European Cooperative Programme for Plant Genetic Resources
ECP/GR

Report of a Working Group on Documentation and Information

Ad hoc Workshop, 18-19 September 2024, Tallinn, Estonia

S. Weise, K. Annamaa and L. Maggioni

The **European Cooperative Programme for Plant Genetic Resources (ECPGR)** is a collaborative programme among most European countries aimed at contributing to rationally and effectively conserving *ex situ* and *in situ* plant genetic resources for food and agriculture, providing access and increasing utilization (http://www.ecpgr.cgiar.org). The Programme, which is entirely financed by the member countries, is overseen by a Steering Committee composed of National Coordinators nominated by the participating countries. The Coordinating Secretariat is hosted by The Alliance of Bioversity International and CIAT. The Programme operates through Working Groups composed of pools of experts nominated by the National Coordinators. The ECPGR Working Groups deal with either crops or general themes related to plant genetic resources (documentation and information, *in situ* and on-farm conservation and cryopreservation). Members of the Working Groups carry out activities based on specific ECPGR objectives, using ECPGR funds and/or their own resources.

The geographical designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of The Alliance concerning the legal status of any country, territory, city or area or its authorities, or concerning the delimitation of its frontiers or boundaries.

Mention of a proprietary name does not constitute endorsement of the product and is given only for information.

Citation

Weise, S., Annamaa, K. and Maggioni, L. 2025. Report of a Working Group on Documentation and Information. Ad hoc Workshop, 18-19 September 2024, Tallinn, Estonia. European Cooperative Programme for Plant Genetic Resources, Rome, Italy

Cover illustration

Participants in the Ad hoc Workshop of the ECPGR Working Group on Documentation and Information, 18-19 September 2024, Tallinn, Estonia.

© ECPGR 2025

CONTENTS

RELATED PRESENTATIONS ARE AVAILABLE ONLINE

HERE

SUMMARY REPORT OF THE MEETING

Introduction

An *ad hoc* workshop of the Working Group (WG) on Documentation and Information (Doc&Info) of the European Cooperative Programme for Plant Genetic Resources (ECPGR) was held on 18-19 September 2024 at the Kalev SPA Hotel in Tallinn, Estonia. The meeting was organized in collaboration with the Centre of Estonian Rural Research and Knowledge (METK).

The meeting was intended to 1) reactivate the Doc&Info WG, 2) revise the general direction/objectives of the WG and 3) develop a work plan for ECPGR Phase XI. The latter directly supports the achievement of the ECPGR objectives regarding the documentation of plant genetic resources for food and agriculture (PGRFA). Based on the relevant ECPGR objectives and targets for Phase XI, the following targets needed to be discussed at the meeting:

- Coherent, comprehensive, coordinated and centralized documentation of CWR and WFP
- European inventory of on-farm genetic resources
- Centralized ordering system for AEGIS material via the European Search Catalogue for Plant Genetic Resources (EURISCO)
- Strengthening comprehensive information system for PGR (documentation in general)
- Promoting sustainable use of PGR

In preparation, the following expectations and considerations were compiled and submitted to the ECPGR Executive Committee for approval of the meeting:

- With regard to the first point, the form in which the EURISCO module for *in situ* crop wild relatives (CWR) data should be expanded needs to be discussed. While the data standard, infrastructure and upload/check procedures are already in place, the public web application only offers limited search and filter options. Their continuous expansion should be discussed. With regard to the goal of expanding EURISCO to wild food plants (WFP) and wild harvested plants (WHP), a general discussion must be held. Should data on WFP be treated as a subset of the *in situ* CWR data or separately? In the latter case, a strategy needs to be developed to establish the necessary network of data providers, a separate data standard and the appropriate technical infrastructure.
- With regard to the second point, the meeting offers a good platform to discuss the usefulness and feasibility of an on-farm extension. In particular, it is important to assess whether there is sufficient interest on the side of on-farm maintainers and whether sufficient data is available. It is also important to agree on how often data should be submitted/updated in order to achieve a comprehensive and up-to-date inventory. In addition, minimum sets of passport descriptors and minimum sets of characterization descriptors must be defined and the technical parameters for implementation must be agreed.

- With regard to the third point, the advantages and disadvantages of a centralized ordering system must be discussed, particularly with regard to how the necessary requirements can be met. In addition, a protocol must be defined and an application programming interface (API) must be implemented on the EURISCO side. The APIs required on the side of the respective genebanks represent the greatest challenge. Ways of supporting the respective responsible parties must be explored.
- In addition, the meeting intends to discuss general topics related to documentation, including the support of data providers, the availability of phenotypic data in EURISCO, increasing the compliance of data with findable, accessible, interoperable, and reusabe (FAIR) principles and the development of EURISCO towards a trustable repository. A discussion is also planned about what support can be provided alongside EURISCO to make data and accessions available to users. This concerns in particular the support of crop portals that can obtain parts of their data automatically from EURISCO.
- Other points that have relevance for the work of the Doc&Info WG include further development of standards and exchange formats. There are a number of unresolved issues with the EURISCO-MCPD[1] standard in particular. In addition, the current revision of the exchange format for phenotypic data should be agreed. Furthermore, opportunities must be sought to strengthen lobbying for the use of DOIs, to implement established interfaces such as Breeding API (BrAPI) and to address linking with other sources, particularly with genetic information.

Due to the size of the Doc&Info WG with over 80 members, it was necessary to limit the number of participants to a smaller group. Participants were selected to bring together the most active members of the Working Group to ensure a productive meeting. In addition, the expertise of the participants was considered in relation to the items on the agenda.

A one-day meeting of the EURISCO Advisory Committee (AC) after the Doc&Info WG meeting was also planned to increase the synergies and effectiveness of both meetings.

Opening of the meeting

Stephan Weise, Chair of the WG, welcomed the meeting participants gathered from nine countries, as well as the ECPGR and FAO Treaty Secretariats' representatives. He noted that the previous meeting of this type was held more than ten years ago, at that time hosted at the Crop Research Institute in Prague, Czech Republic, by the late and missed friend Iva Faberová.

