions for metadata.....	13
2.2. Making data openly accessible.....	13
2.3. Making data interoperable.....	13
2.5. Making data re-useable.....	13
3. Allocation of resources.....	14
4. Ethical aspects.....	14

The aim of the ECOBOOST's Data Management Plan (DMP) is to identify the project's research data and to describe how to make them findable, accessible, interoperable and re-usable (FAIR). Following the H2020 Data Management Plan Template available on the Participant Portal, all the partners involved in the research activities were asked to provide detailed information about the data generated during the entire project as reported in 1. Data Summary.

In this third DMP (1.4c), we do not have significant changes to report respect the previous version (1.4b - DMP). The main difference refers to: 1) changes in the status of the document from "confidential" to "public" following the recommendation of the project officer. The document is now also published on the project website (<https://ecoboost-prima.com/resources/>); 2) data interoperability and 3) timeline for re-use of data: both sections were previously not defined in the DMP. For the other sections, and considering the data generated so far, all PIs agreed that the data management plan is well detailed, and it specified the purpose of the data collection, the formats, the origin of data, the size and the data utility. Thus the document contains an analysis on how we intend to manage the amount of data produced in the project, the resources required to do so and the ethical aspects.

1. Data Summary

Hereafter, we report the tables where the information about the contents and the collection/generation of the data provided by each ECOBOOST Partners. At this early stage of the project, the reported information will be processed to determine the general specifications and structures of the metadata that will be generated within the DMP. The table for UNIRC can also be applied to UPMC1.

Coordinator - UNIPA	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	UNIPA generates data related to A) identification of Mediterranean accessions of solanaceous plants resistant to feeding damage and egg deposition of insect pests. B) identification of wild Mediterranean plants that attract and/or increase longevity of pest natural enemies; C) use of beneficial soil microbes (<i>Trichoderma</i> spp) that boost plant defenses against insect pests. Both A, B aim to screen plant material, while C aims to use beneficial microbes, that can be implemented in sustainable crop production under Mediterranean conditions
<i>What types and formats of data will the project generate/collect?</i>	Mainly documents (doc, docx, pptx, etc.), illustrations (png, jpg, etc.), drawings (svg, dwg, etc) and raw data (csv, txt, xls, raw, etc.).
<i>Will you re-use any existing data and how?</i>	Data from literature, technical reports, previously existing data from partners on tomato genetic diversity (WP2) and on wild Mediterranean plants (WP3) (comparison, reference).
<i>What is the origin of the data?</i>	Data originate from A) performance assays to test tomato genetic resistance against stink bug

	eggs; B) behavioral and chemical investigation of parasitoid responses towards Mediterranean flowering plants using Y-tube olfactometer assays. Plant volatiles will be analyzed with gas-chromatography coupled with mass spectrometry (GC-MS).
<i>What is the expected size of the data?</i>	Most of the data are expected to be around hundreds of megabytes
<i>To whom might it be useful ('data utility')?</i>	The data are of fundamental importance for some partners of the ECOBOOST consortium (especially for WP2, WP3 and WP4) and can be useful for possible scientific production (publications, meetings, etc.).

Partner 1 – UNIRC	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	UNIRC generates data related to A) the identification of other Mediterranean accessions of solanaceous plants tolerant to low-N and salt stress, alone or in combination. B) the assessment of the biodiversity and structural genetic variants of solanaceous collections from different ECOBOOST partners by using SNP array and genome sequencing; C) the identification of candidate genes related to tolerance/resistance to abiotic and biotic stressors by transcriptomic approaches. Both A, B aim to screen germplasm collections and to select new breeding lines of interest, while C aims to identify genes and/or QTL useful for molecular assisted selection (MAS) and further studies for tolerances/resistances mode of action
<i>What types and formats of data will the project generate/collect?</i>	Documents (doc, docx, pptx, etc.), illustrations (png, jpg, etc.), drawings (svg, dwg, etc) and mainly raw data of DNA sequencing (csv, txt, xls, raw, fastq, etc.).
<i>Will you re-use any existing data and how?</i>	Data from scientific literature and technical reports from ECOBOOST partners, but not only, on tomato, eggplant, and pepper genetic diversity (SNP and sequencing data).
<i>What is the origin of the data?</i>	Data originate from A) screening of germplasm collections in controlled condition (hydroponic and pot systems; B) SNP array analysis on germplasm collections, cluster analysis and genome sequencing of the most interesting genotypes for tolerance/resistance to stressors; C) RNA-seq experiments and comparative

