PRO-GRACE

PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

DELIVERABLE 4.4

Interconnection of the different phenotype databases with the central EURISCO information system

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

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

Promoting a plant genetic resource community for Europe

Deliverable No. 4.4

Interconnection of the different phenotype databases with the central EURISCO information system

Contractual delivery date:
M33

Actual delivery date:
M33

Responsible partner: INRAE

Contributing partners: CREA, IPK, IPGRI, 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	**D4.4**
Deliverable title	**Interconnection of the different phenotype databases with the central EURISCO information system**
Type	DEC
Dissemination level	PU
Work package number	4
Author(s)	Kritika Adhikari, Véronique Lefebvre, Ignazio Verde, Michael Alaux, Cyril Pommier, Lorenzo Maggioni, Giuseppe Aprea
Keywords	Plant Genetic Resources, FAIR, Interoperability, Interconnection, Phenotype

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.

Abbreviations

AI	Artificial Intelligence
API	Application Programme Interface
AWS	Amazon Web Services
BrAPI	Breeding Application Programme Interface
CLARISA	CGIAR Level Agricultural Results Interoperable System Architecture
CGIAR	Consortium of International Agricultural Research Centers
CO	Crop Ontology
CSV	Comma Separated Values
CWR	Crop wild Relatives
DCAT	Data Catalog Vocabulary
DOI	Digital Object Identifier
ECPGR	European Cooperative Programme for Plant Genetic Resources
ELT	Extract-load-transform
EMBL-EBI	European Bioinformatic Institute
EMPHASIS	European Infrastructure for Multi-Scale Plant Phenotyping and Simulation for Food Security in a Changing Climate
ETL	Extract-Transform-Load
EURISCO	European Search Catalogue for Plant Genetic Resources
EVA	European Evaluation Network
FAIR	Findable, Accessible, Interoperable and Reusable
FAO-WIEWS	World Information and Early Warning System
FDP	Fair Data Point
FAIDARE	FAIR Data-finder for Agronomic Research
FTP	File Transfer Protocol
GA4GH	Global Alliance for Genomics and Health
GBIF	Global Biodiversity Information Facility
GCP	Google Cloud Platform
GnPIS	Genoplante Information System
GraphQL	Graph Query Language
GRIN	Germplasm Resources Information Network
GWAS	Genome-wide Association Studies
HTP	High Throughput Phenotyping
HTTP	Hypertext Transfer protocol
IBM	International business Machines
INRAE	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
IPK	Leibniz Institute of Plant Genetics and Crop Plant Research
ISA-TAB	Investigation-Study-Assay Tab Delimited
JPEG	Joint Photographic Experts Group
JSON	JavaScript Object Notation
MaizeGDB	Maize genetics and Genomics Database
MARLO	Managing Agricultural Research for Learning and Outcomes

MCPD	Multi-crop Passport Descriptors
MEL	Monitoring Evaluation and Learning
MIAPPE	Minimum Information About Plant Phenotyping Experiments
NCBI	National Center for Biotechnology Information
NFP	National Focal Points
NI	National Inventories
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OWL	Web Ontology Language
PGR	Plant Genetic Resources
PGRFA	Plant Genetic Resources for Food and Agriculture
PNG	Portable Network Graphics
PUID	Persistent Unique Identifier
QTL	Quantitative Trait Loci
RDF	Resource Description Framework
REST	Representational State Transfer
SOAP	Simple Object Access Protocol
SQL	Simple Query Language
SRA	Sequence Read Archive
TIFF	Tag Image File Format
UAV	Unmanned Aerial Vehicle
URI	Uniform Resource Identifier
USDA-ARS	United States Department of Agriculture- Agriculture Research Service
WGS	Whole Genome Sequencing
XML	Extensible Markup Language

Terminologies

Data providers	Institute or organizations that curate data.
Data source	Point of origin of data.
Digital asset	Anything that exists in digital form and provides information with distinct usage rights or distinct permission for use.
External sources/system	Data providers that exist outside a central or core infrastructure.
Fault tolerance	Ability of a system/architecture to properly function even when one or more components fail.
Governance	Framework of policy, rule and agreements to ensure effective management, security and ethical management of a body that could be an organization or network of organizations.
Harmonized data	Data that is consistent across sources in terms of structure, format, terminologies and meanings.
Integration	Process of combining data from multiple sources into a unified system to facilitate coherent access, harmonization and analysis.
Interconnection	Process of establishing communication links between different systems or components to enable exchange of data.
Interoperability	The ability of different systems, applications or organizations to exchange, interpret and use data effectively. It is the system's ability to interconnect, integrate and interpret data from different sources usually in an automated manner.
Linked Data	Structured data published using standard web technologies (URIs and RDF) interlinked with other data enabling machine readable relationships and semantic querying across datasets.
Metadata	Data summary describing the property of the data it represents i.e. attributes, origin, explanation etc.
Node(s)	An individual component or an endpoint in a networked system that produce, process or transmit data.
Query	A form of questioning or request for information retrieval.
Schema	A rule or standard of sorting or structuring data to make it 'queryable'.
Ontology	An extended controlled vocabulary. Usually, a list of descriptors or terms agreed within a given community. The vocabularies terms have names, description, synonyms and semantic links among them allowing hierarchical or graph organisation.

Executive summary

Plant genetic resources (PGR) form the foundation of crop improvement, agriculture resilience and global food security. They represent the diversity that can accelerate crop innovation and sustainability. With the technologic advances on phenotype intricate characteristics and high-resolution genetic maps, our ability to characterize these resources have tremendously expanded, creating a great opportunity to transform genetic diversity to knowledge and innovations. However, for this potential of genetic resources to be fully realized, information generated from PGR should be effectively organized, findable and accessible.

In Europe, the digitalization and access of data in PGR have progressed significantly in terms of genotypic information through centralized repositories such as European Nucleotide Archive (ENA). However, phenotypic data remains fragmented, unrepresented or have limited accessibility. The issue is not a lack of data, but the absence of a coherent framework that enables different actors (gene banks, research infrastructures, breeders) and other stakeholders to share and exchange information in a standardized, interoperable, and **FAIR** (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) compliant manner. Phenotypic data lacks a common hub where data can be easily discovered, queried or linked. In this context, developing an interconnected network is more practical and sustainable approach to address this limitation given the heterogeneity of data from measurements, descriptors to institutional policy and governance models. Such a network should link distributed resources, maintain autonomy at the source and maintain FAIR principles without introducing conflicts on data ownership and governance.

This deliverable takes in account the current challenges in phenotypic data exchange particularly in interconnecting diverse phenotypic databases to facilitate data sharing and integration. It outlines the conceptual basis and technical infrastructure requirements for connecting distributed data, including relevant components such as data sources, data formats communication mechanisms and mediation layer. Furthermore, it explores possible mechanisms of connection of various external phenotypic systems to EURISCO using mechanisms like standardized API, metadata harvesting etc. such that EURISCO remains the central hub for discovery of European PGR. The document also emphasizes the importance of adoption of common domain standards or efforts to harmonize phenotypic data using consistent vocabularies and unique identifiers to achieve the purpose of interconnection.

1. Introduction

The landscape of Plant Genetic Resources (PGR) research has become increasingly data-driven marked by an exponential rise in the generation of biological data and a growing diversity of data sources (**Ghamkhar et al., 2025**; **Lee et al., 2005**; **Saint Cast et al., 2022**; **Sima et al., 2019**). As data are produced across a wide range of institutions, research disciplines, and geographic regions, the PGR data systems have also grown both in complexity and scale. It now functions as a dynamic evolving ecosystem where people, systems, and data interact continuously, with flows of information supporting scientific advancement, innovation, and long-term strategic decisions (**Cobb et al., 2013**; **Rosenqvist et al., 2019**). Hence, data is no longer a static resource; something stored and accessed occasionally. Instead, it must be treated as a continuous stream of information, which is constantly available and indispensable to the advancement of scientific knowledge (**Woody et al., 2020**). This requires infrastructure and governance models that allow for real-time, flexible, and secure access to data spread across various institutes while maintaining integrity and relevance. In other words, data to truly fuel research and innovation must be made **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (FAIR). This requires transparent exchange of data across systems which is only possible when the systems in which the data exist are technically and semantically capable of communicating with one another (**Adam-Blondon et al., 2016**; **Papoutsoglou et al., 2017**; **Pommier et al., 2023**). Data interoperability thus becomes fundamental for leveraging data as a shared resource and attaining the FAIR principles, which relies on establishing interconnection mechanisms between diverse data sources.

The goal of interconnecting PGR information systems is to make diverse datasets accessible and usable beyond the boundaries of data domains or of the institutions that generate them. This promotes the core principle of open science, accessibility, transparency, reproducibility, and inclusive participation in the creation of scientific knowledge (**Ghamkhar et al., 2025**; **Halewood et al., 2018**). As data continues to grow in volume, velocity, and variety, complemented by high-throughput technologies, the ability to access and combine these data meaningfully becomes increasingly essential (**Arend et al., 2020)**. Currently, much of the PGR data is often held in isolated collections, developed and maintained by individual researchers or institutions sometimes referred to as “data siloes” (**Bayer & Edwards, 2020**; **Pommier et al., 2019a**; **Selby et al., 2019**; **Ugochukwu & Phillips, 2022**). While these datasets are valuable, their utility is significantly reduced if they cannot be discovered or integrated by other users. This fragmentation leads to inefficiencies, duplication of effort, and missed opportunities for new insights and collaboration. In such a setting, even well-documented data can become invisible and underutilized when they are not linked to a larger network.

Connecting databases across institutions and platforms also enables researchers to cross-reference genotypic, phenotypic, geographic, and environmental data leading to richer analyses and more informed decision-making in breeding programs, conservation planning, and policy development (**Adam-Blondon et al., 2016**; **Papoutsoglou et al., 2020**; **Wilkinson et al., 2016**). It also lays the groundwork for large-scale meta-analyses and comparative studies that are increasingly vital in addressing complex global challenges such as climate change and food insecurity. Importantly, interconnection extends participation of smaller organizations or those with limited technical capacity to contribute to and benefit from interconnected data environments, with minimal compatibility and documentation. Even in cases where datasets are not fully harmonized in structure or terminology, integration can be facilitated by leveraging shared metadata standards, ensuring that data can always be located, interpreted, and used in context. Moreover, connecting systems also help to enhance trust in the source and accuracy of information further reducing the risks of outdated versions being used in analyses (**Pommier et al., 2019**). This contributes to a more reliable and efficient scientific ecosystem, where data is not only abundant but also relevant and usable. These goals of interconnectivity and interoperability are not unique to the PGR domain. Across the biological sciences, data integration initiatives have long sought to address similar challenges, especially as biological data continues to grow. However, many integration systems still struggle with issues of data scalability, sustainability, and heterogeneity (**Gligorijević & Pržulj, 2015**; **Johansson et al., 2024**; **Lawler et al., 2015**). In PGR specifically, these challenges are magnified by the heterogeneity of data systems and standards, underscoring the need for intentional and coordinated interconnection. Interconnecting PGR information systems is more than a technical enhancement; it is a strategic necessity. It ensures that valuable PGR data resources are not scattered and isolated, but instead shared and integrated, potentially encouraging innovation and global collaboration.

In Europe the activities and utilisation of plant genetic resources for food and agriculture (PGRFA) have been supported by the European Cooperative Programme for Plant Genetic Resources (ECPGR) since the early 1980s. The PGR landscape in Europe includes the diverse network of more than 400 PGR collections across different countries, along with *in situ* resources (**Maxted et al., 2011**). With the aim to develop a comprehensive system that allows sharing and easy access to European plant genetic resources, ECPGR developed the European Search Catalogue for Plant Genetic Resources (EURISCO) which is now maintained by the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) on behalf of ECPGR (**Kotni et al., 2023**; **Weise et al., 2017**). EURISCO serves as a key gateway to PGR passport data. To further strengthen EURISCO as a central access point for European PGRFA information and harness its full potential, one area for development lies in enhancing interoperability with external phenotypic databases/information systems that are not currently integrated to EURISCO (**ECPGR, 2021**). This deliverable aims to address this limitation and discuss the recommendations of methods and techniques to interconnect different phenotypic information systems with EURISCO with a vision of developing EURISCO as a central information system for European PGRFA.

