PRO-GRACE

PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

DELIVERABLE 4.3

Unified, crop-specific standards, protocols and descriptors for the evaluation of the phenotypes and agronomic characteristics of PGR, incorporating the ECPGR, MIAPPE, Crop Ontology, EMPHASIS and final user recommendations and methodologies

This deliverable has been submitted and is currently pending approval by the European Commission.

Call identifier: HORIZON-INFRA-2022-DEV-01-01
PRO-GRACE
Grant agreement no: 101094738

Promoting a plant genetic resource community for Europe

Deliverable No. 4.3

Unified, crop-specific standards, protocols and descriptors for the evaluation of the phenotypes and agronomic characteristics of PGR, incorporating the ECPGR, MIAPPE, Crop Ontology, EMPHASIS and final user recommendations and methodologies

Contractual delivery date:
24

Actual delivery date:
25

Responsible partner:
(CREA)
Contributing partners:
(CREA, INRAE, IPGRI, ENEA, EMPHASIS-RI ….)

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

Grant agreement no.	Horizon Europe – 101094738
Project full title	PRO-GRACE – Promoting a plant genetic resource community for Europe

Deliverable number	**D4.3**
Deliverable title	**Unified, crop-specific standards, protocols and descriptors for the evaluation of the phenotypes and agronomic characteristics of PGR, incorporating the ECPGR, MIAPPE, Crop Ontology EMPHASIS and final user recommendations and methodologies**
Type	
Dissemination level	
Work package number	4
Author(s)	I. Verde (CREA), V. Lefebvre (INRAE), K. Adhikari (INRAE), C. Marchal (INRAE), E. Vendramin (CREA), S. Micali (CREA), F. Guzzon (IPGRI), P. Vaccino (CREA), R. Pieruschka (IBG/EMPHASIS), C. Pinheiro (UCIBIO/EMPHASIS), M. Alaux (INRAE), E. Mazzucotelli (CREA), M. A. Palombi (CREA), G. Giuliano (ENEA), P. Ferrante (ENEA)
Keywords	Plant Genetic Resources (PGR), Phenotyping, Trait and Crop Ontology, Harmonization, Metadata template, Data file format, Data repository

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**.**

Table of contents

List of abbreviations 4

Executive summary 6

1. Introduction 7

2. Activities 8

2.1. Goals of FAIRisation of phenotypic data 8

2.2. Difficulties to build unified standard protocols for phenotyping 8

2.3. Alternative proposition of harmonizing metadata and format of data files instead of standardizing 9

2.4. Designing essential recommendations to improve the reusability of phenotyping data 10

3. Results 10

3.1. A uniform, concise, simplified and rigorous metadata file associated to each phenotypic data file 10

3.2. A guideline for ensuring reproducibility and quality of phenotypic data 15

3.3. A guideline for ensuring the FAIRness of phenotypic or agronomic data deposited in data repositories 17

3.4. Expected added value of following these recommendations 18

4. Deviations 19

References 20

Annex 1. Proposed metadata template 22

List of abbreviations

AGENT	Activated Genebank Network
AI	Artificial Intelligence
API	Application Programming Interface
BrAPI	Breeding Application Programming Interface
CC-BY	Creative Commons Attribution
CC0	Creative Commons Zero
CO	Crop Ontology
CRD	Completely Randomized Design
CSV	Comma Separated Values
DMPs	Data Management Plans
DOI	Digital Object Identifiers
ECPGR	European Cooperative Programme for Plant Genetic Resources
ELIXIR	European Life-science Infrastructure
EMPHASIS	European Infrastructure for Multi-scale Plant Phenomics and Simulation for Food Security in a Changing Climate
EU	European Union
EURISCO	European Search Catalogue for Plant Genetic Resources
EVA	European Evaluation Network
FAIR	Findable, Accessible, Interoperable, and Reusable
FAO	Food and Agriculture Organisation
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
GxE	Genotype by Environment Interaction
GO	Gene Ontology
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
IPGRI	International Plant Genetic Resources Institute (currently Biodiversity International)
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
MCPD	Multi Crop Passport Descriptors
MIAPPE	Minimum Information About Plant Phenotyping Experiments
NA	Not available

OA	Open Access
ORCID	Open Researcher and Contributor Identification
PGR	Plant Genetic Resources
PGRFA	Plant Genetic Resources for Food and Agriculture
PUID	Passport Unique Identifier
RCBD	Randomized Complete Block Design
RI	Research Infrastructure
SSD	Single Seed Descent
TO	Trait Ontology
TSV	Tab-Separated Values
WIEWS	World Information and Early Warning System
XML	Extensible Markup Language

Executive summary

Plant phenotyping is an essential procedure not only for identifying and classifying plant species, but also for describing the diversity of accessions of a given species, based on observable or measurable differences. Over time, the methods and scales for taking phenotypic measurements have evolved, from visual observations, in categories, such as colours, or measured using simple tools, such as scales, to increasingly fine instrumented measurements. These phenotypic data, which take advantage of advances in emerging techniques such as omics or precision agriculture, play an essential role in the management of plant genetic resources (PGR) and support crop improvement strategies.

However, the unique characteristics and specialized requirements of different plant species pose significant challenges to the establishment of a universal framework for the collection and evaluation of phenotypic data. The heterogeneity of plant phenotypic data, due to inconsistent measurement methods and scales between research teams or experimental trials - ranging from manual notations or using human-calibrated instruments to imaging technologies - complicates their integration and sharing as part of collaborative data-driven research projects.

Additionally, the lack of standardized protocols, controlled vocabularies, and metadata descriptors across repositories limits the interoperability and comparison of data from diverse sources. Current gaps in research infrastructure and ontology databases exacerbate this issue, making it challenging to adopt a unified approach to phenotypic data management. These challenges highlight the need for harmonized frameworks and standards that ensure consistency, usability, and accessibility of phenotypic data for all stakeholders involved in the collection, management, and conservation of plant genetic resources (PGR).

In this deliverable, we analyse the challenges of creating a unified standard for characterizing PGR (as outlined in D4.1), while analysing gaps in the existing frameworks, and we recommend strategies to make harmonization of phenotypic data more achievable. We emphasize adherence to the FAIR (Findable, Accessible, Interoperable, Reusable) principles and suggest recommendations for experimental designs, data collection protocols and methods, data repositories. We also propose an extended template that builds on and complements existing metadata template formats. Finally, we advocate the adoption of a research infrastructure to facilitate the implementation of integration by fostering cross-institution collaboration.

1. Introduction

Deliverable D4.3 is built on in-depth analysis of D4.1 in which all the initiatives, including phenotypic methodologies and ontologies available in four groups of crops have been inventoried and reviewed. The four crop groups were: fruit trees, fruit vegetables, leafy vegetables, and grains (cereals and legumes). Gaps and redundancies, notably the heterogeneity of phenotypic descriptors and relative ontologies, as well as the absence or the scarcity of metadata, were also identified in D4.1.

