

Rethinking Plant Genetic Resources Documentation in the Age of Data-Driven Science



Catherine Hazel M. Aguilar

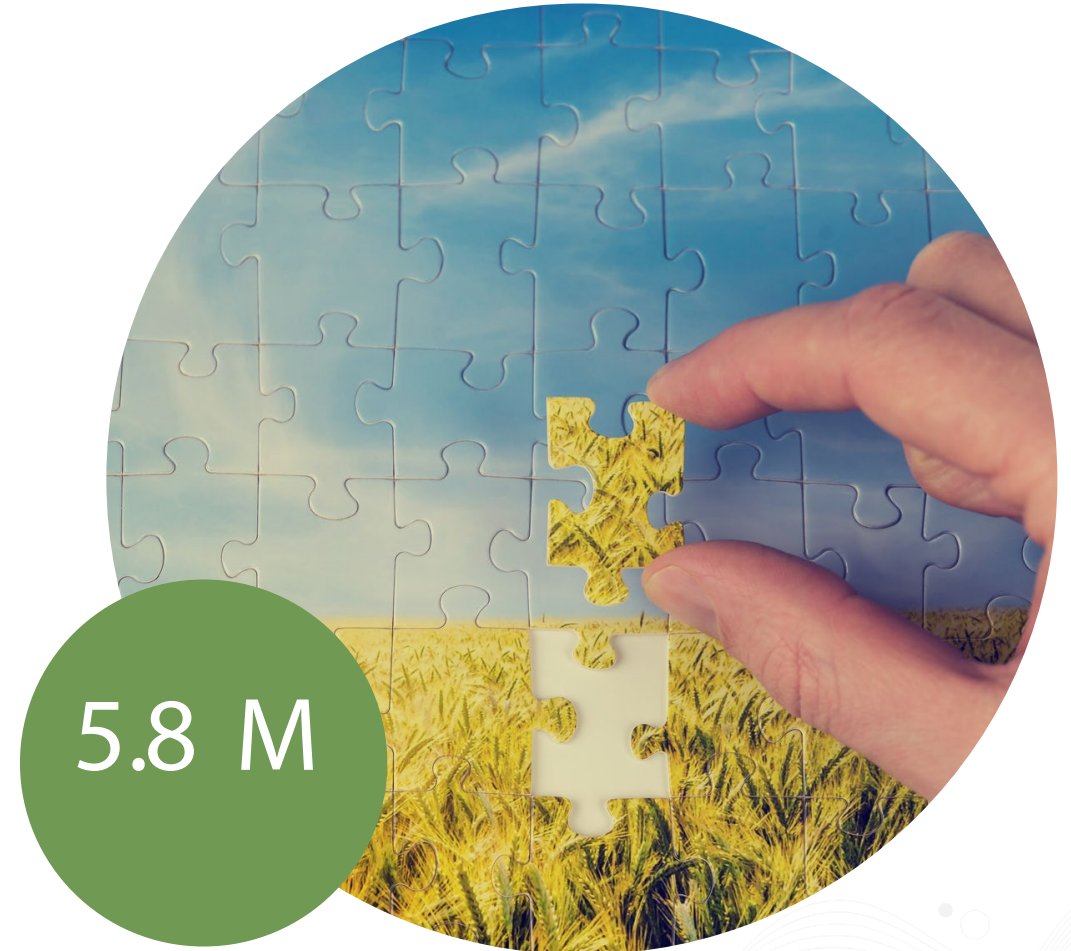
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)

Joint PRO-GRACE/EMPHASIS policy symposium and
workshop about plant genetic resources and phenotyping
28 June 2024

Millions of accessions stored in over 800 genebanks in 115 countries (Hanson et al., 2024)

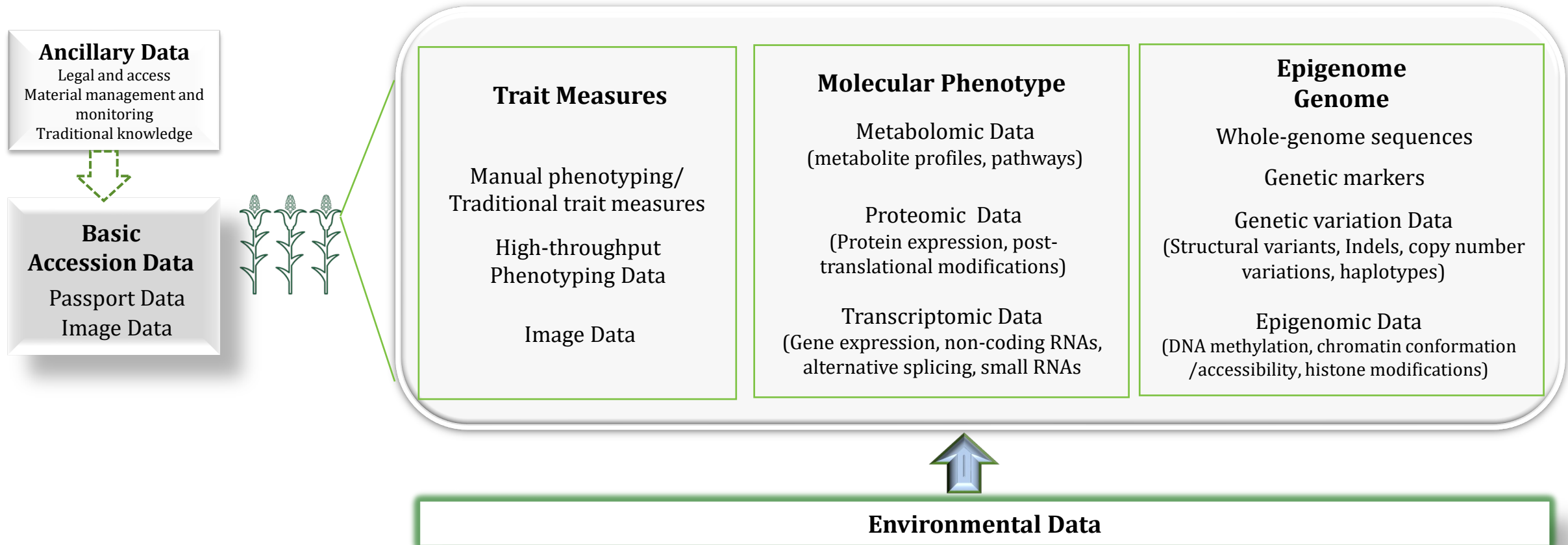
Why do genebank collections remain largely untapped?

