PRO-GRACE

PROMOTING A PLANT GENETIC
RESOURCE COMMUNITY FOR EUROPE

DELIVERABLE 1.5

Inventory of PGR information not yet represented in EURISCO and unified strategy for the interfacing of different information systems with EURISCO

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. 1.5

Inventory of PGR information not yet represented in EURISCO and unified strategy for the interfacing of different information systems with EURISCO

Contractual delivery date:
M28

Actual delivery date:
M28

Responsible partner:
IPK
Contributing partners:
IPGRI, WUR, UOB, UPV, KIS, VURV, CNR, ENEA

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	**D1.5**
Deliverable title	**Inventory of PGR information not yet represented in EURISCO and unified strategy for the interfacing of different information systems with EURISCO**
Type	Report
Dissemination level	
Work package number	1
Author(s)	Catherine Hazel Aguilar, Stephan Weise, Nigel Maxted, Sandra Goritschnig, Theo van Hintum, Gabriele Bucci, Jaime Prohens, Jelka Šuštar Vozlič, Vojtech Holubec, Giovanni Giuliano
Keywords	Information systems, plant genetic resources, documentation, EURISCO, FAIR

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

ABBREVIATIONS AND ACRONYMS

API	Application programming interface
BrAPI	Breeding API
C&E	Characterization and evaluation
CWR	Crop Wild Relatives
DDBJ	DNA Data Bank of Japan
DiSSCo	Distributed System of Scientific Collections
DOI	Digital Object Identifiers
DwC	Darwin Core
ECPGR	European Cooperative Programme for Genetic Resources
ELIXIR	European Life-Science Infrastructure
EMPHASIS	European Infrastructure for Multi-Scale Plant Phenotyping and Simulation for Food Security in a Changing Climate
ENA	European Nucleotide Archive
ENVO	Environment Ontology
EURISCO	European Search Catalogue for Plant Genetic Resources
EVA (EMBL-EBI)	European Variation Archive of the European Molecular Biology Laboratory-European Bioinformatics Institute
EVA	European Evaluation Network
FAIDARE	FAIR Data-Finder for Agronomic Research
FAIR	Findable, Accessible, Interoperable, Reusable
FAO	The Food and Agriculture Organization of the United Nations
GBIF	Global Biodiversity Information Facility
GLIS	Global Information System
GRC	Genetic Resource Centre
HTP	High-throughput phenotyping
INSDC	International Nucleotide Sequence Database Collaboration
ISA-tab	Investigation/Study/Assay tab-delimited format
ITPGRFA	International Treaty on Plant Genetic Resources for Food and Agriculture
MCPD	Multi-Crop Passport Descriptors
MIAME	Minimum Information about a Microarray Experiment
MIAPPE	Minimum Information about a Plant Phenotyping Experiment
MIxS	Minimum Information about any (x) Sequence
MLS	Multilateral System
NCBI	National Centre for Biotechnology Information
NGS	Next-generation Sequencing
NFP	National focal point
NI	National inventory
PECO	Plant Experimental Conditions Ontology
PGR	Plant genetic resources
PO	Plant Ontology
PRIDE	Proteomics IDEntifications Database
PUID	Persistent unique identifier
REST	Representational State Transfer
RI	Research Infrastructure
WIEWS	World Information and Early Warning System

Executive Summary

Europe maintains a diverse array of plant genetic resources (PGR), conserved and managed by an extensive network of institutions including public genebanks, research institutes, universities, community seed banks, orchards, botanical gardens, non-profit organizations, genetic reserves, protected areas and on-farm conservation systems. For over two decades, the European Search Catalogue for Plant Genetic Resources (EURISCO) has functioned as the central aggregator for *ex situ* passport data and main information gateway for PGR in the region. The catalogue has progressively expanded its scope to incorporate characterization and evaluation (C&E) data and, notably in 2024, extended its purview to include *in situ* crop wild relative (CWR) passport data from selected pilot countries, thereby broadening its remit beyond traditional core descriptors. Nonetheless, significant volumes of PGR-related data remain outside EURISCO's current framework.

Concurrent with EURISCO’s development, many contemporary research and conservation initiatives are generating increasingly diverse and high-volume PGR-related data that extend well beyond conventional passport and basic C&E descriptors. These include multi-environment high-throughput phenotyping, and multi-omics (genomic, transcriptomic, proteomic, and metabolomic) datasets essential for targeted conservation efforts and advanced (pre)breeding research. The consolidation of these heterogeneous datasets into a single repository presents insurmountable technical barriers due to their exponential scale, inherent structural complexity, domain-specific metadata requirements, and specialized computational infrastructures necessary for their analysis. Data interoperability challenges further compound this issue, as different research domains employ distinct ontologies, controlled vocabularies, and analytical frameworks that resist standardization within a unified schema. A technically viable solution necessitates maintaining these specialized datasets in domain-optimized repositories while establishing programmatic linkages with EURISCO through standardized application programming interfaces (APIs) and persistent, unique identifiers (PUID).

This deliverable systematically examines the current gaps in EURISCO's coverage, both in terms of institutional representation and data types, and analyzes the underlying governance, institutional, geopolitical, and scientific/ domain-specific factors contributing to these shortfalls. It addresses the foundational question of determining precisely which PGR information and materials should be catalogued within EURISCO, considering both the spectrum of PGR-related data categories and the diverse typologies of germplasm collections. The deliverable then presents a coordinated strategy for a hybrid federated infrastructure that preserves EURISCO's core focus while establishing efficient connections to complementary specialized repositories. The proposed framework delineates technical prerequisites and integration protocols necessary for implementation across key data categories, with repository-specific connection mechanisms tailored to each data domain's unique requirements.

Subsequent sections of the document identify potential synergies with existing pan-European research infrastructures and relevant international PGR initiatives. The deliverable concludes by proposing a multi-faceted approach that strengthens intra-country collaboration, addresses critical data gaps, and expands EURISCO's functional role without disrupting established local workflows and practices. A dedicated PGR research infrastructure, i.e., GRACE-RI can facilitate these developments by providing structured services, specialized training, and technical support. The final recommendations recognize that EURISCO operates within a defined scope determined by National Inventory contributions, and national and regional policy frameworks, while proposing concrete technical and organizational steps to enhance cross-repository connectivity, standardized data exchange protocols, and coordinated information system development in alignment with both European priorities and global PGR information initiatives.

1. Introduction

An estimated 5.9 million accessions of plant genetic resources (PGR) are conserved *ex situ* in over 871 genebanks worldwide (FAO, 2025 Hanson et al., 2024), making essential genetic materials readily available for utilization while effectively hedging against current and future food and environmental sustainability problems. These collections, encompassing landraces, heritage varieties, crop wild relatives (CWR), improved cultivars, and specialized breeding materials, constitute a critical genetic reservoir that underpins global efforts to ensure food security, improve nutritional quality, restore degraded ecosystems, and enhance socio-economic resilience, aligning with the 2030 Agenda for Sustainable Development (SDG 1 on Zero Poverty, SDG 2 on Zero Hunger, SDG 13 on Climate Action and SDG 15 on Life on Land) (Ulian et al., 2020; Lusty et al., 2021; Pathirana & Carimi, 2022; Ebert et al., 2023). However, lack of access to greater breadth of genetic diversity remains a major constraint on crop improvement (McCouch *et al*., 2013; IPCC, 2014). Paradoxically, only a fraction of genebank accessions end up in breeding pipelines and formal crop improvement programs (Fowler & Hodgkin, 2004; Engels & Visser, 2006; Anglin et al., 2018). This failure to utilize the entire spectrum of genetic variability in PGR is partly attributable to the limitations and inadequacies associated with PGR documentation and data management, among other contributory factors.

For instance, *ex situ* and *in situ* PGR conservation are inherently uncoordinated, as various institutions, ministries, conservation agencies, and national programs follow distinct mandates, standards, and funding structures (Anglin et al., 2018; Iriondo et al., 2021; van Hintum et al., 2021; Hanson et al., 2024), resulting in widely disparate practices and data protocols. In addition, the prioritization of urgent core operations such as germplasm regeneration, active *in situ* population management, and viability testing over thorough documentation leaves large portions of the collections poorly described. This data shortfall is further exacerbated by short-term, project-based funding cycles which, upon termination, result in fragmented and transient databases that lack the sustained commitment needed for effective long-term data stewardship. In certain cases, vital information (such as passport details, phenotypic evaluations, genomic profiles, time series monitoring data, and environmental descriptors) remains siloed in offline spreadsheets, local databases, outdated software, project reports, or publications, making data integration exceedingly more complex. Where data do exist, they are often housed in heterogeneous, non-interoperable systems with inconsistent or incomplete metadata, further limiting overall availability, accessibility, and utility (Anglin et al., 2018; Halewood et al., 2018).

For *in situ* and on-farm conservation, data-related challenges are much more apparent. Unlike formal *ex situ* PGR programs, *in situ* conservation operates under frameworks managed by governmental agencies or ministries with broader mandates spanning biodiversity management or agricultural extension, and, in some cases, collaboratively with genetic resource centres (GRC). However, individual CWR or LR populations are typically managed in genetic reserves or on farms, where managers and farmers rarely have the remit, resources or appropriate skills to describe, characterise and evaluate germplasm at the level required for advanced germplasm utilization (Maxted et al., 2025). Although a recent review of the relationship among *in situ*, on-farm and *ex situ* conservation strategies has advocated for a closer, integrated approach to PGR conservation, led by the GRC, where all *in situ* and on-farm conserved populations are backed up *ex situ* to provide (a) long-term security, (b) facilitate characterization and evaluation, and (c) aid end user access (Maxted et al., 2025), these steps remain largely conceptual and have not been widely adopted. Meanwhile, on-farm conservation typically evolves through *de facto* practices carried out by local communities, grassroots networks, or non-profit organizations (de Boef et al., 2010; Bellon et al., 2017; Engels and Thormann, 2020). In these settings, even when documentation exists, particularly where environmental stewardship measures support on-farm cultivation, it often relies on ad hoc or non-standardized methods (e.g., local surveys, farmer field notes, anecdotal observations, or coarse-scale biodiversity-monitoring protocols) (Jarvis et al., 2011). Moreover, local heritage or folk naming systems may not align with established taxonomic references (Zeven, 1998), and records of traits are often anecdotal or missing altogether. Such piecemeal, legacy datasets may limit the scope and depth of subsequent analyses and can be difficult to integrate with the curated, systematically catalogued data maintained in *ex situ* repositories.

In recent years, major advances in high-throughput phenotyping (HTP), multi-omics (including genomic, transcriptomic, proteomic, and metabolomic approaches), advanced computational tools, predictive characterization and detailed environmental monitoring have rapidly increased the volume and complexity of PGR data (Thormann et al., 2014; Krajewski et al., 2015; Halewood et al., 2018; Nguyen & Norton, 2020; Iriondo et al., 2021). Maximizing the value of these multifaceted datasets for conservation planning, pre-breeding research, and varietal development necessitates stable and scalable data infrastructures, standardized protocols, comprehensive indexing and discovery tools, and consistent curation practices (Halewood et al., 2018; Deng et al., 2023) that draw on concepts such as FAIR (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016). Many genebanks, national programs, regional projects and related initiatives, however, operate within intermittent or shifting funding cycles, making it difficult to ensure long-term data management and stewardship. When financial support wanes or institutional priorities shift, data platforms often become transient or remain inadequately maintained, thereby jeopardizing vital genetic information in these facilities. Furthermore, inconsistent metadata and the limited adoption of data standards further impede interoperability, ultimately diminishing the impact of these high-value datasets.

The European Search Catalogue for Plant Genetic Resources (EURISCO) (http://eurisco.ecpgr.org/) serves as a regional aggregator and gateway for PGR data (Weise et al., 2017; Kreide et al., 2019; Kotni et al., 2023). Established in 2003 under the European Cooperative Programme for Plant Genetic Resources (ECPGR), EURISCO initially consolidated *ex situ* passport and characterization-and-evaluation (C&E) data from national inventories (NI), enabling genebank curators, researchers, breeders, and policymakers to access information on more than two million accessions from over 400 institutes. In 2024, EURISCO broadened its scope by including *in situ* CWR data from pilot countries to capture a wider range of PGR information (van Hintum & Iriondo, 2022). This effort successfully demonstrated how *in situ* CWR (and potentially LR on-farm) passport data might be collated at the national level and uploaded to EURISCO through national focal points (NFP), at least in principle, making these population data available to potential user communities and showing how *in situ* germplasm might be accessed by end users. Nonetheless, significant challenges remain regarding how a predominantly centralized information system might incorporate large-scale, rapidly evolving datasets such as HTP outputs or multi-omics profiles. Concomitantly, less formalised *in situ* CWR data and on-farm landrace information risk remaining siloed, fragmented or be inconsistently documented, particularly when some may lie outside national PGR frameworks.

This deliverable addresses two primary objectives in response to these persistent gaps.

1. First, it presents an inventory of PGR data not presently included in EURISCO, detailing their categories, scale, and scientific or conservation significance.
2. Second, it proposes a unified strategy for interfacing disparate information systems with EURISCO.

This approach builds on the repository's strengths, provides a technically feasible solution for direct links with established external archives that host large, complex PGR-related datasets, and identifies points of synergy with pan-European research infrastructures and global PGR initiatives. Ultimately, these measures will enable EURISCO to evolve into a single, trusted repository for stakeholders seeking to discover and utilize the full breadth of Europe’s PGR landscape.

2. Methodology

2.1 Cross-Referencing Institute-level and Country-level Information across EURISCO, FAO WIEWS, Genesys, and National/Country Compliance Reports

Publicly available information from EURISCO provided the initial basis for identifying contributing institutions at the national level, including the specific types of data they submitted. The latest upload dates associated with each institution were recorded to evaluate the recency of these data submissions and to pinpoint any prolonged gaps in updates.

Data from the World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture (WIEWS) of the Food and Agriculture Organization of the United Nations (FAO) (https://www.fao.org/wiews/) were subsequently reviewed, noting that confirmed updates from WIEWS extend only through 2022, in order to determine whether institutions recognized at the national or international level might be underrepresented in EURISCO. This cross-reference revealed instances where institutions documented by FAO–WIEWS did not appear in EURISCO. Concurrently, Genesys (https://www.genesys-pgr.org/) was consulted to ascertain whether it listed any additional institutions, under specific countries, not accounted for in EURISCO.

To obtain more detailed insights into national-level institutional structures, mandates, and responsibilities, official country reports, policy/ treaty compliance documents (https://www.fao.org/plant-treaty/areas-of-work/compliance/compliance-reports/en/), and related publications were reviewed for information on formal recognition, governance frameworks, resource allocations, and coordination mechanisms pertinent to PGR data management. By comparing these sources with the previously collated EURISCO, FAO–WIEWS, and Genesys data, a more comprehensive view of the institutional landscape emerged, allowing verification of whether the documented entities were sufficiently represented and whether they contributed PGR-related data to their country's National Inventory. Finally, a structured questionnaire was administered to national focal points and key genebank managers to validate and expand upon the findings from these preliminary data checks and institutional listings.

2.2 Development of a Unified Strategy for Interfacing Different Information Systems with EURISCO

A systematic review was carried out to determine how PGR-related information systems, data archives, and domain-specific data management platforms can be potentially interfaced with EURISCO. Emphasis was placed on mapping each platform's data exchange mechanisms, i.e., how different information systems communicate (e.g., application programming interfaces (APIs), standardized web services, and custom file formats), and appraising the extent to which persistent unique identifiers (PUID) or equivalent referencing schemes support the reliable cross-linking of information across systems. Technical and governance documentation provided further clarity into data structures, semantic frameworks, and curation processes that shape how information is organized, stored and updated. In parallel, potential synergies with pan-European research infrastructures and global PGR initiatives were delved into, assessing established protocols, governance models, and data-sharing practices that could inform EURISCO's further development at both regional and international levels. By synthesizing these insights, the review laid the groundwork for a cohesive strategy that supports the comprehensive integration of diverse systems with EURISCO, which, when implemented, may improve data accessibility, promote consistency, and broaden the practical utility of PGR-associated datasets.

3. Overview of EURISCO's Current Structure and Aggregation Model

3.1 EURISCO's Data Architecture and Data Flow

EURISCO's data architecture (Figure 1) follows a centralized aggregation model fed by decentralized NI (Weise et al., 2017; Kreide et al., 2019; Kotni et al., 2023). *Ex situ* passport records follow the FAO/IPGRI Multi-Crop Passport Descriptors (MCPD) v2.1 (Alercia et al., 2015), supplemented by additional fields to capture key accession-level details. For *in situ* CWR, these same MCPD principles are extended with specialized descriptors that capture the unique considerations relevant to populations remaining in their natural or cultivated habitats. While each *ex situ* record pertains to a discrete genebank accession, each *in situ* record represents a population that is actively maintained and managed on site. References to responsible institutions, i.e., holding institutes or liaison institutes (for *in situ*) are also embedded within the schema using FAO WIEWS codes, to ensure that each data record clearly identifies where and by whom the material is curated or coordinated. This mechanism provides a direct point of contact for users who wish to seek further information, request germplasm, or inquire about management practices. Meanwhile, C&E data adhere to a simplified phenotyping schema, ensuring that each data point is accompanied by metadata about how and when it was collected. EURISCO's C&E templates represent a "minimum consensus" for phenotypic data exchange. They intentionally capture key elements like trait names, values, and basic experimental info with no enforced standard ontology or units.

It is important to note that EURISCO was designed for discovery-level information. It does not store operational curatorial data, active conservation management details, or granular utilization records. Rather, it acts as a gateway, which enable stakeholders to locate relevant PGR materials and connect with those who maintain them *in situ* or *ex situ*. *(Section 4 provides a broader review of the entire spectrum of PGR data types, including those not reflected in EURISCO's current schema)*

In each member country, a designated National Focal Point (NFP) compiles the NI and submits updates via a dedicated web interface. Upon submission, EURISCO implements automated validation checks to verify completeness and format consistency, flagging discrepancies (matching taxonomic names to authoritative references e.g., USDA GRIN Taxonomy or Mansfeld's World Database, and standardizing institute codes through FAO–WIEWS identifiers) for NFP review. Once the NFP validates any flagged items and confirms the upload, the revised data are incorporated into the EURISCO catalogue and become globally accessible. Most countries provide updates at least annually, ensuring that EURISCO reflects the status of decentralized national collections under a unified, quality-controlled framework.

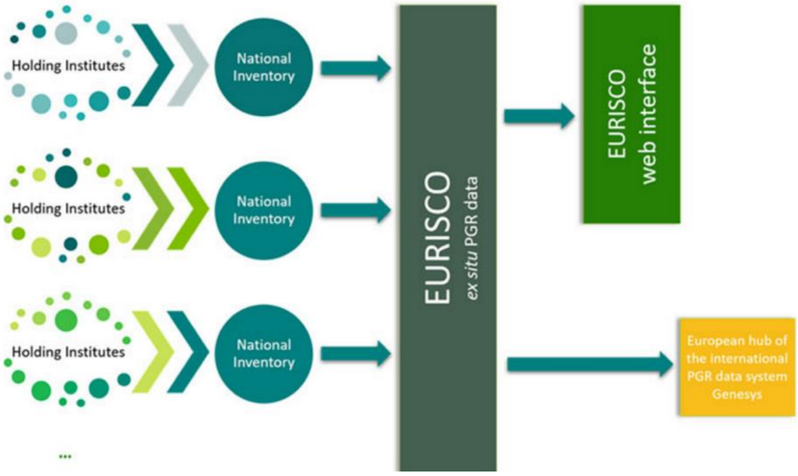

Figure 1. System Architecture of EURISCO (Source: Arend et al., 2022)

3.2 Governance and National Data Sovereignty

EURISCO's governance structure balances decentralized data ownership with centralized oversight. Data remains in a federated environment at the national level. Each country retains sovereign authority over its PGR data, including responsibility for content quality and the right to determine what information is shared. NFP manage technical submissions and must explicitly authorize each data release, guaranteeing that no national dataset is disseminated through EURISCO without local approval. At the regional level, the ECPGR oversees policy coordination, while the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK-Gatersleben, Germany) hosts EURISCO's infrastructure on behalf of the ECPGR Secretariat. IPK adheres to the guidelines of the ECPGR Steering Committee and maintains stringent principles on data use, thereby safeguarding national rights. A EURISCO Advisory Committee (https://www.ecpgr.org/contacts-in-ecpgr/ecpgr-contacts/eurisco-advisory-committee-members/) provides counsel on system enhancements and ensures that developments reflect collective European interests while respecting individual national priorities.

3.3 International Alignment and Data Exchange

In addition to its European mandate, EURISCO serves as a strategic node in global germplasm information networks by serving as a principal data provider (*ex situ* passport data) to Genesys (https://www.genesys-pgr.org/), the international information system aggregating genebank records worldwide. This centralized exchange mechanism obviates the need for individual European genebanks to upload data to multiple international systems, thereby minimizing administrative workload while ensuring consistent international visibility for European genetic resources. This streamlined approach also supports the Global Information System (GLIS) (https://glis.fao.org/glis/) established under the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA), facilitating compliance with treaty obligations and advancing broad access to curated data. Cooperation with Genesys, the FAO, and other global initiatives is upheld through cross-representation on advisory committees, which in turn reinforces alignment on data standards and strategic objectives. Simply put, EURISCO is a pillar in the emerging network-of-networks for PGR data, bridging European NI with global data exchange efforts.

4. Review of Data Types Involved in PGR Science

The data that underpins the documentation, management and use of PGR spans a remarkably broad range, from fundamental passport information to high-resolution "omics" outputs. This diversity mirrors the multitude of activities and objectives inherent in PGR science, including maintaining *ex situ* collections (where viability tests, storage procedures, genetic integrity checks and regeneration events must be closely tracked), monitoring *in situ* populations (where ecological and socio-cultural nuances affect both data collection and sharing), and conducting advanced research (where large-scale phenotypic and omic datasets address specialized scientific questions). Given these distinct contexts, it is neither practical nor beneficial to centralize all data types within a single information system. Because each data type relies on distinct curation workflows, data structures, domain-specific analytical methods, and update intervals, merging them into a single system would compromise the precision and usability of the data, strain the platform's capacity, and ultimately undermine the specialized scientific and management needs each data category is meant to address. Instead, each data-holding system, whether a genebank database, a national or regional repository, or a specialized research platform, fulfils a distinct yet equally vital role, ensuring that each category of PGR data is curated, preserved, and leveraged by those with the appropriate expertise, infrastructure, and mandates to maintain its integrity and maximize its impact. The tables that follow present a structured overview of PGR-associated data that are grouped into:

1) Cross-cutting data types (Table 1), such as passport information, fundamental C&E descriptors and legal or policy (meta)data, which form the backbone of nearly all PGR documentation systems. These data are universally relevant because they establish the identity, origin and traceability of each accession or population;
2) *Ex situ*–specific data types (Table 2), capturing routine genebank procedures and metrics (e.g., seed viability, storage details, regeneration logs, and distribution records) that are indispensable for maintaining and supplying germplasm in formal collections;
3) *In situ*–specific data types (Table 3), focusing on population monitoring, habitat conditions, and community-based conservation, which are central when PGR remain in wild habitats or farmers' fields;
4) Research-oriented data types (Table 4), encompassing the advanced, high-volume information generated by modern breeding and biotechnological investigations, from genomic sequences to high-throughput phenotyping.

Deliverables 1.1, 1.2, 1.3, 1.4, 5.1 also provided an extensive discussion about specific PGR data types.

Table 1. Cross-cutting PGR data types: Descriptions, Context, Standards, Formats and Interoperability

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
Passport data	The fundamental identification and origin data for a PGR. This includes the accession's unique identifier (in the case of *in situ*, population-level information) and name, taxonomic information (scientific genus/species, common name), biological status (wild, landrace, breeding line, cultivar, etc.), origin details (country and locality of collection or provenance, latitude/longitude of collecting site, elevation), date of collection or acquisition, the donor or collector information, and other basic descriptors. This is the "who/what/where/when" baseline information for any accession.	Universal across *ex situ* and *in situ* contexts. Every genebank accession has passport data documented, and even *in situ* conservation sites maintain passport-like info for the populations (taxon, location, etc.). It is crucial for genebank management (distinguishing accessions, tracking duplicates), for users (to select material by origin or traits associated with origin), and for legal purposes (knowing the	FAO/Bioversity Multi-Crop Passport Descriptors (MCPD) v2.1 is the internationally accepted standard for recording passport data. MCPD defines a set of fields and coding schemes (e.g. country codes, biological status codes) that genebanks worldwide use, enabling data exchange. For in situ conservation of Crop Wild Relatives (CWR), MCPD has been adapted with additional descriptors to document wild populations in their natural habitats. Darwin Core's germplasm extension maps closely to MCPD, so that passport data can also be shared via biodiversity networks (GBIF, etc.). The International Treaty's Global Information System (GLIS) also leverages passport data when assigning DOI to accessions (passport info is part of the minimum metadata). Genebank databases and PGR information systems (EURISCO, Genesys) require passport data in MCPD format for uploads.	Typically stored in structured databases or spreadsheets, one row per accession with multiple columns (fields). passport datasets are routinely exchanged in MCPD-compliant CSV or Excel files. Darwin Core Archive (DwC-A) is used to publish passport data to biodiversity databases, with terms aligned to MCPD (ensuring easy mapping).	**Yes**

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
		country of origin for access and benefit sharing (ABS) compliance). In essence, passport data travels with the germplasm and underpins all other data.			
Characterization data	Data describing observable, heritable traits of an accession measured in a common environment. Characterization usually covers morphological or agronomic traits that are largely genetic in control and expressed across environments (e.g., flower color, plant height, seed shape etc). Often collected by growing the accession in field or greenhouse and recording standardized descriptors.	Key for both conservation and use: genebanks characterize accessions to catalogue diversity and make it easier for breeders/ researchers to find materials with desired traits. Because characterization traits are stable, this data is used to differentiate	Internationally agreed Descriptor Lists (usually crop-specific) define the traits and how to score themhttps://openknowledge.fao.org/server/api/core/bitstreams/62d6d5d3-f7a3-47f3-abaf-9d2d5e230e51/content. Developed by FAO and Bioversity International, these provide standard trait names and scales (e.g., 1–9 scoring). Many genebanks and networks use these to ensure consistency. The Crop Ontology (CO) project also formalizes trait definitions for interoperability. For experimental contexts, MIAPPE (Minimal Information About Plant Phenotyping Experiment) provides metadata standards to record how characterization trials are done.	Stored in spreadsheets or databases as tables of accessions × traits. However, the structure and interpretation of descriptors can vary considerably as descriptor lists are often modified, used inconsistently, or applied without formal trait ontologies.	**Yes.** Aggregated i EURISCO as C&E data, but coverag depends on individual institutions submitting these records. Not all d so, and data protocols/ trait definitions/ scoring scales are not yet harmonized.

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
		accessions and to do preliminary grouping (e.g., by maturity class or seed type). It’s applicable to *ex situ* collections and also to on-farm or *in situ* observations of traits in traditional varieties.		Metadata describing how and where the traits were recorded is often absent or unstructured, limiting reusability. MIAPPE now provides a robust metadata model, but community-wide adoption is still limited. While many datasets are shared in CSV, greater interoperability depends on using shared ontologies (e.g., Crop Ontology) and adopting metadata standards like MIAPPE for	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
				broader FAIR compliance.	
Evaluation data	Trait data generated from observing how PGR perform or respond under specific environmental, agronomic, or experimental conditions. This includes quantitative or scored traits related to agronomic performance (e.g. yield, biomass), abiotic stress tolerance (e.g. drought, salinity), biotic stress resistance (e.g. pests, pathogens), quality attributes (e.g. nutritional content, grain hardness, flavor), and end-use suitability (e.g. storability, cooking properties). Evaluation data may also include subjective assessments (e.g. farmer preference ratings, taste panels). It reflects the interaction of genotype and environment, making it	Collected in a wide range of contexts, making it a cross-cutting data type: • *Ex situ*: genebank-managed evaluation plots or multi-location trials involving curated subsets (e.g. core or mini-core collections). • On-farm: participatory variety selection trials with farmers scoring varieties under real farming conditions, often considering	• FAO/ Bioversity crop-specific descriptor lists often include evaluation traits (e.g. disease resistance, drought tolerance), but usage varies widely. • Ontologies (e.g. CO) provide standardized trait names, definitions, measurement methods, and scales that are critical for trait harmonization across sources. • MIAPPE offers a comprehensive metadata model for capturing experimental design, environmental parameters, and measurement protocols, which are essential for making evaluation data reusable and comparable. • BrAPI (Breeding API) supports standardized access to trait observation data via APIs, especially within breeding information systems. • MCPD allows only basic evaluation flags or notes, not structured trait data. • Other standards like ISA-Tab may be used for packaging data and metadata, particularly in research trials.	Managed in relational databases or spreadsheet formats, with rows for observations and columns for accession ID, trait name, trait value, unit, location, date, trial environment, and replicate/block. In practice, data structure and content vary widely: • Many genebanks collect a limited set of traits and may store them with little or no environmental metadata.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
	distinct from basic morphological characterization.	cultural or culinary preferences. • *In situ*: population-level assessments of traits such as fecundity, stress response, or productivity in wild or semi-managed habitats. • Research: replicated trials under experimental designs to quantify trait-environment interactions, often linked to genomics, breeding, and phenotyping projects. Evaluation data supports both conservation priorities and		• On-farm and participatory data may include farmer scores, preferences, and rankings, often in formats not harmonized with scientific trait lists. • Research institutions may generate large multi-environment datasets, but often use bespoke formats or store in local breeding platforms. Interoperability is challenged by: • Inconsistent trait definitions, scoring scales, and units across	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
		downstream utilization.		institutions. • Lack of standardized metadata describing trial design, environmental conditions, and protocols (though MIAPPE addresses this). • Infrequent use of trait ontologies (e.g. CO) or persistent identifiers for traits and units. Where adopted, MIAPPE and BrAPI substantially improve interoperability. BrAPI enables direct exchange of trait data via APIs, allowing integration	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
				across genebank, breeding, and phenotyping systems. Evaluation data is increasingly being shared via EURISCO and Genesys (as evaluation datasets), but full harmonization and reusability still depend on community-wide use of trait ontologies, controlled vocabularies, and (meta)data standards.	
Image/ media documentation	Photographs and other media documenting the germplasm or associated knowledge. This includes images of plants (e.g. field photographs of an accession's phenotype,	Cross-cutting usage: genebanks use images as part of characterization (e.g. a photo of	Darwin Core supports media through terms like dwc:associatedMedia, linking specimens to media files. Additionally, TDWG's Simple Multimedia Extension provides a structure for describing media (with fields for URL, description, creator, rights, etc.). Genebanks and biodiversity repositories often follow general metadata standards for images: e.g. EXIF data for technical details, and Dublin Core or other library standards for descriptive metadata. No PGR-specific image standard exists, but initiatives like the Crop Ontology	Stored as digital files (JPEG, PNG, MP4, etc.) referenced in databases by a URL or file path. Many PGR	**No.** EURISCO doe not currently host or aggregate image/audio/vide files. Its scope is primarily textual accession-level (e

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
	herbarium voucher scans, seed photos for morphology), as well as possibly audio recordings (e.g. pronunciation of local variety names, interviews about the variety) or video clips (such as farmer interviews or field management practices). These media files serve as a visual or auditory record complementing textual data.	a fruit or ear of corn for each accession in an evaluation trial), *in situ* projects use photos to document populations and habitats, and socio-cultural work might record audio of traditional stories about a plant. Media data helps experts and laypersons alike to better understand and verify the material (for instance, a photo can confirm a plant's traits or identity). It's also increasingly used for	sometimes include image examples for descriptors. In practice, many genebanks just adhere to internal guidelines (like photographing at specific angles or scales).	databases (e.g. Genesys, GRIN) allow uploading images for accessions. These can be exported or shared by providing links. The main interoperability issue is metadata, making sure each image is associated with the correct accession and carries captions or tags.	*situ*) and population-level (*in situ* CWR) data

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
		machine learning (digital phenotyping) and broader outreach (showcasing crop diversity).			
Collecting Missions / Ecogeographic Surveys	Large-scale, often systematic field expeditions or survey campaigns designed to collect new germplasm samples and gather detailed environmental, ecological, and socio-economic data about target areas. These missions typically aim to: (1) identify and collect priority PGR for *ex situ* conservation, (2) document *in situ* populations and their habitats, and (3) map ecogeographic parameters (climate, soils, land use).	• *Ex situ*: Collecting missions feed genebanks with new accessions, ensuring that priority species or landraces at risk are safeguarded. They also capture contextual “passport++” data (precise location, habitat, notes on local uses). • *In situ*: Ecogeographic surveys can reveal previously	• Bioversity/FAO Collecting Manual: provides guidelines on how to plan and conduct germplasm collecting, record associated data, and engage local communities. • FAO Genebank Standards: emphasize rigorous documentation of site conditions and collecting protocols. • DwC: can capture occurrence/observation data for wild populations encountered during surveys. • ECPGR: relevant references on collecting strategies, sampling methods, and preliminary *in situ* evaluations.	• Typically recorded in field survey forms, Excel sheets, or specialized applications (including GPS-based tools) for georeferenced data. Data includes both structured fields (coordinates, collector info, species ID) and unstructured observations (local name, ecological notes).	**No.** Detailed ecogeographic surveys are typically held in separate institutional or research databases

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
		unknown wild populations or landraces, assess threats, and inform *in situ* conservation plans (e.g., establishing genetic reserves or supporting on-farm maintenance). They also help identify gaps for future collecting.		• GIS integration is common if collecting teams overlay data onto climate/land-use maps. • Core attributes (species name, location) can align with DwC or MCPD for *ex situ* upload, but richer environmental or socio-economic details often remain in separate survey reports or internal databases.	
Legal documentation	The physical or digital documents that accompany germplasm to prove legality of acquisition and use. This	Applies whenever germplasm is collected, transferred, or	Nagoya Protocol (CBD) and national ABS laws set requirements for permits and mutually agreed terms (MATs) when accessing genetic resources. ITPGRFA sets the Standard Material Transfer Agreement (SMTA) as a requirement for Annex I crops in the Multilateral System (MLS). FAO Genebank Standards explicitly state that all samples should be acquired legally with relevant	These documents are usually stored as PDFs or paper copies in	**No**

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
	category includes collection permits and agreements (e.g. a permit from a national authority to collect wild samples, often with Prior Informed Consent (PIC) documents from local communities), material transfer agreements (MTAs) for exchanges (including the Standard MTA for Treaty-regulated transfers), phytosanitary certificates for shipments, and any other contractual documents stipulating terms (e.g. a letter of agreement with a donor specifying usage restrictions). Essentially, these are the signed papers and certificates ensuring compliance with laws and agreements.	distributed, so it cuts across *ex situ* and *in situ*. Genebanks must obtain legal docs when adding new material (to ensure it was acquired legally under national laws) and provide the correct documents when sending material out. Researchers working *in situ* similarly need permits to access certain areas or traditional knowledge. This documentation is crucial for demonstrating that access and use of PGR is done ethically	documentation. While each country or provider has its own permit form or MTA text, the SMTA is a standardized form used globally for a large subset of PGR exchanges. Many organizations have templates for MTAs (the SMTA or others for non-Annex I). Phytosanitary certificates follow International Plant Protection Convention (IPPC) models. So, although the documents themselves are varied, the frameworks (CBD/Nagoya, ITPGRFA, IPPC) are the guiding standards.	genebank archives, often referenced in the database by an ID or note. They are not part of routine data exchange, due to sensitive content and variability. However, key metadata from them (permit ID, SMTA number, date, issuing authority) may be recorded in a database. The SMTA has a digital system (Easy-SMTA) that genebanks use to log agreements, which is a step toward standardized digital records but is not a	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
		and legally; lack of it can halt utilization or lead to disputes.		public format for all to use. Legal docs are kept as unstructured files linked to accessions.	
ABS compliance records	Metadata and logs related to Access and Benefit-Sharing (ABS) obligations for an accession or set of germplasm. This includes tracking the country of origin and its ABS status, the conditions under which the germplasm was acquired (e.g. terms MAT document), whether the accession is part of the Multilateral System of the Treaty (MLS) or not, any restrictions on use (e.g. "for research only"), and benefit-sharing commitments (like a requirement to share royalties or to send duplicate samples back to the provider). It may also	Relevant at both collection (acquisition) stage and distribution/use stage. Genebank managers maintain these records to ensure they honor the conditions tied to each accession. For example, if an accession was acquired pre-Nagoya with no conditions, it might be freely available; if post-Nagoya	Nagoya Protocol (2010) provides the framework – countries implement it via national legislation requiring prior informed consent and MAT. As a result, genebanks often have to record for each accession whether it has an Internationally Recognized Certificate of Compliance (IRCC) from the ABS Clearinghouse (which is a standardized code confirming PIC/MAT). The ITPGRFA framework simplifies ABS for Annex I crops by designating them under the MLS – many genebanks flag accessions as "MLS" or "non-MLS". For MLS accessions, using the SMTA and reporting in Easy-SMTA is the norm (thus the SMTA itself is part of ABS records). Some initiatives like CETAF (for herbaria) have developed best-practice checklists of data to record (e.g. source country permit ID, etc.), which genebanks can adapt. In short, while no single standard data format exists, common elements (origin, legal status) are well-understood.	Often implemented as flags or notes in genebank databases: e.g. a field "MLS status = yes/no", "Restrictions = none/MAT details", or a link to a permit record. The data is mostly internal. However, it's increasingly needed to share at least some of this when transferring material	**Minimal.** EURISCO can capture whether an accession is part o the MLS (a key AB indicator). However, detailed ABS conditions, IRCC numbers, or benefit-sharing logs are not part o EURISCO's schema. Most ABS details remain managed internally due to complexity and confidentiality.

