D2.4. Gap analysis

Authors Merlijn Morisse, Darren Wells, Pierre-Etienne Alary, Jean-Eudes Hollebecq, Clement Saint Cast, Roland Pieruschka, Stijn Dhondt

21.11.2019

EUROPEAN INFRASTRUCTURE FOR PLANT PHENOTYPING

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°**	D2.4	**Title**	**Gap analysis**
Work Package	**N°**	WP2	**Title**	Infrastructure setup, operation and funding

Date of delivery	**Contractual**	30/06/2019	**Actual**	DAY/MONTH/YEAR (Month 36)
Dissemination level	x	**PU Public, fully open, e.g. web**		
		CO Confidential, restricted under conditions set out in Model Grant Agreement		
		CI Classified, information as referred to in Commission Decision 2001/844/EC.		

Authors (Partner)				
Responsible author	**Name**	Merlijn Morisse	**Email**	Merlijn.morisse@psb.vib-ugent.be

Version log

Issue Date	Revision N°	Author	Change
14/10/2019	1	Merlijn Morisse, Darren Wells, Pierre-Etienne Alary, Roland Pieruschka, Jean-Eudes Hollebecq	first version provided for review
17/12/2019	2	Stijn Dhondt	Draft review
18/12/19	3	Merlijn Morisse	Finalizing report

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.

Executive Summary

Objectives

The objective of EMPHASIS is to develop a long term, distributed, pan-European infrastructure equipped with state-of-the-art plant phenotyping experimental installations, which provide access to user community and enable excellent science that aims to improve crop performance and address future grand challenges. There is an increasing demand for use of phenotyping infrastructure which requires different infrastructure categories as described in the deliverable D2.1. Criteria list for plant phenotyping infrastructure. Access to phenotyping facilities and services requires a coordinated infrastructure, linked with an integrated data management system for storing and analysing (meta) data, and with modelling platforms associated with the phenotyping platforms. To be able to provide services related to plant phenotyping infrastructure considering the comparability and/or differences between installations, it is essential to map the new and existing plant phenotyping facilities that use non-destructive, image-analysis based determination of the phenotype of plants and allow for a characterization of plant traits. This mapping has been performed and the results are summarized in the EMPHASIS-PREP deliverable D2.3: mapping of existing and upcoming infrastructures.

This deliverable, D2.4, will extensively evaluate the gaps and limitations based on the mapping activities. Moreover, where possible, the gaps will be strategically addressed by presenting in what perspective EMPHASIS could tackle these gaps and enable services to the benefit of excellent plant phenotyping science in Europe.

Rationale

Both the mapping and the analysis of gaps has been done by EMPHASIS-PREP partners in extensive collaborations and discussions with the plant phenotyping community in Europe driven by four regional and three topical workshops, two surveys, and workshops during the annual support group meetings. During the workshops, breakout sessions were organised to extensively discuss about the gaps of plant phenotyping in Europe and how to address these limitations. Furthermore, EMPHASIS-PREP developed two surveys, which had the purpose to assess the phenotyping landscape, and included and dedicated questions on how the participants see the future of phenotyping, what gaps there identify and how to tackle these gaps. The development of a virtual map in the scope of the mapping deliverable was very effective to find regional gaps for specific plant phenotyping infrastructure in Europe.

Main Results:

1. Gaps in the mapping data extraction

We mapped the plant phenotyping infrastructures in Europe in details (Deliverable 2.3) and were able to identify limitations that need to be addressed in the future. The mapping involved the plant phenotyping community from academia and industry at large but the results may be biased for installations from academia. Additional efforts to evaluate plant phenotyping within the industry sector may be required.

Moreover, countries with a well-established national networks have a higher number of installations in the database. It could be assumed that well-connected national communities can be easier addressed and provide feedback on their installations in contrast to facilities that are in the process of establishing national communities. Thus further and continuous evaluation of the dynamic phenotyping landscape will be required and an important tool for service provision by EMPHASIS.

2. Gaps in harmonizing and innovating plant phenotyping infrastructure

The need for a collective harmonisation of protocols and experimental design strategies to ensure reproductive and interoperable data was clearly expressed. Moreover, innovation through knowledge and technology transfer is very much needed to ensure qualitative sustainable phenotyping. To address these limitations a harmonisation and innovation pilots were initiated and are under development (Deliverable 6.3).

3. Gaps in plant phenotyping categories

Controlled condition plant phenotyping

The survey indicated that there are gaps in geographical location of installations under controlled conditions. There are further limitation of required throughput rates to address mapping populations potentially driven by limited automation and the use of more advanced sensor technologies. The capacity for root phenotyping is less-developed than for shoot and canopy studies.

Intensive field

Geographical gaps have also been identified for intensive fields. There was a limited response with respect to intensive field and further identification and characterisation of intensive field sites is required.

Networks of lean field phenotyping

Networks of fields are very difficult to map. Some institutes own land for field experiments, while others have annual renting opportunities with local land owners. The latter causes the overall total of fields in the database to be substantially underestimated. Although minimal equipment, like UAVs for imaging analysis in lean field phenotyping is more commonly used, there appears to be a limitation with respect to the analysis of these imaging data. Multi-climatic trials across Europe are essential to breed for new crop varieties, needed to cope with the changing climate. Access procedure to networks field sites is rather challenging needs to be developed, a field pilot that is under development (Deliverable 6.3) will support this development.

Modelling

The plant model database populated through the review of the literature indicated that there are gaps in the plant models inventoried (e.g. models developed by industry are not published) and the plant model availability (i.e. not directly available). Moreover, the diversity observed between the plant models will constitute challenges to improve the interoperability between the phenomics data and the plant models through a unique portal. EMPHASIS, through the next modelling pilot that is under development (Deliverable 6.3), will encourage the modelling community to adopt and develop harmonized standards for interfacing models with phenomics data sources.

Data management systems

The survey revealed that a high number of local infrastructures are not using data management system that will allow interoperability of data. Thus, FAIR principles are not applied yet in many information systems. Several steps are required before these principles can be effective, enabling interoperability. The data pilot (Deliverable 6.3) that is under development will provide users some tools and standards to enable interoperability.