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



5.8 M

Bridging the gap between conservation and utilization



The Complexity of Plant Phenotyping

High costs
and resource
intensity

Specialized
phenotyping
requirements

Heterogeneity of
crops, traits,
experimental
contexts,
environment

Extensive,
complex
datasets

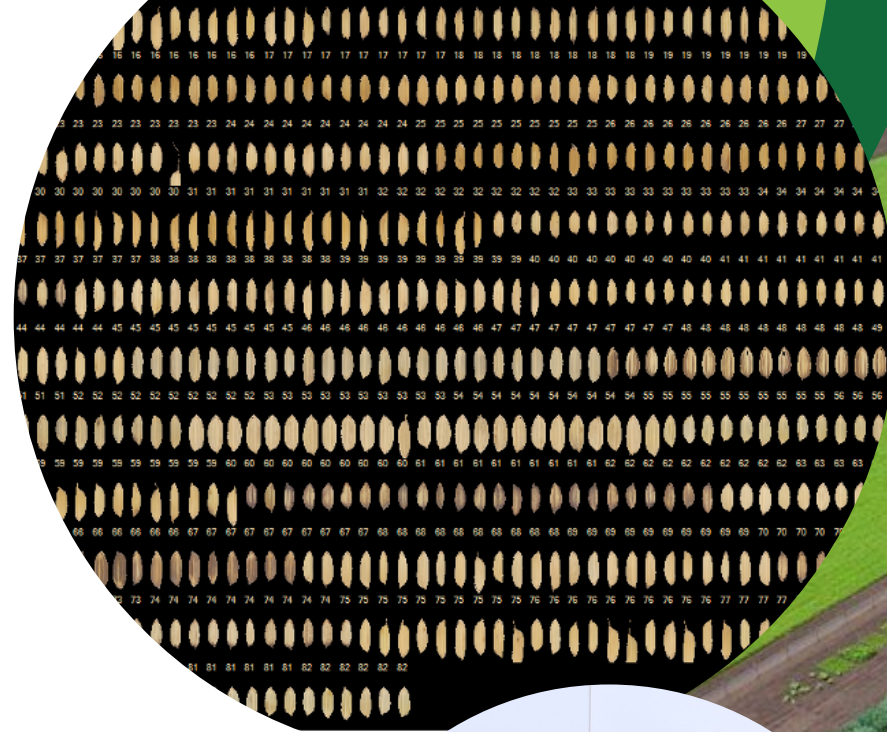
Reproducibility
and accuracy



The Complexity of Plant Phenotyping

Large number of genotypes

- Heterozygous and heterogeneous accessions
- Homozygous and heterogeneous accessions
- Heterozygous and homogeneous accessions
- Homozygous and homogeneous accessions



