



EXCELERATE Deliverable D7.1

Project Title:	ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences	
Project Acronym:	ELIXIR-EXCELERATE	
Grant agreement no.:	676559	
	H2020-INFRADEV-2014-2015/H2020-INFRADEV-1-2015-1	
Deliverable title:	Datasets annotated for at least 1 woody plant, cereal and solanaceous species; genotype, phenotype and sample metadata submitted to appropriate public archives	
WP No.	7	
Lead Beneficiary:	1. EMBL	
WP Title	Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants	
Contractual delivery date:	31 August 2017	
Actual delivery date:	31 August 2017	
WP leader:	Paul Kersey, Celia Miguel	1. EMBL, 18. iBET
Partner(s) contributing to this deliverable:	CIRAD, EI, EMBL, FCG, iBET, INRA, NBIC, NIB, VIB	

Authors and Contributors:

Anne-Françoise Adam-Blondon, Inês Chaves, Frederik Coppens, Bruno Costa, Richard Finkers, Carlos Horro, Paul Kersey, Celia Mihotey, Célia Miguel, Evangelia Papoutsoglu, Cyril Pommier, Ziva Ramsak

Table of contents

1. Executive Summary	2
2. Impact	3
3. Project objectives	4
4. Delivery and schedule	4
5. Adjustments made	4
6. Background information	5
7. Appendix 1: Report. Datasets annotated for at least 1 woody plant, cereal and solanaceous species; genotype, phenotype and sample metadata submitted to appropriate public archives	8
Task 7.1 Development/adoption of appropriate controlled vocabularies for annotating plant phenotypic data	8
M7.1: Community workshop held to agree common ontologies	8
M7.2: First set of agreed ontologies for annotation of phenotype in crop and forest plants	10
Task 7.2 Annotation of key plant phenotypic datasets with agreed controlled vocabularies	11
Task 7.3 Submission of exemplar genomic and phenotypic datasets to appropriate public repositories [Office3]	12
8. Appendix 2: List of Presentations on the Work Described	15
9. Appendix 3: Proposal for MIAPPE v1.1	16

1. Executive Summary

Plant omics data is often not published, or dispersed among multiple repositories and, owing to poor annotation of submissions with descriptive metadata, effectively not FAIR (i.e. Findable, Accessible, Interoperable and Re-usable). In this work package, our goal is to make plant -omics data FAIR, and the first part of this work has involved the development of standards for the annotation of data sets, and their exemplary application to diverse data from diverse species. Finally, the annotated data has been made available in the public repositories.

In developing these standards, we have built from previous work in the field, most notably the development of the Minimal Information about a Plant Phenotyping experiment (MIAPPE) standard (<http://www.miappe.org>). We have taken MIAPPE as our starting point, and extended and revised it according to the characteristics of our exemplar data.

2

Because a proposed standard becomes a real one through its adoption, we have actively participated in the wider community developing MIAPPE, with the goal of incorporating our developments within the official MIAPPE specification.

A particularly important issue is the identification of material used in an experiment. Materials used within and between experiments may be related to each other (by time, space, condition or genealogy), and multiple assays may be taken from the same material; moreover, what is often of interest is to connect experimental data to the physical material assayed (or at least, related physical material, such as a stock held in a germplasm repository). We have taken special care in aligning the MIAPPE standards with the FAO Multi Crop Passport Descriptor Data standard used by the gene banks and used the BioSamples database (<http://www.ebi.ac.uk/biosamples>) as an integrative layer to allow us to capture the relationships between samples and the measurements that have been made upon them.

2. Impact

The purpose of this deliverable is to provide a set of exemplary data annotated and published in accordance with community-approved standards (some of which were developed in the course of the work). Impact will depend on public awareness of the standards and the data, and adoption of those standards by a wide range of actors. To achieve this, we have proposed the standards we have developed to the steering committee of the MIAPPE project (which developed a previous standard that provided our baselines), with the goal that they should be promoted to the wider community. Indeed, members of this work package have played an active role in establishing the formal governance structure for MIAPPE, to serve as a persistent body to integrate, validate and promote efforts to develop standards in this area, whichever project they derive from; and two participants from EXCELERATE have been chosen to serve (along with 5 other individuals from other interested projects, such as the EPPN2020 and EMPHASIS projects) on the initial steering committee.

We have additionally participated in the Breeding API (BrAPI) project which aims at developing and implementing a Web Service API for data exchange focused on plant material, phenotyping and genotyping data types, mainly for breeding purposes. This project, initiated by the centres of the Consultative Group on International Agricultural Research, the James Hutton Institute, and others, has been gaining traction in the community and we have been working to align the activities on MIAPPE and BrAPI. By now, BrAPI can be considered an implementation of the MIAPPE data standard. Two EXCELERATE nodes, ELIXIR-FR and ELIXIR-NL, have actively participated in the BrAPI specification process.

Moreover, the work in this work package has been directly promoted in numerous international meetings, national meetings and invited seminars (see appendix 2).

We have engaged with industry at an EBI-organised Industry Workshop on "[Ontologies in Agriculture, Food and Nutrition](#)" and at the national level through the GIS-BV research

group of interest (<http://www.gisbiotechnologiesvertes.com/en/>) where the progresses of the work package have been presented.

We have also engaged with the eROSA project, which is tasked with the construction of an e-infrastructure roadmap for open science in agriculture.

Work package participants are implementing the standards in their own nodes. Moreover, partner 27, INRA, is implementing the work package 7 guidelines in five large Public - Private Projects (that together involve most of the seed companies working in France) in which it has data management responsibilities. Finally, advice has also been given to the nascent Finnish national phenotyping centre concerning the implementation of MIAPPE.

Members of the work package have also participated in the Genetic Resources/Bioinformatics EU-China Working group, to advise on areas of priority for future work, and have advised participants in the CORBEL project as they consider similar challenges to ours in the field of marine microorganisms of medical importance.

3. Project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives:

No.	Objective	Yes	No
1	Deliver world-leading data services for academia and industry: Establish a distributed genotype–phenotype annotation that supports agriculture research and industrial development	X	
2	Make data interoperable (in accordance with the 'FAIR' principles specified in WP5) through the development of controlled vocabularies and standardised APIs, proving the concept of a common phenotypic API through which any participant in an open network can advertise the availability of their data in a common domain.	X	
3	Annotate and submit key exemplar datasets to relevant public archives	X	

4. Delivery and schedule

The delivery is delayed: • Yes No

5. Adjustments made

N/A.