2. Activities

For the purpose of the deliverable, a series of comprehensive studies were carried out. An extensive literature review was conducted, covering information systems components and architecture, and methods of data exchange, with a focus on their applicability within PGR networks. Sources included peer-reviewed papers and technical reports in the fields of network development, computer science, enterprise systems, PGR and data engineering. This review aimed to identify suitable methods for interconnecting distributed data sources, assess their infrastructure and technical requirements, and evaluate their relevance to the broader PGR ecosystem. In parallel, an analysis of existing data sources in the PGR domain, including EURISCO's current operational workflows, was performed to pinpoint gaps and limitations in achieving the goal of establishing a centralised information system. Finally, as a proof of concept, a demo connection was implemented between the wheat accession (Germplasm: Barbu du Finistere) hosted on the GnpIS information system (https://urgi.versailles.inra.fr/gnpis) and EURISCO (https://ecpgr.org/eurisco/), to demonstrate the feasibility and practical considerations of data integration across systems.

3. Mechanisms of data exchange

A functional PGR information ecosystem requires streamlined data exchange mechanisms. This goes beyond than just system interconnection, it demands careful coordination between the data providers, the data architecture that defines how data is structured and governed, data type, and the communication mechanisms. If we narrow down to a single institute or organization, the mechanisms of data exchange are typically more straightforward: data sources are governed under a unified policy framework, and can be centrally managed, with communication protocols standardized across internal systems. The Harvard Enterprise Architecture framework identifies three fundamental pillars of data exchange in an organizational framework: architecture pattern (design and movement of data), data formats, and communication mechanisms (**Charest & Rogers, 2020**). This model provides a foundation for understanding interconnection design within a single organisation. However, in PGR systems, when we consider multiple data sources, the data formats, communication mechanisms and architectural pattern is not homogeneous. Therefore, to achieve an interconnected and interoperable information system, it is essential to consider not only these three pillars but also the characteristics and diversity of the data sources themselves and coordination between them. In this document, we draw on the concept of the Harvard framework to structure our discussion of data exchange mechanisms from a perspective of a larger network of systems and define the following terms in data exchange mechanisms:

A. Architecture
B. Data sources and formats
C. Communication mechanisms and
D. Mediation layer

3.1. Architectural model

An architectural model defines how a data system is structured, how data flows within it, and how governance is distributed. The choice of architecture depends on several factors, including infrastructural capacity, the purpose of data exchange, data ownership, sensitivity, and scalability needs. In the context of PGR, systems must accommodate high-volume, heterogeneous data (e.g., phenotypic traits, sequences, images), support standardization, and enable data contributors to retain control while allowing users to discover and interact with the data.

It is useful to briefly review key data architecture models to understand the compatibility of different architecture models with virtual interconnection in context of PGR data. Table 1 introduces five common models namely centralized, decentralized, distributed, federated, and hybrid models to illustrate the different ways in which independent external data sources might be organized.

Table 1. Summary of the different architectural patterns in data exchange systems

Architecture	Data management and governance model	Advantages	Limitations
Centralized architecture	Data is stored in a central system or repository.	Lower resource maintenance and update costs	High maintenance burden in the central node

	Governance mostly centralized following a top-down approach but can be decentralized according to requirement.	Data operations can be executed at once.	More risks of security breach and losses if central system gets affected Data ownership becomes complex if external system is merged Poor adaptability to high-volume heterogeneous data
Decentralized architecture	There are no central authority and entities (nodes) are fully autonomous. The governance model follows a bottom-up approach. Relevant in global, multi-institutional domains	Authority lies within independent nodes Faster adaptation to new technologies and higher scalability Encourages system interoperability efficiently	Harder to implement unified data discoverability due to inconsistent data formats and standards across nodes Needs for broad consensus on standards Demands high maintenance
Distributed architecture	Nodes are independent and coordinated with a central authority. Data authority, decision making and governance can be centralized or involve coordination between nodes and central entity.	Parallelizing work or data handling improves system efficiency Upscaling or downscaling of the system is easy.	The architecture can become complex as it grows Errors and updates can be difficult to trace Proper configuration and maintenance are required
Federated architecture	Nodes are independent and have their own authority but agree to share or commit to common standards. The architecture is more a product of governance than a data storage method.	Allows independent entities to collaborate on a shared standard. Flexible to changes in the system or the design.	Management complexity, especially in governance and decision-making. Heterogenous sources can lead to data inconsistency. Higher demands on resources and maintenance
Hybrid architecture	Combination of two or more architecture types	More flexibility in terms of standards and governance Integration of emerging data types and technologies can be easier and efficient	Can be very complex with growth of the system

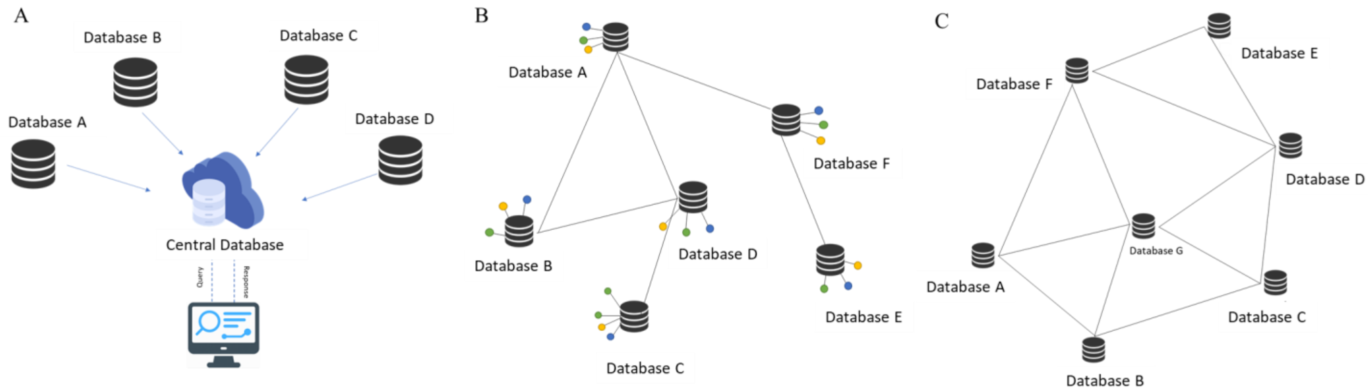


Figure 1. Overview of different data architectural patterns A. Centralized architecture, B. Decentralized architecture and C. Distributed architecture

3.1.1. Federated architectures

Federated architecture adheres to the concept of governance and authority rather than how data are spread and is designed to enable collaborations among independent data providers, while preserving their autonomy (**Bollam, 2025**; **Heimbigner & McLeod, 1985**). In a such framework, each participating institution or database retains full control over its own data or digital assets, which physically remain at the source. However, they collaborate or contribute to a common objective, such as improving discoverability and data integration, by adhering to common standards and protocols (**Feeney et al., 2010**; **Sheth & Larson, 1990**). Each node may adopt shared metadata standards (*e.g.*, **M**inimum **I**nformation **A**bout **P**lant **P**henotyping **E**xperiment (MIAPPE) for phenotypic data and multi-crop passport descriptors (MCPD) for passport data) and expose data exchange interfaces or application programme interface (API) endpoints that allow external systems to interact with them in a structured way.

Federated architectures can be implemented over decentralized or distributed architectures, depending upon how governance and control are organized (**Bustamante et al., 2023**; **Fernandez et al., 2003**). In the case of cross-institutional collaboration for a PGR data network, a federated architecture can be built over a distributed system. In this set up, each node retains data sovereignty and control while participating in a coordinated and collectively defined set of communication and interoperability standards (*e.g.*, MIAPPE, MCPD, common API specifications). Data can then be accessed through a central discovery layer, such that the interconnected system appears as a single system to the user; for instance, EURISCO specialized for Plant Genetic Resources and FAIDARE for Plant Research Data in general. In contrast, a fully decentralized federated system removes the central discovery layer, enabling each participating node to act both as a data provider and as a peer in a network, working in a peer-to-peer manner.

In terms of having a central aggregator or portal like EURISCO, this architecture can allow the interconnection with external databases such that, the central aggregator (EURISCO) only need to impose some data requirement/ formats needed by their portal and data is independently allowed to be accessed by the databases.

3.2. Data sources and formats

3.2.1. Data sources

Data sources form the foundational layer of an information system. They serve as repositories of digital assets and may be aggregated or remain independent, selected resources based on the type of architecture of the system. For PGR ecosystem, these data sources are diverse and span across conservation, breeding, genotyping, phenotyping and agronomic domains. Below (Table 2) is the non-exhaustive list of sources of data in a PGR landscape and the kind of data they hold, this is discussed in detail in the deliverable 1.5 (Aguilar et al., 2025).

Table 2. Non-exhaustive list of different data sources in PGR landscape and the type of data they host

Data sources	Types of entities	Data type
Gene banks/ *Ex situ* repositories	- National and international gene banks (e.g., CGIAR centres, Centres de resources Biologiques-CRB-INRAE, France) - Regional repositories - University or institutional collections (e.g., Royal botanic gardens, Plant departments in universities) etc.	- Passport data (origin, taxonomy) - Primary descriptors (not influenced by the environment, such as fruit colour, hairiness, etc.) - Image data - Collection location and environment data - Storage, germination and viability data - Legal and policy information
In situ and on-farm conservation sources	- Farmer's cooperatives - Local/community seed banks - Participatory breeding projects etc.	- Landrace information - Wild relatives information - Local adaptation traits - Varietal records - Habitat description - Community seed bank and local farm repository data - Access and benefit sharing protocols - Community agreement and protocols
Genomic and molecular data repositories	- NCBI GenBank - EMBL-EBI - Sequence Read Archives - Ensembl Plants - Crop-specific databases (e.g., Sol Genomics, MaizeGDB, Gramene, Genome Database for Rosaceae - GDR)	- Whole Genome Sequences (WGS) - Molecular markers, Variant data - RNAseq/Transcriptomic data, GWAS results - Quantitative Trait Loci (QTLs) - Genetic maps - Molecular markers (SSRs, AFLPs, SNPs, etc.) - Reference genomes
Breeding Institutes/Companies	- Private seed companies - CGIAR breeding platforms - National breeding institutes	- Pedigree data - Multi-location trial data - Genotype x environment effects (G x E) - Yield performance and quality trait assessments - Breeding lines and hybrid line data - Segregating populations

		- Genomic selection,
Metadata and ontology standards	- Planteome - Crop Ontology (CO) - MIAPPE - FAIRsharing.org - BioPortal	- Standardised trait descriptors - Ontologies - Experimental metadata
Environmental and geospatial sources	- WorldClim - Global biodiversity information facility (GBIF) - SoilGrids - National agro-meteorological services	- Climate data (temperature, precipitation, etc.) - Soil data (pH, nutrients, etc.) - GPS geolocation for samples - Vegetation indices and agro-ecological zones - Insect/pest/disease spread/distribution data
Data repositories	- Zenodo - Figshare - FAIRDOMhub - Recherche.data.gouv.fr - e!DAL	- Datasets linked to experiments - Supplementary data files for diverse experiments

3.2.2. Data formats

Data formats define how data are structured, encoded, and described, and enable their storage, exchange, and analysis in computational systems. To be processed efficiently, phenotypic data must be represented in machine-readable formats that follow consistent conventions for organizing observations, descriptors, and associated metadata. Over time, advances in data management and the growing importance of data integration have led to the development of more standardized formats to facilitate their reuse, sharing and interoperability (**Lapatas et al., 2015**). The table below (Table 3) lists the common but not limited to formats of data for plant phenotype.