The Complexity of Plant Phenotyping

Different crops, diverse range of traits and growth patterns





The Complexity of Plant Phenotyping

Different interactions

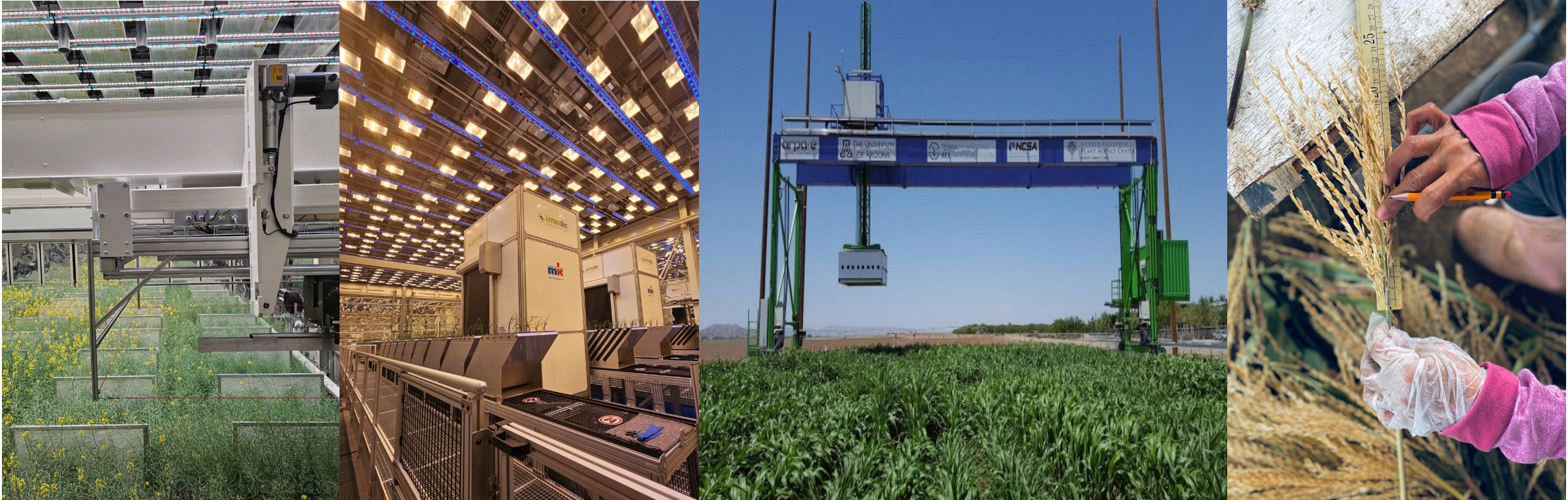
The Complexity of Plant Phenotyping

Different scales



The Complexity of Plant Phenotyping

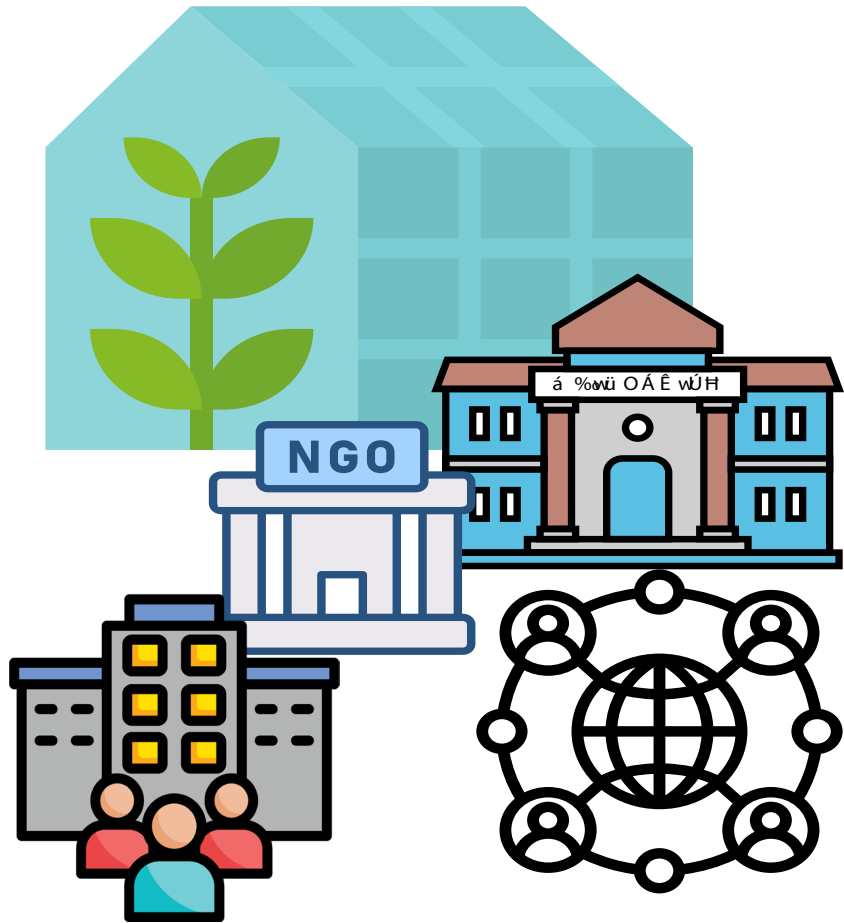
Contextual Variability



IPK PhenoSphere

Newcomb and Shakoor, 2022

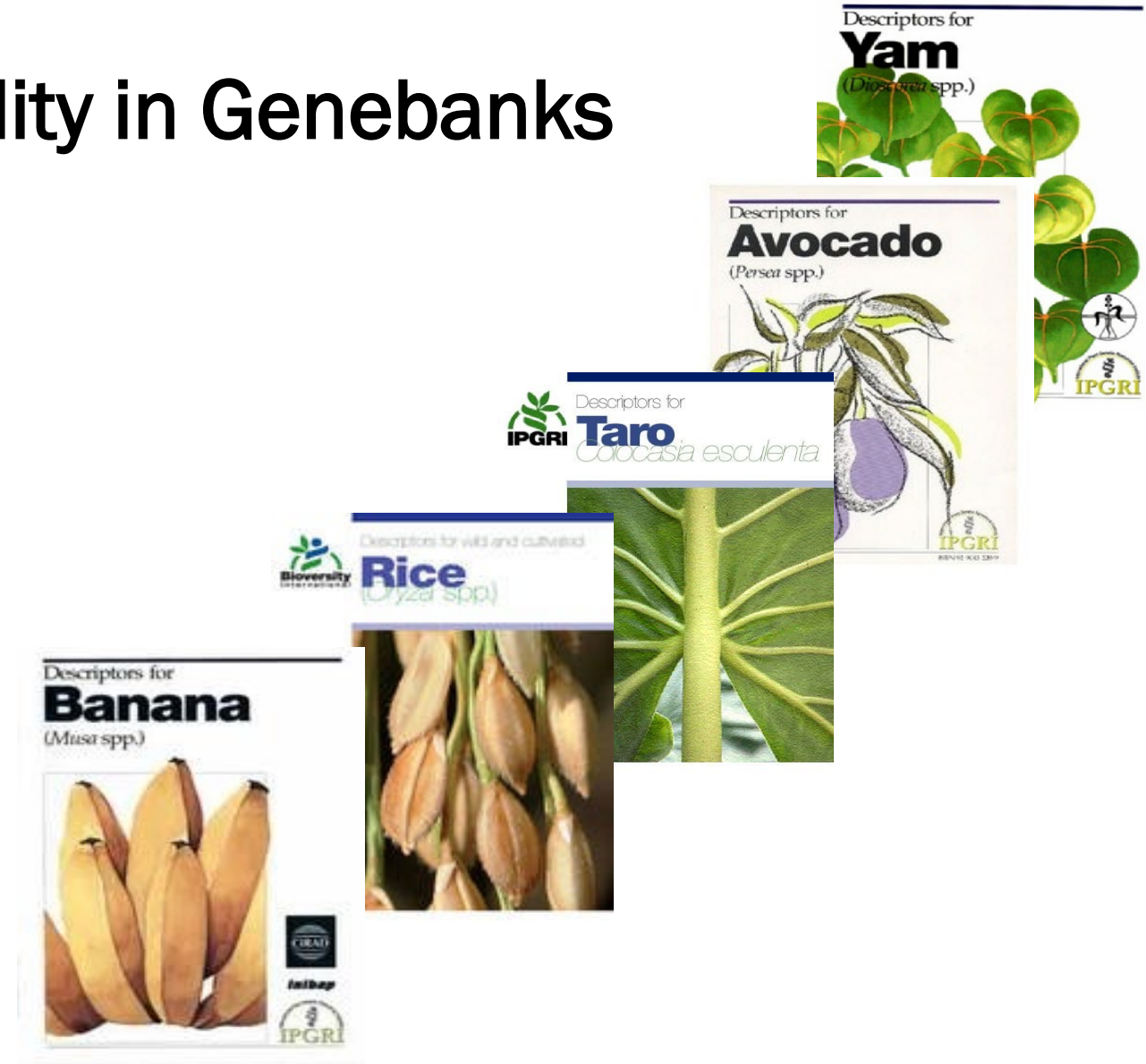
The Complexity of Plant Phenotyping





**Different stakeholders and
diverse data contributors**

Phenotypic Data: Today's reality in Genebanks

- Crop-specific descriptors
- Varied data collection methodologies
- Variable data reporting and exchange formats
- Variable naming and semantics
- Different information management systems
- A multitude of historical datasets