6. Background information

Work package number	WP7	Lead beneficiary	1 - EMBL
Work package title	Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants		
Start Month	1	End Month	48
Work Package Lead	Paul Kersey (UK) and Celia Miguel (PT)		
Objectives			
<p>The main objective of WP7 is to design and test an infrastructure to allow genotype-phenotype analysis for crop plants based on the widest available public datasets. To achieve this, the specific objectives for WP7 are to:</p> <ul style="list-style-type: none"> • Make data interoperable (in accordance with the 'FAIR' principles specified in WP5) through the development of controlled vocabularies and standardised APIs, proving the concept of a common phenotypic API through which any participant in an open network can advertise the availability of their data in a common domain. • Annotate and submit key exemplar datasets to relevant public archives. • Engage industry in defining priorities in genotype/phenotype annotations, and collaborate with WP13 in showcasing the developed resources to the agroforestry industry. • Collaborate with WP11 in delivering specific training for the use of developed resources. 			
Description of work and role of partners			
<p>WP7 - Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants [Months: 1-48] EMBL, TGAC, NBIC, FCG, IBET, INRA, CIRAD, NIB, VIB</p> <p>This work will facilitate the analysis of many of phenotypes against large panels of crop accessions through the aggregation of locally held data; and thereby, enable more powerful association analysis, opening the way to understanding of function, candidate gene prioritisation, and improved crop breeding. Working on exemplar species, we will establish a sustainable model for the interaction of distributed phenotypic repositories with defined genomic and sample reference data, in which organisations can expose data to the system through conformity with standards for annotation and interface, allowing the subsequent expansion of the approach to other species and domains. It will also provide resources (in the form of standards, ontologies and models for annotation and collaboration) for use within ongoing species-centric (e.g. the Wheat Initiative) and/or national endeavours.</p> <p>Massive sequencing and genotyping of crop and forest plants (and their pathogens and pests) is generating large quantities of genomic variation data. These efforts are likely to accelerate in the near future, with further expected reductions in the cost of sequencing and international efforts (such as the DivSeek Initiative) aiming to catalogue all genetic diversity present in global</p>			

germplasm resources. Such data could serve as a powerful panel in association screens and facilitate precision breeding of increasingly complex traits. But structural variation in most crop plants is enormous (more so than in humans), and phenotypic characterisation data is (i) often inaccessible (ii) diverse and non-standard

(iii) lacks any route of unified access. Indeed, “phenotype” is a broad concept, covering many data types (descriptive data, molecular data, image data) at many scales (laboratory, phenotyping centre, field data) on many species; and moreover, a phenotype exists in the specific concept of an experiment (in contrast to a genotype, which is assumed to be constant in a given sample). Both technical and sociological progress in data definition and sharing are lagging experimental progress.

To address this problem, we will harness the domain-specific expertise and data held in a distributed fashion across many national Nodes with interests in agriculture. Seven ELIXIR Nodes will jointly establish a technical infrastructure and associated social practices to define an open model for the publication and sharing of plant genotype-phenotype data, working on a minimum of 3 exemplar species from different domains of plant life to establish a model applicable in all species. We will establish a scalable, distributed model, transparently integrated through the development and use of common vocabularies and search technologies, adhering to the FAIR principles 60 (WP5), and using established repositories for genomic data and sample meta data. Domain-specific training will be coordinated with WP11 and will include training users and data curators. The expected impact is accelerated research and plant breeding through the exploitation of an interoperable commons of public data.

Task 7.1: Development/adoption of appropriate controlled vocabularies for annotating plant phenotypic data (34PM)

The use of controlled vocabularies, to define the material assayed, the form of the experiment and the observed phenotype are critical to enabling diverse datasets to be interrogated and compared. A number of initiatives have developed controlled vocabularies that can be used for the annotation of plant phenotypic data, including the Plant Ontology (<http://www.plantontology.org>), Crop Ontology (<http://www.croponontology.org>), Plant Trait Ontology (http://www.obofoundry.org/cgi-bin/detail.cgi?id=plant_trait), Environment Ontology (<http://environmentontology.org>), XEML (<http://xeml.codeplex.com>). Different vocabularies apply in different species, with some specific and some overlapping features; in many countries, national lists of traits found in specific crop varieties are in use by breeders (distinct from the terms in use in academia). Slow-growing forest species have overlapping needs to annual food crops.

Together with representatives from the appropriate infrastructure resources, communities and ontology developers, and industrial/SME partners, we will work on establishing common guidelines for ontology usage when annotating crop and forest species. We will hold a workshop within the first 4 months of the project bringing experts together from all participating Nodes and key industrial participants, working on the target species to agree on a common set of vocabularies to be used in the project (by the end of the month 6). Existing ontologies will be extended where necessary, and cross-references established between corresponding high-level terms in the recommended vocabularies (e.g. between common anatomical concepts in different species-specific ontologies). Outputs will be regularly showcased to industry and list reconsidered at 6 monthly intervals.

Partners: EMBL-EBI, BE, FR, NL, PT, SI, UK

Task 7.2: Annotation of key plant phenotypic datasets with agreed controlled vocabularies (51 PM)

At least 3 exemplar species will be chosen, including one cereal species, one woody species, and one other crop species, each of which is of interest to at least 2 ELIXIR partners (maize, oak and

potato have been identified as initial candidates). Phenotypic data is very varied and we will work on a variety of data types, including quantitative trait loci, association genetics (laboratory and field data), and biotic stress. Participating Nodes will collect and curate sample and experimental meta data and phenotypic description data to agreed standards using the vocabularies identified and extended in Task 7.1.

Partners: BE, FR, NL, PT, SI, UK

Task 7.3: Submission of exemplar genomic and phenotypic datasets to appropriate public repositories (33.9PM)

Annotated datasets will be submitted to appropriate repositories wherever possible, using existing platforms designed for such scope; for example, the European Nucleotide Archive (developed by EMBL-EBI) (for genomic and transcriptomic data), or phenotypic archives e.g. PIPPA (the PSB Interface for Plant Phenotypic Analysis, in development at VIB), BreeDB (in development at NBIC / WUR), GoMapMan (in development at NIB), and Ephesis (in development at INRA). Sample identification will be handled through the BioSample DB at EMBL-EBI, or, where the sample is an accession from a public gene bank, by cross-references to EURISCO, the European catalogue of plant collection data (<http://eurisco.ipk-gatersleben.de>). We will work closely with EURISCO and the gene banks to ensure that appropriate meta data is reliably, non-redundantly associated with samples, and that genomic and phenotypic data held in different resources but derived from the same biological material can be reliably identified.

Partners: EMBL-EBI, BE, FR, NL, PT, SI, UK

Task 7.4: Development and implementation of agreed public APIs for access to data in participating repositories and exposure via public computational infrastructures (51PM)

We will develop a common API for data query and retrieval, in close collaboration with WP5. We will build from the experiences already gained by partners in developing repositories and interfaces and will utilise established standards for programmatic data access (e.g. REST, RDF, etc.). The API will be implemented at each partner holding a genotypic, phenotypic or sample repository, allowing users to query a single end point that will return results meeting a common standard from dispersed resources. The API will be exposed to users via the ELIXIR computational infrastructure and other computational infrastructures in the plant sciences (for example, on the iPLANT infrastructure in the United States). The first version of the API will support identification and query of datasets held in distributed repositories matching specified experimental and sample meta data. The API will be later be extended to encompass the querying of resources based on the phenotypic descriptions. We would also like to enhance the interpretability of results and lower the barrier of computer competency required by users to access data queries from this platform. We will use the data served by the API to develop intuitive visualisation components to allow researchers to mine phenotypic data accessible through the API. These components will be developed within existing frameworks e.g., PIPPA or BioJS (a JavaScript library of open source components for biological visualisation), allowing their re-use in other contexts.