Table 3. Different data formats

Format class	**Common data formats**
Plain text/Unstructured	txt, markdown, log files
Tabular	Comma separated values (CSV), tab separated values (TSV), Spreadsheet (.xlsx/.xls),
Structured tables/Relational	Structured query language (SQL), Relational database (PostgreSQL, MySQL), SQLite
Semantic/Ontology based	Resource Description Framework (RDF), Web ontology language (OWL)
Structured/Hierarchical	Extensible markup language (XML), Javascript object notation (JSON), YAML
Experimental metadata	MIAPPE templates, ISA-Tab (investigation/study/assay)
HTP/Sensor Data	HDF5, GeoTIFF, TIFF, JPEG, PNG, SVG, NetCDF

3.3. Communication mechanisms

Communication mechanisms are the technical methods or protocols used to enable interconnected systems exchange, expose and retrieve data. They are essentially the channel that makes interoperability possible. In this context, communication mechanisms help external databases/information systems connect and share data with EURISCO.

3.3.1. Static file export/ File transfer protocol (FTP/SFTP)

An application stores data in a file, which is then transferred to a destination location and loaded into the target system. These files may use formats such as JSON, XML, CSV, or other text-based or binary file formats. This system is used when real-time access is not required. This mechanism however transfers data physically from one place to other which is not the scope of this document.

3.3.2. Application Programming Interface (API) and web services

An API is a set of defined rules and protocols that act as a bridge connecting two systems, such as databases, software applications or devices (**De Souza et al., 2004**; **Sohan et al., 2015**; **Woody et al., 2020**). By enabling standard communication between systems, API supports data findability and accessibility and thus facilitating data integration. Furthermore, they make it easier to implement modular, flexible architectures where new data sources or applications can be connected incrementally, reducing the dependency on monolithic systems and enabling a more flexible, future-ready infrastructure. API exist in multiple forms and functionalities and can be built for libraries, operating systems, databases, or for services over the network (web). Additionally, depending upon the intended audience, an API may be open access (OpenAPI), private (internal to an institute/business), or a partner API (shared between partner organisations). For plant genetic data sharing and exchange, web-based APIs are the most relevant, as they facilitate access and exchange of information over the internet using standardized protocols and data formats (**Araya & Singh, 2017**; **Petcu et al., 2011**). While these APIs may vary in their architecture and data formats, they typically rely on well-established protocols, standards, styles and languages, to allow smooth interoperability between systems.

Common Web API protocols

An important aspect of APIs is the communication protocol they use. A protocol defines how the API connects to the internet and how it transmits information (**Araya & Singh, 2017**; **Goodwin, 2024**). There are different protocols and interaction models of web APIs. The most common ones are described below.

- RESTful APIs (Representational State Transfer) https://ics.uci.edu/rest_arch_style.htm
  RESTful APIs employ standard HTTP requests to access and use data (GET, POST, PUT, DELETE). They return data in lightweight formats like JSON or XML. RESTful APIs are widely adopted due to its simplicity and expandability (**Fielding, 2000**).

- SOAP APIs (Simple Object Access Protocol) https://www.w3.org/TR/soap/
  SOAP APIs is a more rigid protocol that uses only XML data format, and it follows strict standards to send and receive requests and responses. As a result, SOAP puts an overhead

burden for the client with significant setup and processing overhead (**Yates et al., 2015**). However, it offers higher security and formal standardization.

API in PGR data systems

Standardized RESTful APIs can facilitate automated retrieval of passport data, trait observations, or genetic data records, ensuring that information remains up-to-date and harmonized across platforms. Various PGR specific repositories and information system use APIs to enable interconnection facilitating standardized and efficient data exchange. These are usually custom APIs or webservices or broader domain-specific APIs.

Domain-specific API: Domain specific APIs are tailored for use within a specific field. A commonly used domain-specific API is BrAPI (Breeding API), which is commonly used for the exchange of plant breeding and genetic data across institutions and platforms. BrAPI is a RESTful web API designed to enable access and exchange of germplasm information data, trial metadata, phenotypic observations, and genomic marker data (**Selby et al., 2019**). It is being implemented by numerous databases such as Germinate, Breedbase, GnpIS and data portals (e.g. FAIDARE).

Custom API: Custom APIs are developed by individual projects or databases to meet their own data exchange requirements or sometimes to manage legacy system constraints. These are locally used, does not follow broader community standards and can be public or private. Some examples of these are

- *Genesys PGR API* (https://www.genesys-pgr.org/documentation/apis): Genesys uses custom RESTful APIs allowing access to passport data and descriptor information on millions of accessions.
- *CGIAR genebank API* (CLARISA): CGIAR uses CLARISA (CGIAR Level Agricultural Results Interoperable System Architecture), which is a RESTAPI that enables different CGIAR systems like MARLO (Managing Agricultural Research for Learning and Outcomes) and MEL (Monitoring evaluation and learning) to exchange information on research data as well as various other data among themselves.

3.3.3. Semantic Query protocols (Linked data)

SPARQL protocol: SPARQL Protocol and RDF Query Language (SPARQL) is a query language and protocol for linked open data and RDF databases allowing retrieval and manipulation of data stored in RDF format in a distributed or decentralised system. It requires each source to expose SPARQL endpoints to allow access to its data. For effective use of SPARQL protocol, all the data sources should use RDF format and rely on shared, stable vocabularies or ontologies which makes it less common in phenotypic data or current PGR landscape.

3.4. Mediation layer

A mediation layer in modern information systems is a software component that ensures smooth system operations by automating complex workflows, managing communication, and coordinating tasks across different services and infrastructure components **(Osborn, 2025**). The challenge in accessing data from multiple sources lies not only in interconnection, but doing so in a way that is efficient, unified, and interpretable. Consider a plant breeder or geneticist aiming to study a disease resistance trait or evaluate a QTL associated with resistance to a particular pathogen. The relevant datasets, such as phenotype scores, environmental conditions, cultivar performance, and trial metadata, may be distributed across multiple databases hosted by different gene banks, research institutes, and breeding companies. In a general interconnected system where each database exposes its data through APIs, users must query each source individually, and the responses may be fragmented, uncoordinated and redundant **(Neven & Van de Craen, 2006)**. In cases where data sources do not have API endpoints (access), relevant data may remain inaccessible through just API requests. As the system grows in complexity, particularly when integrating heterogeneous and geographically distributed datasets, a central mediation layer becomes essential **(Lapatas et al., 2015; Krajewski et al., 2015**; **Schantz & Schmidt, 2007)**. Depending upon the complexity and goal of the architectural system, the mediation layer may incorporate several components, such as 'middleware' to bridge incompatible systems, 'workflow engines' to manage task sequencing in data processing or 'API gateways' to manage incoming request and route them to correct services components. The mediation layer becomes increasingly valuable when real-time data retrieval, unified queries and flexible workflow execution are required across multiple diverse PGR systems. For a sustainable and user-friendly interconnection of several PGR databases to a central system, a minimal mediation architecture should include the following components.

3.4.1. Middleware

Middleware is a broad class of software which act as an intermediary between applications and data systems, providing services of communication, protocol translation, and data harmonisation. It effectively "glues" disparate systems together **(Bernstein, 1996; Verma, 2022)**. In the context of PGR where multiple systems operate independently, Middleware helps integration without requiring significant changes to the original systems, thereby preserving local autonomy and legacy infrastructure.

Middleware can provide various services depending on the system design:

- **Database middleware:** when databases speak different query languages, it translates the request from a user or application into something the database understands, retrieves the required data, and send it back in a usable format**.**
- **API middleware:** help to define standard ways to ask and receive information, especially if the original system lacks API endpoints.
- **Message-oriented middleware**; manages and organizes many requests at once or queues requests, etc. (**Schantz & Schmidt, 2007**).
- **Adapters and connectors:** transform data formats (*e.g.*, converting custom XML to BrAPI-compliant JSON) and establish connections to various resources, such as databases, file systems, or external services.

3.4.2. Query federation and aggregation tools

In a decentralized or distributed system, data often stored across multiple, independent databases are made accessible through interfaces such as API, web portals, etc. To ensure interconnection across sources, a system must distribute the query to multiple databases in parallel and aggregate the corresponding responses, a process made possible by a component known as a query mediator.

A query mediator is a software mechanism that orchestrates the entire process of query federation. It receives the user query, breaks it down into sub-queries specific to each specific database, sends them simultaneously, and finally aggregates the results into a single output. Importantly, query federation is achieved without physically moving data from its source, thus preserving institutional data sovereignty.

3.4.3. Central interface system

The central interface serves as the primary user-facing entry point enables query submission, data discovery and curation across distributed data sources. While much of the data exchange and integration happens behind the scenes (through backend components, such as API interfacing, the orchestration layer and databases), the interface is the visible front end through which users interact with the system. To maintain interoperability, the interface through which data is made accessible to the user should also have infrastructures to align with proper data exchange.

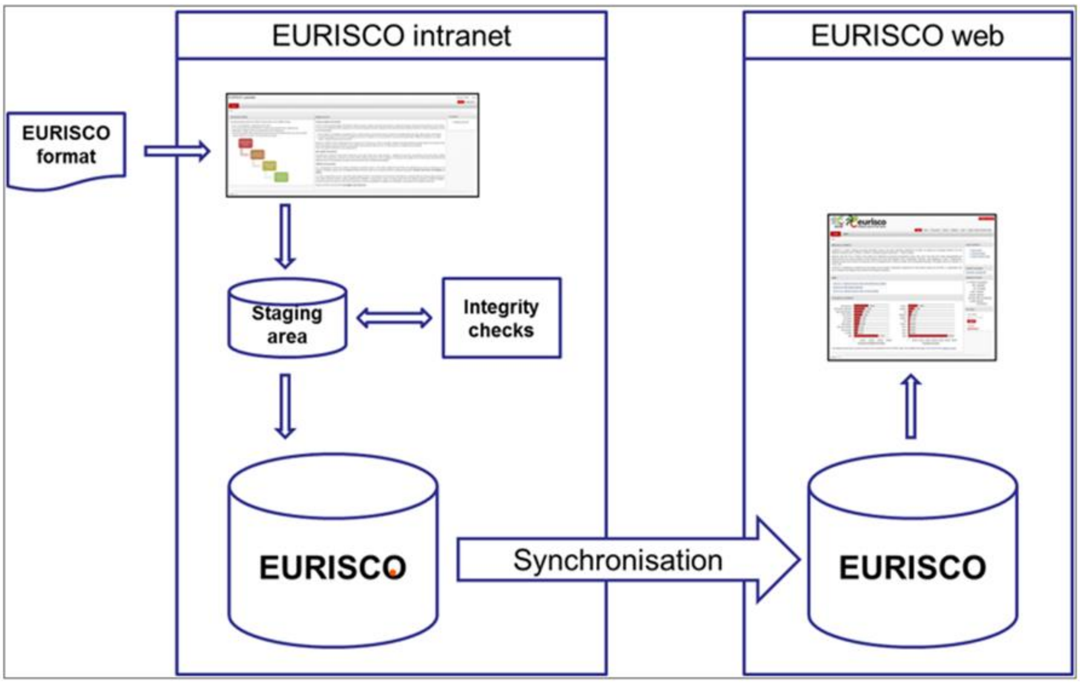
4. EURISCO current status

The European Search Catalogue for Plant Genetic Resources (EURISCO) is the central information gateway for accessing data on plant genetic resources (PGR) across Europe. It was launched in 2003 by the European Cooperative Programme for Plant Genetic Resources (EPCGR), with support of an EC-funded project to establish a European Plant Genetic Resources Information Infra-Structure, EPGRIS (EPGRIS). It is currently maintained by the Leibniz Institute of Plant Genetics and Crop Research (IPK). It functions as the European node of Genesys (https://www.genesys-pgr.org/), and plays a crucial role in the international co-ordination and visibility of European PGR (ECPGR, 2021). Currently, EURISCO compiles and maintains data that includes passport data and phenotypic data through collaboration with National Focal Points (NFP) from 43-member countries of the geographical Europe. The catalogue documents 2,106,681 *ex situ*-conserved accessions from 419 individual collections and 6,384 *in situ*-conserved populations of crop wild relatives (CWR), the whole encompassing 6,775 genera and 45,424 species (derived from Deliverable 1.5; Aguilar et al.,2025). While EURISCO represents one of the most comprehensive European PGR catalogues, a substantial data gaps persists. Many PGR collections across Europe, especially those maintained by smaller institutes, universities, regional programs, remain either underrepresented or absent. This underrepresentation is driven by combination of factors including institutional and infrastructural limitations, fragmented data management systems, inconsistent legal and policy frameworks for data sharing, and the inherent complexity of managing and standardizing diverse PGR datasets, particularly in the scope of phenotypic information. These numerous challenges and detailed list of underrepresented data in EURISCO are elaborated in the Deliverable 1.5.