Contents

Document information 3

Executive Summary 4

Objectives 4

Rationale 4

Main Results: 5

1. Gaps in the mapping data extraction 5
2. Gaps in harmonizing and innovating plant phenotyping infrastructure 5
3. Gaps in plant phenotyping categories 5

Controlled condition plant phenotyping 5

Intensive field 5

Networks of lean field phenotyping 5

Modelling 6

Data management systems 6

The Plant Phenotyping infrastructure landscape in Europe: gap analysis 9

Gaps in the mapping data extraction 9

Gaps in harmonising and innovating plant phenotyping infrastructure 10

Gaps in the different plant phenotyping pillars 11

1. Phenotyping installations in (semi-)controlled conditions for high-resolution and high-throughput phenomics. 11

Gap analysis - geographical distribution 11

Phenotyping focus 12

Automation and throughput 12

Sensors 13

2. Intensive field experimental sites for high throughput phenomics 14
3. Field experiments using minimal equipment, linked as a network of fields 15

Gap analysis - geographical distribution 16

Field phenotyping equipment 17

Multi-climatic fields networks 18

Breeding companies 19

4. Modelling 21

Model findability and geographical distribution 21

Model diversity, interoperability and harmonization 23

5. Data management systems 23

Conclusion, discussion and next steps 25

Glossary 27

Annex 1: Check list 29

The Plant Phenotyping infrastructure landscape in Europe: gap analysis

Gaps in the mapping data extraction

In the deliverable D2.3: “mapping the existing and upcoming plant phenotyping infrastructures”, EMPHASIS-PREP identified 182 plant phenotyping installations. The details are stored in the EMPHASIS-PREP database with a virtual map that indicates their locations of the installations in Europe (https://emphasis.plant-phenotyping.eu/database). A separate database was built for the modeling installations and is also on the EMPHASIS website (https://emphasis.plant-phenotyping.eu/modelling).

The approach to find and analyze gaps started with these mapping activities, in order to first identify currently established or upcoming installations in Europe. Both the mapping and the analysis of gaps has been done by EMPHASIS-PREP partners in extensive collaborations and discussions with the national plant phenotyping community in Europe by using surveys and workshops. Four regional workshops in different regions of Europe and three topical workshops have been organized. During these workshops in dedicated breakout sessions, we discussed extensively the phenotyping landscape including opportunities and limitations. Additionally, during the annual support group (SG) meetings EMPHASIS-PREP organized, breakout sessions were also used to discuss about plant phenotyping in Europe and the potential gaps.

Furthermore, EMPHASIS-PREP developed two surveys, which aimed at identifying gaps and how to tackle these gaps.

The development of a virtual map, in the scope of the mapping, was very effective to find regional gaps for specific plant phenotyping infrastructure categories in Europe.

Although the efforts to map the plant phenotyping research activities in Europe were effective, it needs to be noted that the approach, using surveys and organizing workshops has some limitations. We addressed mostly participants from academia, while the industry was underrepresented (see Figure 1). Moreover, some countries, with active national communities, were better represented in the workshops and surveys then other countries.

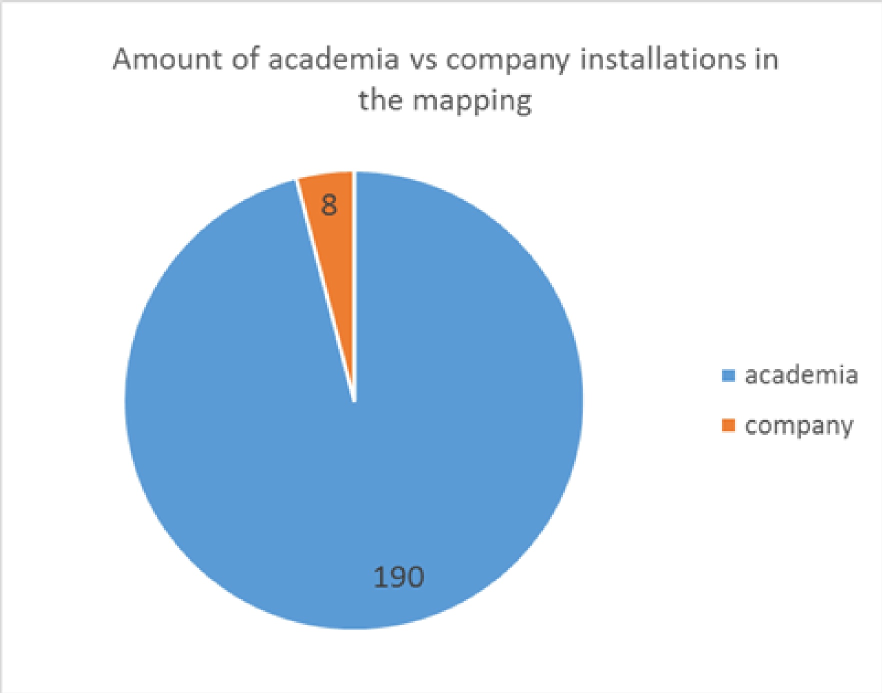


Figure 1: Amount of academia versus company installations in the mapping database (November 2019)

The maturity of the national phenotyping communities, national networks, and the involvement in I3 projects such as EPPN and EPPN2020 also played a role in providing feedback on installation details. Installations from countries that have national phenotyping networks in place more input, for example the surveys, was provided, indicating well established communication channels. As a consequence of these limitations in the mapping, it needs to be recognized that the mapping results, as described in the deliverable D2.3. "Mapping upcoming and existing infrastructures", may be biased. This gap has to be tackled by mapping the phenotyping installations and activities as a continuing process in the future of EMPHASIS operations. Therefore, an updating process of the mapping should be established. Moreover, the input and involvement of companies could be improved by e.g. organizing specific forums for SME's (see below innovation pilot).

Gaps in harmonising and innovating plant phenotyping infrastructure

It has been identified that to ensure long-term sustainability with high quality of services such as access to the facilities, generation and utilization of data, EMPHASIS must encourage innovation in all aspects of plant phenotyping technology. Moreover, harmonizing the installations, not only in perspective of the generated data through the FAIR principals, but also in the experimental design through protocol harmonization. The community expressed their needs for a collective harmonization of protocols of experimental design to ensure reproductive and interoperable data. This lead to the development of harmonization and innovation pilot services in EMPHASIS-PREP. The pilots will be described in more details in the Deliverable 6.3.

