

## **DELIVERABLE 1.1**

### **Standards for collecting and displaying phenotypic data and images**



Call identifier:  
PRO-GRACE  
Grant agreement no: 101094738

## Promoting a plant genetic resource community for Europe

### Deliverable No. D1.1

Standards for collecting and displaying phenotypic data and images

Contractual delivery date:  
M11

Actual delivery date:  
M11

Responsible partner:  
UPV

Contributing partners:  
(CREA, CNR, INIAV, IPK, NASC, IPGR, NORDGEN)



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094738.

|                            |   |
|----------------------------|---|
| <b>Grant agreement no.</b> | Horizon Europe – 101094738  |
| <b>Project full title</b>  | PRO-GRACE – Promoting a plant genetic resource community for Europe   |
| <b>Deliverable number</b>  | <b>D1.1</b>   |
| <b>Deliverable title</b>   | <b>Standards for collecting and displaying phenotypic data and images</b>   |
| <b>Type</b>                | R – Document, report  |
| <b>Dissemination level</b> | PU - Public   |
| <b>Work package number</b> | WP1   |
| <b>Author(s)</b>           | María José Díez, Jaime Prohens, Domenico De Paola, Paolo Direnzo, Lorenzo Maggioni, Sandra Goritschnig, Jan Svensson, Kjell-Åke Lundblad, Pawel Chrominski, Maria Antonietta Palombi, Ana Maria Barata, Sean May, Clara Pons, Antonio Granell, Catherine Hazel Aguilar, Stephan Weise |
| <b>Keywords</b>            |   |

**The research leading to these results has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101094738.**

The authors are solely responsible for its content, it does not represent the opinion of the European Commission and the Commission is not responsible for any use that might be made of data appearing therein.

## Introduction

A plant **phenotype** is the set of structural, physiological, and performance-related traits of a genotype in each environment. The term **phenotyping** covers any procedure of measuring plant characteristics that can be expressed quantitatively or qualitatively, at the level ranging from single cells, through to whole plants. **Phenomics** is the discipline of gathering high-dimensional phenotypic data at multiple levels of organization towards the full characterization of the complete set of phenotypes of a genome (Houle et al., 2010). Given that a **phenome** consists of the set of all possible phenotypes of a given genotype, plant **phenomics** could be considered as the study of phenomes of multiple genotypes (Dhondt et al. 2013). Besides providing the tools to perform phenotyping itself, **plant phenotyping systems** usually comprise the means to grow plants in environments which are either defined and controlled, or semi-controlled, or uncontrolled and measured. The characteristics of the phenotyping system determine its capacity in terms of the number of genotypes and the range of environmental conditions or treatments, therefore its suitability for phenomics.

A growing number of experiments investigating similar plant systems on a large scale, calls for the development and deployment of technology that supports integration of data coming from different sources. Such integrations are feasible only if standards concerning definition and recording of the phenotypes are agreed upon and widely used.

There have been previous projects aiming at recommendations for standards about phenotypic observations. It is common practice that each existing repository of phenotypic data defines the set of metadata and data format in the instructions for data submission; this was done, for example, for the MaizeGDB, Triticeae Toolbox, Phenopsis DB, and Ephesis databases (Krajewski et al., 2015). More recently, two European infrastructure projects, trans-PLANT (Trans-national Infrastructure for Plant Genomic Science, <http://transplantdb.eu>) and EPPN (European Plant Phenotyping Network, <http://plant-phenotyping-network.eu>) have developed new recommendations, seeking to combine them with existing standards to address the content, meaning (semantics), and format of the old and novel phenotypic data types, to support the integrative analysis of multiple types of phenotypic data (Krajewski et al., 2015; Ćwiek-Kupczyńska et al., 2016). Other issues, such as the influence of phenotyping methodologies, the need for good ontologies to achieve semantic interoperability of data commons, and clear demand for standard(s)-aware tools and services, have also been addressed by these authors.

Images are also phenotypic observations of genetic resources registered with standardized descriptors. However, the lack of standardized protocols specifically suited for collecting and annotating images of genetic resources can impede their integration, comparison, and collaboration across different projects and databases. Here we present a foundation proposal to establish standards for collecting images for plant phenotyping databases, with a focus on European projects.

## Activities

To start addressing this deliverable, two activities were carried out: a meeting between IPK and UPV to discuss and reach an agreement about the content and orientation of the deliverable, and a bibliographic review to explore different strategies developed for standardising collections of data and display of phenotypic data and images. During the meeting, we expanded on text from D1.1 included in the project proposal alongside information presented at the kick-off meeting regarding the aim of the deliverable. It was evident that a change of orientation, from **collecting phenotypic data and images** to the **establishment of the criteria for generating a harmonised system of standards and descriptors for phenotypic data and images** was needed.

A review was then performed, collating strategies developed by previous projects for collecting phenotypic data (Poorter et al., 2012; Krajewski et al., 2015; Pommier et al., 2018, 2019; Pieruschka and Schurr, 2019; Kim, 2020). The methodology proposed by Krajewski et al. (2015) was generally followed.

## Results

For the standardisation, capture and description of phenotypic information we chose to adopt the methodology proposed by Krajewski et al. (2015), which considers three layers of information:

- a) Standardisation of metadata: Minimum Information about Plant Phenotyping Experiments standard (MIAPPE) (<https://www.miappe.org/>).
- b) Standardisation of vocabulary and recommendations regarding the use of ontologies.
- c) Format for phenotypic data exchange as flat files and web services (ISA-Tab for Phenotyping).

### a) Standardisation of metadata: Minimum Information about Plant Phenotyping Experiments standard (MIAPPE) (<https://www.miappe.org/>)

MIAPPE is an open, community-driven, data standard designed to harmonise data from plant phenotyping experiments. MIAPPE provides a specification including a checklist and a data model of metadata required to adequately describe plant phenotyping experiments ([https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE\\_Checklist-Data-Model-v1.1/MIAPPE\\_Checklist-Data-Model-v1.1.pdf](https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE_Checklist-Data-Model-v1.1/MIAPPE_Checklist-Data-Model-v1.1.pdf)).

MIAPPE aims in particular to harmonise the description of experimental and computed data in order to enable its sharing, publication and reuse. It is designed to handle datasets of crops and woody plants grown in greenhouses, single fields or experimental networks over one to several years. The main elements of a MIAPPE datasets consist of the **Dataset** itself (or Investigation following ISA), the **Study**, the **Biological Material** and the **Observation Variables** (Table 1)

#### Dataset (Investigation)

Contains the basic information shared by any type of datasets: title, authors list, description, a DOI for citation, etc....

#### Studies

A MIAPPE dataset contains one to many studies, each study being one experiment in one location over one to several years. These multiannual capabilities allow to handle perennial plants such as trees or perennial forages, etc... A study can also be virtual to hold the result of a computation over one to several experimental studies. It also contains the geographical coordinates of the study.