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
	cover records of benefits shared (e.g. monetary benefit dates, technology transfer provided).	with a special agreement, that must be noted (so that any requester is informed or certain uses are barred). It's cross-cutting because even *in situ* efforts need to track ABS (e.g. a community conserving a variety might have an agreement with a research institute on benefit-sharing). ABS records ensure transparency and compliance in utilization of PGR.		between genebanks or to users (so the next holder knows the conditions). Interoperability is emerging: for example, when DOIs are assigned to accessions in GLIS, one can attach the MLS status and even link SMTA events to those DOIs. The ITPGRFA's systems and the ABS Clearinghouse use structured data, but these are separate systems. There isn't yet a unified exchange format for ABS metadata	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
				between genebank databases, so most information is communicated via accompanying documents or institutional databases.	
Policy/agreement metadata	Data about broad policy designations or agreements that affect the management of PGR collections. This includes noting if an accession or collection is part of an international network or agreement (e.g. part of the Article 15 international collections under the ITPGRFA like those of CGIAR centres, which means they're automatically under the MLS; or designated as part of AEGIS (European Collection) which implies agreed conservation	This is cross-cutting and typically at the collection or subset level rather than each accession (though often applied per accession). Genebank administrators use it to manage obligations (e.g. CGIAR genebanks must make Article 15	Frameworks include the ITPGRFA, e.g. CGIAR centers have agreements with the Treaty and thus tag their accessions accordingly. AEGIS has its own criteria and maintains a registry of European Accessions. Each such program often defines what data to tag: AEGIS requires an "AEGIS flag" on accessions accepted into the European Collection. While these frameworks exist, the implementation in data terms is usually institution-defined (like a checkbox "Article 15 collection" in a database).	Usually recorded as boolean flags or list membership. E.g., a genebank database might have a field "International status" with options like MLS, Article 15, AEGIS, etc. These metadata can be shared: for instance, EURISCO and Genesys	**Partial**. EURISCO includes an AEGIS flag, MLS status, and some policy indicators in its accession records. Other specialized designations (e.g., national heritage variety, IP-protected) are not systematically captured because they are highly country- or institution-specific.

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
	responsibilities and availability). It could also record if certain germplasm is under Plant Breeders' Rights or other IP, or if it's considered a national heritage variety under special policy. Essentially, this metadata tags germplasm with any higher-level policy status.	material freely available and report to the Treaty, AEGIS accessions must meet quality standards and be available to all Europe). For users, such tags can indicate how freely they can request material or any formalities needed. It helps align genebank operations with international policies and ensures clarity on which ruleset applies to which accessions.		indicate which accessions are part of AEGIS or the MLS. The format is just an attribute of the accession in those databases. Interoperability is straightforward when the definitions are clear – e.g. everyone in Europe knows the AEGIS flag meaning. Many of these designations are also reported in international information systems: e.g. WIEWS collects data on MLS holdings. Technically, this	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO Coverag
				data is easy to exchange (it’s categorical), but one must agree on controlled vocabulary. As policies evolve, new categories may be added. Overall, this metadata can be integrated into exchange formats (MCPD v2 has a field for MLS status, for example), making it reasonably interoperable.	

Table 2. ***Ex situ*****-specific data types: Descriptions, Context, Standards, Formats and Interoperability**

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO coverage
Seed viability data	Information on the germination ability and vigour of stored seeds (e.g. germination percentage, seedling health). Typically obtained via periodic germination tests to monitor seed longevity. Critical for deciding when to regenerate samples or if they are suitable for distribution.	Used by genebank curators in *ex situ* seed banks to ensure collections remain alive over time. High viability means seeds can be safely stored longer; declining viability triggers regeneration or withdrawal from distribution. Essential for maintaining the usefulness of stored accessions.	FAO Genebank Standards prescribe regular viability monitoring at intervals based on seed longevity. Testing protocols follow (or in many cases, modify) International Seed Testing Association (ISTA) rules for germination. Genebank management policies set viability thresholds (e.g. ≥85% for distribution).	Typically recorded in genebank databases as a test event (date, method, sample size, % germinated). Often stored in relational tables linked to the accession. Not commonly exchanged publicly, though internal systems (e.g. GRIN-Global) have standard fields. Could be shared in reports or as CSV if needed, but no widely used external format.	**No.** All of these data types are highly dynamic, institution-specific, and primarily operational in nature, which are necessary for internal genebank management but are not required by most users at a first point of access. They are updated frequently, reflect operational protocols and lack standardized global exchange protocols. Keeping this information accurate in a centralized system would require significant coordination and resources across multiple institutions,
Seed health data	Data on phytosanitary status of seeds or plant materials, including results of health tests for pathogens (fungi, bacteria, viruses, insects) and seed purity. Ensures material is pest- and disease-free for safe conservation and exchange.	Relevant in *ex situ* genebanks before storage and especially prior to distribution. Seed health labs test accessions (e.g. for viruses or seed-borne diseases) to prevent spreading pests. Important for international shipments (phytosanitary certificates require these data) and for maintaining collection integrity (avoiding cross-contamination in storage).	IPPC and national phytosanitary regulations provide the legal framework for required tests. Standard testing protocols by ISTA or FAO genebank guidelines are used for seed health. Genebank Standards mandate that samples are acquired and distributed with proper phytosanitary documentation.	Often recorded as lab reports or database fields (e.g. “virus X: not detected” on a given date). Formats are typically internal; data may be stored as binary pass/fail flags or detailed results. Phytosanitary certificates are issued as PDFs. Data usually stays in-house or attached to shipments. If shared, it’s via PDFs or plaintext reports rather than a structured global database.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO coverage
Storage data	Details of storage conditions and locations for conserved materials. Includes type of storage (seed dry freezer, cryopreservation, in vitro, field bank), temperature and humidity settings, container type, accession inventory code, etc. Also indicates whether sample is in long-term (base) or medium-term (active) storage.	Used by genebank managers to ensure each accession is kept under optimal conditions for longevity. For example, orthodox seeds are stored at low temperature and moisture; clonal crops might be in cryo or in vitro. Knowing where and how an accession is stored (freezer unit, vial number) is essential for retrieval and management.	FAO/Bioversity genebank standards define recommended conditions (e.g. seeds dried to low moisture and stored at -18°C for base collections). MCPD include fields like Storage type to indicate status (e.g. long-term, short-term). Internal genebank SOPs and quality management systems (e.g. ISO 9001) ensure storage data is recorded consistently.	Managed in inventory databases, often with fields for location (e.g. cold room/shelf ID) and environment parameters. Not commonly shared externally. Some systems might include a general flag for "Accession storage" or availability, but precise storage conditions aren't exchanged. Data are usually exported as reports or spreadsheets for backup or audits. Interoperability is mainly via adherence to standards like MCPD codes (e.g. code for long-term storage).	without matching the core purpose of EURISCO, which is to help users discover and request genetic resources at a broad, accession-level scale. Furthermore, certain pieces of distribution data can be sensitive, meaning open publication is not always appropriate. As a result, maintaining operational information at the institutional level and sharing it through more specialized channels remains the most straightforward and reliable approach.
Regeneration data	Information on regeneration events (the process of growing plants from stored material to produce fresh seed or propagules). Data include regeneration date and location, field or greenhouse conditions, number of plants grown, pollination control measures, harvested seed quantity, and any genetic integrity safeguards. It documents how an accession was renewed.	Used by genebank curators to maintain sufficient seed stocks and genetic integrity. *Ex situ* collections regenerate accessions when seed stocks are low or viability falls below thresholds. This data is crucial for tracking how often an accession has been regenerated (which can affect its genetic makeup) and for planning future regenerations. Also ensures the accession's identity and traits are	Genebank guidelines (e.g. Bioversity's seed handling manual) emphasize careful regeneration to minimize genetic drift or contamination. FAO Genebank Standards call for documenting regeneration processes for each accession. There is no universal exchange format for regeneration events; however, internal databases (e.g. GRIN-Global) have standard fields for regeneration	Stored in genebank information systems as event records linked to the accession. A typical format is a log entry: accession X regenerated at site Y on date Z, producing N seeds. Not shared publicly; data are used internally. If needed for exchange, it could be included in a report or as part of an accession's metadata. Different institutions record different details, but all adhere to the concept of logging regeneration events for transparency.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO coverage
		maintained (or notes if any deviations occurred).	history (such as regeneration number, location, method).		
Safety duplication data	Records about safety backup copies of accessions stored in an alternate location as a security measure. Includes the existence and location of duplicate samples (e.g. seeds deposited in the Svalbard Global Seed Vault or another genebank), the date of duplication, and accession identifiers at the duplicate site.	Used to ensure *ex situ* collections are safeguarded against accidental loss. Curators maintain this to know that if the primary collection sample is lost (due to freezer failure, natural disaster, etc.), a duplicate exists elsewhere. Applies to orthodox seeds (commonly duplicated in black-box storage at another genebank or Svalbard) and can also apply to in vitro or cryo collections (duplicate cultures in a separate lab).	FAO Genebank Standards require safety duplication for *ex situ* collections (each accession should have a backup at a geographically distant site). Agreements like the Standard Deposit Agreement with Svalbard outline procedures for safety backups. Documentation typically follows genebank-internal conventions, noting the holding institute of the duplicate (often using FAO WIEWS institute codes) and the storage conditions there.	Often maintained as a simple field or linked record (e.g. "Safety duplicate at XYZ genebank, Svalbard vault ID#..."). Format is usually textual or coded (yes/no for safety duplicate, plus location). Some global databases (e.g. EURISCO or Genesys) may indicate if an accession is safety duplicated, but details are usually internal. If using common institute codes and accession identifiers, another genebank can identify the material. However, there's no single global registry of all safety duplicates aside from the Svalbard inventory.	
Inventory (meta)data	The quantitative and logistical data about stored germplasm samples. Includes the number or weight of seeds in stock, number of propagules or plants (for clonal material), container IDs, lot numbers, and current status (available, depleted, etc.).	Primarily an *ex situ* genebank concern, though community seed banks maintain similar records. Curators rely on inventory data to manage distribution (knowing how much can be sent) and to plan regenerations (when stock is low). Each	Institutional genebank database schemas (e.g. GRIN-Global) define standard fields for inventory management (quantity, lot ID, viability per lot, etc.). FAO documentation standards note that accession information should include	Managed in relational databases with tables for inventory linked to the accession table. Data format is structured (numeric fields for quantities, text for lot IDs). Not usually shared outside the institution; instead, availability or stock status might be shared. If exporting, CSV or Excel inventories	

Data Type	Description	Scope/Context	Standards/Frameworks	Format & Interoperability	EURISCO coverage
	Also covers metadata about each seed lot such as its origin (original collection or a regeneration batch), and its link to the parent accession.	accession can have multiple inventory lots (e.g. a base lot and an active lot); tracking each is important for management efficiency and to prevent loss.	inventory data. There is no external standard exchange format for inventory, but genebanks often adhere to internal standards for recording units (e.g. number of seeds vs. weight).	can be produced for backups or audits.	
Distribution data	Records of the distribution and use of germplasm from the collection. Includes details of orders and shipments: who the recipient is (institute or researcher), what accession and how much was sent, date of shipment, purpose of use, and the terms under which it was sent (e.g. SMTA number). May also include feedback from users or follow-up information.	Used by genebanks to track how materials are utilized and to fulfill reporting obligations. Every time seeds are sent out, a record is kept, which is crucial for accountability (knowing what left the genebank) and for benefit-sharing (under the ITPGRFA, certain uses must be reported). It also helps measure impact and demand for specific accessions. This data type links *ex situ* collections with the user community (breeders, researchers, farmers).	The ITPGRFA mandates the SMTA for exchanges of materials in the MLS, and genebanks use the SMTA as a standardized legal framework. Genebank Standards emphasize documenting distribution and making such data available to enhance use. Internally, genebank databases have order management modules (e.g. to generate shipping lists and MTAs).	Maintained in internal systems; each distribution is an entry (with fields for accession ID, recipient info, SMTA or MTA reference, quantity). Formats are often relational (one table of distributions linking accession and recipients). Some summary distribution data may be shared (e.g. number of samples distributed per year, or via reporting to FAO). The ITPGRFA's reporting is done via an online portal (not a public format). There's currently no universal data exchange standard for distribution events; however, if needed, data can be exported (CSV, etc.). Interoperability is mainly at the conceptual level (everyone records similar info) rather than via automatic data exchange.	

Table 3. *In situ*-specific data types: Descriptions, Context, Standards, Formats and Interoperability

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
Population Size/ Abundance Data	Measurements of population size and abundance, such as number of individuals, population density, or area occupied by the genetic resource. For cultivated varieties, may include number of farmers or fields and area under cultivation.	Critical for *in situ* CWR (counts of wild plants, population census) and on-farm conservation (e.g., how many households grow a landrace). Used to monitor viability and risk of loss.	International Union for Conservation of Nature (IUCN) Red List criteria use population size/trend as key metrics; on-farm monitoring guidelines count farmers/area as proxies for landrace populations.	Typically, numeric data in monitoring reports or databases; often time-series in spreadsheets to track changes.	**NO.** Similar to ex situ curatorial-related datasets, in situ management datasets, which are primarily used on the ground by conservation practitioners and local communities, are not covered
Population Structure & Demography Data	Data on population structure and dynamics, such as age classes, life stages, regeneration status, or sex ratios in wild populations; for landraces, generational turnover or distribution across communities. Indicates reproductive health and stability of populations.	Primarily for wild *in situ* populations (e.g., age or size class distribution, seedling vs. mature plant counts); sometimes considered for on-farm (e.g., continuity of a variety across generations of farmers).	IUCN assessment guidelines include population structure as an indicator of trends; specialized monitoring protocols in conservation biology for demographic studies.	Collected via field surveys (e.g. tagging and measuring individuals) and stored in research databases or spreadsheets. May be visualized in demographic models; no single exchange standard beyond structured survey data.	
Threats & Risk Assessment Data	Information on threats to the population or variety and its risk status. Includes types of threats (habitat loss, climate change, replacement by modern varieties, etc.), their severity, and any risk	Relevant to both wild and on-farm contexts: e.g. habitat degradation and overharvesting for wild plants, or socio-economic pressures causing farmers to abandon a traditional	IUCN Threat Classification Scheme and Red List categories for wild species; adapted threat assessment frameworks for landraces (applying IUCN-like criteria to on-farm populations);	Often recorded as qualitative codes or narratives in surveys and status reports. Interoperable if using standardized threat categories (IUCN threat codes) in databases; can be linked to conservation status assessments.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
	evaluation (e.g. extinction risk or genetic erosion level).	variety. Helps prioritize conservation actions.	FAO indicators of genetic erosion.		
Habitat & Site Condition Data	Description of the habitat or environment where the PGR occurs. For wild populations: ecosystem type, vegetation, soil, climate, disturbance levels; for on-farm: agro-ecosystem setting, cropping system, soil fertility, etc. Also notes on habitat quality or degradation.	Applies to *in situ* sites of CWR and wild food plants (natural habitat profiles) and traditional farming systems (e.g. terraced fields, home gardens). Provides context for survival requirements and threats.	Ecological classification systems (e.g. EUNIS habitat types); Darwin Core "habitat" term for free-text descriptions; FAO ecogeographic survey descriptors for environmental conditions.	Often captured in field survey forms or GIS-linked databases. May include habitat codes or textual descriptions; can be integrated with GIS (environmental layers) for analysis.	
Associated Species & Ecological Interactions Data	Records of other species and ecological interactions relevant to the target PGR. Includes associated flora and fauna (e.g. neighboring plant species), pollinators, seed dispersers, symbionts, pests or diseases. Helps understand ecological dependencies (for example, a wild relative's obligate pollinator).	Mainly for wild populations (documenting ecosystem context, e.g. pollination or competition). In on-farm systems, can note companion crops or common pests affecting landraces.	Documented via ecological surveys; Pollination syndromes or pest/disease incidence might follow agricultural survey standards.	Stored as field notes or database links (species IDs of associated taxa). Interoperable through biodiversity databases or pest/disease databases if standardized names are used.	
Reproductive Biology Data	Details on the reproductive system and biology of the plant population: e.g. mating system (self vs. cross-pollinating), flowering/fruiting phenology, seed dispersal mechanism, and seed viability *in situ*. For cultivated varieties,	Important for both CWR and landraces: understanding reproduction informs management (e.g. need for pollinator conservation if outcrossing, or seed renewal practices if saved	Often recorded in species biology accounts (e.g. IUCN Red List species info includes breeding system and phenology); crop-specific guides document breeding system of traditional varieties.	Typically, narrative or categorical data in documentation; not a standardized exchange format, but Darwin Core notes can capture phenological state. Usually stored in reports or databases alongside other population data.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
	includes farmers' propagation practices (vegetative propagation, seed saving interval).	by farmers). Often noted in species profiles or community knowledge.			
Site Protection & Status Data	Information on the legal or conservation status of the site/population. Notes if the population is inside a protected area, designated genetic reserve, or other conserved area, and any official status (e.g. IUCN protected area category, UNESCO heritage site).	Mainly for wild populations (e.g. CWR located within national parks or dedicated genetic reserve sites). For on-farm, could include if the area/community is part of a recognized conservation program or a custodian farmer network.	IUCN Protected Area categories and WDPA codes for sites; concept of "Genetic Reserve" (specific *in situ* site for target taxa) in national CWR strategies. National PGR programs provide frameworks for *in situ* site designation.	Recorded in conservation databases and reports (site name, coordinates, protection status).	
Land Tenure & Management Regime Data	Details on land ownership and management at the site. Indicates whether the habitat is on public land, private farm, community land, etc., and the land use regime (e.g. nature reserve, agricultural land, communal grazing land). Provides context for access and management authority.	Relevant to all *in situ* sites, e.g. distinguishing a wild population on communal land vs. one in a state reserve, or a landrace maintained on family-owned farm vs. community fields. Influences how conservation actions are implemented.	Often noted per site in conservation plans (e.g. protected area management plans, community land agreements). Some descriptor sets include a field for land management type.	Stored as part of site metadata. Sometimes linked to GIS layers indicating land ownership boundaries.	
Conservation Interventions Data	Documentation of conservation and management actions applied. For wild sites: actions like habitat restoration, fencing, controlled burns, invasive species removal, population	All *in situ* contexts, captures what is being done to actively manage and protect the PGR population. Used to evaluate and adjust	IUCN Conservation Actions Classification (a standard list of action types) can be used for wild species (e.g. action codes for site management).	Recorded in management plans and activity logs, often as text descriptions or coded action types.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
	reinforcement, etc. For on-farm: interventions such as training in seed selection, provision of diversity seed kits, improved storage facilities, or community bylaws for conservation.	management strategies over time.			
Monitoring & Trend Data	Time-series observations and indicators that track changes in the PGR population or variety over time. Includes repeated surveys of population size, genetic indices, number of cultivating households, or diversity scores, allowing calculation of trends (increasing, stable, declining).	Used in both wild and on-farm monitoring schemes, e.g. annual population counts of a CWR, or periodic surveys of how many landraces and farmers persist in a community. Enables early warning of genetic erosion or recovery.	Monitoring frameworks in conservation biology (long-term ecological monitoring protocols); Often aligned with CBD indicators (e.g. SDG 2.5.1 on genetic diversity).	Data typically stored in spreadsheets or databases with time-stamped records.	
Seed Exchange & Seed System Data	Information on the social system of seed supply and exchange. Records how farmers or communities obtain, share, and renew seed of the PGR: e.g. seed saving practices, exchange networks, participation in seed fairs, frequency of introducing outside seed.	Specific to on-farm conservation of cultivated materials. For landraces, resilient informal seed systems (farmer-to-farmer exchange, local markets, community seed banks) are key to maintenance. Also, if wild plant seeds are harvested for use, notes on community seed sharing or sale are included.	FAO seed system assessment frameworks differentiate formal vs. informal seed systems in national reports. Social network analysis methods are sometimes used to model farmer seed exchange networks.	Data collected via farmer surveys, interviews, and community mapping. Often qualitative with some quantification (who exchanges with whom, how often). Represented in databases or network diagrams; not standardized globally, usually shared as case study data or reports.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
Community Seed Bank & Local Repository Data	Data from community-managed seed banks or biodiversity registers. Includes inventories of local varieties conserved (entries with variety name, origin, quantity of seed, viability), membership information, and records of seed distribution or regeneration.	Applies to on-farm/community contexts where local seed banks or community repositories are established. Complements *in situ* field conservation by providing a backup and accessible source of seed locally.	Community seed bank guidelines (e.g. by Bioversity International) outline standard data to record (passport info, storage conditions, viability tests). Often aligned with genebank standards (using MCPD fields adapted to community context).	Typically maintained as ledger books or simple databases by the community or NGOs. Data can be exported in spreadsheet format; interoperability with national genebank systems is possible if using compatible data fields (allowing integration or safety duplication of samples).	
Traditional Knowledge & Cultural Data	Qualitative data on local knowledge and cultural practices related to the PGR. Encompasses farmers' knowledge of cultivation techniques, selection criteria, folk taxonomy, culinary preparation, storage methods, and cultural/religious significance of the landrace or wild plant. Often captured through ethnobotanical surveys and oral histories.	Critical for on-farm conservation (documenting how and why farmers maintain landraces) and useful for wild plant conservation (indigenous knowledge on wild food uses and habitat management). Preserves intangible heritage associated with PGRFA.	Adheres to Traditional Knowledge documentation guidelines (e.g. requiring Prior Informed Consent and respecting community protocols). Uses tools like community biodiversity registers and farmer-developed variety catalogues (with local descriptors) . Also guided by frameworks under the CBD	Recorded via interviews, participatory workshops, audio/video recordings, and written narratives. Stored in qualitative databases or community archives. Interoperability is limited (due to sensitivity and context), but summaries can be linked to scientific data (with consent) or archived in digital libraries for preservation.	
Ethnobotanical Use Data	Specific data on the uses of the PGR by local people. Details how the plant or its products are utilized, e.g. culinary uses (recipes, food products), medicinal uses (traditional remedies), fodder,	Relevant to both cultivated and wild PGR: e.g. documenting traditional dishes made from a landrace, or medicinal/cultural uses of a wild relative. Provides	Ethnobotanical data collection methods (e.g. economic botany descriptors of use categories) are used to systematically record uses. Some national PGR	Collected through interviews and participant observation. Data stored in databases or reports with categorical fields (food, medicine, etc.) and narrative details. Interoperability via standardized use categories when integrating	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
	fiber, or other material uses, and associated preparation/storage methods.	insight into the resource's local value and supports benefit-sharing arguments.	inventories include use categories for landraces. Traditional knowledge frameworks (e.g. WIPO toolkit) emphasize capturing use information with proper community attribution.	into larger PGR databases or knowledge portals.	
Farmer Variety Description Data	Descriptions of landraces or traditional varieties from the farmer's perspective. Captures the traits farmers observe and value, e.g. drought tolerance, flavor, cooking quality, yield stability, often in farmers' own terminology. May include origin stories or how the variety has been maintained over generations.	Applies to on-farm conserved germplasm (farmers' varieties/landraces). Supports participatory cataloguing of diversity, ensuring that farmer-identified characteristics are recorded (complementing scientific characterization).	Traditional variety catalogues and community registers often use farmer-friendly descriptors. Programs implementing Farmers' Rights (ITPGRFA Art. 9) encourage documentation of farmer varieties in registries with their distinct traits and names.	Documented in written or digital catalogues (often bilingual: local language and scientific terms). These records are usually stored in local databases or reports intended for community and national use.	
Community Agreements & Protocols	Locally developed agreements or protocols that govern the conservation and use of PGR *in situ*. Examples include community conservation agreements (where communities commit to maintain certain landraces or protect wild populations) and biocultural community protocols outlining rules for access and benefit-sharing.	Usually at community level in on-farm projects or in indigenous territories with wild PGR. They formalize roles and responsibilities, e.g. a village agrees to conserve a traditional crop and not share its seed outside the community without consent, often in exchange for support or benefits.	Aligned with legal frameworks like the Nagoya Protocol (which recognizes community protocols in ABS). Often facilitated by NGOs or government programs following established guidelines.	Exist as written documents (signed agreements, protocol booklets). Not a "data" format per se, but a key conservation output kept on file (PDFs, print copies). Some metadata (e.g. existence and date of agreement) may be noted in project databases.	

Table 4. Research-oriented data types: Description, Scope, Standards and Formats

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
Genomic data	Genetic sequence data and molecular markers related to PGRFA. This ranges from discrete markers (like SSR or SNP profiles of accessions) to high-throughput DNA sequences (whole genome sequencing, resequencing, gene sequences, genomic libraries). Also includes any characterized genetic variants (SNPs, insertions, etc.) and sometimes epigenetic data (DNA methylation, etc.).	Allows in-depth analysis of genetic diversity and relationships. Genomic data is used to assess genetic variation within and between accessions or wild populations, identify genes or alleles for important traits, and guide breeding (marker-assisted or genomic selection). For conservation, it can reveal genetic erosion or structure in wild populations (informing how many samples to conserve). In the era of Digital Sequence Information (DSI) debates, genomic data of PGRFA also has policy significance.	Rich set of standards: FASTA format for DNA sequences; GenBank/EMBL/DDBJ accession standards for sequence submission; VCF (Variant Call Format) for SNPs/variants. The Genomics Standards Consortium provides MIGS/MIMS/MIXS (Minimal Info for Genomic/Metagenomic Data) ensuring metadata like source organism and methods are recorded. Data often deposited in international repositories (NCBI, EMBL-EBI, etc.) which enforce their formats.	Highly structured and very interoperable within bioinformatics, sequences and variant data are easily exchanged in standard formats, and large databases index them. However, interoperability with traditional PGR databases is still developing (often requires manual linking via identifiers). Genomic data files are stored outside genebank systems, but metadata bridges (like a field noting “BioSample ID” for an accession’s sequence) can be used	**No.** Although these advanced data types lie largely beyond EURISCO’s current remit, there is a clear vision to link or interface them in the future, as outlined in Objective 2 of this deliverable.
Transcriptomic/ Proteomic/ Metabolomic data	High-dimensional “omics” data capturing gene expression (transcriptomes), protein profiles (proteomes), and biochemical compounds (metabolomes) of PGRFA	Sheds light on functional diversity and trait mechanisms in PGRFA. For example, transcriptomic data can identify stress-responsive genes in a wild relative;	Each omics field has its own standards: MIAME for microarray gene expression and MINSEQE for sequencing-based transcriptomics ensure	The data are typically large matrices or sequence files (for transcripts) and are stored in specialized repositories (GEO for gene expression, PRIDE for	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
	samples. These are sometimes called molecular phenotypes, as they reflect functional traits at the molecular level (e.g., which genes are upregulated under drought, what metabolites a plant produces).	metabolomic profiles can reveal nutritional or flavor compounds in a landrace. This data is mainly used in research to connect genotype to phenotype and to discover novel genes or compounds. It's less directly used in routine genebank management, but it adds value to accessions by characterizing them beyond the genome. It can guide breeders to candidates for traits.	consistent metadata; proteomics has standards like MIAPE and repositories (e.g., PRIDE); metabolomics has MIAMET guidelines and databases like Metabolights. Controlled vocabularies and ontologies are used to annotate experiments. These are generally not PGRFA-specific but science-wide standards that PGRFA researchers adopt.	proteomics, etc.). Interoperability within each domain is good (common formats like FASTQ for RNA-seq, mzML for metabolite spectra). However, integrating these with other PGR data requires good metadata: experiments must reference the plant material (accession ID, voucher, or DOI). If that link is made, then cross-dataset integration is possible via databases or analytical platforms, but this remains a challenge.	
High-Throughput Phenotyping (HTP) & Image Data	Massive phenotypic datasets collected via automated systems and sensors, for example, drone or satellite imagery of field trials, time-lapse photos in a phenotyping chamber, or multispectral/ hyperspectral sensor readings. These provide quantitative measures of plant traits (height growth curves, canopy temperature etc.) at	Revolutionizing how germplasm is evaluated, by enabling screening of large collections for traits like drought tolerance or growth rate with minimal labor. For PGRFA, HTP can be used in common gardens or field genebanks to characterize collections in detail, or in breeding nurseries to pick out promising genotypes. It generates far more data points per accession	Still emerging: MIAPPE provides an overall framework for recording phenotyping experiment metadata (so HTP experiments should use MIAPPE to describe setup and variables). For the data streams themselves, there are some efforts and specialized databases to handle HTP data.	Data format is often heterogeneous: thousands of images (terabytes of data) plus derived numeric features. Typically stored in dedicated data lakes or databases at research institutions. Interoperability is limited by size and complexity, usually only derived results (like processed trait values) are shared. However, when those results are distilled	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
	scales and frequencies not feasible manually.	(sometimes thousands of data points or images per plant), offering a deeper phenotypic profile. This helps in discovering subtle trait differences and doing genome-to-phenome studies.		(e.g., "accession X had 20% higher canopy temperature under heat stress"), they can be incorporated into traditional databases. APIs like BrAPI are being extended to handle links to image data and large-scale trait data, but this is a frontier area.	
Pre-breeding & Breeding data	Information from breeding programs that have utilized PGR, especially wild or unadapted materials. This includes pedigrees (family trees of crosses), selection histories, trial data of breeding lines, and data on introgressed traits (e.g., a wild donor of a disease resistance gene and the lines derived from it). Essentially, it tracks the flow of genes from conserved material into improved germplasm.	Illustrates the utilization pathway of genetic resources. By tracing breeding data, genebanks and researchers can see how an accession contributed to new varieties or research lines. Pre-breeding data often involves intermediate lines (like bridging crosses to transfer a trait from a wild species into a usable form). Having this data helps avoid duplication of efforts (knowing a trait is already transferred from a certain CWR) and ensures credit sharing. It's mainly relevant in research and plant breeding contexts, but genebanks increasingly want	There are community-driven standards: BrAPI has been increasingly adopted schema/API by breeding database systems to exchange data (covers trials, genotypes, pedigrees). The International Crop Information System (ICIS) and its successors provided data models for pedigrees and breeding workflows. No global mandate, but CGIAR and others have breeding data management platforms that standardize how crosses and trials are recorded.	Usually housed in breeding management systems (like IBM Breeding Management System or breeder's custom databases). Data is structured (each cross, each generation with identifiers, each trial with plot-level data). With BrAPI, interoperability is improving, e.g., one can query a variety's pedigree or performance from another program's database if both use BrAPI. However, these systems are separate from genebank databases. Integration requires mapping an accession's ID in the genebank to its ID in breeding programs.	

Data Type	Description	Scope/Context	Standards/Frameworks	Format/Interoperability	EURISCO coverage
		this feedback to demonstrate impact.			

As noted in Section 3.1, EURISCO maintains a targeted, purpose-driven scope that highlights essential standardized descriptors that are necessary for genebank personnel, breeders, researchers, community workers and other stakeholders to discover and request PGR accessions. In emphasizing discoverability, EURISCO consciously excludes more specialized or granular curatorial information that, although vital to day-to-day conservation efforts, have limited utility for broader user queries. A short-lived drop in germination rates, for example, signals an operational priority to genebank curators but would not necessarily guide a breeder seeking accessions for drought-resistance trials. Retaining these operational data in genebank-specific systems ensures they remain accurately interpreted and up to date, while EURISCO provides a clear, succinct resource for external users.