	analyses of transcripts abundances for the identification of hub genes of interest.
<i>What is the expected size of the data?</i>	SNP, RNA-seq and genomes sequencing data are expected to be around hundreds of gigabytes
<i>To whom might it be useful ('data utility')?</i>	The data might result of key importance for the ECOBOOST partners, but not only, and will be useful for scientific production (publications, meetings, etc.).

Partner 2 – UGOE	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	UGOE Purpose of the data is A) to obtain (identify) <i>Trichoderma</i> strains locally adapted to Mediterranean climate. B) Identify role of <i>Trichoderma</i> volatiles on plant defense against nematodes; C) Identification of <i>Trichoderma</i> strains with protective effect against nematodes. D) Identification of <i>Trichoderma</i> strains that boost the defense of selected solanaceous accessions (provided by UNIRC and UNIPA) against leaf pests. All data aims to test <i>Trichoderma</i> strains to strengthen plant resistance to animal pathogens and to researching the underlying mechanisms
<i>What types and formats of data will the project generate/collect?</i>	Documents (doc, docx, pptx, etc.), illustrations (png, jpg, etc.), drawings (svg, dwg, etc), raw data of phylogenetic analysis, gene expression, volatile analysis (csv, txt, xls, FASTA, pcrd etc.).
<i>Will you re-use any existing data and how?</i>	Data from literature, Data from ECOBOOST partners, Genbank sequences
<i>What is the origin of the data?</i>	Data originates from A) collection and phylogenetic analysis of <i>Trichoderma</i> strains from Greece, B) Nematode behavior in 4-arm olfactometer assays and volatile collection of <i>Trichoderma</i> and <i>Trichoderma</i> inoculated plants; <i>Trichoderma</i> volatiles will be analyzed with gas-chromatography coupled with mass spectrometry (GC-MS). C) Screening of protective effect of Mediterranean <i>Trichoderma</i> stains against nematodes and leaf pests, and D) Physiological and molecular responses of plants to <i>Trichoderma</i> inoculation and nematode and herbivore attack
<i>What is the expected size of the data?</i>	Hundreds of megabytes
<i>To whom might it be useful ('data utility')?</i>	Data about A) <i>Trichoderma</i> strains will be used by ECOBOOST partners, in particular DUTH, and for the WP 6 in the field trials. Additionally, it will contribute to understand the mode of action of the beneficial <i>Trichoderma</i> , and will be used for scientific divulgation (publications, congress, etc)

Partner 3 – DUTH	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	DUTH generates data related to A) the sampling of solanaceous plants and soil; B) the assessment of wild Mediterranean plants that attract or enhance the performance of natural enemies; C) the use of beneficial soil microbes and volatiles that boost plant defenses against arthropod pests; D) the application and demonstration of microbes, volatiles and coated seeds under field conditions; and E) the distribution of questionnaires related to the socioeconomic analyses of the projects. A, B aim to sample plant and soil material, while C and D aims to use beneficial microbes and volatiles, that can be implemented in sustainable crop production under Mediterranean conditions. E aims to assess the environmental and socioeconomics of ECOBOOST tools.
<i>What types and formats of data will the project generate/collect?</i>	Mainly documents (doc, docx, pptx, etc.), illustrations (png, jpg, etc.), drawings (svg, dwg, etc.) and raw data (csv, txt, xls, raw, etc.).
<i>Will you re-use any existing data and how?</i>	Data from literature, technical reports, previously existing data from partners on wild Mediterranean plants (WP3) and on beneficial soil microbes and volatiles (WP4) (comparison, reference).
<i>What is the origin of the data?</i>	Data originate from A) samplings of soil and solanaceous plants; B) performance assays to test herbivore performance on microbe-inoculated plants; B) behavioral and chemical investigation of natural enemies' responses towards Mediterranean flowering plants or microbe-inoculated plants using Y-tube olfactometer assays. Plant volatiles will be analyzed with gas-chromatography coupled with mass spectrometry (GC-MS); C) molecular study of plants resistance and priming as affected by microbes and seed coating; D) questionnaires distribution to end-users.
<i>What is the expected size of the data?</i>	Most of the data are expected to be around hundreds of megabytes.
<i>To whom might it be useful ('data utility')?</i>	The data are of fundamental importance for some partners of the ECOBOOST consortium (especially for WP2, WP3, WP4, WP5 and WP7) and can be useful for possible scientific production (publications, meetings, etc.).