#### Biological Material

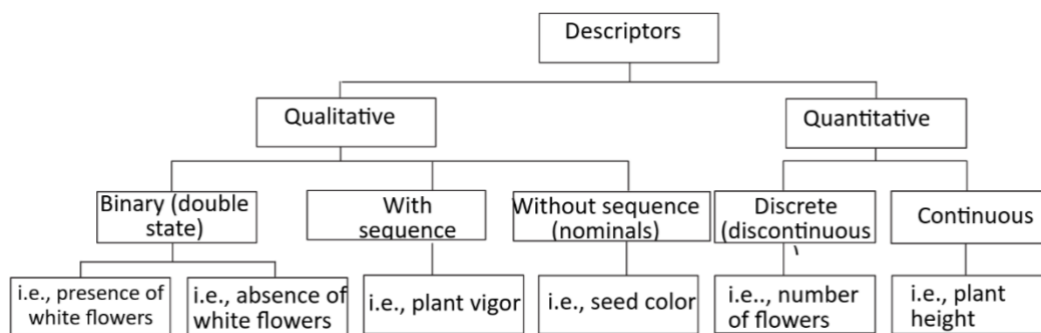
This is one of the two most essential descriptions to enable FAIR (Findable, Accessible, Interoperable, Reusable) data sharing. It refers to the biological material being studied (e.g., plants grown from a certain batch of seed, or plants grown in a particular field) including identification and traceability. The original source of that material (e.g., the seeds or the original plant cloned) is called the material source, which, when held by a material repository, should have its stock identified. In this way, the use of DOIs can provide a unique and persistent identifiers that have become a quasi-standard in the field of plant genetic resources in recent years (Alercia et al., 2018).

#### Observed Variable

An observed variable describes how a measurement has been made. It typically takes the form of a measured characteristic of the observation unit (plant or environmental trait), associated with the method and unit of measurement. Multiple variables with the same combination of trait, method and scale can be used in association with different plant parts (e.g., leaf 1, leaf 2), when this distinction is necessary for observations referring to different parts of the same observation unit.

Quantitative phenotypic traits emerge from complex interactions between heredity (genome and epigenome) and various environmental factors. This calls for observations of plants in multiple environmental contexts if broadly applicable conclusions are to be drawn, and for application of multifactorial experimental designs in both field and greenhouse trials. To disentangle the interactions of genotype and environment and to interpret the growth models, a detailed characterization of the environment in which the experiments are conducted is necessary (Tardieu and Tuberosa, 2010; Poorter et al., 2012). Such characterizations must be expressed in a standardized way to be of value for the community.

There are different categories of data, depending on the expression of the descriptor, which can be qualitative or quantitative (**Figure 1**). If expressed qualitatively, binary data (double-state), data with sequence (ordinal), and data without sequence (nominal) can be generated. If expressed quantitatively, the generated data can be discontinuous (discrete) and continuous (Hidalgo, 2003).



**Figure 1.** Types of data and descriptors.

The following suggestions help in practical data recording:

- For qualitative data of the binary type, each descriptor presents two states (present = 1, absent = 0). For example, presence of white flowers (1), or absence of white flowers (0).
- For qualitative data of the ordinal or sequence type, the descriptor is recorded using a series of predefined states; for example, for plant height: 1 = short (<0.5 m), 3 = intermediate (>0.5 <1.5 m), 5 = tall (>1.5 m).
- For nominal or non-sequential qualitative data, the descriptor is recorded using a series of previously defined states; for example, 1 = white, 2 = yellow, 3 = red.
- For quantitative data of the continuous type, the descriptor is recorded with the standard units of the International System of Units (SI = Système International), for example, height of the plant = 0.9 m; weight of 100 seeds = 250 g.

MIAPPE has developed a checklist in Excel format that includes all the guidelines for each category of information ([https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE\\_Checklist-Data-Model-v1.1/MIAPPE\\_Checklist-Data-Model-v1.1.tsv](https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE_Checklist-Data-Model-v1.1/MIAPPE_Checklist-Data-Model-v1.1.tsv)).

MIAPPE is being adopted as the main standard for describing plant phenotyping metadata. Several international initiatives such as the transPLANT project, the European Plant Phenotyping Network (EPPN), ELIXIR and EMPHASIS are promoting their adoption. MIAPPE has been implemented as an essential part of the ELIXIR RDM Kit plant domain ([https://rdmkit.elixir-europe.org/plant\\_sciences](https://rdmkit.elixir-europe.org/plant_sciences)) and ELIXIR Plant Phenomic Tool Assembly ([https://rdmkit.elixir-europe.org/plant\\_phenomics\\_assembly](https://rdmkit.elixir-europe.org/plant_phenomics_assembly)) a series guidelines to standardize, organize and integrate phenotypic data with other omic data. In addition, some tools and databases have adhered to MIAPPE standards to ensure proper documentation and reporting of plant phenotyping experiments, while others, such as GerminateDB (<https://germinateplatform.github.io/get-germinate/>) are in the process of adhering. Some of the MIAPPE-compliant metadata databases and tools are:

- **Phenotyping Hybrid Information System** (PHIS; <http://www.phis.inra.fr/>): is a platform developed to support data management and analysis for plant phenotyping experiments. PHIS is designed to assist researchers in collecting, storing, and analyzing data related to plant phenotyping (Neveu et al., 2018)
- **Genoplante Information System** (GnpIS; <https://urgi.versailles.inra.fr/Tools/GnpIS>): a multispecies integrative information system dedicated to plant and fungi pests. It bridges is an international information system that links phenomic, genetic, and genomic data for species of agronomical interest (Steinbach et al., 2013; Pommier et al., 2019).
- **PSB Interface for Plant Phenotype Analysis** (PIPPA; <https://pippa.psb.ugent.be/>): 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. It provides access to the raw phenotypic data and linked metadata (<https://pippa.psb.ugent.be/>)
- **Collaborative OPen Omics** (COPO; <https://copo-project.org/>): a data management platform specific to plant science. It allows to describe, store and retrieve data using community standards and public repositories
- **pISA-tree** (<https://github.com/NIB-SI/pISA-tree>): pISA-tree and its accompanying R packages (*pisar* and *seekr*) provide a user-friendly system that guides researchers towards organisation of their research projects using a standardised directory tree without the requirement of any advanced systems maintenance.
- **FAIRDOME-SEEK platform** (<https://seek4science.org/>): an open-source, web-based cataloguing and commons platform, for sharing heterogeneous scientific research datasets, models or simulations, processes and research outcomes. A publicly available instance of a FAIRDOME-SEEK commons is available - as the FAIRDOMEHub (<https://fairdomhub.org/>)
- **BioSamples** (<https://www.ebi.ac.uk/biosamples/>): the biological sample metadata database and resource provided by the European Bioinformatics Institute (EBI). BioSamples is designed to facilitate the storage, organization, and retrieval of metadata information associated with biological samples used in various research experiments and projects. BioSamples integrates with other biological and genomic data resources, such as the European Nucleotide Archive (ENA), European Variation Archive (EVA), Array Express and Bioimage Archive.
- **Bioimage Archive** (<https://www.ebi.ac.uk/bioimage-archive/>): a free, publicly available online resource which stores and distributes biological images. It accepts submissions of data from any imaging modality, as long as the data are either associated with a peer-reviewed publication, or are of value beyond a single experiment.

Further, to facilitate interoperability and implementation, MIAPPE maps with other metadata exchange schemas such as ISA-tools (in particular ISA-MIAPPE, Breeding API (BrAPI) ([https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE\\_Checklist-Data-Model-v1.1/MIAPPE\\_mapping/MIAPPE\\_Checklist-1.1-with-mapping.tsv](https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE_Checklist-Data-Model-v1.1/MIAPPE_mapping/MIAPPE_Checklist-1.1-with-mapping.tsv))).