A parallel logic governs the treatment of in situ conserved populations, which can be subject to rapid and unpredictable changes. While passport data, including custodial details, for select in situ CWR populations do appear in EURISCO, the dynamic, context-specific management information central to their ongoing stewardship remains outside its scope. In situ populations can experience abrupt changes in population structure, habitat quality, or threat levels, driven by environmental factors, local agricultural practices, or community-based decisions. Capturing and revising these fluctuating details in a global catalogue would not only burden the system with constant updates, but also risk disseminating partial or outdated information. Consequently, such records, covering annual monitoring, risk assessments, and interventions, are typically housed in a patchwork of local or national databases, community-led archives, or project-specific repositories. Nevertheless, there is a recognized priority to establish a consolidated, taxon-level inventory of in situ CWR and on-farm conserved landraces, so that a wider community of potential users gains a clear, centralized overview of what is being conserved in situ and can identify where to direct initial inquiries.

BOX 1. Role of EURISCO: End-User Data vs. Management Data

EURISCO's core mission has been to provide an information portal for well-managed, accessible germplasm samples (accessions *(ex situ)* and populations *(in situ)*) held in Europe and adjacent regions . By design, EURISCO focuses on passport data (what the accession/ population is, where it's kept, origin, etc.) and also phenotypic data from characterization and evaluation trials. It deliberately does not handle day-to-day genebank management details because those are outside its scope of informing users about available diversity. Likewise, EURISCO was never intended to be a comprehensive system for all PGR-related data in Europe. Attempting that would dilute its effectiveness. Instead, EURISCO excels as a central entry point for users (breeders, researchers, policymakers etc.) to find genetic resources that can be requested for use.

The exclusion of other data types, i.e., curatorial and management data, does not mean they are unimportant. On the contrary, they are critical for national program managers, institute curators, and conservationists. To clarify this distinction:

- **<u>End-users (breeders, researchers, educators, policymakers)</u>** typically use EURISCO to find genetic materials with certain traits or from certain origins. They are interested in what species/varieties exist, where they are conserved, and some traits or categorical info. They usually do not need to know how the genebank keeps it alive or how many farmers grow it. They just need access to the diversity. For example, a breeder searching for drought-tolerant barley landraces might query EURISCO by country or environment of origin. They might value a note that says "traditional variety from dry region X" but wouldn't benefit from seeing the seed's last moisture content reading or that only 5 farmers still grow it. Those details don't directly inform breeding decisions (and could be distracting or misinterpreted). The focus for end-users is on genetic and phenotypic information that helps select materials.
- **<u>Managers (genebank curators, national coordinators, conservationists)</u>** have the opposite need: they often know what they have, but need data to ensure they maintain it effectively. For them,

viability drop from 95% to 80% is an alarm bell to regenerate seeds; knowing only 5 farmers still grow a landrace is a call to launch a support program; tracking distribution requests can show which accessions are in high demand (and perhaps should be multiplied more). These data help guide decision-making for conservation priorities and resource allocation. They might also be needed for reporting purposes, e.g. countries reporting to FAO on the state of their PGRFA will include number of accessions regenerated, threat statuses of landraces, etc. Such stakeholders often operate with internal databases or spreadsheets, and share summaries in reports (like National Reports for the State of the World's PGRFA). They do not necessarily want all this raw data public, but they need it in-house.

5. Inventory of PGR Information Not Yet Represented in EURISCO

5.1 Geographic and Institutional Gaps

EURISCO currently aggregates substantial data, comprising 2,102,951 accessions in total as of February 2025. This includes 2,096,567 *ex situ* accessions from 417 collections and 6,384 *in situ* CWR populations, representing 6,753 genera and 45,424 species across 43 countries. Notwithstanding, notable geographic and institutional gaps remain in the current representation of PGR information within the system. Numerous institutions involved in conservation, research, education, breeding, community development or grassroots initiatives, hold valuable germplasm collections that are currently either underrepresented or entirely absent from EURISCO. This section provides a systematic inventory of these intra-country gaps, pinpointing specific regions and institutions whose PGR data are not yet represented (Table 5). The present inventory was compiled based on the collections and data sources that became readily visible or accessible during the course of the project. As a result, certain collections (such as working collections maintained by research institutes, breeding stations, or university departments, botanical gardens and arboreta with CWR collections) were included when we encountered them in the literature, through web searches, or via direct correspondence with NFP and curators. This pragmatic approach allowed us to capture an initial and meaningful snapshot of relevant genetic resources, although it did not systematically screen every possible working collection in the region. Clearly documenting these gaps establishes the groundwork for targeted future initiatives aimed at enhancing EURISCO's comprehensiveness and reliability.

It is important to recognize, however, that the ultimate decision to incorporate these institutions rests at the national level. In addition to the factors covered in Section 4.3, several considerations may limit the inclusion of certain collections in a country's NI, including legibility, legal or jurisdictional constraints, regulatory barriers, unclear or restricted germplasm access, and institutional restrictions. Nevertheless, countries that see strategic value in bringing additional legible collections into their NI, whether for improving overall data visibility and access, strengthening conservation efforts, ensuring compliance with the ITPGRFA, or enhancing local and collaborative research and breeding, are more likely to take the necessary steps for inclusion. By highlighting existing gaps and underrepresented institutions, this analysis encourages stakeholders to consider, where appropriate, ways to integrate these collections into their NI, and, subsequently, EURISCO.

BOX 2. Defining Plant Genetic Resources in EURISCO: Scope, Coverage and Proposed Considerations

In contextualizing its scope, EURISCO builds on a broad definition of Plant Genetic Resources (PGR) drawn from the Convention on Biological Diversity (CBD) and the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA), encompassing any plant-derived genetic material that holds present or potential value for food and agriculture, including cultivars, landraces, farmers' varieties, breeding lines, genetic stocks, and research material. Its core data structure is provided by National Inventories (NIs), which individual European countries compile and maintain in accordance with their commitments under the ITPGRFA, the FAO Global Plan of Action, the

Clearing-House Mechanism (CHM) of the CBD, and the Global Biodiversity Information Facility (GBIF). These NIs typically include a wide range of materials such as food crops, forages, wild and weedy species, and medicinal plants. EURISCO primarily features germplasm managed within Europe, including holdings in overseas territories and extraterritorial research sites belonging to European countries. In certain cases, species not strictly classified as PGRFA may also be included if they are deemed critical for advancing agricultural or scientific research. All such materials are considered "potentially available" for research, breeding, and conservation, assuming they meet acceptable standards of documentation, stewardship, and accessibility.

Proposed considerations when including certain collections:

(1) **Working Collections**: typically maintained by research institutes, breeding stations, or universities and may comprise breeding lines, single-seed descent (SSD) materials, near-isogenic lines, and other specialized genetic materials. While they are not always designated as formal *ex situ* collections, they are frequently the primary source of in-depth characterization and evaluation data, including HTP, genomic, transcriptomic, proteomic, and metabolomic (omics) analyses.
Include or prioritize the inclusion of a working collection in the NI only if:
  - Uniqueness or non-redundancy. The collection harbours germplasm that is not already conserved in a national or international genebank, thereby contributing novel diversity.
  - The materials have established or potential value for plant breeding, genetic research, or conservation (e.g., unique traits, novel alleles).
  - Breeding lines in question are actively utilized in research projects, indicating that they could be of interest to the wider community.
  - Minimum data standards are met. Sufficient passport data and any available pedigree information are provided in a consistent format.
  - The collection can be shared without violating intellectual property agreements or other confidentiality constraints.
  - The hosting institution demonstrates sufficient infrastructure (e.g., secure storage, regeneration protocols, staffing) to ensure the collection's long-term physical maintenance.
  - A commitment exists to keep the lines available for distribution under appropriate Material Transfer Agreements ((S)MTA) or equivalent mechanisms, ensuring that researchers or breeders can actually access these resources if needed.
  - An appropriate workflow (e.g., standard operating procedures for data updates) is in place to handle the dynamic nature of working collections efficiently.

(2) **Non-PGRFA materials**, including model organisms (e.g., *Arabidopsis thaliana*), ornamental species, medicinal plants, and other taxa not traditionally regarded as crops for food or fodder.
Include or prioritize the inclusion of non-PGRFA species in EURISCO only if:
  - Demonstrable or potential agricultural/research value. The species displays traits or genetic insights potentially applicable to crop improvement, pathology, stress tolerance, or other agricultural objectives.
  - There is documented usage in breeding, genetic mapping, or foundational research that underpins improvements in PGRFA.
  - The material aligns with broader interpretations of PGR under the CBD or ITPGRFA, which acknowledge any plant-derived material of value for agriculture or research.
  - Inclusion does not conflict with EURISCO's core focus on food and agriculture, but rather complements it (e.g., bridging research on ornamental/medicinal species to food crops).

- Taxonomic, ecological, or genetic information can be recorded in a way that fits existing EURISCO data fields or adaptable templates.
- The institution responsible for these non-PGRFA materials can provide complete, standardized passport data.
- The relevant authorities support or do not object to the inclusion of these species (e.g., species subject to special permits).
- The inclusion of non-PGRFA does not lead to excessive dilution of the system's primary mandate.
- A clear framework is in place to prioritize non-PGRFA materials with the greatest relevance to plant breeding, genetic research, or broader agricultural systems.

(3) **Botanical gardens and arboreta** maintain living plant collections primarily for conservation, research, education, and public engagement. Many hold diverse taxa, including rare species, local landraces, wild relatives of crops, and potentially non-PGRFA species. However, the breadth and heterogeneity of these holdings require careful evaluation when considering their incorporation into EURISCO.

Include or prioritize the inclusion of a botanical garden or arboretum collection in EURISCO only if:

- The collection contains species or accessions with present or potential value for food, agriculture, or breeding (e.g., crop wild relatives, traditional varieties, medicinal or ornamental species known to improve crop traits).
- The garden or arboretum maintains consistent and sufficiently detailed passport data, including taxonomic identification, provenance, and any relevant phenotypic or genotypic descriptors.
- A system (e.g., a living collection database) is in place to update records over time as plants are added, removed, or relocated within the facility.
- Regeneration or repropagation protocols exist for species that require periodic renewal to ensure they remain viable and accurately identified.
- The garden or arboretum is willing and able to distribute plant material (e.g., seeds, cuttings, grafts) under appropriate (MTAs) or similar guidelines.
- The garden or arboretum meets regulatory obligations related to plant collection, introduction, and exchange (e.g., CITES for endangered species, the Nagoya Protocol on access and benefit-sharing).
- No unresolved intellectual property, ownership, or provenance issues that would conflict with open documentation in EURISCO.
- The NI or equivalent authority supports the inclusion of the botanical garden/arboretum collection, recognizing its importance for national PGR strategies or biodiversity conservation objectives.

(4) Geographical Coverage (overseas or extraterritorial materials)

- The collection is officially managed or supported by a European country's national program, or by an institute under European jurisdiction.
- International treaties and relevant bilateral agreements do not prohibit the documentation or exchange of these materials.
- The overseeing entity in the overseas territory is prepared to adhere to EURISCO data standards and protocols.
- As with in-country collections, a minimum threshold of passport data and maintenance records exists.
- Materials from overseas territories bring distinct genetic diversity (e.g., from different agroecological zones) that complements European-based collections.

Table 5. Institutions Holding PGR Collections Not Yet Represented in EURISCO

WIEWS CODE	Institute	Custodian/ Institution Type	Conservation Focus	Notes
ALBANIA				
ALB019	Agricultural Technology Transfer Center – *Lushnja*	Governmental (public research institution)	Field crop landraces (wheat, beans, vegetables)	~8,000 accessions, considered a working collection (source: Ibraliu, 2018)
ALB022	Agricultural Technology Transfer Center – *Shkodra*	Governmental (public research institution)	Regional crop diversity (e.g. maize, vegetables	~700 accessions, working collection maintained for regional research
ALB023	Agricultural Technology Transfer Center *Fushë Kruja*	Governmental (public research institution)	Field crops (cereals, forages)	Working collection
ARMENIA				
ARM010	Armenian Botanical Society	NGO/ CSO	Endemic wild plants and local heirloom varieties; community seed bank	145 accessions (Available in Genesys c/o CWR project); not officially part of national PGR program
ARM011	Scientific Center of Viticulture, Fruit-Growing & Winemaking (Nalbandyan, Armavir province)	NGO	Field collections: vineyards, orchards (field genebank); Indigenous grape cultivars	~300–400 grapevine accessions (wild V. vinifera sylvestris and traditional cultivars); dozens of fruit tree varieties (apricot, pear, etc.); Passport information for some accessions are found in the *European Vitis Database*
	Green Lane Agricultural Assistance (Yerevan; community seed banks in Kotayk, Tavush, etc.)	NGO	Traditional farmer varieties and rare landraces (e.g. beans, tomato, pepper, cereals); medicinal and culinary herbs from local gardens	
AUSTRIA				
AUT031	University of Natural Resources and Life Sciences (BOKU) – Dept. of Crop Sciences	University (academic collection)	Breeding lines and research materials (especially cereals, pulses, alternative crops)	Materials are used for academic studies on agronomy, stress tolerance, etc.; data usually remain in departmental databases or project-specific archives.

AUT046	Arche Noah, Schiltern	NGO/CSO – Community seed bank & living collections	Heirloom vegetable varieties, local cereals, and fruit trees (orchards). Emphasizes local Austrian heritage crops and endangered cultivars.	More than 6500 accessions. Maintains its own seed registry, with partial data shared through local networks. Currently not included in the NI due to separate NGO–government data arrangements and a preference for community-based seed exchange rather than formal genebank integration (including use of SMTA).
AZERBAIJAN				
AZE017	Central Botanical Garden, Baku	Governmental (public research institution)	Collection consisting of about 2500 species, forms and varieties of native and exotic trees, shrubs and herbaceous plants from the flora of Azerbaijan and different botanical and geographical areas of the world	
AZE095	'ARAZ' Science-Production Unit, Nakhchivan	Governmental, under the Ministry of Agriculture of Nakhchivan Autonomous Republic	Wheat, barley, alfalfa, apple, pear, quince, fig, walnut, hazelnut, apricot, alycha, plum, sweet cherry, cherry, pomegranate and other fruits and berries	
AZE100	Nakhchivan Bioresources Institute	Governmental (public research institution)	More than 1000 seed accessions	
BELGIUM				
BEL017	Department of Plant Production, Universiteit Gent (University of Ghent)	University	broad living plant collection for research and education; Working collection of cereals (e.g., Triticale)	
BELARUS				

BLR012	Department of Genetics, Faculty of Biology, Belarussian State University, Minsk	University (academic collection)	Lupinus spp. with ~33 accessions; Triticosecale and minor crop collections	
BLR014	Institute of Experimental Botany of the National Academy of Sciences of Belarus, Minsk	Governmental (public research institution)	Seed bank and living collections (indigenous or historically significant to Belarus (e.g., rare forest herbs, medicinal plants, or regionally important wild relatives of crops)	
BLR015	Republican Unitary Enterprise 'The Institute of Flax', Uste, Orsha district, Vitebsk Region	Governmental (public research institution)	Fiber flax (*Linum usitatissimum*) and allied species.	
BLR016	Republican Unitary Enterprise 'Research and Practical Center of the National Academy of Sciences of Belarus for Potato, Fruit and Vegetable Growing', Samokhvalovichi, Belarus	Governmental (public research institution)	Potato landraces and breeding lines; also holds some vegetable and fruit genetic resources	~1,500 (including 855 potato accessions – around 55% of the region's potato germplasm).
BLR017	Republican Unitary Enterprise 'Institute for Fruit Growing', Skiershchyna	Governmental (public research institution)	Fruit tree and berry germplasm (Malus, Prunus, Ribes, etc.,)	~1,000 accessions (as of 2022)
BLR018	Republican Unitary Enterprise 'Research station of sugar beet of the National Academy of Sciences of Belarus	Governmental (public research institution)	Beta vulgaris (sugar beet) breeding lines and wild beet relatives	

BLR019	Central Botanical Gardens of the National Academy of Sciences of Belarus, Minsk	Governmental (public research institution)	Wild plant species, including crop wild relatives and medicinal plants.	
BLR021	Institute of Vegetable Growing, Samokhvalovichi	Governmental (public research institution)	Vegetable crop landraces and breeding lines (brassicas, cucurbits, solanaceous crops).	
BLR026	The Polessye Institute of Plant Growing, Polessye Region		Genetic resources of crops suited to Polessye region (e.g., sorghum, maize, other cereals).	
BLR029	Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk	Governmental (public research institution)	Genetic research, including *ex situ* collections of mutant lines and genetic stocks for crops like barley or tomato	
BOSNIA AND HERZEGOVINA				
	Faculty of Agriculture and Food Sciences (FAFS), Univ. of Sarajevo – Butmir (Ilidža), Sarajevo, Federation BiH. Includes Genebank Butmir (Center for Plant Genetic Resources).	University; Federation-run genebank	Seed genebank; field experimental plots. Crop landraces (cereals, legumes, vegetables); some medicinal/industrial crops.	Historically operated separately under Federation institutions.
	Federal Agro-Mediterranean Institute (FAZ) – Mostar, Federation BiH.	Independent federal administrative organization	Field collections (orchards, vineyards); figs, pomegranates, almonds; local vegetable and herb seeds.	Operates under Federation ministry; not formally linked to the national PGR program.
	Biotechnical Faculty, University of Bihać – Bihać, Federation BiH.	University	Working collections (trial fields, gardens). Crop landraces suited to	Collections kept for academic use (research and teaching)

			Northwestern BiH (vegetables, maize, potato)	
	Faculty of Agriculture, University of East Sarajevo – East Sarajevo/Pale (Republika Srpska).	University	Field collection (orchard) and seed samples. Fruit tree landraces from Eastern Bosnia; some cereals.	Not part of the Genetic Resources Institute genebank structure. Its PGR activities are project-based (education, extension)
	Sjemenčica – Community Seed Bank – Nationwide (coordinated by NGO “Nešto Više”; initial base in Mostar).	NGO/CSO	Seed library (community-managed, distributed to gardeners; Vegetables, herbs, medicinal and aromatic plants, and some grains – a broad range of traditional garden crops from across BiH. Emphasis on farmer varieties and rare local cultivars.	More than 1,500 accessions (heirloom samples) of various crops
	Alica Foundation – Balkan Seed Network (BSN) partner, Banja Luka	NGO/ CSO	Seed collections (heirloom varieties, saved by members); seed exchange network. Traditional field crops and vegetables (e.g. old corn, beans, peppers, etc.) mainly from rural Banja Luka/Krajina areas.	Part of the broader Balkan Seed Network promoting on-farm conservation. Operates as an advocacy and networking body. PGR holdings are diffuse (with farmers)
BULGARIA				
	Institute of Agriculture, Kyustendil	Governmental (public research institution)	Field genebanks (orchards) and *in situ* conservation plots;rich collections of small fruit crops (berry collections) and stone fruits; Temperate fruit cultivars and rootstocks, including landraces (e.g. the famous	The NI submitted to EURISCO did not include field-grown clonal repositories, as evidenced by no Malus/Pyrus accessions in the NI listing. These fruit collections are maintained by crop-specific institutes under the Agricultural Academy, which may not have provided passport data to IPGR Sadovo.

			‘Kabyle’ plum) and breeding selections.	
	Institute of Fruit Growing-Plovdiv	Governmental (public research institution)	Field genebanks; Fruit cultivars and landraces, e.g., apples, plums, cherries, apricots traditionally grown in Bulgaria, conserved as trees.	
	Viticulture and Enology Institute (Pleven)	Governmental (public research institution)	Vine field genebank; known to have an ampelographic collection	Similar to fruit trees, the grape collection data wasn’t part of the EURISCO submission.
	Maritsa Vegetable Crops Research Institute (Plovdiv)	Governmental (public research institution)	Working collections (field and seed) of vegetable landraces (Capsicum, Solanum) and breeding lines	
	Maize Research Institute, Knezha	Governmental (public research institution)	Seed collection (working); Bulgarian maize landraces and breeding lines adapted to local conditions	
CROATIA				
	Institute for Adriatic Crops, Split	Governmental (public research institution)	Field genebank and seed accessions; Mediterranean crop collections, e.g. Dalmatian olive groves, fig collection, almond and subtropical fruits; local vegetable landraces adapted to coastal climates.	Historically, Croatian PGR efforts focused more on continental crops via Osijek and Zagreb. The Split institute’s collections (especially clonal tree crops) may not have been fully documented in the NI.
CZECH REPUBLIC				
CZE011	Faculty of Horticulture in Lednice, Mendel University of Agriculture and Forestry, Brno	University	Working collections of various horticultural crops, including fruit trees, grapevines, medicinal	Member of Czech National PGR programme

			plants, perennial vegetables, and selected ornamental flowers.	
	Research and Breeding Institute for Pomology, Holovousy	Private Research Institute	Collection of apple landraces additional to that included in the documentation system GRIN Czech/EURISCO. Collection of Ribes from Sudet region	Member of Czech National PGR programme
	Gengel Seed Saver Network, Czech Republic	NGO	Landraces and old Czech varieties (cereals, vegetables) maintained by farmers and gardeners; seed exchange network	
	Botanical garden of Botanical Institute, Pruhonice	Public Research Institution	Working collection of Iris, Hemerocalis and Paeonia additional to that included in the documentation system GRIN Czech/EURISCO	Member of Czech National PGR programme
DENMARK				
DNK020	University of Copenhagen Pometet (Taastrup): Pometum	University	Field genebank (orchard) for fruit and berry crops: Nordic and Danish heritage fruit cultivars and a reference collection of global fruit varieties.	Over 750 fruit varieties – including ~400 apple varieties, 250 pear, plus cherries, plums, and berries (currant, gooseberry) maintained in living trees/bushes.; NordGen (and hence EURISCO entries) primarily covers seed accessions; the clonal fruit collections of the Nordic countries are managed nationally
DNK059	Department of Bioscience, University of Aarhus	University	living plant collections/ working research collections	A vegetative collection of horticultural crops, notably vegetables and herbs propagated by tubers, bulbs, or cuttings. This includes historic Danish clones of rhubarb, horseradish, Jerusalem artichoke, shallots (potato onion), hop, etc. which are maintained at the Aarhus University research station in Årslev
	Danish Seed Savers (“Frøsamlerne”),	NGO	Community seed bank (distributed among	~2,000 accessions of vegetables and grains have been maintained and exchanged by members

			members); Heirloom vegetables (peas, beans, tomatoes, etc.), heritage cereals (old barley, rye landraces), and ornamentals	
ESTONIA				
EST015	Botanical Garden of the University of Tartu	University	Seed bank of wild flora and living plant collection; Native flora, including some crop wild relatives (wild grasses, wild berries)	
FINLAND				
FIN016	Natural Resources Institute Finland Piikkiö	Governmental (public research institution)	Field genebanks, in vitro and cryopreserved stocks of vegetatively propagated species	Thousands of accessions, e.g. Luke (Natural Resources Institute Finland) holds Finnish and Nordic apple, pear, plum varieties (some 300+ apples), berry cultivars (currants, strawberries, sea buckthorn, cloudberry clones), ornamental perennials, and vegetable clones (e.g. rhubarb, horseradish); Finland maintains significant clonal collections under Luke, outside NordGen's seed mandate; The procedures for material transfer agreements and practices to arrange access to the plant material need to be developed.
FIN034	Natural Resources Institute Finland Loppi			
FIN036	Natural Resources Institute Finland Suonenjoki			
FIN037	Natural Resources Institute Finland Jokioinen			
FIN038	Kainuu Region Vocational College, Kajaani, Kainuu	Governmental/ vocational school	Field collections of Luke (fruits and berries, vegetables, herbs and landscaping plants)	
FIN039	Livia College Tuorla, Kaarina		Field collections of Luke (fruits and berries, vegetables, herbs and landscaping plants)	

FIN043	Botany unit, Finnish Museum of Natural History, University of Helsinki	University	Living Plant Collections: Maintained at the Kaisaniemi and Kumpula Botanic Gardens; genomic resources collections (DNA bank)	These collections serve both research and educational purposes; genomic resources collections are available on loan for university research
	Maatiainen ry (Finnish Heritage Plant Association) (nationwide)	NGO	Decentralized community seed bank and living collections; Traditional Finnish landraces of garden and field crops and ornamentals	
FRANCE				
FRA014	Centre de Coopération Internationale en Recherche Agronomique pour le Développement	Governmental (public research institution)		
FRA064	Amélioration génétique et adaptation des plantes méditerranéennes et tropicales, INRAE-CIRAD	Governmental (public research institution)	Mediterranean & tropical crops (e.g. sorghum, millets, cotton, etc.)	Joint INRAE-CIRAD research collections of genetic resources. Historically fragmented data management
FRA098	Station de la Réunion, CIRAD-FLHOR	Governmental (public research institution)	Field accessions; Indian Ocean tropical fruit crops (e.g. lychee, mango, vanilla)	not part of the national PGR program (focused on export crops)
FRA099	Station de la Martinique, CIRAD, Campus agro-environnemental Caraïbe (CAEC)	Governmental (public research institution)	Caribbean crops (banana and plantain clones, roots/tubers, tropical forages)	Major field genebank for Musa (banana/plantain) and root crops in the French West Indies; these international collections were outside the French national inventory process
FRA109	Génétique et Amélioration des Plantes, Plant Biology and Breeding, INRAE Antilles-Guyane	Governmental (public research institution)	Tropical root and tuber crops, local landraces (yam, cassava, etc.); field and in vitro collection	Historically omitted in the NI due to fragmentation and no formal recognition until 2022 (national collection status).

FRA201	Station de la Guadeloupe, CIRAD-FLHOR	Governmental (public research institution)	Tropical crops (banana, plantain, taro, yam, tropical forages)	~2000 accessions; Hosts the international Musa germplasm backup and other crops; regarded as an international/base collection not originally reported through the French national system. Only recently has France moved to include such materials in the MLS and national catalogue
FRA215	Groupe d'Étude et de contrôle des Variétés et des Semences, Unité expérimentale de Brion	Governmental	Forage and cereal heritage collections (working collections for variety trials)	GEVES maintains reference samples of old varieties (e.g. forage grasses, wheat) for DUS and conservation. These were outside the genebank networks historically
FRA254	Institut de Recherche pour le Développement	Governmental	Underutilized and wild species (e.g. African traditional crops, crop wild relatives)	IRD holds some seed samples (and herbarium specimens) from its tropical research (e.g. fonio, wild yams); Institutional focus on research (not conservation) and regulatory uncertainties (ABS for wild species)
FRA309	Conservatoire Végétal d'Aquitaine	Private	Field genebanks (historic orchards); Fruit heritage varieties	CVR Aquitaine holds hundreds of fruit tree accessions, including dozens of local cultivars for each of apple (Malus), pear (Pyrus), plum (Prunus domestica), cherry, kiwi, walnut, etc.
FRA311	Union pour les ressources génétiques du Centre-Val de Loire	NGO	local cereal/vegetable landraces on-farm; wild crop relatives *in situ*	
FRA321	Réseau semences paysannes	Umbrella network of ~80 grassroots groups	Facilitates on-farm/*in situ* conservation through local "Maisons de la semence paysanne" (peasant seed houses), community seed bank	Emphasizes farmer-led seed selection and landrace maintenance, seed exchange, training, and participatory plant breeding
	Les Croqueurs de Pommes ("Apple Crunchers")	NGO/CSO	Focuses on *in situ*/on-farm conservation in orchards: identifying old local fruit cultivars, propagating them by grafting, and establishing conservatory orchards.	
GEORGIA				
GEO002	Tbilisi Botanical Garden	Governmental		2 accessions in Genesys

GEO028	Agricultural University of Georgia	University	Includes more than 2,000 unique samples of traditional and local selective varieties, endemic and wild species, as well as field and vegetable crops.	
GEO040	Certification and Standardization Department, PGR Gene Bank of Field Crops	Governmental	Field crops, including cereals and food legumes	
GREECE				
GRC002	Tobacco Department, Institute of Plant Breeding and Genetic Resources	Public Research Institute	More than 30 accessions of *Nicotiana tabacum* (As of 2022)	
GRC006	Plant Production Department, Institute of industrial and forage crops	Public Research Institute	industrial, forage crops, and pulses	
GRC009	Cotton Department, Institute of Plant Breeding and Genetic Resources	Public Research Institute	*Gossypium hirsutum*, othet cotton species (*G. barbadense*)	
GRC016	Institute of Olive Trees, Subtropical Crops and Viticulture	Public Research Institute	Field genebank of olive cultivars and other subtropical fruit (e.g. citrus). Focus on Greek olive variety conservation and evaluation	More than 100 olive accessions (Greek and some foreign varieties); also maintains local fig, citrus, avocado, etc. (small collections).
GRC017	Medicinal and Aromatic Plants Department, Institute of Plant Breeding and Genetic Resources	Public Research Institute	Medicinal and aromatic plants	
GRC020	Olive, Fruit and Vegetables Department of Kalamata	Public Research Institute	Local orchard collections; focuses on olive, and unique local horticultural crops (e.g. figs)	Mostly *in situ*/on-farm observations rather than formal *ex situ* holdings. The institute's work is more survey-oriented, with no formal genebank facility for germplasm exchange

GRC044	Balkan Botanic Garden of Kroussia, Institute of Plant Breeding and Genetic Resources	Public Research Institute	Botanical garden & seed bank for native flora; focuses on wild plant species of Greece and the Balkans, especially rare, endemic and threatened taxa	hundreds of seed accessions of rare species, collected via EU-funded projects
GRC047	Vegetable Department, Institute of Plant Breeding and Genetic Resources	Public Research Institute	Conservation, evaluation, and breeding of various vegetable crops.	
GRC050	AEGILOPS-Network for Biodiversity and Ecology in Agriculture	NGO	Participatory on-farm conservation network with focus on landrace improvement and participatory plant breeding. Works mainly with cereals, pulses, and vegetable landraces, engaging organic farmers to evaluate and select local varieties	Dozens of local landraces actively maintained and improved on farms across multiple regions. AEGILOPS facilitates regional seed “focal points”, seed schools, and community seed banks, so its holdings are dynamic rather than a single static collection
GRC057	Aristotle University of Thessaloniki	Governmental/ University	Small, project-based collections	ad hoc research materials, not maintained as long-term genebank collections.
GRC059	University of Thessaly (Volos) – Medicinal/Aromatic Plant Collection	Governmental/ University	Research collection of medicinal and aromatic plants, including CWR.	Used for pharmacological/agronomic research and on-farm conservation studies. Involved in *in situ* CWR projects
GRC073	National and Kapodistrian University of Athens	Governmental/ University		University teaching and research collection, seed samples of various plant species (including crop relatives and model plants) for education and study
GRC100	CIHEAM Mediterranean Agronomic Institute of Chania (MAICh)	Educational and research institute/ International	Regional genebank focusing on Cretan and Mediterranean crop germplasm. Long-term seed	More than 2000 accessions of endangered endemic Cretan plant species and old vegetable varieties

			storage for endemic wild plants of Crete and heirloom vegetable landraces	
	Peliti Seed Network	NGO	Farmer seed network conserving traditional Greek landrace varieties on-farm. Focus on collecting, cultivating, and sharing heirloom seeds of vegetables, cereals, pulses, etc., through a community of volunteer farmers	
	Julia and Alexander Diomedes Botanical Gardens	Governmental	Various sections including (1) section of historic plants (Olives and wild olives, fennel) (2) medicinal and aromatic plants (3) Plants and economy section (tobacco, cotton, flax, traditional grape varieties, fruit trees)	Biggest botanical garden in Greece and Eastern Europe
	Archipelagos Institute – Aegean Seed Bank	NGO	Regional seed bank for Aegean islands' landraces.	Focus on island landraces like Santorini tomato, local beans, etc. The bank works with a wide network from Thrace to Crete for seed collection and multiplication; Its focus is partly educational (school gardens) and regional
HUNGARY				
HUN018	Research Centre, Debrecen Univ. of Agriculture	Governmental/ University	Field crops, landraces (multi-crop)	University-held collection; lack of resources and mandate to submit data

HUN019	Cereal Research Institute	Public Research Institute	Field crops (cereals, maize, etc)	
	MATE Faculty of Horticulture and Rural Development	Governmental/ University	Fruit tree cultivars (apple, pear, plum, etc.); Ampelographic collection of Hungarian varieties	Maintained in the field genebank; Over time, several Hungarian agricultural research and higher education institutions merged and formed MATE in 2021 (including National Agricultural Research and Innovation Centre (NARIC) Fruitculture Research Institute, and Szent István University)
	Hungarian Association of Arboreta and Botanic Gardens (Magyar Arborétumok és Botanikus Kertek Szövetsége (MABOSZ))	Non-governmental, professional association that brings together Hungary's arboreta and botanical gardens.	each member arboretum or botanical garden maintains its own collection focus, the collective holdings under MABOSZ include protected or threatened native species, some of which are crop wild relatives (e.g., wild relatives of fruits, cereals, and medicinal plants).	
IRELAND				
IRL034	Department of Agriculture Food and the Marine (Backweston)	Governmental	Cereals and CWR	Formerly IRL029 (Department of Agriculture, Fisheries and Food, National Crop Variety Testing Centre, no updates since 2011)
IRL036	Department of Agriculture Food and the Marine, Potato Centre	Governmental	Potatoes	
	Trinity College, Dublin	Governmental/ university	seeds of threatened native plant species including CWR; vegetables; Living collection (arboretum)	Irish Threatened Plant Seed Bank has 200 accessions of 59 threatened native species
	National Botanic Gardens, Dublin	Governmental	Ornamental, horticultural, wild relatives, and rare species	Managed by the Office of Public Works (OPW). Holds living collections and an herbarium. Focuses on conservation of Irish native and exotic plant species, botanical research, and public education. Part of the international botanic garden network.
ITALY				

