Project:	EURISCO	Date:	2024-01-23

From:	Stephan Weise (IPK)	To:	Lorenzo Maggioni (ECPGR Secretary) Theo van Hintum (Head, EURISCO Advisory Board)
CC:	Markus Oppermann (IPK)		

This document describes the activities carried out in 2023 for hosting and maintaining the European Search Catalogue for Plant Genetic Resources (EURISCO), and for coordinating the EURISCO network.

Content

1 EURISCO development

The position of EURISCO developer expired on 15 March 2023, the position of the developer for the extension to include *in situ* CWR data on 31 May 2023. Development tasks were then taken over by the EURISCO coordinator.

1.1 Extension of the intranet support mechanism for data providers (permanent activity)

1.1.1 EURISCO update mechanisms for *ex situ* passport data and for phenotypic data

Uploads of *ex situ* passport data and phenotypic data are initiated via the same web application. Various PL/SQL packages are used in the background for this purpose. Both the web application and the PL/SQL packages were continuously maintained and further developed. This concerns minor enhancements in terms of content, but in particular bug fixing. The revision of the upload mechanism for phenotypic data, which was completed in 2022, has proven to be very stable.

1.1.2 Extension of EURISCO backend for *in situ* CWR passport data

The adaptation of EURISCO for *in situ* CWR data started in October 2022 within the project 'Extension of EURISCO for Crop Wild Relatives (CWR) *in situ* data and preparation of pilot countries' data sets', funded by the German Federal Ministry of Food and Agriculture and coordinated by the ECPGR Secretariat. Based on the white paper 'Principles for the Inclusion of CWR Data in EURISCO' (van Hintum & Iriondo 2022), a data exchange standard was already defined and a first version of a mechanism to support the upload process was implemented and tested. As the group of data providers of *in situ* CWR data is not necessarily the same as that of *ex situ* passport data, a separate web interface for uploading *in situ* CWR data is used for this purpose.

Building on this initial work, the extension of EURISCO for *in situ* CWR data was continued in 2023. The focus here was on finalising various PL/SQL packages for upload and data integrity checks as well as extensive testing of these. Procedures for the subsequent integration of the data into the EURISCO database schema were newly developed and also tested. The PL/SQL packages were then integrated into the intranet web interface for *in situ* CWR data (Figure 1).

Based on feedback from *in situ* CWR data providers, extensions were made to the data standard. This affected the descriptors SITEPROT and CONSACTION and resulted in changes to the database schema. To allow multiple values (only one value was possible in the first version of the standard), additional bridging tables had to be added to the database schema. The procedures for integrity checks and data integration also had to be adapted in order to be able to handle multiple values.

The pilot group of data providers involved in the project provided real data in order to test the overall system of the *in situ* extension under real conditions. Any errors that occurred were fixed. Once the tests in the development environment were completed, the *in situ* extension was rolled out in the production system.

At the end of 2023, the first production *in situ* CWR data from four countries was imported into EURISCO and made publicly available (see section 2 (EURISCO coordination)).