### **Recommendations for the GRACE-RI**

The MIAPPE checklist comprises 83 descriptors (**Table 1**) divided into 11 categories. There are training spreadsheets available with definitions, examples, and formats for all the descriptors. All this information should be studied in detail to select the most suitable descriptors to be implemented in PGR databases, such as EURISCO (Kotni et al., 2023). Also, the use of DOIs as stable and unique identifiers of the biological material is a recommended important prerequisite for the operation of aggregated PGR databases, allowing the information in each database to be linked easily to information in other data sources, especially where it concerns genomics and other omics data, that usually are not incorporated in PGR databases.

**Table 1.** Number of descriptors of the MIAPPE checklist included in each category.

| Category            | Number of descriptors | Metadata stored   |
|---------------------|-----------------------|---|
| Investigation       | 8                     | Investigation title, authors list, description, a DOI for citation, etc....   |
| Study               | 19                    | Experiments description, locations, years, geo-coordinates, growth conditions   |
| Person              | 5                     | Name, affiliation and address of people involved in the investigation   |
| Data file           | 3                     | Link to files or digital objects holding observation data recorded during one or more assays of the study, typically in tabular form  |
| Biological material | 17                    | Biological material info: source, taxonomy, ID (compliant with MCPD), BiosampleID   |
| Environment         | 2                     | Environmental parameters that were kept constant throughout the study and did not change between observation units or assays.   |
| Experimental factor | 3                     | Biotic or abiotic condition whose impact is studied   |
| Event               | 4                     | Information of natural or planned event at a particular time of the experiment  |
| Observation unit    | 5                     | Information of objects (plots, blocks, plants, organs, tissues) that are subject to instances of observation and measurement  |
| Sample              | 6                     | Information about the portion of plant tissue harvested, non-harvested or extracted from an observation unit for the purpose of sub-plant observations and/or molecular studies |
| Observed variable   | 12                    | How a measurement has been made. An observation variable is composed by the combination of a trait+method+scale   |

#### **b) Standardisation of vocabulary: recommendations regarding the use of ontologies**

To integrate data properly, it is necessary to identify the objects we are working with, and the observations made on those objects. In this way, Krajewski et al. (2015) define interoperability pivots which allow the comparison of trials from different data sets, or the construction of integrative data sets for meta-analysis or genetic analysis. The interoperability of the observations (phenotypic and environmental variables) can be achieved through many ontologies, some of which are linked, and some under active development (see, for example, the Phenotype RCN project at <http://phenotypercn.org>, or the Planteome project at <http://planteome.org>).

The pivot object (i.e., the observation variable) consists of a trait, a method, and a scale. The phenotypic trait or environment variable can be decomposed into an entity and a quality (EQ model; Mungall et al., 2010; Arnaud et al., 2012; Deans et al., 2015). For example, the trait 'leaf area' is related



to the entity 'leaf' and the quality 'area'. Therefore, strong references and definitions for entity and quality are needed.

These are provided by Plant Ontology (PO) and Phenotype And Trait Ontology (PATO) (<http://bioportal.bioontology.org/ontologies/>). Trait dictionaries/lists can also be shared, as in PO, with a structured vocabulary and database resource. PO establishes links between plant growth, anatomy, morphology, development, and plant genomics data. The PO is under active development to encompass terms and annotations from all plants (<https://bioportal.bioontology.org/ontologies/PO>). **Figure 2** below demonstrates how to access the Plant Ontology BioPortal.

The figure consists of three screenshots of the Plant Ontology BioPortal interface, showing the main page, the 'Ontologies Used' section, and the 'About Us' section.

**Top Screenshot: Main Page**

The main page displays the Plant Ontology (PO) overview. It includes a table of releases and a line graph showing the number of views over time.

| Version            | Released   | Updated    | Downloads               |
|--------------------|------------|------------|-------------------------|
| release/2017-08-13 | 08/16/2017 | 08/16/2017 | 100 (CSV   REXML   DIF) |
| release/2019-09-05 | 09/05/2019 | 09/05/2019 | 100 (DIF)               |
| release/2020-09-04 | 09/04/2020 | 09/04/2020 | 100 (DIF)               |
| release/2019-09-05 | 09/05/2019 | 09/05/2019 | 100 (DIF)               |
| release/2018-09-25 | 09/25/2018 | 09/25/2018 | 100 (DIF)               |

**Middle Screenshot: Ontologies Used**

The 'Ontologies Used' section lists various ontologies used in the Plant Ontology project. A red arrow points to the 'Plant Ontology' entry.

- Basic Formal Ontology
- Cell Ontology
- Central Plant Development Ontology
- Central Plant Gross Anatomy Ontology
- Crop Ontology
- Discipline Development Ontology
- Foundational Model of Anatomy
- Gene Ontology
- Major Gross Anatomy Ontology
- Medical Subject Headings
- NCIT Controlled Vocabulary
- Phenotype Quality Ontology
- Plant Anatomy
- Plant Experimental Conditions Ontology
- Plant Ontology
- Plant Structure Development Stage
- Plant Trait Ontology
- Relations Ontology

**Bottom Screenshot: About Us**

The 'About Us' section provides information about the Plant Ontology project, including contact details and a list of products.

**Products:**

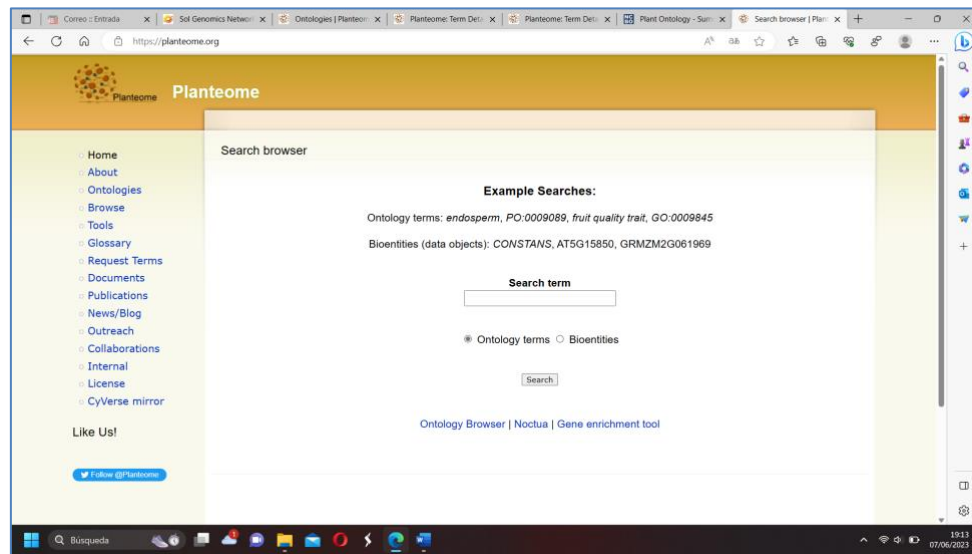
- BiPortal
- BiPortal R2D API
- BiPortal Virtual Appliance
- NCBO Web Widgets

