PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Funded by the European Union

Minimum Information about a Plant Genetic Resource (MI-PGR)

Catherine Hazel M. Aguilar

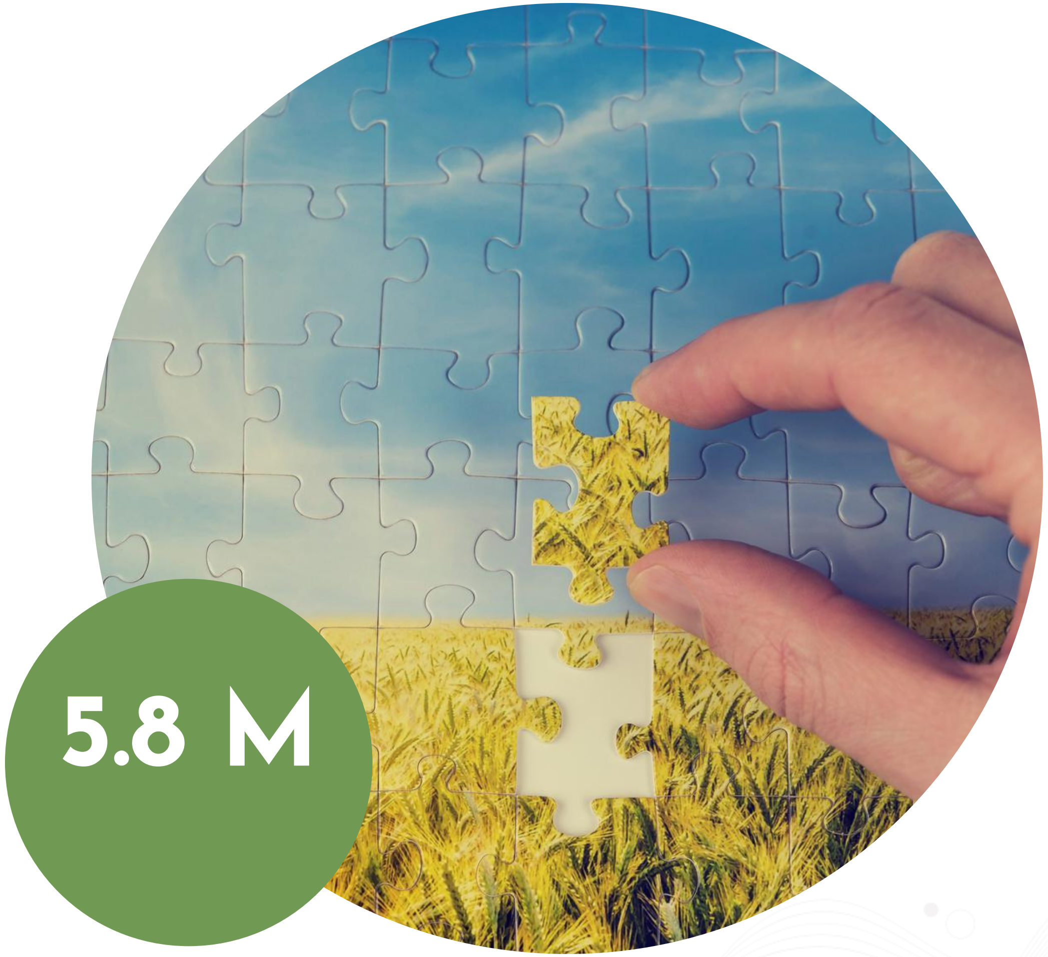
Markus Oppermann, Anne-Françoise Adam-Blondon, Joanna Magos Brehm, Nigel Maxted, Jaime Prohens, Antonio Granell, Clara Pons, Theo van Hintum, Lorenzo Maggioni, Paul Shaw, Bela Bartha, Vojtech Holubec, Luigi Guarino, Giovanni Giuliano, Stephan Weise

08 October 2025

Millions of accessions stored in over 800 genebanks in 115 countries (FAO, 2025)
https://doi.org/10.4060/cd4711en

Why do genebank collections remain largely untapped?

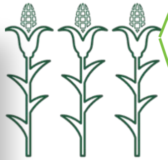
How can we better translate genebank conservation into actionable agricultural solutions?



Bridging the gap between conservation and utilization

Ancillary Data
Legal and access
Material management and monitoring
Traditional knowledge

Basic Accession Data
Passport Data
Image Data



Trait Measures

Manual phenotyping/ Traditional trait measures

High-throughput Phenotyping Data

Image Data

Molecular Phenotype

Metabolomic Data
(metabolite profiles, pathways)

Proteomic Data
(Protein expression, post-translational modifications)

Transcriptomic Data
(Gene expression, non-coding RNAs, alternative splicing, small RNAs

Epigenome
Genome

Whole-genome sequences

Genetic markers

Genetic variation Data
(Structural variants, Indels, copy number variations, haplotypes)

Epigenomic Data
(DNA methylation, chromatin conformation /accessibility, histone modifications)

Environmental Data

“Minimum Information” concept

Specify the smallest complete set of fields that must accompany a dataset, experiment, or object so that:

- The data can be discovered and interpreted
- Results can be reproduced or verified.
- Datasets can be integrated with others using shared semantics.
- Downstream users can make valid decisions and inferences.