Historical datasets

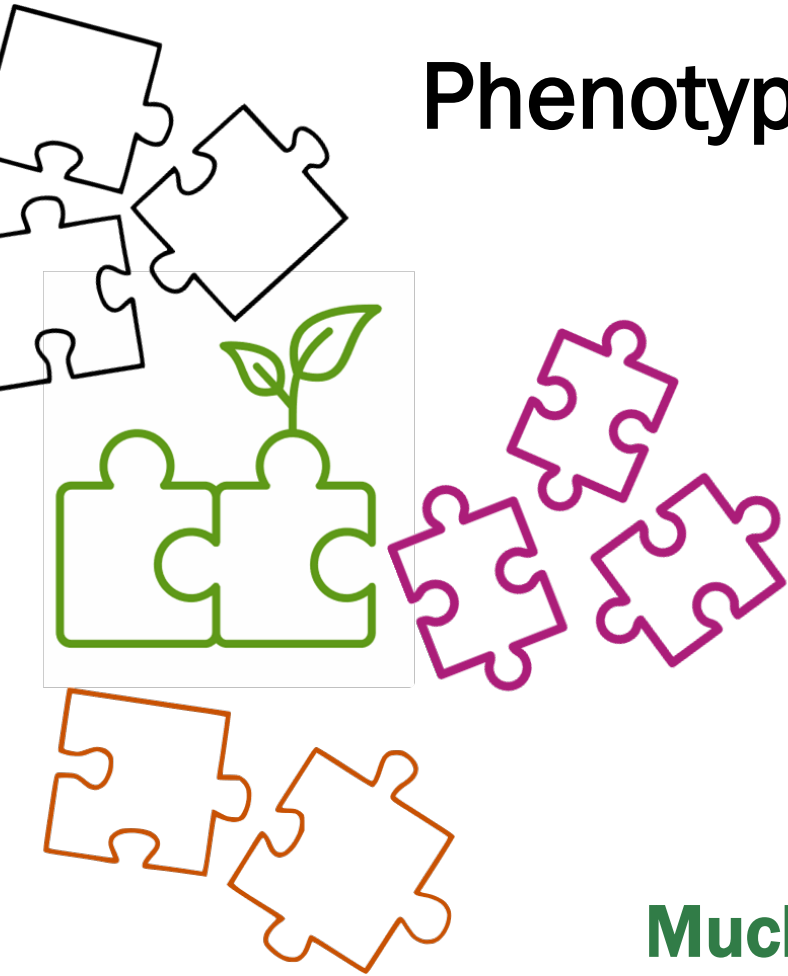
Home | Filter experiments

- Home
- Passport data
- C&E data
- C&E data
- Experiment search
- Filter by trait
- Export data
- Statistics & documents
- About
- News
- Newsletter subscription
- Intranet ex situ
- Intranet in situ CWR
- Imprint
- Data protection policy

1 - 10 of 134

Experiment description	Dataset remark	Experiment start year	Experiment end year	Details
Scoring of barley accessions 1991 - 1992.	Characterisation data (1946 - 2012) of barley accessions from DEU146	1991	1992	contained traits
experiment name: LOL_ESP99_EVA2004.Elicitation of evaluation data.	Evaluation data of DEU271	2004	2006	contained traits
experiment name: LOL_HRV96-97_pr-eva2003.Elicitation of evaluation data.	Evaluation data of DEU271	2003	2005	contained traits
experiment name: LOL_IRL2002_EVA2008.Elicitation of evaluation data.	Evaluation data of DEU271	2008	2009	contained traits
experiment name: LOL_BGR98_pr-eva2002.Elicitation of evaluation data.	Evaluation data of DEU271	2002	2004	contained traits
C-Daten POA Vor-Projekt_SZS	This dataset contains C&E data of Poa pratensis	2000	2002	contained traits
C-Daten POA Vor-Projekt_NPZ	This dataset contains C&E data of Poa pratensis	2000	2002	contained traits
Scoring of barley accessions 1945 - 1946.	Characterisation data (1946 - 2012) of barley accessions from DEU146	1945	1946	contained traits
Scoring of barley accessions 1946 - 1947.	Characterisation data (1946 - 2012) of barley accessions from DEU146	1946	1947	contained traits
Scoring of barley accessions 1947 - 1948.	Characterisation data (1946 - 2012) of barley accessions from DEU146	1947	1948	contained traits

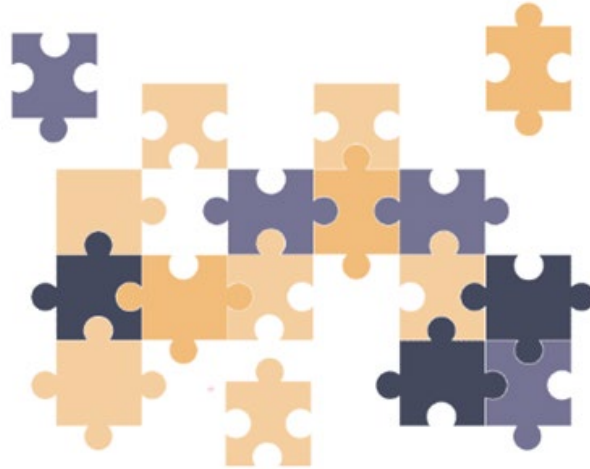
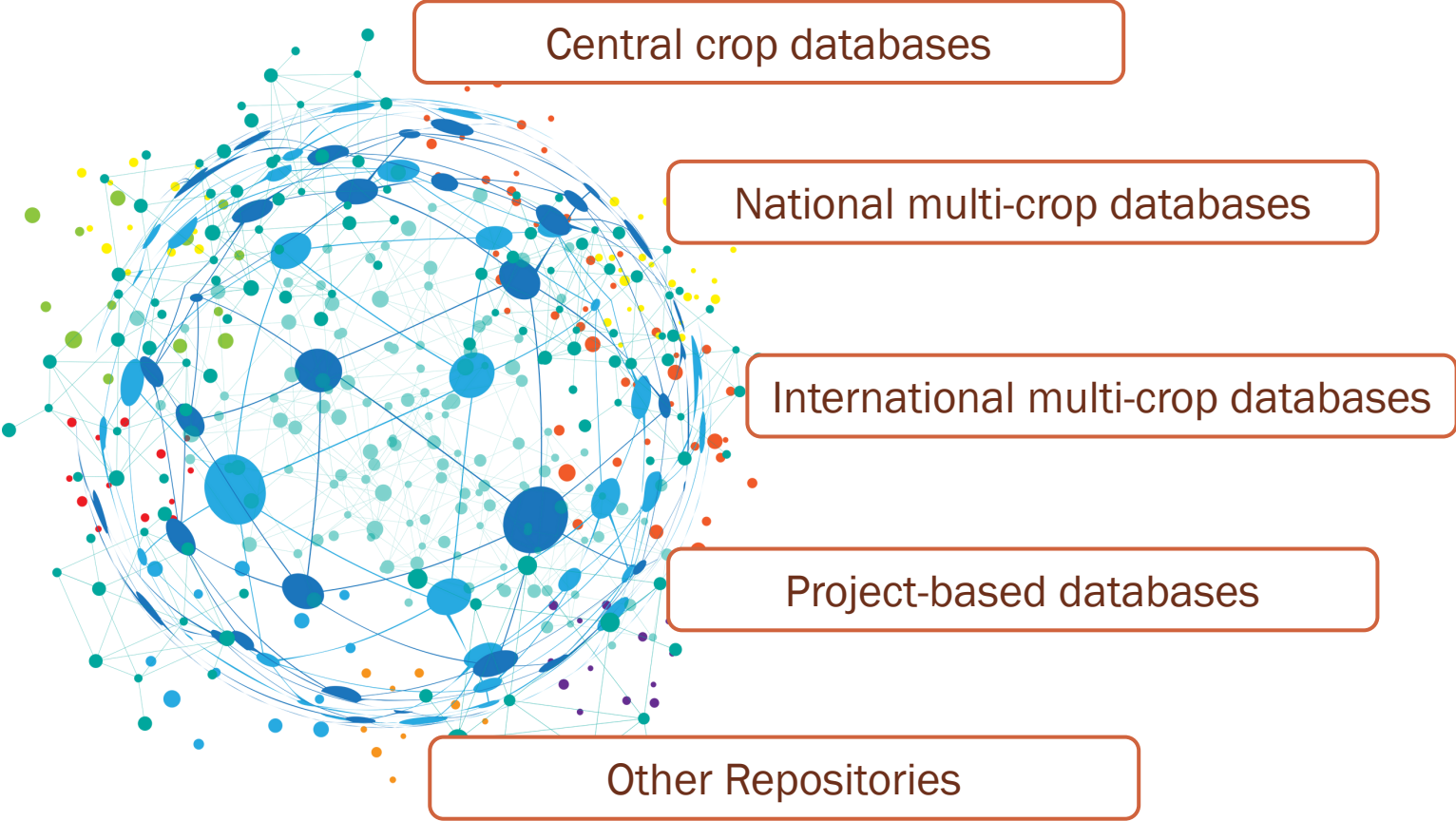
Phenotypic Data: Today's reality