Marko Kass, Research Director of METK, welcomed the group to the medieval city centre of the Estonian capital Tallinn and wished success for the meeting. He indicated that plant breeding has been practised here for more than 100 years, while the genebank has existed for 25 years. METK is devoted to three pillar activities: 1) research related to plants, carried out in Jõgeva by 150 staff members, 2) variety testing in three locations and lab services, performed at a recently renovated site close to Tallinn, and including soil analysis and feed analysis, and

[1] Multi-crop passport descriptors

3) innovation and entrepreneurship. Estonia is quite diverse in terms of soil type, sandy in the south and clay in the north. The climate is also different, with seasons being two weeks earlier in the south. The challenge is the creation of new varieties, especially cereals and potatoes, that are suitable for the different locations, also facing new conditions of spring drought and heavy rains.

Külli Annamaa, Estonian National Coordinator and vice-Chair of the WG, also welcomed the participants and told a few words about the local hero Kalev (the oldest potato accession still preserved in the genebank of 1934 is named after him), the history of the Oleviste church of the 13th century, the tallest church in the Baltic countries, and the geography of the highest mountain of the Baltic countries, reaching 318 m!

Participants presented themselves with a short round of introductions.

Coherent, comprehensive and centralized documentation of *in situ* CWR and WFP conservation

Nigel Maxted, University of Birmingham, UK

N. Maxted, Chair of the Crop Wild Relatives WG, summarized the objectives of the WG, namely to effectively and sustainably conserve the breadth of wild PGRFA (CWR and WFP) *in situ* (in nature or on farm), with complementary *ex situ* activities, and to provide and increase access to the conserved resources for crop improvement, research and other uses. The principles of conservation and use were also reaffirmed, including the need to maximize diversity for the long term, actively using complementary techniques and ensuring that the resources remain available for utilization. Documentation is a key factor, since the better it is gathered, stored and displayed, the more material can become available for diverse user communities. The link between *in situ* conservation and utilization was particularly emphasized because if this link is weak the justification for *in situ* action is less critical and sustainable.

For *in situ* conservation, several types of data are essential for conservation planning (prioritization-related and ecogeographic data), and for conservation implementation (data enabling reserve design and management implementation and monitoring). Additionally, resource description (characterization and evaluation) and promotion of utilization data are necessary. Further data need to be collected at taxa and population levels and stored at national and/or European levels. Practically given that the *in situ* population managers are unlikely to have specific PGR skills, this will involve collaboration between Genetic Resource Centres (GRC) and PGR population maintainers, who will together focus on *in situ* data collection, management and sharing at different levels. A proposal was made for data responsibilities with the *in situ* process (see Table 1) and discussion followed.

Table 1. Proposed *in situ* population data management collaboration. LR, landraces

CWR, WHS or LR *in situ* population conservation	
National GRC role	**PGR population maintainer's role**
International, national and local policy development	Preparation, implementation and periodic revision of site management plan
National conservation planning	Management of target populations
Target population national network management	Monitoring of target populations
Target population characterization and evaluation	Periodic collection of target populations to make representative *ex situ* backup samples, for backup, characterization and evaluation (C&E), and user access
Ensuring user access to *in situ* conserved resources (via the *ex situ* backup sample)	Promotion of PGR integration into the broader biodiversity community
Lead and participate in PGR *In Situ* Management Committee	Participate in PGR *In Situ* Management Committee

Coordinating *in situ* actions through a network structure offers several advantages. Additionally, the networking activity is associated with specific sets of data to be recorded (related to coordinating activities, communities of practice, ABS regimes, etc.).

The correct destination for all the relevant data at the different levels needs to be well-defined.

The ongoing German-funded project working on the extension of EURISCO for CWR *in situ* data is focusing on the 'use potential' of the *in situ* populations. Sets of *in situ* population data to be gathered in EURISCO have been defined through a list of EURISCO-specific descriptors, and pilot countries have already started to populate EURISCO. As a result of the discussion, the original list was amended and the final set of descriptors for uploading *in situ* CWR passport data to EURISCO is available from the ECPGR website (here).

The PRO-GRACE project (https://www.grace-ri.eu/pro-grace) is developing *in situ* ontologies and descriptors 'gap analysis', with the objective to identify gaps in the documentation instruments, such as missing *in situ* ontologies and descriptors needed to facilitate the *in situ* process. The deliverable will fill the gaps and produce an integrated guide to PGR *in situ* data management, although the implementation of these more population management-based descriptors will then require further discussion between the Doc&Info and CWR WGs.

A further item requiring clarification is the documentation of WFP, which are wild species differing from CWR in terms of their use, which is direct consumption rather than indirect use through breeding. The assumption that will need to be tested is that there is no logical reason to apply different conservation techniques, documentation tools and procedures for WFP than those already used for CWR conservation techniques and documentation.

Discussion

The meeting took stock of the developments made by the German-funded project on *in situ* CWR population data in EURISCO. This project was based on a number of principles that were eventually validated and are recommended for practice by all European countries. In particular, only data related to germplasm, that could potentially be made available to users should be entered in EURISCO. Datasets included in EURISCO should refer to 'actively conserved' populations, based on the criteria indicated in Box 2, page 7 of Hintum and Iriondo (2022)[2].

It is assumed that each country will define its specific criteria for inclusion of *in situ* CWR data into EURISCO. It is expected that by common sense, datasets would not include all the possible existing observations, but a subset that the country will prioritize as sharable material. Examples of data sets and criteria developed by the first pilot countries can serve as guidance for additional countries. For example, the solution envisaged by The Netherlands involves inclusion in EURISCO of 1,912 records, of which 298 are prioritized populations based on niche modelling and climate change scenarios. Other 1,614 are diffused populations that are common in the Netherlands. Each record is a combination of the species and the flora district in which it is present.

The Doc&Info WG and the ECPGR Secretariat were encouraged to make available a compilation of lessons learned in the preparation of *in situ* CWR National Inventories and to invite contributions from all countries to the *in situ* CWR dataset in EURISCO.

It is expected that the systematic preparation of CWR National Inventories by all countries is encouraged and coordinated by the CWR WG, based on existing guidelines, recently summarized in Hintum and Iriondo (2022). A possible task for the CWR WG would be to set up a link on the ECPGR website, connecting to all the useful information and tools to help countries develop *in situ* National Inventories, including a link to existing checklists of CWR per country. Such a task could be elaborated and supported as part of an ECPGR Grant Scheme activity.

