



EXCELERATE Deliverable 7.2

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:	Phenotypic data exposed through integrating API by all participants	
WP No.	7	
Lead Beneficiary:	18. IBET	
WP Title	Use Case B: Integrating Genomic and Phenotypic data for Crop and Forest Plants	
Contractual delivery date:	31 August 2018	
Actual delivery date:	31 August 2018	
WP leader:	Célia Miguel, Cyril Pommier	18-. IBET, 27- INRA
Partner(s) contributing to this deliverable:	28 - CIRAD, EI, 1- EMBL, 16- FCG, 18 - IBET, 27- INRA, 6 - NBIC, 33 - NIB, 37 - VIB	

Authors and Contributors:

Célia Miguel, Cyril Pommier, Richard Finkers, Frederik Coppens, Dan Bolser, Manuel Ruiz, Anne-Françoise Adam-Blondon, Kristina Gruden, Živa Ramšak, Evangelia Papoutsoglou, Daniel Faria, Bruno Costa, Inês Chaves, Robert Davey

Table of contents

1. Executive Summary	2
2. Impact	3
3. Project objectives	4
4. Delivery and schedule	5
5. Adjustments made	5
6. Background information	5
Appendix 1: Phenotypic data exposed through integrating API by all participants	9
A1.1 Introduction	9
A1.2. Report	9
Appendix 2: List of Presentations on the Work Described	16
A2.1. Presentations in International Conferences	16
A2.2 Posters in International Conferences	17
A2.3 Other Presentations	17

1. Executive Summary

High-throughput “omics” technologies are widely used and increasingly important to support plant biology research and breeding of diverse plant species for production of food, feed, fibre and other biomaterials, and bio-energy. Significant advances in plant science can be obtained from the integration of available genomic and genotyping data with diverse types of phenotyping data, including field or greenhouse experimental data, molecular, -omics and image data. Although most -omics data, and especially phenomic data, are being generated in increasing scale, either from public or private research institutes, the dispersion of datasets and metadata among multiple repositories and their often poor description and annotation, make their use and exploitation still challenging or even unapproachable.

To help unlocking the full potential of a multi-omics approach to plant science, the overarching goal of this workpackage is to make plant data interoperable in accordance with the FAIR principles (i.e. Findable, Accessible, Interoperable and Re-usable). Hence, several standards have been built these past years for the annotation of data sets, and the Elixir plant community is coauthoring most of them. Their use has been demonstrated by exemplary application to data from diverse species published in public repositories. Building from this previous work (Deliverable 7.1) we have now implemented an architecture for supporting access to distributed datasets annotated according to the developed standards. This architecture uses the Breeding API (BrAPI, www.brapi.org), an API for accessing data relevant for plant breeding developed by the international plant community.

We have implemented 7 API endpoints over datasets held at the several ELIXIR nodes participating in the Plants Use Case, thereby publicly exposing these datasets and making them accessible and contributing for their interoperability and re-use. We further contributed to the evolution of BrAPI specification by working on the alignment with the MIAPPE (www.miappe.org) standard for plant phenotypic data, based on the experience with our exemplar data sets, and the needs of the Use Case. This work represents an important intermediate step towards the full implementation of an ELIXIR Plant Data Lookup Service currently under development.

2. Impact

This deliverable aims to expose MIAPPE-compliant phenotypic datasets by all participating ELIXIR nodes through the integrating Breeding API (BrAPI). The goal of the BrAPI project is the development and implementation of a Web service API for data exchange focused on plant material, phenotyping and genotyping data types, mainly for breeding purposes. It has been initiated by Consultative Group on International Agricultural Research (CGIAR), the James Hutton Institute and others with an early and strong contribution from EXCELERATE WP7. The implementation of BrAPI endpoints described in this deliverable involved a close interaction with the BrAPI community. Our implementation and previous and ongoing work to align the activities on MIAPPE and BrAPI has been simultaneously contributing to the evolution of BrAPI specification. A publication with the recent updates on BrAPI specification is currently under preparation for submission, with co-authorship of several members of this workpackage (ELIXIR-FR, ELIXIR-NL, ELIXIR-PT). It is expected that the contribution and involvement of the EXCELERATE members in this global community effort have a positive impact in enabling the interoperability among plant databases held at different locations, which is the ultimate the goal of the workpackage. Moreover, this work has been presented in several international meetings, national meetings and invited seminars (see Appendix 2). The overall impact will depend on the adoption and use of BrAPI by the wide community of plant data producers and users. This adoption is strongly encouraged within ELIXIR, EMPHASIS (European Plant Phenotyping Infrastructure) and the CGIAR. The ELIXIR partners are strongly collaborating with EMPHASIS on BrAPI evolutions and adoption.