**Support:**

- Contact Us
- Documentation
- NCBO Wiki

**About Us:**

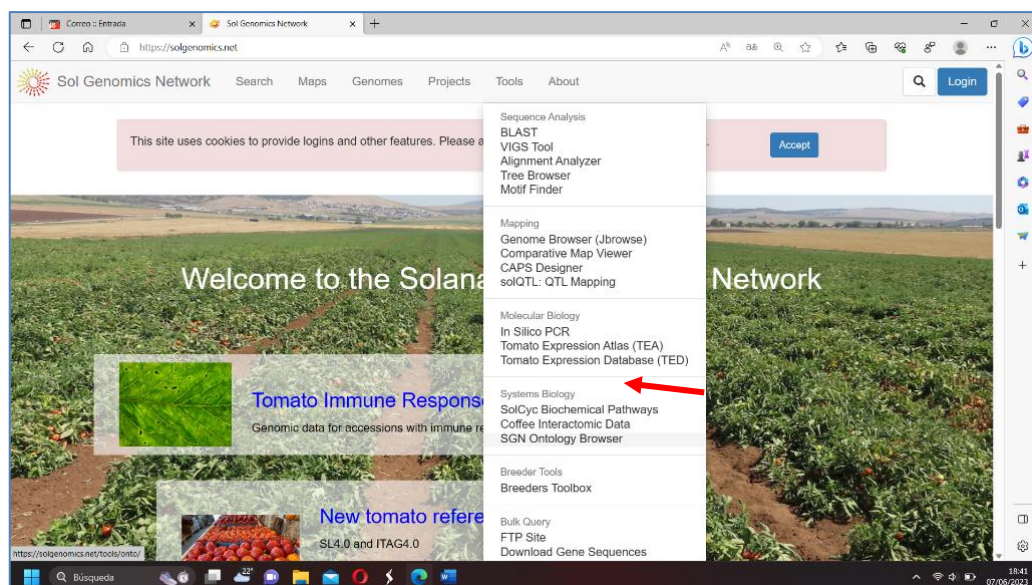
- About Us
- Mission & Vision
- Team
- Search



**Figure 2.** Steps for accessing the Plant Ontology BioPortal.

Trait dictionaries/lists can also be shared, as exemplified in Crop Ontology (<http://cropontology.org>). This serves as an excellent repository of species-specific trait ontologies, offering a comprehensive collection of measurement protocols and notation scales. Such ontologies assist multiple communities, because a simple trait, such as ‘spike density’, can have different measurement methods and units in different contexts and/or projects, and can be classified as a morphological trait in one project and as a grain quality trait in another.

As an example, **Figure 3** below shows the Solgenomics ontology browser (<https://solgenomics.net/tools/onto/>) developed by the Solanaceae Genomics Network.



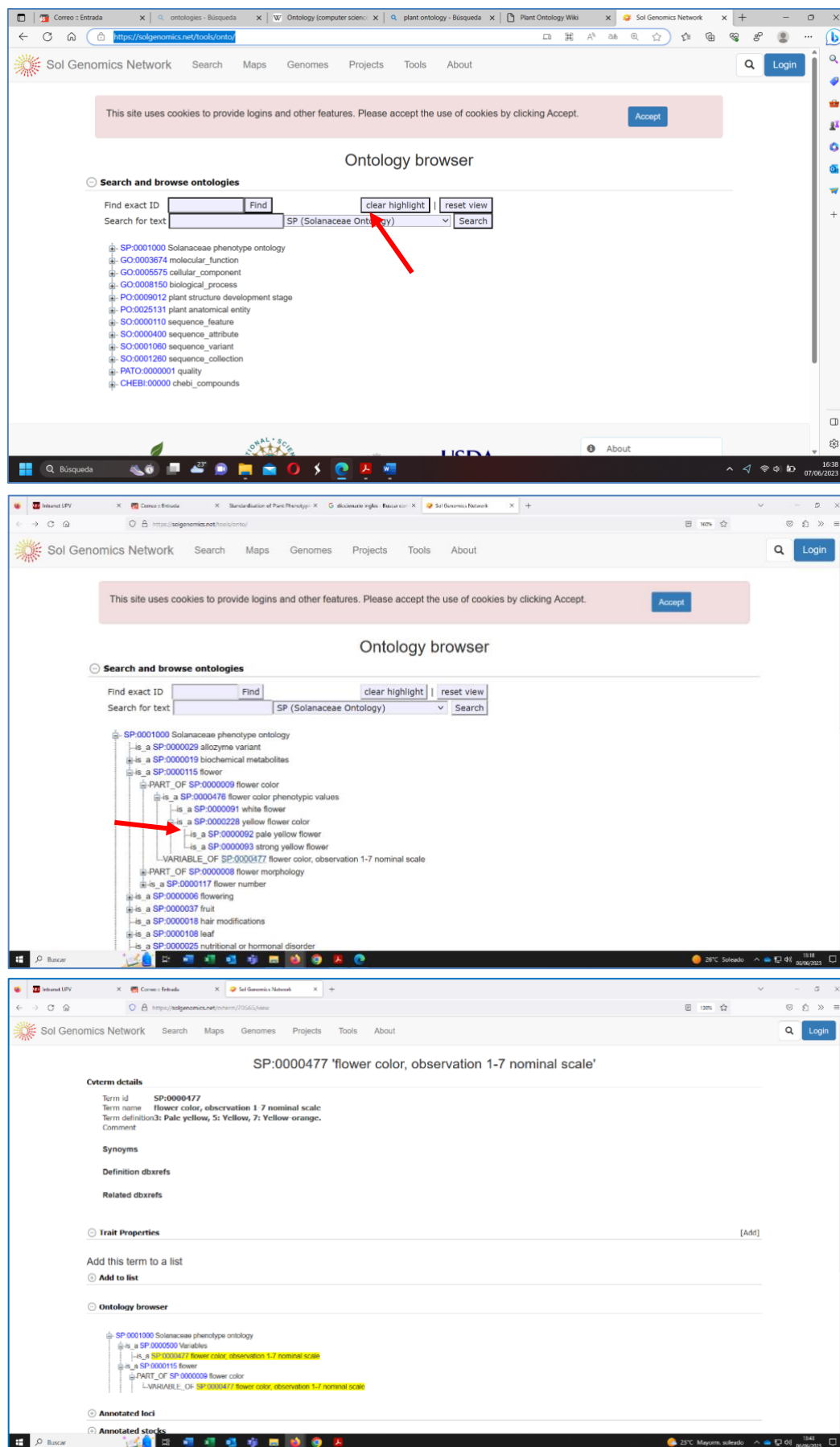


Figure 3. Steps for accessing the Solgenomics ontology browser.

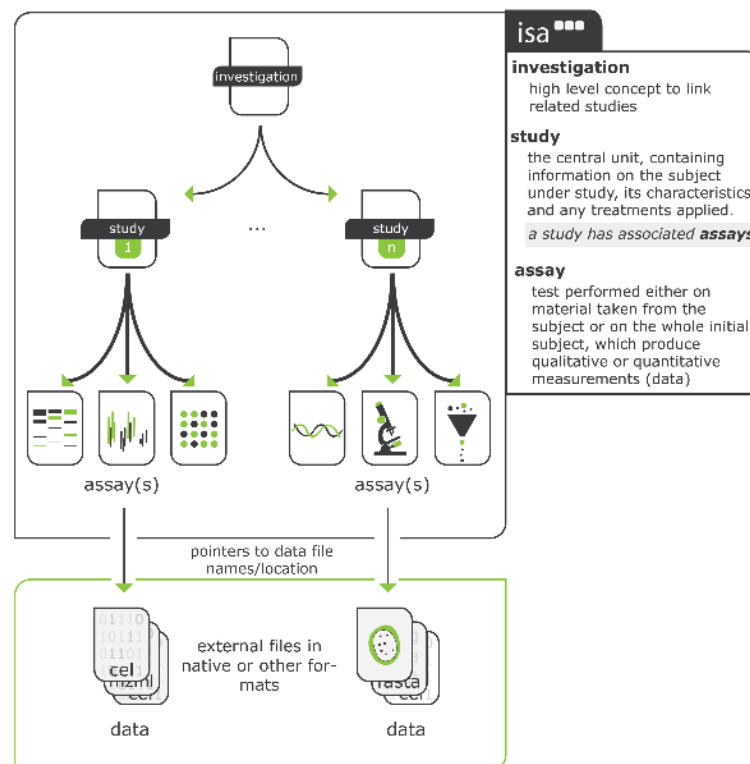
### Recommendations for the GRACE-RI

The level of precision in existing ontologies varies greatly across different crops and is generally much more developed in economically important species. Therefore, a first step should be the creation of working groups specializing in different broad crop groups. These groups should study, complete and propose ontology-based names for the most relevant traits that will be included in EURISCO.