The difference between CWR, WFP and WHF was noted, indicating that it mainly rests in the different use, keeping in mind possible overlaps of these domains. As far as documentation of wild material in EURISCO is concerned, it didn't seem relevant to prepare specific descriptors to distinguish the different uses of CWR, WFP and WHF.

On the other hand, the use of different descriptors could be evaluated to verify whether their current elaboration is still valid and comprehensible (i.e. wild, wild/natural, wild/sown). This task might be a topic for discussion in a dedicated webinar (see below).

[2] Hintum T, Iriondo J (2022). Principles for the Inclusion of CWR Data in EURISCO

European inventory of on-farm genetic diversity

Parthenopi Ralli, Hellenic Agricultural Organization - DIMITRA Institute of Plant Breeding and Genetic Resources Thermi, Thessaloniki, Greece

EURISCO provides passport information about more than 2 million *ex situ* accessions in European genebanks and increasingly also phenotypic data. In addition, the EURISCO infrastructure has been recently extended to receive *in situ* CWR population data. It should be explored whether an extension for landraces on-farm data could also be prepared.

So far, limited information on on-farm/in-garden conserved landraces has been made available (Veteläinen et al. 2009; 2012)[3] through activities carried out during EC-funded projects:

• 'An Integrated European In Situ Management Work Plan: Implementing Genetic Reserves and On-Farm Concepts' (AEGRO)

• 'Novel characterization of crop wild relative and landrace resources as a basis for improved crop breeding' (http://vnr.unipg.it/PGRSecure/), and 'Descriptors for web-enabled national *in situ* landrace inventories' were developed for recording such data, ('PGR-SECURE' project),

https://www.pgrsecure.bham.ac.uk/sites/default/files/documents/helpdesk/LRDESCRIPTORS_PGRSECURE.pdf

The 'Farmer's Pride project' of the Horizon 2020 EU Programme aimed at gaining a detailed view of landraces still maintained on farms/in gardens in Europe, since no conservation and promotion of use can be carried out without knowing where landraces are, which species they belong to, why and how they are still maintained. Initially, project activities focused on the creation of a European inventory of on-farm conserved landraces and the collection of detailed information on landrace case studies across Europe. A subset of the 'fields' listed in the web-enabled template developed for collecting anonymous data on on-farm conserved landraces was produced. Only information on the country inventory, the taxon and the landrace (name, location and area) were asked. The idea of using only a subset of the total available fields was to maximize the number of possible answers (i.e. the number of recorded on-farm landraces) while reducing the time needed by the respondents to fill in the information.

As a deliverable of the Farmer's Pride project, the largest ever-produced database of *in situ* maintained landraces was created (Raggi et al., 2022[4]). It has a total of 19,335 records, including forages, cereals, pulses, garden crops and fruit trees from 14 European countries.

---

[3] Veteläinen M, Negri V, Maxted, N (2009). European landraces: on-farm conservation, management and use. Biodivers. Tech. Bull. 15.

Veteläinen M, Negri V, Maxted N (2012). A second look at the European strategic approach to conserving crop landraces. 10.1079/9781845938512.0181.

[4] Raggi L, Pacicco L, Caproni L, et al (2022). Analysis of landrace cultivation in Europe: A means to support in situ conservation of crop diversity. Biological Conservation 267:109460

Additionally, another product of the Farmer's Pride project is the *In situ* landraces best practice evidence-based database, which provides access to evidence-based information on the benefits, opportunities and practices of landrace cultivation to help in decision-making and to promote their *in situ* maintenance as a means of conserving and diversifying PGR for food, nutrition and livelihood security. This includes examples of *in situ* management practices and of adding value to landraces for different crops and socio-cultural, environmental and economic contexts (Raggi et al., 2021)[5].

Another activity dedicated to creating on-farm landrace inventories is the ECPGR Grant Scheme Activity 'Inwheatory' (Inventorying wheat on-farm diversity). In this activity, partners agreed to use appropriate templates to collect *in situ* occurrences data of wheat landraces and case studies of successful examples of wheat landrace cultivation and use.

To complete the picture of on-farm documentation in Europe, the EU catalogue of registered conservation varieties should be mentioned. Since the implementation of the European Commission Directives 2008/62/EC, 2010/60/EC and 2009/145/EC, several countries have registered various landraces as conservation varieties (more than 400 from agricultural plant species and almost 200 from vegetable species varieties). Recent reviews on achievements on landraces *in situ* (on-farm) conservation in Europe and on landrace legislation in the world with emphasis on the EU system are available in Raggi et al. 2024[6] and Thanopoulos et al. 2024[7].

Despite the work that has been done already and the information gathered, a comprehensive European inventory of *in situ* maintained landraces is still lacking. However, landraces are still widely grown in different European countries and biogeographic regions and inventories are required because, without knowing the extant, it is rather difficult for governments to properly plan and implement their systematic conservation and use. In addition, countries that ratified the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) are required to "survey and inventory PGRFA" (Art. 5.a) (FAO, 2001).

EURISCO is suggested to help in keeping track of landraces and cultivation sites, also enabling to matching these with the landraces conserved *ex situ*. Similarly to the actions implemented for the inclusion of *in situ* CWR data, EURISCO could be extended to enable public display of passport data of European landraces conserved on farm. Initial sets of data could be prepared by a few pilot countries, offering examples for all other countries to follow. The descriptors for on-farm inventories and the results produced in the framework of other projects could help as a starting point for data exchange, and/or a new agreement for the type and requirements of populations to be inventoried should be made. The link of the on-farm data with the existing

[5] Raggi L, Caproni L, Negri V (2021). Landrace added value and accessibility in Europe: what a collection of case studies tells us. Biodiversity and conservation 30:1031-1048.

[6] Raggi L, Spataro G, Negri V (2024). Landrace in situ (on-farm) conservation: European Union achievements. Biodiversity and Conservation 33(10):2709-2738.

[7] Thanopoulos R, Negri V, Pinheiro de Carvalho MA, Petrova S, Chatzigeorgiou T, Terzopoulos P, Ralli P, Suso MJ, Bebeli PJ (2024). Landrace legislation in the world: Status and perspectives with emphasis in EU system. Genetic Resources and Crop Evolution 71(3):957-997.

ex situ data in EURISCO will improve the *ex situ*/on-farm conservation interface.