To help promoting the adoption and use of BrAPI, the workpackage members have been actively involved in several initiatives.

Together with WP5, ELIXIR BE and ELIXIR NL WP7 partners organized a BrAPI BYOD/hackathon in Ghent, Belgium (May 30 - June 1 2017), to showcase the potential of FAIR data in the context of plant research and a hackathon on how to make the interoperability platform within the plant domain (BrAPI) FAIR reusing the general accepted MIAPPE standards and assessing data validation with ISA-TOOLS. This event was attended by 22 participants. ELIXIR-FR and Phenome-EMPHASIS.fr also organised

another BrAPI hackathon in Versailles (February 2018) for 30 attendees from CGIAR, EMPHASIS and ELIXIR. This allowed ELIXIR partner to make major contribution to phenotyping specifications and to integrate this with MIAPPE.

EXCELERATE WP7 partners attended four BrAPI Hackathon, with invitation funding from the Bill & Melinda Gates Foundation for three of them. There were between 25 and 35 attendees, with 2 to 5 ELIXIR attendees.

As part of the ELIXIR capacity exchange program, teams from ELIXIR-FR, ELIXIR-PT, ELIXIR-NL and ELIXIR-IT have been collaborating for enhancing the implementation of BrAPI standard in ELIXIR and semantifying it. Two F2F meetings took place in Wageningen (NL, Dec 2018) and Oeiras (PT, April 2018) to (1) transfer tools and knowledge to the Italian node (ELIXIR-IT) for implementing a BrAPI endpoint, and (2) Improve and facilitate BrAPI global adoption, further establishing it as the standard to make plant breeding data interoperable. During this staff exchange program, significant progress was achieved towards the alignment of MIAPPE, BrAPI and ISA and the establishment of a formal and standardized controlled vocabulary for BrAPI.

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

The Divseek initiative aims at sharing Plant Genetic Resources data worldwide through application of FAIR data principles. This is being implemented by reusing the existing BrAPI endpoints and helping the implementation of new ones. It will add a semantic layer to enable FAIR for the machine through JSON-LD. Several EXCELERATE WP7 partners are part of Divseek.

In total EXCELERATE WP7 partners attended or organised 8 workshops related to this deliverable, with approximately 30 attendees for each.

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

None

6. Background information

Background information on this WP as originally indicated in the description of action (DoA) is included here for reference.

Work package number	7	Start date or starting event:	month 1
Work package title	Use Case B: Integrating Genomic and Phenotypic data for Crop and Forest Plants		
Lead	Cyril Pommier, INRA; Célia Miguel, IBET		
Participant number and person months per participant			
1- EMBL 12.00; 3 - TGAC 21.20; 6 - NBIC 0.00, DLO (LTP) 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			
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.

Work Package Leads: Cyril Pommier, INRA (since 25/01/2018); Célia Miguel, IBET (since 01/03/2017)

Description of work and role of partners

WP7 - Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants [Months: 1-48]

INRA, EMBL, TGAC, NBIC, FCG, IBET, CIRAD, NIB, VIB

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.