Gaps in the different plant phenotyping pillars

1. Phenotyping installations in (semi-)controlled conditions for high-resolution and high-throughput phenomics.

Phenotyping installations under controlled conditions allow the investigation of the variability of measured plant traits as a response to well-defined and monitored environmental conditions with a capacity of several hundreds to thousands of plants. Facilities may also be linked to high precision platforms for deep phenotyping with lower throughput (tens to hundreds of plants) with measurements over shorter timescales (weeks) and time steps (minutes to hours). For detailed definitions, see Deliverable 2.1; https://emphasis.plant-phenotyping.eu/Infrastructure_Categories.

As part of the mapping in EMPHASIS (see Deliverable 2.3 document for details: "List of existing/upcoming infrastructures") a total of 112 installations in 19 countries were identified as phenotyping installations under controlled conditions (Figure 1, data from https://emphasis.plant-phenotyping.eu/database.). 109 are based in academia, three in industry.

Gap analysis - geographical distribution

66 (59%) of these installations are located in three countries (Belgium, Germany and the UK, see Table 1 and Figure 1). This most likely represents of the missing information from new and emerging installations that are not yet well linked on the national and European level, rather than the actual status. This will be addressed in future updates of the phenotyping landscape. Similarly, the number of installations from industry (3 of 112) is probably not representative and should also be addressed in future surveys (or by a targeted activities to address industry).

Country	***Number of installations***
Austria	4
Belgium	20
Czech Republic	3
Denmark	4
Estonia	2
Finland	3
France	8
Germany	20
Greece	1
Hungary	1
Ireland	1
Israel	1
Italy	3
Netherlands	4
Norway	3
Slovakia	1
Spain	2
Sweden	5
UK	26

Table 1: Number of controlled condition installations in European countries.

Figure 2: Pan-European plant phenotyping installations under controlled conditions. Data from emphasis.plant-phenotyping.eu/database.

Phenotyping focus

The majority of installations focus on canopy and shoot measurements, with ~10% focusing on roots (and ~10% on both). Nearly a third of the respondents in the 2018 EMPHASIS survey rated “root phenotyping” as the largest challenge facing plant phenotyping in the future. This indicates a demand for root phenotyping that is currently not met.

Automation and throughput

Surprisingly, nearly a third of the installations reported manual phenotyping, potentially limiting the throughput. Moreover, nearly 50% of installations reported a capacity between 100 and 500 plants per experiment. Upgrading the phenotyping installations with automated systems may substantially increase the throughput. Annual throughput varied from 25 to 19000 plants per year, with ~30% of respondents reporting throughput in the range 1000-5000 plants per year. This annual throughput can be increased either by increasing the numbers of plants per experiment or by running more experiments each year. As the latter is often set by biological considerations, a focus on increasing throughput will be either increasing the number of plants per experiment, which may be limited by the capacity of the installation, or by increasing the number of installations.

Sensors

A relatively low number of installations (31) reported about the specific sensors and thus it is difficult to draw firm conclusions. However, RGB cameras were the most common sensor reported indicating that advanced imaging systems (multispectral, hyperspectral, 3D) are not widely implemented, which needs to be further evaluated.

2. Intensive field experimental sites for high throughput phenomics

Intensive field sites are installations set up to allow the detailed study in the field (natural light condition) of hundreds of plant micro-plots through frequent measures of several plant traits. Intensive field sites are highly equipped in order to monitor both the plants phenotype and the environment during the plant growth cycle. The quasi-continuous data acquisition paired with the storage of time courses enables the study of the plant growth dynamics, which can be analyzed with respect to the dynamics of environmental variables.

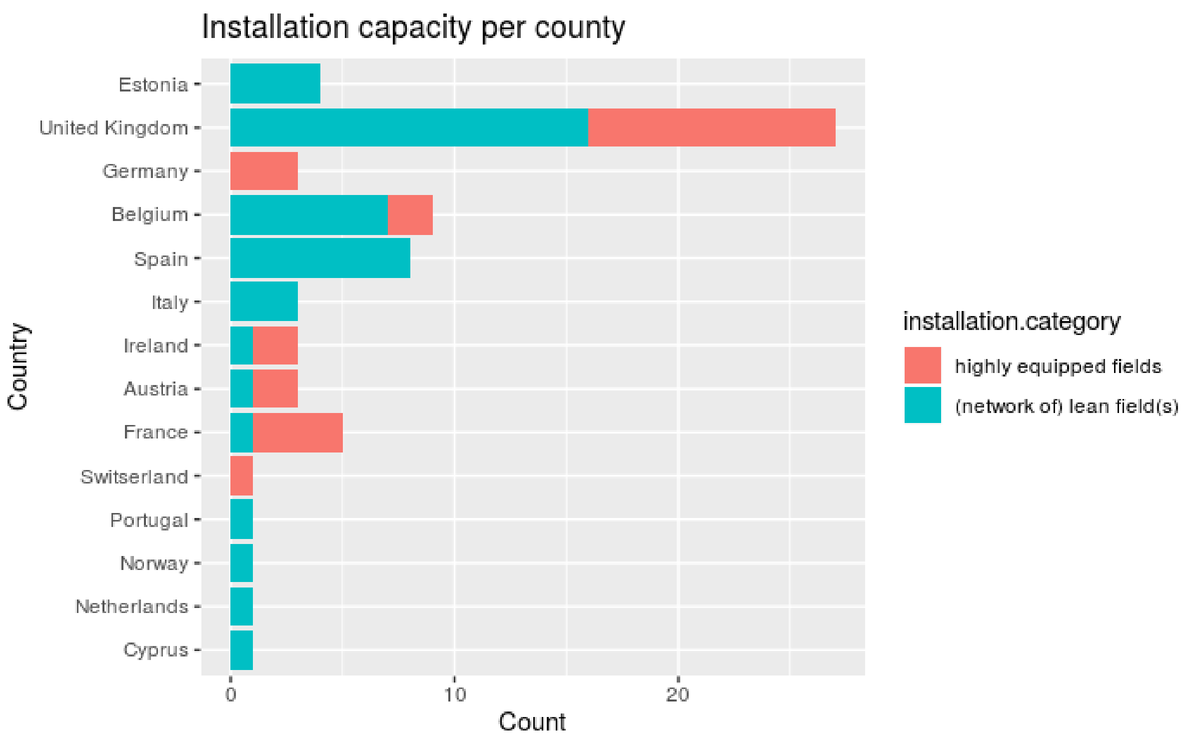


Figure 3: Field phenotyping installations per European country, split for networks of fields and highly equipped fields

The vast majority of the intensive field installations are found in the UK and only very little installations are found in the Netherlands or Spain (Figure 3). This is quite surprising and we can suspect some bias (e.g. due to the low feedback rate from providers of field experiments in different countries). The variety of actors is certainly underrepresented as well. A lot of small companies are using installations to perform experiments, but unfortunately did not answer our solicitation. As already stated, the incentives for private companies to share their information are limited and not very well perceived.