Discussion

The establishment of a European inventory of on-farm material was discussed. The low sustainability and very complex compilation and maintenance of an inventory of germplasm managed on-farm was reiterated, also considering the absence of a strong and stable network of maintainers that would be interested in providing the data. Also, the value for the user was questioned, considering that access to such material could be guaranteed only through the deposit of samples into an *ex situ* genebank. On the other hand, knowledge of the existing diversity on farm, with the approximate geographic distribution, size area of cultivation, and the most valuable traits of landraces and other heterogeneous material, would be useful, especially at national level, to monitor, promote and manage on-farm diversity, including through incentives.

It was suggested before trying to prepare a full inventory of all European landrace populations, that the priority action of the On-farm WG should be to create a European inventory of landraces, made of nationally prepared datasets, where the information units would be the landrace and genus plus species names. The amount of information to be associated with these names could be defined by a list of proposed descriptors for national landrace inventories. Each country would select and work on its own list of landraces, based on local priorities. The European list of 'names' could be linked to EURISCO, with each name being associated through a 'hot link' to the related accessions conserved in European genebanks. This virtual link could serve to monitor the level of *ex situ* conservation of on-farm material, as well as for gap analysis of what still needs to be secured into a genebank. The Doc&Info WG made itself available to offer advice for the compilation of appropriate descriptors for on-farm material, as well as to propose the technical mechanisms to implement the link between the European landrace inventory and EURISCO.

A centralized ordering system for AEGIS material via EURISCO

Discussion introduced by S. Weise and M. Oppermann, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) – Gatersleben, Germany

The usefulness of introducing an ordering system was reiterated to enable the user to directly create an 'ordering basket' while browsing EURISCO. The MCPDs do not include a descriptor related to the availability of the material. This was never introduced upon consideration that maintaining it up to date would probably be unrealistic for many National Inventories. The risk of placing many orders that would not be successfully processed would be very high. On the other hand, a descriptor that indirectly indicates 'availability' does exist and it is the AEGIS[8] descriptor. In fact, all the accessions that are part of AEGIS should by default be available under the Standard Material Transfer Agreement (SMTA). It was therefore suggested to start implementing the ordering system for the AEGIS accessions, which could be associated

[8] A European Genebank Integrated System, https://www.ecpgr.org/aegis

with an ordering button. In addition, only those accessions for which the maintaining genebanks agree to receive orders via EURISCO should be available for ordering.

It was confirmed that it could be possible for EURISCO to simply interconnect the requestor with the recipient genebanks, without storing any data related to the requestor or the ordered accessions. This would require providing a specific application programming interface (API) to the genebank information systems and thus redirecting specific orders, which would be received by the genebank in exactly the same way as any other order submitted from outside of EURISCO, and enter the regular ordering system (possibly including a click-wrap SMTA approval etc.). This procedure would only be suitable for those genebanks that are endowed with an autonomous ordering system, or that are ready to invest to establish one.

For the case of genebanks that are not organized with an information system that could install the API, EURISCO could generate an email that would be sent to the suitable address where the genebank wishes to receive this type of orders. The latter mechanism would have the drawback that EURISCO should maintain an updated list of emails of genebanks and this requires manual maintenance and is prone to errors. Also in this case, the genebank would need to implement a specification in its own protocol.

It was suggested that the EURISCO team prepare a plan on how to implement the system, starting with the involvement of those genebanks that already have an ordering system in place, while the possibility of generating an email will be considered later on. This proposal should also be discussed during the EURISCO Advisory Committee meeting on 19 September.

Strengthening a comprehensive information system for PGR

R. Kowalik, National Centre for Plant Genetic Resources Plant Breeding and Acclimatization Institute – National Research Institute (IHAR), Radzików, Błonie, Poland

It was reminded that the framework in which the WG should operate in the next few years should refer to the targets to be achieved by 2030 according to the ECPGR Strategy for Europe:

2030 Targets for strengthening a comprehensive PGR information system:

1. EURISCO network of NFPs[9] is optimally supported.
2. High-quality of passport data of all European *ex situ* collections + progressively extended of *in situ* CWR populations and appropriate on-farm landraces data.
3. NFPs assure access to all publicly available quality phenotypic data related to the conserved PGR. Access provided via inclusion in EURISCO.
4. Genebanks and other PGR holders can improve their data management system practices (access, use of tools, resources, services, adopting FAIR principles, part of the open data community).

[9] National Focal Points

5. Data in EURISCO and associated IT infrastructure are compliant with FAIR principles.
6. EURISCO becomes a trustable repository in the arena of European and global open-access databases with high standards.

The following considerations can be made regarding the 2030 targets:

1. Support of the NFPs requires regular training, which is very useful for new users and new staff (documentation specialists, curators in genebanks, new members of ECPGR WGs, etc.). The possibilities range between in-person, online and hybrid meetings. Feedback received from trainees during and after these meetings is often very relevant to orienting necessary improvements.
2. High-quality passport data of *ex situ* collections need measurement of data quality, that is their completeness, correctness and reliability, amongst others. Completeness can be measured with the Passport Data Completeness Index (PDCI) (Hintum et al. 2011)[10], which is an index ranging between 0 (low quality) and 10 (high quality) and can be applied to the 45 *ex situ* and 28 *in situ* passport descriptors. A tool to help complete the data could be a template with descriptors of different categories highlighted in different colours (red for mandatory descriptors, yellow for collected accessions, light grey for bred accessions, light green for those acquired from other institutions, etc.). The most common mistakes reducing correctness are the use of wrong country codes (scroll-down validation tools might help), the use of improperly converted coordinates or switching the order of LAT and LON. Reliability can be improved when data acquired from other genebanks perfectly match with the corresponding information published by the respective data donor.
3. To assure access to all publicly available quality phenotypic data related to the conserved PGR, the NFPs need to be well interconnected with curators and PGR holders and actively promote the data flow, making use of the data exchange standard for C&E data. Regarding the quality of phenotypic data, many curators do not know which descriptors to use and why. There are also often ambiguous states that would need explanation.
4. To improve data management practices, training workshops should be organized at genebank level or national level.
5. The FAIR principles should always be kept in mind by the information specialists, ensuring that the data are easy to Find, and to get Access to, Interoperable (data format needs to interoperate with applications for analysis) and Reusable (well described, can be replicated/combined in different settings).