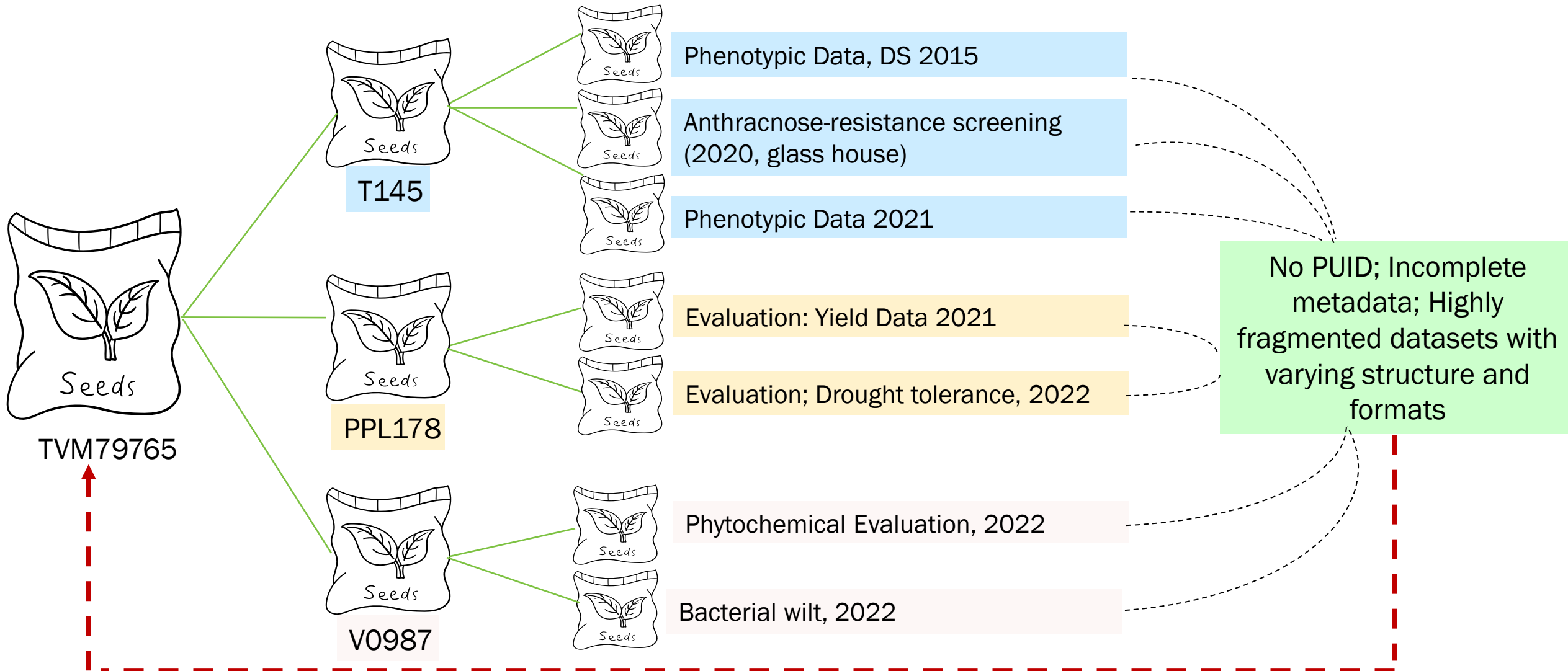


- Ambiguous IDs
- Data fragmentation
- Compromised and incomplete (meta) datasets
- High volume and complexity
- Limited access and availability

Much of the valuable data generated remains underutilized

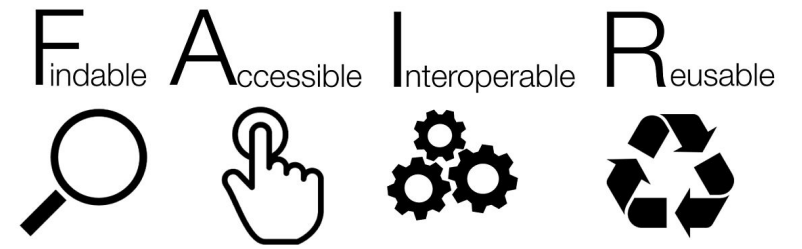
Data Fragmentation





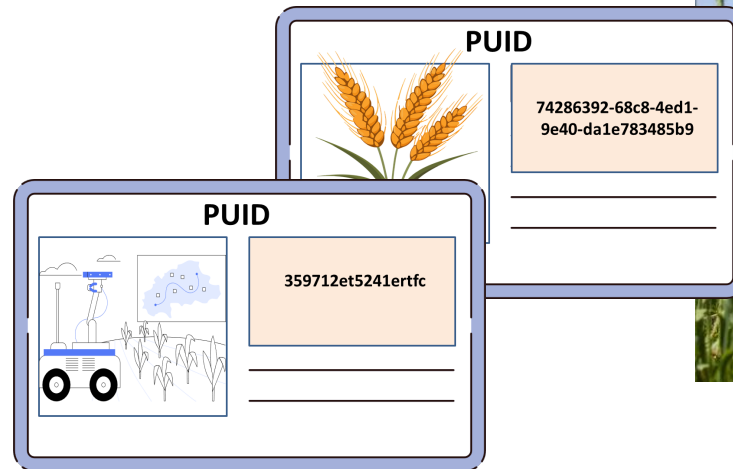
How will you integrate these diverse datasets to accurately attribute them to the specific accession?