Partners: EMBL-EBI, BE, FR, NL, PT, SI, UK

Relation to other WPs

Propose phenotype resources to service registry (WP3)

Exposure on the ELIXIR cloud resources (WP4)

Adhering to and develop vocabularies and API compatible to the FAIR principles (WP5) Training in the development of plant ontology development and in the use of resources (WP11)

Partner number, short name and effort: 1- EMBL 12.00; 3 - TGAC 21.20; 6 - NBIC 0.00, DLO

9.00; 16 - FCG 2.00; 18 - IBET 48.00; 27 - INRA 24.70; 28 - CIRAD 12.00; 33 - NIB 30.00; 37 - VIB 11.00

7. Appendix 1: Report. Datasets annotated for at least 1 woody plant, cereal and solanaceous species; genotype, phenotype and sample metadata submitted to appropriate public archives

The work towards the achievement of this deliverable was divided into 3 tasks, as follows:

Task 7.1 Development/adoption of appropriate controlled vocabularies for annotating plant phenotypic data

We identified appropriate controlled vocabularies for annotating plant phenotypic data, which includes the definition of materials assayed, the form of the experiment and the observed phenotype. The term 'phenotyping' is here used in its broad scope, and data types collected include both descriptive and molecular data.

After initial discussions involving all partners to exchange knowledge about available vocabularies, a community workshop ("PhenoHarmoniS") was co-organized with other interested projects (Milestone **M7.1**). Based on these internal and external discussions, work then proceeded to specify the standards for plant phenotyping data to be met in the project. The initial specification (Milestone **M7.2**) has allowed work to undertake task 2 (the annotation of data sets to the standard required), and produce the initial project deliverable.

M7.1: Community workshop held to agree common ontologies

Phenotyping is at the heart of all research in plant biology and breeding and has changed tremendously in recent decades thanks to progress in imaging and robotics at many scales. This has led to a considerable increase in the throughput of data acquisition and has pushed forward the necessity to better organize and standardize phenotyping data. No international archive is yet available for these data, which remain very heterogeneous compared to genomic data. Data management thus relies on an emerging network of community repositories and the purpose of these repositories is to ensure that data meets the FAIR principles (Findable, Accessible, Interoperable and Reusable) by applying common international ontologies and standards.

To assist in this process, EXCELERATE partners participated to an initiative led by Bioversity International for the organization of a 5-day workshop "Semantics for Harmonization and Integration of Phenotypic and Agronomic Data" gathering specialists

of data standardization, data managers and data producers. The PhenoHarmonIS workshop was held in Montpellier (2016, May 9th-13th) and all the information (organizing committee, detailed program, presentations and list of attendees) is available at <http://tinyurl.com/hzsho6v>.

PhenoHarmonIS was co-organized and sponsored by international consortia promoting FAIR data such as the Research Data Alliance and DivSeek but also by the European plant phenotyping infrastructure (EPPN/EMPHASIS), large European and international infrastructures about data management and data standardization (ELIXIR, Integrated Breeding Platform, Planteome) and international organizations (CGIAR). The workshop gathered 80 attendees from 17 countries, targeted and invited by the organizing committee. The partners of EXCELERATE served on the organising committee, and contributed specifically to the organization of two sessions:

- *Promoting best practice and developing common approaches to the management and analysis of phenotyping data* on May, Tuesday 10th
- *Hands-on data annotation MIAPPE-ELIXIR* on May, Thursday, 12th

EXCELERATE partners were furthermore active throughout the meeting, talking to community members about their needs and learning about pre-existing solutions and helping to further develop these.

Aside from work on specific vocabularies and data standards, some generic imperatives emerged from the meeting:

- Adopt ontologies and metadata standards useful to the data management throughout the project, to capture all the necessary information for the data to be re-usable, avoiding complex generic ontologies or standards. Ensure that the environment in which the experiment was carried out is well described and captured - this is very important for allowing sound future meta-analysis of experimental data.
- Make ontologies useful for knowledge inference. Some initiatives are starting to develop in this area (e.g. Planteome, AgroLD, KNetMiner) but a major issue is also to progress towards the access to public well standardized data sets.

The workshop allowed the work package to proceed in adopting MIAPPE (Minimal Information About Plant Phenotyping Experiments; <http://www.miappe.org>) as a standard for plant phenotyping data, the Crop Ontology trait dictionary (<http://www.cropontology.org/>) as a framework to capture the data about the phenotyping variables measured by the researchers and the Breeding API (www.brapi.org) as a specification for RESTful web services compliant with the MIAPPE standard. Finally, [ISA-tab](#) was pinpointed as a useful exchange format between information systems. Gaps and improvements to be addressed at the data, metadata or ontology level were also identified: standardization of time series, plant-pathogen interactions and modelisation of developmental stages.

M7.2: First set of agreed ontologies for annotation of phenotype in crop and forest plants

Taking into account the already existing controlled vocabularies (e.g. Plant Ontology, Crop Ontology, Plant Trait Ontology, among others) and standards (e.g. MultiCrop Passport Data for the identification of the plant material) that can be used for the annotation of plant phenotypic data, we worked on establishing common guidelines for annotation of phenotypic data from diverse plant species including species with different characteristics (e.g. annual food crops and woody perennials), and ontology usage when annotating such data.

A first set of guidelines with agreed vocabularies for the annotation of phenotype in crop and forest plants was achieved based on the extension and revision of v1.0 of the MIAPPE standard for specification of required information about plant phenotyping experiments, previously created with the participation of some elements of this work package (Ćwiek-Kupczyńska et al. 2016; DOI: [10.1186/s13007-016-0144-4](https://doi.org/10.1186/s13007-016-0144-4)). An initial document, consisting of a list of attributes grouped into sections (general metadata, timing and location, biosource, environment, treatments, experimental design, sample collection, processing and management, and observed variables) recommended for the description of phenotypic observations, was extensively discussed among all partners (EMBL-EBI, INRA, iBET, EI, VIB, FCG, NBI, NBIC, CIRAD) and used as a basis to create an extended set of recommendations for phenotypic annotation which would accommodate most types of datasets as provided by the participant partners. Work was carried out in two working groups, one specifically focused on standard descriptions of biological materials, led by EMBL-EBI; and one on the remainder of the standard, led jointly by ELIXIR-FR and ELIXIR-PT.

As a result, we have developed a revised list of attributes including new and edited attributes (some of which tagged as mandatory) to better describe actual data sets (see Appendix 3). Special emphasis has been placed on biosource attributes for non-crop species such as forest trees which were not well-represented by the previous specification. A definition of each attribute is provided to guide users and ensure easy and correct interpretation during the annotation process; a list of proposed ontologies and expected datatypes was also added to detail how to fill in each field. The revised specification has been submitted to the MIAPPE steering committee for consideration and potential adoption as MIAPPE v1.1 (Appendix 3).

Plant material identification (biosource attributes) was given special attention due to its importance for integrating diverse data sets corresponding to the characterization of the same plant accessions with different technologies: e.g. genotyping experiments and phenotyping experiments.

Additional developments are the result of specific collaborations among partners to address gaps in existing controlled vocabularies, for describing traits/features not present in all plant species (e.g. wood production), or describing plant experimental assays. A Woody Plant Ontology has been developed by partner 27 (INRA) in collaboration with iBET to cover tree-specific traits not included in current ontologies. A first version is

available at <https://urgi.versailles.inra.fr/ephep/ephep/ontologyportal.do> and in the Crop Ontology portal : http://www.croponontology.org/ontology/CO_357/Woody%20Plant%20Ontology. The Vitis Ontology (http://www.croponontology.org/ontology/CO_356/Vitis) has been thoroughly revised in collaboration with its curator by INRA. Traits and protocols have been added to all variables and new variables have been included to prepare the submission of phenotyping data sets. A work is ongoing between INRA, partner 6 (DLO), and partner 37 (VIB) to add new variables to the maize and tomato ontologies and between INRA and partner 3 (EI) to generate a rapeseed ontology (<https://github.com/Brassica-Trait-Ontology/brato>).