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 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 metadata. 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)

Appendix 1: Phenotypic data exposed through integrating API by all participants

A1.1 Introduction

The improvement of plant characteristics and traits for the benefit of mankind has been ongoing for thousands of years. However, the current capacity to speed up plant breeding is higher than ever before due to the power of current plant -omics technologies. Improvement of traits such as disease resistance or increased yield under specific environmental conditions can be tackled making use of available genomic/ genotyping and phenotyping data for a growing number of crops. The analysis of genotype-phenotype associations can be very complex since desired phenotypic outcomes can result from single or multiple genes, and from their interaction with each other and with the environment. The systematic study of phenotypes on a genome-wide scale, and its association with genomic/genotypic information under a range of environmental conditions is being increasingly adopted within breeding programs in diverse plant species.

Whether phenotypic data are obtained from field or greenhouse experimental settings, or from phenotyping platforms allowing to generate volumes of data that are several orders of magnitude higher, a number of gaps often make genotype-phenotype associations still very difficult to approach. Unlike the publicly accessible and curated repositories for DNA and protein data, there is no equivalent public repository for the deposition of large amounts of phenotypic data, which are often stored on multiple systems. In cases where data generated from heterogeneous sources can be found and accessed, they are often of poor value because they are not standardized and lack an adequate annotation, preventing their re-use.

This workpackage has already produced an extended and revised version of the Minimal Information about a Plant Phenotyping experiment (MIAPPE) standard (<http://www.miappe.org>), which was described in Deliverable 7.1. In line with previous work which provided guidelines for a robust annotation of plant phenotypic data and their application to exemplar datasets, we have now implemented several endpoints of an integrating plant Breeding Application Programming Interface, [BrAPI](#), supporting access and exposing datasets annotated according to the developed standards, from all participating nodes.

A1.2. Report

The work developed to achieve the present deliverable is included in the fourth and last task of the work package.

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

As part of this task, we had previously reached milestone M7.4 (specification of an API for data access). We have explored and chosen BrAPI as the interoperability platform for the

exposure of plant phenotypic data by participant nodes within the plant use-case at this stage.

BrAPI specifies a standardized web service interface for plant phenotype/genotype data access to enable interoperability among plant breeding databases, allowing breeders and researchers to exchange, compare and combine data across databases. It is an open API based on the principles of open linked data and a web-based architecture that has been driven by a collaborative community-based global effort. The specification describes the functions and services available for automated access by a computer program. The inputs that are allowed, the structure of the output data, and the protocol used to pass data to a service are also described in the BrAPI specification.

BrAPI calls are organized into categories in alignment with the major domains needed for exchanging information between plant breeding information systems and client applications. In the current version, BrAPI calls cover information about germplasm, phenotypes, experiments, studies, geographic locations, samples, and genetic markers.

BrAPI is still in an actively developing state, and several partners participating in this workpackage have been involved in its development.

The main content of this deliverable is the implementation of integrating API endpoints by all participants to enable exposure of MIAPPE compliant datasets across ELIXIR nodes.

The following endpoints allowing access to multiple datasets from the several participating ELIXIR nodes have been implemented, and are described below:

	Interface	API endpoint	Datasets
BE	PIPPA https://pipppa.psb.ugent.be	https://pipppa.psb.ugent.be/pippa_experiments/brapi/v1/	maize
FR	GnpIS-Ephesis INRA/GnpIS	https://urgi.versailles.inra.fr/gnpis-core-srv/s/wagger-ui.html	several crop and forest sp.
	TropgeneDB http://tropgenedb.cirad.fr	http://tropgenedb.cirad.fr/api/	rice
NL	EU-SOL BreedDB http://www.eu-sol.wur.nl	http://www.eu-sol.wur.nl/brapi/v1/	several Solanaceae sp.
PT	PHENO https://brapi.biodata.pt	https://brapi.biodata.pt/brapi/v1	cork oak, rice

SI	PISA http://pisa.nib.si/brapi/v1/calls	http://pisa.nib.si/brapi/v1	potato
UK	Brassica Information Portal (BIP) https://bip.earlham.ac.uk	https://bip.earlham.ac.uk/api_documentation	Brassica sp.

ELIXIR-BE

- PIPPA: three maize datasets

PIPPA (<https://pippa.psb.ugent.be>), the PSB Interface for Plant Phenotype Analysis, is the central web interface and database that provides the tools for the management of the plant imaging robots on the one hand, and the analysis of images and data on the other hand.