ITA034	Institute of Plant Breeding and Agricultural Research 'Nazzareno Strampelli'	Governmental	Brassica (cole) leafy vegetables; Legume vegetables (succulent or dried); Cereals and pseudocereals.	More than 600 accessions of Triticum
ITA356	Centro Ricerche Produzioni Vegetali (CRPV) – Cesena	Public-private limited liability consortium	Fruit tree germplasm (heirloom and local cultivars of apple, pear, stone fruits, tree nuts), old local grape varieties, vegetable landraces	Collaborates closely with the Emilia-Romagna Regional Government, Università di Bologna (UNIBO), Università di Modena e Reggio Emilia (UNIMORE), CREA, CNR
ITA374 ITA375 ITA376	Agenzia per la Ricerca in Agricoltura della Regione Sardegna	Governmental	Fruit and nut tree cultivars, Vitis, local landraces (cereals, pulses, legumes, other horticultural crops)	Regional agency of Sardinia (Italy) devoted to agricultural research, innovation, and the conservation of local plant and animal genetic resources. (De Pau et al., 2024)
ITA426	Dipartimento di Scienze Agrarie e Ambientali, Università degli Studi di Milano	University	Vitis collection	Vitis International Variety Catalogue VIVC
ITA456	Banca del Germoplasma della Sardegna, Università di Cagliari	University	Wild Sardinian flora; CWR	Member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian network of germplasm banks for *ex situ* conservation of threatened flora); Ucchesu et al., 2016
ITA459	Centro Interdipartimentale per la Conservazione e Valorizzazione della Biodiversità Vegetale, Università degli Studi di Sassari	University	Wild, autochthonous species (endangered, protected by law, or regionally endemic); traditional or lesser-known crop varieties (depending on funding and specific projects)	
	Civiltà Contadina	NGO/CSO	Heirloom seeds, orchard varieties	Ark of Seeds project (ex-situ seed conservation through the medium-long term conservation of accessions in a cold room at -21°C, inside glass containers or heat-sealed bags),

	Fondazione Archeologia Arborea (Umbria)	Private initiative	Historical fruit orchard (apple, pear, plum etc)	
	I Giardini di Pomona	Private initiative	Fruit orchard conservatory (Figs, pomegranate, apple, pear, citrus, grapes, cherries, aromatic herbs, etc.)	one of the most important Ficus collections (around 600) in Europe and the Mediterranean basin, for quality and variety.
	Edmund Mach Foundation (FEM)	Provincial research institute	Fruit crop genebank (national apple and grape collections)	taly's major apple germplasm collection; extensive grapevine accession
	Banca Regionale del Germoplasma della Valle d'Aosta	Governmental	Seed and cryo bank; Native (autoctona) plant heritage of the Aosta Valley (Wild Alpine and forest species (wild relatives, endemic flora), minor local horticultural crops.	primarily seeds of high-altitude species and local vegetable landraces; Member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian network of germplasm banks for *ex situ* conservation of threatened flora) https://www.reteribes.it/index.asp (Magrini et al., 2022)
	Banca del Germoplasma delle Alpi sud occidentalli (Piedmont)	Governmental	Alpine landraces (cereals, pulses), orchard species (apples, pears), wild relatives	Member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian network of germplasm banks for *ex situ* conservation of threatened flora) (Magrini et al., 2022)
	La Banca del Germoplasma della Piante Lombarde	Governmental	Native (autoctone) wild species, especially those that are threatened or have high conservation priority in Lombardy	Member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian network of germplasm banks for *ex situ* conservation of threatened flora) (Magrini et al., 2022)
	Banca dei semi del Dipartimento di Biologia dell' Università di Pisa	University	Seed genebank (CWR, threatened species etc), in vitro collection	2000 accessions are conserved for a total of over 200 taxa: Managed by University of Pisa; member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian network of germplasm banks for *ex situ* conservation of threatened flora) (Magrini et al., 2022)
	Banca del Germoplasma dell'Orto Botanico di	University	Botanical garden with cold storage facility	Member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian

	Padova, Università di Padova			network of germplasm banks for *ex situ* conservation of threatened flora) (Magrini et al., 2022)
	Banca del Germoplasma di Palermo, Università di Palermo	University	Mediterranean wild plants, both rare/threatened species, CWRs and old local cultivars	Sicilian Plant Germplasm Repository: Member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian network of germplasm banks for *ex situ* conservation of threatened flora); member of Network of Mediterranean Plant Conservation Centers (GENMEDA) http://www.genmeda.net/en/members/current_members/spgr-pa
	Banca del Germoplasma dell'Orto Botanico, Università di Roma	University	Seed bank (rare and threatened species, local and regional flora, CWR, medicinal and aromatic plants); living collections	Member of Rete Italiana Banche del germoplasma per la conservazione *Ex situ* della flora minacciata (RIBES) (Italian network of germplasm banks for *ex situ* conservation of threatened flora) (Magrini et al., 2022)
	Azienda Agraria Sperimentale Stuard – Parma	Non-profit limited liability consortium company	over 100 varieties of ancient cereals and the same number of ancient tomatoes, melons, pumpkins, mustard watermelons and beans.	
	Istituto Tecnico Agrario Superiore "Bocchialini"	Governmental/ Technical school	orchard with more than 600 varieties of old fruit varieties	Works in collaboration with Azienda Agraria Sperimentale Stuard – Parma
	La Banca del Germoplasma di Roma, Dipartimento di Biologia Ambientale, Sapienza Università di Roma	University	Hosting over 3000 species of plants, including 160 varieties of Italian grapes	
LATVIA				
LVA006	Priekuli State Plant Breeding Station	Governmental	Cereals (rye, wheat, barley), potatoes	Working breeding material but also has collection of local landraces and heritage varieties
NETHERLANDS				

	Pomologische Vereniging Noord-Holland (North-Holland Pomological Society)	Non-profit Association	Heritage fruit tree cultivars (apples, pears, plums, cherries, etc.), maintained in historic orchards, home gardens	More than 500 varieties of fruit trees; apple collections are placed under the ITPGRFA’s Multilateral System (MLS) of Access and Benefit-sharing (ABS).
	Noordelijke Pomologische Vereniging (Northern Pomological Society)	Non-profit Association	Heritage fruit tree cultivars (especially apple) in northern provinces (e.g. Friesland, Groningen, Drenthe)	apple collections are placed under the ITPGRFA’s Multilateral System (MLS) of Access and Benefit-sharing (ABS).
	Fruithof Frederiksoord	NGO	*Ex situ* arboretum/orchard of fruit trees (apple, pear, peach, quince)	More than 500 varieties; apple collections are placed under the ITPGRFA’s Multilateral System (MLS) of Access and Benefit-sharing (ABS).
	Het Levend Archief	Collaborative, non-profit initiative	Seeds of native wild plants of the Netherlands, including crop wild relatives and wild food plants	Initiative focusing on wild flora (not primarily crops); with backup at CGN
NORWAY				
NOR003	University of Oslo, Natural History Museum, The Botanical Garden	Governmental/ University	Wide range: fruit, berry, vegetable, herb, ornamental heritage plants	Major botanical garden living collection
NOR009	NMBU Norwegian University of Life Sciences, Faculty of Biosciences , Department of Plant Sciences (IPV)	Governmental/ University	Fruit trees and vegetables (research collections)	University orchard and garden collection
NOR011	NIBIO Apelsvoll	Governmental	Norwegian clonal fruit collections; vegetables	Small specialized collection of perennial herbs
NOR013	University of Bergen, The Arboretum and Botanical Garden, Milde	Governmental/ University	Wide range: fruit, berry, vegetable, herb, ornamental plants	University botanical garden collection

NOR017	NIBIO Landvik	Governmental	Norwegian clonal fruit collections; Berries, vegetables, and perennials	National horticultural genebank site
NOR027	NIBIO Ullensvang	Governmental	Norwegian clonal fruit collections; (extensive apple collection, also pears etc.)	
NOR046	Domkirkeodden - Anno Museum	Museum (Open-air)	Old fruit tree varieties (apple/pear)	
NOR053	Njoes fruit and berry centre	Private	Norwegian fruit and berry genotypes (apples, plums, berries)	
NOR059	University of Agder, Natural History Museum and Botanical Garden	Governmental	fruit, berry, vegetable, herb, ornamental plants	
NOR060	UiT – The Arctic University of Norway , Tromsø Arctic-Alpine Botanic Garden	University	Northern-adapted berries, vegetables, medicinal & spice plants, ornamentals;	Botanical garden living collection; part of national clonal network
NOR061	NIBIO Division of Biotechnology and Plant Health	Governmental	Norwegian clonal fruit collections	
NOR062	The Norwegian Museum of Horticulture , Dømmesmoen	Museum	Living collection of fruits, vegetables, ornamental plants	
NOR064	Kystmuseet Hvaler, Vesterøy	Museum (Open-air)	Very small orchard collection of coastal heritage fruit trees	
NOR066	Lier Bygdetun	Museum (Open-air)	small orchard with older fruit varieties, apples, plums, and pears	
NOR068	Lund bygdemuseum og kulturbank	Museum (Open-air)	Traditional fruit varieties (local apple cultivars)	Community-run orchard gene bank; considered part of national clonal network

NOR069	Musea i Sogn og Fjordane, avd Nordfjord Folkemuseum	Museum (Open-air)	old local fruit and berry varieties adapted to the coastal/inland climate of Nordfjord.	regional folk museum that maintains historic gardens and orchards
NOR070	Vigatunet, Ryfylkemuseet	Museum (Open-air)	Heritage orchards with older apple, pear, or plum cultivars typical of southwestern Norway	
NOR071	Hjeltnes vidaregåande skule/Grønt kompetansesenter, Ulvik	Educational	West Norway fruit varieties	Horticultural school collection; treated as part of national clonal genebank network
NOR072	Bygdøy kongsgård - gartneriet	Governmental	Heirloom vegetable varieties and older fruit tree cultivars,	
NOR073	Ringebu prestegård		Orchard with old fruit varieties	
ROMANIA				
ROM012	SCDA Lovrin, Timiș – Agr. Research Station	Public research institute	Field crops: wheat, barley, sunflower, maize (Western RO) breeding and preservation of local germplasm.	breeding material for ongoing research and conservation of local or historically significant cultivars.
ROM014	SCDA Secuieni, Neamț – Agr. Research Station	Public research institute	Field crops & forages: cereal and fodder crop landraces (Moldova region).	
ROM022	'Ion Ionescu de la Brad' University of Agricultural Sciences and Veterinary Medicine Iasi	University	Academic collections: houses a viticulture collection and various crop trials	Dozens of grape varieties in vine collection; other crops in faculty research.
ROM025	Sericarom S.A. Bucuresti		Mulberry germplasm	
ROM029	Research and Development Station for Fruit Tree Growing Baneasa	Public research institute	*ex situ* collections of various fruit tree genotypes, including older local	Fruit tree breeding and variety improvement (apples, peaches, apricots, cherries, plums, etc.).

			cultivars, breeding lines, and new varieties.	
ROM032	Research and Development Station for Fruit Tree Growing - Bistrita	Public research institute	temperate fruit species suited to the cooler climate of northern Romania (apples, plums, pears, cherries)	local landraces or heirloom fruit varieties from Transylvania.
ROM033	Nursery and Fruit Farm Istrita, Buzau		Fruit rootstocks & viticulture: certified fruit tree nursery and grapevine collection	
ROM037	Research and Development Station for Fruit Tree Growing - Fălticeni	Public research institute	fruit cultivars (apple, plum, pear, cherry)	
ROM076	National Research and Development Institute for Biotechnology in Horticulture	Public research institute	in vitro maintenance of rare or endangered horticultural germplasm (e.g., micropropagation of fruit rootstocks, ornamental plants, and other horticultural species	
SPAIN				
	Fundació Miquel Agustí (FMA)	Non-profit foundation	Catalonian traditional crop varieties (especially vegetables)	More than 1700 accessions (Catalan landrace vegetables)
	Red Española de Bancos de Germoplasma de Plantas Silvestres, or REDBAG network	Network of botanical garden seed banks	dedicated to wild flora and endemic plant conservation, butmay inadvertently include CWR, wild food plants	Include: Real Jardin Botanico, Banco de Germoplasma Vegetal Andaluz, Gran Canaria Seed Bank, Sóller Botanical Garden Seed Bank, Jardín Botánico Atlántico Seed Bank
SWEDEN				

SWE089	Department of Landscape Architecture, Planning and Management/National Genebank Alnarp	Governmental	A large collection of Swedish heritage fruit cultivars and locally adapted varieties.	Maintained as living trees in orchards, often duplicated across multiple sites; Raspberries, strawberries, currants, gooseberries, apples, pears, plums, cherries; Herbaceous perennials, shrubs, and trees
UNITED KINGDOM				
GBR012	NIAB East Malling Research	Governmental	Clonal field collections (orchards and plantations)	A field genebank of fruit trees and berries (apples, pears, quinces, cherries, plums, raspberries, strawberries, etc.),
	Agri-Food & Biosciences Institute (AFBI)		Both seed and field collections. At AFBI Crossnacreevy, seed accessions of forage grasses, forage legumes and cereals; At AFBI Loughgall, clonal field banks include Irish heritage apple varieties (Armagh Orchard Trust) and small collections of other fruits fao.org, as well as wild potato species	The Northern Ireland Horticultural and Plant Breeding Station (under its original name GBR212) no longer exists as a standalone entity, its work and resources continue under AFBI.
GBR017	Garden Organic – Heritage Seed Library	Non-profit foundation	UK traditional crop varieties (especially vegetables)	More than 800 accessions of landrace vegetables
GBR030	National Fruit Collections, University of Reading	Governmental	National Fruit Collections	More than 1,000 accessions of fruits, primarily fruit tress
GBR040	National Institute of Agricultural Botany	Governmental	Diverse field and pasture crops	
GBR045	School of Plant Science, University of Reading		Research collections of Lupins, Cocoa,	
GBR070	Rothamsted Experimental Station, Institute of Arable Crops Research	Governmental	historic cereal germplasm and genetic stocks (Seed archives from long-term field trials)	These materials are maintained for research and historical purposes, not as part of an accessible genebank collection; Rothamsted’s famous Broadbalk Wheat Experiment has preserved wheat seed samples annually since the 1840s, and the

				institute manages specialized resources like a wheat TILLING population (mutant library)
GBR213	Scotland's Rural College (SRUC, formerly Scottish Agricultural College)	Governmental	Seeds of small grain cereals and oil/fiber crops *(ex situ)*	Scottish landrace barley collection (ca. 1,000 accessions of Hordeum landraces and heritage cultivars) and a fiber flax/linseed collection (~350 accessions) were curated by SAC researchers
GBR142	Botanic Gardens Conservation International	Umbrella network	Network of Botanic Gardens	Network of living collections, seed genebanks and associated databases
GBR250	Natural England	Governmental	Coordination of English National Nature Reserves	Manager of 1,000s of *in situ* conserved populations of Englands CWR diversity
GBR252	Plant Heritage	Non-profit foundation	Curator many national collections of PGR	

5.1.1 Trends in the Intra-country Gaps

i. Clonal Field Genebanks and Perennial Crops

A recurring gap in EURISCO's coverage concerns large, field-based collections of fruit trees and other horticultural crops. Institutions like the Armenian Scientific Centre of Viticulture, Fruit-Growing & Winemaking, the Institute of Adriatic Crops in Croatia, UK National Fruit Collection and fruit research stations in Bulgaria, Greece, UK and France hold extensive collections, often of apple, plum, grape, citrus, and similar orchard species, that remain underrepresented in national submissions.

Historically, national PGR strategies have favoured annual, seed-based crops (e.g., grains, legumes, vegetables) due to their relative ease of storage and documentation within conventional seed banks. Consequently, living field genebanks, comprising individual trees and vines, commonly exist outside major databases, even in countries with well-established horticultural breeding programs. Field collections of perennial fruit crops often exist independently of seed-based genebanks, falling under different institutional or administrative arrangements. Such organizational silos risk leaving behind valuable reservoirs of fruit crop diversity, many of which are regionally or locally adapted genotypes that may be endangered unless they are brought under broader conservation initiatives.

In the Nordic region, for example, the Nordic Genetic Resource Centre (NordGen) oversees seed collections for Denmark, Finland, Iceland, Norway, and Sweden and represents them in EURISCO. However, clonal field repositories remain under each country's purview. In Finland, the Natural Resources Institute (Luke) maintains substantial apple, berry, and other fruit collections in field and cryopreservation systems. Norway's Norwegian Institute of Bioeconomy Research (NIBIO) manages orchard-based apple, pear, plum, and cherry germplasm across multiple sites. Sweden likewise conserves a large clonal collection at Alnarp under the Swedish University of Agricultural Sciences, supplemented by local initiatives focusing on heritage fruit varieties. While portions of this material are documented and shared, there is no unified platform encompassing all orchard collections at the Nordic level, unlike the case with seed germplasm.

ii. Botanical Gardens and Arboreta

Botanical gardens and arboreta, long recognized for their emphasis on conserving and studying wild plant diversity, are widely acknowledged as custodians of significant collections, many of which include CWR or horticultural species with traits relevant to agriculture. In fact, a comparison of relative agricultural and botanic garden genebank holdings revealed that botanic gardens house significantly higher numbers of both CWR taxa and population samples than agricultural genebanks (Kell et al., 2008). Even so, their core missions, record-keeping practices, and conservation goals commonly differ from those of institutions responsible for agricultural germplasm. Rather than aiming to distribute breeding materials or align systematically with genebank standards, botanical gardens often prioritize scientific research, public education, and the preservation of rare and endemic flora. Consequently, they rarely submit their accessions to their respective NI.

The University of Tartu Botanical Garden in Estonia, for instance, maintains a seed bank of native species, including some that may overlap with regional CWR. The Balkan Botanic Garden of Kroussia in Greece focuses on preserving and studying wild plant populations of the Balkans, often on account of their threatened status. In Ireland, both the National Botanic Gardens and the Trinity College seed bank collaborate on *ex situ* conservation of Irish flora, while Norway's university botanical gardens (located in Oslo, Bergen, and Tromsø) curate a wide range of northern-adapted plants of horticultural, medicinal, and ecological interest. Spain's REDBAG network, spanning multiple regional botanic gardens, safeguards endangered and endemic species through seed conservation protocols. Meanwhile in Italy, numerous university and regional botanical gardens participate in RIBES (Rete Italiana Banche del germoplasma per la conservazione *ex situ* della flora minacciata), which coordinates seed banks dedicated to rare, endemic, or otherwise threatened Italian species. While the

Millenium Seed Bank at the Royal Botanic Gardens, Kew has by far the largest and most taxonomically diverse collection of CWR taxa and population samples (FAO, 2025).

Despite the presence of species that could hold agronomic or breeding value, these institutions generally document their material using garden-specific information systems, specialized nomenclatural standards, or dedicated conservation registers aligned with international botanical networks (e.g.,Botanic Gardens Conservation International (BGCI) PlantSearch), rather than agricultural genebank platforms. In many cases, the collections are organized according to taxonomic or ecological criteria and accompanied by minimal agronomic or performance data. The curatorial frameworks revolve around taxonomy, phenology, and sometimes *ex situ* cultivation research. Thus, the protocols in botanical gardens do not directly translate into the passport data sets or descriptor systems needed for integration into a crop-focused inventory.

Moreover, because botanical collections are often maintained for research or educational purposes, garden administrators might not perceive clear benefits, or possess the administrative support, to document these plants as part of national PGR programs. Although botanic gardens commonly have extensive genebank collection, their living collections are sometimes short-term, periodically refreshed in public displays, or represented by permanent exhibit plantings rather than germplasm accessions conserved for distribution. While herbaria housed within the same institutions, meanwhile, hold dried specimens that have proved of highly significant value as the foundation for CWR conservation planning yet this information is rarely used the botanic gardens themselves who primarily focus on taxonomic verification and archival reference, rather than any direct conservation goal related to breeding or reintroduction. In selected cases, collaboration does occur when both botanical institutions and agricultural research agencies recognize mutual interests, for instance, in the conservation of a rare CWR or in joint ecological–agronomic studies, but such coordination often depends on individual institutional mandates and national policy frameworks. Consequently, a wealth of potentially valuable genetic diversity in botanical gardens remains underrepresented in formal agricultural PGR inventories and data inter-operativity is not facilitated, which limits the accessibility of traits that could prove critical for crop improvement.

iii. Overseas and Specialty Stations

Overseas and specialty stations often manage significant repositories of germplasm that have evolved separately from mainstream national genebank programs. In the case of France, key examples appear in its overseas departments such as Martinique, Guadeloupe, and Réunion, where the French Agricultural Research Centre for International Development (CIRAD) and partner institutions hold extensive collections of bananas, plantains, yams, cassava, lychee, mango, and other tropical or subtropical crops. These collections have commonly been treated as international, rather than national, resources, with associated administrative and policy frameworks that may not align with conventional pathways for integration into the NI and EURISCO.

Although these territories are politically linked to the metropolitan states, they often operate under separate funding lines and organizational mandates, focusing on tropical agriculture or regional development. Another factor influencing the status of such collections is the historical reliance on international networks or crop-specific consortia. The *Musa* germplasm network, for example, has long provided a collaborative framework for circulating banana and plantain material among research stations worldwide. In this model, genebank managers have generally emphasized distributing germplasm through global platforms or maintaining back-up collections in other countries, rather than submitting detailed passport data to the French NI. Such an approach, while beneficial for global research collaborations, can inadvertently sideline the inclusion of these materials in domestic databases or EURISCO. In practical terms, it means that *ex situ* holdings in Martinique or Guadeloupe, although formally under French jurisdiction, have not always been reflected in listings of French national genetic resources.

As stakeholders continue to recognize the strategic importance of tropical genetic resources, it may become more likely that these historically international collections will see greater inclusion in European inventories. This is a shift that would better reflect the full scope of agricultural biodiversity managed under countries with overseas or regionally specialized research stations.

iv. Research Institutions and Universities with "Working Collections"

Research institutions and universities frequently maintain extensive yet loosely organized "working collections" of plant materials, reflecting their priority on active research rather than long-term conservation. Institutions such as Debrecen University in Hungary, Aristotle University of Thessaloniki in Greece, National Research Council in Italy, and the University of Ghent in Belgium exemplify this trend. Their holdings are typically developed or assembled to meet specific research objectives. Because such collections often revolve around ongoing experiments rather than permanent curation, they are not systematically deposited into national genebanks or reported in consolidated NI and, when the project is completed or researchers retire, these important resources may be threatened.

Unlike dedicated genebank collections, which conform to standardized procedures of long-term storage and data sharing, university-based and research station collections evolve rapidly with the progression of scientific projects. Breeding lines and experimental materials are multiplied, recombined, or phased out once specific studies conclude. These processes can result in a fluid inventory, where certain lines are retained for extended experimentation, while others are discontinued or replaced by more promising genotypes. Consequently, it becomes challenging for institutional curators or individual researchers to invest in the passport data compilation and curation work that would be necessary for formal genebank submission. In many cases, the information about origin, pedigree, and key phenotypic traits, if recorded at all, remains confined to departmental archives, *ad hoc* spreadsheets, or scattered project databases. Even if researchers recognize the broader value of their lines for breeding or biodiversity, the lack of an institutional mandate to deposit material or share standardized data means that potential inclusion in a national PGR information system does not occur. Instead, scientists typically exchange this material within specialized research networks or use it for educational purposes at the university level. On occasion, lines may transition into formal collections if a project yields notable findings or if a national genebank identifies a subset of lines that address recognized gaps in existing *ex situ* collections.

Even institutions that have explicit mandates to conserve PGR typically keep single-seed descent (SSD) lines, introgression lines, and other experimental populations outside their official genebank holdings. Although these lines trace back to genebank materials originally intended for long-term conservation, bringing them into official collections can be difficult. A single accession may spawn multiple new lines over the course of various experiments, effectively multiplying the total number of distinct genotypes. Managing and documenting these derivatives can quickly exceed existing curatorial capacities. As a result, most of these derivatives are never catalogued in *ex situ* systems, even though they underpin much of today's multi-omics research. This situation highlights a persistent gap between the volume of genetic information produced in academic contexts and the resources formally captured by the NI or EURISCO.

v. NGO/CSO and Community Seed Banks

Many landrace, heirloom and farmer varieties are often maintained by single or small, local groups of farmers but are most visibly upheld by seed-saver networks and community seed banks, which facilitate the collection, exchange, and *in situ* maintenance of diverse germplasm outside formal *ex situ* structures. Across Europe and neighbouring regions, numerous civil society organizations (CSO) exemplify this approach. Arche Noah (https://www.arche-noah.at/english/about-arche-noah/) in Austria conserves over 6,500 heirloom accessions in decentralized seed collections, while the Peliti Seed Network (https://peliti.gr/peliti-community/) in Greece, Sjemenčica and the Alica Foundation in Bosnia and Herzegovina, Garden Organic's Heritage Seed Library in the UK and Danish Seed Savers ("Frøsamlerne") (https://www.froesamlerne.dk/forside) collectively maintain thousands of vegetables, cereal, and horticultural varieties through locally managed repositories. Similar examples, such as AEGILOPS in Greece and Gengel in the Czech Republic, combine farmer participation with basic seed conservation, often highlighting the significance of farmer-led selection practices and on-farm breeding to sustain culturally and agronomically valuable crop populations. In France, Réseau Semences Paysannes (https://www.semencespaysannes.org) coordinates dozens of grassroots groups working toward seed autonomy, and in Finland, Maatiainen ry enlists gardeners and farmers in perpetuating traditional landraces adapted to northern climates.

Despite the clear value of this diversity, particularly given its adaptability and direct ties to regional heritage, community-based collections are frequently absent or underrepresented in official inventories like EURISCO. One fundamental reason is that many seed-saver organizations favour local exchanges and participatory conservation models and may not always deposit materials in centralized genebanks or employ SMTA frameworks. These groups often prioritize the agency of small-scale producers, maintaining varieties within their immediate networks to encourage ongoing crop evolution under real farm conditions. Additionally, limited technical resources or administrative capacity can make it difficult to meet genebank data requirements. When documenting thousands of heirloom accessions, the necessary data management tools and expertise, such as cataloguing accession histories, morphological descriptors, or GPS-based collection sites, are not always available.

A further concern involves apprehensions regarding legal frameworks, particularly those governing access and benefit sharing. Some NGOs are cautious about sharing data or genetic material through formal channels, fearing it might lead to diminished local control or entangle them in obligations that diverge from their grassroots ethos and even legal prosecution for sale of unlisted crop varieties. While national genebank systems and community networks can collaborate under favourable circumstances, these groups sometimes prefer limited engagement with what they perceive as intricate policy regimes. The result is that many of Europe’s heirloom seeds, potentially containing rare alleles or regionally honed adaptations, remain only partially visible in broader databases.

vi. Private, Museum, and Other Non-Traditional Holders

Beyond the major public genebanks and community seed networks, a range of museums, cultural institutions, and private orchard initiatives steward valuable PGR in ways that diverge from conventional *ex situ* conservation practices. While these sites frequently manage historically or regionally important crop varieties, their day-to-day operations and overarching missions typically center on cultural interpretation, public engagement, education, or local heritage rather than long-term germplasm conservation for research and breeding. As a result, these holdings, despite their potential value in agricultural programs, are often absent or underrepresented in official inventories.

A key example comes from open-air museums and historic sites in Norway, such as Domkirkeodden and the Ryfylkemuseet, where living orchards serve to illustrate agricultural traditions for visitors. These orchard collections may contain older varieties of apples, plums, cherries and other fruit trees that may be genetically distinct and potentially of interest to breeders. Privately organized orchard projects can also contribute to the undercount of PGR at the national level. The Conservatoire Végétal d’Aquitaine in France, for example, maintains a large orchard collection of regional and traditional fruit cultivars. While there is awareness that certain materials might be of interest for broader agricultural or scientific purposes, the organizational structures and funding priorities in such settings do not always align with formal genebank procedures. Compiling standardized passport data and meeting administrative requirements can be challenging for groups largely devoted to community outreach or orchard upkeep. Questions around data ownership and additional responsibilities may further deter private stewards from collaborating with official PGR networks.

5.1.2 Data Update Gaps

As shown in Table 6, the frequency and consistency of data updates in EURISCO differ considerably among participating countries, with important consequences for both the platform itself and broader

European PGR management. The Greek national genebank (GRC005) illustrates this situation. Its last update in EURISCO occurred in 2012 with 5,355 accessions, while its actual holdings have since nearly tripled to 15,000 accessions (https://ipgrb.gr/greek-genebank/). Similarly, institutional changes often go unrecorded, as in Ireland, where the national genebank formerly operated under the Department of Agriculture, Fisheries and Food, National Crop Variety Testing Centre (IRL029) but has been functioning as the Department of Agriculture Food and the Marine (IRL034, IRL036) (https://www.gov.ie/en/organisation/department-of-agriculture-food-and-the-marine/) since 2011, which was a change incompletely reflected in EURISCO's records. While each country determines its own data submission protocols and timelines according to national priorities and resources, these variations naturally influence the comprehensiveness of EURISCO as a regional resource. Analyses and policy decisions that draw upon this information system necessarily reflect the current state of voluntary national submissions. Furthermore, as EURISCO contributes to Genesys, these data submission patterns have broader international implications for understanding germplasm availability. When participating countries can provide regular updates, EURISCO's capacity to accurately represent Europe's collective PGR collections is enhanced, supporting more informed conservation planning and resource allocation across the region.

Table 6. Data Updates in EURISCO as of February 2025 (year (no. of institutes))

COUNTRY (NI)	LAST ACCESSIONS UPDATE
ALBANIA (ALB)	2025 (1); 2024 (3)
ARMENIA (ARM)	2024 (1); 2022 (2); 2021 (2)
AUSTRIA (AUT)	2024 (13)
AZERBAIJAN (AZE)	2021 (1); 2017 (7)
BELGIUM (BEL)	2023 (1); 2013 (12)
BOSNIA & HERZ. (BIH)	2020 (2)
BELARUS (BLR)	2012 (1)
CYPRUS (CYP)	2024 (2); 2012 (1)
GEORGIA (GEO)	2024 (1); 2008 (3)
GREECE (GRC)	2021 (1); 2012 (3).
IRELAND (IRL)	2020 (1); 2011 (3).
LITHUANIA (LTU)	2019 (4); 2017 (2)
NORTH MACEDONIA (MKD)	2010 (1)
PORTUGAL (PRT)	2010 (4)
RUSSIAN FED. (RUS)	2017 (1)
SERBIA (SRB)	2012 (1)
SWITZERLAND (CHE)	2023 (1); 2013 (22).
TURKEY (TUR)	2008 (1)
UKRAINE (UKR)	2021 (68) (On-hold due to war/ crisis)
UNITED KINGDOM (GBR)	2024 (3); 2022 (1); 2020 (2); 2019 (1); 2018 (3); 2017 (1)

Beyond its effect on data accuracy, sporadic updating poses conservation risks. If an institute experiences a disaster and has not recently backed up its records through EURISCO, those accessions may be lost to the collective knowledge base. Regular data submissions thus serve as a form of insurance, preserving critical information even under adverse conditions. Chronic under-reporting can also signify systemic challenges, like insufficient funding or staffing, within national PGR programs, which might jeopardize the long-term security of these valuable resources.

5.2 Data type and quality gaps

5.2.1 Metadata

i. ***Persistent Unique Identifiers (PUID).*** The adoption of digital object identifier (DOI) among EURISCO member countries remains limited. As of February 2025, only about 14.21% of the more than 2 million *ex situ* accessions and none of the *in situ* conserved populations have been assigned a DOI. This low level of DOI adoption underscores a substantial gap in the robust and reliable identification of accessions. The limited implementation of DOIs significantly hinders effective data tracking, cross-dataset referencing, and thus the broader integration and reuse of accession data. EURISCO, in collaboration with the FAO-GLIS, already provides services to support DOI registration for collection holders. Increasing the uptake of PUID is an essential step toward enhancing traceability and interoperability, eventually improving EURISCO's compliance with the FAIR principles.

ii. ***Taxonomic Discrepancies***. Many PGR collections were established decades ago, resulting in accessions that still carry historical or outdated genus and species names. In other cases, institutions or national programs favour local synonyms or vernacular labels over internationally accepted nomenclature. Furthermore, different organizations rely on various recognized taxonomic references, which leads to divergent taxonomic classifications when datasets from multiple sources are compiled in one platform. This divergence can occur even within the same institution, where historical labels may coexist with more recent nomenclature, which reflects the differences in curatorial approaches or changes in taxonomic frameworks over time. As a consequence, the same species can appear under multiple names, while distinct taxa may inadvertently be conflated, creating confusion for those attempting to identify and compare accessions. Substantial progress in addressing these taxonomic inconsistencies depends on sustained collaboration among curators, taxonomists, and EURISCO's coordination team. Key collaborative actions include refining validation tools, defining a widely accepted taxonomic backbone (possibly GRIN Taxonomy), and conducting regular data reviews.

iii. ***Geographical Coordinates.*** In many datasets, metadata that tie a germplasm sample to its precise collection site, i.e., geographical coordinates, are frequently absent, recorded in non-standard formats, or insufficiently accurate, owing to various historical and practical mishaps. For one, older collections often predate the widespread availability of dependable global positioning system (GPS) technology, leading collectors to rely on approximate descriptors like the nearest town or region, which introduces spatial uncertainty, especially for CWR passport data. Furthermore, germplasm collectors may have limited awareness of the critical importance of these descriptors and lacked training in standardized data recording practices or used coordinate systems that are incompatible with modern geospatial standards (e.g., WGS84). Over the years, these inconsistencies have accumulated, leaving gene bank managers, researchers, and conservationists with notable difficulties in pinpointing and interpreting a large portion of PGR collections. Currently, the completeness of critical geographical coordinate descriptors (DECLATITUDE, LATITUDE, DECLONGITUDE, LONGITUDE, ELEVATION) across all datasets currently ranges only from 13.22% to 13.63%, which highlights the extent of this critical gap.