An ontology to describe the pipeline of manipulations performed from specimens to data (Plant Experimental Assay Ontology, <https://bioportal.bioontology.org/ontologies/PEAO>) has been developed by ELIXIR-PT and is undergoing further improvement for alignment with MIAPPE.

Task 7.2 Annotation of key plant phenotypic datasets with agreed controlled vocabularies

Following the development of the initial set of data standards in task 7.1, work has begun at all participating nodes (with the exception of EMBL-EBI) on annotating an exemplar data set in accordance with these standards. Specifically, the following data sets are being worked on:

ELIXIR-BE

The PSB Interface for Plant Phenotyping Analysis (PIPPA, <https://pippa.psb.ugent.be>) has been improved to align to the MIAPPE standard. The Breeding API has been implemented on PIPPA (https://pippa.psb.ugent.be/pippa_experiments/brapi_docs/) Plant and leaf phenotyping for two maize RIL populations have been made available through PIPPA ([Baute et al. 2015](#), [Baute et al. 2016](#), [Dell'Acqua et al. 2015](#)). Also, transcriptome data is available in ArrayExpress and has been linked to the phenotyping data through BioSamples. Also genotyping data is available for these inbred lines.

ELIXIR-FR (CIRAD)

Observation Trials across Multiple Environments for 60 lines of rice (tropical japonica).

ELIXIR-FR (INRA)

Annotation and curation of several datasets stored in the GnpIS information system (<https://urgi.versailles.inra.fr/gnpis/>) and, in parallel, improvement of GnpIS to better comply to MIAPPEv1.1.

A maize dataset (Bouchet et al 2013: doi:10.1371/journal.pone.0071377): 192 maize lines genotyped with a 50K chip and phenotyped for three traits related to flowering time. The variables have been submitted to the curator of the Maize Ontology and information has been retrieved from the paper to complete the metadata.

A poplar dataset provided consisting of five phenotyping experiments on a population of 350 individuals; the same population was also genotyped. This has been annotated with the new MIAPPE format and the attributes used to describe the data set have been added to the Woody Plant Ontology (http://www.croponontology.org/ontology/CO_357/Woody%20Plant%20Ontology). The notion of lot/sample versus accession has been clarified with the forest genetic resource managers and forest tree researchers at INRA. The data has been published in the GnpIS repository; detailed information can be obtained on the accessions of the population from the catalogue of genetic resources maintain in GnpIS (e.g. <https://urgi.versailles.inra.fr/gnpis-core/#accessionCard/id=doi:10.15454/1.4921786776732773E12>) or through GnpIS portal (e.g. <https://urgi.versailles.inra.fr/gnpis/#result/term=661300224>), which links back to all experiments in which the given accession was characterized.

ELIXIR-NL/ELIXIR-SI

Transcript profiling of tuber samples of 94 genotypes deriving from a segregating diploid potato population, C x E, derived from the diploid parental clones C (USW5337.3 (Bonhomme 2000; Hanneman and Peloquin 1967)) and E (77.2102.37 (Jacobsen 1980)).

Transcriptomic, proteomic and morphological time series profiling of potato plants (cv. Désirée; non-transgenic and transgenic lines) in response to PVY infection, Slovenia, 2010 (Stare et al., 2015, BMC Genomics 16:719).

ELIXIR-PT (IBET)

Developmental stage-specific transcriptional changes during zygotic embryogenesis in maritime pine (*Pinus pinaster* Ait.) (de Vega-Bartol et al 2013, BMC Plant Biology 13:123)

Molecular study of drought response in the Mediterranean conifer *Pinus pinaster* Ait.: Small non-coding-transcript profiling (under revision).

ELIXIR-UK

Annotation of data in the Brassica Information Portal (<https://bip.earlham.ac.uk/>). Two different *Brassica napus* datasets were annotated, one related to seeds (seed mineral concentration data of compost-grown lines of the RIPR panel generated by ICP-MS analysis) and one related to leaves (leaf mineral concentration data of compost - grown lines of the RIPR panel generated by ICP-MS analysis).

Task 7.3 Submission of exemplar genomic and phenotypic datasets to appropriate public repositories [Office3]

As annotation of the exemplar data sets is proceeding, different portions of the data have started to reach the point where they can be submitted to appropriate public archives. This work has begun with a primary focus on the submission of sample data to the BioSamples database (<http://www.ebi.ac.uk/biosamples>), in accordance with the model agreed in Task 7.1. The BioSamples database has a flexible data model that can be used to capture the required information about samples, and the submitter has a lot of

discretion about how to utilise this freedom. We have been standardising the choice of representation, in order to allow data to be subsequently extracted from the database in a consistent manner. This work has been coordinated by EMBL-EBI, working with each of the partners annotating data sets. In some cases, the work has involved new submissions, in other cases it has involved the re-curation of previously submitted samples to meet the new data standards.

The specific data sets that have been submitted to BioSamples (and the contributing partner) are as follows:

ELIXIR-BE: <https://www.ebi.ac.uk/biosamples/groups/SAMEG328646> *Zea mays* data from material of parental lines of the RIL and phenotypic and transcriptomic data linked <https://www.ebi.ac.uk/biosamples/groups/SAMEG328664>.

ELIXIR-FR (CIRAD): <https://www.ebi.ac.uk/biosamples/groups/SAMEG323144>. Data from *Oryza sativa*.

ELIXIR-FR (INRA): <https://www.ebi.ac.uk/biosamples/groups/SAMEG323004>. Data from *Ampelopsi*, *Parthenocissus* and *Vitis* species.

<https://www.ebi.ac.uk/biosamples/groups/SAMEG324828> . Data from *Zea mays*.

ELIXIR-NL: <https://www.ebi.ac.uk/biosamples/groups/SAMEG322905>. Data from a breed between *S. tuberosum* and *S. tuberosum* Group Phureja .

ELIXIR-PT: <https://www.ebi.ac.uk/biosamples/groups/SAMEG323204> (embryos).

<https://www.ebi.ac.uk/biosamples/groups/SAMEG327884> (leaves, stems, roots). Data from *P. pinaster*.

ELIXIR-SI: <https://www.ebi.ac.uk/biosamples/groups/SAMEG322964>. Data from *S. tuberosum*.

ELIXIR-UK: <https://www.ebi.ac.uk/biosamples/groups/SAMEG324805> (seeds).

<https://www.ebi.ac.uk/biosamples/groups/SAMEG322904> (leaves). Data from *B. napus* derived from the Brassica Information Portal.