Deliverable D4.3 aims at providing unified crop-specific standards and methodologies for phenotyping and poses several challenges. The vast diversity of plant species and their phenotypes differing in biology, growth and adaptation to specific environments, poses a challenge to adhere to a general standard that may not capture the unique characteristics of all the accessions or even landraces exhibiting unique traits. For example, protocols built for perennial crops like Citrus may not be relevant to annual crops like wheat. Additionally, experimental design and procedures to measure phenotypes, their methods and scales, are usually specific to the context of the study. For example, to measure drought resistance, the timing, duration and environment features can be highly specific and may differ for a same crop studied in different environments. The evaluation of PGR should also consider the diversity of stakeholders involved and the scale of experiments; large scale breeding experiments may use high precision tools for trait scoring while this can be irrelevant for a small-scale farmer experiment. These complexities make the prospect of unified standards elusive and very difficult, if not impossible to achieve. It is therefore preferable to use the term 'harmonized' rather than 'standardized' and to aim at creating a flexible framework that takes these complexities into account, to make the harmonization of phenotypic data more achievable. On the other hand, harmonization includes not only methods and protocols, but especially the data formats used to assess plant traits across different research groups, breeding programs, and geographic locations. A minimum information set to be included is necessary. This process guarantees consistency, comparability, and interoperability of phenotypic data, which is crucial for drawing accurate conclusions and enhancing collaboration across scientific communities.

Concerns have been raised about the difficulties of standardization. For example, a pilot study carried out as part of The European Infrastructure for Multi-scale Plant Phenomics and Simulation for Food Security in a Changing Climate (EMPHASIS; https://emphasis.plant-phenotyping.eu/) aimed at integrating three information systems resulted in difficulties in achieving this goal, notably due to differences in the names of the variables (R. Pieruschka, personal communication). The importance of having rigorous metadata associated with phenotypic data was highlighted, to allow the data of a specific experiment to be reused in a wider analysis, meta-analysis, modelling or through Artificial Intelligence (AI) (Saint Cast et al., 2022; Papoutsoglou et al., 2023). The importance of trait and crop ontologies for this kind of data has been raised and suggestions were made for using machine-readable trait measures or variables described in ontologies, even though there are redundancies in many existing ontologies (Dumschott et al., 2023).

In D4.3, we took into consideration all the work already done by different institutions and all the positive examples, such as the European Search Catalogue for Plant Genetic Resources (EURISCO) and Minimal Information About a Plant Phenotyping Experiment (MIAPPE) file templates, as well as other recommendations from various Research Infrastructures (RI) like the Breeding API (BrAPI; https://brapi.org/), European Life-science Infrastructure for Biological Information (ELIXIR;

https://elixir-europe.org/), EU recommendations and other existing documents (https://rdmkit.elixir-europe.org/; https://faircookbook.elixir-europe.org/content/recipes/introduction/FAIR-cookbook-audience.html) to improve the FAIRness of data (Wilkinson et al., 2016).

In conclusion, harmonization of phenotyping protocols and associated metadata is essential and crucial to ensure data reliability, foster collaboration, and accelerate progress in plant breeding and research. It enables global efforts to improve crop performance and address challenges such as climate change, food security, and sustainable agriculture.

2. Activities

2.1. Goals of FAIRisation of phenotypic data

The FAIRisation of phenotypic data aims to enhance the accessibility, integration, and utility of data generated in Plant Genetic Resources (PGR) research. At its core, the FAIR—Findable, Accessible, Interoperable, and Reusable— principles ensure that phenotypic data, which are growing rapidly in scale and complexity, can be more easily shared and used. One key goal is to be able to merge data from different independent experiments, ensuring consistency and comparability across datasets from varied sources, environments, and methodologies and making it easier for other researchers to access, review, and reuse the data. This is crucial for conducting collaborative science and reducing duplication of effort. This integration allows researchers to build larger and more robust datasets that can lead to more reliable conclusions and insights into plant performance under various conditions and thus accelerate innovations. Another important aspect is the ability to merge data from different approaches, including phenotypic-related omics (transcriptomics, proteomics, metabolomics, ionomics, etc.). By integrating the phenotype-related omics data with genotypic and sequencing data, it becomes possible to create a comprehensive view of how genotypes are translated into observable traits. This facilitates systems biology approaches and deepens our understanding of the genetic bases for key agronomical traits. In a time where data are growing at an unprecedented rate, ensuring the accuracy, reusability and responsible management of phenomic data is fundamental for advancing agricultural research and addressing global food security challenges. Beyond sharing, the FAIR principles enable data to be reused in modelling across a pipeline. Researchers can apply machine learning or predictive models to phenotypic datasets, enhancing their ability to predict plant responses and performance in different scenarios, which is particularly important for breeding and crop improvement programs (Cembrowska-Lech et al., 2023; Watt et al., 2023).

2.2. Difficulties to build unified standard protocols for phenotyping

Building unified standard protocols and descriptors for assessing the phenotypes and agronomic characteristics of PGR faces significant challenges starting from the simple consideration that each laboratory has its own practices and equipment and internal standardized methods. Several phenotyping standards already exist for different experiments, specific traits, and often for specific purposes. Consequently, these different standards are both redundant for the same species, but at the same time have few differences that make them complementary. The use of these phenotyping standards is therefore very time-consuming since sometimes several standards need to be completed

for the same experiment, depending on the recipient. These difficulties are a real obstacle to their adoption by the scientific community. Besides, attempts like the EMPHASIS pilot study and efforts within the Arabidopsis community have struggled to unify data across different systems, due to variations in protocols, data formats, and technologies. These initiatives have demonstrated that while there are many existing protocols, such as those from the European Cooperative Programme for Plant Genetic Resources (ECPGR), European Evaluation Network (EVA), and International Plant Genetic Resource Institute (IPGRI), the sheer number of available standards causes uncertainty about which one to use in the scientific community, leading to inconsistencies in data collection and evaluation. This inconsistency hinders efforts to merge data collected across different experiments, regions, and platforms.

A connection and integration framework of data discovery solutions and software tools has been developed within EMPHASIS to enable the quick generation of adapter servers that handle the task of connecting a local research institute's data-warehouse to a network of other registered adapters running at respective other research institutes (https://zendro-dev.github.io/). Implementing the adapters for interoperability of these data-warehouses was tested in a pilot study between three organisations. However, the implementation turned rather difficult largely due to differences in data standards. The data interoperability was addressed by a searchable metadata index FAIDARE (https://urgi.versailles.inrae.fr/faidare/) that enables searching and exploring data sets. This points at the relevance of metadata as an important approach to ensure reusability of data. This was complemented by the bottom-up development and ongoing continuous improvement of a metadata standard called MIAPPE (Papoutsoglou et al., 2020; https://www.miappe.org/). In response to these challenges, we propose creating a uniform, concise, and simplified metadata template that would accompany each phenotypic dataset. Our goal is therefore to enhance the adoption of common practices to fill in the metadata template instead of imposing standards that are finally not used.

2.3. Alternative proposition of harmonizing metadata and format of data files instead of standardizing

Instead of establishing a single protocol for each crop and trait, which is almost impossible given the number of "crop x trait" combinations, the emergence of new phenotyping technologies and the enormous possibilities offered by phenomics, we propose to use a metadata file as a standardized record of essential information about the experimental conditions. This metadata file associated to each dataset would enable researchers to work with diverse datasets while guaranteeing data interoperability and would allow to capture essential information about the experimental setup, trait definitions and environmental conditions, ensuring that datasets are comparable, regardless of the specific protocol used. To avoid redundancy and the researcher's need to fill in several similar/redundant templates according to the data repository, this metadata template should be unique to ensure its broad adoption by the research community, and it should include both mandatory and optional fields, balancing thoroughness with ease of use. For instance, mandatory fields would cover crucial aspects like species, accession names, the Digital Object Identifier (DOI) for each accession, and the organ considered during phenotyping. Optional fields, such weather data, experimental design, phenotyping protocols, would allow for more detailed metadata when necessary but would not overburden researchers. By keeping the mandatory fields simple and clear, while offering flexibility through optional fields, this approach would encourage widespread use and support the integration of phenotypic data across various research efforts. This concise and flexible metadata template would

promote greater data sharing and facilitate data merging without the need for researchers to change their established practices entirely.