Regarding field sites, in general, it could be assumed that we have a lot of missing data. While the controlled environment installations are already well integrated e.g. by the EPPN and EPPN2020 projects that only addressed controlled environment facilities, field facilities are still very fragmented and not integrated within Europe. Thus, one of the large limitations is the integration of field phenotyping that has to be addressed by EMPHASIS.

3. Field experiments using minimal equipment, linked as a network of fields

Field phenotyping has been identified as one of the biggest challenges in the future of plant phenotyping, according to the EMPHASIS survey (2018).

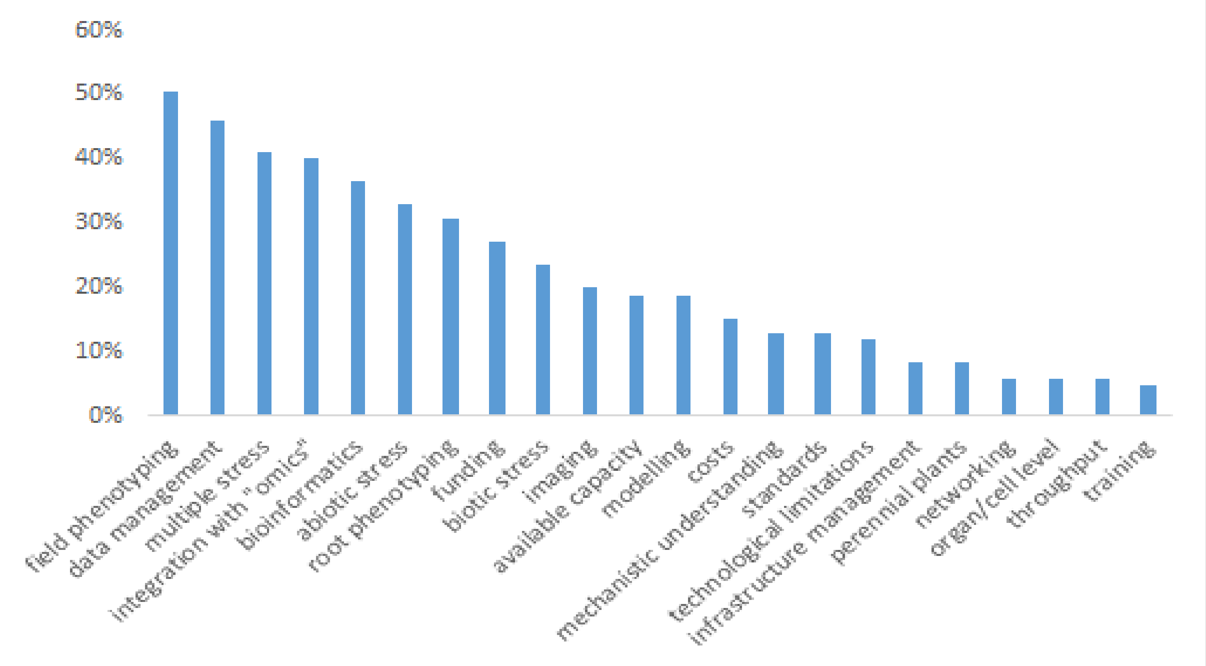


Figure 4: Figure xx: What do you think are the largest challenges for plant phenotyping in the future (83). From 2018 EMPHASIS Survey (320 respondents) the overall user demands for field trials was acquired.

Within plant phenotyping, field experimentation allows plant researchers to test genotypes of interest in agriculture-relevant environments and breeding-like conditions. A network of fields makes it possible to test a range of environmental scenarios as a response to differences in field management, soil composition, water, nutrient availability or other environmental conditions. Testing plants in different environmental conditions allows to monitor plants and their phenotypes to simulate environmental changes, like e.g. wet and drought conditions, in a changing climate. Thus, we mapped field phenotyping network stations across Europe that may be able to be part of a multi-site field experiments across an environmental gradient Europe.

Field phenotyping data is mostly produced by quantitative measurements of variables such as; seedling emergence, flowering time, plant height, biomass and yield components. Simple imaging techniques involving UAVs and, probably, satellite imaging in the near future, become increasingly available and can carry a variety of sensors going from RGB to thermal to hyperspectral cameras.

Gap analysis - geographical distribution

The mapping of the lean field sites in Europe resulted 45 sites in 12 countries (see figure 5 and table 2). These fields are organized a local network and may become part of a multi-site and multi-climatic network of fields for large scale experiments along environmental gradients. A number of field sides may not be included in this mapping because of a large fragmentation of field sites focusing on local single sited experiments. It could be noted that there are certainly more similar fields used for field experiments, for example in breeding companies but also in academic institutions that we were not able to address. Most likely it is due to confidentiality reasons that companies have less to gain of sharing detailed information on their (field) phenotyping equipment or activities. Moreover, the current definition of network of lean fields excluded fields that are not institute owned field networks, but rented fields from for example farmers or kind of landlords.

Country	Number of network of field installations
Austria	1
Belgium	7
Cyprus	1
Estonia	4
France	1
Ireland	1
Italy	3
Netherlands	1
Norway	1
Portugal	1
Spain	8
United Kingdom	16

Table 2: Number of installations of lean field networks per country in Europe.

As stated before, it is essential for to tackle global challenges as climate change, to organize multi-location field sites, addressing different climatic region. Field trail networks with minimal equipment could also be used outside Europe and could be beneficial in the future of phenotyping, and this would then also be a scope of an EMPHASIS service in the future.