[10] Hintum T et al. 2011, Quality indicators for passport data in *ex situ* genebanks, Plant Genetic Resources, 9(3):478-485

Discussion

Trainings of NFPs are currently taking place every two years physically and occasionally online on a bilateral basis. Additional training may be regularly organized at the national level. Moreover, regular webinars could be planned and training materials prepared, such as tutorial videos.

A few topics for webinars or online discussion meetings were proposed, either with the purpose of training or to disseminate the value of the data analyzed from the genebanks:

- Discussion by the Doc&Info WG on the MCPDs (including issues such as the difference between **country of origin** and **country of provenance,** or addressing the meaning and common understanding of the states listed for 'wild material')
- The procedure for the selection of AEGIS material could be discussed from the documentation point of view, inviting responsible curators or documentation specialists of the AEGIS Associate Member Institutions and NFPs
- Uncertainties about the use of descriptors could receive orientation regarding the choice of descriptors, the consequences of using different types of scoring and the need to make sure that the data can be correctly interpreted
- The results of genome-wide association studies (GWAS) applied to genebank data could be presented (e.g. lettuce analyzed at CGN), showing which accessions are predicted to be resistant. This information is now valid for companies with bioinformaticians and should be provided to everyone.

Phenotypic data, FAIR principles, trusted repository

C. H. Aguilar, M. Opperman and S. Weise, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) – Gatersleben, Germany

How to improve the availability of phenotypic data in EURISCO?

The importance of phenotypic data cannot be underestimated, since they add value to germplasm for breeding and research. Provision of these data meets multiple obstacles, considering that several standards are in use to express the traits, as well as a large amount of related metadata and many different means of data management. Data have to be aggregated or exchanged between organizations.

The current approach to data standardization is pragmatic, with no standardization of trait, scale or experimental design and with the import of existing data as-is to reach a critical mass. The only standardization in effect is related to the exchange format, which is designed to be as simple as possible, with as few fields as possible, thereby responding to a 'minimum consensus'. The extension to phenotypic data is available in EURISCO since 2016, and 2,729,636 records of data have been captured from 21 countries, consisting of 74 phenotypic datasets describing 3,919 experiments and 9,764 traits related to 91,443 accessions. Limitations depend on the fact that the EURISCO data exchange format represents just a 'minimum consensus'. It is also difficult to compile the files manually, and the resulting reproducibility

and comparability are very limited. It is proposed here to simplify the data collection, introducing one column per trait to support manual recording. Additional metadata for the experiment, trait and range of values are also proposed (see presentation for details).

How to increase the compliance of data with FAIR principles?

For the data to be **Findable**, it is important to have unique and persistent identifiers and rich accompanying metadata with clear data–metadata links, which make the data better searchable. To overcome the existing limitations, which mainly relate to insufficient standardization and data quality, it is recommended to increase the adoption of unique identifiers. FAO recommends these to be in the form of DOIs. There is great potential for improvement, considering that they are currently adopted for approximately only 20% of the accessions. The application of DOIs can also be extended to phenotypic datasets.

Data are **Accessible** if they are retrievable using a protocol, the protocol is open and implemented. Metadata need to be persistently accessible. Current limitations of access to passport data depend on access restrictions or technical limitations (e.g. obsolete platforms). Phenotypic data are affected by the use of many different storage systems and technical infrastructures, as well as discrepancies in depth and granularity, and use restrictions. Genotypic data are hosted in specialized infrastructures for vast data sets, with dubious sustainability and requiring advanced analytical tools. Image data limitations relate to inefficient or proprietary compression algorithms affecting data retrieval speed and quality and to ethical and regulatory complexities. The path to improvement involves creating aggregators, expanding existing systems and enhancing networking interactions. Trusted repositories, especially for phenotypic data, should further develop and project data should consistently be submitted to public repositories. Stronger cooperation in this regard is required among genebanks.

Data are **Interoperable** when they link to other (meta) data, using consistent vocabularies, formal, accessible, shared and applicable language. In this case, limitations are due to inconsistent adoption of the MCPD standard, which is an obstacle to semantic interoperability. Phenotypic data are limited by the diversity of trait measurements and terminologies, format and standard incompatibility and the lack of proper data management and exchange protocols. Similarly, genotypic data suffer from diverse data formats and standards and a lack of data integration across platforms. Image data often have inconsistent resolution, format, scale and annotations, and there is a lack of universally adopted ontologies for PGR image data. Improvements require increased standardization/harmonization of terminologies and data standards. However, approaches to semantic standardization are still underdeveloped and ineffective, especially for phenotypic data. Additionally, ontologies used for PGR do not work properly.

Data are **Reusable** when there are clear and accessible data usage licences, the provenance is well-detailed and the data descriptions are rich, with accurate and relevant attributes.

Current limitations depend on data legacy issues, and insufficient metadata reducing the potential for reuse. Phenotypic data require comprehensive metadata and documentation,

good data quality and integrity. Genotypic data require (meta) data quality and completeness. Image data also require metadata completeness, standardization, format compatibility, quality assurance and preservation. The way forward requires the adoption of approaches for better description, for example MIAPPE. This also applies to project or legacy data. Early involvement of data stewards is paramount.

How to develop EURISCO towards a trustable repository with acceptably high governance and data management standards?

The proposed way forward requires developing EURISCO into an integrated European PGR information system (PRO-GRACE and beyond), adding missing sources to the system, connecting additional domains, and promoting standards and protocols. This result can be obtained by remaining committed to project cooperation, spreading the word, raising awareness and expanding cooperation with bioinformatics hubs.

Discussion

Following the detailed presentation of the proposed simplified template for phenotypic data gathering, there was consensus for its adoption. It was specified that it can only partially map to MIAPPE compliant standards, but it improves the level of information compared to the present. We need not drift away from other communities (MIAPPE), but at least explain why we can't use MIAPPE and at least partially map to it. This proposed format also represents a compromise. A large part of the phenotypic data available in genebanks and projects is compiled and curated manually. In addition, metadata is not available to the extent required for full MIAPPE compliance, especially for older but still important datasets. The proposal must take this into account.

Regarding data accessibility, it was reiterated that more phenotypic data need to be deposited in EURISCO rather than in project databases.

To improve interoperability, the idea of using ontologies is good, but the way they are used for PGR does not work properly. Crop WGs need to be more closely involved. Project funds would be necessary for this purpose.