Minimum-information standards emerged because high-throughput experimental data were accumulating faster than the ability to document and reuse them

Existing Data Standards

Building on established frameworks

MCPD v2.1

Multi-Crop Passport Descriptors

FAO/Bioversity - Passport data standardization

Darwin Core

Biodiversity Information Standard

TDWG - Biological specimen data

MIAPPE v1.1

Plant Phenotyping Experiment

Phenotyping data standardization

MIxS

Minimum Information about any Sequence

GSC - Genomic data standards

Issue: Independent development → Different terminologies, structures, and formats → Integration challenges

Features of MI-PGR

Key design principles

Hierarchical Levels

Six progressive levels from basic identification to comprehensive genomic data

Integration

Built upon and harmonizes existing standards (MCPD, MIAPPE, MIxS, DwC)

Scalability

Accommodates varying institutional capacities and data complexity

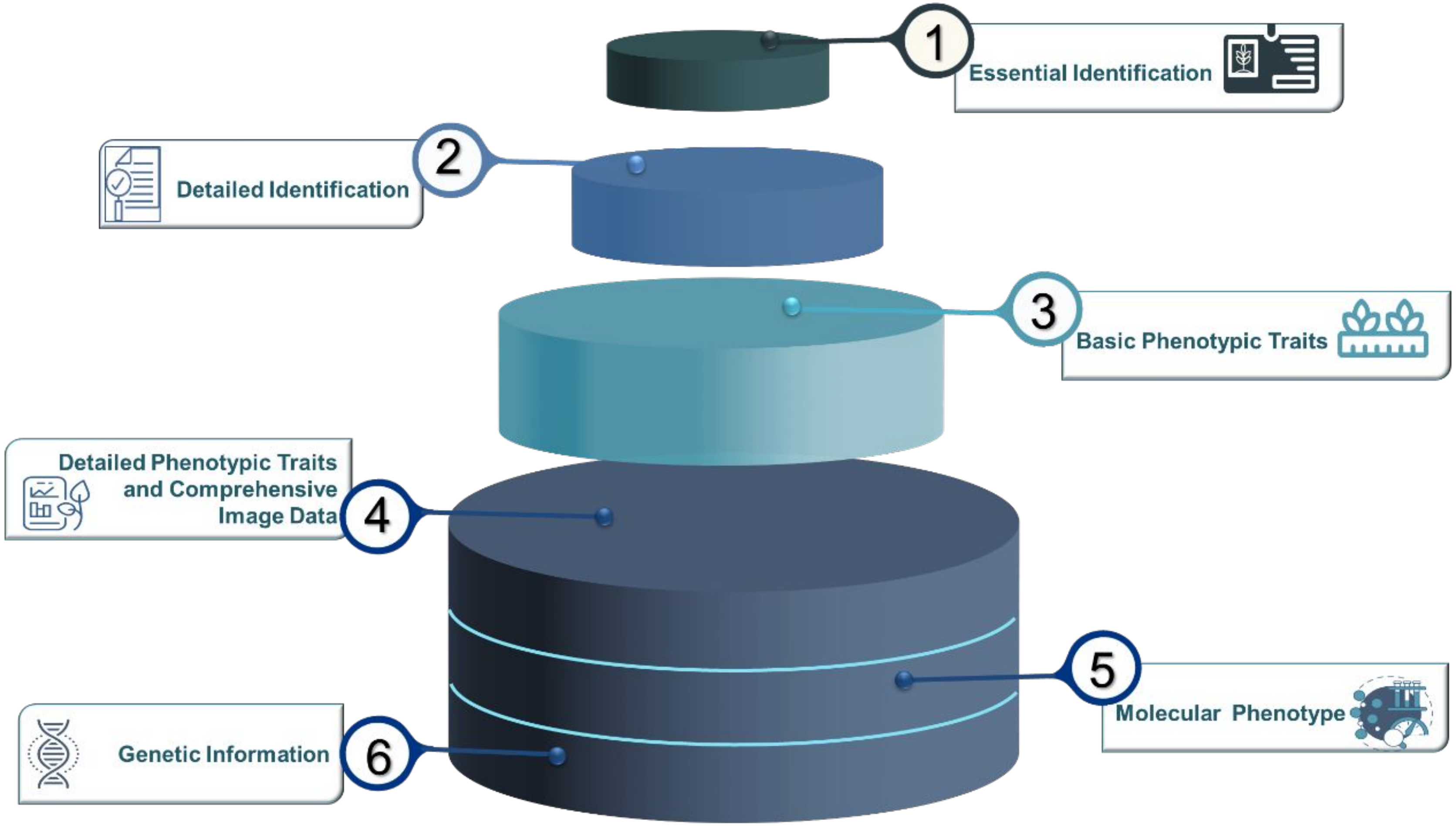
FAIR Principles

Ensures data is Findable, Accessible, Interoperable, and Reusable

FAIRness at all costs? Practical limitations in applying the FAIR principles to plant genetic resources data. Under review.