Partner 4 – BPI

<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	BPI generates data related to A) identification of wild Mediterranean plants with impact on fitness of pest natural enemies and attract pollinators B) characterization of impact of flowering plants on natural enemies and pollinators in the field and C) identification synergisms and trade-offs between species that attract pollinators and those attracting natural enemies. All activities aim to screen plant material, that can be implemented in sustainable crop production under Mediterranean conditions
<i>What types and formats of data will the project generate/collect?</i>	Mainly documents (doc, docx, pptx, etc.), illustrations (png, jpg, etc.), drawings (svg, dwg, etc) and raw data (csv, txt, xls, raw, etc.).
<i>Will you re-use any existing data and how?</i>	Data from literature, technical reports, previously existing data from partners on wild Mediterranean plants (WP3) (comparison, reference).
<i>What is the origin of the data?</i>	Data originate from A) botanical surveys for wild plants that are visited by natural enemies and pollinators B) behavioral and chemical investigation of parasitoid responses towards Mediterranean flowering plants using Y-tube olfactometer assays. Plant volatiles will be analyzed with gas-chromatography coupled with mass spectrometry (GC-MS), C) field observations on presence of natural enemies and pollinators on selected wild plants in field margins. Pollinator species and natural enemies will be collected and identified in the lab.
<i>What is the expected size of the data?</i>	Most of the data are expected to be around tens of gigabytes
<i>To whom might it be useful ('data utility')?</i>	The data might be of key importance for the ECOBOOST partners, but not only, and will be useful for scientific production (publications, meetings, etc.).

Partner 5 - SU/ISA	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	SU generates data related to 1/tomato- below ground microorganism's biodiversity pattern comparing to non cultivable/bulk soil 2/ Evaluation of the enzymatic activities in soil before and after amendments of selected local composts and biochar. Growth and yields parameters of tomatos under the different tested conditions. the selection of the best microbes candidates able to enhance

	soils biodiversity from local agricultural soils (Conventional/Organic)
<i>What types and formats of data will the project generate/collect?</i>	Mainly documents (doc, docx, pptx, etc.), illustrations (png, jpg, etc.), drawings (svg, dwg, etc) and raw data (csv, txt, xls, raw, etc.).
<i>Will you re-use any existing data and how?</i>	Data from scientific literature and technical reports from ECOBOOST partners,
<i>What is the origin of the data?</i>	Data originate from A) screening of soils from different conditions per partner country. B/Identification of key bacterial strains in the different soils C/ characterization of soil enzymatic activities before and after amendments of biochar and local composts
<i>What is the expected size of the data?</i>	Bacterial specific genomes sequencing data are expected to be around hundreds of gigabytes, Soil enzymatic activities related report are expected to be around 10 megabytes
<i>To whom might it be useful ('data utility')?</i>	The data might result of key importance for the ECOBOOST partners, but not only, and will be useful for scientific production (publications, meetings, etc.).