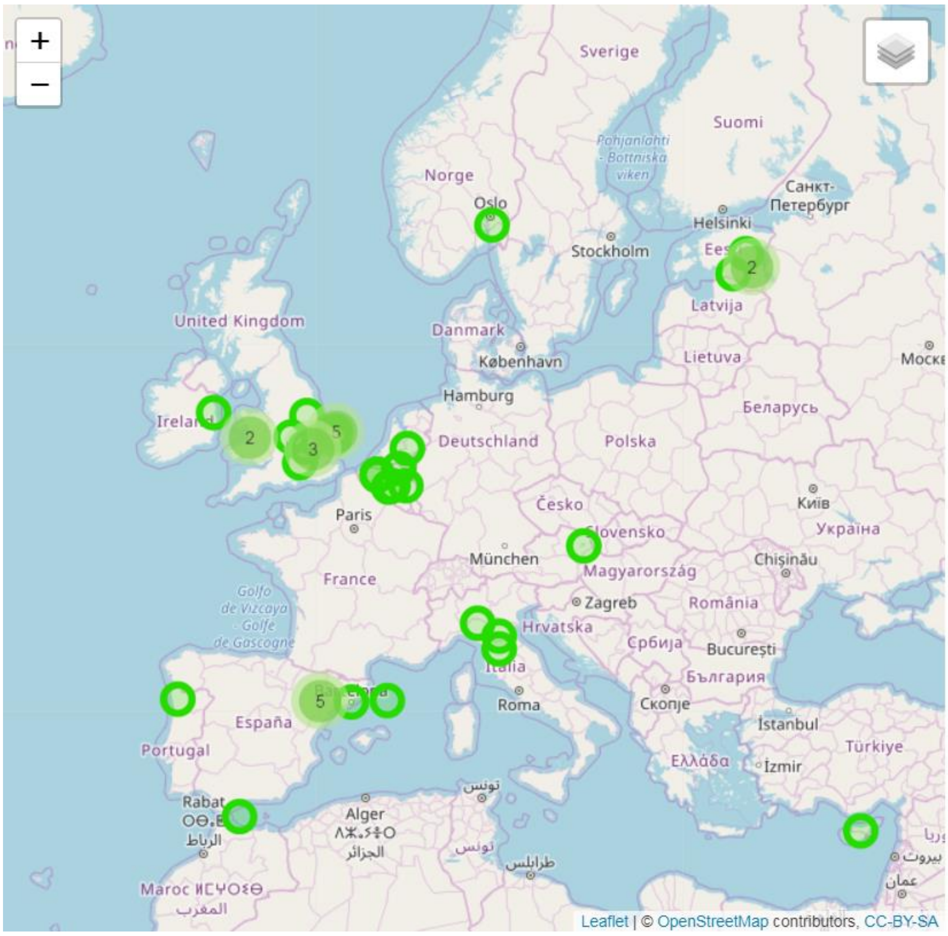


Figure 5: Distribution of networks of fields across Europe, mapped by EMPHASIS-PREP (status of November 2019)

Field phenotyping equipment

As also indicated by the EMPHASIS users in a survey launched in 2018, 129 participants out of 320 declared they currently used in their field site experiments low cost techniques to score morpho-physiological traits, with RGB, multispectral and thermal imaging based systems as key sensors in field phenotyping. Custom systems are also used, (see D2.2 https://emphasis.plant-phenotyping.eu/lw_resource/datapool/systemfiles/elements/files/165e15fe-0ae6-11ea-b1c5-dead53a91d31/live/document/EMPHASIS-PREP__D2.2_Criteria_list_user_demands.pdf).

Remote sensing equipment as UAVs become more largely available and prices are dropping which makes imaging in field phenotyping even more common, using a range of different sensors. Nonetheless, a gap seems to be formed in the field of image analysis of these UVA data. More information on this will be available in the report D4.4 Analysis of imaging approaches.

Multi-climatic fields networks

Field trial sites like experimental stations or experimental fields are essential research infrastructures for environmentally oriented agricultural sciences studying the interactions between cropping systems and the environment, and for plant sciences aiming to learn about the behavior of genetic material in natural environments. Multisite field phenotyping experiments are key to tackle global challenges such as climate change and food security, as it allows testing genotypes of important crops in different climatic conditions. Breeding companies are investing substantially in multi-climatic field experiments to test new crop varieties. However, most industries hold the data confidential for profit motives. Field sites of the public sector are more distributed and currently it is a struggle to organize long-term sustainable and accessible multi-climatic networks of fields. These are currently mainly organized through bilateral collaborations with in kind contributions or in EU funded projects with no long term vision. The administrative process of doing a multisite experiment can be demotivating for scientists to even start preparing such a proposal. An extensive map of field sites and phenotyping equipment can make a huge difference in that perspective. Moreover, on societal level it could be financially beneficial that the administration of bilateral agreements does not need to be done over again and again.

However, and despite the difficulties, multi-climatic field projects across Europe have been established before with success. EMPHASIS-PREP was able to map academic and private research projects that lay on the use of network of fields as summarized in Table 3, (See table in Deliverable D 2.3 for further details). The two examples below show the relevance of multi-climatic field phenotyping experiments:

- the FP7 DROPS project (July 2010 - December 2015) testing maize, wheat and Sorghum in multi-climatic regions in Europe and beyond for drought tolerant traits, as seed abortion, maintenance of vegetative growth, root-system architecture and transpiration efficiency. (www.drops-project.eu)

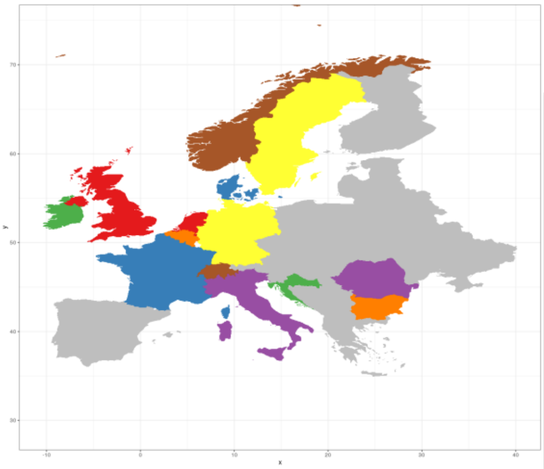
- European Consortium for Open Field Experimentation (ECOFE), stretching from Scandinavia to the Mediterranean, and from Ireland to the eastern border of the EU, which allows European scientists to access a platform for collaboration with agronomy focus while providing them a competitive advantage. A more detailed outline of the ECOFE concept can be found on the website www.ecofe.eu [Stützel, Brüggemann, Inzé, 2016].