As an example, in the frame of the Horizon EU InnOBreed project (Innovative Organic fruit Breeding and uses, grant agreement No 101061028), a common phenotypic trait list has been defined across all the partners belonging to 21 different European institutions. The participants have defined, for each trait and for each species, shared methodologies such as common assessment scales and protocols. Furthermore, meteorological data have been analysed to highlight climate tendencies and to identify the most important traits for enhanced productivity. It is probably starting from a core of prioritized traits that the adoption of harmonized and shared protocols should begin, to maximize the efforts of the various European projects and programs (https://innobreed.eu/).

2.4. Designing essential recommendations to improve the reusability of phenotyping data

To improve the reusability of phenotyping data, several key recommendations can be implemented, aligned with the FAIR data principles:

- Harmonize the **dataset formats** to facilitate merging of data across different studies and institutions, distinguish the raw from the transformed data, and ensure compatibility with other downstream analyses;
- Associate **a unique, rich, descriptive metadata file** to each dataset that is human and machine-readable to enable easy retrieval and understanding of the dataset;
- Assign **a permanent digital identifier** to each plant accession (such as DOI) and to each individual plant;
- Integrate **established trait and crop ontologies** to ensure that the phenotypic terms are used consistently across studies and even across plant species;
- Deposit the **dataset and the metadata files in an open access well-maintained data repository**, such as recognized governmental or international repositories, that offer robust search and download facilities with user-friendly documentation describing the elements of the data. Some examples of this could be research information system such as openAIRE (https://explore.openaire.eu/) or open access government-supported public repositories like https://www.data.gouv.fr/fr/.

Following the listed recommendations, the reusability and sharing of phenotypic data can be significantly improved, which will finally enhance innovation in plant science.

3. Results

3.1. A uniform, concise, simplified and rigorous metadata file associated to each phenotypic data file

In the framework of this deliverable, we compared the phenotypic data collection templates of three European initiatives dealing with phenotyping of plant genetic resources (PGR): EVA, Activated Genebank Network (AGENT) and GERMINATE.

EVA is an international project managed by the ECPGR and supported by the German Federal Ministry for Food and Agriculture. It aims at increasing the use of crop genetic diversity and the diversity of

stakeholders in plant breeding. Through crop-specific public–private partnerships, EVA generates standardized evaluation data (both phenotypic and genotypic) for numerous crop accessions and landraces available in European genebanks. Several crop-specific networks are part of EVA: carrot, legumes, lettuce, maize, pepper, wheat and barley. The data are uploaded in an internal database (EURISCO-EVA, Kumar et al. 2024), with the aim of including these evaluation data in EURISCO after an embargo period. More information is available at https://www.ecpgr.org/eva.

The AGENT project aims to unlock the full potential of the biological material stored in genebanks around the globe by using FAIR international data standards and an open digital infrastructure for the management of plant genetic resources. In AGENT, seventeen partners are evaluating at the genotypic and phenotypic level more than 14 000 accessions of wheat and barley. More information is available at https://agent-project.eu/. The phenotypic data are collected using a metadata template, which is hereafter discussed.

The GERMINATE database is a generic plant genetic resources database and offers facilities to store both standard collection information and passport data along with more advanced data types such as phenotypic, genotypic and field trial data (Raubach et al., 2021). It stores data of 21 different crops. More information is available at https://germinateplatform.github.io/get-germinate/#.

The metadata templates of EVA and AGENT are very similar. These templates include information on plant material, experimental metadata (*i.e.* Trial ID, geographical information on the location, data on the experimental design), treatments (*e.g.* irrigation, insecticide, growth regulators) and traits to be evaluated (including range of permitted values, such as rating scores or metric values). The main difference among the two metadata templates is the possibility of attaching an overview of the experimental design in the AGENT template. The GERMINATE platform uses a more simplified template for the upload of metadata and phenotypic data. The main difference with the EVA and AGENT templates is an initial sheet that provides fundamental information on the submitted data (*e.g.* descriptions, format, contact information, language).

In addition, a new minimum MIAPPE template is currently in development in the frame of the MIAPPE consortium (Pommier et al., https://github.com/MIAPPE/MIAPPE/blob/master/Templates/MIAPPE_Minimal_Spreadsheet_Template.xlsx).

We conducted a comparative analysis of these existing templates to evaluate their strengths and limitations in the context of phenotypic and agronomic data collection. The primary goal was to identify the gaps in these templates and to propose a set of recommendations to address these gaps. We then built an improved metadata template by merging the existing templates with sets of descriptors designed to harmonize the collection, description and sharing of phenotypic and agronomic data.

The recommended template (Annex 1) contains comprehensive sections to describe passport data, experiment design, trait characteristics, etc. It also incorporates the use of ontologies and digital identifiers for plant materials like DOIs and it suggests controlled vocabularies in the sections where it is applicable. Utilizing this new template helps to comply with the FAIR principles making data more accessible and reproducible. Similarly, it also aims to simplify meta-analysis and data aggregation by ensuring compatibility with databases like EURISCO. It is important to retain the detailed information of any experiment to understand and reuse the data, for instance, the scale of measurements, type of phenotype protocol used, and the kind of experiment design adopted. For this, our template emphasizes the role of frequency of treatments, plot level data and environment data which are important for understanding GxE interactions and ensure repeatable experiments. One of the gaps in the above templates was that they focus more on the crop description and less on the experimental

design, trait measurement and treatment descriptions. The recent efforts of AGENT and EVA templates have partially addressed this but still lack comprehensive descriptors for treatment frequencies, scaling methodologies and quality control measures. While our recommendations align with existing standards, it builds upon them by adding detailed experimental metadata, interoperability features and can assist in simplifying data usability and collaboration. However, to achieve the harmonized structure for phenotypic data, periodic validation must be done to check the compliance to the recommended format and correct data entry by the users. Our review of existing standards and the status of phenotypic data collection and sharing also found that existing ontologies often fail to cover a wide range of crops. For example, they are not context-specific for some experiments that incorporate biotic stress and localized climatic variations. Consequently, researchers are forced to define their own terms or provide incomplete information that leads to redundancies or gaps in datasets. Similarly, there are no broadly accepted standards and guidelines for controlled vocabularies for standardizing terms across different types of experiments, trait scoring methods, experimental setups. This poses challenges in harmonizing the data across databases and interoperability of data. Moreover, to make data machine-readable, data formats should align with machine-readable formats like Extensible Markup Language (XML). Many existing databases and datasets lack this integration that further complicates the integration of phenotypic data across databases. Database platforms and repositories can leverage research infrastructure framework to create centralized repositories for ontologies and vocabularies as well as integrate recommendations to adopt ontologies, aligning data formats, improve the standards of data, simplify data submission and sharing.

Based on our review and evaluation of descriptors from various existing standards, we compiled a list of essential descriptors and strongly encourage their use when collection and evaluation of phenotypic data. This list provides a concise overview, with more detailed and comprehensive recommendations available in our template guidelines.

