Project:	EURISCO	Date:	2021-09-01

From:	Stephan Weise (IPK)	To:	Anne-Françoise Adam-Blondon (INRAE) Theo van Hintum (CGN) José María Iriondo Alegría (URJC) Kjell-Åke Lundblad (NordGen) Lorenzo Maggioni (ECPGR) Matija Obreza (Crop Trust) Paul Olson (KWS) Ludmila Papoušková (CRI)
CC:	Francisco Lopez (ITPGRFA), Pragna Kotni (IPK)		

MEMO – EXTENSION OF EURISCO FOR *IN SITU* DATA

1   Background

EURISCO is currently limited to germplasm accessions maintained *ex situ*. At the third EURISCO Advisory Committee meeting on 7 July 2021, the inclusion of further domains in EURISCO was discussed. One focus was on *in situ* material. The EURISCO coordination was mandated to prepare a memo outlining a proposal for the extension of EURISCO to include *in situ* data, as well as the associated effort and the required data flow mechanism.

While the addition of information on well-managed *in situ* CWR seems reasonable, this cannot be answered clearly for on-farm LR data. An important prerequisite for the operation of a central search catalogue such as EURISCO is the regular updating of the data available there. While this can be considered rather uncritical for *in situ* CWR, the associated logistical effort is much greater for on-farm conservation. For this reason, on-farm data should be excluded for the time being and the focus placed only on *in situ* CWR.

The H2020 project Farmer's Pride has already addressed the question of a possible extension of EURISCO for such data. The following considerations are therefore largely based on the concept developed there[1], but also take into account the recently published proposal for *in situ* CWR descriptors[2] of the International Treaty (ITPGRFA).

2   Network/data flow

Currently, EURISCO documents more than two million accessions from about 400 *ex situ* collections from 43 countries. This, especially the regular updating of the data, is only possible by compiling the data in the National Inventories of the member countries in a first step. An analogous structure embedded in the ECPGR is also a necessary prerequisite for *in situ* data.

In the case of *in situ* collections, responsibilities cannot be defined as easily as in the case of *ex situ* collections. Therefore, a feasible way forward in this context could be to gradually build up a

---

[1] Weise et al. 2020, https://more.bham.ac.uk/farmerspride/wp-content/uploads/sites/19/2020/10/D2.5_EURISCO_in_situ_extension_concept.pdf
[2] Alercia et al. 2021, http://www.fao.org/3/cb3256en/cb3256en.pdf

corresponding network of *in situ* National Inventory Focal Points. José Iriondo pointed out during the AC meeting that Spain is working on establishing a register with requirements for official genetic reserves. This and similar efforts could provide a nucleus to feed the first *in situ* data into EURISCO. From EURISCO's point of view, it does not matter how the "*in situ* National Inventories" are organised in the individual countries. The only decisive factor is that, in analogy to the existing network of *ex situ* National Inventory Focal Points, there should be a responsible contact person per country and a data exchange format that is used uniformly by all partners, and of course that the *in situ* data transmitted to EURISCO is kept as up-to-date as possible at all times. In fact, lists of *in situ* and on-farm National Inventory Focal Points were started to be defined already during Phase VII (2004-2008) of ECPGR and are available from the ECPGR web site[3], but these lists will need to be updated, based on Terms of Reference to be defined for the Focal Points.

3 Proposal for extension

3.1 Data exchange

Within the Farmer's Pride project, a concept for a possible extension of EURISCO for *in situ* data was developed (see above). This includes a proposal for the data exchange of *in situ* CWR data. In order to minimise the necessary adaptations of the EURISCO infrastructure, the aim was to achieve the greatest possible conformity with the MCPD standard. The proposal was forwarded as input to the ITPGRFA, which started a project in May 2019 to develop an internationally accepted standard for the exchange of *in situ* CWR data. The results have recently been published (Alercia et al. 2021, see above). As a widely accepted standard is a crucial prerequisite, the ITPGRFA document should form the basis for a EURISCO extension. It was therefore analysed with regard to its feasibility.

The proposal provides a good basis for the exchange of *in situ* CWR data. However, some additional descriptors would be desirable from EURISCO's point of view and a final list of "EURISCO-CWR *in situ* descriptors" should be eventually agreed by the ECPGR Working Groups on 'Documentation and Information' and on 'Wild species conservation in genetic reserves':

Farmer's Pride Descriptors

- NICODE - to support the concept of National Inventories and to make it easier for users to identify the responsible countries.
- COORDUNCERT+GEOREFMETH - together with the geo-coordinates, these descriptors would be more helpful than textual descriptions of the occurrence sites and at the same time provide the possibility to indicate imprecise locations in order to protect rare populations; could in principle replace the ITPGRFA descriptors 7-7.3.
- OTHERNUMB - in principle mandatory, e.g. to make changes in accession numbers or population IDs transparent.

Other Descriptors

- PUIDs/DOIs of *ex situ* accessions derived from the *in situ* population – DOIs are increasingly used for *ex situ* conservation and there is no obvious reason to omit this information here.
- Optionally also PUIDs/DOIs of *in situ* populations, as already discussed in the ITPGRFA proposal.