4.1. Architecture and working model of EURISCO

EURISCO operates in a centralized database architecture with decentralised governance. Its data ingestion and access model rely on NFPs designated by each of the 43-member countries who serve as the primary intermediaries between national and institutional PGR data sources and EURISCO database (**Kotni et al., 2023**).

Data from over 400 contributing institutes are collected, validated and uploaded to the staging area via secure intranet interface where it undergoes additional consistency checks and a quality control process. Following the approval, the validated data are published to the production database and become accessible to users via the EURISCO frontend web interface **(Weise et al., 2017)**. Initially, EURISCO was designed to enable the discoverability of *ex situ* passport data through structured querying, making it a reliable catalogue of Europe's genetic resource holdings.



(Weise et al., 2017)

Figure 2. Working model of EURISCO

5. Results

5.1. Recommended interconnected network

To gradually transform European phenotype landscape into interconnected network with central discovery hub capable of integrating and coordinating broad PGR data ecosystems, a strategic architectural shift is needed. One of the promising outlooks for this network should be to evolve into a hybrid federated architecture where EURISCO functions as a central discovery hub while continuing to receive centralized data submissions from National Focal points. Here, EURISCO would establish virtual connections with independently maintained external information systems, databases etc. This virtual linking would require interfacing/communication mechanisms such as APIs, metadata exchange services, persistent identifiers and semantic mappings. This architecture balances scalability, institutional autonomy and could support real-time interoperability of data. Within this architecture the scope of what should be stored centrally versus what should remain distributed requires careful distinction.

5.1.1. Hybrid federated architecture

Central system for hosting data from National Inventories (NI) and Federated interconnection with external systems

Within its current architecture, EURISCO continues to function as a centralized hub for hosting structured passport and phenotypic data from NFPs, while maintaining decentralized governance. Individual countries remain responsible for collecting, validating, and authorizing their data before submission via manual or semi-automated file transfers, and EURISCO provides a central catalogue that enhances accessibility and visibility.

For phenotypic data, central submission remains most appropriate for compact characterization and evaluation (C&E) datasets generated by local genebanks, community seedbanks, or short-term national projects whose outputs might otherwise be fragmented or lost. While, high-volume and complex HTP datasets should be federated, as this would impose unsustainable burdens if centralised. Beyond the type of data, however, the readiness of repositories must also be a primary consideration. Smaller providers such as seedbanks typically rely on flat files and lack the scalability to support dynamic connections, making central submission the only possible data sharing strategy. Larger national repositories and institutional databases, in contrast, often have the infrastructure to adopt interoperable standards and can gradually evolve dynamic interconnection. For these systems, interconnection strategies should be put in place according to their infrastructural readiness (e.g. API availability, metadata standardization etc.). This architecture of interconnected network of independent data in a federated system is envisioned as an advanced querying or indexing layer that connects to multiple external databases, supporting both metadata discovery, and where permitted, direct access to datasets. Potential external systems for federated connection can include:

- Crop-specific databases and information systems, such as Sol Genomics Network (https://solgenomics.net/), SoyBase (https://www.soybase.org/), GnpIS (https://urgi.versailles.inra.fr/Tools/GnpIS), Genome Database for Rosaceae GDR (https://www.rosaceae.org), Gramene (https://www.gramene.org/)
- Multi-thematic repositories from the European life-sciences infrastructure for biological information (ELIXIR; https://elixir-europe.org/) plant community or the European Infrastructure for Multi-Scale Plant Phenotyping and Simulation for Food Security in a Changing Climate (EMPHASIS; https://emphasis.plant-phenotyping.eu/)
- Institutional or research infrastructures developed by international collaborative projects such as Germinate (https://germinateplatform.github.io/get-germinate/#)

The recommended federated architecture would provide centralized search functionality, allowing users to query distributed data sources and retrieve both metadata and datasets from different systems. It would reduce the burden of navigating separate platforms. To enable such a federated infrastructure, most importantly each external system connected to EURISCO should maintain a minimum level of FAIR compliance.

5.2. Recommendations for interconnection to EURISCO

The idea of interconnecting different phenotypic databases to EURISCO relies heavily on the ability of external data sources to expose their data in standardized, accessible and interoperable ways. Since the concept is to virtually access the data rather than physical ingestion to EURISCO, this is possible only when EURISCO can reference or query data stored in a different system. For the communication interfaces to function reliably underlying data should at least be structured and well described. Thus, for meaningful and FAIR interconnection, external systems (Phenotypic information systems, gene banks, breeding institutes etc) should adopt community agreed standards for data content, expose rich metadata and gradually develop mechanisms supporting connection and data exchange (API endpoints, DOI assignments, metadata harvesting). This implementation should be done from the outset of data curation as the extent to which data is well structured and described directly impacts the feasibility and sustainability of interoperability and integration. Without well described data, structured metadata and identifiers in place from the beginning, efforts to interconnect systems become significantly more complex and fragile.

Structuring captured data

One of the major problems in the phenotypic data landscape is the lack of harmonization in how descriptors, variables, and observational information are recorded. This inconsistency in data content, particularly trait names, methods, and environmental conditions limit interoperability across systems. Without a consistent vocabulary and data model, even basic queries become unreliable. To address this issue, data providers should adopt domain specific standards and vocabularies (e.g. trait ontology, crop ontology, Planteome, XEML Environment Ontology) for structuring the data of phenotyping experiments. By aligning trait names and variable definitions across datasets, it becomes possible to semantically link or federate data from different sources, increasing both discoverability and analytical value. For example, one dataset might record plant height trait as “height”, another uses “plant_height” and a third dataset as “Ht”. Even though these represent the same measurement, automated systems would identify them as three separate traits without semantic alignment. By mapping all three to a shared ontology term like TO:0000207 (the Trait Ontology term for "plant height"), systems can recognize them as equivalent. This allows EURISCO or any central platform to aggregate or federate phenotypic data on "plant height" across multiple databases enabling unified queries even when different terminologies are used in different datasets.

Making data discoverable with standardized metadata and persistent identifiers

To be linked to a central system, the datasets must be first discoverable. This requires exposing well-structured metadata that describes the context, ownership and accessibility of the dataset. A well-structured metadata can allow queryable access through APIs or through harvesting/cataloguing. MIAPPE provides the domain specific standard for describing phenotypic experiments. However, MIAPPE defines what needs to be described not how metadata is formatted or shared. Formatting of MIAPPE compliant datasets should be done in semantic supporting systems like ISA-Tab (Investigation Study Assay) or DCAT (Data Catalog Vocabulary) which allows automated packaging of metadata thus allowing cataloguing or harvesting. These systems allow central systems like EURISCO to interpret, catalogue, and potentially federate metadata even if full data exchange through techniques like API is not established.

Assessing the status of phenotypic systems in terms of data exchange

One of the important principles to consider in phenotypic domain is that the phenotypic data system varies greatly in their capacity in integrating existing technologies and standards for data exchange. This ranges from advanced infrastructures (in current context) that are already ontology-driven, API-enabled and fully aligned with standards like MIAPPE, BrAPI and MCPD to lower-level systems that have little to no adherence to data standards. As a result, interconnection cannot follow a single uniform model that fits for all. Advanced systems tend to converge around shared solutions like ontology alignment, standardized metadata and which makes it straightforward to interconnect with a common approach. However, systems with low readiness are more heterogeneous in data characteristics and technical capacity to connect to other systems. To classify these differences broadly, phenotypic systems can be divided based on their data characteristics and feasible interconnection mechanisms (Table 4). This division should, however, be seen as a continuum rather than rigid categories and the interconnection strategies should be flexible to accommodate future improvements and long-term support for less mature systems.

Table 4. Readiness levels for interconnecting phenotypic systems and their characteristics

Readiness	Data characteristics	Interconnection strategies (Virtual)	Next steps recommendation
Low	Flat files (CSV/XLS/PDF); sparse metadata, little/no MIAPPE coverage, uncontrolled terms for trait and descriptors, local accession codes	Static links to landing pages; this linking is feasible to link dataset from accession page in EURISCO	Gradual adoption of minimal metadata templates; training on systematic PID assignment; map accession IDs to MCPD; Data deposit to long term repositories
Moderate	Relational Databases; Structured catalogue records; partial MIAPPE/metadata standardization; some controlled lists; Partial PID coverage	DOI based accession-to-dataset links; metadata harvest where metadata is/can be exposed; JSON/XML exports	expand MIAPPE and FDP adoption; strengthen PID coverage and ontologies
High	Normalized schemas; MIAPPE-aligned descriptors; versioned metadata; units/methods defined; Persistent study/trial/accession IDs; resolvable URIs; could have API integration	Metadata harvesting for catalogues; Direct API integration where possible, and ontology mapping where applicable	Consistent API (BrAPI) implementations; increase ontology coverage and maintenance protocols
Advanced	RDF/JSON-LD; SPARQL; provenance models; Ontology-driven (CO/TO/Units); MIAPPE complete; machine-actionable semantics; Entity-level URIs/DOIs (accession, sample, dataset) with resolvers	API integration; Semantic linking; cross linking at study/trial and accession level	Broaden federated connection and strengthen sustainability of the system

5.2.1. Linking data in the federated system: services for discoverability of the data

<u>Static linking:</u>

Static linking refers to use of fixed, pre-defined reference links such as DOIs, stable URLs or accession specific landing pages for linking data. In federated system, this can be a straightforward method to connect/link external datasets to EURISCO thus, this method connects accession-to-dataset. To connect phenotypic data of accessions already present in EURISCO (via passport data), these links can be embedded within the accession metadata. For accessions not yet in EURISCO, additional, minimal metadata cataloguing should be established to enable linking.

Some repositories such as Zenodo assign DOIs to phenotypic trial datasets, publication and metadata packages while some with accession-based data assign identifiers or DOIs in accession level. If any form of persistent identifiers are not available, well-maintained and structured URLs to specific dataset or trait pages can be linked to EURISCO accessions. For example, a phenotypic dataset deposited in a national data portal (e.g. Recherche. data.gouv.fr) assigned with DOI can be linked at the EURISCO interface without requiring API protocols or direct data harvesting. This can be particularly useful for legacy datasets or published data from experimental trials, unrepresented institutes, to which connection through other methods may not be very efficient and may cause additional burden. To ensure proper linking, identifiers must be stable, ideally persistent, and clearly point to an accession or dataset landing page. Corresponding metadata must also be indexed in EURISCO to contextualize the link. While this linking does support findability but does not allow dynamic querying or data integration and requires proper cataloguing of the PIDs ensuring it is linked to a proper accession/dataset/metadata. Similarly, URLs may create problems of broken links when the contents are moved or replaced

<u>Metadata harvesting:</u>

Metadata harvesting allows EURISCO to periodically ingest structured metadata from external systems without transferring the full data. This approach can be suitable for the institutes/projects/repositories which can support and expose standard metadata through interoperable services (API, FAIR Data Points (FDP), Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)). EURISCO could maintain a central metadata catalogue where connected databases register their dataset descriptions. This could be elaborated based on the FAIR Data Point (FDP) and FAIDARE approaches. Given the heterogeneity of phenotypic data, a minimum FAIR metadata profile should be defined that all the connecting systems are expected to meet. This metadata structure must align with existing domain standards such as MIAPPE and BrAPI (experimental-level metadata) and generic standards like DCAT or Dublin Core (dataset-level description) that clearly describes its content, structure, licensing, and provenance to facilitate automated federated discovery. This approach also supports semi-dynamic integration, which can scale over time, and reduces the risk of broken links.