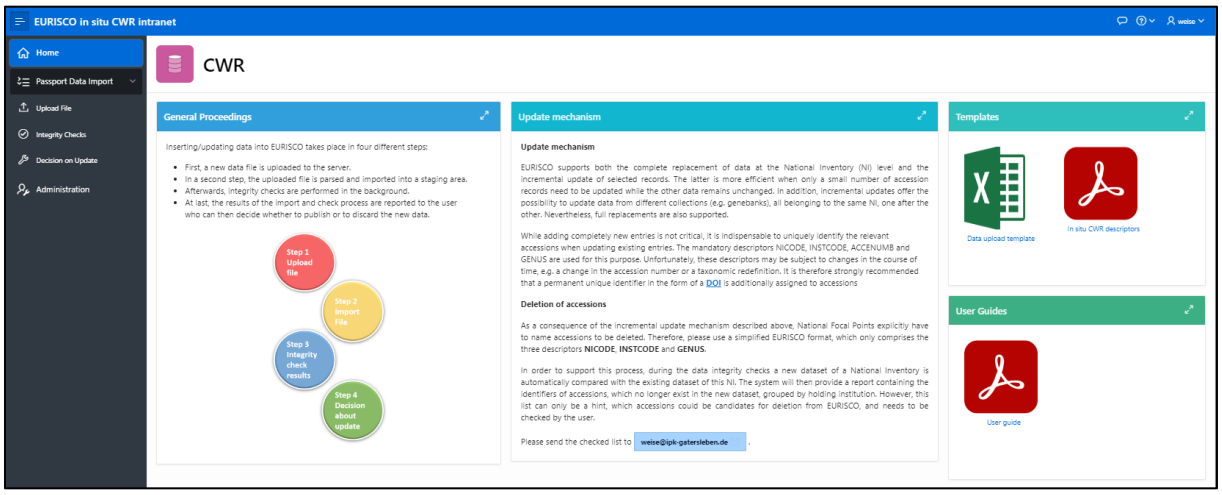

Figure 1: Screenshot from the web application for uploading *in situ* CWR data.

1.1.3 Update how-to documents

How-to documents, descriptions of data exchange formats etc. are continuously updated. A new how-to document describing the upload of *in situ* CWR data was compiled.

1.2 Extension of functionality of the public EURISCO application (permanent activity)

1.2.1 Continuous development of the existing public web interface

The public EURISCO web application was continuously maintained and further developed in 2023. In particular, the following enhancements were made:

A subpage with a data dictionary was created to better explain the column headings used in the various reports in the application. This includes terms from the EURISCO-MCPD standard for *ex situ* passport data and from the standard for *in situ* CWR data.

An additional export mechanism has been implemented to provide a full dump of the passport data in EURISCO as a CSV file. Previously, a full dump was only provided as a MS Access file. Since CSV files are widely used for data exchange in the field of life sciences, the additional mechanism was established (Figure 2).

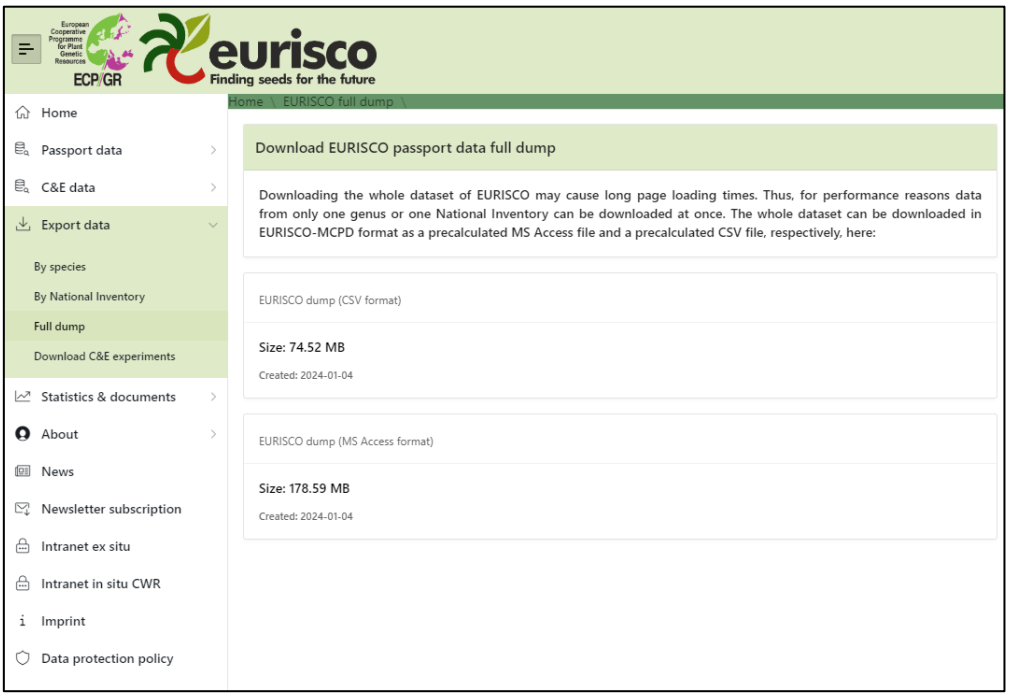

Figure 2: Download facility providing full dumps using different file formats.