#### c) Format for phenotypic data exchange by flat files and web services (ISA-Tab for Phenotyping)

To tackle complex scientific questions, experimental datasets from varying sources often need to be harmonised with regard to structure, formatting and annotation, to open their content to (integrative) analysis. Vast swathes of bioscience data remain locked in esoteric formats, described using nonstandard terminology, lack sufficient contextual information, or are simply never shared due to the perceived cost or futility of the exercise. Sansone et al. (2012) describe a way in which a group of data producers and consumers can work within an invisible metadata framework to enable a coordinated use of reporting standards by service providers. This circumvents many of the problems caused by data diversity. The same framework enables researchers, bioinformaticians and data managers to operate within an open data commons.

Sansone et al. (2012) refer to efforts toward such positive solutions as ‘data commoning’. The emerging commons (**Figure 4**) depends on participants using the metadata categories: ‘Investigation’ (the project context), ‘Study’ (a unit of research) and ‘Assay’ (analytical measurement). This ISA framework is the backbone upon which discovery, exchange and informed integration of data sets articulate with one another. The ISA commons is a growing exemplar ecosystem of data curation and sharing solutions built on a common metadata tracking framework, providing tools and resources to create and manage large, heterogeneous data sets in a coherent manner, allowing users of data sets and their constituent parts to ‘connect the metadata dots’.



**Figure 4.** Building the ‘ISA commons’, a growing ecosystem of resources that work to provide a data commons (Sansone et al., 2012).

**Recommendations for the GRACE-RI**

The pros and cons of the ISA framework system for phenotypic data acquisition, data exchange, integration of multi-source data, and their integration in the current EURISCO framework should be discussed by EURISCO together with a group of experts as needed.

**Standards for collecting images**

Collection of images of plants, economically important organs (fruit, spike, kernel, leaves), tissues, etc. provides relevant information complementary to the one furnished by phenotypic descriptors. An appropriate standardization, including metadata and references of size and possibly of colour, is required for the optimal usage of the information contained in the image, and their exchange and interoperability with other databases, as well as for further processing using software (e.g., PlantCV, PhenoPhyte, SmartGrain, Tomato Analyzer, WinRhizo, etc.) that can extract relevant phenotypic information from images.

Collecting relevant metadata along with each image is crucial for data organization and integration. No standards for image metadata for phenotyping of genetic resources have been published. However, for related fields such as for the reuse of microscopy data in biology, recommended metadata have been published (Sarkans et al., 2021) that can serve as a basis for further development of the minimum standards for image metadata for images of genetic resources. The proposed metadata standards have been prepared taking into account the MIAPPE (Minimum Information About a Plant Phenotyping Experiment) standardized metadata formats to ensure interoperability and facilitate data sharing.

**Leveraging Established Standards and Ontologies for PGR Image Data Standardization**

While there are currently no dedicated standards published specifically for PGR image data, a variety of existing standards and frameworks can be effectively leveraged. These established systems offer a foundation for ensuring consistency, accuracy, and accessibility in the management and analysis of PGR images, thus facilitating enhanced research and conservation strategies. The following discussion outlines key existing standards that can be adapted and applied to the standardization of PGR image data.

**TDWG-Audubon Core (AC).** Designed explicitly for biodiversity content, the Audubon Core is a set of vocabularies developed to represent multimedia resources (Morris et al., 2013a). Its strength lies in the comprehensive detailing of metadata elements specifically tailored for media resources, making it one of the most relevant existing standards for our context. AC provides clarity on licensing models, multimedia format descriptions, content descriptions, and more. Such granularity in media documentation can be beneficial to PGR image documentation (Morris et al., 2013b).

**ANSI/NISO Z39.87-2006 (R2017).** Known as the Data Dictionary- Technical Metadata for Digital Still Images, this standard establishes an essential framework for the documentation of digital imaging specifications, including image assessment metadata, data structure, and rights management. It is specifically designed to define metadata elements for raster digital images, aiding users in creating, sharing, and understanding digital image files. It promotes system interoperability, ensures long-term maintenance of digital image collections, and facilitate sustained access to these resources (NISO, 2017). Its application is vital for the effective technical standardization of PGR images.

**Dublin Core Metadata Element Set (DCMES).** This standard presents a structured yet flexible framework for describing and cataloging PGR images. Comprised of 15 core metadata elements, it is



designed for widespread use in the digital archiving field (Weibel, 2005). One of the major strengths of the Dublin Core is its balance between simplicity and comprehensiveness. It is straightforward enough to be easily adopted without requiring extensive technical knowledge, yet it can capture a wide range of information. This makes it a good fit for the diverse needs of those working with PGR, from botanists to data scientists.

**Darwin Core (DwC).** DwC, an extension of the Dublin Core specifically designed for biodiversity data, offers a refined and detailed framework for cataloging biological specimens and observational data (Wieczorek et al., 2012). DwC's specialized terms add depth to the documentation process, allowing for the incorporation of specific details that are crucial in PGR management and conservation. When applied to PGR images, DwC facilitates the embedding of rich contextual information within each image's metadata such as geographical, ecological, and phenotypic information.

**Access to Biological Collection Data (ABCD).** This schema provides an even more comprehensive data structure, covering extensive details about biological collections, including taxonomy, geography, genetics, and ecological information (Karam et al., 2016). By adopting this standard for PGR image data, there is an opportunity to link each image with a rich tapestry of information about the accession's environmental setting, its genetic makeup, and the interactions it has within its ecosystem. This approach enables a more holistic understanding of each plant specimen, going beyond the visual representation to include a wealth of contextual information.

**ISA Framework.** For PGR image data, contextual metadata is vital. The context in which a plant image was captured, e.g. the soil type, the climate, the growth stage, can significantly influence its interpretation. The ISA framework ensures that all relevant metadata accompanying an image is systematically captured and catalogued (Johnson et al., 2021). This framework ensures that anyone encountering the PGR image in a database or repository understands its full context and the conditions under which it was captured.

**Exchangeable Image File Format (EXIF).** It is a technical standard for storing metadata within digital photographs. This metadata includes camera specifications, capture settings, and geolocation data (Chauhan and Panda, 2015). In the context of PGR image data standards, EXIF's technical capability to embed detailed information such as the geographic coordinates of where a photo was taken, camera settings, and the exact time of capture, makes it highly relevant. It allows for precise documentation and analysis of the environmental conditions and location-specific factors affecting PGR, enhancing the accuracy and utility of botanical databases and research.