<u>API based link</u>

In the recommended federated model, EURISCO acts as the central coordination point, initiating outbound connections to multiple external data sources via different standardized API protocols, persistent identifiers or semantic links. API connection is the most used communication mechanism in data management and analytics nowadays. EURISCO could expose API endpoints to give programmatic access to its data. Reciprocally, data sources equipped with structured databases would need to adopt EURISCO recommended API specifications. This could facilitate establishing EURISCO compatible standards for data exchange formats but on the contrary, it could limit scalability and cross-platform compatibility unless widely adopted by external systems.

For phenotypic PGR data, the most appropriate API standard that could be adopted by EURISCO currently is BrAPI (Breeding API). It has a domain-specific design and is compliant with community standards (such as MCPD, MIAPPE, and the Crop Ontology), thus promoting interoperability and data standardization across systems (**Pommier et al., 2019**; **Selby et al., 2019**). Additionally, BrAPI has a well-established community, active development, and reference implementations, making it well sustainable. The BrAPI 2.0 (https://github.com/plantbreeding/BrAPI/releases/tag/V2.0) defines four major modules: core, phenotyping, genotyping, and germplasm. For implementation, platforms can selectively support only the endpoints relevant to their data and use cases. A growing number of PGR platforms, including repositories, web portals, and crop-specific databases, use BrAPI services by either implementing BrAPI endpoints, integrating BrAPI client libraries, or embedding BrAPI-compliant tools to support interoperability (Table 5).

Table 5. Platforms utilizing BrAPI in PGR landscape

Platform	URL	Primary Use	BrAPI Usage
Breedbase	https://breedbase.org/	Breeding data management and analysis	Uses BrAPI libraries to retrieve and analyse phenotypic and genotypic data from BrAPI compliant systems for analysis and visualization (**Morales et al., 2022**)
Germinate based platforms	https://germinateplatform.github.io/get-germinate/#	Plant resource platform for storing and sharing PGR data	Recent BrAPI implementation focuses on germplasm and genotypic data, Expansion to other BrAPI modules is ongoing (**Raubach et al., 2021**).
FAIDARE	https://urgi.versailles.inra.fr/faidare/	Data portal for federated access to genotype and phenotype datasets	Unified BrAPI-compliant interface aggregating phenotypic and genotype data from distributed repositories through endpoints provided by partner systems.
GnpIS	https://urgi.versailles.inra.fr/gnpis/	Phenotyping data aggregation platform	BrAPI endpoints for phenotypic data, especially trait variables and observations, integrated Crop ontology through tools like trait-ontology-widget
Crop Ontology	https://cropontology.org/	Standardized trait definitions for crops	Provides BrAPI compliant services for trait ontologies, allowing access to standardized variable definitions
Grin-Global	http://grin-global.org/	Genebank data management	Implements BrAPI read-only endpoints

			for germplasm and related accession level data allowing retrieval of passport and inventory information

However, not all external systems that EURISCO could interface with are BrAPI-compliant, structured around breeding data models, or even equipped with API endpoints. While BrAPI is ideal for breeding and phenotypic databases, many other systems such as environmental datasets, gene bank inventory files, or legacy records may rely on custom RESTful APIs, SQL databases, Excel spreadsheets, or flat files.

To bridge these gaps, middleware can serve as a key architectural layer between EURISCO and external systems with different API protocols. Middleware helps with taking request from EURISCO's portal and translating them according to the capabilities (API specifications, query structures) of the target database. For this system to be practical, each middleware service must be specifically configured to align with the technical protocols and data schemas of its corresponding system, which can become labour-intensive and costly over time. Besides, a robust query federation service could be beneficial for EURISCO to support federated querying. This will allow receiving user queries via the EURISCO portal, identifying relevant external resources based on indexed metadata, and routing those queries across multiple databases in parallel. It would then coordinate the aggregation and delivery of responses, facilitating real-time discovery and access across a distributed landscape.

There are several ways to implement such an integrated system, depending on how sources are connected and level of standardization across databases.

- In early stages where data structures vary significantly, a custom-built query mediator within a mediation layer allows to define rules and data transformations, as needed.
- As more databases begin to adopt standard schemas and interfaces (*e.g.*, BrAPI), a federated query engine offer a more automated and scalable solution. These engines handle query splitting and response aggregation with minimal manual configuration.
- Incorporating a metadata catalogue can improve query efficiency by helping the system determine which databases hold relevant information and how to access it.

5.2.2. Community capacity building and Onboarding activities

Creating a hybrid architecture would mean EURISCO would potentially connect with different sources with varying institutional and technical capabilities. For progressive onboarding and improvement of data-sharing capability of the stakeholders, it is essential to strengthen community network and capabilities through training and shared knowledge. EURISCO can develop onboarding tools like metadata templates, sample data with annotations, BrAPI implementation guidelines, examples of compliant submission, validation checklists, etc.

Many platforms still lack important data structure requirements as discussed in the section 5.2, (standard metadata, data formats and identifiers) that allows basic long-term linking and interoperability between independent systems. Here, European research infrastructures and platforms play a pivotal role in shaping the future interconnection of phenotypic data to EURISCO. Infrastructures such as EMPHASIS (ESFRI project for plant phenotyping), EPPN2020 (European Plant Phenotyping Network), and national platforms like PHENOPlant provide harmonized, high-quality environments for field and controlled-condition phenotyping. They play important role in standardizing data generation processes, ensuring MIAPPE compliance, promoting BrAPI implementation, and developing tools like PHIS/OpenSILEX, which facilitate data interoperability which is the emerging necessity in PGR phenotypic ecosystem. These infrastructures also provide long-term sustainability for data by aligning different national platforms under a shared vision and technical framework. For example, projects like EPPN2020 offer access to high-throughput platforms and define experimental protocols that enhance data quality and reuse. Moreover, these infrastructures can serve as intermediary data aggregators, ensuring that even ephemeral project outputs can be archived and made discoverable through future portals or DOI assignments.

5.2.3. Access, authorization and Governance

In federated system when disparate sources are interconnected, clear policies on access control and data governance must be established. Each data ingested or connected to EURISCO should clearly define the access conditions, licensing (e.g. creative commons, embargoes) and technical requirements (e.g. API authentication keys). Based on the access conditions, a tiered access could be put in place to allow databases to gradually adopt the data standards and exchanges. For example, on the first level, sources may only provide access to the metadata which is mandatory, and then further levels of access could be added according to the characteristics and accessibility of the data. This multi-tiered approach allows data providers to participate according to their technical and legal readiness, encouraging wider adoption while maintaining minimum threshold for FAIR metadata.

Besides, EURISCO should formalise its governance outlining but not limited to clear data sharing principles, responsibility division between EURISCO and data providers, operational procedures for onboarding, metadata validation and removal, mechanisms for dispute resolution and sustainable protocols to maintain the overall connected system.

5.2. Potential phenotypic systems to be connected to EURISCO

Table 6 presents a non-exhaustive list of phenotypic data repositories, project-specific databases, and information systems in Europe that could be potentially interconnected with the EURISCO. These systems demonstrate variable technical readiness for interoperability, including alignment with FAIR principles, the use of standardized metadata schemas (e.g., MIAPPE, MCPD), implementation of persistent identifiers (e.g., DOIs, URIs), utilization of controlled vocabularies and ontologies (e.g., Crop Ontology, Trait Ontology), and the availability of programmatic access via REST APIs, BrAPI, or SPARQL endpoints. These features allow either static metadata linking (e.g., embedding external URLs or DOIs in EURISCO's accession records) or dynamic connections (e.g. API access). However, the degree of technical maturity varies across systems. For example, if a project/dat repository project hosts phenotypic data through a dedicated portal, however, it lacks both programmatic interfaces and DOIs. Although static linking is theoretically possible by embedding accession page URLs into EURISCO's passport records, the absence of persistent identifiers increases the risk of broken links and maintenance challenges if URLs change over time. Therefore, each system may require a case-by-case strategy to enable meaningful interconnection.

Table 6. List of selected data repositories/information systems and their characteristics for potential interconnection to EURISCO

Data system	Host/country	Type of system	Characteristics of the phenotypic data	Technical capacity for data exchange

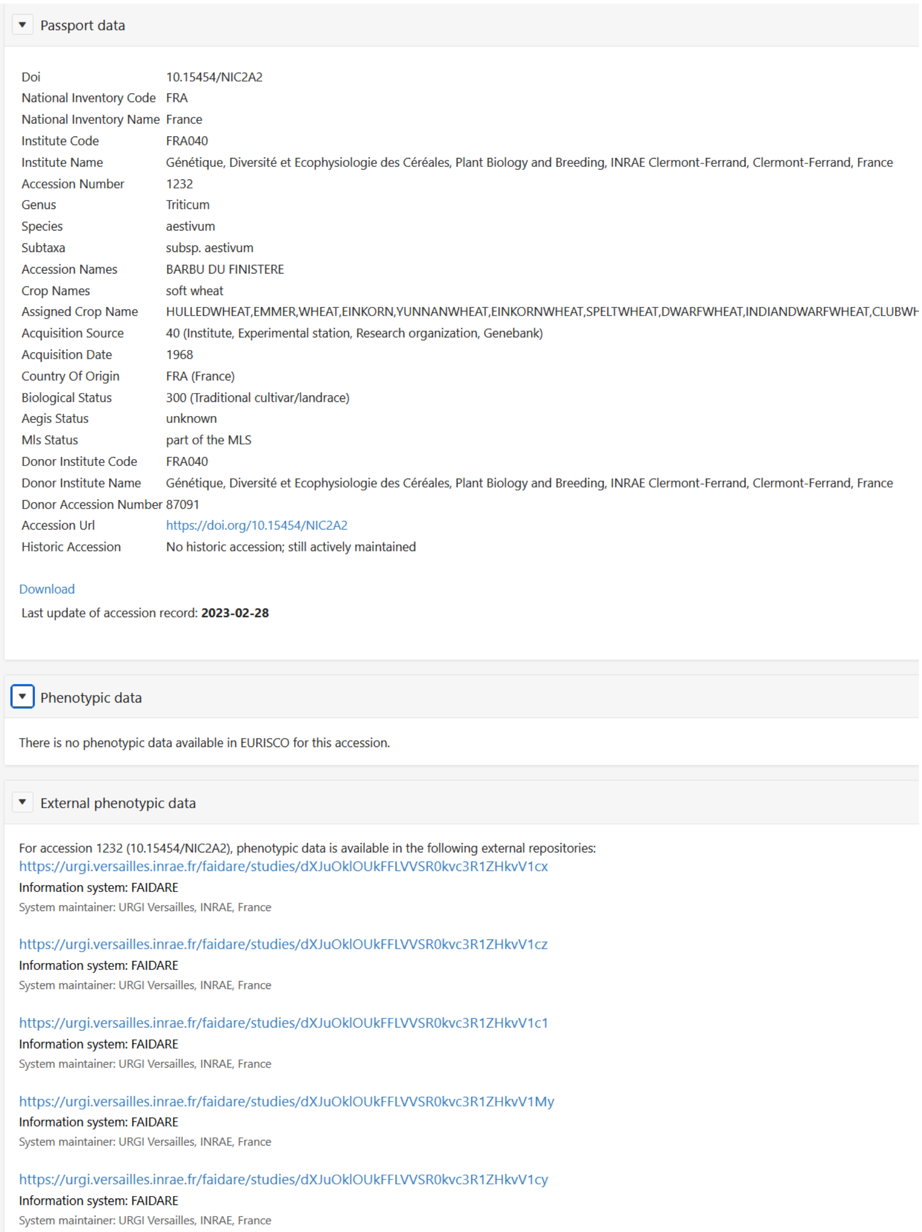
Brassica information portal (BIP; https://bip.earlham.ac.uk/)	Earlham institute, UK	Information portal	Characterization and evaluation data in Brassica species	- Public API; programmatic access - DOI assigned to datasets - Supports controlled vocabularies and ontologies (CO, TO)
e!DAL-PGP (https://edal-pgp.ipk-gatersleben.de/)	IPK, Germany	Data repository	Plant phenomics/genomics research data including imaging data	- FAIR compliant repository, - DOI linking possible to accession page in EURISCO - Currently lacks native support for real-time linking
G2P-SOL gateway (https://www.g2p-sol.eu/G2P-SOL-gateway.html)	Coordination: ENEA, Italy; data hosted on: Phenome networks	Project Database	Accession level trait data of four major solanaceous crops (Potato, tomato, pepper and eggplant)	- No public API endpoints for G2P-SOL gateway however host Phenome network has BrAPI integration - Possible connection through collaboration with Phenome networks or - Static metadata links to the zenodo datasets or landing accession pages
Gen4olive (https://gen4olive.eu/)	Coordination: University of Córdoba, Spain	Project portal	Characterization data of (~1700) Olive germplasm from across Mediterranean basin	- No programmatic access - Static metadata linking can be possible by linking the accession-specific landing pages
Germinate based repositories (https://germinateplatform.github.io/get-germinate/)	Multiple institutions (Decentralized)	Collaborative database	Genotypic, phenotypic, passport and climate data of various crops	- Compliance to MCPD and metadata Standards (Dublin core) - Rest API coverage - Variable BrAPI and MIAPPE implementation at different institutions/database
GnpIS/Ephesis (https://urgi.versailles.inrae.fr/ephesis/ephesis/)	INRAE, France	Information system/Data integration platform	Field and controlled environment trial, perennial and multi-year phenotyping, environmental covariates from research and breeding across crops and French institutions	- Standardized metadata (MIAPPE aligned) - BrAPI v2, REST API endpoints - Supports Ontology mapping and FAIR principles