For each search possibility for passport data, additional filter options were implemented to limit the number of hits to those accessions for which phenotypic data is available (Figure 3).

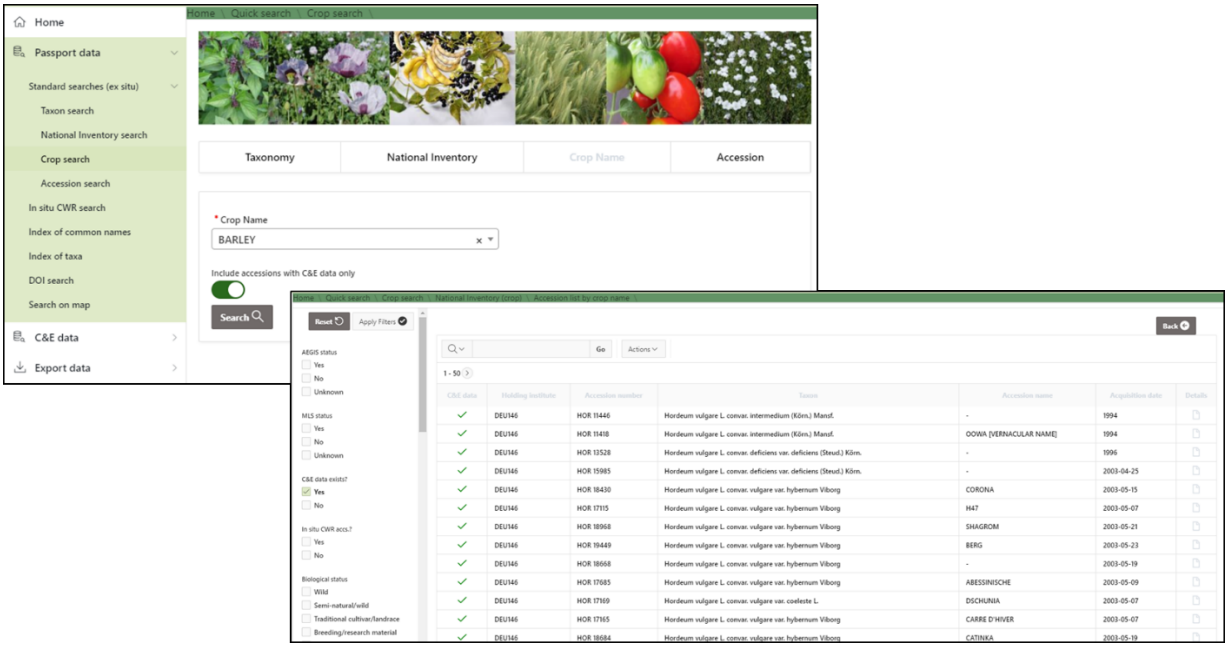

Figure 3: Different possibilities to filter for accessions with phenotypic data only.

As the acceptance of Digital Object Identifiers (DOI) as permanent and unique identifiers for genebank accessions continues to grow, an additional search mechanism for DOIs has been implemented (Figure 4).