PIPPA communicates with the platforms by transferring experiments created in the PIPPA web interface to the platform. Treatment and genotype information for each pot is defined in the database in order to ensure data integrity across experiments. During the runtime of the experiment, PIPPA retrieves weighing and irrigation measurements, environment data, error logs as well as images from the platform. BrAPI implementation for PIPPA can be found at https://pippa.psb.ugent.be/static/swagger/dist/index.html?url=/pippa_experiments/get_brapi_swagger_spec?0.12.2

ELIXIR-FR

- [INRA/GnplS](#): 20 datasets with 1 000+ studies (maize, wheat, forest trees, grape, etc)

GnplS, and specifically [GnplS-Epheis](https://urgi.versailles.inra.fr/epheis/) (<https://urgi.versailles.inra.fr/epheis/>), is used to publish INRA (Elixir-FR) phenotyping data. It has implemented a Breeding API endpoint with a [swagger](#) (<https://urgi.versailles.inra.fr/gnpis-core-srv/swagger-ui.html>) instance that provides both relevant documentation and a web service testing environment.

The web service layer is implemented in Java using Spring Rest. Its storage layer uses Postgresql and its query layer uses Elasticsearch, a NoSQL system fed by an Extract Transform Load (ETL) toolbox written in Python. This later system has been extended to create the Elixir Data Harvester described below.

All necessary calls have been implemented, covering germplasms, studies, observation variables and location. It is compliant with the Breeding API validator [Brava](http://webapps.ipk-gatersleben.de/brapivalidator/) (<http://webapps.ipk-gatersleben.de/brapivalidator/>). It serves as a reference for the other implementations.

The following datasets are available:

- *Brassica napus* dataset from the Rapsodyn national project. Agronomy, Phenology and Biochemistry data on 69 varieties and accessions.
 - *Populus* dataset from the Popyomics dataset (doi:10.1186/1471-2229-12-173). Phenology, Phenology and Disease data on 336 germplasms.
 - *Triticum aestivum* dataset from a French experimental network (doi:10.15454/1.4489666216568333E12). Agronomic, Quality, Disease and Phenology data on more than 10 experimental locations during more than 15 years and for more than 1700 winter wheat genotypes.
 - *Quercus robur* dataset from the Plantacomp experimental network. Morphology and Phenology data on 1416 germplasms.
 - *Pinus pineae* dataset from the Plantacomp experimental network. Morphological and phenological data on more than 30 000 trees from 36 provenances.
 - *Vitis Vinifera* dataset from experimental network. Phenology and Biochemistry data on 816 germplasms over fifty years.
- TropgeneDB: 30 datasets (rice datasets)
TropgeneDB (ELIXIR-FR, <http://tropgenedb.cirad.fr>) is publishing genomic, genetic and phenotypic information on tropical and Mediterranean crops. BrAPI endpoint has been implemented to provide access to rice plant phenotypic datasets (<http://tropgenedb.cirad.fr/api/>). The Web Service layer is implemented in Python using Flask, Flask-RESTPlus, and Swagger UI.

The following datasets are available:

- Agronomic, morphological and technological traits of 430 genotypes of European rice genetic resource collections: Mean value and genotype × environment interactions (RESGEN, Proceedings of the Eurorice Symposium, Krsanodar, Russia. 38 Sept. 2001)
- Evaluation of resistance to 17 strains of *Pyricularia grisea* on 172 European rice germplasms - Variable measured : percentage of leaf area infected - Data provided are the adjusted means over 3 years.
- Evaluation of salinity tolerance among the European rice germplasm (100 accessions in 2008 and 100 additional accessions in 2009)
- Screening for drought tolerance of 120 varieties (2008) and 109 varieties (2009) of the European rice germplasm.
- Phenotyping of indica rice tolerant to heat (167 accessions, 21 traits)

ELIXIR-NL

- EU-SOL BreeDB: (various solanaceae datasets)
BreeDB is the BrAPI endpoint provided by the ELIXIR-NL node. It aims to provide access to plant phenotypic datasets of Solanaceae in general, with the largest datasets in potato.