Minimum Information about a Plant Genetic Resource (MI-PGR)

Basic passport data; 4 mandatory data points based on MCPD v2.1 (*ex situ*) and CWRI v.1 *(in situ* CWR); +1 new data element**, TAXONID**



Taxonomic Ambiguity and Evolution: The Case for TaxonID

Historical names (legacy labels)

Local synonyms / vernacular labels

Different taxonomic references

Effects:

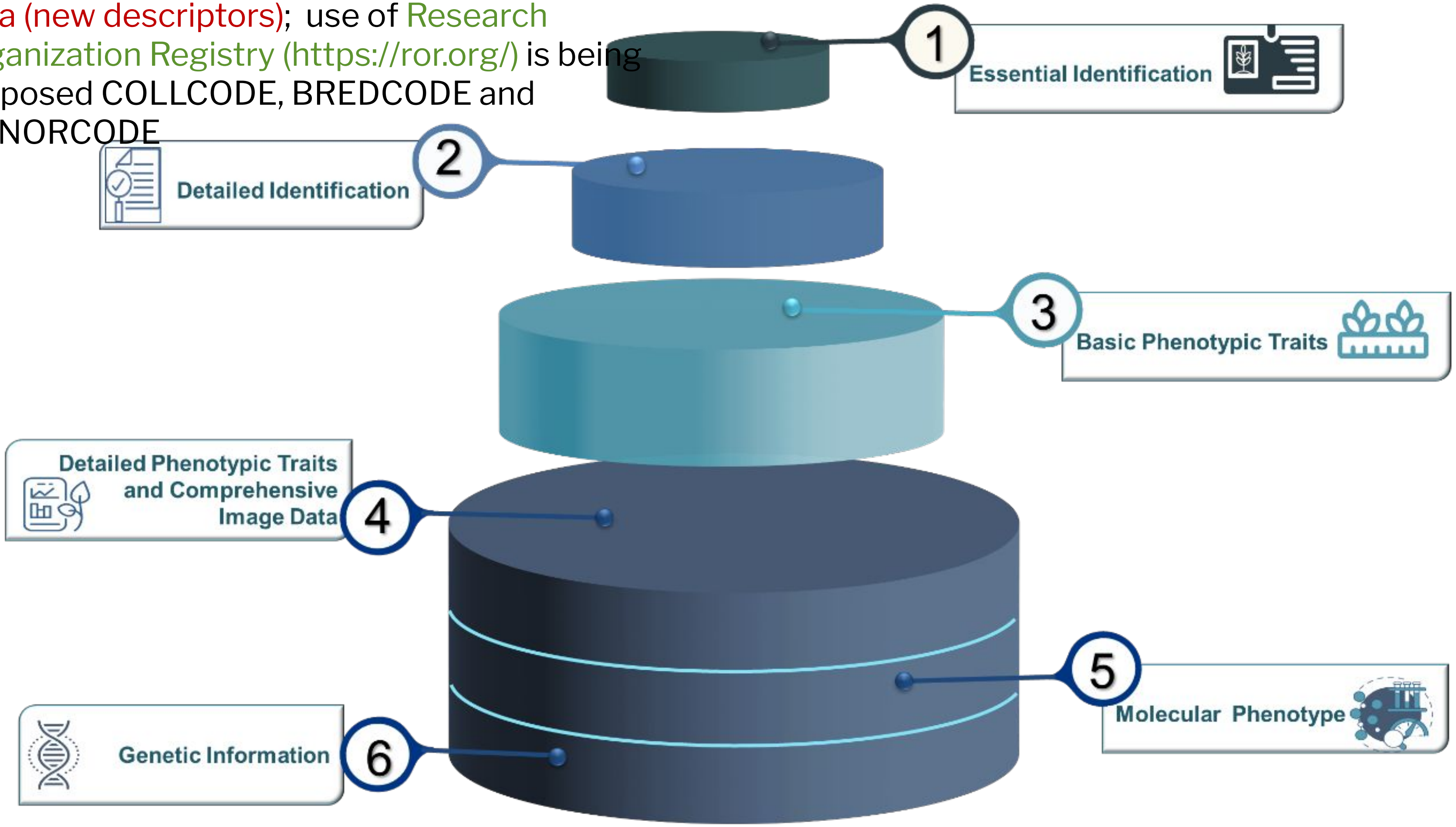
- Same species under multiple names
- Conflated or split taxa inconsistently
- Harder cross-dataset matching/analytics

Proposed Solution:

- Use attribute-value pairs to store multiple TaxonIDs (e.g., TaxonID_Source: NCBI, TaxonID_Value: 4565)
- Link each identifier to its source database
- Enable seamless cross-referencing between taxonomic systems

Minimum Information about a Plant Genetic Resource (MI-PGR)

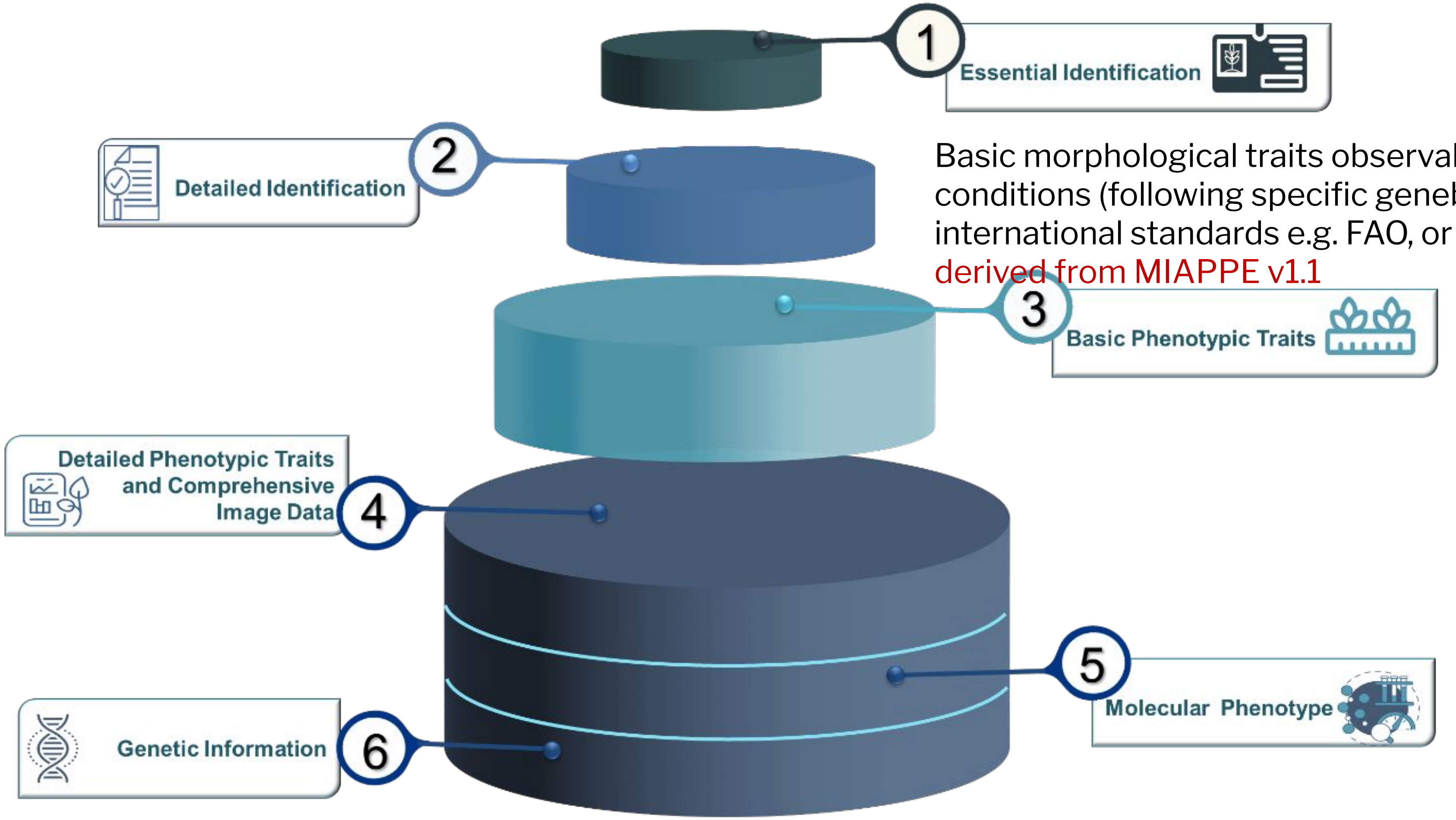
Comprehensive passport data with initial image data (new descriptors); use of Research Organization Registry (https://ror.org/) is being proposed COLLCODE, BREDCODE and DONORCODE



Ir

31. *Material Type (MATYPE)	Type of specimen/ Nature of the plant material 10) Seeds 20) Leaves 30) Stems 40) Roots 50) Flowers 51) Male Flower 52) Female Flower 53) Hermaphroditic 54) Inflorescences 60) Fruits 61) Unripe 62) Ripe 63) Dried 70) Whole Plant 71) Seedling 72) Mature Plant 80) Plant tissue culture 90) Planting material 91) Bulbs 92) Rhizomes 93) Corms 94) Stolons 95) Tubers **99) Other** (elaborate in IMAGEDESCRIPT field)	20		preparations	preparationsText
32. *Record Creator (RECREATOR)	Person/entity responsible for collection/observation for Origin data tracking and source attribution (Recommendation: ORCID if known)	ORCID: 0000-0002-1825-0097			recordedBy
33. *Description of Image Content (IMAGEDESCRIPT)	Brief description of what the image portrays.	Lanceolate leaf, 15 cm length, 4 cm width, with acuminate apex, cuneate base, and serrated margins.			Description