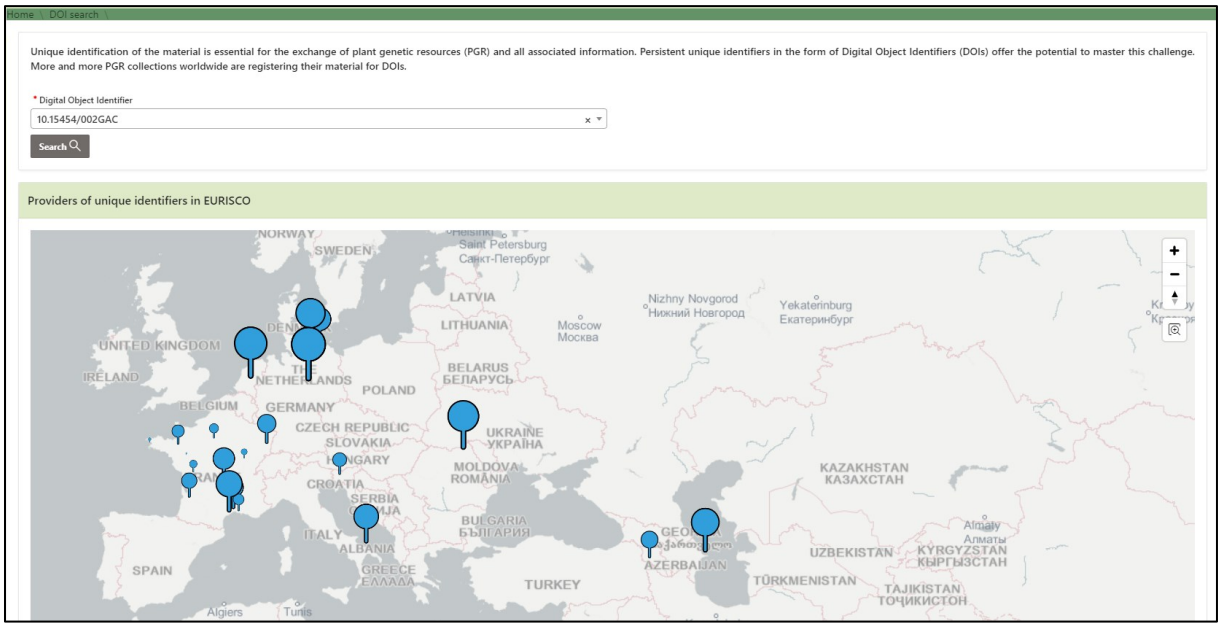

Figure 4: Search mechanism for DOIs.

Furthermore, various options were examined for offering additional searches for passport data at genus level. Although this was possible in the past, this option was deactivated with the introduction of the taxonomic synonym search because the species information is also essential for meaningful results. Following several user requests, an attempt has now been made to offer a pure genus-based search again, which nevertheless enables synonym hits. However, the results were not satisfactory, as the EURISCO coordination believes that they are not user-friendly enough. This option will therefore not be used until further notice.

1.2.2 Extension of EURISCO frontend for *in situ* CWR passport data

In order to make the data of the first *in situ* CWR populations (see section 2) imported into EURISCO searchable, work has begun on expanding the public web application. It is possible to search for all

populations of an *in situ* National Inventory, for specific taxonomic terms or a combination thereof (Figure 5). Additional filters can be applied in a subsequent step using a faceted search.

Figure 5: Search mask to filter for *in situ* CWR populations.

In addition, all previous features that search the entire EURISCO dataset, which includes the *in situ* CWR populations, can also be used. For this purpose, an additional filter has been implemented for each of the existing reports, which allows the number of hits to be limited to *in situ* material (Figure 6). The accession details pages in EURISCO have also been extended to display the additional information (Figure 7).