Plant/Crop Information

Descriptor	Description
Material ID	Each experiment should have its internal identification for accessions. In this context, material ID refers to the unique identifier of germplasm accession, centrally provided within the project.
Accession ID	When the PGR accessions are derived from a genebank, it is essential to put the accession's original identifier provided by the genebank. This helps with easy accessibility and reusability of the samples and related data.
Material type	The biological material that is used in the study should have descriptive information including its pedigree and taxonomic information. Types could be original accession (material used in the same way as received from genebanks, *e.g.* landraces), SSD (single seed descent-line, derived from original accession), cross (testcross population, derived from original accession), controls (checks used in experiments), tester (anonymized parent for crosses).
Institution code	FAO-WIEWS code of the institute maintaining the original material, following Multi-crop Passport Descriptors (MCPD). Use TBD (for "to be determined") for material where institute is unclear. The

	MCPD is a widely used international standard to facilitate germplasm passport information exchange.
Binomial Latin taxonomic name	The full genus names (with the first letter in uppercase) and species names (entirely in lowercase) must be mentioned.
Vernacular name of the organism	Common English name of crop under evaluation, *e.g.* barley, wheat, carrot, pepper, eggplant, tomato...
DOI	A permanent unique identifier (PUID) for the plant Material, following MCPD. Using a DOI is highly recommended in the present context for its usefulness in finding and accessing data easily.
Provenance country	ISO3 code of the country in which the material was collected or bred. For derived material this can be the country of the parent material or remain as NA (not available).
Conservation status	Information on whether the material is conserved ex-situ, in-situ, is breeding material, etc.

Experiment Information

Descriptors	**Description**
Trial Id	Identification of the trial, centrally provided by the project or the institution.
Experiment description	Descriptive name of the experiment set-up including the name and goals of the experiment.
Experiment type	*e.g.* field, greenhouse, growth chamber (phytotron, climatic chamber).
Organisation	Company/Institute/Group organising the trial.
Experimental design information	Information about the experiment setup can help reproduce the experiment and provide broader information in the phenotype data. Data sets should mention the design of the experiment for example: randomized complete block design (RCBD), completely randomized design (CRD), split-plot design, number of blocks, number of plots per block, plot dimensions including spacing within and between rows, etc. The experiment should be planned using statistically relevant designs, with the layout structured to support spatial analysis effectively
Location information	Country, city and/or region of the trial, site name, altitude above sea level, Global Positioning System (GPS) coordinates of latitude and longitude of the experimental sites; this increases the reusability of data and accurate information on climatic conditions can be retrieved. If the experiment is conducted in a closed environment, information should be also provided about the setting (e.g. greenhouse, polyethylene tunnel, growth chamber), the application of any artificial means to control environmental conditions (e.g. artificial illumination, thermoperiod, humidity control) and the conditions applied (e.g. photoperiod duration, light intensity at canopy level, temperature and humidity levels during day and night).
Experiment timing	Timing of experimental procedures like date of sowing, dates of beginning and end of the data collection, date of experiment termination; if possible, each collected data point should have an associated date/time.

Phenotypic protocol	Methods used to capture phenotype data (manual/use of machines). With the advent of modern technologies and portable machines in crop studies, proper information about the phenotypic protocols and calibration are beneficial for maintaining consistency in trait scales and consequent values within experiments.
Weather data	Datafile or relevant website links about information of weather data during the experiment where applicable.

Treatment and trait information

Descriptors	Description
Trial ID	Unique identifier of the trial/study/experiment (see above, centrally provided by the project or the institution).
Treatment description	Type of treatment (biotic stress, nutrition supplement, hormonal treatment etc.), and a brief description of the treatment applied.
Application method and frequency	Date of application (if frequent, date of first application) and number of times the treatment is used (if reapplied).
Control description	Description of the status of the control(s) applied in the experiment.
Trait ID	Identification number of trait.
Trait description	Description of the trait, type and significance.
Trait ontology	Identification of trait from ontology databases (*e.g.* Length of fruit: CO: 333:2000224).
Measurement method and frequency	The method of measurement of trait is recommended to be included whether it was manually done or with portable machines or drones. Similarly, the frequency of measurement is essential to understand the timepoints of application (*e.g.* 0, 30 and 60 days after sowing).
Trait scoring protocol	The trait scoring protocol describes the scales used to measure the trait, it can be a rating score or a metric value.
Observation values	The trait values obtained during the experiment are the so-called "phenotypic data" from the experiment. The method and scale used for the measurements must be specified, to ensure that the study is reproducible and the data reusable.

3.2. A guideline for ensuring reproducibility and quality of phenotypic data

Reproducibility and quality of phenotypic data is crucial for robust scientific research. Here we report a guideline of the minimal recommendations leading to more reliable and impactful scientific findings.

Study design

1. Establish the specific research questions and hypotheses;
2. Identify the phenotypic trait(s) you want to assess and treatment(s) of interest (also called the factor(s)) on the assessed trait(s);
3. Choose the appropriate statistical methods and the statistical model you will apply, the software to use and report the rationale behind the chosen statistical methods;
4. Design your experiment, with minimum 2 independent trials, minimum 2 blocks and 2 replicates per block, to have statistical significance on the collected data. Use randomization in treatment (factor) assignment to reduce bias.

PGR sampling and controls

5. Ensure that the sample size of PGR is adequate and representative of the overall population;
6. Use the same methods for sample collection across all independent experiments (trials);
7. Include positive and negative plant controls in each experiment. Positive controls are samples known to produce a specific response or outcome and are useful for experimental validation; negative controls are samples not expected to produce a response and are useful in

identification of background noise, potential contamination, and/or experimental errors. Both positive and negative controls should be common in all independent experiments. If possible, add internal controls that provide a baseline for comparison and are useful to measure the variability within a specific experiment or across different experiments. Replicate at least a few accessions as biological replicates to ensure robustness and account for variability.

Data acquisition

8. Use consistent methods and tools to measure each phenotypic trait to standardize measurements and reduce ambiguities and produce a detailed experimental protocol for each trait to minimize variability (the protocol can be loaded in the publication and in the data repository);
9. Certify that the instruments you use are well-calibrated and train personnel to minimize human error and ensure uniformity in data collection. Consistent methods and scale of measurement should be established across trials such as utilising a catalogue book for disease score in plants, or colour, size grading of fruits, or guidelines to measure height of plant especially when it is manually collected. This minimizes the error in data and allows comparability. Some examples for these kind of guidelines for calibration could be colour charts (https://www.cn.nysed.gov/sites/cn/files/tomatocolorchart.pdf), diagrammatic disease score (Rios et al., 2013; Del Ponte et al., 2017).
10. In terms of controlled experiments, keep environmental conditions constant across all independent experiments as much as possible (e.g., temperature, humidity, light, etc.). In field conditions however, it is not possible to keep uniform gradients of environment. In these conditions, a proper documentation of local weather conditions for the duration of experiment is encouraged. Accurate recording of geographical coordinates and data collection dates, and usage of Application Programming Interfaces (API) that provide high resolution of local historical weather data (https://openmeteo.substack.com/p/historical-weather-api-with-highhttps://openmeteo.substack.com/p/historical-weather-api-with-high or https://www.weatherbit.io/api/historical-weather-apihttps://www.weatherbit.io/api/historical-weather-api) to document environmental conditions for experiments allows broader interpretation to the phenotypic data.
11. Repeat measurements on at least three samples as biological replicates, to ensure the accuracy of the measurements.