The corresponding genotype and phenotype data submitted to the public archives is described below:

	Genotype	Phenotype (could include transcriptomic data)	Sample meta data
Cereal: maize (BE+FR)	VIB: Markers Imputed SNPs INRA: Genotyping Experiment S1P9	VIB: phenotyping experiments Transcriptomics: E-MTAB-3173	VIB: SAMEG328646 SAMEG328664 INRA:

	Bouchet 2013	E-MTAB-3758 E-MTAB-3965 INRA : ASSO S1P9 FFL W8 Average:Gif-sur-Yvette,Saint-Martin-de-Hinx,Einbeck	SAMEG324828
Solanaceae: potato (NL+SI)	WUR: http://edepot.wur.nl/165211	WUR: E-MTAB-701/ NIB: GSE58593/E-GEOD-58593/TID(FAIRDO MHUB)	WUR: SAMEG322964 NIB: SAMEG322905
Woody plant: pine (PT)		Embryos: E-GEOD-32551	Embryos: SAMEG323204 Leaves, stems, roots: SAMEG327884
Woody plant: poplar (FR)	https://urgi.versailles.inra.fr/GnpSNP/snp/genotyping/form.do#results/experimentIds=14	https://urgi.versailles.inra.fr/ephehis/ephehis/viewer.do#dataResults/trialIds=822,823,824,825,826	
Brassicaceae: rapeseed (UK)	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA309367 / https://bip.earlham.ac.uk/data_tables?model=plant_lines&query[plant_populations.id]=176	Seeds: https://bip.earlham.ac.uk/trial_scorings/48 Leaves: https://bip.earlham.ac.uk/trial_scorings/47	Seeds: https://www.ebi.ac.uk/biosamples/groups/SAMEG324805 Leaves: https://www.ebi.ac.uk/biosamples/groups/SAMEG322904

8. Appendix 2: List of Presentations on the Work Described

Presentations in International Conferences

Keynote Presentation at the Integrated Bioinformatics/NETABB meeting, Bari, Italy, October 2015

Presentation at the Plant and Animal Genomes Meeting, San Diego, USA, January 2016, in the workshop on *Challenges and Opportunities in Plant Science Data Management*.

Presentation at the Plant Biology Europe EPSO/FESPB 2016 Congress, Prague, Czech Republic, 26-30 June 2016, in the session *Use of Data Standards in Plant Science*.

Presentations at the Plant and Animal Genomes Meeting, San Diego, USA, January 2017 in the workshops on *Connecting Crop Phenotype Data* and *Interoperability and Federation Across Bioinformatic Platforms and Resources*.

Posters in International Conferences

Durand S et al (2017) Findable, Accessible, Interoperable and Reusable data linked to genetic resources for plant biology and breeding. Eucarpia Genetic Resources, Le Corum, Montpellier, France, 8–11 May 2017

Horro Marcos C et al (2017) Brassica Information Portal and Elixir Excelerate. Elixir All-hands 2017, Rome, Italy, March 21-23, 2017.

Durand S et al (2016) Contribution to a European Roadmap for data standardisation in Plant Sciences. EPSO Plant Biology Europe EPSO/FESPB 2016 Congress, Prague, Czech Republic, June 26–30, 2016

Miguel CM et al (2016) Annotation of plant phenotypic data. IUFRO Genomics and Forest Tree Genetics, Arcachon, France, 30 May-3 June 2016.

Chaves I et al (2016) ELIXIR Portugal: Integrating genomic and phenotypic data for woody plants. EU COST Action FA1306 - the quest for tolerant varieties: phenotyping at plant and cellular level. 2nd general meeting, Copenhagen, Denmark, 18-20 April 2016.

Presentations to Specific Projects

Presentation to the CORBEL project use case 4, Villefranche, France, February 2016.

Presentation to the Trees4Future project final conference *Designing trees for the future: data is the cornerstone*. Brussels, Belgium, April 2016.

Annual meeting of the French crop genebanks, San Giuliano, Corsica, France, 2016, November 27-28th

Presentation to EBI Industry Workshop on Ontologies in Agriculture, Food and Nutrition, Hinxton, UK, April 2017.

Presentation at the Wheat Initiative Phenotyping Expert Working Group, Vienna, Austria, April 22th 2017.

Presentations (2) at the first e-ROSA Stakeholder Workshop, Montpellier, France, July 2017.

Invited Seminars

Seminar at the Gregor Mendel Institute, Vienna, Austria, October 2015.

Biotalent lecture, Institute for Plant Genetics-Polish Academy of Sciences, Poznan, Poland, March 2017.

Other Presentations

Presentation to the Genetic Resources/Bioinformatics EU-China Working group, Brussels, Belgium, September 2016.

9. Appendix 3: Proposal for MIAPPE v1.1

MIAPPE v1.1 proposal for assessment

An extended set of recommendations for metadata and phenotypic data annotation was developed by the EXCELERATE work package 7 team based on the publicly available version of MIAPPE (Ćwiek-Kupczyńska et al. 2016; DOI: 10.1186/s13007-016-0144-4), as presented below. Some Attributes were edited and new ones were added. Definition for Attributes in each of the 10 sections (Study metadata, Timing and location of study, Biosource, Environment-growth facility, Environment-rooting conditions, Environment-nutrients, Treatments, Experimental design, Sample collection, processing and management, Observed variables) were included, as well as examples and ontologies/data types that should be used to describe the attributes. Some general guidelines are also provided below for a consistent use of the standard.

Conventions for the Sample MIAPPE metadata:

- 1) If there is a "**Derived Material**" attribute, it should be unique in each sample of the investigation
- 2) If there is not unique value or non-existent values for "**Derived Material**", then the "**Material Source**" should be unique per sample in the investigation UNLESS in the investigation there are time series
- 3) For trees: if there is value for either **latitude** or **longitude** there should be value for both.
- 4) For trees: if there is value for **altitude** there should be value for all three: **latitude, longitude and altitude**.
- 5) for the definition of the ontologies, the FULL URL of the term is expected
- 6) "**Material Source**" and "**Derived Material**" cannot have the same value in 1 sample
- 7) The "**Environment: XXX**" sections should describe the environmental parameters that are measured but are not modified as an experimental factor
- 8) The "**Treatment**" section should describe a plant experimental condition (EO:0007359) or set of conditions describing the application of an abiotic (EO:0007191) or biotic plant treatment (EO:0007357) or the combinatorial application thereof. The treatment (or Factor) is declared and described at the study level (e.g. Experimental field X) and varies in the different assays of the study (e.g. different plots or plants in the experimental field).