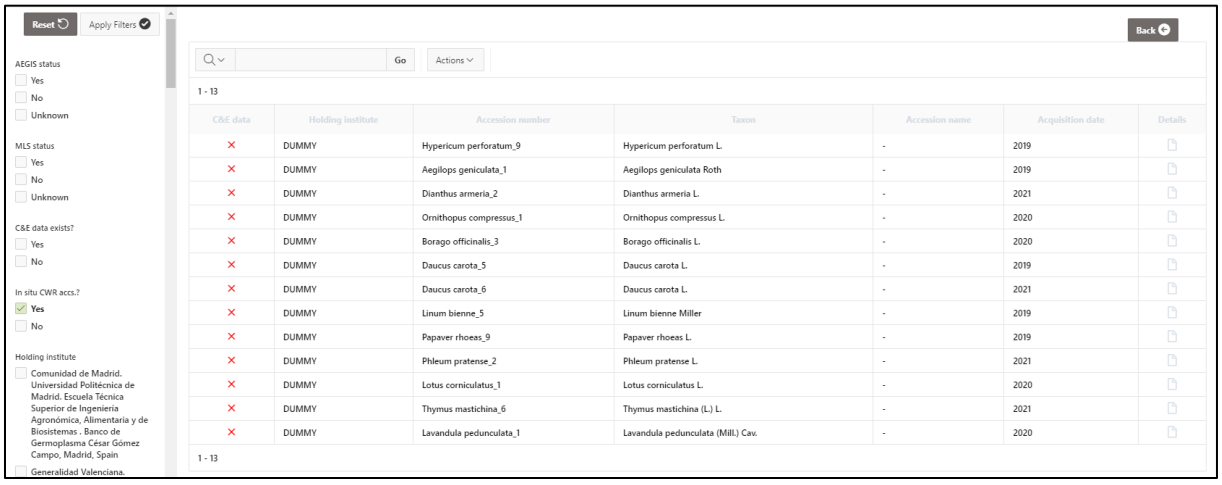

Figure 6: Filter limiting search results to *in situ* CWR material only.

