

D4.2: Strategy: Data Management Plan

Roland Pieruschka, Sven Fahrner, Francois Tardieu, Björn Usadel | 20 December 2019



Document information

EU Project N°	739514	Acronym	EMPHASIS-PREP
Full title	Preparation for EMPHASIS: European Infrastructure for multi-scale Plant Phenomics and Simulation for food security in a changing climate		
Project website	emphasis.plant-phenotyping.eu		

Deliverable	N°	D4.4	Title	Strategy: Data Management Plan
Work Package	N°	4	Title	Information system and imaging workflows

Date of delivery	Contractual	31/12/2019	Actual	31/12/2019 (Month 36)
Dissemination level	<input checked="" type="checkbox"/>	PU Public, fully open, e.g. web		
	<input type="checkbox"/>	CO Confidential, restricted under conditions set out in Model Grant Agreement		
	<input type="checkbox"/>	CI Classified, information as referred to in Commission Decision 2001/844/EC.		

Authors (Partner)				
Responsible author	Name		Email	

Version log			
Issue Date	Revision N°	Author	Change

This project has received funding from the European Union's Horizon 2020 Coordination and support action programme under grant agreement No 739514. This publication reflects only the view of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

Contents

Document information	3
Executive Summary.....	5
Objectives.....	5
Main results	5
1. The goal of the Data Management Plan.....	6
2. Data description.....	6
3. Data publication	8
4. Legal status of data	8
Annex 1: Check list	9

Documents used in the preparation of this deliverable:

- EPPN2020 Data management plan: https://eppn2020.plant-phenotyping.eu/Data_Policy
- Data management H2020: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Executive Summary

Objectives

EMPHASIS should establish rules on data property, sharing and right of the first use, including necessary metadata for fair sharing in accordance to the EU policy of open data within the data management plan. This involves the property of data, duration of confidentiality for data involving private companies, principles that address consortium agreements in EU or national projects and of novel methods and techniques.

Main results

An operational Data Management Plan (DMP) is currently implemented within the related EPPN2020 project providing transnational Access and addressing data management. Within EMPHASIS a DMP which can serve as a template for further partner activities will be elaborated to reduce the loss of data and increase the availability and reusability of data according to the legal and ethical standards. In general plant phenotyping data are generated during the research process, that are used for it, or that are the result of the research process and may occur in different media types and formats (as numerical data, images, documents, texts etc.) during the data life cycle. The metadata in this context are important for the re-use of research data ensuring also the possible verification of research results by third parties addressed by Minimum Information on Plant Phenotyping Experiments (MIAPPE). In general research data should be published and EMPHASIS will strongly encourage that data should be published in a way that access to these data is enabled and possible, e.g. via a web interface or a citable data publication.

Research activities between two or multiple partners will be supported by agreements including data management issues, currently a template of such an agreement is used within the EPPN2020 projects and will be adapted for EMPHASIS.

1. The goal of the Data Management Plan

A Data Management Plan (DMP) describes the management of datasets generated and processed during and after the use of EMPHASIS services such as access to plant phenotyping facilities to perform dedicated experiments. As EMPHASIS goes beyond a simple H2020 project, there are specific requirements and sustainability issues that need to be addressed. This DMP shall develop a general framework to help manage data, meet funder requirements, and facilitate multiple use of data by the scientific community. As such, the DMP ranges from data acquisition to the evaluation, processing up to archiving and publication of the data, where appropriate. The goal is to make these data “findable, accessible, interoperable and reusable” (FAIR), which is becoming increasingly important in many research projects, requiring a proof of structured data management embedded in a sound data policy during and after completion of a research project. Specifically, the ERC Scientific Council recommends to *“all its funded researchers that they follow best practice by retaining files of all the research data they have produced and used during the course of their work, and that they be prepared to share these data with other researchers whenever they are not bound by copyright restrictions, confidentiality requirements, or contractual clauses.”*¹

As such a DMP should:

- reduce the risk of data loss
- make data available and reusable
- promote the implementation of ethical standards and principles of Good Scientific Practice
- create legal certainty
- improve data exchange within research groups

Practical implementation of a DMP in plant phenotyping and as such highly relevant for EMPHASIS has been performed in the I3 project EPPN2020: https://eppn2020.plant-phenotyping.eu/Data_Policy

2. Data description

The general aim is that data are interoperable between platforms and that querying mechanisms are standardized. All platforms will collect data in such a way that this is eventually possible, and tools will be progressively deployed in the project sites. The scientific managers of phenotyping facilities shall be responsible for managing the data and ensure that the DMP is carried out.

¹ https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

Research data, in very general terms, refer to all data that are generated during the research process, that are used for it, or that are the result of the research process. Depending on the specific research question and methods used, data are generated, obtained or collected, observed, simulated, derived, validated, processed, analysed, published and finally archived. These data therefore may occur in different media types and formats, the aggregation and quality of these data depends on the stage in the life cycle of the data.