Minimum Information about a Plant Genetic Resource (MI-PGR)



Basic morphological traits observable under standard conditions (following specific genebank protocols or international standards e.g. FAO, or specific crop consortia); derived from MIAPPE v1.1

Minimum Information about a Plant Genetic Resource (MI-PGR)

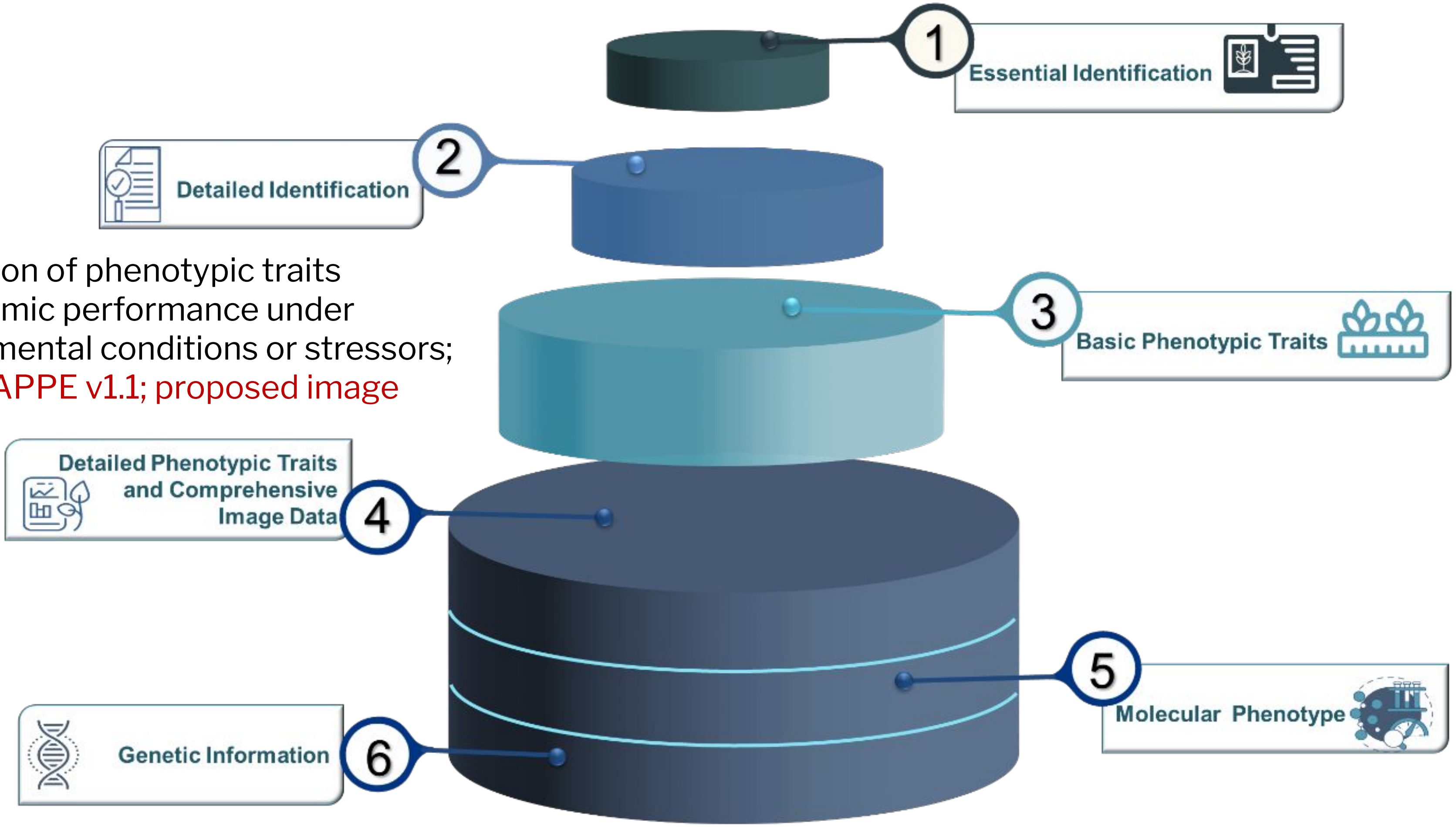
Level 3

focuses on capturing specific data elements about the physical and observable characteristics of a PGR accession under standardized conditions (i.e. **traditional morphological characterization** following specific genebank protocols, international standards by FAO or specific crop consortia).