Despite occasionally being overlooked, the implications of missing or inadequate geographic coordinates are significant. Accurate coordinates allow researchers to contextualize each accession within its local environment, encompassing crucial factors like soil composition, temperature ranges, rainfall patterns, and topography. Without these coordinates, it becomes substantially harder to identify how local environmental conditions might have shaped the genetic diversity of a population, a knowledge that is essential for recognizing adaptive and stress tolerance traits, such as drought or disease resistance. This insight underpins conservation planning, landscape genomics (which integrates geospatial and genomic data to reveal genotype–environment relationships and elucidate how populations develop traits suited to their native conditions) and predictive characterisation (where the presence of certain traits is inferred based on local adaptation). Spatial data also guides evidence-based approaches in germplasm conservation. Tools like geographic information systems (GIS) help identify biodiversity hotspots, reveal underrepresented regions in collections, and limit redundancy by isolating duplicate accessions across different genetic diversity zones (GDZ). In this manner, reliable location data enhances the effectiveness of resource allocation, ensuring that collection or site designation and conservation efforts focus on the most genetically valuable, complementary or at-risk populations. Furthermore, standardized coordinates are crucial for predictive modelling, particularly in light of climate variability. As weather patterns shift, crop viability and the resilience of landraces will increasingly hinge on how well we can predict their future environmental conditions. With accurate coordinates, scientists can map projected changes in these environments, identifying accessions that are likely to harbour traits essential for breeding climate-adaptive varieties. In contrast, data gaps hamper our capacity to model future distribution patterns or prepare for potential losses of wild relatives and traditional crop variants.

5.2.2 Characterization and Evaluation (C&E) Data Gaps

i. ***Limited Coverage***. A major gap in EURISCO is the underrepresentation of C&E data (i.e. phenotypic data). Only ~91,779 accessions or about 4.4% of EURISCO's holdings have any C&E data associated with them. The limited number of countries (21 out of 43 member nations) (Figure 2) providing such data indicates that more than half of the contributing nations have not submitted any phenotypic trait information. This limited coverage substantially diminishes EURISCO's utility for researchers, breeders, and other users who depend on detailed trait information to identify germplasm suited to specific agricultural or research needs. Despite extensive C&E efforts conducted through national programs and various project initiatives, significant quantities of valuable data remain inaccessible, often restricted to local databases, publications or unpublished repositories rather than aggregated within a common system.

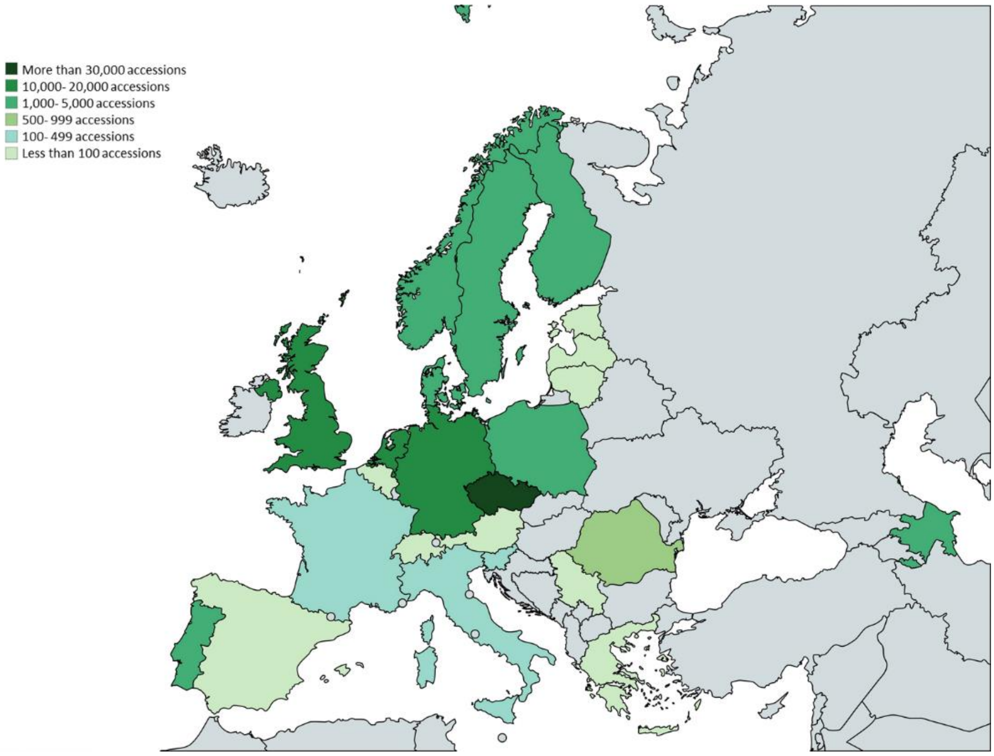

Figure 2. Map illustrating the countries (in different shades of green) that have submitted characterization and evaluation (C&E) data to EURISCO.

ii. ***Lack of Standardization***. Beyond the quantitative shortfall, EURISCO faces considerable qualitative challenges related to data standardization. While passport data benefit from the use of MCPD, phenotypic (C&E) data are inherently heterogeneous due to diverse methodologies, descriptors, measurement scales, and experimental conditions applied across different GRC and research institutions. This heterogeneity significantly complicates data harmonization and restricts comparability between datasets. While ongoing efforts such as EURISCO's minimum consensus C&E data exchange format and adherence to guidelines like MIAPPE mark initial steps towards enhanced uniformity, achieving robust and universally applicable descriptor standards remains challenging. Although the existing minimum consensus format effectively encourages data submission by lowering entry barriers, it limits the practical usability of C&E data due to reduced comparability and specificity. Addressing this fragmentation and further developing comprehensive yet feasible standards remains a key objective within the community.

To gradually address these gaps, ECPGR initiatives focus on progressively enhancing the availability and standardization of C&E data. Notably, the EURISCO-EVA information system developed within the frame of the ECPGR European Evaluation Network (EVA) project extends EURISCO to incorporate data from collaborative multi-site evaluation trials. EURISCO-EVA provides a dedicated information system for partners (genebanks, breeders, researchers) to upload and share phenotypic data in a standardized template. By coordinating trials across multiple countries and crops, and using agreed protocols, EURISCO-EVA is consolidating much-needed evaluation data in a harmonized way. As of 2024, EURISCO-EVA had already captured over 500,000 phenotypic data points for 4,845 accessions across 8 crops, gathered from about 3,800 experiments in 33 countries. These data are curated with metadata on traits, methodologies, and environmental conditions, improving their interoperability (Kumar et al., 2024).

5.2.3 In situ CWR Passport Data

EURISCO has recently expanded its scope to include *in situ* CWR data. Through a collaborative effort coordinated by ECPGR, EURISCO now incorporates standardized passport records for *in situ* populations submitted by NFP. Each participating country provides data following an agreed *in situ* CWR data exchange standard (van Hintum and Iriondo, 2022), enabling users to search EURISCO for occurrences of wild populations (treating each CWR population as a type of "accession" in the database) and access samples for use via the *in situ* backup sample deposited in designated GRC, alongside traditional genebank-held material. The data inclusion is still at an initial stage. As of early 2025, eleven European countries have submitted *in situ* CWR datasets for inclusion in EURISCO. These include pilot countries such as Bulgaria, Cyprus, Germany, Italy, the Netherlands, Spain, UK and several others that were involved in the first phase of data integration (ECPGR, 2024). Nevertheless, this development begins to fill a historical gap. EURISCO is evolving into a more comprehensive catalogue by covering *in situ* conserved diversity alongside *ex situ* collections. It's worth noting that search tools for this new data type are being developed (initially a basic search mask, with more advanced query functionality will be forthcoming).

5.2.4 On-farm Conservation Data

EURISCO was originally conceived as an information resource to aggregate and provide data on genetic materials that are, in principle, "professionally managed" and, crucially, "accessible" to request by users. Practically, this has involved listing accessions maintained by genebanks and similar institutes that can reliably distribute samples, typically through formal agreements such as the Standard Material Transfer Agreements (SMTA). Thus, every accession listed in EURISCO corresponds to conserved seeds or plants that users can, in principle, obtain.

On-farm conserved landraces, however, do not neatly fit this access model. These varieties are maintained by individual farmers or local communities, which makes their availability to external users uncertain. In Spain, for example, there is a robust on-farm conservation network called Red de Semillas "Resembrando e Intercambiando," encompassing farmer-led seed networks and community seed banks that maintain local varieties and landraces *in situ*. These include regional seed-saving networks in nearly every autonomous community, with the NGO/community seed bank sector conserving several thousand unique landrace accessions (aggregating across all regional networks), especially vegetable and cereal landraces adapted to local conditions. Yet despite this rich conservation activity, unlike genebank-held accessions, breeders cannot simply order these on-farm landraces through standardized channels. Access typically requires direct communication and negotiation with the farmer, a process often challenging or impractical. Relying on direct farmer-to-user access is often impractical, as most farmers or local custodians do not see provision of the conserved resource as part of their role and are generally unwilling or unprepared to handle external seed requests. In other words, farmers maintaining landraces may be focused on their own use of the variety, and may lack the time, resources, or incentives to distribute seeds widely. This contrasts with genebanks, whose mandate includes distributing germplasm. Nevertheless, information about these on-farm conserved landraces is extremely valuable for understanding crop diversity, adaptation to local conditions, and traditional agricultural knowledge, making their documentation an important goal. Also, as with CWR populations conserved *in situ*, it is recommended that on-farm conserved landrace populations are backed up in a designated genebank and as for *in situ* conserved CWR backup samples so on-farm conserved landrace populations backup samples might be provided to users with the maintainers' agreement.

Beyond these access limitations, the very nature of on-farm landraces presents fundamental compatibility issues with EURISCO's framework. On-farm landraces are dynamic by nature. Their populations evolve over time through continuous cultivation and selection. They are often heterogeneous and can change from year to year. This dynamism means a landrace grown on-farm is not a fixed, preserved sample, but a moving target of genetic diversity. One season's planting might differ genetically from the next, and if a farmer stops cultivating it, the resource could disappear. EURISCO's current framework is built around stable accession records (with an accession ID, storage location, etc.), which aligns with static *ex situ* samples, not with continuously evolving on-farm populations. The private, localized maintenance of landraces (often by small-scale farmers on their own holdings) further means there is no guarantee those resources will continue to exist or be obtainable in the future, which again clashes with EURISCO's emphasis on secure and accessible conservation. Due to these fundamental differences, data on on-farm landrace populations have thus far been excluded from EURISCO. Nevertheless, various initiatives have made progress toward documenting Europe's on-farm conserved landraces, including Horizon projects Farmers' Pride and Dynaversity, and grassroots networks such as Let's Liberate Diversity (LLD), illustrating ongoing efforts and challenges in capturing and supporting this dynamic genetic heritage. These initiatives have also found that the existence of the majority of landraces maintained on-farm is not as precious as may previously have been thought the norm and many have been maintained by families or communities for extended periods (Veteläinen et al., 2009). Furthermore, the dynamism of on-farm maintenance could be seen as an advantage ensuring the landrace is adapted to a changing agro-environment, and any user wishing to obtain a genetically static sample could be supplied by the proposed backup on-farm *ex situ* sample.

5.2.5 Multi-Omics Data

EURISCO's current scope excludes genomic and other omics data (DNA sequences, markers, transcriptomics, metabolomics, proteomics, phenomics) and instead focuses exclusively on accession-level metadata and conventional phenotypic information. Researchers seeking genotypic data for accessions catalogued in EURISCO must consult external repositories, creating a significant gap in functionality as genomic characterization becomes increasingly essential for plant genetic resources work. Users frequently need integrated access to genetic markers, genome sequences, and allelic information alongside traditional passport data. Recognizing this limitation, the second part of this deliverable outlines strategic approaches for connecting EURISCO with external omics information systems to provide comprehensive, integrated data access that meets modern research requirements.

5.3 Factors Contributing to PGR Information Gaps in EURISCO

The information gaps in EURISCO stem from a confluence of adverse variables rather than discrete, isolated causes. The correlation between these factors is seldom linear. Rather, they create complex interdependencies in which challenges in one area often magnify limitations in others. This multifaceted nature explains why addressing documentation gaps requires comprehensive approaches rather than isolated and piecemeal solutions. To visualize these contributing factors, an Ishikawa diagram (Figure 3) has been developed that maps the root causes across several key dimensions. The subsequent sections explore these factors in depth, illustrating how their interactions shape the overall completeness and reliability of EURISCO.

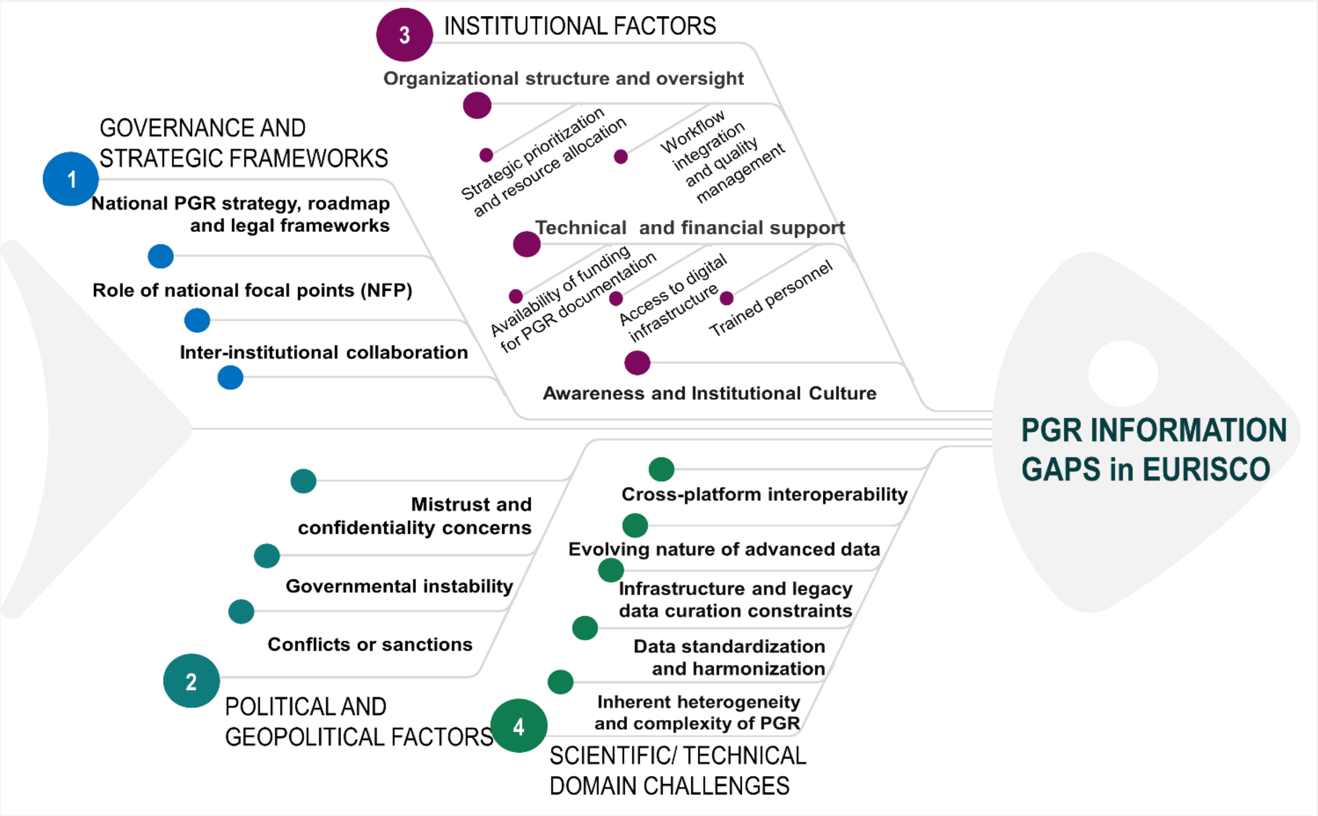

Figure 3. Ishikawa diagram of the root causes of PGR information gaps in EURISCO.

5.3.1 Governance and Strategic Frameworks

i. ***National PGR Strategy, Roadmap and Legal Frameworks***. Historically, PGR collection and conservation evolved organically through the independent efforts of breeders working across diverse research institutions (Diez et al., 2018; van Hintum et al., 2021). This development progressed without centralized planning or structural frameworks, growing somewhat haphazardly as individual scientists recognized the importance of conserving genetic diversity. In a sense the recent development of *in situ* and on-farm had an advantage of permitting more systematic planning of implementation and building on the *ex situ* model. The result is today's complex landscape where many countries like Spain, France, Italy and Ukraine house numerous separate genebanks, each with its own history, protocols, and specialized collections. This institutional fragmentation, while rich in diversity and local expertise, creates coordination challenges that extend to every aspect of PGR management, including documentation practices and information exchange. As institutions developed independently, so too did their approaches to recording, cataloguing, and sharing data about their collections. This documentation inconsistency underscores why strong governmental intervention through comprehensive national PGR strategies, detailed implementation roadmaps, and cohesive legal frameworks has become essential.

Where strong national PGR strategies or enabling laws exist, institutions tend to document PGR collections regularly and contribute data to national and regional databases. However, the presence and implementation of these frameworks vary widely. Some countries have dedicated PGR programs with clear roadmaps, whereas others have only recently formulated strategies.

Without robust legal mandates or comprehensive national strategies, documentation often becomes informal, fragmented, and inconsistent, which leads to substantial data gaps and irregular submissions to EURISCO. These challenges are frequently intensified by limited funding and capacity constraints. ECPGR (2021) highlighted that insufficient legislative support and resource limitations significantly hinder data-sharing efforts, causing the chronic underrepresentation of valuable accessions.

Dedicated government funding for PGR infrastructures enables institutions to conduct thorough documentation and submit more consistent and current information to EURISCO. In contrast, when PGR documentation depends on broader agricultural or scientific research budgets without dedicated funds, both data quality and submission frequency decline. Even internationally binding agreements, such as the Nagoya Protocol or the ITPGRFA, significantly depend on national implementation. Ambiguous national regulations or unclear institutional practices, for instance, uncertainty among germplasm holders about documentation of legal status or data-use conditions, can discourage data sharing and directly contribute to gaps in EURISCO. Effective national implementation typically includes clear procedures (e.g., consistent use of SMTA), which facilitate confident and comprehensive data sharing.

In situ and on-farm conservation further enhances but also complicates national PGR management. CWR in protected areas often fall under environmental authorities whose conservation priorities and regulatory frameworks might not align with agriculture-oriented institutions. Similarly, landrace conservation on farms involves diverse stakeholders, including local communities and extension services, which operate differently and engage variably with national PGR programs. Without coordination or clearly articulated supporting policies, these essential *in situ* resources often fail to enter formal documentation pipelines, resulting in incomplete NI, poor conservation and a complete lack of wider utilisation.

A well-structured national strategy supported by clear legal frameworks, stable funding, and effective inter-agency collaboration can address the complexities of *ex situ* and *in situ* conservation. Clearly defined institutional responsibilities, refined data standards, and coherent coordination among ministries and stakeholders enable countries to document their genetic resources more accurately. Such alignment enhances national reporting and strengthens a country's ability to sustainably manage and safeguard its PGR.

ii. ***Role of National Focal Points (NFP).*** As discussed in section 3, the NFP serves as essential intermediaries between national institutions holding germplasm collections / populations and EURISCO, ensuring that submitted data align with agreed standards and formats. When NFP have clearly defined mandates and institutional backing (e.g. as part of a national program or relevant ministry), data updates tend to be regular and reliable. Conversely, frequent turnover or lack of official support can disrupt the continuity of data submissions, reflecting the importance of maintaining both expertise and authority over time. While each country independently determines its frequency of updates and internal quality management, those embedding the NFP role in stable institutions generally achieve greater consistency in data contributions. Without such institutional embedding, maintaining regular updates and comprehensive national participation becomes significantly more challenging.

iii. Inter-institutional Collaboration. A consistent approach to national PGR documentation depends on effective coordination among ministries, GRC, research centers, protected area managing agencies, farmer's networks and other pertinent stakeholders. Countries with structured coordination mechanisms or formalized PGR networks typically streamline data collection and verification processes, minimize redundancy, and efficiently identify coverage gaps. These collaborative structures facilitate inclusive participation from diverse institutions, ensuring that both major and minor collections are accurately represented. In contrast, in countries with fragmented institutional structures and informal collaborations, i.e., where individual GRC or research centres maintain isolated records with inconsistent data-sharing practices, NI can suffer from data inconsistencies, overlapping or omitted accessions. Regional partnerships like the ECPGR play a vital role in mitigating some of the challenges that arise from such fragmentation. By promoting harmonized data standards and facilitating cross-border knowledge exchange, they enhance the effectiveness of national PGR programs. Nonetheless, without a clear, centralized entity to oversee and coordinate data exchange, collections risk being underreported or updated only sporadically. Creating a cohesive national coordination mechanism, such as a dedicated PGR program or a formal network led by the national GRC, helps ensure that institutions maintain a comprehensive and updated view of their holdings, *in situ* or on-farm conserved accessions or populations, while also improving the reliability and timeliness of data shared with regional or international platforms.

5.3.2 Institutional Factors

i. Organizational Structure and Oversight

a. ***Strategic Prioritization and Resource Allocation.*** When institutions clearly define PGR documentation as an organizational priority and support this commitment with dedicated funding, qualified staff, and efficient documentation systems, the resulting data are typically comprehensive, consistent, and frequently updated. For example, the Netherlands' Centre for Genetic Resources (CGN) has strategically embedded documentation into its institutional mission, enabling consistent investment in specialized staff and robust data-management platforms. Similarly, Nordic institutions coordinated by NordGen illustrate how sustained institutional focus and shared resources foster regular validation, standardized workflows, and timely data updates.

However, as discussed in the previous section, not every organization holding agriculturally significant genetic resources aligns with a formal germplasm conservation and distribution mandate. Many universities, research institutes, protected area networks, botanical gardens and farmers primarily focus on research, education, public engagement, or crop production. In these cases, PGR documentation may be secondary, and as a result, record-keeping can be fragmented, with limited data sharing, causing certain valuable collections to remain underrepresented in NIs and EURISCO.

But even institutions with a formal mandate to conserve and manage PGR do not always give documentation the same level of attention as other core activities, such as protected area managers active maintaining CWR population or farmers maintaining unique trait diversity. While systemic constraints, like limited budgets and staffing concerns, undeniably shape operational decisions, they often reflect an underlying lack of recognition that systematic record-keeping is just as vital to fulfilling the institution’s mandate. Without explicit directives, dedicated resources, and well-defined workflows, data management can become fragmented and sporadic, leaving important accession information incomplete or outdated. Over time, these gaps undermine the comprehensiveness and reliability of national and European PGR repositories, reducing their value for both conservation and research. If institutional mandates are to be fully realized, i.e., ensuring maximum PGR diversity conservation and facilitating conserved germplasm is readily available for use, documentation must be acknowledged as a key operational priority and supported by the appropriate structures and investments.

b. ***Workflow Integration and Quality Management.*** The efficiency with which an institution's databases interface with EURISCO’s reporting framework is another determinant of data gaps. Some institutions still rely on older or custom-built platforms not originally designed for external data sharing and rely on proprietary structures and terminologies misaligned with EURISCO standards. Other institutions employ contemporary specialized genebank software such as GRIN-Global yet still face integration hurdles due to varying data models and terminologies. This technical diversity, which spans advanced database solutions, spreadsheets, and even paper records, complicates standardization efforts. Institutions adding new data fields for specialised research create further difficulties since they call for extra mapping to meet EURISCO's standardised descriptors.

Converting institutional data into EURISCO-compliant formats frequently introduces significant workflow bottlenecks. Many institutions rely on manual reformatting for national inventory submissions, which heightens the risk of errors, inconsistencies, and delays. Smaller facilities that lack dedicated IT support often submit data sporadically or incompletely, further contributing to information gaps. Data workflows within institutions also tend to be fragmented by departmental silos. Collection, characterization, and distribution units may each maintain separate systems with limited interoperability, thereby creating multiple transfer points at which information can be lost. These challenges are compounded by a common disconnect between curatorial staff who hold deep biological knowledge and the IT teams managing the information systems, an organizational gap that increases the likelihood of errors during EURISCO submission.

Institutional priorities likewise shape the completeness and accuracy of data submissions. C&E activities often serve specific research objectives, leaving valuable data locked in project-specific formats rather than merged into the institution's main database. In some cases, institutions see minimal benefit in expending the additional effort to prepare comprehensive submissions for EURISCO, thereby perpetuating data coverage gaps in the broader network.

Robust quality assurance (QA) practices at the institutional level are another critical yet often underutilized mechanism for reducing information gaps. Although national-level validation processes exist, data reliability ultimately depends on institutional QA steps that uphold accuracy, consistency, and completeness before data enters the NI pipeline. Adhering to community standards is itself a QA step. Forward-thinking institutions also adopt custom interfaces with built-in validation rules that, for example, flag taxonomic inconsistencies, detect implausible collection dates or locations, and ensure mandatory passport fields are properly populated. This proactive strategy prevents error propagation far more effectively than reactive checks later in the process. However, QA protocols vary markedly across the continent. Some institutions implement formal policies and standard operating procedures, while others rely on ad-hoc approaches that allow inaccuracies to persist.

Technical capacity also differs substantially. Well-resourced institutions typically operate modern database systems that can enforce referential integrity, maintain controlled vocabularies, and track data provenance, which reduces the frequency of systematic errors. In contrast, smaller or underfunded institutions often rely on manual data entry or spreadsheet-based validation, which can leave many errors undetected and contribute to uneven data quality. Incorporating domain expertise into QA is similarly crucial. Taxonomists, GIS specialists, and crop experts can pinpoint subtle inconsistencies, such as incompatible environments for particular species or contradictory trait information, beyond the scope of automated checks. However, disparities in access to this specialized expertise might lead to widely varying data quality across institutions.

Building on these QA strategies, thorough documentation of procedures, validation criteria, and data confidence levels provides essential context for both contributors and end users. Comprehensive metadata helps data users assess suitability for specific purposes and guide NFP during compilation. Conversely, weak documentation practices obscure data limitations and usage contexts, creating metadata gaps that diminish the overall value of information incorporated into EURISCO.

ii. Technical and Financial Support

a. ***Availability of Funding for PGR Documentation.*** Although maintaining PGR documentation is an ongoing task, many countries do not have a dedicated budget allocation for this core function (Genetic Resources Strategy for Europe, 2021). Consequently, some NFP often update EURISCO on an *ad hoc* basis, fitting it around other duties due to the absence of earmarked funding or dedicated personnel for data management. These financial constraints result in infrequent updates (e.g. a country might go several years without submitting new data, see Table 2) or incomplete data (if manpower is not available to gather and format all required information). Conversely, when funding is available, either through national programs or external grants, countries can adequately invest in data curation. The initial establishment of EURISCO, for instance, was supported by an EU project (European Plant Genetic Resources Information Infra-Structure (EPGRIS)) (Weise et al., 2016), which enabled many countries to develop their NI and submit large datasets. Some countries have since secured national funds to maintain these inventories and sustain regular updates, typically annually, demonstrating the impact of consistent funding on data availability and quality. Continuous financial support facilitates hiring dedicated data managers or allocating sufficient time for NFP to collaborate closely with collection holders, perform systematic data quality assessments, and implement necessary system upgrades. In the long term, inadequate funding not only leads to data entry backlogs and delayed technological improvements but also jeopardizes institutional memory, especially when experienced staff retire without replacements due to budget restrictions. However, there is a positive trend in many European countries, where PGR documentation has increasingly been recognized as a crucial obligation under the ITPGRFA. Such commitment ensures at least basic, continuous financial support, critically underpinning EURISCO's capacity to provide current, reliable, and comprehensive data.

One PGR component that may be better funded in terms of documentation is where the *in situ* or on-farm resource is supported financially by some form of environmental stewardship funding. Funding is provided from governmental sources to support the public good value of maintaining a particular CWR or landrace resource as happens in many European countries (Martin et al., 2023). The maintainer payment is necessarily associated with strict description of the resource to avoid fraudulent claims and such data could be of benefit to the broader PGR community.

b. ***Access to Digital Infrastructure.*** Technical capacity for data management differs greatly across institutions, influencing their ability to provide accurate and timely information. Some institutions benefit from modern, integrated database systems and reliable high-speed internet connectivity, which simplifies data compilation and submission processes. Smaller or less resource-rich collections, on the other hand, may continue to rely on basic tools such as Excel sheets or outdated software solutions. Overall, considerable progress has been made in Europe toward digitizing PGR data, greatly reducing data exchange burdens. Where advanced infrastructure exists, an NFP may frequently extract an updated dataset with a few clicks (for example, running a query in a national genebank information system that outputs the MCPD fields for all accessions). Some countries have even implemented automated or semi-automated pipelines, which streamline the extraction and formatting processes, and allow NFP to quickly validate and upload data. However, without digital infrastructure, the data submission procedure becomes more labour-intensive, involving manual aggregation of records from multiple institutions, complex format conversions, and resolving encoding discrepancies. Poor IT conditions, such as slow internet speeds, insufficient server storage, and inadequate data backup solutions, can deter frequent updates, as NFP may be concerned about data loss or struggle with transferring large files. Equally important to hardware and infrastructure are trained IT personnel capable of maintaining databases, troubleshooting data issues, and integrating new technologies. The use of standardized software has also been encouraged. Over the years, adoption of GRIN-Global, an open-source genebank data management system, has helped strengthen local capacities in several European genebanks. For example, national genebanks in the Czech Republic, Portugal, and the United Kingdom have adopted GRIN-Global (https://www.grin-global.org/) to manage their germplasm collections more effectively. Finally, continual investment in both technical infrastructure and human capacity is critical. Strengthening these elements not only facilitates timely, large-scale data contributions to EURISCO but also ensures continuous improvement and sustainability of PGR documentation systems.

c. ***Trained Personnel for Data Management*.** Even with good hardware and software, skilled human resources are required to manage PGR data effectively. ECPGR has prioritized capacity building in documentation through regular EURISCO training workshops aimed at enhancing the skills of NFP. These trainings address essential topics such as data quality, the EURISCO upload interface, handling phenotypic data, and other important curation aspects like assigning DOIs to accessions. Such initiatives directly reduce disparities in expertise among member countries. Given that not all countries have dedicated data specialists, continuous sharing of knowledge and best practices remains critical. Trained personnel at the national level significantly influence data quality and contribution frequency. Their expertise ensures efficient standardization, data cleaning, and adherence to evolving EURISCO guidelines. On the flip side, insufficient training can lead to hesitancy in updating data or inconsistencies in submissions. However, training is not a one-off need. Recognizing that these needs evolve alongside information systems and standards, EURISCO employs ongoing capacity-building activities, including webinars, helpdesks, and mentoring from the coordination team. Additionally, some countries have established documentation teams or committees instead of relying on individuals, distributing tasks and knowledge effectively and preventing bottlenecks due to limited expertise.

iii. ***Awareness and Institutional Culture.*** The degree to which data management is prioritized within an institution strongly affects the thoroughness and reliability of its documentation. Even when infrastructures, such as GLIS for assigning DOI or EURISCO's support for integrating these identifiers, are in place, the decision to implement these tools often hinges on whether staff and leadership view data as a foundational resource or as a peripheral administrative requirement. Institutions that cultivate a data-centric culture typically invest in specialized documentation roles, structured training programs, and regular knowledge exchange. These measures encourage the timely adoption of emerging standards, thorough descriptor completion, and proactive QA protocols.

By contrast, organizations that treat documentation as an afterthought rarely allocate sufficient funding or administrative support for sustained data improvement. PID and other recommended practices may be recognized in principle but remain sporadically applied if staff are neither trained nor incentivized to integrate them into daily workflows. Cultural biases can also create systematic variation in descriptor quality: taxonomically focused institutions often excel in species-level detail while overlooking geospatial or phenotypic fields, whereas conservation-oriented entities may do the reverse. Such selective emphasis leads to predictable coverage gaps in larger data repositories. Institutional adaptability to evolving documentation requirements is also closely tied to organizational learning structures. Cross-departmental committees, routine workflow reviews, and feedback loops between curators and data managers facilitate the rapid integration of new descriptors and practices. By contrast, rigid institutional cultures often adhere to longstanding approaches and resist adopting fields or standards that deviate from established norms.

Likewise, professional incentives can either strengthen or undermine these documentation practices. Where performance metrics recognize data completeness and quality, staff have compelling reasons to adopt new standards, descriptor fields, and keep records up to date. Conversely, where metrics prioritize aspects like collection size or publication output, thorough documentation is rarely incentivized, hindering uptake of otherwise valuable tools.

5.3.3 Political and Geopolitical Factors

EURISCO operates on the principle that each country retains sovereignty over its data. Governments decide what to share, how much detail to provide and how often to update. However, the same principle can also create intentional or unintentional data gaps under a range of political and geopolitical conditions.

i. ***Conflicts or Sanctions.*** In conflict zones, both physical collections and their information systems face severe threats, exemplified clearly by the ongoing war in Ukraine. The national genebank in Kharkiv was endangered by military action, field trial sites were destroyed, institutional budgets were drastically reduced, and staff were displaced (Crop Trust, 2024). Under these circumstances, humanitarian concerns understandably take precedence over anything else. Protecting genetic materials from immediate harm and destruction becomes the subsequent priority, while documentation and information management inevitably fall to a lower priority level. When conflict disrupts normal operations, genebank information systems often stagnate. Damaged or inaccessible servers, compromised records, or disrupted IT infrastructure prevent regular information flow to networks like EURISCO. This creates growing gaps that may take years to address even after conflicts subside. Beyond direct conflict, international sanctions and diplomatic strains can also curtail data sharing, not through formal bans but by restricting funding, travel, and technology transfers. This *de facto* isolation can be seen in countries under sanctions (e.g., Russia or Belarus), where participation in networks like ECPGR may be limited in practice. Historically, data exchange often rebounds only when diplomatic relations improve. Consequently, whether through warfare or broader political tensions, geopolitical instabilities create enduring obstacles to the timely documentation and maintenance of PGR information, ultimately leaving critical gaps in both physical collections and their associated databases.

ii. ***Government instability***. Political and institutional volatility can significantly contribute to PGR information gaps, as shifting administrative priorities, frequent leadership changes, and unstable budgets often derail the continuous data management needed for accurate inventories. While official documents rarely attribute these gaps exclusively to political upheavals, accounts from documentation and stakeholder interviews across Southeast and Eastern Europe indicate that national contributions to EURISCO tend to dip when ministries undergo reorganizations or research institutes face funding disruptions. This correlation, though not always explicitly documented, underscores how turbulent governance can impede data collection, processing, curation and submission over extended periods.