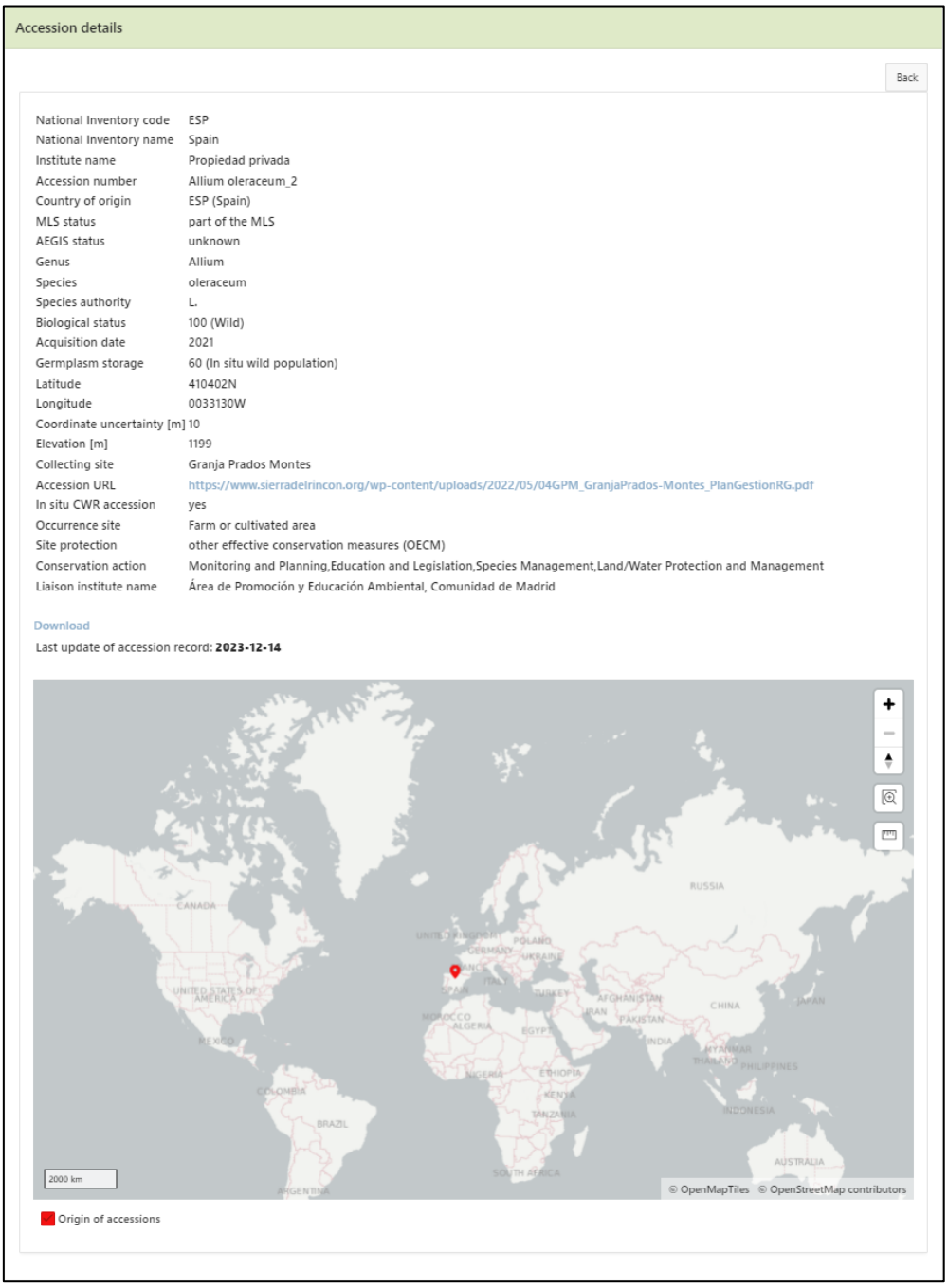

Figure 7: Accession details page showing data on an *in situ* CWR population.

Further search and visualisation options for *in situ* CWR populations will be realised in the course of 2024. A separate export functionality that only includes *in situ* CWR data does not yet exist; it will also be implemented in the course of 2024.

1.3 Continued hosting of the EVA infrastructure

The intranet infrastructure of the European Evaluation Network (EVA) continued to be hosted after the project funding ended in order to enable the various EVA consortia to import their data. The EVA infrastructure will continue to be hosted in the future and adapted if required.

1.4 Documentation and planning of the next steps

The developments described in sections 0–1.3 were specified and all developments were documented.

2 EURISCO coordination

As in previous years, in addition to the actual network activities, an important focus of work consisted of collaboration with various projects, especially under the EU's Horizon programmes.

2.1 EURISCO network maintenance and coordination of the EURISCO development

The contact with (potential) EURISCO stakeholders was also intensified as a continuous activity in 2023 in order to demonstrate the potential of this joint European approach.

At the BOLD WP1 workshop, which was jointly organised by CropTrust and IPK, three lectures were given on documentation and data management, which also covered aspects of EURISCO (22–24 May 2023, Gatersleben, Germany).

A presentation on the progress of EURISCO during ECPGR Phase X and prospects for Phase XI was given during the 17th Meeting of the ECPGR Steering Committee, 30 May–1 June 2023, Oeiras, Portugal.

A presentation on the importance of information management for genetic resources conservation in genebanks was given at the GPZ workshop "Activating plant genetic resources in the crosshairs of plant breeding", 5–6 July, Gatersleben, Germany.

A talk on the FAIRness of European PGR information was given at the "International workshop on the sustainable management of plant genetic resources", 3–4 October 2023, Chania, Greece. (Talk given by C.H. Aguilar)

A presentation on phenotypic data in EURISCO was given at the kick-off meeting of the project "Improvement of Fruit Tree Data Inclusion in EURISCO (FRUITTREEDATA)", 8 December 2023 (online).

A presentation on *in situ* CWR populations in EURISCO was given at the final meeting of the project "Extension of EURISCO for Crop Wild Relatives (CWR) *in situ* data and preparation of pilot countries' data sets", 12 December 2023 (online).

An article about EURISCO was published in the 2023 edition of the renowned database issue of *Nucleic Acids Research* (Kotni et al., 2023). This special issue is published once a year with articles on the most important databases in the field of life sciences.

In addition, together with colleagues from the Global Crop Diversity Trust and the James Hutton Institute, among others, a book chapter on database solutions for genebanks and germplasm collections was written, in which EURISCO was also discussed (Shaw et al., 2023).

The conservation of the rapeseed gene pool in European genebanks was analysed on the basis of data from EURISCO and other sources. In addition to a gap analysis of accessions of the various species, this also included the use of a niche modelling approach to investigate how the natural distribution ranges of these species are expected to change by the end of the century, assuming different climate change scenarios (Weise et al., 2023).

The EURISCO newsletter was sent around in December 2023. This is considered very important for providing feedback to the EURISCO users.

In the course of 2023, 36 production updates of passport data from various National Inventories were performed, either partially or completely. Overall, this involved updating passport data for 535,225 accessions. The total number of accessions documented in EURISCO increased by 10,312 to 2,092,387.