Data Management

12. Use standardized formats for data entry (comma separated values, CSV; tab separated values, TSV) to prevent errors;
13. Document metadata using standardized formats, including methods, conditions, and any deviations from protocols;
14. Implement version control for datasets to track changes over time;
15. Deposit the checked dataset and the associated metadata files in a well-maintained repository.

By implementing the listed recommendations and adhering to technical standards, researchers can significantly enhance the reliability and reproducibility of their phenotypic data, facilitating their reuse and sharing for ultimately leading to more credible and impactful scientific outcomes.

3.3. A guideline for ensuring the FAIRness of phenotypic or agronomic data deposited in data repositories

Ensuring the FAIRness of phenotypic data refers to making the data Findable, Accessible, Interoperable, and Reusable. These principles help maximize the utility of data for research and scientific discovery. Here's a summary of guidelines to ensure FAIRness in phenotypic data.

Findability

1. Assign persistent identifiers: each dataset should have a DOI (Digital Object Identifier) or another persistent identifier to ensure it can be reliably found and cited;
2. Metadata standards: use standardized metadata formats, as the format we propose in D4.3, to describe data content, ensuring it is discoverable through search engines and data catalogues;
3. Make metadata searchable: ensure that phenotypic data is indexed by appropriate data repositories with rich, descriptive metadata to enable easy searching.

Accessibility

4. Open Access (OA): data should be openly accessible, if possible, or provide clear information about access requirements (e.g., through user agreements, licenses, or requests for controlled access);
5. Data availability: ensure that data can be downloaded or accessed through standard protocols [e.g., Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP) or Application Programming Interface (APIs)];
6. Clear licensing: clearly define usage terms for public as well as commercial usage, e.g., via open licenses like Creative Commons Zero (CC0) or Creative Commons Attribution license (CC-BY).

Interoperability

7. Use standardized formats: ensure the data are represented in standardized formats (e.g., CSV, JSON, XML) that are widely used in the domain, allowing it to be easily combined with other datasets;
8. Ontologies and controlled vocabularies: use established ontologies (e.g., Trait Ontology, Crop Ontology, Gene Ontology (GO)) to annotate phenotypic data. This enhances cross-referencing and harmonization across datasets;
9. Cross-referencing: include links to relevant external resources (e.g., gene databases, agriculture databases) whenever possible.

Reusability

10. Clear documentation: provide detailed documentation, including data dictionaries, descriptions of the phenotypic variables, methodologies, and any assumptions or limitations in the data;
11. Data provenance: maintain records of the data's origin, methodology, processing steps, and any transformations they underwent;
12. Versioning: ensure version control of the data, so users can track changes or updates over time and access previous versions if needed;

13. Data quality: include quality control measures and metrics to ensure data reliability and completeness.

Additional best practices

14. Ethical considerations: respect privacy and ethical guidelines when sharing phenotypic data (by adhesion to General Data Protection Regulation (GDPR) rules);
15. Data integration: facilitate integration with other data types (e.g., genomic, environmental) to promote multidisciplinary analysis and use; Use crop and trait ontologies as much as possible, at least for the basic common traits, and not create new ones. For novel traits or very specific traits that were never described before, such as robustness of plant immunity in changing environment, it is necessary to well describe the measurements and the calculation steps of the new trait and deposit the data and the script to calculate them in data and script repositories (*e.g.* Billaud et al 2024).

3.4. Expected added value of following these recommendations

We need to encourage the implementation of these recommendations to foster the collaborative research. The construction of GRACE-RI (and GRACE-ERIC) will give the opportunity to translate this advice in specific guidelines and protocols since there will be a constant exchange of information among genebanks and institutions to implement these recommendations. Thus, GRACE-RI should bridge the gap between theoretical recommendations and practical implementations. With the collaborative environment provided by GRACE-RI, the usability, scalability, effectiveness and limitations can be assessed and evaluated. This fosters more research engagements with shared interests in data-driven research and agreed refinements to meet practical needs. Furthermore, promotion of the recommendations within a RI framework should improve efficiency and reachability in terms of scale of information.

3.5. Drivers to convince stakeholders to adopt the recommendations

The framework recommended in this deliverable aims to address the gaps in the existing initiatives as well as to develop a way to align the present infrastructures to make plant phenotypic data collection, management, and presentation more harmonized with compliance to FAIR (Findable, Accessible, Interoperable, and Reusable) principles. This provides an opportunity for researchers to integrate data from multiple sources efficiently, allowing better utilization and understanding of PGR data which can foster more robust analyses in research and promote cross-disciplinary collaborations thereby maximising the value of phenotypic data. With the massive generation of phenotypic, genotypic, and environmental data, research initiatives are looking to accelerate the acquisition and alignment of these multi-dimensional data. The ease of integration offered by controlled vocabulary and the use of ontology indexes can be useful for both the public and private sectors to accelerate decision-making in conservation, breeding or evolutionary studies.

Since our recommendations adhere to pre-existing standards (MIAPPE, EVA, AGENT), they ensure consistency of data formats (types of data to be entered, *e.g.*: biological material, experiment data etc.) while also limiting complexity by focusing on key metadata and descriptor variables and increasing the flexibility of secondary variables. This prevents end-users from being overwhelmed by the requirements to follow for the collection and sharing of data.

Adopting the above recommendations would foster the formatting of open access phenotyping data in existing standards to increase data reliability, their use, their sharing and collaborations. For achieving this, it is necessary to validate and cross-check the compatibility of the proposed framework with different types of experimental data, across different end-users and sufficient feedback system should be established to implement the improvements needed. Promoting collaboration with stakeholders (PGR databases, public/private end-users) is another essential action. This can be achieved through communication to all possible actors through courses, webinars, publications, etc. Besides, for adoption, our framework should provide adequate guidelines in usage and dissemination of collection, management, and presentation of data, so that it is easy to use. A non-exhaustive list of actions that could be implemented to further encourage the adoption of our recommendations is reported here:

1. Encourage collaborative data sharing from the outset, setting expectations for data management, storage, and sharing through trainings or documentation on sharing to ensure consistency across contributors;
2. Many funders and institutions now require Data Management Plans (DMPs) as part of grant proposals to ensure data is effectively managed and shared. Thus, preparing a draft and following DMP outlining how data should be collected, shared and protected can ensure better reliability and usability;
3. Publish datasets in data journals, as supplementary material in scientific journals, or in a well-maintained data repository. This describes the dataset and its relevance, enhancing discoverability and potential use by other researchers and enables citation of datasets;
4. Use ORCID IDs for researchers to link their identities to datasets and publications;
5. Use DOIs or other persistent identifiers for datasets to ensure they remain accessible and citable over time;
6. Use collaborative platforms (e.g., Zenodo, Figshare) that support versioning and commenting on datasets. Some platforms even allow others to fork datasets, enabling derived studies;
7. Engage with online communities or forums in your field to promote your data and gather feedback or suggestions for improvements.

4. Deviations

We deviate slightly from the original objective by promoting the harmonization of data files and the use of a metadata model, rather than the standardization of protocols, which is an almost impossible goal to achieve based on our studies.