All calls needed for exchange of passport data and phenotype data are implemented.

BreeDB is implemented in JavaEE and uses MySQL as the persistent data store. User web access is provided via the url: <http://www.eu-sol.wur.nl> and the BrAPI endpoint under <http://www.eu-sol.wur.nl/brapi/v1/>

ELIXIR-PT

- [PHENO](#): 2 datasets, cork oak and rice

PHENO (<https://brapi.biodata.pt>) is the BrAPI endpoint (<https://brapi.biodata.pt/brapi/v1>) provided by the Portuguese node of ELIXIR. It aims to provide access to plant phenotypic datasets with a major focus on the woody plant domain, although not restricted to it.

PHENO is implemented in Node.js, a JavaScript run-time environment, and data persistence is ensured by MySQL. Express is used as the web application framework to route the calls. Sequelize, a promise-based ORM for Node.js, is used to connect to the database.

In the current version of PHENO, 13 BrAPI calls are implemented including Call Search, Search Studies, Study Germplasm Details, Study Details and Study Observation Variables.

The following datasets are available:

- Cork quality traits in 3 populations of cork oak (*Quercus suber*) (<https://doi.org/10.1371/journal.pone.0169018>).
- Salt stress physiological traits in 392 rice accessions (<https://doi.org/10.1111/pbi.12010>).

ELIXIR-SI

- [PISA](#): 2 solanaceae datasets

PISA (<http://pisa.nib.si/>) is the BrAPI endpoint (<http://pisa.nib.si/brapi/v1/>) implemented at the Slovene ELIXIR node. It aims to provide access to plant phenotypic datasets, primarily focused on Solanaceae plant species described using the MIAPPE standard.

PISA is implemented in Python language using the Django web framework, on the backbone of PostgresSQL object-relational database management system.

Currently, 13 BrAPI calls are implemented, covering primarily germplasm, locations, programs, studies and trials searches. The code for the implementation was made available at [GitHub](#), and is licensed under the GPLv3 license.

ELIXIR-UK

- ELIXIR-UK (EI)

The Brassica Information Portal (BIP) provides a HTTP RESTful API which uses JSON to represent resources. Resources represent database model elements (tables) which constitute BIP Schema. Full access to the API is allowed for all

registered users. BIP (<https://bip.earlham.ac.uk>), which already provides access to phenotypic datasets for several species, has initiated the implementation of BrAPI. This work isn't finished yet, but due to the presence of a Brassica Portal specific API it will be rather easy to finish the BrAPI implementation once the Brassica Portal has filled the necessary positions.

In addition to the Brassica Portal, the infrastructure used for the UK Wheat Portal (<https://grassroots.tools>) as part of the UK-wide wheat improvement programme, Designing Future Wheat (<https://designingfuturewheat.org.uk>), will be added to this list of available endpoints. It will be integrated in a latter deliverable (D7.3). However, it is already hosting a number of crop datasets within an iRODS data grid (<https://irods.org>) in bulk download form via HTTP, WebDAV, or S3 protocols. We will be expanding functionality to support phenotyping datasets marked up with MIAPPE v1.1, deposited within the DFW repository using COPO (<https://copo-project.org>), and an ElasticSearch query layer, within the coming year.

ELIXIR-EMBL

- ELIXIR-EMBL (EBI)

The Elixir Plant Phenotyping datasets need to be linked with the genetic data published at the European Bioinformatics Institute (EBI) to reach the Plant community goal of phenotype to genotype integration. The EBI dataset publication mechanism is centered on “biosamples”. All EBI hosted genetic data are linked to Biosample entries which hold its source description and metadata. Those will be used to enable future search and integration among phenotype and genetic datasets. The EBI does not produce any experimental data itself but it is an archive for primary/raw and interpreted/analysis molecular data. It will therefore be a central resource for the upcoming WP7 federated search to be delivered thanks to D7.3.