HARNESSTOM (http://harnesstom.eu/en/index.html)	Agencia Estatal Consejo Superior de Investigaciones Científicas, Spain	Project portal/Database	Genotypic, Phenotypic data including fruit quality, disease resistance characteristics) in cultivated Tomatoes in European Union	- Accessions already mapped to EURISCO passport records, static linking to phenotypic database with DOIs on accession page is possible - Well-structured data - No API integration available
PHIS (http://www.phis.inra.fr/)	INRAE, France	Information system	Ontology-based, High throughput phenotypic data (sensor data, imaging, time series), along with environmental data	- Ontology centered architecture - Standardized metadata and semantic annotations - Utilizes RESTful webservices, exposes BrAPI and SPARQL endpoints - High technical maturity, requires semantic expertise to deploy the linking
PIPPA (PSB-Interface for Plant Phenotype Analysis); https://www.psb.ugent.be/phenotyping/pippa	VIB, Belgium	Phenotypic data Platform	Large raw phenotypic and imaging data from HTP platforms	- FAIR data based on MIAPPE standards - Public API and BrAPI implementation
FAIDARE (https://urgi.versailles.inrae.fr/faidare/)	INRAE, France	Data portal	Datasets search in several Plant databases including e!DAL-PGP, GnpIS, PIPPA, PHIS	- It will soon replace GnpIS/Ephesis for dataset integration and meta-analysis - Standardized metadata (MIAPPE aligned) depending on the source database - BrAPI v2, REST API endpoints - Supports Ontology and FAIR principles

5.3. Demo interconnection of Wheat germplasm Barbu du Finistere dataset to EURISCO

For the proof of concept of interconnection, a demonstrative connection was implemented in the beta version of EURISCO (not publicly available). Static links (resolvable URLs) of phenotypic data from Barbu du Finistere accession from the European project Whealbi were embedded within the metadata page of the accessions in EURISCO. Here, the accession page in EURISCO (Fig 3a) provides links as "External phenotyic data" to the Phenotyping studies where this accession has been used. The links lead the user to FAIDARE (Fig 3b.) to display all the metadata (location, accessions list, traits) and from FAIDARE to GnpIS/Ephesis (Fig 3c.) to display the phenotypic data (phenotypes values and dates) and options to export in standard format (Fig 3d.). Thus, through the links a user can discover phenotypic data links in EURISCO that are in fact located in GnPIS system.

A live demo is available on EURISCO beta:
https://eurisco.ipk-gatersleben.de/apex/eurisco_ws_dev/r/pro-grace-demo



Passport data

Doi	10.15454/NIC2A2
National Inventory Code	FRA
National Inventory Name	France
Institute Code	FRA040
Institute Name	Génétique, Diversité et Ecophysiologie des Céréales, Plant Biology and Breeding, INRAE Clermont-Ferrand, Clermont-Ferrand, France
Accession Number	1232
Genus	Triticum
Species	aestivum
Subtaxa	subsp. aestivum
Accession Names	BARBU DU FINISTERE
Crop Names	soft wheat
Assigned Crop Name	HULLEDWHEAT,EMMER,WHEAT,EINKORN,YUNNANWHEAT,EINKORNWHEAT,SPELTWHEAT,DWARFWHEAT,INDIANDWARFWHEAT,CLUBWH
Acquisition Source	40 (Institute, Experimental station, Research organization, Genebank)
Acquisition Date	1968
Country Of Origin	FRA (France)
Biological Status	300 (Traditional cultivar/landrace)
Aegis Status	unknown
Mls Status	part of the MLS
Donor Institute Code	FRA040
Donor Institute Name	Génétique, Diversité et Ecophysiologie des Céréales, Plant Biology and Breeding, INRAE Clermont-Ferrand, Clermont-Ferrand, France
Donor Accession Number	87091
Accession Url	https://doi.org/10.15454/NIC2A2
Historic Accession	No historic accession; still actively maintained

Download

Last update of accession record: **2023-02-28**

Phenotypic data

There is no phenotypic data available in EURISCO for this accession.

External phenotypic data

For accession 1232 (10.15454/NIC2A2), phenotypic data is available in the following external repositories:
https://urgi.versailles.inrae.fr/faidare/studies/dXJuOklOUkFFLVVSR0kvc3R1ZHkvV1cx
Information system: FAIDARE
System maintainer: URGI Versailles, INRAE, France

https://urgi.versailles.inrae.fr/faidare/studies/dXJuOklOUkFFLVVSR0kvc3R1ZHkvV1cz
Information system: FAIDARE
System maintainer: URGI Versailles, INRAE, France

https://urgi.versailles.inrae.fr/faidare/studies/dXJuOklOUkFFLVVSR0kvc3R1ZHkvV1c1
Information system: FAIDARE
System maintainer: URGI Versailles, INRAE, France

https://urgi.versailles.inrae.fr/faidare/studies/dXJuOklOUkFFLVVSR0kvc3R1ZHkvV1My
Information system: FAIDARE
System maintainer: URGI Versailles, INRAE, France

https://urgi.versailles.inrae.fr/faidare/studies/dXJuOklOUkFFLVVSR0kvc3R1ZHkvV1cy
Information system: FAIDARE
System maintainer: URGI Versailles, INRAE, France

Figure 3a. Static links embedded in EURISCO accession page

Study Phenotyping Study: Wheat_Winter_INRA

Origin site Collecting site Evaluation site Multi-purpose site

Identification

Name	Wheat_Winter_INRA
Identifier	dXJuOklOUkFFLVVSR0kvc3R1ZHkvV1cx
Source	GnpIS
Data link	Link to this study on URGI GnpIS
Project name	Whealbi
Active	No
Seasons	2015
Location name	INRA Clermont-Ferrand

Genotype

Accession number	Name	Taxon
1747	114/62	Triticum aestivum subsp. aestivum
92	11IWSWSN14	Triticum aestivum subsp. aestivum
347	2838-39	Triticum aestivum subsp. aestivum
421	3716-1	Triticum aestivum subsp. aestivum
514	6-1-3	Triticum aestivum subsp. aestivum
524	60293	Triticum aestivum subsp. aestivum
546	664-258-18	Triticum aestivum subsp. aestivum
39115	79TK098-511	Triticum turgidum subsp. dicoccum
39111	859	Triticum turgidum subsp. dicoccum
39113	97	Triticum turgidum subsp. dicoccum
748	A.4	Triticum aestivum subsp. aestivum
39259	ABI	Triticum aestivum subsp. aestivum
38989	ACTROS	Triticum aestivum subsp. aestivum
794	ADMONTER	Triticum aestivum subsp. aestivum

Variables

Variable ID	Variable short name	Variable long name	Ontology name	Trait description
WIPO:0000276	CarbonContent	Carbon content	Wheat Inra Phenotype Ontology	
WIPO:0000279	Picture	Picture of the accession/field	Wheat Inra Phenotype Ontology	Name of the file or archive where photo(s) of the accession/field can be found