Specifically, data collected in phenotyping experiments can be numerical data, images, documents, texts or manual measurements. The data need to be complemented by metadata that describe the research data. This may include includes authorship, contact data, time of creation, it defines the data formats used and contains the context that led to the data. Additionally, metadata should include settings of experimental sensors or instruments, environmental conditions, comments, and measurement uncertainties etc. The metadata is essential for the re-use of research data ensuring also the possible verification of research results by third parties.

In plant phenotyping the Minimum Information About a Plant Phenotyping Experiment (MIAPPE) comprises both a conceptual checklist of metadata required to adequately describe a plant phenotyping experiment, and software to validate, store and disseminate MIAPPE-compliant data.² Within EMPHASIS, we strongly encourage to follow the recommendations outlined in MIAPPE.

Examples of data categories and sources that need to be managed in a phenotyping experiment are:

- Resource description: information of genotype, seed, accession
- Facilities: installations, sensors, cameras, conveyors and specific devices
- Trait recovery work-flows: sensor and image analysis methods and software tools used to extract traits from raw image and other data
- Phenotypic data at plant or population level, e.g., image-based traits, phenological stages, manual measurements
- Environmental conditions as collected by sensors (e.g. soil water status, air temperature or evaporative demand)
- Date and description of management events (e.g. irrigation, pruning, sampling) and of observations (e.g. plant disease, accidents)
- Characteristics of experiments e.g. design, protocol and organisation

² <https://www.miappe.org/>

3. Data publication

Basic utilisation of research data is a publication, specifically in the academic context. Data are exploited and published by the scientists responsible for them, unless there are substantial grounds for not doing so. This first use of data is usually a publication in a specific journal, the extent and duration of the first exploitation can be determined in the DMP and the end of the first exploitation phase should be context-dependent and justified. For example within the EPPN2020 project, “data will be published whenever possible, and made available either after publication or at the latest three years after the last day of the experiment.”³

Within EMPHASIS we will strongly encourage that data should be published in a way that access to these data is enabled and possible, e.g., via a web interface or a citable data publication. An important technical tool for publishing research data is a data repository, a server service on which data can be uploaded by the data creators. This data should get a worldwide unique identifier (e.g. a DOI) and can be searched and downloaded. There is a number of different repositories e.g.:

- topic specific repository (esp. for plant science and phenotyping), e.g. e!DAL⁴
- generic repository, e.g. zenodo⁵, dryad⁶

4. Legal status of data

Research data may have an intellectual level of creation worth of protection. In order to ensure legal certainty, it should always be assumed that protection is important, i.e. that usage and exploitation rights are clarified contractually with external partners. EMPHASIS will strongly support the establishment of contractual agreements in research projects to ensure fair exploitation of data. Currently, there is a template for an agreement for bilateral project within the I3 project EPPN2020 to be used within the Transnational Access and to clearly define the roles as well as rules for data use and reuse by the access provider and user.

³ https://eppn2020.plant-phenotyping.eu/Data_Policy

⁴ <https://edal.ipk-gatersleben.de/index.html>

⁵ <https://zenodo.org/>

⁶ <https://datadryad.org/stash>

Annex 1: Check list

Deliverable Check list (to be checked by the “Deliverable leader”)

	Check list	Comments
Before	I have checked the due date and have planned completion in due time	<i>Please inform Management Team of any foreseen delays</i>
	The title corresponds to the title in the DOW	<i>If not please inform the Management Team with justification</i>
	The dissemination level corresponds to that indicated in the DOW	
	The contributors (authors) correspond to those indicated in the DOW	
	The Table of Contents has been validated with the Activity Leader	<i>Please validate the Table of Content with your Activity Leader before drafting the deliverable</i>
	I am using the EMPHASIS deliverable template (title page, styles etc.)	<i>Available in “New EMPHASIS Logo, Templates, CI” on the collaborative workspace</i>
The draft is ready		
After	I have written a good summary at the beginning of the Deliverable	<i>A 1-2 pages max. summary is mandatory (not formal but really informative on the content of the Deliverable)</i>
	The deliverable has been reviewed by all contributors (authors)	<i>Make sure all contributors have reviewed and approved the final version of the deliverable. You should leave sufficient time for this validation.</i>
	I have done a spell check and verified the English	
	I have sent the final version to the WP Leader and to the Project coordinator (cc to the project manager) for approval	<i>Send the final draft to your WPLLeader and the coordinator with cc to the project manager on the 1st day of the due month and leave 2 weeks for feedback. Inform the reviewer of the changes (if any) you have made to address their comments. Once validated by the 2 reviewers and the coordinator, send the final version to the Project Manager who will then submit it to the EC.</i>