References

Adam-Blondon, A.-F., Alaux, M., Brouwer, M., Kersey, P., Lange, M., Le Floch, E., Pommier, C., Schüler, D., Stein, N., & Weise, S. (2024). AGENT Guidelines for dataflow [Data set]. Zenodo. https://doi.org/10.5281/zenodo.12625360

AGENT documents: https://agent-project.eu/

Billaud, W., Hirsch, J., Ribaut, V., Tamisier, L., Massire, A., Szadkowski, M., Lopez-Lauri, F., Moury, B., & Lefebvre, V. (2024). Unveiling pepper immunity's robustness to temperature shifts: Insights for empowering future crops. *Horticulture Research*, *11*(11), uhae239. https://doi.org/10.1093/hr/uhae239

Cembrowska-Lech, D., Krzemińska, A., Miller, T., Nowakowska, A., Adamski, C., Radaczyńska, M., Mikiciuk, G., & Mikiciuk, M. (2023). An Integrated Multi-Omics and Artificial Intelligence Framework for Advance Plant Phenotyping in Horticulture. *Biology*, *12*(10), Article 10. https://doi.org/10.3390/biology12101298

Del Ponte, E. M., Pethybridge, S. J., Bock, C. H., Michereff, S. J., Machado, F. J., & Spolti, P. (2017). Standard Area Diagrams for Aiding Severity Estimation: Scientometrics, Pathosystems, and Methodological Trends in the Last 25 Years. *Phytopathology®*, *107*(10), 1161–1174. https://doi.org/10.1094/PHYTO-02-17-0069-FI

Dumschott, K., Dörpholz, H., Laporte, M.-A., Brilhaus, D., Schrader, A., Usadel, B., Neumann, S., Arnaud, E., & Kranz, A. (2023). Ontologies for increasing the FAIRness of plant research data. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1279694

EVA documents: https://www.ecpgr.org/eva

GERMINATE: https://germinateplatform.github.io/get-germinate/#

Kumar, S., Guzzon, F., Goritschnig, S., & Weise, S. (2024). The EURISCO-EVA Information System, an innovative approach to the data management of multi-site crop evaluation data. *Genetic Resources*, *5*(10), Article 10. https://doi.org/10.46265/genresj.IHXU5248

MIAPPE documents: https://www.miappe.org/

Papoutsoglou, E. A., Athanasiadis, I. N., Visser, R. G. F., & Finkers, R. (2023). The benefits and struggles of FAIR data: The case of reusing plant phenotyping data. *Scientific Data*, *10*(1), 457. https://doi.org/10.1038/s41597-023-02364-z

Papoutsoglou, E. A., Faria, D., Arend, D., Arnaud, E., Athanasiadis, I. N., Chaves, I., Coppens, F., Cornut, G., Costa, B. V., Ćwiek-Kupczyńska, H., Droesbeke, B., Finkers, R., Gruden, K., Junker, A., King, G. J., Krajewski, P., Lange, M., Laporte, M.-A., Michotey, C., … Pommier, C. (2020). Enabling reusability of plant phenomic datasets with MIAPPE 1.1. *The New Phytologist*, *227*(1), 260–273. https://doi.org/10.1111/nph.16544

Raubach, S., Kilian, B., Dreher, K., Amri, A., Bassi, F. M., Boukar, O., Cook, D., Cruickshank, A., Fatokun, C., El Haddad, N., Humphries, A., Jordan, D., Kehel, Z., Kumar, S., Labarosa, S. J., Nguyen, L. H., Mace, E., McCouch, S., McNally, K., … Shaw, P. D. (2021). From bits to bites: Advancement of the Germinate platform to support prebreeding informatics for crop wild relatives. *Crop Science*, *61*(3), 1538–1566. https://doi.org/10.1002/csc2.20248

Rios, J. A., Debona, D., Duarte, H. S. S., & Rodrigues, F. A. (2013). Development and validation of a standard area diagram set to assess blast severity on wheat leaves. *European Journal of Plant*

Pathology, *136*(3), 603–611. https://doi.org/10.1007/s10658-013-0191-xhttps://doi.org/10.1007/s10658-013-0191-x

Saint Cast, C., Lobet, G., Cabrera-Bosquet, L., Couvreur, V., Pradal, C., Tardieu, F., & Draye, X. (2022). Connecting plant phenotyping and modelling communities: Lessons from science mapping and operational perspectives. *In Silico Plants*, *4*(1), diac005. https://doi.org/10.1093/insilicoplants/diac005

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O., & Schurr, U. (2020). Phenotyping: New Windows into the Plant for Breeders. *Annual Review of Plant Biology*, *71*(Volume 71, 2020), 689–712. https://doi.org/10.1146/annurev-arplant-042916-041124

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Annex 1. Proposed metadata template

This template was designed by reviewing and incorporating current existing template frameworks (Minimum information about plant phenotyping experiments (MIAPPE), Activated genebank network (AGENT), European evaluation network (EVA) and GERMINATE) for phenotypic and agronomic data collection under different experiment conditions. This document describes different sheets and their fields. The information and field are colour coded for their importance levels. Each section (numbered) highlighted in red are mandatory and the parts highlighted in green are optional. Additionally, the mandatory descriptors in each section are coloured red.

1_Plant material Information

The plant material information section provides important passport data of the biological material used for the experiment such as taxonomic information, its status during use (breeding material, tester etc), its pedigree (if applicable).

1_Plant material information

This section provides the general information of the material studied (e.g. plants grown from a certain bag or seed lot, or plants grown in a particular field). (This section is mandatory)

Descriptors	Material_ID	Accession_ID*	MaterialType*	InstitutionCode*	Genus*	Species*	Organism*
Description	*Unique identifier of germplasm accession within the project, centrally provided by the Work package.*	*Identifier for accessions in genebank (for original accessions). For derived material (SSD, crosses) these are linked to parent accessions provided by genebanks (original accessions)*	*The type of biological material used. possible types are: original accession (material used in the same way as received from genebanks, e.g. landraces), SSD (single seed descent-line, derived from original accession), cross (testcross population, derived from original accession), check (controls used in experiments), tester (anonymized parent for crosses),*	*FAO-WIEWS code of the institute maintaining the original material, following MCPD. Use TSD for material where institute is unclear or if the institute doesnot have FAO-WIEWS code*	*Genus name for taxon. Initial uppercase letter required.*	*Specific epithet portion of the scientific name in lowercase letters. Only the following abbreviation is allowed: 'sp.'*	*common English name of crop under evaluation, e.g. barley, wheat, carrot*
example	GPC050380	14PT201	original accession	ITA391	*Capsicum*	annuum	Pepper
example	GPC049470	14PT22	original accession	ITA391	*Capsicum*	annuum	Pepper

DOI	DOI_Parent	Accession_IDparents	inEURISCO	ProvenanceCountry*	ConservationStatus	RegistrationYear	CollectionYear	GeneticInformation	Remarks
A permanent unique identifier (PUID) for the plant material, following MCPD. Include where available.	*A permanent unique identifier, for parent material, e.g. for the original genebank accession the SSD line was selected from. Include where available.*	*Identifier for parent material. Applicable for the original genebank accession the SSD was selected from. Or the cross was made, in case of cross put in the format Female,Male*	*information on whether material is already included in EURISCO (yes/no)*	*ISO3 code of the country in which the material was collected or bred. For derived material this can be the country of the parent material or remain empty.*	*Information on whether the material is ex-situ or in-situ*	*(for varieties)*	*(for landraces or wild populations)*	*Link to genotypic data (if available)*	*free text remarks on material.*
NA	NA	NA	Y	ITA	ex-situ	1981	NA	NA	NA
NA	NA	NA	Y	ITA	ex-situ	1981	NA	NA	original genebank accession

2_Experiment data

The experiment data section is designed to collect the detailed information about the experiment, such as description of experiment, location, experiment design type, layout, start and end of experiment, type of treatment etc. Depending on the type of experiment this, the section to describe experiment metadata is divided into two sections: **Experiment data_open field (2a)** field and **Experiment data_controlled conditions (2b).** The optional fields may not be required depending on the type of experiment. Each trial should have a unique id provided by the project/work package or institute responsible whichever is applicable. This makes it easy to understand the experiment design.