Figure 3b. Redirection from EURISCO to FAIDARE through the static link

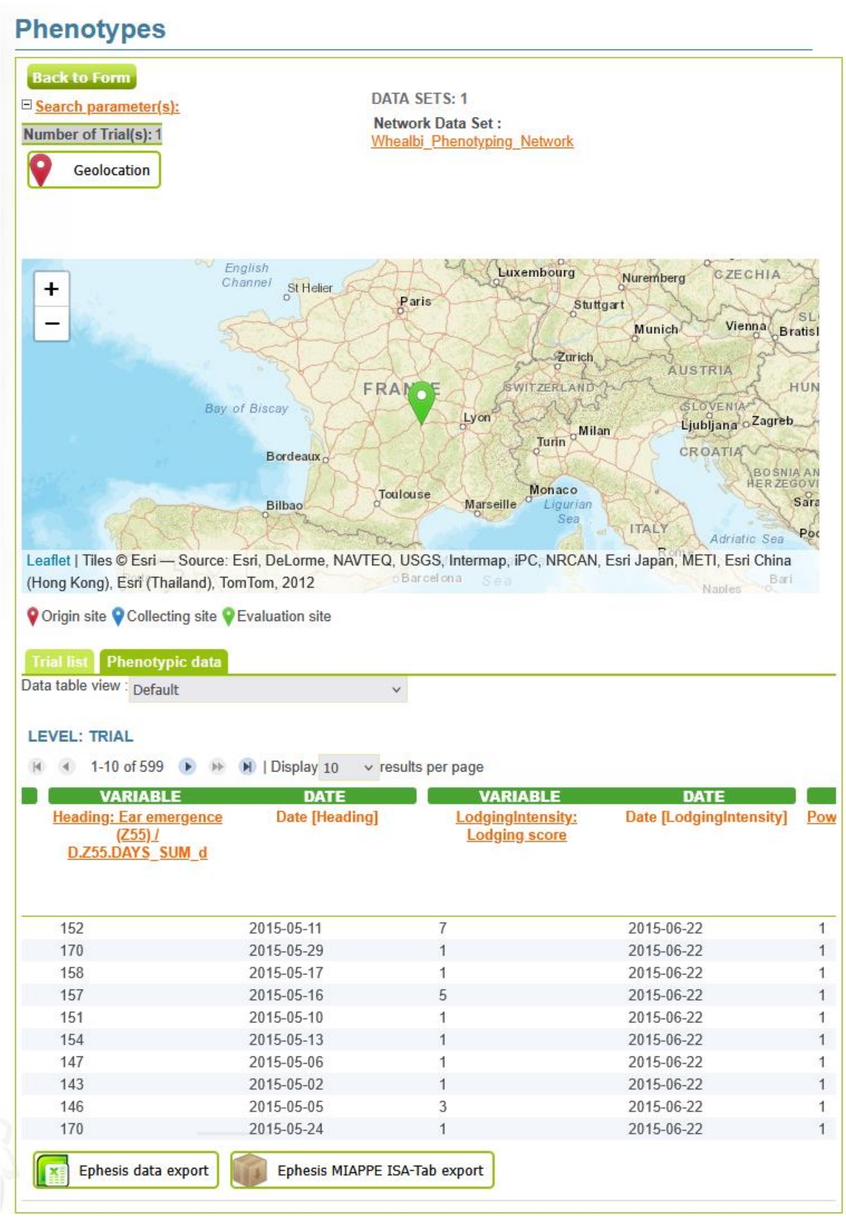
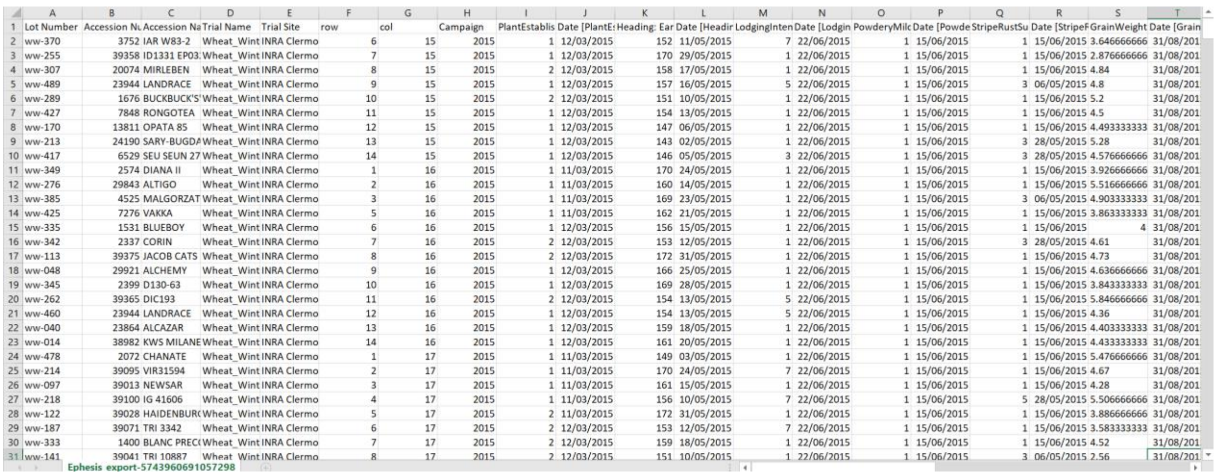


Figure 3c. Phenotypic data in GnPIS/Ephesis



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Lot Number	Accession Nu	Accession Na	Trial Name	Trial Site	row	col	Campaign	PlantEstablis	Date [PlantEs	Heading: Ear	Date [Headir	LodgingInten	Date [Lodgin	PowderyMilc	Date [Powde	StripeRustSu	Date [StripeF	GrainWeight	Date [Grain
2	ww-370	3752	IAR W83-2	Wheat_Wint	INRA Clermo	6	15	2015	1	12/03/2015	152	11/05/2015	7	22/06/2015	1	15/06/2015	1	15/06/2015	3.646666666	31/08/201
3	ww-255	39358	ID1331 EP03:	Wheat_Wint	INRA Clermo	7	15	2015	1	12/03/2015	170	29/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	2.876666666	31/08/201
4	ww-307	20074	MIRLEBEN	Wheat_Wint	INRA Clermo	8	15	2015	2	12/03/2015	158	17/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.84	31/08/201
5	ww-489	23944	LANDRACE	Wheat_Wint	INRA Clermo	9	15	2015	1	12/03/2015	157	16/05/2015	5	22/06/2015	1	15/06/2015	3	06/05/2015	4.8	31/08/201
6	ww-289	1676	BUCKBUCK'S'	Wheat_Wint	INRA Clermo	10	15	2015	2	12/03/2015	151	10/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	5.2	31/08/201
7	ww-427	7848	RONGOTEA	Wheat_Wint	INRA Clermo	11	15	2015	1	12/03/2015	154	13/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.5	31/08/201
8	ww-170	13811	OPATA 85	Wheat_Wint	INRA Clermo	12	15	2015	1	12/03/2015	147	06/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.493333333	31/08/201
9	ww-213	24190	SARY-BUGDA	Wheat_Wint	INRA Clermo	13	15	2015	1	12/03/2015	143	02/05/2015	1	22/06/2015	1	15/06/2015	3	28/05/2015	5.28	31/08/201
10	ww-417	6529	SEU SEUN 27	Wheat_Wint	INRA Clermo	14	15	2015	1	12/03/2015	146	05/05/2015	3	22/06/2015	1	15/06/2015	3	28/05/2015	4.576666666	31/08/201
11	ww-349	2574	DIANA II	Wheat_Wint	INRA Clermo	1	16	2015	1	11/03/2015	170	24/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	3.926666666	31/08/201
12	ww-276	29843	ALTIGO	Wheat_Wint	INRA Clermo	2	16	2015	1	11/03/2015	160	14/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	5.516666666	31/08/201
13	ww-385	4525	MALGORZAT	Wheat_Wint	INRA Clermo	3	16	2015	1	11/03/2015	169	23/05/2015	1	22/06/2015	1	15/06/2015	3	06/05/2015	4.903333333	31/08/201
14	ww-425	7276	VAKKA	Wheat_Wint	INRA Clermo	5	16	2015	1	11/03/2015	162	21/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	3.863333333	31/08/201
15	ww-335	1531	BLUEBOY	Wheat_Wint	INRA Clermo	6	16	2015	1	12/03/2015	156	15/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4	31/08/201
16	ww-342	2337	CORIN	Wheat_Wint	INRA Clermo	7	16	2015	2	12/03/2015	153	12/05/2015	1	22/06/2015	1	15/06/2015	3	28/05/2015	4.61	31/08/201
17	ww-113	39375	JACOB CATS	Wheat_Wint	INRA Clermo	8	16	2015	2	12/03/2015	172	31/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.73	31/08/201
18	ww-048	29921	ALCHEMY	Wheat_Wint	INRA Clermo	9	16	2015	1	12/03/2015	166	25/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.636666666	31/08/201
19	ww-345	2399	D130-63	Wheat_Wint	INRA Clermo	10	16	2015	1	12/03/2015	169	28/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	3.843333333	31/08/201
20	ww-262	39365	DIC193	Wheat_Wint	INRA Clermo	11	16	2015	2	12/03/2015	154	13/05/2015	5	22/06/2015	1	15/06/2015	1	15/06/2015	5.846666666	31/08/201
21	ww-460	23944	LANDRACE	Wheat_Wint	INRA Clermo	12	16	2015	1	12/03/2015	154	13/05/2015	5	22/06/2015	1	15/06/2015	1	15/06/2015	4.36	31/08/201
22	ww-040	23864	ALCAZAR	Wheat_Wint	INRA Clermo	13	16	2015	1	12/03/2015	159	18/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.403333333	31/08/201
23	ww-014	38982	KWS MILANE	Wheat_Wint	INRA Clermo	14	16	2015	1	12/03/2015	161	20/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.433333333	31/08/201
24	ww-478	2072	CHANATE	Wheat_Wint	INRA Clermo	1	17	2015	1	11/03/2015	149	03/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	5.476666666	31/08/201
25	ww-214	39095	VIR31594	Wheat_Wint	INRA Clermo	2	17	2015	1	11/03/2015	170	24/05/2015	7	22/06/2015	1	15/06/2015	1	15/06/2015	4.67	31/08/201
26	ww-097	39013	NEWSAR	Wheat_Wint	INRA Clermo	3	17	2015	1	11/03/2015	161	15/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.28	31/08/201
27	ww-218	39100	IG 41606	Wheat_Wint	INRA Clermo	4	17	2015	1	11/03/2015	156	10/05/2015	7	22/06/2015	1	15/06/2015	5	28/05/2015	5.506666666	31/08/201
28	ww-122	39028	HAIDENBURG	Wheat_Wint	INRA Clermo	5	17	2015	2	11/03/2015	172	31/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	3.886666666	31/08/201
29	ww-187	39071	TRI 3342	Wheat_Wint	INRA Clermo	6	17	2015	2	12/03/2015	153	12/05/2015	7	22/06/2015	1	15/06/2015	1	15/06/2015	3.583333333	31/08/201
30	ww-333	1400	BLANC PREC(	Wheat_Wint	INRA Clermo	7	17	2015	2	12/03/2015	159	18/05/2015	1	22/06/2015	1	15/06/2015	1	15/06/2015	4.52	31/08/201
31	ww-141	39041	TRI 10887	Wheat_Wint	INRA Clermo	8	17	2015	2	12/03/2015	151	10/05/2015	1	22/06/2015	1	15/06/2015	3	06/05/2015	2.56	31/08/201

Ephesis export-5743960691057298

Figure 3d. Complete access of phenotypic data in the tabular form

6. Deviations

The original goal of this deliverable was to implement a functional interconnection between EURISCO and external phenotypic databases. However, it became clear that such integration involves legal, policy, and governance issues beyond the timeframe and capacity of the current project. All these aspects could be addressed in the future GRACE-RI given their complexity and the time needed to tackle them.

First, cross-border data exchange requires formal agreements among all parties, outlining ownership, access rights, and responsibilities. In addition to that a successful interconnection would also require proper governance model in place and legal framework to ensure compliance, transparency and trust. EURISCO's structure already ensures that National Inventories (NI) retain data control, so any broader connection must respect and align with this principle. Secondly, sensitive phenotypic data from breeding and research institutes and project initiatives also require legislation and institutional policy and data protection frameworks. Such alignments could not be achieved within this project's timeframe.

Considering these constraints, the implementation of such goals was reconsidered. Indeed, instead of building a full system, the project developed a proof of concept by integrating a dataset of wheat accession (Germplasm: Barbu du finistere) into EURISCO to show the technical feasibility when proper agreements will be in place.

The document presented here now serves as a proof of concept as well as a concept note. It offers the strategic foundation to guide the establishment of EURISCO as a central information system. It outlines future steps and recommendations toward interconnections and furthermore moving forward the PGR domain should aim for: multilateral formal agreements, common data standards, proper governance structures and legal alignment defining responsibilities for data protection, and sustained funding.