    In Bosnia and Herzegovina, for instance, the country is divided into the Federation of Bosnia and Herzegovina, and Republika Srpska, each with distinct or overlapping agricultural responsibilities. Coordination of PGR data often occurs via institutions in Republika Srpska, particularly the Institute of Genetic Resources at the University of Banja Luka, leading to limited coverage of material from the Federation side, which has its own cantonal ministries and procedures. This fragmented structure makes it difficult to compile a truly comprehensive NI. In Moldova (2009–2016), frequent leadership changes at the Ministry of Agriculture reportedly hampered genebank continuity and led to sporadic data submissions, as local staff struggled with shifting priorities and uncertain budgets. Albania's governmental reorganizations between 2013–2015 coincided with funding interruptions at the Agricultural University of Tirana, likely affecting their capacity to process and submit data.

    By comparison, Bulgaria and Romania, which both underwent repeated agricultural ministry reorganizations in the years leading up to and following EU accession, have since stabilized their systems and now regularly update their NI. EU membership provided a crucial impetus for these reforms, delivering legal alignment, funding, and structural changes that helped strengthen PGR data management. Sustained focal points, administrative continuity, and targeted investments in GRC have also played important roles in supporting more consistent submissions. Croatia, which joined the EU more recently, has similarly developed a robust data management framework and maintains a relatively up-to-date national inventory. While direct cause-and-effect links are not always formally documented, these examples suggest that stable governance, organizational continuity, and adequate resources create a more reliable environment for regular EURISCO data updates.

iii. ***Mistrust and confidentiality concerns.*** Despite international agreements promoting collaboration and fair access to PGR, many data holders remain hesitant to share detailed information. Concerns primarily involve the risk of others using shared data, such as genotypic and phenotype datasets, for commercial or research benefits without proper acknowledgment or benefit-sharing. These concerns intensify when data highlight unique adaptations, valuable landrace traits, or sensitive collection locations. When national laws or enforcement capacities seem insufficient, institutions often respond by sharing only limited descriptors or withholding specific data entirely. Although the ITPGRFA encourages transparency and mutual benefit, ongoing mistrust continues due to past disputes, unclear legal protections, and fear of unauthorized appropriation. Clear legal frameworks, explicit benefit-sharing agreements, and transparent attribution standards can increase stakeholder willingness to share data openly. Without these safeguards, however, confidentiality concerns will likely persist, limiting the effectiveness and scope of PGR data exchange.

5.3.4 Scientific and Technical Domain Challenges

i. ***Inherent Heterogeneity and Complexity of PGR.*** The biological complexity of plants themselves, with varying reproductive systems, life cycles, adaptive characteristics, genetic backgrounds, and taxonomic relationships, inherently creates challenges for consistent documentation, as different types of germplasm require different forms of management, at least partially different descriptors and characterization approaches. The complex nature of PGR also manifests in the multitude of forms of curation, traits, properties, and characteristics that could potentially be recorded, from basic passport data to detailed molecular characterization, phenotypic evaluations, and environmental adaptations. Given limited resources, genebanks must prioritize which data to collect and digitize, inevitably creating information asymmetries across collections. Additionally, the evolutionary and dynamic nature of PGR means that information needs evolve over time as scientific understanding advances and new technologies emerge, which may result in temporal gaps in older collection data. These inherent complexities make it extraordinarily difficult to achieve complete, standardized documentation across the entire spectrum of PGR in EURISCO, despite ongoing harmonization efforts. The result is an information landscape characterized by varying depths of data, inconsistent coverage across taxa and collections, and persistent gaps that reflect both the biological complexity of the materials themselves and the historical development of PGR conservation practices throughout Europe.

ii. ***Data Standardization and Harmonization***. A major source of information gaps in PGR repositories lies in the fragmented approaches to collecting, organizing, and describing data across diverse types, formats, and institutional contexts. Pro-GRACE deliverables 1.1, 1.2, 2.3 and 1.4 examined these challenges in detail, emphasizing how adopting common frameworks and data standards, such as MCPD, MIAPPE, PID and ontologies, can streamline data structures, standardize taxonomic references, and unify metadata practices.

iii. ***Infrastructure and Legacy Data Curation Constraints***. EURISCO's technical infrastructure was originally designed primarily as a centralized *ex situ* passport data repository, built upon a traditional Oracle relational database optimized for robust data storage, integrity, and structured queries. While this foundational architecture has effectively supported EURISCO's core functions, its initial design did not fully anticipate the evolving complexity and diversity of contemporary PGR-related datasets. EURISCO's infrastructure will be further challenged by the inclusion of *in situ* CWR and on-farm landrace dataset, but it is hoped the increased breadth of trait diversity available to users will significantly outweigh the current challenges. Nevertheless, recognizing these evolving demands, EURISCO has been continuously improving and adapting its infrastructure to progressively accommodate new data types, functionalities, and user requirements, and avoid new information gaps. This ongoing evolution ensures that EURISCO remains responsive and relevant to contemporary PGR research needs (as discussed below).

A significant factor contributing to current information gaps is the inherent complexity surrounding legacy data curation across the broader PGR community. Traditionally, genebanks and research institutions developed their collections and databases independently, often driven by localized priorities, available technologies, and evolving documentation practices. This resulted in datasets maintained in a variety of formats, ranging from paper records and spreadsheets to early-generation information systems, each with varying completeness, consistency, and metadata standards. Consequently, when these data are aggregated into EURISCO, significant manual effort is often required to harmonize historical terminologies, interpret outdated descriptors, and map older data classifications to current standards.

Crucially, datasets collected before the establishment of modern PGR and data standards present substantial challenges that extend beyond simple data harmonization. They typically lack precise geolocation coordinates, habitat descriptions, standardized collection methodologies, and comparable phenotypic characterizations, among others. Because these details were never systematically recorded, or have since been lost, filling these gaps is seldom possible, even with current technologies. Although partial information can occasionally be recovered from archived field notes or oral accounts, such efforts rarely compensate fully, leaving fundamental limitations that may permanently diminish the scientific value of these legacy collections.

iv. ***Evolving Nature of Advanced Data.*** Advances in genomics, high-throughput phenotyping, and environmental data collection have introduced new data types that were not envisaged when systems like EURISCO were first designed. Today, researchers may sequence the genomes of genebank accessions or genotype them with tens of thousands of markers. These datasets can be enormous (gigabytes of sequences or matrix data) and are typically stored in specialized repositories (e.g. EMBL/GenBank for sequences, or institutional databases for genotypes) rather than genebank catalogs. EURISCO as an aggregate does not include genomics or other "omics" data in its current scope. This means that while the passport record for an accession is in EURISCO, any genomic characterization of that accession lives elsewhere, making it harder for users to find all relevant information in one place. EURISCO acknowledges this and highlights the importance of linking out to those external data sources via stable identifiers. In practice, this requires assigning DOIs to accessions and ensuring genomic information systems reference those DOI, or building portals that interconnect EURISCO entries with external genomic data. Phenotypic data, too, has evolved rapidly. Multi-site, multi-year evaluation trials and high-throughput phenotyping (HTP) platforms generate large, heterogeneous datasets with extensive metadata needs. While EURISCO has started to include phenotypic observations, which is a huge step forward, the contributed phenotyping datasets were very heterogeneous in how they were described and formatted, especially the older ones. Many lacked crucial metadata (e.g. environment of the trial, measurement protocols). To make such data usable, EURISCO recognized the need to enforce standards like MIAPPE for phenotypic experiments. This demonstrates how the data model must evolve. New fields and reference ontologies are needed to capture experiment design, traits, units, etc., which go far beyond the original passport descriptors. Additionally, storing and querying these data in a relational schema can be complex, especially as the volume grows by order of magnitude. To manage the complexity, one approach has been to develop a separate but linked platform. The EURISCO-EVA system (for European Evaluation Network data) was created to handle the standardized collection and analysis of multi-site evaluation data. EURISCO-EVA's design addresses fragmentation by enforcing common protocols across trials and integrating passport and phenotypic data from those trials in one place (Kumar et al., 2024). This highlights that new kinds of data often require new infrastructure or major extensions to existing ones. Without such adaptations, data from advanced breeding trials would remain fragmented and not integrated into EURISCO's main catalogue.

v. ***Cross-platform Interoperability***. No single system can cover all aspects of PGR data, making interoperability between platforms crucial. EURISCO operates within a global ecosystem of genetic resources databases. It is the European node of the worldwide Genesys information system and a part of the FAO's Global Information System (GLIS) on PGRFA. In theory, this networked approach allows data from regional, global, and crop-specific databases to complement each other. In practice, achieving seamless data exchange and linking remains a challenge. One issue is that EURISCO's current design and data-sharing mechanisms have limitations in real-time interoperability. For example, EURISCO regularly feeds updated European passport data into Genesys (ensuring global visibility of those accessions), but phenotypic data are not passed on and are provided exclusively through EURISCO. This means a user searching Genesys won't see evaluation data that actually exist, unless they know to separately check EURISCO. Without tight integration, users may have to query multiple systems and manually reconcile the results, which is inefficient and can lead to missed information. Essentially, each platform's scope and technical constraints can create silos of data.

   From a technical standpoint, EURISCO was designed primarily as a web portal for human users, so programmatic (machine-to-machine) access is limited. There is no publicly available API, which hinders automated data sharing with other platforms. To address this, EURISCO is considering adopting the Breeding API (BrAPI) (Selby et al., 2019). Implementing BrAPI endpoints would allow external software to query EURISCO programmatically, streamlining cross-platform data flows and paving the way for deeper integration with advanced "omics" repositories.

   As previously discussed, consistent use of PUID is also crucial for referencing the same accession across different databases. However, bridging diverse database infrastructures (e.g., Oracle vs. MongoDB) can still be complex. Projects like GLIS are working to unify these efforts, with EURISCO playing a key role in that network. Realizing true global interoperability will require ongoing development, such as implementing robust APIs, synchronizing updates among information systems, and enhancing user interfaces. Future integration with external "omics" archives similarly depends on stable APIs and shared data models to ensure that large-scale genomic or other specialized datasets link back to EURISCO's accessions. By expanding accession-level connections and leveraging DOIs, EURISCO can ultimately help users access all relevant PGR information in one place, reducing fragmented data silos and information gaps.

6. Unified Strategy for Interfacing Different Information Systems with EURISCO

Building on the challenges discussed previously and reiterated in the subsequent section below, one key question is how EURISCO can remain a cohesive reference point amid the rapidly expanding scope, depth and diversity of contemporary PGR data, especially as these developments preclude the viability of consolidating all of it in a single data infrastructure.

A particularly well-suited approach is a hybrid federated model, in which EURISCO retains its function as a principal PGR data aggregator and discovery hub while interlinking with specialized repositories that focus on deeper curation and analytics. In this approach, a central platform provides unified discovery and indexing, while detailed data remain distributed at source repositories, and are virtually accessed or fetched on-demand via APIs. This balances the benefits of a centralized search (ensuring findability in FAIR) with the flexibility of distributed data ownership (each repository maintains control and updates of its data, aligning with accessibility and local governance). The hybrid model avoids complete data duplication by federating queries to external systems when needed, yet it can cache or index key metadata centrally for performance. Adopting this hybrid federated approach will therefore allow EURISCO to remain current with scientific and technological advances. It reinforces its role as a trusted gateway to Europe's PGR while leveraging the agility and specialization that a decentralized network provides, all within a consolidated platform.

6.1 Rationale for a Hybrid Federated Model

6.1.1 Growing Diversity and Volume of PGR Data

Data generated by contemporary PGR research now extends well beyond the limited scope of standard MCPD descriptors. The application of *in situ* CWR and on-farm landrace population description and management requires news sets of descriptors (Moore et al., 2008; Bioversity and The Christensen Fund, 2009; Iriondo et al., 2012; Negri et al., 2012; Thormann et al., 2013, 2017; Maxted et al., 2016; Magos Brehm et al., 2017a, b; Caproni et al., 2020; Weise et al., 2020; Alercia et al., 2022; Phillips et al., 2025). Also complex workflows combining phenotypic observations, whether from semi-automated sensors, advanced imaging, or high-throughput field evaluations, can produce dense, time-series datasets of considerable size (Rebetzke et al., 2019; Fasoula et al., 2020; Sheikh et al., 2024). At the same time, multi-omics studies (e.g., genomics, transcriptomics, proteomics, metabolomics), leveraging modern sequencing and molecular profiling techniques, routinely generate files of a volume and complexity that far surpass what traditional repositories were built to manage. Continuous efforts to enrich phenotypic and genetic data with environmental, climatic, and agroecological dimensions further add to this explosion in scale and heterogeneity. A single-system approach struggles to absorb and harmonize these multifaceted data streams, creating significant obstacles to interoperability and consistent metadata curation. *(Different PGR-associated data types and their characteristics were discussed in detail in Pro-GRACE Deliverable 1.4).*

6.1.2 Preserving EURISCO's Core Strengths

EURISCO's strength has always been its role of promoting access to conserved PGR resources and as a streamlined aggregator that standardizes passport information and consolidates C&E data from various national sources. What makes it valuable to researchers is its efficient submission process, consistent reference framework for germplasm discovery, and reliable structure for updates. Overloading this system with massive, domain-specific datasets, such as raw imaging files or the entirety of next-generation sequencing outputs, would undermine these core strengths. Data ingestion would become slower, metadata checks more unwieldy, and system maintenance far more labour-intensive, jeopardizing the reliability and responsiveness that researchers currently rely on. To accommodate more intricate or voluminous data, the aggregator's database can instead include minimal reference fields, e.g., persistent unique identifiers, that guide users to the relevant external repositories. This approach allows specialized archives to manage their data on their own terms, using formats and curation practices suited to their particular research domains, without placing undue strain on the aggregator. Researchers still benefit from a centralized resource for PGR data discovery, with immediate links to deeper, domain-specific datasets when needed.

6.1.3 Technical Limitations of a Fully Centralized System

As the scope of PGR research continues to expand, a single repository that attempts to integrate every category of data, ranging from multi-terabyte omics outputs to continuous sensor streams and *in situ* population time series monitoring information, confronts a myriad of escalating challenges. At the most fundamental level, storage infrastructure can become overwhelmed by large and constantly growing datasets, leading to lengthy ingestion processes and higher operational costs for capacity upgrades. Such volume also exerts pressure on database performance, with query times increasing and system responsiveness declining as more users attempt to access and analyze these data. Meanwhile, relying solely on traditional relational engines can exacerbate these problems, since they were not originally designed for rapid ingestion of unstructured or semi-structured data such as streaming field measurements or raw sequencing files. Moreover, centralized systems also encounter difficulties in preserving consistent metadata and in managing version control, especially when multiple research communities each adopt distinct schemas, ontologies, or data standards. Even subtle changes in one domain can ripple through the entire repository, forcing large-scale reconfigurations that raise maintenance burdens and increase the likelihood of errors. Where high-throughput phenotyping might emphasize rapid, iterative updates, omics projects may require intricate version control and annotation processes. Attempting to unify these distinct workflows under one overarching repository can lead to protracted development cycles, synchronization issues, and inconsistent data quality. Ultimately, these combined factors undermine the ability of a single system to offer the reliability, scalability, and adaptability needed for modern PGR research.

6.1.4 Reduced Redundancy and Clear Division of Responsibilities

One of the advantages of federation is the prevention of unnecessary duplication. Instead of importing complex or high-volume data into the aggregator, each external source retains its native files and data structures, ensuring they remain accurate, comprehensive, and up to date. The aggregator, in turn, holds succinct pointers, i.e., commonly PUID (digital object identifiers (DOIs), uniform resource identifier (URIs), or other stable links), to those external repositories. This division of responsibilities also underpins a transparent governance structure. The aggregator's role is to maintain an easily navigable index of European PGR resources and to ensure the consistency of core descriptors. Specialized archives take charge of organizing advanced data, versioning procedures, and domain-specific analysis pipelines. Data ownership and autonomy remain firmly in the hands of the original contributors, be they NFP, research consortia, or community-based programs, who can decide how best to manage their resources over time.

6.2 Centralized vs. Federated vs. Hybrid Federated data

Integrating diverse data systems requires determining which data should be stored centrally and which should remain distributed. Although EURISCO already adopts a partially federated approach by aggregating information from NI, this vision can be expanded to incorporate additional external archives in a broader hybrid federated model (Figure 4). Under this model, EURISCO continues to serve as a central aggregator while allowing data to be retrieved on demand from external repositories when keeping it at the source is more efficient. In a federated system, multiple nodes or databases operate within a network, often guided by a central index that records the location and nature of each data source. Consequently, some data remains fully integrated in EURISCO, whereas other datasets are accessed dynamically from external services. This ensures that users can discover all information through a single-entry point (EURISCO), while data remains wherever it can be best maintained, achieving an optimal balance between integrated discovery and efficient, distributed curation.

- *Central Aggregator (Hub-and-Spoke Model) with Decentralized Governance*
  EURISCO will continue to serve as the central aggregator (the "hub") for PGR datasets (*ex situ* and *in situ* CWR passport and C&E data) fed by decentralized NI (the "spokes"), which remain responsible for data curation and submission, via unidirectional integration (i.e., secure file transfers). These datasets are relatively small per record and broadly shared. This will still include *ex situ* passport, *in situ* CWR passport, and C&E datasets. Furthermore, phenotypic datasets generated by short-term projects and often stored in project-specific or ephemeral databases will be integrated into EURISCO once the projects or their embargo periods conclude, following EURISCO's C&E data exchange standard (*refer to section 5.4.5 for further details on C&E data)*

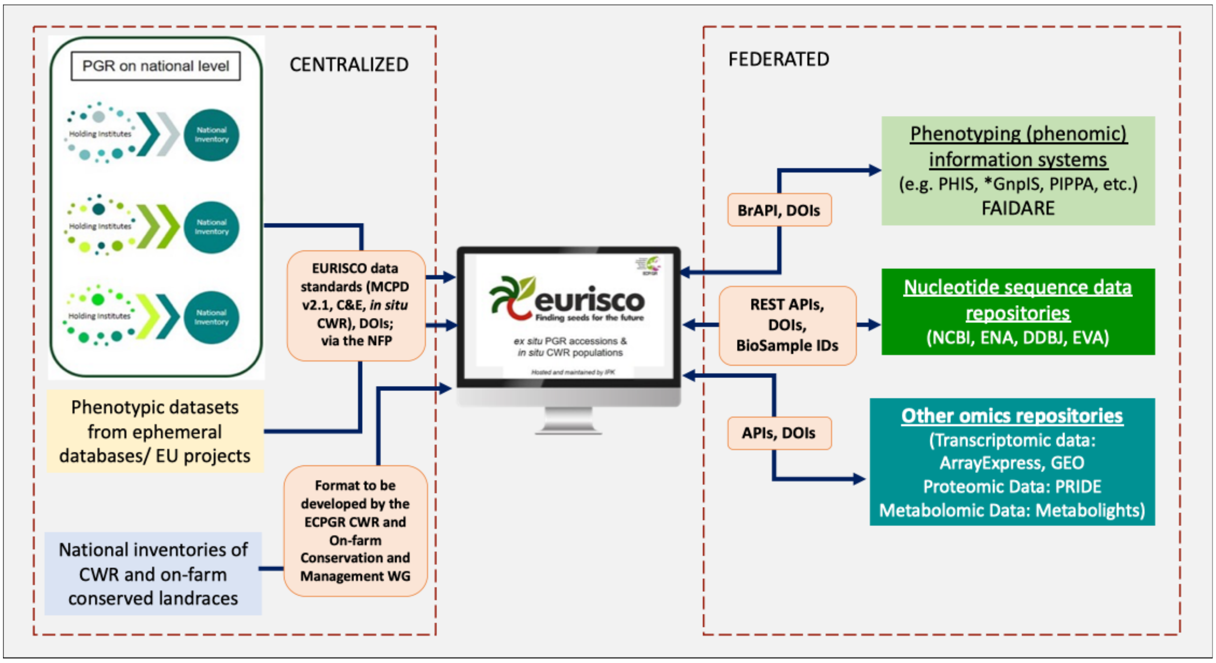

Figure 4. Proposed hybrid federated model for EURISCO, *upholding its central role in aggregating standardized PGR data (MCPD v2.1, in situ CWR passport descriptors, and C&E data) submitted via National Inventories. Additionally, phenotypic datasets from ephemeral databases and concluded EU projects are uploaded per EURISCO's data exchange standard for C&E data, ensuring their long-term accessibility within EURISCO. National inventories for CWR and on-farm conserved landraces (taxon-level), for which the formal inventory format will still be developed, will also be centralized. High-volume, dynamic, or specialized PGR-associated datasets, such as extensive phenotypic records, high-throughput sequencing outputs, genomic variation analyses, and integrative multi-omics data, remain housed in domain-specific repositories optimized for storage, updates, and analysis. EURISCO retains only essential metadata and PUID (e.g., DOIs, BioSample IDs) and leverages standardized Application Programming Interfaces (APIs) (Breeding API (BrAPI), Representational State Transfer (RESTful API)) for on-demand retrieval. This distributed approach enables data providers to update content in real time while EURISCO functions as a unified information system, directing users to the latest authoritative sources without overwhelming its infrastructure*.

***Integrative information system for phenomic, genetic, and genomic data**

- *Federated (Distributed) Data*

  All high-volume, highly dynamic, or specialized data will remain in federated external systems, with EURISCO acting as an integrative interface or information system to them rather than a host. As noted by Engels and Ebert (2021), most omics datasets related to genebank accessions are not managed by the genebanks themselves but generated via external projects and stored in specialized repositories. Keeping these data distributed subsequently will avoid overwhelming EURISCO's infrastructure and leverage the strengths of existing repositories optimized for those data types. Instead of importing these datasets en masse or constantly changing records, EURISCO will store metadata and pointers (links, identifiers) to where the data can be accessed. For example, raw sequence reads, variant call datasets, extensive multi-environment trial measurements, and omics experiments will not be copied into EURISCO. They remain in domain-specific databases that are built to manage such content (for storage, updates, analysis). EURISCO's role is to federate access by indexing key reference metadata and enabling on-demand retrieval. This distributed approach ensures that data providers (e.g. genomic archives) can update their data in real-time, while EURISCO always points users to the latest version. It also respects data sovereignty and security, as sensitive or proprietary datasets stay under the control of the original repository.

6.3 Technical Requirements

6.3.1 Persistent Unique Identifiers (PUID)

A hybrid federated model for integrating PGR-associated data relies on PUID to maintain stable cross-referencing among disparate information systems. A PUID remains valid independently of organizational or technological changes, thereby avoiding the risks associated with transient or locally assigned identifiers. This approach ensures the long-term resolvability of records and supports a scalable framework for sharing passport, phenotypic and multi-omics information (FAO, 2016). Crucially, a one-to-one mapping is enabled by these PUID. Each unique accession in one system corresponds exactly to one record in another, preventing conflation of data from distinct samples or duplication of the same accession under multiple identities.

DOI (https://www.doi.org/the-identifier/resources/handbook/) (Paskin, 2005) have gained particular significance in the domain of PGR. This stable and globally unique identifier ensures that one specific germplasm accession or *in situ* / on-farm population can be recognized and referenced consistently in multiple databases or systems. The ITPGRFA, through its GLIS, facilitates the assignment of DOIs to genebank accessions and *in situ* / on-farm conserved populations. These identifiers benefit from a recognized registration system and resolution services, thereby offering consistent references over time. EURISCO, functioning as a coordinating entity, enables GRCs to register DOIs and automatically updates the associated metadata when passport data changes. This service encourages standardized use of DOIs in platforms such as Genesys and WIEWS, and it addresses potential ambiguity arising from the varied combinations of institute codes and local accession numbers. The growing number of European GRC adopting DOI further demonstrates the feasibility of this approach.

When external datasets, including genomic or transcriptomic records, reference a DOI, integration becomes straightforward. The DOI acts as the permanent link to the corresponding passport record in EURISCO or another PGR catalogue. This direct linkage ensures unambiguous identification of the germplasm in question, allowing researchers to unify data layers, including passport, phenotypic, and omics data, under a single accession identity. If a DOI is unavailable, alternative metadata fields, such as institutional codes, voucher references, or locally assigned accession numbers, can be used to establish a match. While these locally assigned IDs may be unique within their respective genebank or institution, they often do not have global resolvability unless combined with recognized standards (e.g., FAO WIEWS codes). Maintaining clear mappings to these alternate identifiers is essential for bridging data from legacy systems and accommodating regions where DOI usage remains partial.

Nucleotide sequence archives such as NCBI or the ENA commonly provide fields for voucher or sample information that can link back to a genebank identifier. The BioSample ID assigned in these repositories serves as a globally unique reference for a given sample record, but to preserve the one-to-one mapping with the original germplasm, the accession identifier or a persistent ID such as a DOI must also be recorded in the sample metadata. Community standards, including the Minimum Information about any (x) Sequence (MIxS), recommend structured entries in these fields, yet practices vary across research initiatives. The Global Genome Biodiversity Network (GGBN) also promotes consistent referencing, although different institutions show varying levels of adherence.

Taxonomic identifiers add another dimension to this integration strategy. Data standards such as MIAPPE (Minimum Information About a Plant Phenotyping Experiment) and MIxS encourage using stable references like the NCBI Taxonomy ID, which enables unambiguous species-level alignment. Storing taxonomic identifiers in a central taxonomy backbone permits more confident incorporation of external datasets and reduces errors introduced by synonyms or informal naming conventions. By combining taxonomic references with PUID for accessions, researchers and genebank managers can more seamlessly unify data from distinct sources, reinforcing the stability and clarity of the hybrid federated model for plant genetic resources.

6.3.2 Standard Communication Protocols: API Specification

Implementing integration in practice relies on web service APIs, which are clearly defined methods that allow one software system to access the data or functionality of another (Sohan et al., 2015; Krishna & Sharma, 2021). Within a federated model that links EURISCO to external repositories, APIs function as bridges that enable real-time or periodic data exchange (See Table 7 for use cases/ examples). Two categories of APIs are particularly relevant in this context: domain-specific APIs such as BrAPI and generic RESTful APIs or web services made available by various repositories.

- **Breeding API (BrAPI)** is a standardized RESTful API specification developed specifically for plant breeding and genebank data integration (Selby et al., 2019). It defines endpoints for germplasm (accession information), studies/trials (experiment design), observations (phenotypic measurements), genotypes (marker or sequencing data), and more. BrAPI v2 aligns with MIAPPE for trial metadata and uses standard ontologies for traits when possible. Many contemporary plant data platforms implement BrAPI, including breeding databases (e.g., BreedBase, PHIS) and genebank portals. FAIDARE (FAIR Data-Finder for Agronomic Research: the French/ELIXIR agronomic data discovery portal) (https://urgi.versailles.inrae.fr/faidare/) uses BrAPI to integrate data from distributed nodes. Supporting BrAPI on the EURISCO side, whether by offering a BrAPI endpoint for EURISCO data or consuming BrAPI endpoints from external nodes, makes integration “plug-and-play” for any other system that implements BrAPI. This approach guarantees not just the ability to retrieve data, but also a shared schema and consistent terminology, satisfying the Interoperability and Reuse aspects of the FAIR principles.

- **General RESTful Web APIs.** Many repositories publish REST APIs for programmatic data access, as it provides flexibility by handling different types of calls, multiple return data formats and dynamically structured hypermedia (Krishna & Sharma, 2021; Sayers et al., 2022). In the context of federated data integration, these services allow on-demand retrieval of records from external databases, such as:
    - **National Center for Biotechnology Information (NCBI) Entrez E-Utilities** (https://www.ncbi.nlm.nih.gov/books/NBK25501/)- NCBI provides a suite of APIs (E-utilities) for searching and retrieving data across databases such as BioSample or the Sequence Read Archive (SRA). A user can employ e-search to locate relevant records and e-fetch to download the full details in XML or JSON. If a EURISCO accession has a corresponding BioSample or SRA entry, these APIs allow direct retrieval of that information. NCBI also offers more recent RESTful endpoints (e.g., the SRA data access API), along with a powerful query language. Returned data adhere to internal NCBI schemas, often overlapping with MIxS standards for sequence records.
    - **European Molecular Biology Laboratory- European Bioinformatics Institute (EMBL-EBI) APIs** (https://www.ebi.ac.uk/ebisearch/documentation/rest-api)- The European Nucleotide Archive (ENA) exposes REST endpoints for querying sequences by accession or BioSample, with results in XML or JSON. Other EBI services, such as the European Variation Archive (EVA), offer APIs for variant data queries, and ArrayExpress has an API for discovering gene expression experiments by keyword. MetaboLights similarly provides REST endpoints for study searches and metadata access, while PRIDE (Proteomics IDEntifications Database) supplies a REST API for proteomics dataset queries, experiment metadata, and protein/peptide identifications.
    - **FAIR Data Point / SPARQL** (https://www.w3.org/TR/sparql11-query/) **-** Certain systems offer semantic web interfaces via SPARQL endpoints, enabling RDF-based queries. Although not mandatory for most PGR data-sharing scenarios, this approach can be beneficial for advanced use cases, such as cross-repository queries in RDF-based portals. In practice, simpler REST or BrAPI services tend to be more widespread due to their ease of implementation.
- **Domain-Specific APIs and Formats**
    - **GEO and ArrayExpress-** The Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) at NCBI can be accessed using NCBI APIs or specialized

tools like GEOquery in R (https://github.com/seandavi/GEOquery) . GEO organizes data into series (GSE), samples (GSM), and so forth, making it possible to retrieve metadata and even expression matrices programmatically. ArrayExpress (https://www.ebi.ac.uk/biostudies/arrayexpress) offers both a REST API and FTP-based downloads keyed by accession identifiers, while the Expression Atlas provides additional layers of functional annotations. Linking germplasm identifiers to GEO or ArrayExpress records enables cross-referencing gene expression data within an integrated framework.

- **NCBI SRA / ENA-** NCBI's SRA and ENA both provide APIs for locating and retrieving sequence reads. While these repositories often handle large datasets (making full downloads impractical in some integration workflows), storing accession numbers in EURISCO allows outward linking for additional metadata on demand.
- **BrAPI vs. Custom APIs-** BrAPI covers a broad range of plant breeding scenarios, but some domains, such as raw high-throughput sequencing data, fall outside its scope. In these instances, native repository APIs are necessary. Middleware or "connectors" can reconcile one API's accession references with a repository's unique identifiers, returning normalized results to the calling application.

API Orchestration and Integration Layers: When connecting to multiple external APIs, implementing an orchestration layer can significantly streamline the interface or integration process. This layer determines how to route a given query (e.g., a germplasm identifier) to the relevant repositories, collects the outcomes, and consolidates them into a coherent response. By offloading query logic and data aggregation to a specialized service, EURISCO benefits from interacting with one consolidated endpoint rather than managing multiple API connections directly.

Authentication, Access, and Performance: Most public APIs, including those from NCBI, EBI, or BrAPI-based platforms, permit open read access, which aligns with the "Accessibility" principle of FAIR. Nonetheless, some information systems require API keys or tokens for controlled content, necessitating secure credential storage and respect for usage policies. In addition, rate limits and performance constraints can impact federated operations. Caching frequently accessed data in a centralized store can reduce redundant calls to external APIs. For instance, querying ENA for the same accession repeatedly can be avoided by retaining the relevant metadata locally and refreshing it at set intervals. This strategy preserves a near-real-time integration experience while mitigating excessive load on external services.

6.3.3 Structural Data Standards and Semantic Alignment

Interfacing various information systems with EURISCO requires a clear and consistent approach to data standards and semantic alignment. This ensures that data exchanged between systems is both technically compatible and meaningfully consistent. Structural data standards (e.g. MCPD, MIAPPE, MIxS) define a common format and structure for information (Ćwiek-Kupczyńska et al., 2016; Deng et al., 2023; Pommier et al., 2023), which makes it easier for EURISCO to parse, validate, and connect with external datasets. When EURISCO interacts with external repositories, the use of community-agreed data standards simplifies the process of retrieving and processing data via standardized APIs. For example, if external systems follow the same structure for passport information or phenotypic data, then the API calls made by EURISCO can reliably extract the necessary metadata. This uniformity reduces the need for custom data mapping and transformation, thereby lowering the chance of errors during data exchange.

At the same time, semantic alignment, achieved through the use of controlled vocabularies, and ontologies (e.g., Crop Ontology, Trait Ontology) (Shrestha et al., 2012), ensures that the meaning of data elements is interpreted in the same way across different systems. Without it, data coming from diverse sources may use different terms or measurement units for the same concept, leading to ambiguity and misinterpretation. By applying consistent terminology, along with PUID, EURISCO can accurately link related records even when they originate from various sources. This approach supports data accuracy and facilitates cross-system comparisons, ensuring that users receive a coherent and reliable dataset. *(For a detailed discussion on the data standards and ontologies pertinent to PGR, please refer to Pro-GRACE Deliverables 1.1, 1.2, and 1.4.).*

6.3.4 Metadata Registry

Metadata management ensures that data from heterogeneous sources can be reconciled and interpreted with consistency. Each external archive may adopt different file formats, naming conventions, or domain-specific terminologies, which can lead to inconsistencies or misunderstandings when data from different systems is linked or cross-referenced. A metadata registry addresses these challenges by describing the structure, semantics, and access rules of each dataset, and by recording how various schemas map to a shared reference model. This alignment guarantees that identical concepts in different repositories are recognized as equivalent. The registry also incorporates PUID and links to original records, enabling unambiguous cross-referencing of data entities across multiple systems. In addition, it tracks provenance, versioning, and licensing information, making it possible to trace each data update or annotation back to a defined set of rules. Such clarity supports reproducibility by documenting how data is transformed or merged, while also maintaining compliance through proper usage controls. As a result, the integration layer can query, transform, and unify data from multiple sources in a reliable and transparent manner, even as the underlying repositories evolve.