2_Experiment data

The experiment data are divided into two sections each to be used for experiment done either in open field or controlled conditions. More column fields can be added according to the type and requirement of the experiment. Optional columns may not be adapted to all crops. for empty cells put 'NA'. This section is mandatory

2a_Experiment data_openfield

Descriptors	Trial_ID*	ExperimentDescription*	Organisation*	ContactPerson*	ExperimentalDesign	Country*	Location*	Latitude*	Longitude*	HeightAboveSeaLevel*
Description	unique identifier for each experiment/trial	descriptive name of the experiment set-up including the goals of the experiment	company/institute/group organising the trial	person in charge of trial/providing data, include email if relevant	Design of the experiment for example :RCBD, CRD, Split-plot design	country in which the field trial is located	city or region of the trial	GPS coordinates [preferably in decimal degrees, e.g. 40.741895]	GPS coordinates [preferably in decimal degrees, e.g. -73.989308]	[m]
example	**Ecrop_trial#**	barley field trial 2020/2021	Institute of Plant Breeding and Genetic Resources, ELGO-DIMITRA,	john.doe@instt.com	RCBD	Greece	Thermi-Thessaloniki	40.53	23.00	150

2b_Experiment data_controlled conditions

Descriptors	Trial_ID*	ExperimentDescription*	ExperimentType*	Organisation*	ContactPerson*	Country*	Location*	Site	ExperimentalDesign*	StartDate*
Description	unique identifier for each experiment/trial	descriptive name of the experiment set-up including the goals of the experiment	e.g. growth chambers, in vitro, greenhouse	company/institute/group organising the trial	person in charge of trial/providing data, include email if relevant	country in which the field trial is located	city or region of the trial	name of the greenhouse/growth facility, as applicable	(RCBD, CRD, lattice design, latin square design)	**Date of of sowing of the crop/commencement of experiment)**
example	**crop_trial#**	biotic stress [pathogen	lab	JKI	john.doe@instt.com	**Germany**	**Quedlinburg**	JKI	CRD	2022-03-12
example	**202013**	Test Stems/ P. capsici on	growth chamber/lab	INRAE	john.doe@instt.com	France	Montfavet	CCS	CRD	2019-03-07

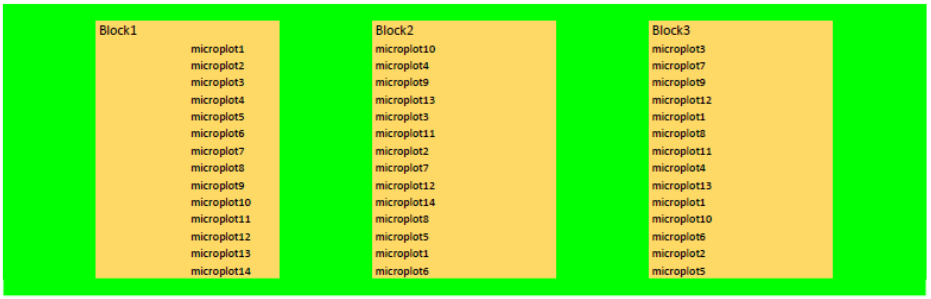
Experiment data_openfield

StartDate*	EndDate*	SoilTexture	NumberOfBlocks	NumberOfplotsPerBlock	Plotlength	PlotWidth	NumberOfRowsPerPlot	DistanceBetweenRows WithinPlots	DistanceBetweenRows BetweenPlots	PhenotypicProtocol*	Irrigation	Weather Data	Remarks
Date of commencement of the experiment in the format [yyyy-mm-dd]	**Termination date of experiment (This should be the date where final data of the study is taken in the field) in the format [yyyy-mm-dd]**	e.g. following FAO soil classification	[replicates]	[replicates]	[m]	[m]		[m]	[m]		[yes/no]	datafile or relevant website links about	free text to provide relevant info on
2024-10-10	2025-01-17	Sandy Clay	2	10	2	0,25	4	0,25	0,25	Manual	NA	NA	50 seeds/row; Hand sowing

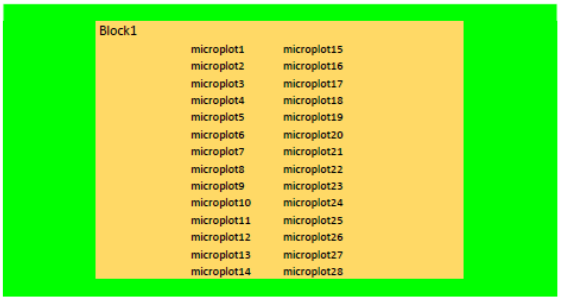
Experiment data_controlled conditions

EndDate*	RelativeHumidity	Potsize	LightDuration	LightIntensity	AverageTemperature Day	AverageTemperature Night	PhenotypicProtocol*	Remarks
Termination date of experiment in the format [yyyy-mm-dd]	[%]	cm ×cm (diameter ×depth)	[h/d]	[µmol/m2s]	[°C]	[°C]	protocols used for phenotypic data collection (manual, imaging, sensory machines etc)	free text to provide relevant info on trial, e.g. deviations from protocol, problems with scoring
2022-09-29	40	20 ×30	6	NA	22	18	Manual scale	NA
2020-07-16	60	9 ×15	12h	NA	25	17	Manual scale	

TrialID: Ecrop_trial#



TrialID: 202013



Replicate	Block	Plot
1	2	1-2-2001
1	2	1-2-2002
1	2	
1	2	
1	2	
1	2	
1	2	
1	2	
1	2	
1	2	
1		
1		
1		
1		
1		

2d_Treatment

2d_Treatment

The treatment section gives information of the treatment and control used in the study. this section is mandatory

Descriptors	Trial_ID*	TreatmentName*	TreatmentDescription*	TreatmentType	Date*	ApplicationMethodandTimepoints*	Amount	Unit	Concentration	Product	ControlDescription*	Remarks
Description	unique identifier of the trial/study/experimetn	name of the treatment applied	Detailed description of the treatment	e.g. biotic stress, abiotic stress, nutrient supplement	date of application (if frequent, date of first application)	timepoint should include how frequent the application was done e.g. 30 days after planting and 60 days after planting	The total volume or quantity of specific treatment used in the experiment	the scale of application for the treatment amount, expressed in terms of measurement scale (e.g. per hectare, per litre, per plant)	Concenration of treatment being applied. in this contest it implies the density or strength. e.g. 500ppm (hormone), 0.040% active ingredient etc)	if any commercial product was used	type of control and description of how the control was used and maintained in the experiment	
example	WP4_trial#	InsecticideTreatment	Application of Abamectin for control of fruit borer	biotic stress	2022-09-29	sprayed on leaf surface 30 days after sowing and	0.5 ml	l/ha	0.5ml/L	qualipro (1.5EC)	NA	applyed 1 day before heavy rain
example	WP4_trial#	GrowthRegulatorTreatment	Application of gibberellic acid (GA_3) to see its affect on growth and fowering	Hormonal treatment	2020-08-05	foliar spray at three leaf stage and repeated weekly for 3 weeks	800	l/ha	100ppm	NA	Control plants were treated with an equivalent amount of the solvent or carrier (e.g., distilled water) without the addition of GA_3	applyed end of the day
example	202013	Pathogen Inoculation	Inoculation of 2 strains of *P. capsici* in vitro	biotic stress	2020-03-03	inoculation after 20 days after planting	20ml	per plant	$10^{\wedge}5$ CFU/ml	NA	Negative Control (Sterile agar plug placed at inoculation site as a control)	

3_Trait information

This section provides metadata (3a_Trait metadata) and descriptive information (3b_Trait values) on the trait of interest or the trait under observation. Trait metadata focuses on the method used to describe the trait, the frequency of measurement and reference from where the method was derived or created de novo. It may comprise data of one or more plants and/or their environment.