References

Adam-Blondon, A.-F., Alaux, M., Pommier, C., Cantu, D., Cheng, Z.-M., Cramer, G. R., Davies, C., Delrot, S., Deluc, L., Di Gaspero, G., Grimplet, J., Fennell, A., Londo, J. P., Kersey, P., Mattivi, F., Naithani, S., Neveu, P., Nikolski, M., Pezzotti, M., … Quesneville, H. (2016). Towards an open grapevine information system. *Horticulture Research*, *3*, 16056. https://doi.org/10.1038/hortres.2016.56

Araya, C., & Singh, M. (2017). Web API protocol and security analysis. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-208934

Arend, D., König, P., Junker, A., Scholz, U., & Lange, M. (2020). The on-premise data sharing infrastructure e!DAL: Foster FAIR data for faster data acquisition. GigaScience, 9(10), giaa107. https://doi.org/10.1093/gigascience/giaa107

Bayer, P. E., & Edwards, D. (2020). Machine learning in agriculture: From silos to marketplaces. Plant Biotechnology Journal, 19(4), 648–650. https://doi.org/10.1111/pbi.13521

Bernstein, P. A. (1996). Middleware: A model for distributed system services. Communications of the ACM, 39(2), 86–98. https://doi.org/10.1145/230798.230809

Bollam, A. (2025). The rise of federated systems in cloud-native architectures. World Journal of Advanced Research and Reviews, 26(3), 207–215. https://doi.org/10.30574/wjarr.2025.26.3.2124

Bustamante, P., Gomez, M., Krishnamurthy, P., Madison, M. J., Murtazashvili, I., Palanisamy, B., Palida, A., & Weiss, M. B. H. (2023). On the Governance of Federated Platforms. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4528712

Charest, G., & Rogers, M. (2020). Data Exchange Mechanisms and Considerations. Harvard University Information Technology. https://enterprisearchitecture.harvard.edu/data-exchange-mechanisms

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics, 126(4), 867–887. https://doi.org/10.1007/s00122-013-2066-0

De Souza, C. R. B., Redmiles, D., Cheng, L.-T., Millen, D., & Patterson, J. (2004). Sometimes you need to see through walls: A field study of application programming interfaces. Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, 63–71. https://doi.org/10.1145/1031607.1031620

ECPGR. (2021). Plant Genetic Resources strategy for Europe. European Cooperative Programme for Plant Genetic Resources. https://www.ecpgr.org/fileadmin/bioversity/publications/pdfs/PGR_STRATEGY_LP_22_Nov_revised.pdf

Feeney, K., Brennan, R., Keeney, J., Thomas, H., Lewis, D., Boran, A., & O'Sullivan, D. (2010). Enabling decentralised management through federation. Computer Networks, 54(16), 2825–2839. https://doi.org/10.1016/j.comnet.2010.07.006

Fernandez, G., Zhao, L., & Wijegunaratne, I. (2003). Patterns for Federated Architecture. The Journal of Object Technology, 2(3), 135. https://doi.org/10.5381/jot.2003.2.3.a4

Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures [Ph.D]. University of California.

Ghamkhar, K., Hay, F. R., Engbers, M., Dempewolf, H., & Schurr, U. (2025). Realizing the potential of plant genetic resources: The use of phenomics for genebanks. PLANTS, PEOPLE, PLANET, 7(1), 23–32. https://doi.org/10.1002/ppp3.10570

Gligorijević, V., & Pržulj, N. (2015). Methods for biological data integration: Perspectives and challenges. Journal of the Royal Society Interface, 12(112). Scopus. https://doi.org/10.1098/rsif.2015.0571

Goodwin, M. (2024, April 9). What Is an API (Application Programming Interface)? | IBM. IBM. https://www.ibm.com/think/topics/api

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 Phytologist, 217(4), 1407–1419. https://doi.org/10.1111/nph.14993

Heimbigner, D., & McLeod, D. (1985). A federated architecture for information management. ACM Transactions on Information Systems, 3(3), 253–278. https://doi.org/10.1145/4229.4233

Johansson, L. F., Laurie, S., Spalding, D., Gibson, S., Ruvolo, D., Thomas, C., Piscia, D., de Andrade, F., Been, G., Bijlsma, M., Brunner, H., Cimerman, S., Dizjikan, F. Y., Ellwanger, K., Fernandez, M., Freeberg, M., van de Geijn, G.-J., Kanninga, R., Maddi, V., … Solve-RD consortium. (2024). An interconnected data infrastructure to support large-scale rare disease research. GigaScience, 13, giae058. https://doi.org/10.1093/gigascience/giae058

Kotni, P., van Hintum, T., Maggioni, L., Oppermann, M., & Weise, S. (2023). EURISCO update 2023: The European Search Catalogue for Plant Genetic Resources, a pillar for documentation of genebank material. Nucleic Acids Research, 51(D1), D1465–D1469. https://doi.org/10.1093/nar/gkac852

Krajewski, P., Chen, D., Ćwiek, H., van Dijk, A. D. J., Fiorani, F., Kersey, P., Klukas, C., Lange, M., Markiewicz, A., Nap, J. P., van Oeveren, J., Pommier, C., Scholz, U., van Schriek, M., Usadel, B., & Weise, S. (2015). Towards recommendations for metadata and data handling in plant phenotyping. Journal of Experimental Botany, 66(18), 5417–5427. https://doi.org/10.1093/jxb/erv271

Lapatas, V., Stefanidakis, M., Jimenez, R. C., Via, A., & Schneider, M. V. (2015). Data integration in biological research: An overview. Journal of Biological Research, 22(1), 9. https://doi.org/10.1186/s40709-015-0032-5

Lawler, M., Siu, L. L., Rehm, H. L., Chanock, S. J., Alterovitz, G., Burn, J., Calvo, F., Lacombe, D., Teh, B. T., North, K. N., Sawyers, C. L., & on behalf of the Clinical Working Group of the Global Alliance for Genomics and Health (GA4GH). (2015). All the World's a Stage: Facilitating Discovery Science and Improved Cancer Care through the Global Alliance for Genomics and Health. Cancer Discovery, 5(11), 1133–1136. https://doi.org/10.1158/2159-8290.CD-15-0821

Lee, J. M., Davenport, G. F., Marshall, D., Ellis, T. H. N., Ambrose, M. J., Dicks, J., van Hintum, T. J. L., & Flavell, A. J. (2005). GERMINATE. A Generic Database for Integrating Genotypic and Phenotypic Information for Plant Genetic Resource Collections. Plant Physiology, 139(2), 619–631. https://doi.org/10.1104/pp.105.065201

Maxted, N., Dulloo, M. E., Miguel A. A. Pinheiro de carvalho, Iriondo, J., Ford-Lloyd, B. V., & Frese, L. (2011). Agrobiodiversity Conservation Securing the Diversity of Crop Wild Relatives and Landraces. CABI.

Morales, N., Ogbonna, A. C., Ellerbrock, B. J., Bauchet, G. J., Tantikanjana, T., Tecle, I. Y., Powell, A. F., Lyon, D., Menda, N., Simoes, C. C., Saha, S., Hosmani, P., Flores, M., Panitz, N., Preble, R. S., Agbona, A., Rabbi, I., Kulakow, P., Peteti, P., … Mueller, L. A. (2022). Breedbase: A digital ecosystem for modern plant breeding. G3 Genes|Genomes|Genetics, 12(7), jkac078. https://doi.org/10.1093/g3journal/jkac078

Neven, F., & Van de Craen, D. (2006). Optimizing Monitoring Queries over Distributed Data. In Y. Ioannidis, M. H. Scholl, J. W. Schmidt, F. Matthes, M. Hatzopoulos, K. Boehm, A. Kemper, T. Grust, & C. Boehm (Eds.), Advances in Database Technology—EDBT 2006 (pp. 829–846). Springer. https://doi.org/10.1007/11687238_49

Osborn, T. (2025, May 9). Data Orchestration 101: Process, Benefits, Challenges, And Tools. Data Orchestration 101: Process, Benefits, Challenges, And Tools. https://www.montecarlodata.com/blog-what-is-data-orchestration/

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

Papoutsoglou, E., Kaliyaperumal, R., Hintum, V., Visser, R. G. F., Athanasiadis, I. N., & Finkers, R. (2017). Toward better data sharing methods for genebanks. https://agris.fao.org/search/en/providers/122575/records/64746aebbf943c8c79801231

Petcu, D., Craciun, C., & Rak, M. (2011). Towards a Cross Platform Cloud API - Components for Cloud Federation. Proceedings of the 1st International Conference on Cloud Computing and Services Science (CLOSER-2011), 166–169. https://www.researchgate.net/publication/220865558_Towards_a_Cross_Platform_Cloud_API_-_Components_for_Cloud_Federation

Pommier, C., Coppens, F., Ćwiek-Kupczyńska, H., Faria, D., Beier, S., Miguel, C., Michotey, C., D'Anna, F., Owen, S., & Gruden, K. (2023). Plant Science Data Integration, from Building Community Standards to Defining a Consistent Data Lifecycle. In H. F. Williamson & S. Leonelli (Eds.), Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development (pp. 149–160). Springer International Publishing. https://doi.org/10.1007/978-3-031-13276-6_8

Pommier, C., Michotey, C., Cornut, G., Roumet, P., Duchêne, E., Flores, R., Lebreton, A., Alaux, M., Durand, S., Kimmel, E., Letellier, T., Merceron, G., Laine, M., Guerche, C., Loaec, M., Steinbach, D., Laporte, M. A., Arnaud, E., Quesneville, H., & Adam-Blondon, A. F. (2019). Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS. Plant Phenomics, 2019, 1671403. https://doi.org/10.34133/2019/1671403

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

Rosenqvist, E., Großkinsky, D. K., Ottosen, C.-O., & van de Zedde, R. (2019). The Phenotyping Dilemma—The Challenges of a Diversified Phenotyping Community. Frontiers in Plant Science, 10. https://doi.org/10.3389/fpls.2019.00163

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

Schantz, R. E., & Schmidt, D. C. (2007). Middleware for Distributed Systems: Evolving the common structure for network-centric applications. In B. W. Wah (Ed.), Wiley Encyclopedia of Computer Science and Engineering (1st ed.). Wiley. https://doi.org/10.1002/9780470050118.ecse241

Selby, P., Abbeloos, R., Backlund, J. E., Basterrechea Salido, M., Bauchet, G., Benites-Alfaro, O. E., Birkett, C., Calaminos, V. C., Carceller, P., Cornut, G., Vasques Costa, B., Edwards, J. D., Finkers, R., Yanxin Gao, S., Ghaffar, M., Glaser, P., Guignon, V., Hok, P., Kilian, A., … The BrAPI consortium. (2019). BrAPI—an application programming interface for plant breeding applications. Bioinformatics, 35(20), 4147–4155. https://doi.org/10.1093/bioinformatics/btz190

Sheth, A. P., & Larson, J. A. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Comput. Surv., 22(3), 183–236. https://doi.org/10.1145/96602.96604

Sima, A. C., Stockinger, K., de Farias, T. M., & Gil, M. (2019). Semantic integration and enrichment of heterogeneous biological databases. Methods in Molecular Biology, 1910, 655–690. Scopus. https://doi.org/10.1007/978-1-4939-9074-0_22

Sohan, S. M., Anslow, C., & Maurer, F. (2015). A Case Study of Web API Evolution. 2015 IEEE World Congress on Services, 245–252. https://doi.org/10.1109/SERVICES.2015.43

Ugochukwu, A. I., & Phillips, P. W. B. (2022). Data sharing in plant phenotyping research: Perceptions, practices, enablers, barriers and implications for science policy on data management. The Plant Phenome Journal, 5(1), e20056. https://doi.org/10.1002/ppj2.20056

Verma, P. (2022). A Brief Study of Middleware Technologies: Programming Applications and Management Systems. Novel Research Aspects in Mathematical and Computer Science Vol. 1, 174–181. https://doi.org/10.9734/bpi/nramcs/v1/16138D

Weise, S., Oppermann, M., Maggioni, L., van Hintum, T., & Knüpffer, H. (2017). EURISCO: The European search catalogue for plant genetic resources. Nucleic Acids Research, 45(D1), D1003–D1008. https://doi.org/10.1093/nar/gkw755

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

Woody, S. K., Burdick, D., Lapp, H., & Huang, E. S. (2020). Application programming interfaces for knowledge transfer and generation in the life sciences and healthcare. Npj Digital Medicine, 3(1), 1–5. https://doi.org/10.1038/s41746-020-0235-5

Yates, A., Beal, K., Keenan, S., McLaren, W., Pignatelli, M., Ritchie, G. R. S., Ruffier, M., Taylor, K., Vullo, A., & Flicek, P. (2015). The Ensembl REST API: Ensembl Data for Any Language. Bioinformatics, 31(1), 143–145. https://doi.org/10.1093/bioinformatics/btu613