The elements of the general information, image acquisition, data on plant part, image preprocessing and additional notes that we consider should be associated to an image of genetic resources for being included in databases are presented in **Table 2** below.

**Table 2.** Proposed elements, description, sites where they can be found, purpose of the elements and their relevance for the general information, image acquisition, data on plant part, image preprocessing and additional notes considered for image data associated to genetic resources.

**General Information**

| Element                                 | Description   | Found In   | Purpose                  | Relevance<br>(Essential,<br>Recommended,<br>Optional) |
|---|---|--|--------------------------|---|
| <b>Unique Identifier<br/>(i.e. DOI)</b> | A distinct identifier for each specimen or observation. | Dublin Core, EXIF, MCPD, MIAPPE, ISA, ABCD, Audubon Core | Distinct identification. | Essential   |

|   |  |                                       |   |             |
|---|--|---------------------------------------|---|-------------|
| <b>Institute Code</b>                   | Code of the institute where the image data is stored.        | MCPD, ABCD                            | Ownership and Custodianship (aligns with international standards, i.e. FAO WIEWS) | Essential   |
| <b>Genus</b>                            | Genus name for taxon   | Darwin Core, MCPD, ABCD, Audubon Core | Broader taxonomic level classification  | Essential   |
| <b>Species</b>                          | Specific epithet portion of the scientific name              | Darwin Core, MCPD, ABCD, Audubon Core | Clear identification/<br>Taxonomic clarity  | Essential   |
| <b>Species authority</b>                | Authority for the species name                               | MCPD, ABCD                            | Validation of scientific species naming   | Recommended |
| <b>Subtaxa</b>                          | Subdivision of the species (i.e.. Subspecies, variety, form) | MCPD, ABCD                            | Detailed taxonomic information provision  | Recommended |
| <b>Subtaxa Authority</b>                | Authority naming the subtaxa                                 | MCPD, ABCD                            | Validation of subtaxa classification  | Optional    |
| <b>Accession number</b>                 | Unique identifier for a specific genetic resource.           | MCPD, ABCD                            | Distinct identification of a genetic resource.                                    | Essential   |
| <b>Experiment ID</b>                    | Unique identifier for a specific experiment.                 | MIAPPE, ISA                           | Distinct identification of experiment.  | Essential   |
| <b>Collector/Observer</b>               | Person/entity responsible for collection/observation.        | ABCD, Audubon Core                    | Origin data tracking, Source attribution  | Recommended |
| <b>Date of Capture/<br/>Observation</b> | Date and time of capture or observation.                     | EXIF, Dublin Core, ABCD, Audubon Core | Temporal context.   | Essential   |
| <b>Country</b>                          | Country where the image was captured                         | EXIF, MCPD, ABCD, Audubon Core        | Identification of the country where the image was captured                        | Essential   |
| <b>Locality</b>                         | Specific location where the image was captured               | EXIF, MCPD, ABCD, Audubon Core        | Detailed description of the image capture site                                    | Essential   |
| <b>Latitude</b>                         | Latitude coordinates of the image capture site               | EXIF, MCPD, ABCD, Audubon Core        | Geographical coordinates for precise location of image capture                    | Essential   |
| <b>Longitude</b>                        | Longitude coordinates of the image capture site              | EXIF, MCPD, ABCD, Audubon Core        | Geographical coordinates for precise location of image capture                    | Essential   |

|                         |   |                                |   |             |
|-------------------------|---|--------------------------------|---|-------------|
| <b>Altitude</b>         | Altitude of the image capture site              | EXIF, MCPD, ABCD, Audubon Core | Elevation information of the image capture site | Recommended |
| <b>Image type</b>       | Type of image (e.g. photograph, scan)           | EXIF                           | Nature and format categorization of the image   | Recommended |
| <b>Image annotation</b> | Notes of labels added to the image post-capture | ISA, EXIF                      | Enhanced information delivery                   | Optional    |

**Image acquisition**

| <b>Element</b>                                 | <b>Description</b>  | <b>Found In</b>        | <b>Purpose</b>   | <b>Relevance to PGR Image Data</b> |
|--|---|------------------------|--|------------------------------------|
| <b>Image Source Device</b>                     | Device or equipment used to capture image   | EXIF, Audubon Core     | Image quality and device-specific characteristics.                         | Essential                          |
| <b>Image format</b>                            | File format of the image (e.g. JPEG, PNG, TIFF)   | EXIF                   | Digital preservation and compatibility (affect usability across platforms) | Essential                          |
| <b>Resolution</b>                              | Image dimensions or resolution.   | EXIF, Audubon Core     | Image quality and detail level indication                                  | Optional                           |
| <b>Bit Depth</b>                               | Color depth of the image.   | EXIF, ANSI/NISO Z39.87 | Image quality and color details.   | Optional                           |
| <b>Lighting Conditions</b>                     | Lighting details at the time of capture (e.g. natural or artificial light, time of day) | ANSI/NISO Z39.87       | Quality and consistency of image lighting                                  | Recommended                        |
| <b>Color Calibration/ Color reference card</b> | Used to maintain color consistency across devices                                       | ANSI/NISO Z39.87       | Image quality, color accuracy and consistency                              | Recommended                        |
| <b>Scale indicator</b>                         | Presence of a scale to denote the size in the image                                     | ANSI/NISO Z39.87       | Size reference and proportionality   | Recommended                        |

**Data on Plant Part**

| <b>Element</b>                   | <b>Description</b>  | <b>Found In</b>  | <b>Purpose</b>                        | <b>Relevance to PGR Image Data</b> |
|----------------------------------|---|--|---------------------------------------|------------------------------------|
| <b>Specimen Type</b>             | Type of specimen  | ABCD, Audubon Core, Plant Ontology (PO)                    | Nature of the plant material          | Essential                          |
| <b>Description of Plant Part</b> | Descriptive notes or annotations specific to plant parts. | Dublin Core, ABCD, Audubon Core, Plant Trait Ontology (TO) | Additional context on the plant part. | Essential                          |



|                                |   |                                       |   |             |
|--------------------------------|---|---------------------------------------|---|-------------|
| <b>Plant Part/Organ</b>        | Specific part or organ of the plant being imaged.   | Plant Ontology (PO)                   | Detailed description of the plant component.                  | Essential   |
| <b>Phenotypic Quality</b>      | Observable and measurable traits or attributes of the plant part.                                 | Phenotypic Quality Ontology (PATO)    | Specific observable features of the plant part.               | Essential   |
| <b>Developmental Stage</b>     | The growth or developmental phase of the plant or organ when imaged.                              | Plant Ontology (PO)                   | Context on the life cycle stage.                              | Essential   |
| <b>Image Magnification</b>     | Level of magnification used for the image   | Custom Field, Bio-Formats             | Detail of the image at the microscopic level                  | Recommended |
| <b>Associated Genetic Data</b> | Links or references to genetic data or studies associated with the specimen.                      | GenBank, EMBL, Sequence Ontology (SO) | Understanding genetic background or traits.                   | Optional    |
| <b>Ecological Interactions</b> | Information on known interactions with other organisms (e.g., pollinators, symbiotes, predators). | Interaction Network Ontology (INO)    | Context on ecological relationships affecting the plant part. | Recommended |
| <b>Image Orientation</b>       | Direction or angle from which the image was taken (e.g., top view, cross-section).                | Custom Field                          | Providing perspective context to the viewer.                  | Recommended |