Global Push for Data FAIRness



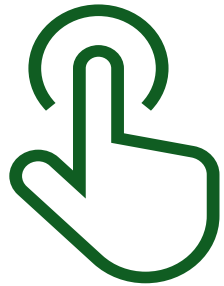
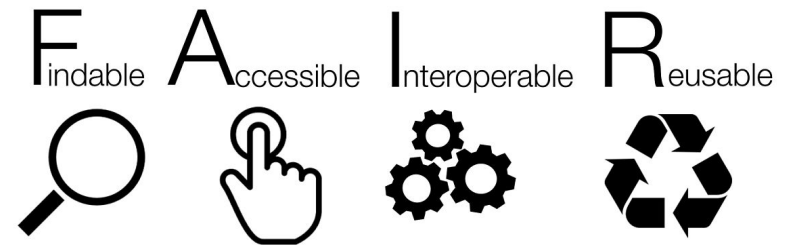
Findable

Unique & persistent identifiers (e.g. DOI)
Rich metadata; Clear data-metadata links
Searchable



<https://www.rothamsted.ac.uk/>

Global Push for Data FAIRness

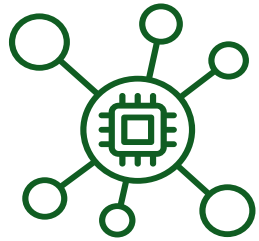
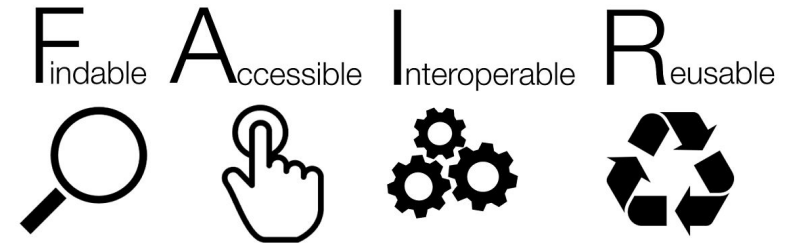


Accessible

Retrievable using open and standardized protocols; allows authorization when needed
License rights



Global Push for Data FAIRness

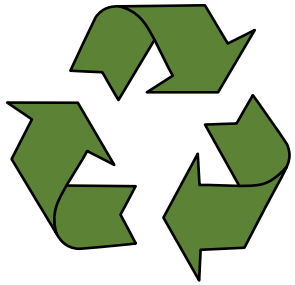
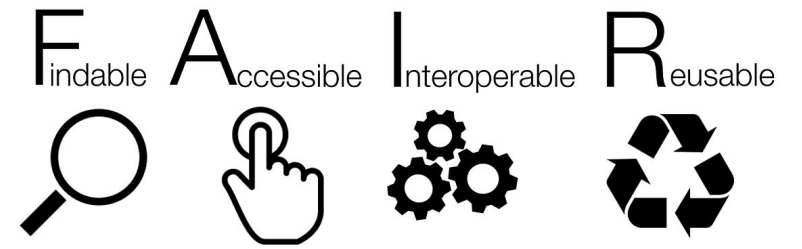


Interoperable

Consistent vocabularies
Formal, accessible, shared and
applicable language
Shared standardized formats

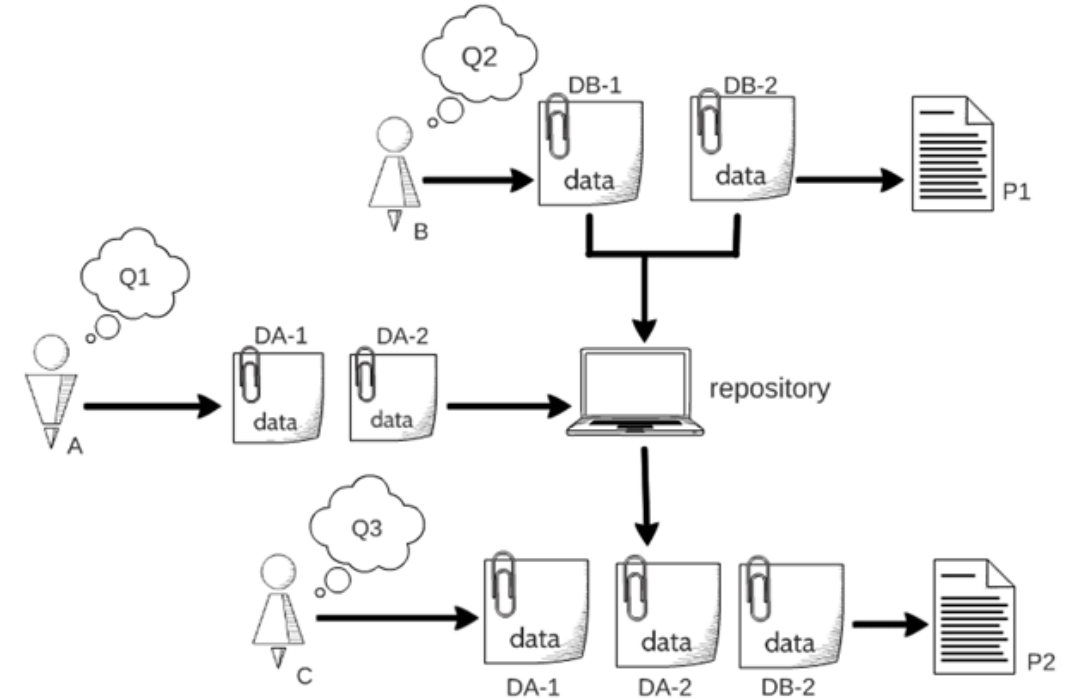


Global Push for Data FAIRness



Reusable

- Detailed provenance
- Rich descriptions with accurate and relevant metadata
- Meet domain-relevant standards