MIAPPE Check list	Definition	Example	Ontology/data type
Study metadata			
Unique identifier of study¹	Key-value pair comprising the unique name of the institution/database hosting the submission of the study data, and the identifier of the study in that institution. To be filled by the submitter	EBI:12345678	xsd:string
Submission date*	Date of submission of the study data to the host institution (given by the host database)	02-26-2006	xsd:date
Title of study*	Human readable string summarising the study	Adaptation of Maize to Temperate Climates: Mid-Density Genome-Wide Association Genetics and Diversity Patterns Reveal Key Genomic Regions, with a Major Contribution of the Vgt2 (ZCN8) Locus	xsd:string
Description of study¹	Human readable string describing the study	Time series response of potato cv. Désirée, which is tolerant to PVY infection, was analysed in both inoculated as well as upper non-inoculated leaves. Additionally, transgenic plants deficient in accumulation of salicylic acid (NahG-Désirée) were studied in the same	xsd:string

		setting.	
Public Release date*	Date of first public release of the study data. This is filled in by the database when the submission becomes public	02-30-2006	xsd:date
Associated publication/s	DOI of literature publication where the study is described	DOI: 10.1186/1471-2229-13-123	DOI
Name and address of the laboratory*	Name and address of the laboratory where the study took place	Forest Biotech Lab, IBET, Av. da República, Quinta do Marquês, 2780-157 Oeiras Portugal	xsd:string
Data submitter contact (email)	The mail address of the data submitter	ichaves@itqb.unl.pt	xsd:string (valid email address)
Data submitter identifier	The ORCID id of the data submitter	orcid.org/0000-0002-7054-800X	ORCID id
Study data file link	This is Study level metadata. It is the link to the data files of the study in the appropriate database	http://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-32551/	xsd:anyURI
Timing and location of study			
Timing: start of study*	Date and, when relevant, time on which the experiment started	09-27-2006	xsd:dateTime
timing: end date of study	Date and, when relevant, time on which the experiment ended	12-27-2006	xsd:dateTime
Geographic location of study (country)*	Defines the country/ies where the experiment took place (list-valued attribute)	Porto (Portugal), Cambridge (UK)	Geonames
Experimental site name	Natural site/experimental field/greenhouse/phenotyping facility name where the experiment took place (if applicable)	Thurso research station, north field / redwood forest , California, USA	xsd:string
Geographic location of study (latitude)*	Degrees and minutes followed by north (N) and south (S) of the natural site/experimental field/greenhouse/phenotyping facility where the experiment took place. This can be a random GPS location of the site or the GPS of its entrance.	39°4'N	xsd:string
Geographic location of study (longitude)*	Degrees and minutes followed by east (E) and west (W) of the natural site/experimental field/greenhouse/phenotyping facility where the experiment took place	8°44'W	xsd:string
Geographic location of study (altitude)	Elevation expressed in meters (m) above sea level of the natural site/experimental field/greenhouse/phenotyping facility where the experiment took place	100 m	xsd:string

Biosource			
Organism*	The taxon id of the species as defined by NCBI	4577	NCBI Taxonomy id
Intraspecific name	Key (name of the rank): value (value of the rank) pairs. Ranks can be among the following terms: subspecies, cultivar, variety, subvariety, convariety, group, subgroup, hybrid, line, form, subform	subspecies:vinifera;cultivar:Pinot noir	xsd:string
Time factor	Significant date and time, e.g. planting date. key-value pairs to describe the type of date.	http://www.croponontology.org/terms/CO_715:0000033:04-06-2010	Crop Ontology:'CO_715:0000006' + xsd:dateTime
Life stage	Describes the life stage of the sample at the time of the experiment. Takes values of BBCH scales or PO ontology	http://www.croponontology.org/terms/PO:0009009;BBCH-17	BBCH / Plant Ontology
Material Source: Holding Institute/Stock Centre, accession*	Key value pairs of holding Institute (institute/database that gives the accession number or name), accession number or name (describes a record in a genebank or laboratory). In forestry, provenance or region of provenance may be used when accession is not available. Naming of accessions derived from sexual reproduction in plants:: "mother_accession X father_accession". Only 2 parents allowed only ; if father is unknown, format is "mother_accession X UNKNOWN".	INRA:W95115_inra ; ICNF:PNB-RPI	xsd:string
Material source: DOI	Digital Object Identifier (DOI) of the accession or provenance for trees	doi:10.15454/1.4658436467893904E12	DOI
Derived Material'	Key-value pair of derived material identifier and holding institute (institute that gave the identifier to the derived material) ; derived material from an accession: seed or plant lots, any sample collected on or from the accession and that has been phenotyped.	INRA:W95115_inra_2001;INRA:inra_kernel_2351;Rothmasted:rres_GK090847	xsd:string
Treatment	Treatment/s made for all the samples of the study	http://purl.obolibrary.org/obo/EO_0007210:PVY(NTN);	Plant Environment Ontology + xsd:string
Derived material for trees: Geographic location	Degrees and minutes followed by north (N) and south (S) of the location of the source sample	39°4'N	xsd:string

(latitude)			
Derived material for trees: Geographic location (longitude)	Degrees and minutes followed by east (E) and west (W) of the location of the source sample	8°44'W	xsd:string
Derived material for trees: Geographic location (altitude)	Elevation expressed in meters (m) above sea level of the location of the source sample	10 m	xsd:string
Environment: Growth facility			
Type of growth facility*	Environment in which the trial or experiment or characterization or evaluation is carried out.	field environment condition; greenhouse environment condition; greenhouse then field condition; growth chamber (GC)	Crop Ontology: "CO_715:0000005"
Average day temperature	The air temperature during the day (light conditions).	22 °C	xsd:float
Average night temperature	The air temperature during the night (dark conditions)..	18 °C	xsd:float
Change over the course of experiment	Difference between the maximum air temperature recorder and the minimum.	0.75 °C	xsd:float
Average daily integrated photosynthetic photon flux density (PPFD) measured at plant or canopy level.	Photosynthetic photon flux density (PPFD) over a 24-h period	plant PPFD: 61 mol m ⁻² d ⁻¹ ; canopy PPFD: 40 mol m ⁻² d ⁻¹	xsd:string
Average length of the light period	Average length of the light period in h.	16	xsd:float
Light intensity	Intensity of total light	[μmol m ⁻² s ⁻¹]	xsd:float
Range in peak light intensity	Range in peak light intensity.	[μmol m ⁻² s ⁻¹]	xsd:float
Fraction of outside light intercepted by growth	Fraction of outside light intercepted by growth facility components and surrounding structures.	[μmol m ⁻² s ⁻¹]	xsd:float

facility components and surrounding structures			
Type of lamps used	Nature of the light source for controlled environments. XEO:00137	fluorescent tubes; high intensity discharge (HID) lamps; light emitting diodes (LED)	xsd:string
R/FR ratio	Red light to far red light ratio. XEO:00036	[mol mol ⁻¹]	xsd:float
Daily UV-A radiation	Defines the intensity of UVA radiation (320-400 nm); XEO:00037	[W m ⁻²]	xsd:float
Daily UV-B radiation	Defines the intensity of UVB radiation (290-320 nm); XEO:00038	[W m ⁻²]	xsd:float
Total daily irradiance	Defines the intensity of total light (XEO:00034).	[W m ⁻²]	xsd:float
Atmospheric CO2 concentration	Denotes whether the atmospheric CO2 concentrations were controlled during the experiment.	controlled; uncontrolled	xsd:float
Average CO2 during the light and dark periods	Defines the concentration of CO2 in the air during the light and dark periods (XEO:00023)	light period: 390 mL-1; dark period: 450 mL-1	xsd:string
Average VPD_{air} during the light period	The Vapour Pressure Deficit in the air defines the difference between the maximal amount of water in the air minus the actual amount during the light period in kPa (XEO:00021)	2 kPa	xsd:float
Average relative humidity during the light period	The relative humidity describes the amount of water vapor in the air, generally expressed as the percentage of the maximum water vapor during the light period (XEO:00020)	30%	xsd:float
Average VPD_{air} during the dark period.	The Vapour Pressure Deficit in the air defines the difference between the maximal amount of water in the air minus the actual amount during the dark period in kPa ((XEO:00021)	2.7 kPa	xsd:float
Average relative humidity during the dark period	The relative humidity describes the amount of water vapor in the air, generally expressed as the percentage of the maximum water vapor during the dark period (XEO:00020)	33%	xsd:float
Environment: Rooting conditions			
Rooting medium	An abiotic plant treatment (EO:0007191) involving the use of a solid or liquid substrate for growing plants or tissue-cultured plant samples.	hydroponic plant culture media; in vitro liquid growth medium; in vitro solid growth medium; soil environment	Plant Environment Ontology:'EO_0007147'