Partner 6 - UM5R	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	UM5R generates data related to A) the identification of the soil microbiome composition across the consortium countries to identify new microbes that play a key role in functional biodiversity by promoting plant health. B) the selection of the best microbes candidates able to enhance soils biodiversity. C) the inoculation of tomato plants with isolated and selected microbe (<i>in situ</i> and <i>ex situ</i>). D) The assessment of the effect of the tomato plants inoculation by measuring the growth and quality parameters
<i>What types and formats of data will the project generate/collect?</i>	Documents (doc, docx, pptx, etc.), illustrations (png, jpg, etc.), drawings (svg, dwg, etc) and DNA sequencing data (csv, txt, xls, raw, fastq, etc.).
<i>Will you re-use any existing data and how?</i>	Data from literature, reports, data from NCBI (gene bank), data of methodology of seeds coating with fungi (WP4)
<i>What is the origin of the data?</i>	Data originate from A) Soil microbiome assessment by High throughput sequencing and microbes isolation. B) test of the biological activities of the isolates (auxin, siderophores, phosphate solubilization ...) c) inoculations tests in pots (laboratory) and field (<i>in situ</i>). D) evaluation of the inoculation by measuring the growth parameters and chlorophyll content.

<i>What is the expected size of the data?</i>	Genes and genomes sequencing data are expected to be around hundreds of gigabytes
<i>To whom might it be useful ('data utility')?</i>	The data are of fundamental importance for some partners of the ECOBOOST consortium and can be useful for possible scientific production (publications, meetings, etc.).

Partner 7 – INRA	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	INRA will generate data on the production system and households of the platform. Data concerns structural and functional variables of the farmers producing eggplant, tomato and pepper. In principal baseline will survey the ECOBOOST team to characterize the production system, estimate cost and benefits of the three crops, and conduct some impact assessment
<i>What types and formats of data will the project generate/collect?</i>	Data generated will be in SPSS format or Excel format.
<i>Will you re-use any existing data and how?</i>	Secondary data will be used for national and regional analyses. These data will be collected at the ministry level
<i>What is the origin of the data?</i>	Surveys, reports, secondary data, time series
<i>What is the expected size of the data?</i>	The size of data depends on the sample that we are going to adopt
<i>To whom might it be useful ('data utility')?</i>	In general data will be used by all partners for specific crop management, economic and social analysis.

Partner 8 – IUAS	
<i>What is the purpose of the data collection/generation and its relation to the objectives of the project?</i>	IUAS will conduct farmer surveys to collect the required data for socio-economic analyses of ECOBOOST. IUAS will set up trials that will also serve for the demonstration and technological transfer of the new knowledge to end-users and will provide for ECOBOOST living lab sessions (WP 6) Data obtained from will be use farmer survey In order to understand under which conditions they are willing to accept agroecological practices in greenhouse tomato, pepper and eggplant production. These data will be analyzed for determining which socioeconomic variables are effective in the acceptability of farmers' agroecological practices. The application level of agroecological practices in small farmer conditions will be determined with the survey

	study to be conducted in target regions. In addition, socioeconomic factors that affect the acceptability of farmers' agro-ecological practices will be determined. (WP 7) Also, IUAS will obtain data from three greenhouse trial, namely tomato, pepper and eggplant. Data obtained from greenhouse trial will be used to determine the economics of agroecological practices (<i>bio- pesticide, pest-killing plants, beneficial insects, flowering plants, natural enemies (predators and parasitoids), pollinators and etc.,</i>) under farmer conditions by IUAS. Farmers who applied and do not apply (3 greenhouse traditional farmers) agroecological practices will be compared in terms of economic. (WP 7)
<i>What types and formats of data will the project generate/collect?</i>	Documents (doc, docx, pptx, xlsx. spo), illustrations and picture (png, jpg, etc.),
<i>Will you re-use any existing data and how?</i>	Data from scientific literature.
<i>What is the origin of the data?</i>	farmer surveys and greenhouse trials
<i>What is the expected size of the data?</i>	Approximately of the data are expected to be around hundreds of megabytes
<i>To whom might it be useful ('data utility')?</i>	The data might result of importance for the ECOBOOST partners and will be useful for scientific production (publications, meetings, etc.). Also, data will be useful for target farmers, agricultural extension services, policy makers, stakeholders and farmer associations etc.