Acronim of the project	Name of the field networks	Website	Topic	Species	# Sites	Sites locations	Approach
Public research institutions							
ECOFE	European Consortium for Open Field Experimentation	https://www.ecofe.eu/	agricultural sciences	wheat, maize	12	Europe	networking existing field stations across Europe
ALTER-NET	A Long-Term Biodiversity, Ecosystem and Awareness Research Network	ttp://www.alter-net.info/ms	ecological and socio-ecological research			Europe	calls to co-financed diverse multi-site research (MSR) activities simultaneously conducted across Europe with a significant added value due to its pan-European character
Research trials	Author: Canè et al 2014	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4257993/	Root QTL analyses	wheat	5	Europe , North Africa	
ITEX	International Tundra Experiment	https://www.gvsu.edu/itex/	impacts of warming on tundra ecosystems				They ensure multiple locations but setting standars to obtain the warming of the
Private companies							
SOLVAY	STIMUL project	https://www.solvay.com/en/innovation/open-innovation/european-life-projects/life-stimul-project/organization-2018-field	Organization of field trials for biostimulants application	sunflower, beetroot, soybean, corn, rapeseed	91	Europe	
Eurofins	Eurofins agroscience services	https://www.eurofins.com/agroscience-services/	soil and crop health, fertilisation, feed value and food safety				
Agrolab		http://www.agrolab.dk/?pag	provide within the Regulatory landscape and Field phase development of chemicals, including pesticides, bio pesticides or biocides			EU North zone	GEP accreditation, can arrange pan-european multi site trials
SGS		https://www.sgs.be/en/agriculture-food/seed-and-crop/contract-research-services/field-trials/glp-field-trials					GEP and GLP field trials
Charles river laboratories		https://www.criver.com/sites/default/files/resource-files/Field_Trials.pdf	chemicals, agrochemicals, bio pesticides			north and south europe	GLP and GEP accredited field bases

Table 3: Mapping of private and public networks of fields in Europe.

Breeding companies

52 of the European EMPHASIS stakeholders belonging to private sector (See the map in Figure 6) make use of field trials for their activities in European countries.



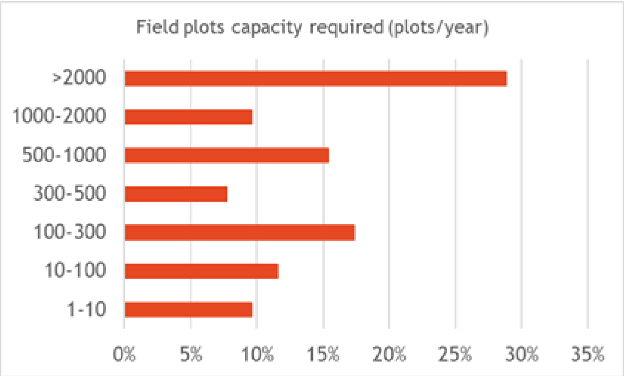


Figure 6: European EMPHASIS industry stakeholders using field trails and field networks (left). Field capacity requirements from private companies (Right).

54% of them require a field capacity ranging from 500 to 2000 plots a year for their activities, and 29% of them more than 2000 plots a year (Figure 6).

These data are consistent with the categories of private companies reached with the surveys (Figure 7) that account for plant breeders as the major representative of private industries interested in plant phenotyping.

Figure 7: Field of work of the private companies involved and interest in field phenotyping

Whether access to the fields is received in kind, rented from selected farmers, or owned by the companies, as well as characteristics like the size of the fields, is not available from the survey and from the available literature. This currently represents a big gap in the field phenotyping mapping.

While the picture of public research fields is clear and easily accessible, the fields and field networks owned by industry (e.g. in breeding companies) are usually not published, likely for confidentiality reasons. EMPHASIS operations should in the future establish an even better relationship with private partners as breeders to be able to innovate both private and public sector and to tackle the gap of multi-climatic field phenotyping research in the scope of a changing climate.

4. Modelling

The plant phenotyping space is too immense to observe all combinations of genotypes by growth stage by environmental conditions. Therefore, capturing the essence of the observed phenomics data in models turns out as a pivotal approach. Feeding phenome data into structural plant models (SPMs), functional-structural plant models (FSPMs) and process-based crop simulation models (CSMs) is a way to derive predictions of integrated (e.g. yield) or functional traits (e.g. root system architecture) for existing or new genotypes and across a wide range of target environments or management practices.

The identified models are summarized in a database and presented in a website, quantitative-plant.org, with each model concisely in a consistent framework. Aa total of 116 plant models (11 SPMs, 34 FSPMs and 71 CSMs) originating from 26 countries have been identified (Deliverable 2.3, Fig. 4.1).

Model findability and geographical distribution

The plant model database was populated with models whose description was initially scattered across a variety of scientific publications and journals, thus requiring a thorough review of the literature. However, plant models developed by industry (e.g. in breeding companies) are usually not published, likely for confidentiality reasons. Therefore, the list of plant models may not be exhaustive. As part of the preparation phase of EMPHASIS, a pilot action was initiated to turn this database into an online portal referencing plant models to raise awareness about the diversity of models and their applications. This portal (quantitative-plant.org) provides high visibility and findability for plant models, and is also made available through the EMPHASIS website.

The database indicates that a large proportion of the SPMs and FSPMs are developed in Europe (73% and 78% of the SPMs and FSPMs, respectively; Fig. 4.1 of the Deliverable 2.3). On the contrary, CSMs are mainly developed in the United States (27%; Fig. 4.1 of the Deliverable 2.3), illustrating a gap in this model category in Europe. Besides their large dispersal, a high proportion of models are not directly available to users, limiting their accessibility to other scientific communities (44% of the models are available "upon request"; Fig. 8).

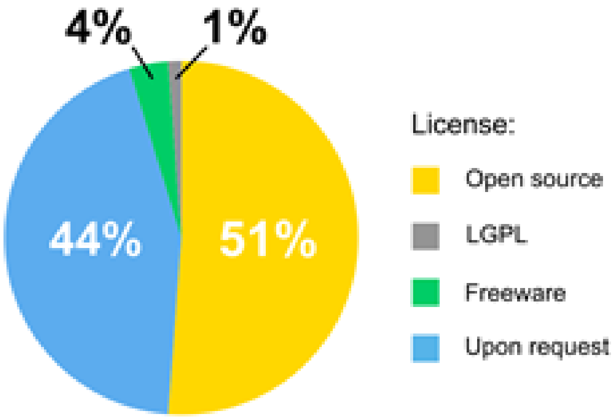


Figure 8: Proportion of license type by plant model (estimated considering all model categories).