Regarding DOIs, it was clarified that the Global Information System (GLIS) is the place where all DOIs are conserved, even those assigned by other institutions (e.g. IPK). However, INRAE, France does not provide their DOIs to GLIS.

It should be important to enable the visualization of all the information related to each DOI that can be made available to the users and in this way improve the understanding of their value.

Answering a question whether DOIs should be assigned to on-farm material, M. Marsella (ITPGRFA) answered that this is under discussion, but also the conclusions emerging during this meeting regarding documentation of on-farm material do not push in that direction.

It was highlighted that the Doc&Info WG is providing a platform and support to make local data in genebanks FAIR. Governments should be informed that this is happening for their reporting to FAO, since this effort is in line with international recommendations. At the same time, this WG should promote the use of the platform that we are providing to make our

phenotypic data available.

Promoting sustainable use of PGR

Külli Annamaa, Centre of Estonian Rural Research and Knowledge (METK), Jõgeva, Estonia

A short overview was given about targets related to documentation and information under the section 'Promoting sustainable use of PGR' in the PGR Strategy. These mainly relate to making data available to users through open centralized information systems, including crop portals. Setting up crop portals, based on crop-specific WG proposals, was accepted as a secondary priority for ECPGR action by the Steering Committee. A few central crop databases (CCDBs) were mentioned with a comment that (most of) these are no longer being updated. Finally, the example of the Forage Portal was brought up. For discussion, the question was raised whether data on available accessions should be provided mainly via EURISCO or also via crop-specific portals (under the control of the Crop WGs).

Discussion

Crop portals can be designed to automatically receive MCPD passport data and phenotypic data from EURISCO, which are automatically updated. WGs can provide additional information not managed in EURISCO (such as Most Original Samples lists). The Forages WG has created such a crop portal, but there has been no interest so far in adding more content, apart from some bioclimatic data.

The question was discussed whether there is a need for dedicated crop portals. The original idea was to provide information about variety lists, commercial sources or other information directly useful to users. On one hand, this requires a lot of time and dedication, on the other hand, we did not get so much feedback, possibly because commercial breeders don't want to share information.

It was concluded that there is no sufficient interest or capacity to invest in these portals at the moment. The API can still be provided to interested WGs, but they would need to invest time and effort.

EURISCO in the world

Theo van Hintum, Centre for Genetic Resources, the Netherlands (CGN), Wageningen University and Research Centre, Wageningen, The Netherlands

Historically, EURISCO was set up to replace the need for European Central Crop Databases. At the time, technology allowed data pooling and EURISCO was, together with SINGER (CGIAR), the first aggregated PGR information system. Subsequently, SINGER developed into Genesys, which expanded its scope to the entire world, using EURISCO for the European regional data.

At present, Genesys is the most complete, well-maintained and accessible PGR aggregator information system, which now includes CGIAR collections, EURISCO, USDA and various

other (inter-)national genebanks. On the other hand, WIEWS (FAO) was not created to serve users, but is oriented towards compiling the 'FAO State of the Worlds PGR' and monitoring the relevant Sustainable Development Goals. EURISCO is loosely linked to Genesys, to which it supplies its data. The two system representatives serve on each other's advisory committee.

The role of EURISCO can be questioned: should it remain a standalone data aggregator system or mainly serve as a channel to feed Genesys? Is its function to maintain a community of PGR documentation experts in Europe or to serve as the interface for PGR users? The relationship between EURISCO and Genesys should be better defined.

Discussion

It was commented that it makes sense, in the long run, to further integrate EURISCO with Genesys or use the same software. However, it would not be wise to abandon the name EURISCO, since it is a well-established and successful brand. Moreover, while the web interface can be replaced, the level below, which is the aggregated data, including phenotypic data, obtained from the European countries, is not replaceable.

It was also observed that if the EURISCO brand should disappear, possibly many countries would no longer be happy to contribute their data. Additionally, Genesys is run by the Crop Trust, which explicitly focuses on *ex situ* material. This would bring the risk of losing the *in situ* part of EURISCO.

There was general agreement that we still need the EURISCO backend infrastructure to integrate the data, curate and clean them. However, the interface could be technically merged with Genesys, without eliminating the EURISCO branding as a major data supplier. With this in mind, all the possibilities that Genesys could provide to avoid duplication of effort should be explored. This will free up capacity in the medium term to focus more strongly on the network behind EURISCO and the gradual improvement of the quality of the integrated data.

GLIS and DOIs for PGRFA

Marco Marsella, Secretariat of the FAO ITPGRFA

The community agreed on the need to accurately and permanently identify PGRFAs in the increasingly critical global context. A permanent identifier enables keeping track of the relationships among the accessions, independently from changes in their accession numbers, WIEWS codes etc. The ITPGRFA Secretariat was asked to provide guidance on which type of Permanent Unique Identifier (PUID) was best for PGR. The Secretariat conducted a study on available technologies, a survey among over 200 experts worldwide and a further validation study with 23 selected experts. The outcome of the process was the adoption of Digital Object Identifiers (DOIs) by the Scientific Advisory Committee (SAC) on the Global Information System (GLIS). GLIS is a place where all PGR DOIs are centrally stored. DOIs are assigned by GLIS free of charge. They are fully opaque, very compact, guaranteed unique and they never change. Opacity is a critical property of PUIDs and is defined as the impossibility of deriving any property of the object by just looking at the PUID. This is a critical aspect since objects that

are not opaque are subject to pressure for change when the meaning embedded in the identifier is no longer correct.

With DOIs, it is possible to know how the PGRFA was obtained and the DOI(s) of the progenitor(s) allow GLIS to build genealogy graphs.

DOIs are also assigned by GLIS to crop wild relatives conserved *in situ* and partners are invited to provide their datasets for uploading to GLIS.

Currently, a Query API enables access to all DOI information in GLIS using content negotiation for XML, JSON, JSON-LD, DarwinCore Archive and BrAPI v1.3. All partners interested in testing the Query API are welcome. GLIS offers an XML-based integration protocol that allows the registration of new DOIs and the update of metadata of already registered DOIs. EURISCO implemented the protocol in 2019. This is also used by GRIN-Global, GG-CE and Web-SDIS. The ITPGRFA Secretariat is happy to support EURISCO in promoting DOIs to member genebanks. So far, a total of 32,830 DOIs have been registered for Romania (48%), Azerbaijan (26%), Italy (17%), Armenia (6%) and Slovenia (3%).