Simplified MIAPPE

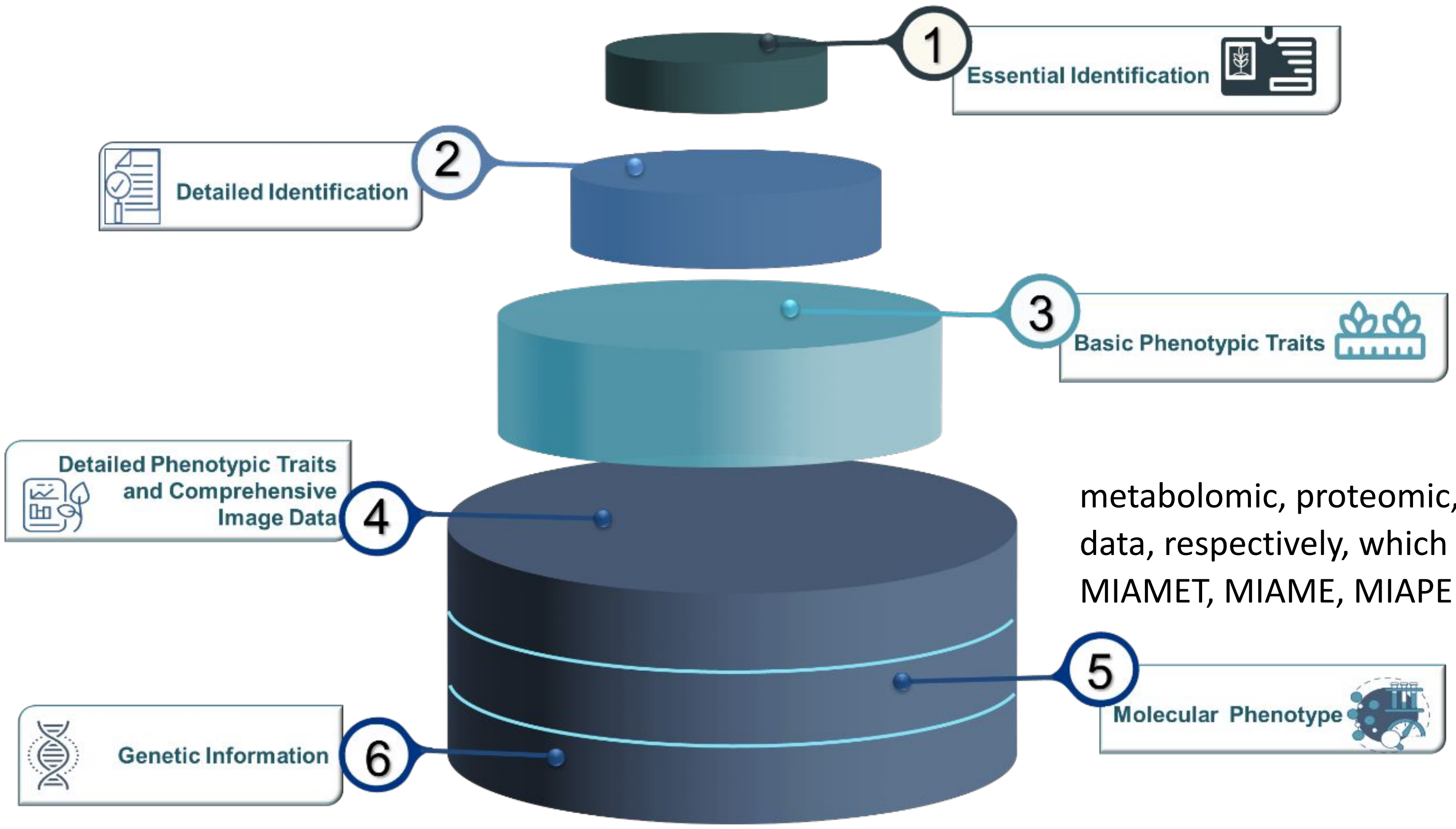
- Study
- Data File
- Biological material (level 1 and 2)
- Environment
- Observation unit
- Observed variable

Minimum Information about a Plant Genetic Resource (MI-PGR)



Detailed evaluation of phenotypic traits including agronomic performance under specific environmental conditions or stressors; derived from MIAPPE v1.1; proposed image data descriptors

Minimum Information about a Plant Genetic Resource (MI-PGR)



metabolomic, proteomic, and transcriptomic data, respectively, which were built upon MIAMET, MIAME, MIAPE and MINSEQE.