(Sandt et al., 2019)

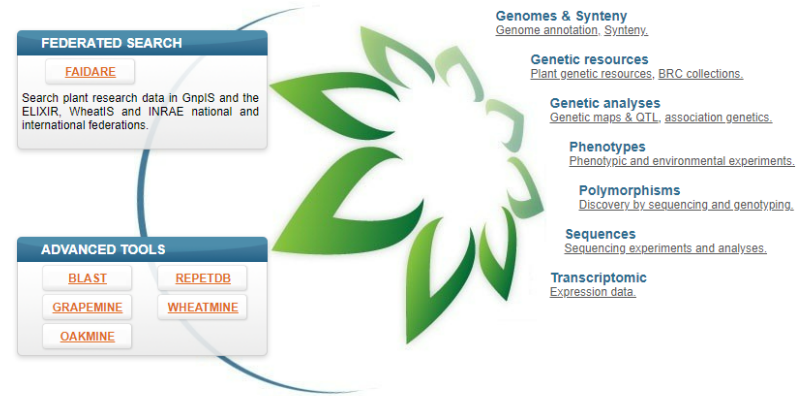
Phenotypic Data: Progressing Towards FAIRness



Shaw et al., 2017



König et al., 2020



Pommier . et al. (2019)



Neveu et al. (2019)