2. FAIR Data.

This section relies on the standard H2020 template for FAIR data ("Guidelines on FAIR Data management in Horizon 2020" - European Commission, 2016).

2.1 Making data findable, including provisions for metadata

2.1.1 Discoverability/identifiability of data (metadata provision)

All the data will possess associated metadata documents (stored as a .txt file) which will describe the key aspects of the data.

2.1.2 Naming convention used for internal purposes

The naming of each document is established in this section, together with its proper management rules. An incorrect use of the naming convention and/or its application may generate unwanted mistakes and confusions, implying additional workload and costs. The naming of the documents may use the following fields:

- Document identifier
- Document name
- Time identifier, Partner, and Contributor
- Version

Each document is identified by a unique document identifier written with capital letters and depending on the type or nature of the document:

Identifier document type	
MM	Meeting Minutes
QR	WPL Quarterly Progress Report
PP	Project presentation
TS	Time sheet
D	Project deliverables
PMTR	Project mid-term Report
PFR	Project Final Report

The name of the document will be the description of the content, especially for documents type ID (indicating the issue treated in the document), MM (indicating the official name of the meeting) and PP (indicating the presentation name, or event/conference where the presentation is to be shown).

Each document has a time identifier that could be a month Mxx for a quarterly or a progress report (e.g., M12) or a date yy_mm_dd for a meeting minute, an internal peer review or a project presentation, (e.g. 17_02_05).

When necessary, the contributor identification will follow the format Partner- Contributor, composed by the partner short name as established in the project proposal and CA and it may be followed (to avoid confusion) by the contributor's identifier

The version identification of the document will follow the format vx.y, where x.y is the current version of the document and will be the same as the one stated in the front page of the document (e.g. v0.2 for a second version of a draft)

The owner is the only person that may generate the different versions of the document. Each contributor may generate additional working versions with different naming in order to avoid confusion, but these ones will not be considered as released versions.

As far as file-naming conventions is concerned, the following table presents the structure that should be followed when creating any new document of contractual or operational nature.

Document type	Convention	File name example
WPL quarterly report	QR_Month_WPnumber_Version	QR_M3_WP1_v1
Meeting Minute	MM_MeetingName_MeetingDate_Partner-Contributor_Version	MM_Kick Off_171022_UNIPA-AC_v1
Project Presentation	PP_Presentation/eventName_EventDate_Partner-Editor_Version	PP_7 th International Entomophagous insect conference _170423_UNIPA-AC_v1

2.1.3 Naming convention used for publicly available data

The convention above is convenient for the members of the ECOBOOST consortium. For all files made publicly available we will use a simplified convention: such files will reference ECOBOOST in

their name, with the recommendation that the convention ECOBOOST_XXXXXXX_[date], where “XXXXXXX” is a meaningful short description of the file.

Photographs and audio/visual recordings are be named ECOBOOST _[event]_[date of event]_[description of event/photograph content, e.g. workshop/dinner/group meeting].

We provide search keywords for optimizing possibilities for re-use

2.2 Making data openly accessible

2.2.1 Which data will be made openly available

The only data which will not be made openly accessible will be data which contains personally identifiable information (e.g., individual forms for socio-economics surveys) and data underlines deliverables that are covered by confidentiality.

The personal data processed in the project are not made publicly accessible but kept closed and inaccessible to third parties.

2.2.2 How the data will be made available

Data will be available to all the consortium partners via Zenodo (with the exception of individual questionnaires, which will be stored at each partner’s premises). The access to Zenodo repository for ECOBOOST is restricted to project partners. Should other individuals wish to access the data for research purposes during the project, it will be openly shared on request.

2.2.3 What methods or software tools are needed to access the data

Data will be published using standard file formats (txt, pdf, csv etc.). All the data will be accessed using standard tools. Software relevant to access the data would be made available, but it is not seen as being a requirement. Should it be needed, we will provide the required open source to access and analyse the data.