A second pilot action will be initiated to streamline and improve exchange between the phenomics and modeling communities, taking advantage of the EMPHASIS layer. The EMPHASIS portal will be developed to connect the phenomics and modeling communities (Figure 9) and, more specifically, facilitate the integration of phenomics data in plant models and offer guidelines and tools improving model-data compatibility.

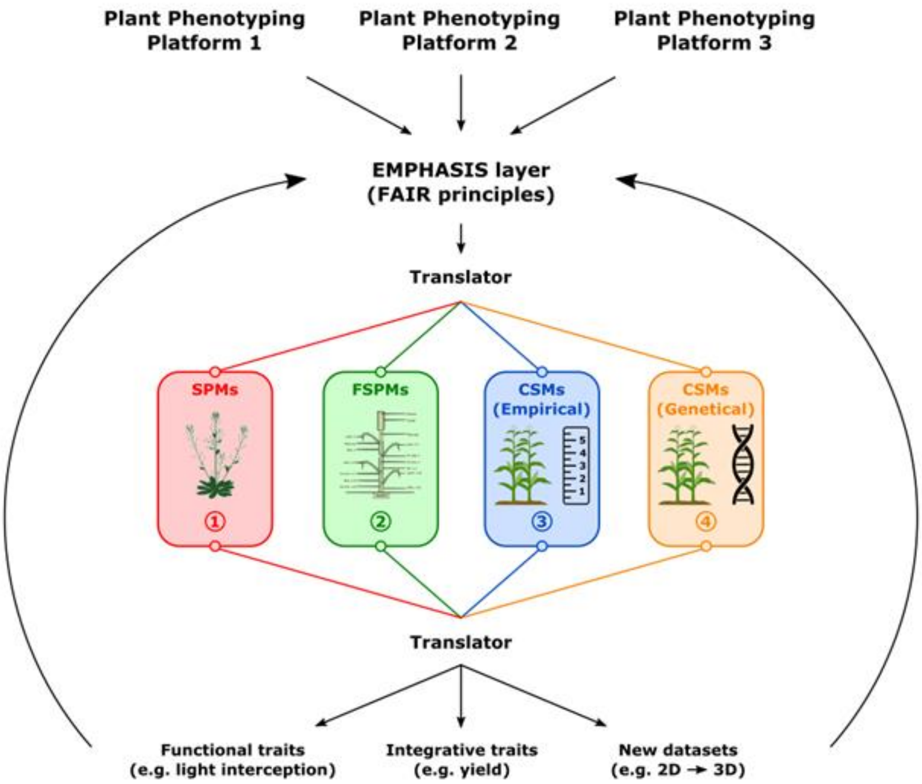


Figure 9: Illustration of the EMPHASIS portal connecting phenomics data to plant models. Phenomics data and plant models are centralized in the EMPHASIS information systems and can be connected by translator tools, linking the EMPHASIS layer with the model input

Model diversity, interoperability and harmonization

Plant models are developed by different groups and for different aims, leading to a considerable diversity between and within each model category. From the point of view of model/model and model/data interoperability, this diversity can be observed at two levels:

1. the structural framework, the modeling approach and the scopes of the models (e.g. operating system, programming language, model inputs, model outputs…),

2. the use of different standards in terms of definitions, units, and interrelations between the variables and the parameters of the models.

These particularities, specific to each model, constitute challenges to improve the interoperability between the phenomics data and the plant models through a unique portal. EMPHASIS, as an infrastructure, may support and contribute to the definition of data and models for improved articulation between the two communities (e.g. by coordinating the development of common ontologies of variables and parameters involved in the phenomics and modeling communities). EMPHASIS will encourage the modeling community to adopt and develop harmonized standards for interfacing models with phenomics data sources.

An EMPHASIS pilot (to be described in Deliverable 6.3) responds to a need expressed by these two communities to meet and organize common trainings and workshops bring together modelers and experimenters, thereby promoting the co-evolution of models, phenotyping platforms, and methods/data standards.

5. Data management systems

EMPHASIS aims to create centralized access to phenotyping data by building and integrating compatible, consistent information systems that will provide methods and interfaces for the interoperability of datasets to manage, share, reuse and visualize heterogeneous, high-throughput plant phenotyping data stemming from different sources, often in an interdisciplinary context.

Before having common tools and practices within the different installations, understanding the practices of the data management within the phenotyping installations (from all kind) is important. This precise question has been the topic of another Deliverable D4.1 Map of information systems.

There are gaps within the survey of data management. It first comes to the fact that the questions were designed and addressed to data managers. Most of the time, these managers were not solicited to provide the input to the questions and the results in missing information or even worse, false information. As a proportion, the data managers that gave some answer only represent 10-15% of the total of the installations.

Another origin of missing information is the fact that questions were asked in different waves, the first one serving mostly for calibration of the survey. That is to say, everyone has not been filling the exact same survey, leading to scattered missing data. In the future of EMPHASIS operation, this gap could be solved by not only continue to map existing and upcoming data management systems, but even more, establish data management training and advice for handling phenomics data and provide a portfolio of existing, and quality approved, data management systems.

Concerning the gaps in the data management practices, it shows that practices are not FAIR yet. To reach FAIR data the local data management systems of local infrastructures should be made accessible and findable, and even better, be linked with each other through an online portal. EMPHASIS-PREP propose to develop such portal, called the EMPHASIS-layer, which will make data findable and accessible from local infrastructures, and by including metadata, the data would be comparable, interoperable and reusable. The establishment of the EMPHASIS-layer will be piloted during the implementation phase of EMPHASIS to better understand the needed resources and requirements to develop a long-term data management system benefiting the European plant phenotyping research community.

FAIR data requires a proper identification system, if you want to gather data from various sources. Such an identification system would be easy if coming from a central service as the EMPHASIS Layer, but mostly requires the actors to agree on a common schema. This identification could for example be based on Uniform Resource Identifier (URI), that is proposed in a sideway document Identification of objects with Uniform Resource Identifier (URI): recommendations for application in plant phenotyping available within the EMPHASIS community.