Container type	Defines the type of container used to grow/treat the plants.XEO:00040	pot; Petri dish; well; tray	xsd:string
Container volume	Defines the volume that is available to the roots. XEO:00113	[L]	xsd:float
Container height	Defines the height of the container.	[m]	xsd:float
Number of plants per containers	Defines the number of plants per container. XEO:00112	X/container	xsd:integer
Plot size	Description of experimental sites.	higher-level landform; land element and position; slope;	Crop Ontology:'CO_715:0000058'
Sowing density	Sowing density.	x/plot	Natural Resource and Environment Ontology
Rooting medium replenishment	Frequency and volume of replenishment or addition of the rooting medium.		xsd:string
pH	Value of soil pH, separated by a colon, the depth (cm) from where soil sample was taken. Multiple values are separated by semicolon. For hydroponics, leave the depth empty.	7.7:40-60; 6.5; 4.3:10-20	xsd:string
Porosity	A permeability quality inhering in a bearer by virtue of the bearer's disposition to admit the passage of gas or liquid through pores or interstices. PATO:0000973	[%]	xsd:float
Medium temperature	Temperature of the replenishment medium.	[°C]	xsd:float
Soil penetration strength	Soil penetration strength as measured by the standard penetration test (SPT; ISO 22476-3), the cone penetrometer test (CPT), in-situ vane shear tests, and shear wave velocity measurements.	[Pa m ⁻²]	xsd:float
Water retention capacity	Defines the potential energy of water per unit mass of water in the soil.XEO:00126	[g g ⁻¹ dry weight]	xsd:float
Organic matter content	Proportion of organic matter in the soil. XEO:00117	[%]	xsd:float
Environment: Nutrients			
Médium composition	Concentration of the nutrients	Ca (XEO:00058): 5 mg/L	XEML Environment

			Ontology:'XEO_00042' + xsd:float
Extractable N content per unit ground area before fertiliser added	Extractable N content per unit ground area before fertiliser added	[mg/m ²]	XEML Environment Ontology:'XEO_00054' +xsd:float
Type and amount of fertiliser added per container/m²	The current practice in field /greenhouse management for fertilization	nitrogen: [concentration]; phosphorus: [concentration]	Crop Ontology:'CO_715:0000 204' + xsd:float
Concentration of [nutrient] before start of the experiment	Concentration of a nutrient at the start of an experiment.	Ca (XEO:00058): 5 mg/L	XEML Environment Ontology:'XEO_00042' + xsd:float
Extractable N content per unit ground area at the end of the experiment	Extractable N content per unit ground area at the end of the experiment	[mg/m ²]	XEML Environment Ontology:'XEO_00054' +xsd:float
Volume and frequency of water added per container/m²	A defined volume of water supplied to pots of a defined size.	[L/m ²]	xsd:float
Matrix potential	Range in water potential for soil.	-10 to -30 kPa	xsd:float
Watering regimen	The treatment involving an exposure to watering frequencies.	irrigation from top; irrigation from bottom; drip irrigation	xsd:string
Composition of nutrient solutions used for irrigation	For all nutrients, the ontology term with concentration.	Ca (XEO:00058): 5 mg/L	XEML Environment Ontology:'XEO_00042' + xsd:float
Composition of the salts	For all nutrients, the ontology term with concentration.	[mol L ⁻¹]	XEML Environment Ontology:'XEO_00042' + xsd:float
Electrical conductivity	A conductivity quality inhering in a bearer by virtue of the bearer's ability to convey electricity.	[dS m ⁻¹]	xsd:float

Treatments			
Seasonal environment	A plant treatment (EO:0001001) involving an exposure to a given conditions of regional seasons.	Spring season; dry season	Plant Environment Ontology:'EO_0007038'
Air treatment regime	The treatment involving an exposure to wind/air with varying degree of temperature, which may depend on the study type or the regional environment.	28/25°C (Day/Night)	Plant Environment Ontology:'EO_0007161'
Soil temperature regime	A physical plant treatment (EO:0007316) involving an exposure to varying degree of temperature, which may depend on regional environment.	27/25°C (Day/Night)	Plant Environment Ontology:'EO_0007161'
Soil treatment regime	The treatment (EO:0007049) involving growing plants and exposing them to soil growth media with varying contents	sand content (10% v/v)	Plant Environment Ontology:'EO_0007161'
Antibiotic regime	A chemical treatment (EO:0007189) involving the use of antibiotic for selection purposes.	actinomycin D; 20mM;20ml per plant; Every week	Plant Environment Ontology:'EO_0007041'
Chemical administration	An abiotic plant treatment (EO:0007191) involving the application of chemical(s).	Bion; 13,5mM; 5ml per plant; Every 15 days.	Plant Environment Ontology:'EO_0007189'
Biotic treatment	A plant treatment (EO:0001001) involving the application of a biotic or biological factor such as a microbe, insect, animal, or plant or a combination thereof	rice tungro bacilliform virus (RTBV) 2.5 µl, incubated at room temperature for 10min	Plant Environment Ontology:'EO_0007357'
Fertilizer regime	A plant nutrient treatment (EO:0007241) involving the use of a fertilizer, a combination of plant nutrients.	Potassium phosphate; 50 Kg P.Ha/y 50 Kg K.Ha/y	Plant Environment Ontology:'EO_0007085'
Fungicide regime	A treatment (EO:0007167) involving the application of a fungicide; a chemical entity or mixture of chemical entities.	Benzothiadiazole; 10mM; 1ml; Every month	Plant Environment Ontology:'EO_0007268'
Gaseous regime	A physical plant treatment (EO:0007316) involving the application of a gas or a combination of gasses.	Carbon Dioxide; 20ppm	Plant Environment Ontology:'EO_0007023'
Gravity	The treatment involving use of gravity factor to study various types of responses in presence, absence or	Zero gravity (International space station)	Plant Environment Ontology:'EO_0007146'