Table 7. Technical Requirements for the Proposed Hybrid-Federated Data Model (strengths, considerations, and example use cases)

Technical Requirements	Strengths	Considerations	Use Cases
BrAPI (Breeding API) (Selby et al., 2019)	• **Standardized for plant data-** BrAPI is an international standard API tailored to plant breeding data, ensuring consistent data structures (germplasm, trials, observations) across providers. • **Interoperable & FAIR-** Aligns with community standards like MIAPPE for metadata, enhancing interoperability and data FAIRness (Findable, Accessible, Interoperable, Reusable). • **Widely adopted in plant domain-** Many crop databases and breeding programs implement BrAPI, meaning a single client (like EURISCO) can integrate with many data sources using the same logic • **Low integration effort for providers-** Once a database implements BrAPI, its data can be indexed by portals like FAIDARE or accessed by any BrAPI client with minimal additional work.	• **Scope is specific-** BrAPI, especially v1/v2, is focused on germplasm, phenotypes, genotypes, etc., but not designed for raw sequence reads or complex omics data. It may not cover all data types (for example, no direct calls for proteomics or transcriptomics raw data). • **Version variations-** Different organizations might implement different BrAPI versions or subsets of calls, occasionally leading to compatibility issues. Keeping up with version updates requires effort. • **Performance considerations-** If large phenotypic datasets are served via BrAPI, paging through many records over an API can be slower than a direct database query (though this can be mitigated with filtering and indexing).	Best for distributed phenotypic and genotypic data integration in the plant domain. BrAPI is ideally suited when connecting multiple breeding/phenotyping databases in a federation. **Example**: EURISCO can use BrAPI to pull trait data from breeding trial databases in real time. Use BrAPI when data providers are willing to expose data via a standard API, as it provides a plug-and-play integration. A practical case is FAIDARE indexing: any repository that implements BrAPI (with MIAPPE-compliant data) can be instantly indexed in the FAIDARE portal, meaning EURISCO (as a BrAPI client or server) can both consume and provide data in a ready-to-federate manner.
REST APIs (Custom/Repository-specific) (Perez-Riverol et al., 2018; Sayers et al., 2022) https://fairsharing.org/	• **Flexible & Powerful-** REST APIs are provided by virtually all major bioinformatics databases (ENA, GenBank, GEO, ArrayExpress, PRIDE, MetaboLights, etc.), allowing programmatic retrieval of rich data. They often support complex queries (filters, search terms) to pinpoint relevant data (e.g., find all sequences for a given	• **Heterogeneity-** Each external API has its own schema, endpoints, and quirks. Integrating many APIs means writing custom connectors for each (one for ENA, one for GEO, one for PRIDE, etc.), increasing development and maintenance effort. There's no single "language", unlike BrAPI which is uniform, so the integration code must handle multiple formats and data models. • **Evolving endpoints-** APIs can change or be	Best for retrieving data from large, specialized databases where data are already well-curated and frequently updated. Use REST APIs when direct live access to the latest data is needed, or when the data type is outside BrAPI's scope. **Example**: to get genomic sequences, EURISCO would call ENA's API with a

Technical Requirements	Strengths	Considerations	Use Cases
	BioSample). • **No need to reinvent data store-** By using repositories' own APIs, EURISCO leverages the existing infrastructure. This means always up-to-date data from the source (e.g., as new sequences get added to ENA, they become available via API immediately). • **Language/platform agnostic-** REST APIs use standard HTTP, JSON/XML, etc., so any programming environment can use them. EURISCO's backend can call these APIs and integrate results seamlessly, treating remote data as if it were an extension of its own. • **Detailed data access-** Repository APIs often allow retrieval of detailed records and even bulk downloads. For example, PRIDE's API gives access to all metadata and files of a proteomics project, which a more generic API like BrAPI wouldn't directly provide.	updated. If a service alters its URL structure or output format, EURISCO's integration might break and require updates. Monitoring and maintaining compatibility is ongoing work. • **Rate limits & performance-** Public APIs might have rate limits or slower response times for large queries. EURISCO must be careful to fetch data efficiently (possibly caching results) to avoid slow user queries. Heavy reliance on external APIs means EURISCO's performance is partially at the mercy of those external services' performance and uptime. • **Authentication and access-** Some APIs (especially for sensitive data) require API keys or authentication. Managing credentials for multiple APIs and ensuring secure access can be complex. However, many public genetic databases allow open read access for public data (making this less an issue for most PGR data).	BioSample ID query. Use cases: Fetching a list of all sequencing runs for an accession's BioSample; querying GEO for any transcriptome study mentioning a given accession; pulling a metabolomics study's metadata from MetaboLights by study ID. In such scenarios, writing a small adapter for the repository's API yields immediate integration. Essentially, REST APIs are the go-to method for cross-domain integration, connecting EURISCO to the wider world of omics databases and repositories.
DOIs (Digital Object Identifiers) (Paskin, 2005) (https://www.doi.org/the-identifier/resources/handbook/)	• **Persistent & Citable-** DOIs provide a permanent, globally unique identifier for a resource (publication or dataset). Once an accession or dataset has a DOI, it can be referenced unambiguously. For EURISCO, linking an accession to a DOI means that even if websites or database URLs change, the DOI will still resolve to the correct information. This longevity is	• **Not a query mechanism:** A DOI itself isn't a way to search or retrieve data in a structured manner; it's an identifier to be resolved. To find *which* DOIs pertain to a given accession, one still needs a catalog or index. EURISCO must maintain the mapping (e.g., accession X ↔ DOI Y) by curating that info. If the DOI is not known, you can't discover it via algorithmic means easily (except via external search engines or if the DOI	Best for linking published studies or datasets that pertain to an accession, especially when those datasets are in external repositories or literature. Use DOIs when: you want to provide a stable link to an external resource without needing to ingest that resource's data.

Technical Requirements	Strengths	Considerations	Use Cases
	crucial for long-term data linking. • **Integration via reference-** DOIs allow a lightweight integration: EURISCO doesn't need an API call for every piece of data if a DOI link suffices. For example, embedding a dataset's DOI in EURISCO provides users a route to that dataset's landing page (which often contains rich information). This is particularly handy for **published datasets in literature or domain repositories** that might not have an easily queryable API but do have a DOI. • **Global interoperability-** Because DOI is an international standard, any DOI used in EURISCO can be resolved by users worldwide. It also means if other databases reference the same DOI (for instance, a journal article linking an accession's DOI), these references are talking about the exact same object, enabling cross-system interoperability through a common identifier. • **Encourages data publication-** By using DOIs, EURISCO aligns with practices of data publishing. Genebanks assigning DOIs to their accessions (through EURISCO's assistance or the ITPGRFA's GLIS system) are essentially publishing those as citable entities. Similarly, datasets with DOIs are considered	metadata is indexed elsewhere). • **Requires consistent metadata-** The usefulness of a DOI link depends on the dataset metadata. For instance, if a proteomics dataset DOI is linked to an accession, we assume the dataset indeed is about that accession. If metadata are poor (e.g., the accession isn't clearly mentioned in the dataset), the DOI link might not be obvious to establish. It often needs manual curation or at least coordination with data producers to ensure the connection is made. • **Multiple DOIs per accession-** One accession could be referenced by many dataset DOIs (multiple studies using the same accession). Conversely, one DOI (say for a large multi-accession study) might relate to dozens of EURISCO entries. Handling these many-to-many relationships in the interface could be challenging. EURISCO needs to display potentially multiple external links per accession in a user-friendly way. • **Resolution dependency-** While DOIs are persistent, their resolution depends on the DOI registry and target site. If, for example, a repository moves content without updating the DOI target, the link could break. However, this is uncommon if DOIs are managed properly. Also, some users may not immediately recognize a DOI link, so a user interface should clarify it (e.g., "Dataset DOI").	**Example**: if a metabolomics study in MetaboLights has DOI X and includes accession *ABC123*, EURISCO can list that DOI for *ABC123*. This alerts the user to external data and lets them click through for details. DOIs are also ideal for linking genebank accessions themselves: many genebanks now obtain DOIs for their accessions, which EURISCO can store (in the MCPD field for other identifiers). Those accession DOIs resolve to the FAO GLIS portal, providing a globally recognized identity for the material. In summary, DOIs are used in the federated model as anchors to external knowledge, (lightweight to store, heavy-weight in information content (when resolved)).

Technical Requirements	Strengths	Considerations	Use Cases
	published, creditable objects. This can incentivize data sharing and acknowledgement.		
BioSample ID (Sample Accession) (Gostev et al., 2012; Sayers et al., 2022) https://www.ncbi.nlm.nih.gov/biosample	• **Cross-database linking of omics data-** A BioSample ID is a unique accession for a biological sample in databases like EBI BioSamples or NCBI BioSample. The same BioSample ID can tie together sequence data, expression data, etc., across different archives. This is extremely powerful: by knowing one ID, you can fetch all related data from ENA, ArrayExpress, PRIDE (if integrated), etc. It essentially serves as a hub identifier for multi-omics on the same sample. • **Consistent metadata and traceability-** BioSample records store sample attributes (organism, tissue, origin, even an alias like an accession number). It enforces structured metadata for samples. When all omics submissions reference the BioSample, it ensures traceability. One can trace that all these datasets came from the exact same plant sample. For EURISCO, this means confidence that genomic and transcriptomic data being linked to an accession truly belong to that accession's plant material, not just a namesake. • **Query facilitation-** Many databases allow searching by BioSample ID. For	• **Potential for duplicates:** If different projects independently create BioSample entries for the same accession (not realizing an ID already exists), there could be multiple BioSample IDs for one accession. This complicates integration. EURISCO would need to track and possibly merge those identifiers. Ideally, coordination (like through DOI and careful metadata) avoids this, but it can happen if data producers don't check existing records. • **Not human-friendly-** BioSample IDs are alphanumeric codes with no intrinsic meaning (unlike an accession number or DOI which might encode institute info or be easier to communicate). For end-users, seeing "SAMEA123456" requires an extra step to interpret. EURISCO's interface should hide this complexity, using BioSample ID behind the scenes, or provide a translation (e.g., show it as "Sample ID" link for advanced users). • **Coverage not universal-** Some data types or repositories might not integrate with BioSample yet. For example, a proteomics dataset in PRIDE might not list a BioSample ID (they often list sample descriptions, but not necessarily link to BioSamples database). So BioSample ID primarily aids integration for sequence, genomic, transcriptomic data, but is less useful for data	Best for unifying multi-omics data around a single accession/sample. Use BioSample IDs when you have or expect multiple data sets from the same biological material and want to pull them together. **Example**: a genebank accession is sequenced (BioSample created) and also used in an RNA-seq experiment and maybe a variation study. If all those used the same BioSample ID, EURISCO can reliably fetch all relevant ENA entries, SNP variant files, and ArrayExpress experiments by querying that one ID. Another use case is linking back to phenotypic data: if phenotyping experiments (especially those in the future following the ELIXIR FONDUE approach) register BioSample IDs for the plant samples used, then even field observations could be connected to genomics via that ID. In summary, BioSample IDs act as the linchpin for data integration in the hybrid model, tying EURISCO's world of accessions to the bioinformatics world of sequence

Technical Requirements	Strengths	Considerations	Use Cases
	example, ENA's portal API can retrieve all runs or assemblies for a given BioSample accession in one query. Similarly, one could search ArrayExpress for a BioSample accession (if their data is synced with BioSamples). This means EURISCO can programmatically gather data from various sources simply by storing one key (the BioSample ID) per accession, instead of storing multiple query keys for multiple databases. • **Global uniqueness and stability-** BioSample IDs (like SAMCx... from EBI or SAMN... from NCBI) are stable accessions. They are part of international databases that synchronize (EBI and NCBI exchange BioSample info). This reduces ambiguity; even if different researchers or databases use slightly different names for an accession, if they use the same BioSample ID, we know it's the same entity.	types that haven't embraced it. In those cases, other identifiers (like DOIs or accession numbers) must be used.	data. Whenever available, EURISCO will use them to make cross-referencing data sources accurate and automated.

6.4 Integration Strategy for Each Data Type (or Repository)

6.4.1 Genomic Data (DNA Sequences and Variants)

Genomic data in this context ranges from raw DNA sequence reads and assembled genomes to variant calls (SNPs, etc.) associated with specific germplasm samples. Rather than storing large sequence files, EURISCO will link out to archival repositories while capturing essential metadata (Table 7). The integration leverages the global infrastructure of the International Nucleotide Sequence Database Collaboration (INSDC) (Karsch-Mizrachi et al., 2018) and related archives (Benson et al., 2013; Ogasawara et al., 2020; Yuan et al., 2023).

In practice, genomic integration works by having each EURISCO accession record store a reference (an ID) to external datasets. A useful method is to register each accession as a BioSample in databases such as ENA or the NCBI, using its DOI or a stable genebank ID. This single BioSample ID serves as a universal link. Any future sequencing or variant data for that accession will reference the same BioSample, ensuring that archives like ENA, EVA, and others automatically group all related datasets. EURISCO simply keeps the BioSample ID and can retrieve or display an integrated “omics” section when needed.

Table 7. Technical Details and Proposed Integration Strategy when Linking Genomic Data (DNA Sequences and Variants) with EURISCO

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
INSDC (Europe: ENA; USA: NCBI GenBank; Japan: DDBJ) https://www.insdc.org/ https://www.ebi.ac.uk/ena/browser/home https://www.ncbi.nlm.nih.gov/genbank/ https://www.ddbj.nig.ac.jp/index-e.html	*API-* Robust RESTful APIs and query tools (e.g. ENA provides a Portal API and Browser for programmatic search/download of sequences; NCBI offers Entrez APIs). *Data Formats-* Standard bioinformatics formats – FASTA for sequences, FASTQ for reads, BAM/CRAM for alignments, GFF for annotations, etc. Metadata in INSDC is in XML/JSON and includes sample information. *Protocols-* Data submission via Webin (ENA) or BankIt/Sequin (GenBank) with synchronized nightly exchange between INSDC partners. Each dataset gets accession IDs (e.g. sequence accessions, BioProject, BioSample IDs).	*Metadata Linkage* – EURISCO can store links and identifiers. For example, if an accession’s whole-genome sequence is submitted to ENA, EURISCO can record the ENA accession ID (run, study, BioProject) in the accession’s record. This requires that submitters indicate the origin of their samples (e.g. the genebank accession code or DOI) in the submission metadata; EURISCO can capitalize by searching INSDC for its accession identifiers. *API Integration* – Using ENA’s API, EURISCO can query by organism, strain/cultivar name, or DOI to find sequences related to a given accession. For instance, a BioSample in ENA might correspond to a genebank accession; if BioSample metadata include the accession number or DOI, EURISCO can automatically fetch and display all sequence projects for that accession. *Interoperability*- As INSDC is global and open, linking EURISCO to it ensures global accessibility of European germplasm genomic data. ENA/GenBank serves as the long-term archive, while EURISCO adds domain-specific context (linking sequence to genetic resources). We adhere to INSDC submission standards so that country of origin, taxonomy, and accession identifiers in sequence records align with EURISCO records. This way, anyone retrieving a

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
		sequence from ENA can trace it back to the EURISCO accession (and vice versa).
European Variation Archive (EVA) – EMBL-EBI archive for genetic variants (SNPs, indels, etc.) across all species https://www.ebi.ac.uk/eva/	*API-* RESTful endpoints and file transfer protocol (FTP) for bulk download of variant datasets; Data Formats: VCF for variant data; metadata in JSON. *Protocols-* Open submission of variant datasets from studies. Provides accessioned variant IDs and study IDs. EBI's EVA took over responsibility for non-human SNP accessioning from NCBI's dbSNP, ensuring a single authoritative source for plant variant data.	*Variant Data Integration* – For accessions that have genotyping or resequencing data, variant calls might be archived in EVA. EURISCO can store references to EVA Study IDs or Analysis IDs associated with an accession; EURISCO can link each relevant accession to that study. *Retrieval* – Through the EVA API, a user on EURISCO could be offered a direct link or query to fetch variants for a chosen accession. Since EVA indexes data by species and maybe population, direct per-accession queries might require that the submitter annotated the accession in the metadata. Encouraging submitters to mention EURISCO ID or DOI in variant study metadata would enable this linkage. *Metadata standards* – Integration requires aligning on reference genome identifiers and variant annotation standards. EURISCO would note the reference genome (perhaps via links to Ensembl Plants or NCBI Assembly) so that users know how to interpret the variant positions. All variant data remains in EVA (which guarantees open access and long-term preservation, while EURISCO serves as the catalog that tells breeders which accessions have variant data available.

6.4.2 Transcriptomic Data (Gene Expression)

Transcriptomic data includes gene expression measurements (from RNA-Seq, microarrays, etc.) often organized per experiment. Key repositories are those that archive gene expression data with Minimum Information About a Microarray Experiment (MIAME)/ Minimum Information About a Next-generation Sequencing Experiment (MINSEQE)-compliant metadata. The integration strategy is similar to genetic/ genomic data, which is to link rather than host raw data, but possibly index experiment metadata in EURISCO (Table 8). ArrayExpress (Europe) (Parkinson et al., 2006) and GEO (US) (Barrett et al., 2004; Barrett & Edgar, 2006; Barrett et al., 2006) are major repositories that provide open access to gene expression data. Many studies are available in both, as they regularly exchange data. By integrating with EURISCO, users searching for a specific genebank accession can easily discover relevant transcriptomic datasets through EURISCO's accession page. The actual storage and technical requirements (e.g., MIAME compliance) are handled by these repositories, while EURISCO focuses on linking the datasets to germplasm records, making it easier for researchers to find and interpret the data in the context of PGR.

Table 8. Technical Details and Proposed Integration Strategy when Interfacing Transcriptomic Data (Gene Expression Studies) with EURISCO

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
EMBL-EBI ArrayExpress https://www.ebi.ac.uk/biostudies/arrayexpress	*API-* Provides search and retrieval via REST (JSON or XML) and FTP for bulk downloads. Also accessible via R/Bioconductor (ArrayExpress package). *Data Formats-* MAGE-TAB format for metadata; raw data in FASTQ (for RNA-seq) or CEL files (for microarrays); processed data in expression matrices. *Protocols-* MIAME/MINSEQE guidelines enforced for submissions (detailed sample annotation, protocols, etc.). Raw sequence reads are automatically brokered to ENA for long-term storage, with ArrayExpress storing the experiment design and processed results.	Using ArrayExpress's API, EURISCO could periodically search by *cultivar/ accession name* or *BioSample ID* to find experiments. *Cross-Reference* – Once identified, a link to the ArrayExpress accession is stored in EURISCO under the relevant accession(s). The user can click through to ArrayExpress to see detailed gene expression results. Because ArrayExpress already links to ENA for raw data, EURISCO doesn't need to manage the files, only the references. *Interoperability* – If accession DOIs are cited in the ArrayExpress metadata, those could be used to automatically match EURISCO records. Encouraging researchers to use standard identifiers (DOIs or stable accession names) when depositing expression data will facilitate automated integration. EURISCO can also use standard ontologies for tissues and treatments to allow search across datasets (leveraging the Experimental Factor Ontology used in ArrayExpress).
NCBI Gene Expression Omnibus (GEO) https://www.ncbi.nlm.nih.gov/geo/	*API-* Entrez Programming Utilities (E-utilities) for search and download of GEO data (returns data in XML/JSON). Also web UI and GEOquery R package. *Data Formats-* SOFT and MINiML (XML) formats for metadata; raw data in text or FASTQ (for sequencing), processed data in matrix or supplementary tables. *Protocols-* Accepts submissions of array-based and high-throughput sequencing experiments; MIAME/MINSEQE compliance required GEO curates submissions into Series (experiments) and Samples, with each sample having metadata (attributes like genotype, treatment).	*Comprehensive Coverage* – GEO, being international, might contain datasets relevant to EURISCO accessions that are not in ArrayExpress (though many are shared). NCBI GEO profiles each sample with attributes, so a search by accession name or cultivar may find matches. EURISCO could use the GEO API to find such samples or use the GEOmetadb. *Integration Approach* – Similar to ArrayExpress: store the GEO Series ID and perhaps Sample IDs in the accession's record. Provide links for users to view those on GEO. Since GEO and ArrayExpress exchange data, many studies are mirrored; EURISCO can choose either as the link source. *Epigenomics Data* – GEO also archives epigenomic data (ChIP-Seq, methylation profiling). Thus, linking to GEO covers both transcriptomic and some epigenomic experiments for PGR.

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
		EURISCO will not differentiate the two in terms of linking mechanism – both are “experiment data”. But metadata tags (e.g. experiment type) can be stored so users know if it’s RNA-Seq vs. ChIP-Seq. *Standard IDs* – As with ArrayExpress, leveraging consistent naming is vital. If an accession’s name is ambiguous, DOIs or unique identifiers should be included in the GEO metadata. Cooperation with data submitters (via instructions to authors) can improve this. Over time, as DOIs become common for accessions, EURISCO could even query GEO for specific DOI references.

6.4.3 Proteomic Data (Protein Expression and Identification)

Proteomic data (e.g., seed proteome profiles, gel bands identification, etc.) are typically stored in specialized mass spectrometry repositories. Integration involves referencing these datasets when they pertain to specific accessions or samples. This data type is less common in genebanks but valuable for studying how genetic traits translate into actual protein functions. The biggest challenge is ensuring that each proteomics dataset clearly connects to the correct genebank accession. This can be solved by using DOIs or BioSample IDs in the dataset’s metadata (Table 9). One possible approach is to register genebank samples in the BioSamples database (EBI), allowing PRIDE (Perez-Riverol et. al., 2029; Perez-Riverol et. al., 2022) to link its datasets to those BioSample IDs. EURISCO can then use these IDs to cross-reference accessions with available proteomics data.

Table 9. Technical Details and Proposed Integration Strategy when Linking Proteomic Data with EURISCO

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
PRIDE (PRoteomics IDEntifications Database) (MS-based proteomics repository, part of ProteomeXchange). https://www.ebi.ac.uk/pride/	*API-* RESTful PRIDE Archive API for programmatic access to dataset metadata and files. Supports queries by species, sample details, etc. *Data Formats-* Raw instrument data (e.g. mass spectra in vendor formats or open mzML), peptide/protein identifications in mzIdentML or text, and metadata in PRIDE XML or JSON. Follows Proteome Standards (PSI) for format and annotations. *Protocols-* Data submission via ProteomeXchange pipeline – get a PX DOI and PRIDE accession.	*Cross-reference by Dataset ID* – When proteomic studies involve specific germplasm, EURISCO can store the ProteomeXchange ID (PXD number) or PRIDE ID for the study. This allows users to click out to PRIDE’s repository page to see proteins identified. *Metadata Matching* – PRIDE metadata includes species and often sample descriptions with cultivar or strain names. EURISCO can exploit the PRIDE API to search for datasets mentioning a particular accession name or ID in the sample description. Automating this will require that studies explicitly reference germplasm IDs. If a DOI was assigned to the plant sample (i.e., via GLIS), one ideal

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
	Datasets are validated and made public with extensive metadata (species, tissue, instrument, etc.). PRIDE is a ProteomeXchange founding member, meaning datasets are globally discoverable.	solution is for proteomics researchers to include that DOI in the PRIDE metadata. EURISCO could then query PRIDE for that DOI. *Standard Formats* – Direct data integration, such as copying protein identification data into EURISCO, will not be implemented. Instead, EURISCO will link accession records to external repositories by including unique identifiers that reference the corresponding datasets. These external repositories will provide data in widely accepted standard formats. Additionally, by leveraging the PRIDE API, EURISCO will be able to fetch up-to-date metadata (such as study titles, authors, etc.) to display alongside the accession record. ProteomeXchange network – PRIDE is mirrored by other repositories (e.g. MassIVE in the US). Because of the consortium, a PXD ID is valid globally. EURISCO thus needs only to store the PXD and not worry which node (PRIDE, MassIVE, etc.) holds the data. Users will be directed appropriately. This simplifies interoperability: one common identifier system for proteomics analogous to accession DOIs.

6.4.4 Metabolomic Data (Metabolite Profiles)

Metabolomic data includes NMR or mass spectrometry profiles of plant metabolites, and metabolite identifications. The integration of this data type relies heavily on detailed metadata, as factors like growth conditions, sample preparation, and instrument settings significantly affect results. In line with the approach for other data types, EURISCO will not duplicate these datasets but will instead link accession records to external repositories that manage them. EURISCO's strategy will focus on linking studies directly to accession records, ensuring that users can trace metabolomics findings back to the original plant material without duplicating all experimental metadata (Table 10). To date, the primary repository in Europe for this data is MetaboLights (Haug et. al., 2020; Yurekten et al., 2024).

Table 10. Technical Details and Proposed Integration Strategy when Interfacing Metabolomic Data with EURISCO

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
EBI MetaboLights – (cross-species). https://github.com/EBI-Metabolights	*API*- REST API for accessing studies. Also supports downloading ISA-Tab files and associated data.	*Study Metadata Linking* – Similar to expression data, EURISCO will link to metabolomics study IDs when relevant. For example, a metabolomic study on

Repository	Technical Details (API, Formats, Protocols)	Integration Strategy with EURISCO
https://www.ebi.ac.uk/metabolights	*Data Formats*- ISA-Tab / ISA-JSON for study metadata capturing sample info, protocols, and data files. Raw data can be mass spectra (mzML) or NMR spectra (nmrML), with processed data in tables. *Protocols*- Submission requires detailed metadata per ISA model (ensures info on samples, treatments, instrumentation). Each study gets an MTBLS accession ID and DOI. MetaboLights is endorsed by journals and ELIXIR as the recommended metabolomics repository.	tomato fruit from heirloom varieties (with those seeds in a genebank) would have a MetaboLights ID. EURISCO can list that under each relevant accession. The study title, metabolites analyzed, etc. could be fetched via the MetaboLights API to display context. *Sample Identification* – MetaboLights's ISA-Tab includes a Source Name and Sample Name which could be the germplasm identifier. If submitters use accession identifiers in those fields, EURISCO can automatically find matches. Close collaboration with metabolomics researchers (perhaps via ELIXIR bridging groups) can promote the inclusion of standard IDs in submissions. Alternatively, linking may occur manually or via curation. *Interoperability* – Because MetaboLights uses ISA-Tab, it can accommodate links to external resources. EURISCO can leverage this by ensuring any exported data to MetaboLights (if EURISCO ever feeds data) is MIAPPE/ISA compatible. Conversely, when pulling data, EURISCO can map ISA fields to its own. For instance, "Organism" and "Variety" fields in MetaboLights can map to species and accession name in EURISCO. No conversion of actual metabolite data is done. Users will go to MetaboLights for spectra or metabolite IDs. But EURISCO can store key summary info (like "X metabolites profiled; Y identified") in the accession's entry to entice users to explore further. *Global Repositories* – In addition to MetaboLights, the Metabolomics Workbench (USA) (https://www.metabolomicsworkbench.org/) could contain relevant studies. It has a similar concept to study IDs. EURISCO's approach would be identical: include a reference and link. Since MetaboLights and Workbench exchange data to some extent, focusing on MetaboLights (which is cross-technique and cross-species) covers most needs, with Workbench as a supplement for any missing studies.

6.4.5 Phenotypic Data

Phenotypic data, in contrast to the relative uniformity of genomic sequences, can vary widely in format, frequency and context. While some measurements follow traditional protocols, like basic morphological descriptors logged at regular intervals, others rely on cutting-edge, high-throughput systems that capture continuous time-series or high-resolution images using drones, cameras, and specialized phenotyping platforms. Moreover, environmental conditions, genotypic background, and management practices can drastically influence trait expression and data interpretation, which means phenotypic data require extensive metadata to keep the data meaningful and reproducible.

A substantial volume of these phenotypic data was generated through short-term research consortia or grant-funded initiatives. These projects typically assemble sizeable datasets, encompassing both traditional field observations and advanced, high-throughput sensor-based measurements. While these project-specific databases are well supported during the active phase of a project, the infrastructure and the staff responsible for curation and maintenance frequently lose funding once project objectives are met. Without sustainability plans or data migration strategies, the resulting "ephemeral" databases may remain stranded on temporary servers or cloud spaces, without guaranteed curation or backup. Over time, this precarious status can render the data increasingly difficult, if not impossible, to locate or access. In turn, this time-limited nature complicates efforts to federate phenotypic data. Although federation can, in principle, grant broader access to phenotypic datasets, directly connecting to ephemeral databases often proves unreliable. Once a short-term project concludes and its online resources are decommissioned, links within the federated network point to non-functional endpoints. Such broken links or "dead" references make aggregated searches less reliable, thereby discouraging in-depth analyses that rely on consistent data availability.

Institutionally managed repositories (Table 11) provide a more dependable alternative. These platforms maintain sustainable funding, professional data governance, and robust technical support, making them far less prone to abrupt discontinuation. Many of these stable repositories expose BrAPI endpoints, which define a RESTful interface for sharing germplasm, trial, study, and observational data. In practice, an aggregator like EURISCO can interface with these BrAPI-enabled repositories, obtaining records in a uniform format instead of making ad hoc connections to short-lived portals. Although coordinating BrAPI versions (e.g., v1.3 vs. v2.0) can introduce transitional hurdles, the standard's endorsement by ELIXIR underscores its growing acceptance within the plant science community.

In response to the challenges posed by ephemeral, project-based databases, a strategy has been discussed between EURISCO and certain consortia, such as AGENT, whereby phenotypic datasets from short-lived platforms will be migrated to EURISCO once the projects conclude. By aligning these datasets with EURISCO's evolving C&E standard and enhancing the portal's search and retrieval functions, this approach ensures that highly valuable phenotyping outputs remain systematically archived, discoverable, and accessible well beyond the lifespan of any single grant.

Table 11. Technical Details and Proposed Integration Strategy when Selected Phenotyping Information Systems & Future Interfaces for a One-Stop EURISCO

System	Host / Country	Data Focus	Key Technical Details	Standards Adoption (BrAPI, MIAPPE, MCPD, etc.)	Key Strengths and Limitations	Proposed Future Interfacing with EURISCO
__*GnpIS__	INRAE (France)	Multi-species *integrative information system (phenotyping, genotyping, QTL mapping etc)	- Modular architecture with multiple relational databases for different data types - REST APIs for genotype/phenotype/QTL - Query GnpIS BrAPI endpoints through FAIDARE - Utilizes Crop Ontology, Trait Ontology (TO), Plant Ontology (PO) for structural annotations	- BrAPI, early adopter provides endpoints for germplasm, phenotypes, trials (often aligns with v1.3+) - Aligns field trials with MIAPPE v1.1+ - Passport descriptors are mapped to MCPD fields wherever applicable - Some major datasets in GnpIS receive DOIs through institutional repositories; not all data is minted due to scale	- Broad Scope (covers genetics to phenotyping) -Strong support (institutional backing) from INRAE, ensuring stability - Ontology-based data curation for consistent trait definitions -Limitation: partial DOI coverage	- Full BrAPI Pipelines (Expand end-to-end BrAPI export to a future EURISCO phenotypic module) - Ontology Coordination (to unify trait definitions into a shared registry) - Systematic DOI linkouts (Assign DOIs to major datasets so EURISCO can reference detailed data in GnpIS repositories)
Germinate-based Repositories	Multiple Institutions Worldwide (James Hutton Institute origin)	Germplasm collections (passport data), phenotypes, genotypes (SNPs, SSRs) for various crops Crop-agnostic	- Open-source platform (MySQL/PostgreSQL) - Optional BrAPI modules (e.g., /germplasm, /observations) - Visualization for allele frequencies and trait distributions - Ontologies: typically Crop Ontology, can incorporate Trait Ontology (TO), Plant Ontology (PO), or custom trait dictionaries	- BrAPI, coverage depends on instance; many adopt v1.3 - MIAPPE adoption varies by local curation - Supports the core MCPD fields, but enforcement or completeness of MCPD descriptors varies - Some Germinate instances assign DOIs to large, published datasets, but no universal policy	- Flexible & modular (widely adopted for diverse crop communities) - Global Reach, large, active user base -Limitations: Decentralized, each instance maintains its own data standards; Quality Varies; MIAPPE/BrAPI compliance differs across installations	- Federated BrAPI layer, develop real-time data pulls from multiple Germinate repositories into EURISCO - Central MIAPPE validator: provide uniform pipelines to ensure standardized trait/environment metadata - Universal DOI policy: Encourage DOIs for major dataset releases, enabling deep links from EURISCO to each local Germinate instance

System	Host / Country	Data Focus	Key Technical Details	Standards Adoption (BrAPI, MIAPPE, MCPD, etc.)	Key Strengths and Limitations	Proposed Future Interfacing with EURISCO
WUR Tomato (BreeDB)	Wageningen University & Research (Netherlands)	Tomato breeding data, germplasm, pedigrees, morphological & disease-resistance traits	- Emphasizes pedigree management & breeding trials - Uses Sol Genomics trait sets or Crop Ontology for tomato - Partially implements BrAPI for phenotypes	- BrAPI, robust for germplasm/genotype; phenotypic endpoints still evolving - MIAPPE, gradual integration of greenhouse/field trial metadata	- Breeder-Focused (Pedigree tracking + robust trait management) - High data quality, standardized trait definitions for tomato - Limitations: Tomato-centric	- Ontology harmonization, Bridge Sol Genomics terms to a unified EU-wide trait reference - Expanded Phenotypic BrAPI, finalize endpoints enabling direct push of trial data to a future EURISCO phenotypic layer
PHIS	French Phenome-EMPHASIS (France)	High-throughput phenotyping (sensor data, imaging, time-series) for field & controlled environments	- Big-data frameworks (HPC, cloud) for large sensor/imaging - Detailed environment logs (time-series, sensor metadata) - Uses PECO, ENVO, plus Crop Ontology for trait definitions	- BrAPI, some phenotypic endpoints, but real-time data is challenging - MIAPPE, strong alignment for environment/protocol - Large, finalized experiments often minted with DOIs	- Big data expertise (Can handle massive sensor/imaging datasets) - Rich Metadata, thorough environment & protocol capture - Limitation: Partial BrAPI, real-time or near-real-time coverage still limited	- Publish minimal metadata in EURISCO, referencing a DOI or URI back to PHIS. -Partial BrAPI for key traits: PHIS can summarize time-series into discrete trait values (e.g., final plant height) and expose them via BrAPI endpoints, letting EURISCO retrieve only essential phenotypes. -SPARQL/REST Gateways, a specialized “gateway” can pull curated (aggregated) PHIS data at set intervals, generating standardized CSV or JSON outputs for EURISCO.