To develop the EMPHASIS-layer a common API needs to be used, which could be found in the already existing and phenomics community driven Breeding API (BrAPI – see website: https://brapi.docs.apiary.io). Such an API can be carried by the EMPHASIS Layer and be a major step toward FAIR data management in Europe. Moreover, for compiling meta-data, EMPHASIS proposed to use MIAPPE that 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 (https://www.miappe.org/).

Last but not least, a common vocabulary is essential to exchange data. This requires conventions and naming patterns for everything, from objects to variables used. Ontologies are not enough to tackle this problem as the number of “concepts” and “instances” concerned is limited. Instead, conventions and naming patterns would help. Again the EMPHASIS Layer has a role to play, being a platform used to share a vocabulary and avoid duplicates.

Once all of these issues have been addressed, interoperability between different information systems will be easy. That is not the case yet.

Conclusion, discussion and next steps

EMPHASIS will facilitate standardized and interoperable plant phenotyping for all pillars of plant phenotyping through services as for example quality control, access, harmonization and innovation to the benefit of tackling the global challenges of climate change and food security.

Although the mapping results are high quality, and the mapping database (https://emphasis.plant-phenotyping.eu/database) will be further developed, the methodology of mapping the existing and upcoming installations was noticed to be biased. In the future of EMPHASIS operation, the mapping needs to be a continuous process and a process for revision of the mapping should be installed in the governance of EMPHASIS, with fixed time lines and responsibilities. Extending the EMPHASIS database with more details and develop an easy process for the users to populate the database will be key, and is a process already started in EMPHASIS-PREP. Furthermore, industry involvement in innovation forums and collaborative projects could establish better insight in the installations and capacities of the private plant phenotyping sector.

The gaps identified in provision of infrastructures for phenotyping in (semi-)controlled conditions will be addressed in part by services offered by EMPHASIS. Proposed services to be piloted in the implementation phase include access and technology innovation that will address the identified gaps in geographical location, throughput and focus.

In the highly equipped fields, the geographical gaps were identified, that either come from the mapping method, the lack of such installations or confusion with network of lean fields. Due to the fact that highly equipped installations are very specialized installations, and the survey failed to capture all the different specificities. These gaps could be tackled through a better installation database (see above).

The increased need for a coordinated European network of field trials for multi-climatic experiments is strongly emerging as tool to i) address the need of access to open field trials (See Deliverable D 2.5); ii) develop standard protocols and harmonizing field management procedures in order to facilitate the exchange of data; iii) address the big themes of plant biological and agro ecological research in Europe. A network structure would allow for a coordinated development of the individual sites with the necessary specialization and optimal resource allocation and will guarantee an up to date technical equipment. Therefore, it is needed to extensively map the available field phenotyping stations across Europe to be able to find the correct collaborative partners for multi-site field experiments, preferably with different climatic regions and enough details, e.g. environment condition, pedoclimatic conditions, size of the fields and available equipment. This will determine whether the fields fits the requirements for a proposed multi-site project.

The gaps identified for the plant models (e.g. the incomplete inventory of models, the limited model access or the large model diversity) will be addressed in the next parts of the EMPHASIS phase. The “quantitative-plant.org” website will provide great opportunities to diffuse information, to incorporate other models and to get the modeling communities involved. However, in the future, a long-term cooperation between the phenomics and modeling communities towards harmonizing data formats will be needed to enable transparent data exchange from models to experiments and vice-versa.

Concerning the data management practices, several gaps have been identified, most notably the lack of FAIR local information management systems and the lack of interoperability between users. To be able to share data, the interoperability requires a proper identification pattern, a common language (API) to access the different information systems of the different users, and finally a common vocabulary when exchanging data. In order to tackle these gaps, the EMPHASIS pilot of data management has integrated tools for identification and proposes naming rules and standards, and proposes to develop an EMPHASIS-layer that will establish communication between local data management systems.

Glossary

API: Application Programming Interface

Canopy: canopy is more than one plant - in CE setups this is the difference between a top-down camera imaging a stand of plants and a conveyor measuring plant by plant…

CSM: process-based crop simulation model

EMPHASIS: European Infrastructure for Multi-Scale Plant Phenotyping And Simulation for Food Security in a Changing Climate- ESFRI listed project

EMPHASIS-PREP: H2020 preparatory phase project of EMPHASIS

ENVRI: ENVRI is the community of the Environmental research infrastructures, projects and networks as well as other diverse stakeholders interested in the environmental research infrastructure matters

EOSC: European open science cloud

ERC=European Research Council

ESFRI: Eureopean Strategy Forum for Research Infrastructure

FACE: (Free-Air CO2 Enrichment)

FAIR: findable accessible interoperable and reusable of digital assets

FSPM: functional-structural plant model

Installation: An INSTALLATION is the elementary level for data acquisition in a specific type of experiments. It stands for other frequently used terms such as ‘platform’, ‘facility’ or others.

Local infrastructure: A LOCAL INFRASTRUCTURE is a group of installations (see §1.3) located in one site depending on one institution (or more), which share governance committees, a common (or at least highly interoperable) information system, common principles for cost calculation and pricing and a common tool for user access.

RGB: The RGB color model is an additive color model in which red, green and blue light are added together in various ways to reproduce a broad array of colors.

SEB: The Society for Experimental Biology

Shoot: phenotyping focusing on imaging

SPM: structural plant model

SQL database: Structured Query Language for databases

TB: in the context of data management systems: Terabytes

UAV: Unmanned Aerial Vehicle, mostly drones

URI: Uniform Resource Identifier

Web service: a layer of abstraction between the database technology and the user. This layer facilitates the interaction between user and provider.

WP: work package

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		*Please inform Management Team of any foreseen delays*
	The title corresponds to the title in the DOW		*If not please inform the Management Team with justification*
	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		*Please validate the Table of Content with your Activity Leader before drafting the deliverable*
	I am using the EMPHASIS deliverable template (title page, styles etc.)		*Available in "New EMPHASIS Logo, Templates, CI" on the collaborative workspace*
	The draft is ready		
After	I have written a good summary at the beginning of the Deliverable		*A 1-2 pages max. summary is mandatory (not formal but really informative on the content of the Deliverable)*
	The deliverable has been reviewed by all contributors (authors)		*Make sure all contributors have reviewed and approved the final version of the deliverable. You should leave sufficient time for this validation.*
	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		*Send the final draft to your WP Leader 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.*