The system under development will routinely index the EBI Biosamples database for annotated plant samples using the NCBI taxonomic classification system. Those biosamples will be sent as JSON files including the sample identification, available metadata and URL links to genomic, transcriptomic, expression and variation data in the core EBI databases including the European Nucleotide Archive (ENA), Array Express and the European Variation Archive (EVA). No genetic enabled BrAPI endpoint will be necessary to achieve this.

In addition to the exposure of data in the listed API endpoints, a system for harvesting data from the endpoints and indexing has been already implemented and published (<https://github.com/elixir-europe/plant-brapi-etl-data-lookup-gnpis>). The harvester is developed using the Python language and feeds an Elasticsearch central index which is the backend of the central plant data search engine that enables data discovery on the current result. This backend is available currently in beta (<https://urgi.versailles.inra.fr/beta/gnpis-core-srv/swagger-ui.html>) and will be available in

production (<https://urgi.versailles.inra.fr/gnpis-core-srv/swagger-ui.html>) for the final deliverable (D7.3) of this workpackage, the ELIXIR Plant Data Lookup Service.

A1.3 Conclusion

The main objective of workpackage 7 is the implementation of an infrastructure to allow plant genotype-phenotype analysis based on the widest available public datasets. This deliverable (D7.2), consisting on the exposure of plant phenotypic datasets through the several BrAPI endpoints implemented by the participants, represents an important step to achieve that objective. In the previously reported deliverable (D7.1), we have worked on the development of controlled vocabularies and submitted exemplar datasets to public repositories annotated in accordance with the adopted standards. Now, we have supported access to distributed plant phenotypic datasets using an API developed by the international community with a common interest in plant breeding. Six BrAPI endpoints are currently available from several ELIXIR nodes for accessing diverse phenotypic datasets from crops and forest plants. Additionally, the Brassica Portal, which already provides access to datasets from several species, has also initiated BrAPI implementation.

The reported work contributes for making plant data FAIR and it is needed for the full implementation of the ELIXIR Plant Data Lookup Service. This service is currently under development and will be the subject of the next deliverable (D7.3), due at the end of the project.

Appendix 2: List of Presentations on the Work

Described

A2.1. Presentations in International Conferences

Alaux M, Pommier C, Dzale Yeumo E, Durand S, Flores R, Kimmel E, Letellier T, Michotey C, Mohellibi N, Quesneville H, Adam-Blondon A-F (2018) Improving the « FAIRness » of Inra's Data for Plant Biology and Breeding. Presented at PAG XXVI - Plant and Animal Genome Conference.

Kersey, P.J. (2018) Invited speaker at Meeting of the Society for Experimental Biology, 3rd-6th July, Florence, Italy

Pommier, C. (2018). FAIR Data publication for phenomics data. Presented at Integrative Bioinformatics Symposium 2018, Rothamsted, GBR (2018-06-13 - 2018-06-15).

Pommier, C. (2018). Standards for distributed Plant Phenotyping data integration: challenges and solutions.. Presented at COST FA1306 "The quest for tolerant varieties - Phenotyping at plant and cellular level" - WG1 Meeting "Phenotyping: from the lab to the field", Leuven, BEL (2018-03-20 - 2018-03-21).

Pommier, C., Cornut, G., Letellier, T., Michotey, C., Neveu, P., Ruiz, M., LARMANDE, P., Kersey, P. J., CWIEK KUPCZYNSKA, H., Coppens, F., FINKERS, R., Laporte, M.-A., Faria, D., Miguel, C., Chaves, I., Costa, B., Adam-Blondon, A.-F. (2018). Data standards for plant phenotyping: MIAPPE and its implementations. Presented at PAG XXVI - Plant and Animal Genome Conference.

Pommier, C. (2017). Distributed plant phenotyping data integration: challenges and feedback from French and international consortiums. Presented at Data Scientist Big Data Plant Breeding, Wageningen , NLD (2017-12-11 - 2017-12-11).