System	Host / Country	Data Focus	Key Technical Details	Standards Adoption (BrAPI, MIAPPE, MCPD, etc.)	Key Strengths and Limitations	Proposed Future Interfacing with EURISCO
e!DAL-PGP	IPK Gatersleben (Germany)	Data publication & archiving for plant genomics & phenomics (cross-domain, e.g., multi-omics, imaging, phenotypic tables)	All datasets are described with a minimum set of metadata and can be cited using DOI - Versioning & flexible metadata - Encourages, but does not enforce, recognized ontologies	- BrAPI, not inherently supported (bridging or adapters needed) - MIAPPE: Submissions can be MIAPPE-compliant, though not mandatory - every published dataset can be minted with a DOI	- Robust DOI Infrastructure: Systematic data citation and versioning - Scalable: Can handle large multi-omics/imaging sets - DOIs, versioning, and flexible metadata support - Limitation: No native BrAPI, must rely on external adapters for direct data exchange	- DOI Cross-Linking. Because e!DAL-PGP systematically assigns DOIs, EURISCO can reference these DOIs in a “Dataset” or “Experiment” field, directing users back to the archived data. - BrAPI Bridge: custom adapters must transform e!DAL-PGP metadata and data into BrAPI-compliant endpoints
PIPPA (PSB Interface for Plant Phenotype Analysis)	VIB-UGent (Belgium)	Image-based morphological and molecular phenotyping in controlled environments	- HPC-based imaging pipelines (e.g., rosette growth, root morphology) - Automated image segmentation and trait extraction - Often references Plant Ontology (PO) for structure-based phenotypes	- BrAPI, partial coverage for morphological trait outputs; expansions under development - Incorporates MIAPPE core fields for controlled-environment metadata	- Advanced Imaging, handles high-resolution morphological/architectural phenotyping - Scalable workflow, HPC backend supports large image data processing	- Full MIAPPE Coverage: Extend environment descriptors to align with broader EURISCO phenotyping scope - BrAPI Export of Image-Derived Traits: Standardize data models for direct integration - Ontology Alignment: Map PO-based morphological terms to a unified EURISCO trait registry, enabling cross-repository queries for image-derived phenotypes

6.4.5 On-farm Conserved Landrace Taxon-level Inventory Data

A coordinated strategy is being implemented to integrate on-farm conserved landrace taxon-level data into EURISCO. Rather than maintaining separate siloed inventories, this approach links this data type with EURISCO. A recently approved initiative led by the ECPGR On-Farm Conservation and Management Working Group, in collaboration with the Documentation and Information Working Group, is driving this effort. Notably, the "Annual & Biennial Crop Landraces Catalogue of Mediterranean Countries" project (approved in 2025) exemplifies this strategy by compiling on-farm landrace inventories for integration into EURISCO. This builds on parallel efforts to include CWR population records from pilot countries.

Technical aspects center on adopting or creating a widely accepted reference taxonomy and ensuring that national or regional synonyms map correctly to these standards. In practical terms, a shared protocol and data format will be developed to guide contributors in submitting taxon names and related metadata, minimizing errors and inconsistencies. Close coordination with NFP, who often maintain their own country-level lists, will be crucial to ensure high-quality inputs. A dedicated pipeline mechanism will be established to channel national-level taxon lists into EURISCO. Each country will designate or confirm a national contact point responsible for compiling and submitting its inventory using a pre-defined template. Submitted data will undergo a validation process and the system will also support versioning and updates, which recognizes that inventories may evolve over time with improved documentation or changes in conservation priorities. Modifications to EURISCO will be made to accommodate this new type of entry. While not linked to specific accessions, the taxon-level data will be clearly distinguishable within the database and searchable through dedicated filters. Importantly, these entries will provide visibility to under-documented resources and serve as reference points for identifying where and how landraces are conserved on-farm across Europe.

6.5 Points of Synergy with Pan-European Research Infrastructures and PGR-related Global Initiatives

The envisaged central PGR information hub must engage effectively with the broader ecosystem of plant and biodiversity research infrastructures to deliver a truly comprehensive, interoperable platform for PGR. By focusing on points of synergy (Figure 5), EURISCO can align its data practices, standards, and services with pan-European RIs such as ELIXIR, EMPHASIS, DISSCO and LifeWatch ERIC *(For a more comprehensive discussion of RIs, please refer to Deliverable 5.1)*, as well as with international initiatives such as GBIF, and Genesys. This approach ensures that each collaboration leverages the unique strengths and capabilities of these platforms, whether in bioinformatics, multi-scale phenotyping, global, biodiversity indexing and global PGRFA visibility, among others (Table 9). Ultimately, these points of synergy will allow EURISCO to expand its functionality, enrich its data offerings, and strengthen its role at the intersection of European PGR governance and advanced research across the life-science and biodiversity domains.

Figure 5. EURISCO's synergies with pan-European RIs and global PGRFA-related initiatives. *This illustrates EURISCO's core position within Europe's PGR landscape. Blue lines depict its points of synergy with four pan-European research infrastructures, namely ELIXIR, EMPHASIS, LifeWatch ERIC, and DiSSCo, to enable advanced PGR research and data integration across Europe. In parallel, orange lines show EURISCO's connections with international initiatives, GBIF, Genesys (Crop Trust), FAO WIEWS, and GLIS, which ensure global data sharing, monitoring, and policy coordination for PGR.*

Table 12. Points of Synergy between EURISCO & pan-European RIs

Research Infrastructure	**Domain**	**Points of synergy**
1. ELIXIR https://elixir-europe.org/	Genomics / Omics Integration FAIR Data & Federation	**Omics Repositories** (ENA, EVA, BioSamples) to link genotype or multi-omics data with EURISCO's accession-level passport records. Linking the EURISCO with ELIXIR's distributed hubs (e.g., EMBL-EBI data repositories, training resources, and computational tools) provides one-stop access for scientists seeking information on both PGR and associated molecular data. **BioSample IDs** **FAIR Standards**- ELIXIR's emphasis on data findability (DOIs), interoperability (BrAPI), use of ontologies (e.g., Crop Ontology) and reuse protocols that EURISCO can incorporate. **FAIDARE Portal**- A federated platform for searching plant data across multiple BrAPI endpoints, enabling direct discovery of EURISCO accessions in tandem with other genotype and phenotype sources.
2. EMPHASIS https://emphasis.plant-phenotyping.eu/	High-Throughput Trait Data (Phenotyping) Standardized Metadata	**Structural standard alignment: MIAPPE** **BrAPI** endpoints for phenotypic data exchange, referencing the same accession IDs or DOIs used by EURISCO. **Common ontologies; Synchronized descriptors**
3. DiSSCo https://www.dissco.eu/	Herbarium / Museum Linkage Taxonomic & Historical Context	**openDS / Digital specimen model** **Persistent Identifiers** (DOIs) that unify voucher specimens and genebank accessions, clarifying historical or taxonomic continuity. Museum taxonomic expertise for refining species names or synonyms in genebank records.
4. LifeWatch ERIC https://www.lifewatch.eu/	Ecosystem & E-Science Integration Data-Driven Environmental Modeling	**Advanced VREs** (Virtual Research Environments): HPC-based e-science tools for analyzing or simulating biodiversity and ecological scenarios (genebank/ *in situ* CWR data can be integrated) **Environmental layers & remote sensing** (overlays of climatic or habitat data (standardized metadata) e.g., on mapped *ex situ* accessions and *in situ* populations) Gap or Distribution Analyses that place PGRFA in a broader biodiversity context (*in situ* vs. *ex situ*).
5. *GBIF https://www.gbif.org/	Global Discovery & Biodiversity Context	**Darwin Core.** Publish genebank holdings as "occurrences" via the GBIF extension for germplasm, making them globally visible and analyzable alongside wild records. - **Aggregated biodiversity indexing**. Use GBIF's indexing system so that searching for a species in the global information system reveals both *in situ* observations and *ex situ* EURISCO accessions. **Citation & Dataset DOIs**. Ensuring open licensing and standardized referencing.

*Global

RECOMMENDATIONS

The effective conservation and use of PGR involves a complex network of stakeholders, each requiring reliable, up-to-date information. Although substantial progress has been made through the establishment of NIs and EURISCO as a central discovery information system, several persistent challenges continue to limit the visibility, completeness, and practical usefulness of germplasm-associated records. These challenges span from fragmented institutional coordination to uneven data coverage, inconsistencies in (meta)data, and a limited connection to emerging 'omics' resources. To address them, a multi-pronged strategy is needed: one that deepens intra-country collaboration, closes critical data gaps, expands EURISCO's role in ways that remain compatible with existing workflows, and promotes ongoing communication among the wider PGR community. A prospective research infrastructure dedicated to PGR, i.e., GRACE-RI, is well-positioned to facilitate these improvements by providing structured services, training, and technical support across multiple dimensions. The recommendations that follow outline how such efforts can be pursued in a realistic and incremental manner.

- **Strengthening Intracountry Coordination and Prioritization**

  A primary concern lies in inconsistent data flows within countries, where GRC, universities, breeding programs, community initiatives, and other data contributors often work with limited dialogue or shared objectives. By creating or refining a National PGR Coordination Mechanism, for instance, a formal committee, stakeholders can set national priorities (e.g., focusing first on unique or underrepresented accessions, or ensuring that all materials "available for distribution" are properly flagged). This mechanism would periodically assess the status of ex situ holdings, in situ CWR, and on-farm landraces to identify coverage gaps and manage the process of data updates.

  Role of GRACE-RI:
    - Facilitate best-practice sharing, by which GRACE-RI can host virtual forums or workshops where national focal points and genebank managers discuss how to integrate multiple data streams into a cohesive NI.
    - Through technical support providing standardized "blueprints" for inventory structures, software modules, and curated taxonomic backbones, GRACE-RI can accelerate adoption of consistent workflows across countries.
    - As part of its networking services, GRACE-RI can coordinate peer-mentoring, pairing advanced digital-infrastructure countries with those seeking to refine their NI approach.

- **Address Data Gaps and Ensure Metadata Completeness**

  Even when data are submitted to NI, inconsistencies in collection site details, taxonomic references, and pedigree or trait information can reduce the overall reliability of the record. A practical approach is to audit metadata fields regularly, identifying missing or ambiguous values and relaying clear error reports to data providers for correction. Adopting a recognized taxonomic backbone, possibly adapted from recognized global references, further helps unify synonymous species names and reduce confusion. Additionally, PUID such as DOI can greatly enhance traceability. Although EURISCO already offers a DOI registration service that updates metadata in GLIS, many genebank curators have not leveraged its simplicity. Targeted outreach (e.g., a short helpdesk or online tutorial) could encourage broader adoption, as could policy requirements (e.g., funders stipulating that all germplasm used in projects must be registered with a DOI). Over time, these measures close data gaps and streamline how accessions are tracked and referenced in external databases or publications.

  Role of GRACE-RI:
    - GRACE-RI can develop or host automated scripts and dashboards as (meta)data audit tools that detect anomalies or missing metadata in real time, generating clear reports for NFP.

  - GRACE-RI could operate a specialized helpdesk, e.g., DOI technical hub, to guide genebank staff through DOI registration. It might also integrate with EURISCO's existing DOI service, streamlining metadata updates in GLIS.
  - Through collaborations with European funders or the European Commission, GRACE-RI can champion policies that require or strongly encourage DOI assignment for germplasm used in research projects.

- **Enhance C&E Data Integration and Capture Project-Derived Phenotypic Data**
  To move beyond the "who/where" focus of passport data, C&E records should be incorporated more systematically into EURISCO. While it already permits C&E uploads, widespread usage remains limited due to inconsistent trait definitions and minimal experimental metadata. Building on the successes of EURISCO-EVA's standardized multi-site trials, a pragmatic approach might define a minimal yet flexible schema, borrowing elements from MIAPPE, that encourages data providers to share trait data in a consistent format without overly complicating submission processes. In parallel, there should be mechanisms to capture project-derived phenotypic data once breeding trials or research initiatives conclude, preventing valuable results from being lost or siloed. By expanding C&E data, EURISCO can better serve breeders and researchers seeking specific agronomic or environmental adaptations.

  Role of GRACE-RI:
  - GRACE-RI can host an online repository of preferred trait descriptors (controlled vocabularies and ontologies) and data-sharing templates. This resource would draw on successful models like EURISCO-EVA, guiding genebanks toward uniform trait definitions.
  - Through specialized training sessions, GRACE-RI can facilitate capacity development among genebank curators, breeding programs, and project leads how to structure phenotypic data for straightforward integration into EURISCO and ensure that knowledge persists beyond the lifecycle of specific grants or trials.
  - Develop user-friendly pipelines for project-derived phenotypic data.
  - GRACE-RI might manage a centralized "C&E clearinghouse," where large multi-location trials or breeding initiatives deposit data, facilitating an eventual push to EURISCO once minimal metadata requirements are met.

- **Expand EURISCO's Scope Without Undermining Its Discovery Role**
  EURISCO's hallmark is its role as a discovery-level platform, helping users quickly identify accessions and find the relevant custodians. Rather than transforming it into a detailed management system, future expansions can remain streamlined. One possibility is to integrate a concise, taxon-level inventory of in situ crop wild relatives or on-farm landraces, listing essential descriptors (e.g., taxon, approximate area, responsible institution) but leaving dynamic stewardship data, to national or local databases.

  Role of GRACE-RI:
    - GRACE-RI can provide a standardized framework, i.e., inventory module support, for how in situ references are formatted and periodically synced with EURISCO.
    - GRACE-RI could offer GIS-based mapping services or lightweight mapping tools that indicate broad in situ distributions, without jeopardizing sensitive location data or requiring constant updates of dynamic fields.

- **Link PGR Data with Omics and Other Advanced Research Tools**
  As genomic, and other "omics" methods advance, interconnections between passport data and molecular information become increasingly valuable. While EURISCO is not intended to store raw sequence data, pilot linkages could demonstrate how certain accession IDs map to external platforms hosting SNP information or reference genome assemblies. By showcasing a

few targeted examples, stakeholders would see the practical benefits for breeding and research. If these pilots succeed, expansions to other systems or more elaborate genotype–phenotype correlations could follow, encouraging consistent use of accession identifiers across multiple repositories.

- **Emphasize User Consultation, Training, and Community Capacity Building**
  Many of the above recommendations hinge on stakeholder involvement, from NFP to local communities and from data providers to end-users in breeding and research. Ensuring regular feedback loops can take multiple forms: workshops, surveys, and interactive forums that capture real-world needs (e.g., advanced search features, new crop coverage, simplified data templates). Training events modelled on previous EURISCO sessions can boost data provider confidence in adopting tools like DOIs or standard trait definitions. Additionally, a community of practice, under GRACE-RI, would allow data managers to exchange problem-solving techniques (e.g., handling outmoded synonyms or mapping legacy fields to updated descriptors), bolstering overall data quality and the speed of integration into EURISCO.

  Role of GRACE-RI:
    - GRACE-RI can maintain an active virtual space for real-time Q&A, best-practice sharing, and tutorials (online forum and regular webinars)
    - Focus on diverse user groups: By including community seed banks, private breeders, and national focal points, the community ensures knowledge dissemination across different levels of capacity and infrastructure.

Taken together, these improvements align directly with the European PGR Strategy 2030, which calls for robust and integrated information platforms. By strengthening intranational coordination, filling data-related gaps, and enhancing EURISCO’s usability and reach, the PGR community can create a more comprehensive, user-centric, and future-proof system. Researchers will find higher-quality metadata and expanded trait data; genebank managers will streamline their curation tasks with persistent identifiers; and the entire network of stakeholders, including policymakers, local communities, private breeders, will benefit from a clearer, more cohesive view of Europe’s plant genetic heritage. GRACE-RI’s role in offering shared infrastructure such as tools for automated audits, DOI registration assistance, trait ontology development, data-management training, and pilot interconnections with omics repositories, would make these efforts feasible at scale. In this manner, EURISCO remains not only a gateway for discovery but also a foundational pillar for sustainable PGR conservation and use in the decades to come.

REFERENCES

- Alercia, A., Diulgheroff, S., & Mackay, M. (2015) FAO/Bioversity Multi-crop Passport Descriptors V. 2.1 [MCPD V. 2.1]-December 2015. Bioversity International. [WWW document] URL https://cgspace.cgiar.org/handle/10568/69166
- Alercia, A., López, F., Marsella, M., and Cerutti, A.L. (2022) *Descriptors for Crop Wild Relatives conserved in situ (CWRI v.1.1) Revised version*. Rome, FAO on behalf of the International Treaty on Plant Genetic Resources for Food and Agriculture. Available at: https://doi.org/10.4060/cb3256en
- Anglin, N. L., Amri, A., Kehel, Z., & Ellis, D. (2018). A Case of Need: Linking Traits to Genebank Accessions. Biopreservation and biobanking, 16(5), 337–349. https://doi.org/10.1089/bio.2018.0033
- Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash A.E., Fujibuchi, W., & Edgar, R. (2005). NCBI GEO: mining millions of expression profiles--database and tools. Nucleic Acids Res. 33(Database issue):D562-6. https://doi.org/10.1093/nar/gki022.
- Barrett T, Edgar R. (2006) Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*. Methods Mol Biol. 338:175-90. https://doi.org/10.1385/1-59745-097-9:175.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res. 35(Database issue):D760-5. https://doi.org/10.1093/nar/gkl887. Epub 2006 Nov 11.
- Bellon MR, Dulloo E, Sardos J, Thormann I, Burdon JJ. (2017) *In situ* conservation-harnessing natural and human-derived evolutionary forces to ensure future crop adaptation. Evol Appl. 2017 Sep 6;10(10):965-977. https://doi.org/10.1111/eva.12521.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. (2013) GenBank. Nucleic Acids Res. 41(Database issue):D36-42. https://doi.org/10.1093/nar/gks1195. Epub 2012 Nov 27.
- Bioversity and The Christensen Fund, (2009) *Descriptors for farmers' knowledge of plants.* Bioversity International, Rome, Italy and The Christensen Fund, Palo Alto, California, USA.
- Cámara Ballesteros, A., Aguayo Jara, E., Verykaki, E. S., Pastor Del Olmo, G., Moreno Vázquez, S., Torres, E., & Wilkinson, M. D. (2024). The FLAIR-GG federated network of FAIR germplasm data resources. Scientific data, 11(1), 1386. https://doi.org/10.1038/s41597-024-04243-7
- Caproni L., Raggi L., Negri V. (2020) *In situ landrace propagation management and access guidelines. Farmer's Pride: Networking, partnerships and tools to enhance in situ conservation of European plant genetic resources*. Farmer's Pride project deliverable. University of Birmingham, Birmingham, UK. Available at: https://more.bham.ac.uk/farmerspride/wp-content/uploads/sites/19/2020/09/D2.4_In_situ_landrace_propagation_management_guidelines.pdf (Accessed 30.01.2025).
- Ćwiek-Kupczyńska, H., Altmann, T., Arend, D. et al., (2016) Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods 12, 44. https://doi.org/10.1186/s13007-016-0144-4.
- de Boef W.S., Dempewolf, H., Byakweli, J.M., & Engels, J.M.M. (2010) Integrating genetic resource conservation and sustainable development into strategies to increase the robustness of seed systems. J Sust Agric 34:504–531. ttps://doi.org/10.1080/10440046.2010.484689.
- Deng CH, Naithani S, Kumari S, Cobo-Simón I, Quezada-Rodríguez EH, Skrabisova M, Gladman N, Correll MJ, Sikiru AB, Afuwape OO, Marrano A, Rebollo I, Zhang W, Jung S. (2023) Genotype and phenotype data standardization, utilization and integration in the big data era for agricultural sciences. Database (Oxford). https://doi.org/10.1093/database/baad088.

- De Pau, L., Fernandes de Oliveira, A., Frau, A.F., Rigoldi, M.P., Di Salvo, R.; Scanu, G.; Satta, D. (2024) Biodiversity of Sweet Cherry in Sardinia. Diversity, 16, 767. https://doi.org/10.3390/d16120767
- Díez MJ, De la Rosa L, Martín I, Guasch L, Cartea ME, Mallor C, Casals J, Simó J, Rivera A, Anastasio G, Prohens J, Soler S, Blanca J, Valcárcel JV and Casañas F (2018) Plant Genebanks: Present Situation and Proposals for Their Improvement. the Case of the Spanish Network. Front. Plant Sci. 9:1794. https://doi.org/10.3389/fpls.2018.01794
- Ebert, A.W,, Engels, J.M.M., Schafleitner, R., van Hintum, T., & Mwila, G. (2023) Critical Review of the Increasing Complexity of Access and Benefit-Sharing Policies of Genetic Resources for Genebank Curators and Plant Breeders–A Public and Private Sector Perspective. Plants. 12(16):2992. https://doi.org/10.3390/plants12162992
- Engels, J.M., & Visser, L. (2003). A Guide to Effective Management of Germplasm Collections; International Plant Genetic Resources Institute: Rome, Italy.
- Engels, J.M.M., & Thormann, I. (2020). Main Challenges and Actions Needed to Improve Conservation and Sustainable Use of Our Crop Wild Relatives. Plants (Basel). 9(8):968. https://doi.org/10.3390/plants9080968.
- FAO (2022) WIEWS: *Ex situ* (SDG 2.5.1) – Overview. Rome. Accessed 15 October 2024. Available at https://www.fao.org/wiews/data/ex-situ-sdg-251/overvew/en
- FAO (2025) The Third Report on The State of the World's Plant Genetic Resources for Food and Agriculture. FAO Commission on Genetic Resources for Food and Agriculture Assessments, 2025. Rome. https://doi.org/10.4060/cd4711en.
- FAO WIEWS Accessed May 2023- December 2024. Available at https://www.fao.org/wiews/data/domains/monitoring-framework/en/
- Halewood M, Chiurugwi T, Sackville Hamilton R, Kurtz B, Marden E, Welch E, Michiels F, Mozafari J, Sabran M, Patron N, Kersey P, Bastow R, Dorius S, Dias S, McCouch S, & Powell W. (2018) Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. New Phytol. 217(4):1407-1419. https://doi.org/10.1111/nph.14993. Epub 2018 Jan 23.
- Haug K, Cochrane K, Nainala VC, Williams M, Chang J, Jayaseelan KV, O'Donovan C. (2020). MetaboLights: a resource evolving in response to the needs of its scientific community. Nucleic Acids Res. 48(D1):D440-D444. https://doi.org/10.1093/nar/gkz1019.
- Hintum, T. V., Engels, J. M. M., & Maggioni, L. (2021). AEGIS, the Virtual European Genebank: Why It Is Such a Good Idea, Why It Is Not Working and How It Could Be Improved. Plants (Basel, Switzerland), 10(10), 2165. https://doi.org/10.3390/plants10102165
- Iriondo, J.M., Magos Brehm, J., Dulloo, M.E., Maxted, N. (eds.) (2021). Crop Wild Relative Population Management Guidelines. Farmer's Pride: Networking, partnerships and tools to enhance *in situ* conservation of European plant genetic resources. ECPGR, Rome, Italy. Available at University of Birmingham, Birmingham, UK. https://more.bham.ac.uk/farmerspride/key-documents/crop-wild-relatives/
- Iriondo, J.M., Maxted, N., Kell, S.P., Ford-Lloyd, B.V., Lara-Romero, C., Labokas, J. and Magos Brehm, J. (2012) Quality standards for genetic reserve conservation of crop wild relatives. In: Maxted, N., Dulloo, M.E., Ford-Lloyd, B.V., Frese, L., Iriondo, J.M. and Pinheiro de Carvalho, M.A.A. (eds.) *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces*. Pp. 72-77. CAB International, Wallingford.
- Jarvis, D. I., Hodgkin, T., Sthapit, B. R., Fadda, C., & Lopez-Noriega, I. (2011). An Heuristic Framework for Identifying Multiple Ways of Supporting the Conservation and Use of Traditional Crop Varieties within the Agricultural Production System. Critical Reviews in Plant Sciences, 30(1–2), 125–176. https://doi.org/10.1080/07352689.2011.554358
- Karsch-Mizrachi I, Takagi T, Cochrane G. (2018). International Nucleotide Sequence Database Collaboration. The international nucleotide sequence database collaboration. Nucleic Acids

Res. 46(D1):D48-D51. https://doi.org/10.1093/nar/gkx1097. PMID: 29190397; PMCID: PMC5753279.

- Kell, S.P., Knüpfer, H., Jury, S.L., Ford-Lloyd, B.V. & Maxted, N., (2008). Crops and wild relatives of the Euro-Mediterranean region: making and using a conservation catalogue. In: Maxted, N., Ford-Lloyd, B.V., Kell, S.P., Iriondo, J., Dulloo, E. & Turok, J. (eds.) Crop Wild Relative Conservation and Use. Pp. 69-109. CAB International, Wallingford.
-
- Krishna, S.M.H., & Sharma, R. (2021). Survey on application programming interfaces in software defined networks and network function virtualization. Global Transitions Proceedings 2(2): 199-204. https://doi.org/10.1016/j.gltp.2021.08.018.
- Kotni P., van Hintum T., Maggioni L., Oppermann M., and Weise S. (2022). EURISCO update 2023: the European Search Catalogue for Plant Genetic Resources, a pillar for documentation of genebank material. Nucleic Acids Research, 51(D1):D1465-D1469, 2023. https://doi.org/10.1093/nar/gkac852
- Kreide S., Oppermann M., and Weise S. (2019). Advancement of taxonomic searches in the European search catalogue for plant genetic resources. Plant Genetic Resources, 17(6):559-561. https://doi.org/10.1017/S1479262119000339
- Kumar, S., Guzzon, F., Goritschnig, S. and 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), pp. 117–125. https://doi.org/10.46265/genresj.IHXU5248
- Lusty C, Sackville Hamilton R, Guarino L, Richards C, Jamora N, Hawtin G. (2021). Envisaging an Effective Global Long-Term Agrobiodiversity Conservation System That Promotes and Facilitates Use. Plants. 10(12):2764. https://doi.org/10.3390/plants10122764

- Magos Brehm, J., Kell, S.P., Thormann, I., Gaisberger, H., Dulloo, E. and Maxted, N. (2017a) *Occurrence data collation template v.1*, Harvard Dataverse, V1. Available at: https://doi.org/10.7910/DVN/5B9IV5
- Magos Brehm, J., Kell, S.P., Thormann, I., Maxted, N. and Dulloo, E. (2017b). *Template for the Preparation of a Technical Background Document for a National Strategic Action Plan for the Conservation and Sustainable Use of Crop Wild Relatives*. Harvard Dataverse, V2. Available at: https://doi.org/10.7910/DVN/VQVDFA

- Martin, P., Shoemark, O., Scholten, M., Wishart, J., Drucker, A.D., Maxted, N.**,** (2023). Trends, challenges and opportunities in the *in situ* conservation of cereal landraces in Scottish islands. *Genetic resources*, 4(7): 32-45. https://doi.org/10.46265/genresj.QGSB7051/
- Maxted, N., Adam-Blondon, A.-F., Aguilar, C.H., Barata, A.M., Bartha, B., Bocci, R., De Paola, D., Fitzgerald, H., Fresta, L., Fusani, P., Giuliano, G., Guzzon, F., Holzherr, P., Holubec, V., Iriondo, J.M., Labokas, J., Maggioni, L., Magos Brehm, J., Palmé A.16, Phillips, J., Prohens, J., Raggi, L., Ralli, P., Ruņģis, D.E.., Sarikyan, K., Šuštar Vozlič, J., Thormann, I. and Zdunić, G., (2025). . Genetic Resources, In Press.

- Maxted N., Amri. A., Castañeda-Álvarez, N.P., Dias, S., Dulloo, M.E., Fielder, H., Ford-Lloyd, B.V., Iriondo, J.M., Magos Brehm, J., Nilsen, L-B., Thormann, I., Vincent, H. and Kell, S.P. (2016) Joining up the dots: a systematic perspective of crop wild relative conservation and use. In: Maxted, N., Ehsan Dulloo, M. and Ford-Lloyd, B.V. (eds.), *Enhancing Crop Genepool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement*. Pp. 87-124. CAB International, Wallingford, UK.
- Moore, J.D., Kell, S.P., Iriondo, J.M., Ford-Lloyd, B.V., Maxted, N. (2008) CWRML: representing crop wild relative conservation and use data in XML. *BMC Bioinformatics*, 9:116. Available at: https://doi.org/10.1186/1471-2105-9-116
- Negri, V., Maxted, N., Torricelli, R., Heinonen, M., Veteläinen, M. and Dias, S., (2012). *Descriptors for web-enabled national in situ landrace inventories*. PGR Secure project deliverable. University of Birmingham, Birmingham, UK. Available at: http://vnr.unipg.it/PGRSecure/data/LRDESCRIPTORS_PGRSECURE.pdf (Accessed 26.01.2025)

- Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R, Rayner T, Sharma A, William E, Sarkans U, Brazma A. (2007). ArrayExpress--a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 35(Database issue):D747-50. https://doi.org/10.1093/nar/gkl995. Epub 2006 Nov 28.
- Pathirana, R., & Carimi F. (2022) Management and Utilization of Plant Genetic Resources for a Sustainable Agriculture. Plants. 11(15):2038. https://doi.org/10.3390/plants11152038
- Paskin N. (2005). Digital Object Identifiers for Scientific Data. Data Science Journal 4: 12-20. https://doi.org/10.2481/dsj.4.12
- Perez-Riverol, Y., Bai, M., da Veiga Leprevost, F., Squizzato, S., Park, Y. M., Haug, K., Carroll, A. J., Spalding, D., et al., (2017). Discovering and linking public omics data sets using the Omics Discovery Index. Nature biotechnology, 35(5), 406–409. https://doi.org/10.1038/nbt.3790
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., … Vizcaíno, J. A. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic acids research, 47(D1), D442–D450. https://doi.org/10.1093/nar/gky1106
- Perez-Riverol Y., Bai J., Bandla C., García-Seisdedos D., et al., (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, Nucleic Acids Research, Volume 50, Issue D1, Pages D543–D552, https://doi.org/10.1093/nar/gkab1038
- Phillips, J., Maxted, N., Magos Brehm, J., Adam-Blondon, A.-F., Avagyan, A., Fresta, L.J., Forycka, A., Gailïte, A., Gay, L., Iriondo, J., Labokas, J., Palmé, A., Raggi, L., Thormann, I., Weise, S., Zdunic, G. (2025) D2.3 Annex 1 Plant Genetic Resources *in situ* descriptors v1. Deliverable 2.3 EC funded Pro GRACE Project. Agenzia Nazionale per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile, Roma, Italy.
- Pommier, C. Coppens, F., Ćwiek-Kupczyńska; H. et al., (2023). Plant Science Data Integration, from Building Community Standards to Defining a Consistent Data Lifecycle. In: Williamson, H.F., Leonelli, S. (eds) Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development. Springer, Cham. Pp. 149-160. https://doi.org/10.1007/978-3-031-13276-6_8
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., et al., (2022). Database resources of the national center for biotechnology information. Nucleic acids research, 50(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112
- Selby P, Abbeloos R, Backlund JE, Basterrechea Salido M, Bauchet G, Benites-Alfaro OE, Birkett C, Calaminos VC, et al. (2019) BrAPI-an application programming interface for plant breeding applications. Bioinformatics. 35(20):4147-4155. https://doi.org/10.1093/bioinformatics/btz190.
- Shrestha, R., Matteis, L., Skofic, M., Portugal, A., et al. (2012). Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. Front. Physio.3, 326, https://doi.org/10.3389/fphys.2012.00326.
- Sohan, S.M., Anslow, C., & Maurer, F. (2015) A Case Study of Web API Evolution, 2015 IEEE World Congress on Services, New York, NY, USA, pp. 245-252, https://doi.org/10.1109/SERVICES.2015.43.
- Thormann, I., Parra-Quijano, M., Endresen, D.T.F., Rubio-Teso, M.L., Iriondo, M.J. & Maxted, N. (2014). Predictive characterization of crop wild relatives and landraces. Technical guidelines version 1. Bioversity International, Rome, Italy. Available online at: http://www.bioversityinternational.org/index.php?id=244&tx_news_pi1%5Bnews%5D=4967&cHash=7cd3c6c2b8360927b83fa6ef7cc28d99

- Thormann, I., Alercia, A., Dulloo, M.E. (2013).*Core descriptors for in situ conservation of crop wild relatives v.1*. Bioversity International, Rome, Italy.
- Thormann, I., Kell, S., Magos Brehm, J., Dulloo, E. and Maxted, N. (2017) *CWR checklist and inventory data template v.1*, Harvard Dataverse, V4. Available at: https://doi.org/10.7910/DVN/B8YOQL
- Ulian, T., Diazgranados, M., Pironon, S., Padulosi, S., et al., (2020). Unlocking plant resources to support food security and promote sustainable agriculture. Plants, People, Planet, 2(5), 421–445. https://doi.org/10.1002/ppp3.10145
- van Hintum, T. and Iriondo, J.M., (2022). Principles for the Inclusion of CWR Data in EURISCO. European Cooperative Programme for Plant Genetic Resources (ECPGR), Rome, Italy. Available online at: https://www.ecpgr.org/resources/ecpgr-publications/publication/principles-for-the-inclusion-of-cwr-data-in-eurisco-2022 (accessed 06.11.2024).
- Veteläinen, M., Negri, V., Maxted, N. (eds.), (2009). *European Landraces: On-farm conservation, Management and Use.* Bioversity Technical Bulletin 15. Pp. 1-359. Bioversity International, Rome, Italy. ISBN: 978-92-9043-805-2.
- Weise, S., Oppermann M., Maggioni L., van Hintum T., and Knüpffer H. (2017). EURISCO: The European Search Catalogue for Plant Genetic Resources. Nucleic Acids Research, 45(D1):D1003-1008. https://doi.org/10.1093/nar/gkw755 Weise, S., Kreide, S., and Maxted, N. (2020) *Concept for a possible extension of EURISCO for in situ crop wild relative and on-farm landrace data*. Farmer's Pride: Networking, partnerships and tools to enhance *in situ* conservation of European plant genetic resources. University of Birmingham, Birmingham, UK. Available at: https://more.bham.ac.uk/farmerspride/wp-content/uploads/sites/19/2021/09/D2.5_EURISCO_in_situ_extension_concept.pdf (Accessed 26.01.2025).
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18
- Yuan, D., Ahamed, A., Burgin, J., Cummins, C., Devraj, R., Gueye, K., Gupta, D., Gupta, V., Haseeb, M., Ihsan, M., Ivanov, E., Jayathilaka, S., Kadhirvelu, V. B., Kumar, M., Lathi, A., Leinonen, R., McKinnon, J., Meszaros, L., O'Cathail, C., Ouma, D., … Cochrane, G. (2024). The European Nucleotide Archive in 2023. Nucleic acids research, 52(D1), D92–D97. https://doi.org/10.1093/nar/gkad1067
- Yurekten O., Payne T., Tejera N., Amaladoss F.X., Martin C., Williams M., O'Donovan C., (2024) MetaboLights: open data repository for metabolomics, Nucleic Acids Research, Volume 52, Issue D1, Pages D640–D646, https://doi.org/10.1093/nar/gkad1045
- Zeven, A. (1998) Landraces: A review of definitions and classifications. Euphytica 104, 127–139 https://doi.org/10.1023/A:1018683119237