2.2.4 Where data and associated metadata, documentation and code are deposited

During the project, the personal data will be stored on local secured server of the partner responsible of taking care of this. Data to be public are stored on Zenodo. Sequencing data will be deposited into the European Nucleotide Archive, which is responsible for making the world’s public sequencing data freely available (<https://www.ebi.ac.uk/ena>)

2.4 Making data interoperable

In order to make data interoperable as much as possible the following practices are currently implemented: 1) Use Standard Data Formats (for a list of data format, see section “Data Summary”) which avoid the use of proprietary or custom formats that may limit compatibility; 2) use naming convention defined for both internal purposes as well as for publicly available data; 3) use of Digital Object Identifiers (DOIs) for datasets.

2.4.1 What data and metadata vocabularies, standards to be used to facilitate interoperability

In the project, standard file formats as noted above will be adopted. Each dataset will possess standard Zenodo metadata.

2.5 Making data re-usable

2.5.1 How the data will be licensed to permit the widest reuse possible

CC Licenses will be used for all the data to be preserved.

2.5.2 When the data will be made available for re-use

Data for reuse will be made available at specific timelines such as: 1) upon publication of results: in this case the owner of the data will be responsible for depositing the data in a public repository or attaching it as supplementary material to the publication; 2) after project completion: once the project has officially ended, data should be cleaned, documented, and archived for reuse, even if there was no formal publication. Additional aspects that could be considered for the timeline of data sharing refer to: 3) compliance with national found bodies and/or institutional requirements or 4) ethical considerations that allow for safe sharing. Data will be also available 5) upon request, when there are no legal or ethical barriers, such as confidentiality or data sensitivity.

2.5.3 Whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project

All Personal Identifiable Information will be restricted to internal usage and not going to be shared with third parties. For shared information, standard format, open source software, and proper documentation will guarantee re-usability by third parties.

2.5.4 Data quality assurance processes

A quality assurance process will be pursued within the duration of the project.

2.5.5 Length of time for which the data will remain re-usable

10 years after the end of the project.

3. Allocation of Resources.

3.1 Costs for making ECOBOOST data FAIR

In principle, costs for data curation and preservation will be covered by the partner(s) producing the data. However, cost for making data fair is estimated to be zero. Zenodo is funded by the EU, which means archiving data in this repository will be free of charge. However, repositories of specific institutions will not be free. In these circumstances, consortium partners will use their own budgets to archive project's research data in their own repositories.

3.2 Responsibilities for data management in the project

Each beneficiary leading Work Packages is responsible for preparing the datasets to make FAIR the data collected within its own activities, provided that AC, as Project Coordinator is co-responsible for general coordination and supervision. Furthermore, consortium partners have the responsibility to make sure their activities are in line with all applicable local, government and international laws, regulations, and guidelines.

4. Ethical aspects

ECOBOOST has considered the ethical aspects of the proposed work. ECOBOOST plan an exchange of material between EU countries (Italy, Greece) and non-EU countries (Algeria, Tunisia, Morocco) to carry out the following tasks: germplasm collection (T2.1), botanical surveys (T3.1) and soil sample collections (T5.1). All the procedures related to plant and soil sample collections will be performed according to the EU legislation or own countries' legislation for non-EU countries. Turkey was initially involved in such exchange material efforts as written in the proposal but was then excluded from such tasks (T2.1, T3.1, T5.1) during the development of the consortium agreement due to the restriction imposed by the Turkish legislation. Each partner will oversee ethical requirements and standards during the project in their own country. ECOBOOST will not work with human participants. Any personal data collected for Multi-Stakeholder innovation platforms (MIP) and household surveys will be done following strict precautions for data protection and use (adhering to GDPR regulations) and the appropriate disclaimers will be used. Animal species involved are invertebrates (arthropods: insects and mites), not endangered and commonly found in agricultural settings in the countries of the consortium