Pommier, C. (2016) GnpIS-Ephesis, the Phenotypic Data Integration Platform for Inra Networks Experimental Data--Data Discovery and Dataset Building Use Cases, Plant and Animal Genome XXIV Conference, 9th-12th January San Diego

Durand, S., Dzale Yeumo, W. E., Pommier, C., Alaux, M., Cornut, G., Flores, R., Kimmel, E., Letellier, T., Michotey, C., Mohellibi, N., Quesneville, H., Adam-Blondon, A.-F. (2017). Findable, Accessible, Interoperable and Reusable data linked to genetic resources for plant biology and breeding at INRA. Presented at 4. International Symposium on Genomics of Plant Genetic Resources, Giessen, DEU (2017-09-03 - 2017-09-07).

Finkers, R. (2018). FAIR Plant Data Management Challenges. Presented at PAG XXVI - Plant and Animal Genome Conference.

Tyrrell S., Bian X., Davey R.P. (2018). Grassroots: An Infrastructure for sharing Services & Data. Presented at the 14th International Symposium on Integrative Bioinformatics. Rothamsted Research, Harpenden, Hertfordshire, UK (2018-06-13 - 2018-06-15)

Tyrrell S., Bian X., Davey R. P. (2018) Grassroots: An infrastructure for sharing services and data. PhenoHarmonIS Semantics for Harmonization and Integration of Phenotypic and Agronomic Data. Agropolis Scientific Park, Montpellier, FR (2018-05-14 - 2018-05-18)

Eckes A. H., Van Den Bergh E., Minotto A., Davey R.P. (2017) CyVerse UK for Brassica: Performing Associative Transcriptomics By Integrating Sequence and Phenotype Repositories. Plant and Animal Genomes XXV, San Diego, US (2017-01-14 - 2017-01-18)

Davey R. P. (2017) Data Stewardship in the Life Sciences. Plant and Animal Genomes XXV, San Diego, US (2017-01-14 - 2017-01-18)

A2.2 Posters in International Conferences

Bian X., Tyrrell S., Davey R.P. (2017) Grassroots Infrastructure: An interoperable data repository for plant science. Genome 10K & Genome Science, Norwich, UK (2017)

Bian X., Tyrrell S., Davey R. P. (2017) Grassroots Infrastructure: An interoperable data repository for plant science. Research Data Alliance 9th Plenary Meeting, Barcelona, ES (2017)

Papoutsoglou, E., Kaliyaperumal, R., van Hintum, T., Visser, R.G.F., Athanasiadis, I.N., Finkers, R. (2017) [Toward Better Data Sharing Methods for Genebanks](#). Proceedings of the 2nd International Workshop on Semantics for Biodiversity co-located with 16th International Semantic Web Conference ([ISWC 2017](#))

A2.3 Other Presentations

Adam-Blondon A-F (2018) Two European projects with interesting output for the grapevine community: Innovine and ELIXIR. Presented to the partners of the USDA VitisGen2 project (<http://www.vitisgen2.org/>)

Faria, D., Chaves, I., Costa, B.V., Miguel, C.M. (2018) Make Plant Data FAIR. Bioinformatics Open Days 2018, 14th-16th March, Braga, Portugal (oral)

Cardoso, J., Costa, B.V., Faria, D. (2018) Automated MIAPPE Compliance Validation. ELIXIR All-Hands Annual Meeting, 4th-7th June, Berlin, Germany (poster)

Kersey, P.J., Pommier, C. (2018) Discussion with UK project "Designing Future Wheat" at Rothamsted Research, 2nd February, UK

Kersey, P.J. (2018) Presentation of project at UK Emphasis meeting at Rothamsted Research, 19th-21st March, UK

Kersey, P.J. (2018) Presentation of project at 12th CeBiTec-Symposium, 19th-21st March, Bielefeld, Germany

Kersey, P.J. (2018) Invited seminar at East Malling Research, 3rd April, UK

Finkers, R. (2018) Latest developments in the FAIR data ecosystem. Invited seminar at ITQB NOVA, 17th April, Oeiras, Portugal

Papoutsoglou, Cornut, G., E. Pommier, C. Finkers, R. Showcasing JSON-LD for BrAPI