In 2023, passport data on *in situ* CWR populations were integrated for the first time. They come from four countries that are part of the pilot group as part of the project "Extension of EURISCO for Crop Wild Relatives (CWR) *in situ* data and preparation of pilot countries' data sets". Data on 20 populations have been imported for Bulgaria, 66 for Germany, 1,912 for the Netherlands and 27 for Spain.

The number of AEGIS accessions increased only minimally last year to a total of 70,428, while the number of phenotypic records rose by 13,181 to a total of 2,729,780. In total, phenotypic data is available for 91,449 accessions from 21 countries.

This year's EURISCO training workshop for National Inventory Focal Points took place from 12–14 September 2023 in Plovdiv, Bulgaria. It was the first in-person training since the COVID pandemic and brought together 16 Focal Points who were briefed on the latest developments. The main emphasis of the training was put on the submission of passport and phenotypic data to EURISCO. The major focus in this context was on the preparation and uploading of data, in particular the quality and completeness of the data, but also on the data templates to be used and the uploading procedure. Practical exercises on uploading and testing data sets of the individual participants were carried out. In addition, the use of Digital Object Identifiers (DOIs) as unique and stable identifiers for plant genetic resources and their support by EURISCO was discussed. The report of the training workshop is available from the ECPGR website. User feedback is essential to continuously develop and improve a system like EURISCO. This is particularly important to find software bugs, improve existing features and add new ones, as well as to generally increase the awareness of EURISCO and to open up new user groups. As the workshop participants are power users who have a good understanding of both the data provision side and the usage side of EURISCO, a survey was designed in advance of the workshop and conducted during the workshop. It consisted of two parts with ten questions each: 1) infrastructure for data providers (EURISCO intranet), and 2) public EURISCO web application. The system was mostly evaluated positively, but some suggestions for improvement and feature requests were also made. A detailed evaluation of the results of the survey will be presented to the EURISCO Advisory Committee (AC). Suggestions for improvements/functional enhancements will be reviewed for feasibility and included in future work plans in consultation with the AC.

Much effort was invested into providing a helpdesk "behind the scenes". Direct, personal communication took place with National Inventory Focal Points and National Coordinators (e.g. support for updates, provision of specific database queries and special data export formats, discussion about future developments).

2.2 Participation in project proposals

An indispensable task of the EURISCO coordination is to acquire additional funding for the future development of EURISCO. Fortunately, some of the corresponding project applications were approved.

Also in 2023, the EURISCO coordination participated in the ongoing Horizon 2020 project "Activated GEnebank NeTwork" (AGENT), which already started in May 2020. AGENT focuses primarily on wheat and barley and aims to facilitate access to genetic resources for breeders and farmers through standardised protocols for data generation, documentation and provision to users. In this context, the EURISCO coordination is mainly involved in two work packages aimed at developing guidelines and formats for data production, exchange and presentation, and developing the infrastructure for managing and analysing genotypic and phenotypic data on genetic resources. The project thus also represents a sandbox in which possible future extensions to EURISCO can be tested.

In 2023, work has started on the project "Promoting a Plant Genetic Resource Community for Europe" (PRO-GRACE), which is funded under the Horizon Europe programme. On the one hand, it is essential for users of plant genetic resources to have as detailed information as possible in order to select and access the material they need for their research and breeding activities. On the other hand, this is hampered by the fact that data in existing information systems is often incomplete and that large amounts of data have been generated in various projects and need to be integrated. PRO-GRACE aims to close this gap by developing the concept and carrying out proof-of-concept activities for the creation of a large European research infrastructure (GRACE-RI) for PGR. Within the project, the EURISCO coordination is responsible for a work package focussing on an inventory of PGR-related systems and information standards.

The involvement of the EURISCO coordination in the project ‘Extension of EURISCO for Crop Wild Relatives (CWR) *in situ* data and preparation of pilot countries’ data sets’ has already been mentioned above.

The EURISCO coordination continued and continues to actively participate in preparing further project proposals to acquire additional funding for developing certain aspects of EURISCO.