Minimum Information about a Plant Genetic Resource (MI-PGR)

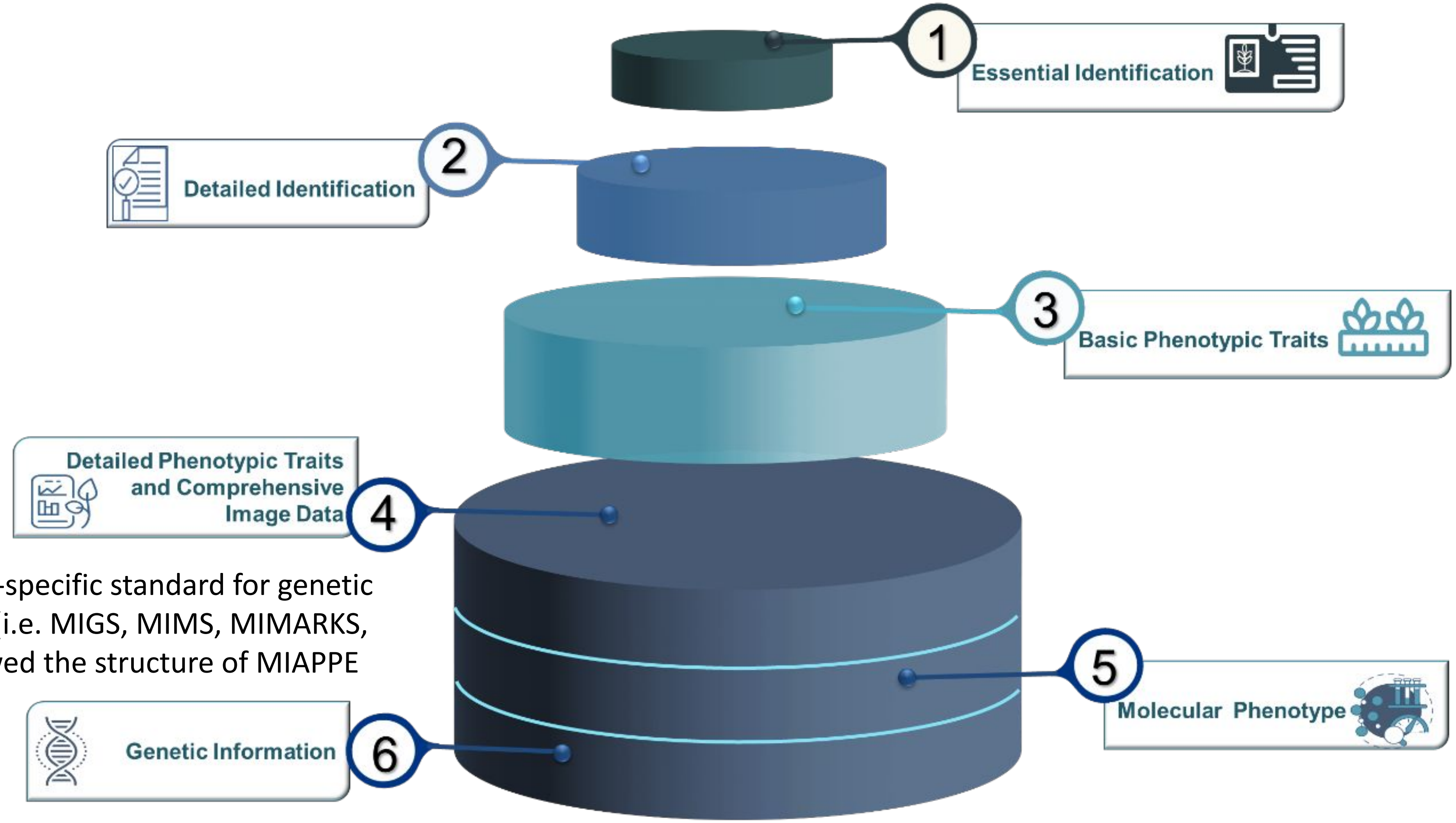
Level 5: Molecular Phenotype (Transcriptomic, Metabolomic, Proteomic Data

Common metadata requirements following ISA-tab structure (Investigation, Study, Assay)

Core descriptors for metabolomic data

Descriptor	Description	Example	Crosswalk Equivalents	Recommendations/ Notes
1. Tissue type	Biological material (tissue or organ) used for metabolite extraction.	Leaf tissue	MSI: Sample type	Plant ontology
2. Sample Preparation	Details on how samples were prepared (extraction method, solvents used, storage conditions).	Methanol extraction at -80°C	MSI: Sample Preparation	
3. Chromatography Details	Description of the chromatography method, if applicable (e.g., column, solvent system).		MSI: Chromatography	
4. Data type	Type of data generated from the instrument.	GC-MS, LC-MS		Standardize data in formats like mzML or NetCDF.
5. Metabolite Quantification	Method for quantifying metabolites (e.g., absolute or relative quantification).		MSI: Quantification	Specify the method (e.g., peak areas, internal standards).
6. Calibration Method	Instrument calibration method used (internal or external standards).		MSI: Calibration	Include reference standards or external calibration methods.
7. Quality Control (QC)	Measures taken for quality control (e.g., blanks, pooled samples).		MSI: Quality Control	Include details of QC samples, pooled standards, and blanks.