EMPHASIS

<https://emphasis.plant-phenotyping.eu>

EUROPEAN INFRASTRUCTURE FOR PLANT PHENOTYPING

Phenotypic Data: Progressing Towards FAIRness

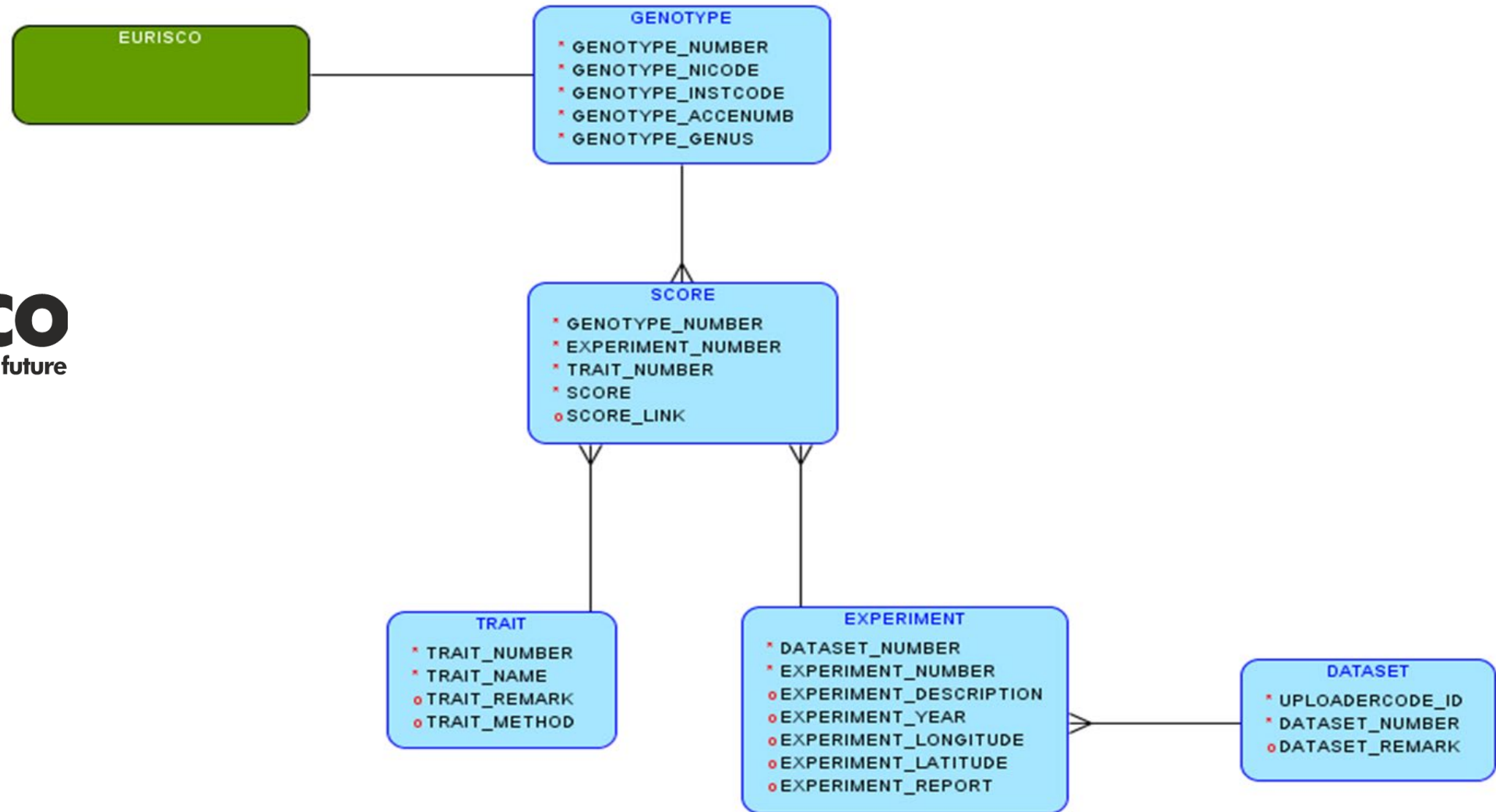


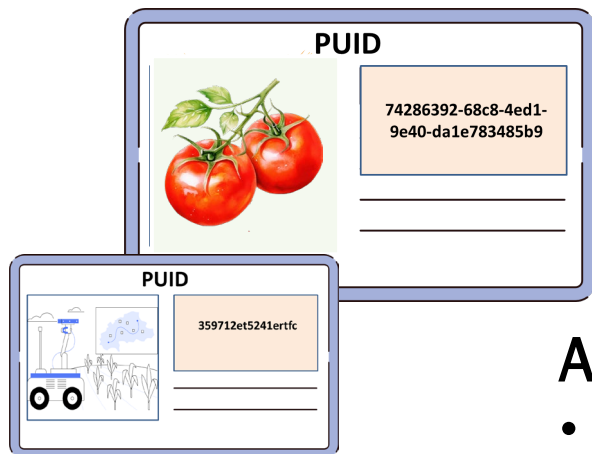
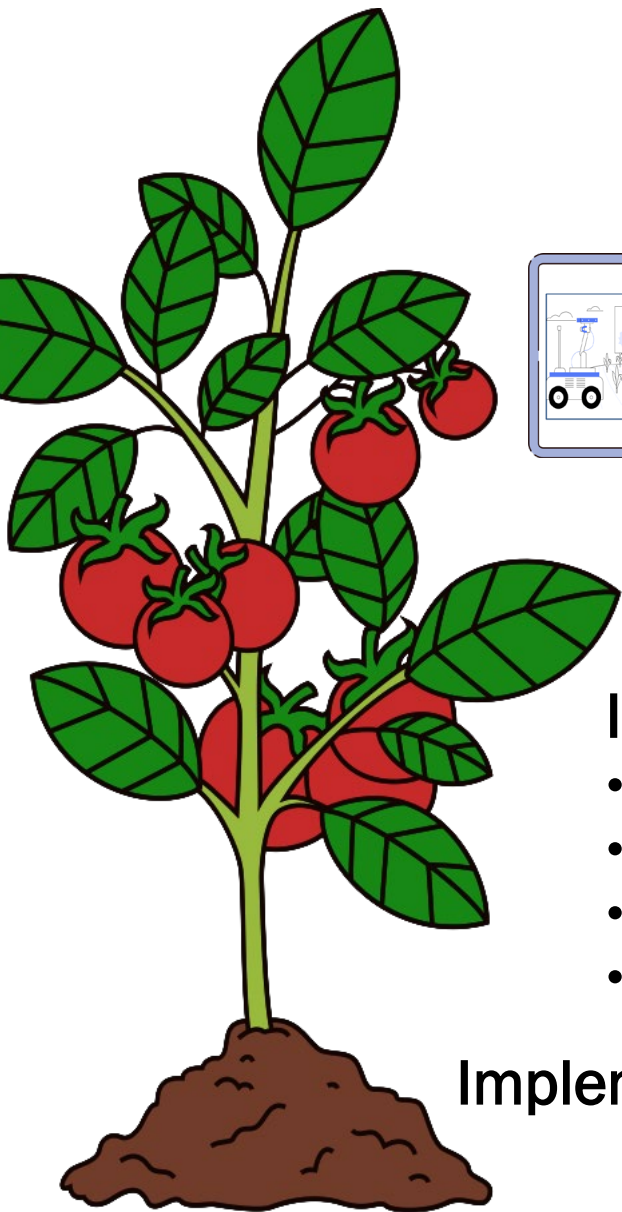
Shaw et al., 2023

Kotni et al., 2023

Kreide et al., 2019

Weise et al., 2017





Implementation of unique identifiers, e.g. DOI

- Use of the infrastructure of the ITPGRFA
- Assignment via EURISCO

Adherence to data standards

- Semantics data standards (controlled vocabularies, ontologies)
- Structural data standards, i.e. MIAPPE



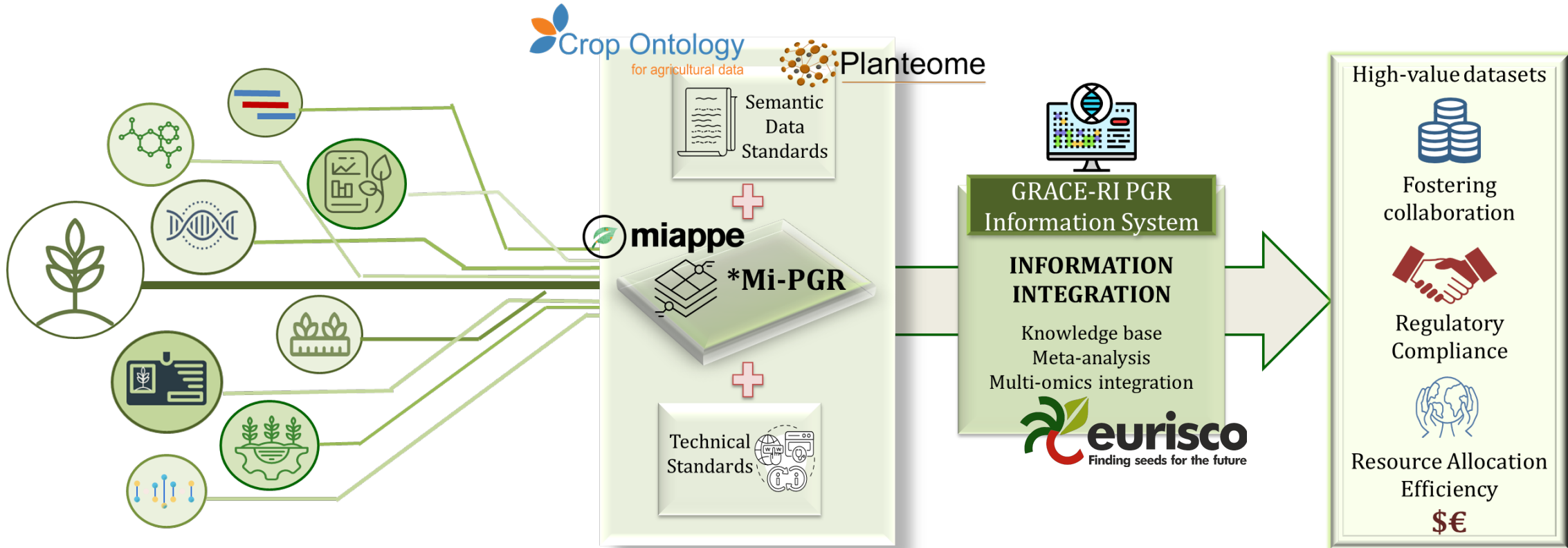
Implementation of best practices

- Establish SOPs for data collection and documentation
- Quality assurance and validation
- Data management plans
- Training and capacity building

Implementation of technical standards for data integration



Phenotypic Data: Progressing Towards FAIRness



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

Documentation is non-negotiable, regardless of institutional capacities

Way Forward

Think big, start small

- Pragmatic approach to data collection and management
- Plan for scalable growth

Future-proofing data

- Foster a culture of FAIRness
- Standardization is key
- Ethical data stewardship

Capacity Building and Collaboration

- Training and Support; Feedback loops
- Collaborate across networks; community engagement

Concluding Thought

In this age of data-driven science, it's not just about having the best tools; it's about bridging gaps in our data practices to ensure our data remains useful not just today, but for the future. With millions of underutilized PGR accessions, our task is clear: turn potential into progress.



Funded by
the European Union

THANK YOU

Catherine Hazel M. Aguilar
aguilar@ipk-gatersleben.de