### Image Preprocessing

| Element                         | Description   | Found In         | Purpose  | Relevance |
|---------------------------------|---|------------------|--|-----------|
| <b>Associated Media</b>         | Related media files or links, including processed versions. | Audubon Core     | Holistic interpretation and comparison; Contextual and supplementary information | Optional  |
| <b>Image Processing Details</b> | Details on any post-processing done on the image.           | ANSI/NISO Z39.87 | Data authenticity and modifications tracking.                                    | Optional  |

### Additional Notes

| Element | Description | Found In | Purpose | Relevance |
|---------|-------------|----------|---------|-----------|
|---------|-------------|----------|---------|-----------|

|                               |   |                           |  |           |
|-------------------------------|---|---------------------------|--|-----------|
| <b>Rights &amp; Licensing</b> | Permissions and licensing associated with data/image (e.g. Copyright details, licensing agreements) | Dublin Core, Audubon Core | Clarity on usage permissions.                            | Essential |
| <b>Usage Restrictions</b>     | Specific restrictions for data/image use. (e.g. commercial use, reproduction limits)                | Audubon Core              | Data access control and guidelines for image utilization | Essential |

### ***Recommendations for the GRACE-RI***

A detailed analysis should be done to check the suitability of implementation into EURISCO or GRACE-RI of the standards for conventional images. The following multi-faceted strategy is proposed for the development of PGR image data standards:

- 1. Adoption of Existing Standards.** An evaluation of established standards is advised to determine their applicability. Standards such as Audubon Core, which offers a structure well-suited to biodiversity content, and the EXIF framework, focused on specific metadata related to images, are potential candidates. Adoption of these standards should be considered if they meet the objectives of EURISCO and GRACE-RI, thereby streamlining the integration process.
- 2. Customization where necessary.** When existing standards closely align but do not perfectly match the specific requirements of PGR image data, customization should be considered. For example, the Dublin Core, with its versatile structure, may need particular adjustments to cater the nuances of PGR data. Similarly, the ANSI/NISO Z39.87 standard, robust for digital still images, might require modifications to emphasize botanical aspects relevant to the context.
- 3. Incorporation of established protocols.** Integrating frameworks from recognized standards can strengthen the data management foundation. Utilizing structures provided by standards such as MCPD (Multi-Crop Passport Descriptors) and MIAPPE (Minimum Information About a Plant Phenotyping Experiment) could enhance the robustness and relevance of the image data standards.
- 4. Integration of Ontologies for Consistency and Interoperability.** Ensuring consistency and interoperability is the key. This involves integrating ontologies facilitate seamless data exchange and interpretation.

Furthermore, an in-depth examination of emerging technologies, including image and video data captured by drones, as well as automated and robotic phenotyping systems, should be conducted. This examination should focus on developing strategies for integrating these technologies into EURISCO's existing structure, ensuring the system remains at the forefront of PGR data management.

### **Discussion**

In this deliverable we have summarized the main standards for metadata, ontologies and data exchange adopted in phenotypic and image data collection and dissemination. In addition, we have proposed a series of recommendations to integrate phenotyping and image data in EURISCO. such as:

- Study in detail the MIAPPE checklist to **select the most suitable descriptors to be implemented** in EURISCO

- Creation of working groups specialized in different broad crop groups to study, **complete and propose ontology-based names for the most relevant traits** that will be included in EURISCO, specially for the species less economically important
- **Discuss with a group of experts** in informatics the best way (ISA-tab, Brapi, JSON) to exchange (meta)data between EURISCO and other systems

The adoption of FAIR (Findable, Accessible, Interoperable, and Reusable) standards for management of phenotypic and image (meta)data within the European setting, similar to other regions, remains a work in progress. While MIAPPE standards are being implemented and enjoying robust promotion by European networks and projects (i.e. ELIXIR, ENPHASIS, EPPN and others) as the preferred choice for germplasm phenotyping data, the same cannot be said for image standards. Currently, there is no consensus regarding the most suitable metadata standard for describing germplasm images. In this deliverable, we have undertaken a review of several image metadata standards that can be integrated into EURISCO. To accomplish this task, we recommend reaching a consensus on the adoption of select existing image standards and adapting them to meet the specific requirements of PGR (Plant Genetic Resources) by incorporating MCPD and MIAPPE metadata, as well as plant-specific ontologies.

## References

- Alercia A, López FM, Sackville Hamilton NR and Marsella M. 2018. Digital Object Identifiers for Food Crops - Descriptors and Guidelines of the Global Information System. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.
- Arnaud E, Cooper L, Shrestha R, et al. 2012. Towards a reference plant trait ontology for modelling knowledge of plant traits and phenotypes. Proceedings of the International Conference on Knowledge Engineering and Ontology Development, 220–225. 4–7 October 2012, Barcelona, Spain.
- Chauhan S and Panda NK. 2015. Chapter 7 – Metadata. Editor(s): Sudhanshu Chauhan, Nutan Kumar Panda, Hacking Web Intelligence, Syngress: Pages 133-146: ISBN 9780128018675, <https://doi.org/10.1016/B978-0-12-801867-5.00007-0>.
- Ćwiek-Kupczyńska H, Altmann T, Arend D, Arnaud A, Chen D, Cornut G, Fiorani F, Frohberg W, Junker A, Klukas C, Lange M, Mazurek C, Nafissi A, Neveu P, van Oeveren J, Pommier C, Poorter H, Rocca-Serra P, Sansone SA, Scholz U, van Schriek M, Seren Ü, Usadel B, Weise S, Kersey P; Krajewski P. 2016. Measures for interoperability of phenotypic data: minimum information requirements and formatting. *Plant Methods* (2016) 12:44. DOI 10.1186/s13007-016-0144-4.
- Deans AR, Lewis SE, Huala E, et al. 2015. Finding our way through phenotypes. *PLoS Biology* 13, e1002033. doi.org/10.1371/journal.pbio.1002033.
- Dhondt S, Wuyts N, Inzeet D. 2013. Cell to whole-plant phenotyping: the best is yet to come. *Trends in Plant Science*, Vol. 18, No. 8. doi.org/10.1016/j.tplants.2013.04.008.
- Johnson D, Batista D, Cochrane K, Davey RP, Etuk A, Gonzalez-Beltran A, Haug K, Izzo M, Larralde M, Lawson TN, Minotto A, Moreno P, Nainala VC, O'Donovan C, Pireddu L, Roger P, Shaw F, Steinbeck C, Weber RJM, Sansone SA, Rocca-Serra P. 2021. ISA API: An open platform for interoperable life science experimental metadata. *Gigascience*. Sep 16;10(9):giab060. doi: 10.1093/gigascience/giab060. PMID: 34528664; PMCID: PMC8444265.
- Hidalgo R. 2003. Variabilidad Genética y Caracterización de Especies Vegetales. In: Franco, TL and Hidalgo R. (eds.). 2003. Análisis Estadístico de Datos de Caracterización Morfológica de