The ITPGRFA proposal provides the possibility to document *in situ* CWR populations that are additionally available in an *ex situ* collection as accessions or for which herbarium vouchers exist. From the point of view of the EURISCO coordination, it would be desirable to exclude herbarium vouchers

---

[3] https://www.ecpgr.cgiar.org/contacts-in-ecpgr/ecpgr-documentation-focal-points

and instead to primarily focus on material that is available in principle and, if possible, also maintained in an *ex situ* collection.

The ITPGRFA proposal contains a descriptor LINKS, which should allow the provision of several URLs. From EURISCO's point of view, it would be desirable to limit this to only one URL. On the one hand, this would be in line with the MCPD standard for *ex situ* data and, on the other hand, experience has shown that link collections become outdated quite fast. Here it would make sense to use a relatively stable link, e.g. the accession URL of a genebank that maintains an *ex situ* sample of the germplasm resource.

3.2   Database infrastructure

For the management of *in situ* CWR data, two possibilities can be considered in principle. On the one hand, an attempt could be made to use the existing infrastructure for *ex situ* data as far as possible and only carry out the necessary extensions. However, this could lead to minor restrictions in the representation of *in situ* CWR data, as the current infrastructure was developed to accommodate specifically *ex situ* data. On the other hand, there would be the option to develop a stand-alone database schema for *in situ* CWR data. This would allow greater flexibility in terms of modelling, but would also require significantly more effort.

Due to the fact that there are relatively high correspondences with the MCPD standard in the ITPGRFA proposal and that the proposal generally only contains the most important *in situ* CWR data, the first variant of extending the existing infrastructure is preferred.

3.3   Import tools and integrity checks

Data exchange between the National Inventories and EURISCO is done by means of MS Excel files using the MCPD standard. An upload tool (currently Java-Swing/WebStart, soon Oracle-APEX) is used to import the files provided into a staging area, where the necessary data integrity checks are carried out. After subsequent approval, the new data is integrated into EURISCO. The procedures for both the checks and the integration are fully implemented as PL/SQL packages.

An upload tool as well as procedures for integrity checks and data integration are also required for the extension for *in situ* CWR data. Due to the existing solutions for the *ex situ* data, it is not necessary to start from scratch, existing procedures should be reused as far as possible. However, this step will require a significant amount of the developer's time.

3.4   User interface

In order to assess the suitability of the existing EURISCO web interface for *in situ* data, a survey was conducted among the project partners within the Farmer's Pride project. It came to the conclusion that the interface is generally also suitable for *in situ* data, but that a number of enhancements would be desirable. The requirements can be summarised as follows:

- Provision of additional information (*in situ* passport data + extra information, such as date of last monitoring, threat status, distribution area etc.)
- Extension of search and filter functionalities with regard to *in situ* data
- Improvement of the presentation of search results, incl. the connection of *in situ* data with climatic, topographic, edaphic or geological information
- Additional types of reports
- Online analyses, such as comparisons of *in situ* populations with *ex situ* samples per species

Further details are given in the concept from the Farmer's Pride project.

4   Effort of the extension

The estimation of the necessary implementation time by the EURISCO coordination office is based on the assumption that a dedicated qualified developer is available full-time. In case the implementation

is to take place in parallel with other work on the EURISCO information system instead, the required development time will increase and depend on ongoing priorities.

Task	Effort in person months
Adjustment/extension of EURISCO's database structures	1
Development of import tools for *in situ* CWR data	2
Development of procedures for data integrity checks and data integration (reuse of existing procedures for *ex situ* data as far as possible)	3
Extension of the EURISCO web interface according to user requirements	6
Total	**12**

This estimate includes the pure implementation time. To this would be added the time for discussions on necessary additions and/or changes to the *in situ* CWR data standard proposed by the ITPGRFA (see above). This should be estimated at a minimum of **3–6 months**.

The time estimate is exclusively based on the extension for passport data of *in situ* CWR. Should phenotypic data or data on climatic, topographic, edaphic or geological conditions of natural populations be considered in the future, this would require additional effort. However, this is not seen as a priority at the moment.

In general, it has to be judged (a) whether it is possible to acquire additional funding for the extension or (b) whether the implementation should take place within the regular EURISCO budget:

a) Although the first case has the advantage that there is no negative impact on EURISCO's work plan and regular budget, it must be pointed out here that the labour market for IT professionals is currently very tense. The recent appointments of the new EURISCO software developer and the EVA developer have shown that it is currently very difficult to fill even 2-year positions. A lot of effort was needed to recruit qualified staff.

b) If, in contrast, the extension is to be carried out within the regular EURISCO budget, this would mean that the entire implementation would have to be stretched out over a significantly longer period of time, as it would be in competition with other tasks. Alternatively, the other tasks could be completely postponed and primarily the *in situ* data extension could be implemented. However, the EURISCO coordination does not consider this to be convenient.

5   Summary

The extension of EURISCO for *in situ* CWR data is technically feasible and can be implemented within a manageable timeframe.

The ITPGRFA proposal is a good starting point for data exchange, but from the point of view of EURISCO coordination, adjustments are necessary.

The basic prerequisite, however, is the willingness of the ECPGR member countries to define a central contact person per country to coordinate the data exchange with EURISCO (in analogy to the *ex situ* Focal Points). Here, it is quite conceivable to start this on a smaller scale with a kind of pilot group.