Minimum Information about a Plant Genetic Resource (MI-PGR)



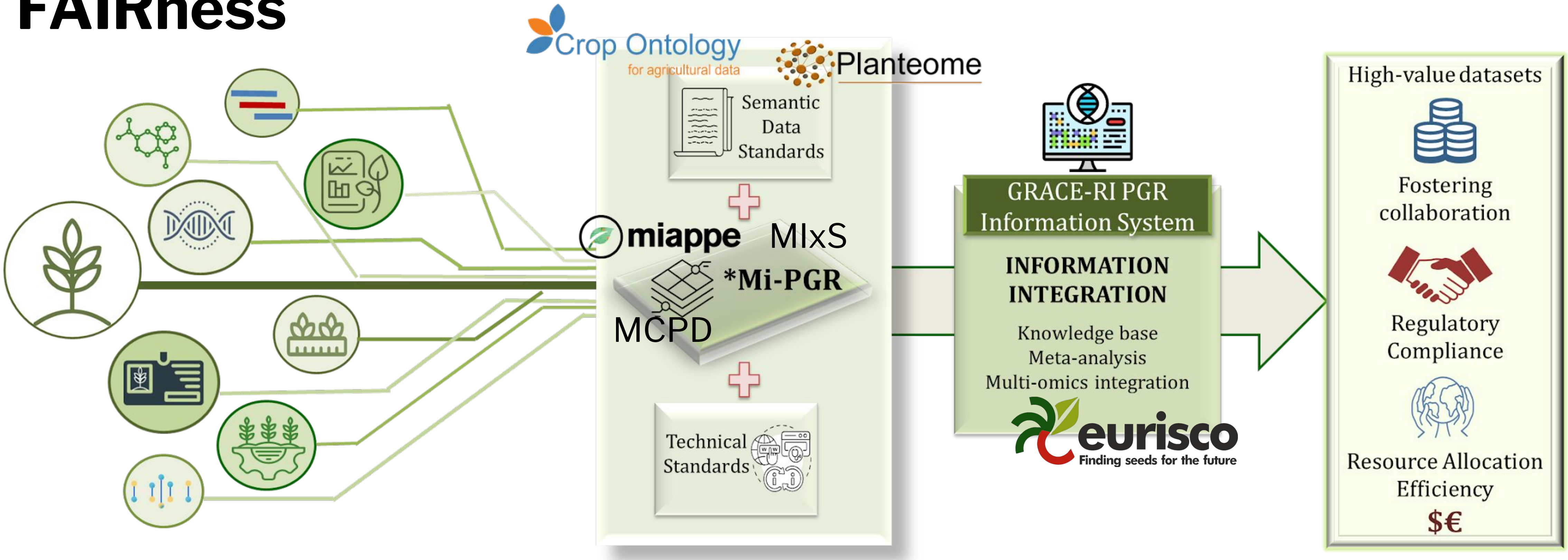
derived from the GnpIS-specific standard for genetic marker data, and MIxS (i.e. MIGS, MIMS, MIMARKS, MISAG, MIMAG); followed the structure of MIAPPE

Minimum Information about a Plant Genetic Resource (MI-PGR)

Level 6: Genetic Information

- The structure of this level follows the framework of MIAPPE (sections covering Investigation, Study, Person, Biological Material, Environment, and specific parts dedicated to Genotyping/Genetic Markers and Genomic Sequences)
- Data elements and terminologies for metadata are aligned with MIAPPE and MCPD wherever applicable.
- The required data elements specific to a dataset type are derived from the GnpIS-specific standard for genetic marker data, and MIxS (i.e. MIGS, MIMS, MIMARKS, MISAG, MIMAG) for genome sequencing data.

PGR- associated Data: Progressing Towards FAIRness



*At present, a proposed integrative framework/ harmonized minimum information checklist

Implementation Considerations

Key factors for successful adoption

✓ Flexibility

Accommodates varying institutional capacities and resources

⚠ Training

Requires capacity building and educational resources

✓ Scalability

Start with Level 1, progressively add detail

⚠ Tools

Need for digital tools and templates

✓ Compatibility

Builds on existing standards and systems

⚠ Legacy Data

Challenges in standardizing historical data

Way Forward

Stakeholder Engagement

Consultations with conservationists, researchers, breeders, data managers

Collaboration

Partner with EMPHASIS, ELIXIR, DiSSCo, and other RIs

Validation Studies

Proof of concept testing in real-world settings

Digital Tools

Development of templates, tutorials, and implementation guides

Expected Impact

Enhanced Data Quality

- Standardized documentation
- Reduced errors and inconsistencies
- Improved completeness

Improved Interoperability

- Seamless data exchange
- Cross-platform integration
- Better collaboration

Accelerated Research

- Easier data discovery
- Faster meta-analyses
- Enhanced reproducibility

Better Conservation

- Informed decision-making
- Reduced redundancy
- Strategic resource use

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Funded by the European Union

THANK YOU