- Recursos Fitogenéticos. Boletín técnico no. 8, Instituto Internacional de Recursos Fitogenéticos (IPGRI), Cali, Colombia. 89 p.
- Houle D, Govindaraju DR, Omholt S. 2010. Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855–866. doi:10.1038/nrg2897
- Karam N, Müller-Birn C, Gleisberg M, Fichtmüller D, Tolksdorf R, Güntsch A (2016) A Terminology Service Supporting Semantic Annotation, Integration, Discovery and Analysis of interdisciplinary Research Data. *Datenbank-Spektrum* 16 (3): 195-205.
- Kim JY. 2020. Roadmap to High Throughput Phenotyping for Plant Breeding. *Journal of Biosystems Engineering* 45:43–55. doi.org/10.1007/s42853-020-00043-0.
- Kotni P, van Hintum T, Maggioni L, Oppermann M, Weise S. 2023. EURISCO update 2023: the European Search Catalogue for Plant Genetic Resources, a pillar for documentation of genebank material. *Nucleic Acids Research*, 51(D1):D1465–D1469. doi: 10.1093/nar/gkac852
- Krajewski P, Chen D, Ćwiek H, van Dijk ADJ, Fiorani F, Kersey P, Klukas C, Lange M, Markiewicz A, Nap JP, van Oeveren J, Pommier C, Scholz U, van Schriek M, Usadel B, Weise S. 2015. Towards recommendations for metadata and data handling in plant phenotyping. *Journal of Experimental Botany*, Vol. 66, No. 18 pp. 5417–5427. doi:10.1093/jxb/erv271.
- Morris, R. A., Barve, V., Carausu, M., Chavan, V., Cuadra, J., Freeland, C., Hagedorn, G., Leary, P., Mozzherin, D., Olson, A., Riccardi, G., Teage, I. & Whitbread, G. 2013a. Discovery and publishing of primary biodiversity data associated with multimedia resources: The Audubon Core strategies and approaches. *Biodiversity Informatics*, 8, 2013, pp. 185-197
- Morris RA, Barve V, Carausu M, Chavan V, Cuadra J, Freeland C, Hagedorn G, Leary P, Mozzherin D, Olson A, Riccardi G, Teage I, Whitbread G. 2013b. Discovery and publishing of primary biodiversity data associated with multimedia resources: The Audubon Core strategies and approaches. *Biodiversity Informatics*. <https://doi.org/10.17161/bi.v8i2.4117>
- Mungall CJ, Gkoutos GV, Smith CL, Haendel MA, Lewis SE, Ashburner M. 2010. Integrating phenotype ontologies across multiple species. *Genome Biology* 11, R2. DOI: 10.1186/gb-2010-11-1-r2.
- Neveu P, Tireau A, Hilgert N, Negre V, Mineau-Cesari J, Brichet N, Chapuis R, Sanchez I, Pommier C, Charnomordic B, Tardieu F, Cabrera-Bosquet LI. 2018. <https://doi.org/10.1111/nph.15385>.
- Papoutsoglou EA et al. 2020. Enabling reusability of plant phenomic datasets with MIAPPE 1.1. *New Phytologist* (2020) 227: 260–273 doi: 10.1111/nph.16544.
- Pieruschka R, Schurr U. 2019. Plant Phenotyping: Past, Present, and Future. *Plant Phenomics* Article ID 7507131, 6 pages. doi.org/10.34133/2019/7507131.
- Pommier C, Cornut G, Letellier T, Michotey C, Neveu P, et al 2018. Data standards for plant phenotyping: MIAPPE and its implementations. 26. Plant and Animal Genome Conference (PAGXXVI), Jan 2018, San Diego, Californie, United States. pp. 24slides. hal-02789754.
- Pommier C, Michotey C, Cornut G, Roumet P, Duchêne E, Flores R, Lebreton A, Alaux M, Durand S, Kimmel E, Letellier T, Merceron G, Laine M, Guerche C, Loaec M, Steinbach D, Laporte MA, Arnaud E, Quesneville H, Adam-Blondon AF. 2019. Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS. *Plant Phenomics*, Article ID 1671403. doi.org/10.34133/2019/1671403.

- Poorter H, Fiorani F, Stitt M, Schurr U, Finck A, Gibon Y, Usadel B, Munns R, Atkin OK, Tardieu F, Pons TL. 2012. The art of growing plants for experimental purposes: a practical guide for the plant biologist. *Functional Plant Biology*, 39, 821–838. [dx.doi.org/10.1071/FP12028](https://doi.org/10.1071/FP12028).
- Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, Fang H, Neumann S, Tong W, Amaral-Zettler L, Begley K, Booth T, Bougueleret L, Burns G, Chapman B, Clark T, Coleman LA, Copeland J, Das S, de Daruvar A, de Matos P, Dix I, Edmunds S, Evelo CT, Forster MJ, Gaudet P, Gilbert J, Goble C, Griffin JL, Jacob D, Kleinjans J, Harland L, Haug K, Hermjakob H, Sui SJH, Laederach A, Liang S, Marshall S, McGrath A, Merrill E, Reilly D, Roux M, Shamu CE, Shang CA, Steinbeck C, Trefethen A, Williams-Jones B, Wolstencroft K, Xenarios I, Hide W. 2012. Toward interoperable bioscience data. *Nature Genetics* 44, 121-126. [doi.org/10.1038/ng.1054](https://doi.org/10.1038/ng.1054).
- Sarkans U, Chiu W, Collinson L, Darrow MC, Ellenberg J, Grunwald D, Hériché, J-K, Iudin A, Martins GG, Meehan T, Narayan K, Patwardhan A, Russell MRG, Saibil HR, Strambio-De-Castillia C, Swedlow JR, Tischler C, Uhlmann V, Verkade P, Barlow M, Bayraktar O, Birney E, Catavittello C, Cawthorne C, Wagner-Conrad S, Duke E, Paul-Gilloteaux P, Gustin E, Harkiolaki M, Kankaanpää P, Lemberger T, McEntyre J, Moore J, Nicholls AW, Onami S, Parkinson H, Parsons M, Romanchikova M, Sofroniew N, Swoger J, Utz N, Voortman LM, Wong F, Zhang P, Kleywegt GJ, Bzoma A. 2021. REMBI: Recommended metadata for biological images—enabling reuse of microscopy data in biology. *Nature Methods* 18, 1418-1446. [doi.org/10.1038/s41592-021-01166-8](https://doi.org/10.1038/s41592-021-01166-8).
- Steinbach D, Alaux M, Amselem J, Choisne N, Durand S, Flores R, Keliét AO, Kimmel E, Lapalu N, Luyten I, Michotey C, Mohellibi N, Pommier C, Reboux S, Valdenaire D, Verdelet D, Quesneville H. 2013. GnpIS: an information system to integrate genetic and genomic data from plants and fungi. *Database*, Article ID bat058, [doi:10.1093/database/bat058](https://doi.org/10.1093/database/bat058).
- Stuart L. Weibel. 2005. The Dublin Core Metadata Element Set ISSN: 1041-5653, Editor(s): Harold Moellering, Henri J.G.L. Aalders, Aaron Crane, In: *International Cartographic Association, World Spatial Metadata Standards*, Elsevier Science, 2005, Pages 493-513, ISBN 9780080439495, <https://doi.org/10.1016/B978-008043949-5/50029-2>.
- Tardieu F, Tuberosa R. 2010. Dissection and modelling of abiotic stress tolerance in plants. *Current Opinion in Plant Biology* 13, 206–212. [Doi.org/10.1016/j.pbi.2009.12.012](https://doi.org/10.1016/j.pbi.2009.12.012).
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>.

## Deviations

No deviations