	modified levels of gravity.		
Growth hormone regime	A chemical treatment (EO:0007189) involving the use of growth hormones to study various types of responses on their extrinsic and/or intrinsic application.	Jasmonic acid; 1mM;20ml;	Plant Environment Ontology:'EO_0007165'
Herbicide regime	A treatment (EO:0007167) involving the application of a herbicide; a chemical entity or mixture of chemical entities.	SUREWET (Polyvinyl polymer and nonionic surfactant); 1,75mM; 5ml per plant; Sprayed every month	Plant Environment Ontology:'EO_0007183'
Mechanical treatment	A treatment involving the application of a mechanical force	Wounding, bending	Plant Environment Ontology:'EO_0007373' / xsd:string
Mineral nutrient regime	A chemical treatment (EO:0007189) involving the application of inorganic chemical(s).	Cd 0.5 mg/L (Hydroponics), CdCl ₂ 15mg.Cd/kg (soil)	Plant Environment Ontology:'EO_0007044'
Humidity regimen	A treatment involving an exposure to varying degree of humidity, which may depend on regional environment.	56%/70% (Day/Night)	Plant Environment Ontology:'EO_0007359'
Non-mineral nutrient regimen	Treatment involving the exposure of plant to molecular forms of nutrient as supplement to study various types of responses.	Low Carbon - Sucrose concentration	Plant Environment Ontology:'EO_0007043'
Radiation (light, UV-B, X-ray) regime	A physical plant treatment (EO:0007316) involving an exposure with a radiation type, intensity or quantity. EMR is classified according to the frequency of its wave. The electromagnetic spectrum, in order of increasing frequency and decreasing wavelength, consists of radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays and gamma rays. (from Wikipedia).	200-280nm; 30min; every day	Plant Environment Ontology:'EO_0007151'
Rainfall regime	Treatment involving an exposure to a given amount of rainfall.	79 rainfall events; 15,6mm (mean size)	Plant Environment Ontology:'EO_0007181'
Salt regime	This treatment may be used to simulate the growth conditions of sea coast regions and saline/sodic soils.	150mM	Plant Environment Ontology:'EO_0007185'

	A chemical treatment (EO:0007189) involving use of salts as supplement to liquid and soil growth media to study various types of responses on their application.		
Watering regime	Treatment involving an exposure to watering frequencies.	20ml every 3 days	Plant Environment Ontology:'EO_0007383'
Water temperature regime	Treatment involving an exposure to water with varying degree of temperature, which may depend on regional environment.	20°C	Plant Environment Ontology:'EO_0007160'
Standing water regime	The treatment involving an exposure to standing water during a plant's life span. This also results in anaerobic soil conditions for either long or short periods.	Flooding water, Deep water	Plant Environment Ontology:'EO_0007282'
Pesticide regime	A chemical treatment (EO:0007189) involving the application of a pesticide; a chemical entity or mixture of chemical entities.	Glyphosate; 1.68 kg acid equivalent (a.e.) / ha	Plant Environment Ontology:'EO_0007167'
pH regime	The treatment involving exposure of plants to varying levels of pH of the growth media.	acidic pH soil environment	Plant Environment Ontology:'EO_0007171'
Other perturbation			xsd:string
Experimental design			
Spatial coordinates: ID	name of the column head of the data file giving the spatial coordinates of an assay in the study	ex: latitude, longitude or X, Y	xsd:string
Spatial coordinates: type	georeference or line/column	can be repeated: ex: long, lat	xsd:string
Unit Spatial coordinates: unit	Unit used to describe the spatial coordinate of an assay in the study (e.g.degree, minute, number)		xsd:string
Time coordinates - ID	Times series or cinetics: name of the column head giving the time at which a set of measures has been collected		xsd:string
Time coordinates - Unit	Times series or kinetics: unit of time used	can be repeated. Date Timestamp Growing degree day (GDD)	xsd:string
Experiment description*	Text description of the unit of observation or assay (=		xsd:string

	genotype * factor combination) in a study.	1/ each maize line is observed on 15 rows with a density of 6 plant per square meters 2/ Observation on the assay are mean values of all repetition for a given genotype * factor combination	
Replication - Technical replication ID	IDentifier of each level of replication for a given assay	ex: block:1, plot:894, rep:1	xsd:string
Replication - technical replication level hierarchy	hierarchy of the different levels between each others (text: no ontology)	ex: block>rep>plot	xsd:string
Type of statistical design	Short description of the statistical design	ex: completely randomized design (CO_715:0000146)	xsd:string
Sample collection, processing, management			
BiosampleID	When the sample submitted to Biosamples DB this is the unique Biosample id	SAMEA4202911	xsd:string
Plant structure development stage	A stage in the life of a plant structure (PO:0009011) during which the plant structure undergoes developmental processes.	Fruit ripening stage	Plant Ontology:'PO_0009012'
Plant product	A portion of organism substance that is or was part of a plant	Resin	Plant Ontology:'PO_0025161'
Organism count	The number/(amount) of entities of this type that are part of the whole	10 roots; 100mg of roots	xsd:integer
Sample temperature	Temperature at Collection / Harvesting	20°C	xsd:float
Oxygenation status of sample	A setting datum that specifies the oxygenation inside a container participating in a storage process.	Vacuum	xsd:string
Sample salinity	A setting datum that specifies the salinity inside a container participating in a storage process.	20mM NaCl	xsd:string
Sample storage duration			xsd:string
Sample storage location	Top level of all location classes.	Campus, Building, Room	xsd:string
Sample storage	A setting datum that specifies the temperature inside a container participating in a storage process.	-18 °C to -35 °C, -60 °C to -85 °C, Liquid nitrogen, Room temperature etc	xsd:string

temperature			
Sampling time	The date-time when the sample was collected / harvested	2005-08-15T15:52:01+00:00	xsd:dateTime
Observed variables			
Trait	Name of the targeted trait or id in a relevant ontology	ex: Anthesis time (CO_322:0000030) or reproductive growth time (TO:0000366)	Plant Trait Ontology / Crop Ontology / XML Environment Ontology xsd:string
Variable name*	https://bioportal.bioontology.org/ontologies/CO (Trait_Dictionary_V5 definition): Name of the variable following the convention <trait abbreviation>_<method abbreviation>_<scale abbreviation>. Variable name must be unique.	ex: Growing Degree days to anthesis	Crop Ontology / xsd:string
Variable id	https://bioportal.bioontology.org/ontologies/CO	ex: CO_322:0000260	Crop Ontology / xsd:string
Source of the variable id	version of the ontology describing the variable	ex: Maize Trait Dictionary in template 5 - CIMMYT- December 2015	xsd:string
Variable Method*	https://bioportal.bioontology.org/ontologies/CO (Trait_Dictionary_V5 definition): (Short) name of the method	ex: MFLW8	xsd:string
Method description	https://bioportal.bioontology.org/ontologies/CO (Trait_Dictionary_V5 definition): Textual and generic description of the method. Optionaliy : extensaion of the Method of the Ontology with growth stage, inoculation precise organ (leave number)	ex: 1/ From Ritchie J, NeSmith D (1991) Temperature and crop development. Modeling plant and soil systems American Society of Agronomy Madison Wisconsin USA. doi:10.2134/agronmonogr31.c4 with TBASE=8°C and T0=30°C 2/ Extension of Method CO_321:0000456 (Plant Height measured with ruler) : Plant height	xsd:string

		measure at 5 years, one year after Botritis inoculation.	
Reference associated to the method	https://bioportal.bioontology.org/ontologies/CO (Trait_Dictionary_V5 definition): Bibliographical reference describing the method.	ex: Ritchie J, NeSmith D (1991) Temperature and crop development. Modeling plant and soil systems American Society of Agronomy Madison Wisconsin USA. doi:10.2134/agronmonogr31.c4	DOI / xsd:string
Scale*	https://bioportal.bioontology.org/ontologies/CO (Trait_Dictionary_V5 definition): Name of the scale associated to the variable	ex: GDD: Growing Degree-Days	xsd:string

¹Mandatory information when preparing a submission to BioSamples database (<http://www.ebi.ac.uk/biosamples>)

*Mandatory information