Some examples of how the DOIs received from EURISCO/GLIS can be used: in publications (listing the DOIs of the materials as bibliographic references); in datasets (entering the DOIs of the materials in the metadata submitted for registration of the dataset); in the SMTAs that are issued (listing the DOIs in Annex 1 to the SMTA).

Discussion

Some clarifications were offered about the assignment of DOIs:

- There is a need to assign a new DOI when a sample is received from another donor (similar to the assignment of a new accession number)
- After regeneration, a new DOI may or may not be assigned. CGN and IPK do not do that (also accession number is not changed). For most genebanks, it should be sufficient to assign DOIs only at accession level
- The principle is to assign a DOI whenever it is necessary to discriminate between two objects
- Samples from *in situ* populations entering a genebank should receive a new DOI
- There is no problem with assigning two DOIs to the same thing, but it is an error to assign the same DOI to two different things
- It is not recommended to assign DOIs to intermediate breeding lines, since these will be discarded, only the starting and the final product should receive a DOI
- Ideally, DOIs could be assigned to genotypic data, but the issue of access and benefit sharing for Digital Sequence Information (DSI) is currently under debate. Future decisions are expected by the Governing Body of the ITPGRFA.

The meeting recommended strongly encouraging the assignment of DOIs.

Discussion on handling of SSD data

The issue of inclusion in EURISCO of single seed descent (SSD) lines data was thoroughly discussed. These lines are increasingly prepared for genotyping, and phenotyping is also frequently carried out on these same lines to facilitate GWAS. This is the case of the AGENT project and the EVA Networks, which have systematically created, genotyped and phenotyped SSD lines of various crops with the commitment to make the phenotypic data available through EURISCO. Different approaches exist regarding the treatment of SSD lines. Many SSDs are not maintained in the long run. ICARDA, for example, does not plan to maintain them, they are all mapped to the original accession. In the Czech genebank, new accession numbers are assigned to SSD lines created as part of the AGENT project and thus additional 1,000 accessions have been added to the collection. At IPK, 20,000 SSD lines created for barley genotyping are kept in a fridge, but are neither included in the genebank collection nor in EURISCO, except a core collection of 1,000 that represent the entire barley genome (i.e. pangenome).

As the SSD lines are often not conserved by genebanks or not maintained with the purpose of free availability, they are generally not included in EURISCO, or they remain as a subset of the accessions from which they are derived. It is, therefore, a challenge to include in EURISCO phenotypic data of accessions that either do not physically exist anymore or are not included in EURISCO as distinct entities. There was a consensus that data collected on SSD lines should not be directly linked to the original accessions, as the SSDs often represent only a subset of the diversity contained in accessions. The proposed solution to this problem is that 'dummy' or 'virtual' entries to EURISCO are created for SSD lines, labelled as research material and historical material, and linked with the original genebank accessions. As this is only an internal technical solution, the NFPs do not need to be involved. In this way, phenotypic data of SSD lines can be imported, but the SSDs would be excluded from passport data searches. In the results of phenotypic searches, it can then be correctly displayed 'line selection from…' or 'derived from…'. This approach was considered feasible, but it would also be important to maintain a reference to the context in which the lines were described, also recording whether the physical SSD samples exist somewhere, in case someone wished to access them. For this, it will be useful that the SSD lines have a DOI assigned. A solution was agreed upon in principle to create virtual entries for SSD lines in order to import their phenotypic data in EURISCO. This approach was confirmed the following day by the EURISCO Advisory Committee. The EURISCO Coordinator, also in liaison with Genesys, received the task to refine the procedure, and verify and test its technical requirements. It will remain necessary that the respective NFPs authorize the upload of the phenotypic data of SSD lines derived from National Inventory accessions.

Conclusion

The WG participants congratulated the Chair, S. Weise, for the smooth planning and management of the meeting, in collaboration with Vice-Chair K. Annamaa. The very efficient set-up of the meeting and the warm hospitality ensured by the Estonian hosts was much appreciated, including the guided visit to the city of Tallinn. The WG's work plan is included in Appendix I of this report.

APPENDICES

Appendix I. Working Group's work plan

- *Secretariat to make available a compilation of the lessons learned in the preparation of* in situ *CWR national inventories and to invite contributions from all countries to the* in situ *CWR dataset in EURISCO.*
- *The CWR WG to encourage/coordinate preparation of* in situ *National Inventories by all countries, based on existing guidelines, recently summarized in Hintum and Iriondo (2022)*[11].
- *The CWR WG to set up a link on the ECPGR website, connecting to all the useful information and tools to help countries develop* in situ *national inventories, including a link to existing checklists of CWRs per country (to be supported through an ECPGR Grant Scheme activity).*
- *The On-farm and the Doc&Info WGs to better elaborate the idea of a European inventory of landraces, based on national datasets and a list of suggested descriptors. The inventory should enable links between landraces names and actual accessions in EURISCO. A Task Force may be established, and/or the activity could be funded with contribution from the ECPGR Grant Scheme. S. Weise and J. Sustar-Vozlic volunteered to steer this activity.*
- *Organize Doc&Info webinar series on various topics of interest, e.g.:*
    - *Appropriateness of states in EURISCO descriptors 19. Biological status of accessions (SAMPSTAT), e.g.: wild, wild/natural, wild/sown*
    - *Meaning of country of origin vs. country of provenance*
    - *Procedure for selection of AEGIS material*
    - *Uncertainties about the use of descriptors*
    - *Other*
- *The EURISCO team to make a plan for the implementation of the ordering system. Implementation could start with genebanks already endowed with an autonomous system that could be connected by an API to requests generated through EURISCO.*
- *The use of a simplified template for the provision of phenotypic data should be adopted, making it available on the EURISCO website and promoting it through various channels for phenotypic data gathering.*
- *The Secretariat and other bodies are recommended to encourage assignment of DOIs to* ex situ *accessions and* in situ *populations.*

The EURISCO Coordinator, also in liaison with Genesys, to refine the procedure to handle SSD phenotypic data in EURISCO, and verify and test its technical requirements.

[11] Hintum T, Iriondo J (2022). Principles for the Inclusion of CWR Data in EURISCO.