The trait values describe how a measurement has been made. It typically takes the form of a measured characteristic of the trait (plant or environmental trait), associated to the method and unit of measurement. It has two sections, depending on the scale of observation of trait; categorial values such as rating scores when trait is measured in a predefined set of scores e.g. disease infection score (1-5) where 1 is resistant and 5 is highly susceptible or fruit size with scores small, medium or big. Another parameter is for metric scales or quantitative value of the trait for e.g. plant height, fruit weight etc.

3a_Trait metadata

3a_trait metadata

Descriptors	Trait_ID*	TraitName*	Traitontology/CropOntology	MeasurementDescription*	Unit*	Timepoints*	TraitClass	MethodReference	Remarks
Description	(unique identifier, centrally provided)		Corresponding term from trait ontology databases	brief summary of method, for additional detail refer to scoring protocol	(e.g. n for scoring scales; ASCII unit for measurements)	time points when measurements were taken (example: 30 days after plantation, 60 days after inoculation)	(e.g. agronomic, biotic stress, quality, yield)	Reference of the publication(s) describing the method for measuring the trait, if derived from pre-existing protocol. If the method is de-novo created, put the link of the publication describing this new trait.	
example	EWB_1000	1000 kernel weight (g);	NA	g; determined as the weight of 1000 grains sampled from 100% clean harvest	g	once after final harvest	yield trait	NA	optional
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	NA	IPGRI descriptor 8.2.4. Average of percentage of infected leaves per plot, symptom expression as 1-9 scores; 1= least symptoms 9= most symptoms	na	15,30,45,60 days after sowing	biotic stress trait	https://www.genebanks.org/resources/publications/descriptors-wheat/	see standard protocols in shared folder

3b_Trait values

3b_trait values

The trait values describes how a measurement has been made. It typically takes the form of a measured characteristic of the trait (plant or environmental trait), associated to the method and unit of measurement.

Trait values (rating)

Descriptors	TraitAcronym*	TraitName*	RatingScore*	Value*	Remarks
Description	(unique identifier, centrally provided)		only the below values are allowed for specific traits	(characteristic associated with that rating score	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp.*	1	no symptoms (resistant)	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	2	less than 1% leaf surface affected	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	3	less than 3% leaf surface affected	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	4	less than 5% leaf surface affected	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	5	~10% leaf surface affected	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	6	~20-30% leaf surface affected	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	7	~40-50% leaf surface affected	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	8	~60% leaf surface affected	
example	EWB_WBG	powdery mildew (*Blumeria graminis f. sp. tritici*)	9	~70% leaf surface affected	

Trait values (metric)

Descriptors	TraitAcronym*	TraitName*	MinimumValue*	MaximumValue*	Unit*	Remarks
Description	(unique identifier, centrally provided)		only values between the minimum and maximum values specified below are allowed for specific traits		unit scale of the measurement of trait values	
example	**EWB_1000**	1000 kernel weight (g);	0	100	g	g; determined as the weight of 1000 grains sampled from 100% clean harvest
example	**EWB_Y**	yield	0	10	t/ha	t ha^{-1}; Grain yield is measured by harvesting each plot and converting the weight to tons per

4_Observed values

This section provides the information of the trait values obtained in the study. The cells in light blue are to be edited or renamed according to the type of trait name or ID. There should be 3 columns or at least 2 columns per variable, one for the value, one for the date on which the trait value was recorded and one for time of record of trait value (if applicable).

4_observedvalues

Trait_ID*	Accession_ID*	Plot	Row	Column	Replicate	Microplot	Pot	Control	Remark	TRAIT_ID_1*	ObservationDate*	ObservationTime	TRAIT_ID_2*	ObservationDate*	ObservationTime
										plant_height30DAS (Days after sowing)			plant_height30DAS (Days after sowing)		
Trait identifier linked to the metadata (Ontology if applicable)	(unique identifier of germplasm accession within the project), centrally provided	continuous numbering of experimental plots	row number of a given plot in field layout	column number of a given plot in field layout	number the replicate plots for each accession in the trial			indicate accessions used as controls as "C"	free text to record any relevant observations, for example if the material is segregating for a specific trait, details can be recorded here. 4000 char max.	Value or description of the trait observed	[yyyy-mm-dd]	hh:mm:ss	Value or description of the trait observed	[yyyy-mm-dd]	hh:mm:ss
202013	13PT6	4	4	1	1	NA	NA			80,00	2024-05-26	12:30:00	154,00	2024-06-11	
202013	14PT1	5	5	1	1	NA	NA			75,00	2024-06-03	12:50:00	155,00	2024-06-11	
201920	14PT149	6	6	1	1	NA	NA			69,00	2024-06-04	14:10:00	149,00	2024-06-11	
202013	14PT164	7	7	1	1	NA	NA			72,00	2024-05-29	11:15:00	154,00	2024-06-11	
202013	14PT17	8	8	1	1	NA	NA			66,00	2024-06-03	16:30:00	154,00	2024-06-11	
201920	14PT188	9	9	1	1	NA	NA			82,00	2024-06-01	09:35:00	155,00	2024-06-11	
201917	14PT201	10	10	1	1	NA	NA			77,00	2024-06-01	18:30:00	149,00	2024-06-11	
201920	14PT22	11	11	1	1	NA	NA			75,00	2024-05-31	09:55:00	154,00	2024-06-11	
201917	14PT23	12	12	1	1	NA	NA			72,00	2024-06-01	10:37:00	154,00	2024-06-11	
201917	14PT28	13	13	1	1	NA	NA			72,00	2024-05-30	11:15:00	155,00	2024-06-11	
201920	14PT29	14	14	1	1	NA	NA			79,00	2024-06-01	12:00:00	154,00	2024-06-11	
201920	14PT355	15	15	1	1	NA	NA			70,00	2024-06-03	11:30:00	155,00	2024-06-11	
201920	14PT358	16	16	1	1	NA	NA			71,00	2024-05-29	11:45:00	149,00	2024-06-11	
201920	14PT383	17	17	1	1	NA	NA			77,00	2024-06-04	11:50:00	154,00	2024-06-11	

5_complementary information

5_complementary

Descriptors	WeatherDataSource	SoilCharacteristics	DataCurator	Versioning	DataAccessPolicy	QualityControlMeasures
description	*Source of weather data provided as a link to data base or website*	*Additional property of soil (pH,texture, type) present/used in the experiment*	*contact of Person or team who curated the data*	*Version number of dataset*	*Conditions for accessing or using the data*	
examples	https://www.worldclim.org/data/monthlywth.html		jane.doe@research.org	6	open access	Proper calibration and removal of outliers