Appendix II. Agenda

Meeting of the Documentation and Information Working Group

Agenda

17 September 2024		
	Arrival on your own	

18 September 2024		
08.30 – 09.00	Registration	
09.00 – 09.15	Welcome	S. Weise, K. Annamaa
09.15 – 09.30	Round of introductions	All
Topic 1: Coherent, comprehensive, coordinated and centralised documentation of *in situ* CWR and WFP		
09:30 – 10.00	Introductory presentation	N. Maxted
10.00 – 10.30	Tea/coffee break	
10.30 – 11.30	Discuss extension of EURISCO module for *in situ* CWR data	All
11.30 – 12.30	Discuss pros and cons of EURISCO extension for WFP data and plan possible implementation	All
12.30 – 13.30	Lunch	
Topic 2: European inventory of on-farm genetic diversity		
13.30 – 14.00	Introductory presentation	P. Ralli

14.00 – 15.30	Discussion
	• General discussion about the usefulness of such an inventory
	  o Type of inventory? Is EURISCO the right place for this?
	  o If so, agree on scope and timeline of EURISCO extension for data about on-farm PGR
	• Discuss and define minimum sets of descriptors + agreement on compiling data standards (the actual compilation will take place at a subsequent stage)
		All

	• Define composition of network of data providers + data flow	
15.30 – 16.00	Tea/coffee break	
Topic 3: Centralised ordering system for AEGIS material via EURISCO		
16.00 – 16.30	Introductory presentation	M. Oppermann/ S. Weise
16.30 – 18.00	Discussion • Pros and cons in general • Define protocol and API (proposal will be shown in the Introductory presentation) • Plan implementation of necessary APIs on EURISCO-side • Possible support for the individual data providers in implementing the API counterparts	All

19 September 2024		
Topic 4: Strengthening comprehensive information system for PGR (documentation in general)		
09.00 – 09.30	Introductory presentation	R. Kowalik
09.30 – 10.30	Get feedback from and discuss support of the EURISCO NFPs, incl. aspects of general training of genebank information system officers	All
10.30 – 11.00	Tea/coffee break	
11.00 – 11.30	Introductory presentation	S. Weise
11.30 – 12.30	Discussion • Availability of phenotypic data in EURISCO • How to increase the compliance of data with FAIR principles? • How to develop EURISCO towards a trustable repository with acceptably high governance and data management standards?	All
12.30 – 13.30	Lunch	
Topic 5: Promoting sustainable use of PGR		
13.30 – 14.00	Introductory presentation	K. Annamaa
14.00 – 15.00	Discussion	All

	• How to make data and accessions available to users? • Via centralised information systems (EURISCO) only or via crop portals? Pros and cons.	
15.00 – 15.30	Tea/coffee break	
Topic 6: Other issues		
15.30 – 15.45	Introductory presentation	T. van Hintum
15.45 – 16.15	Discussion • Where does EURISCO stand in relation to other communities? • EURISCO's perspective: full service vs. stronger integration with other aggregators?	All
16.15 – 16.30	Introductory presentation	M. Marsella
16.30 – 17.00	Discussion on various topics • Lobbying for DOIs • Standards and exchange formats • Implementation of BrAPI • Linking with genetic data • Handling of data collected on SSDs	All
17.00 – 18.00	General discussion • To what extent can/should support be provided to realise the objectives of Phase XI? • Discuss possible strategies for better networking of the Doc&Info WG (incl. a survey regarding the perspective development) • Set up a work plan + assign responsibilities (or contact persons) for the sub-areas	All
	Closing	
	Social dinner (time and location still to be confirmed)	

20 September 2024		
	Departure on your own	

Appendix III. List of participants

Working group members

Stephan Weise **(Chair)**
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) - Gatersleben
Corrensstr. 3
06466 Seeland
Germany
Email: weise@ipk-gatersleben.de

Külli Annamaa **(Vice-Chair)**
Centre of Estonian Rural Research and Knowledge (METK)
J. Aamisepa 1
48309 Jõgeva
Estonia
Email: kylli.annamaa@metk.agri.ee

Ludmilla Papoušková

Crop Research Institute, Gene Bank
Drnovská 507
161 06 Praha 6 – Ruzyne
Czech Republic
Email: papouskova@vurv.cz

Vahur Kukk
Centre of Estonian Rural Research and Knowledge (METK)
Aamisepa 1
48309 Jõgeva
Estonia
Email: vahur.kukk@metk.agri.ee

Markus Oppermann
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3
06466 Seeland
Germany
Email**:** opperman@ipk-gatersleben.de

Parthenopi Ralli
Hellenic Agricultural Organization - DIMITRA (ELGO - DIMITRA)
Thessaloniki-Polygyrou National Road
PO Box 60458
57001 Thermi-Thessaloniki
Greece
Email: pralli@elgo.gr

Dainis Ruņģis
Latvian State Forest Research Institute (LSFRI) Silava
111 Rigas St
LV-2169 Salaspils
Latvia
Email: dainis.rungis@silava.lv

Theo J L van Hintum
Centre for Genetic Resources, the Netherlands (CGN), Wageningen University and Research Centre
PO Box 16
6700 AA Wageningen
The Netherlands
Email: theo.vanhintum@wur.nl

Renata Kowalik
National Centre for Plant Genetic Resources Plant Breeding and Acclimatization Institute – National Research Institute (IHAR)
Radzików
05-870 Błonie
Poland
Email: r.kowalik@ihar.edu.pl

Jelka Šuštar Vozlič
Agricultural Institute of Slovenia
Hacquetova ulica 17
1000 Ljubljana
Slovenia
Email: jelka.sustar-vozlic@kis.si

Nigel Maxted
School of Biosciences
University of Birmingham
Edgbaston
Birmingham B15 2TT
United Kingdom
Email: nigel.maxted@dial.pipex.com

Observers

Catherine Hazel Aguilar
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3
06466 Seeland
Germany
Email: aguilar@ipk-gatersleben.de

Marco Marsella
Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture
Food and Agriculture Organization of the United Nations (**FAO**)
Via delle Terme di Caracalla
00153 Rome
Italy
Email: m.marsella@itworks.it

ECPGR Secretariat
Lorenzo Maggioni
c/o Alliance Bioversity International and CIAT
Via di San Domenico, 1
00153 Rome
Italy
Tel: (39) 066118231
Email: l.maggioni@cgiar.org