European Cooperative Programme for Plant Genetic Resources (ECPGR)

Minutes of the First ECPGR EURISCO Advisory Committee meeting

14 October 2016, Angers, France

Present:
Theo van Hintum, CGN (Chair)
Kjell-Åke Lundblad, NordGen
Lorenzo Maggioni, ECPGR
Matija Obreza, Crop Trust
Paul Olson, KWS
Ludmila Papoušková, CRI
Ian Thomas, IBERS

Observers:
Anne-Françoise Adam-Blondon, INRA
Helmut Knüpffer, IPK
Stephan Weise, IPK (EURISCO Coordinator)

Unable to attend:
José Iriondo, URJC

The Agenda for this meeting is available online (here).

1. Opening

The Chair of the EURISCO Advisory Committee (AC) welcomed all participants, who then briefly introduced themselves. It was reminded that the EURISCO AC was established and endorsed by the ECPGR Steering Committee with the role of monitoring the progress of EURISCO and offering advice for its development. Apart from the ex officio members representing ECPGR and Genesys, each member was selected based on his/her personal capacity and in order to have a balanced representation of different types of expertise and stakeholders views. Each member was expected to serve in the Committee on a voluntary basis and everyone was thanked for his/her availability to participate.

Stephan Weise, EURISCO Coordinator, presented the progress of EURISCO (content, data quality, updates, outcomes and coordination activity) (PPT available here).

He then clarified a few points raised by Committee members:

- "Materialized view": a view is a logical relation (virtual table) in a relational database management system, which is based on a database query. A materialized view is persistently stored in a database and acts as a kind of cache in order to improve query performance.

- The source codes of the system are maintained in a repository (internal IPK CVS[1])) allowing for version tracking.
- How much technical documentation exists for EURISCO? The overall architecture of EURISCO as well as all workflows for updating data are documented. All source codes are commented comprehensively. Documentations (handbooks) for end users are not available yet (except slides of given presentations).
- Monitoring of the use of EURISCO is carried out on the basis of aggregated statistics using AWStats. More detailed information is not possible owing to legal restrictions that do not enable saving the IP addresses.
- Feedback from users is received on occasional and bilateral basis, while no structured interviews have been carried out yet. Ian Thomas confirmed that positive feedback had been received from the UK users.
- The EURISCO Coordinator dedicates half of his time to development and half to networking coordination.
- The old (outdated) French data had been recently removed from the catalogue by the National Focal Point (NFP) and import of new data has started. No other extra-ordinary activities by other NFPs were reported or expected.

M. Obreza presented Genesys and the related role of EURISCO (PPT available here).

He made the following considerations of specific interest:

- The use of Genesys is on average ca. 200 visits per day. It is known that the potential audience for online PGR databases is not very large and this is not considered an issue.
- The Trust invests a lot in Genesys development including annual user interface improvements, as it is an important aspect for usability of the service.
- Genesys receives data either from consortia such as EURISCO and USDA GRIN, or directly from individual genebanks (11 CGIAR collections and others). Data from the Millennium Seed Bank are also received directly from MSB, while the respective data set received from EURISCO is not used. This arrangement allows for frequent publication of information about material collected as part of the CWR project on Genesys.
- |n addition to the Crop Trust IT, about 10 consultants have contributed to the development of Genesys and they are familiar with its architecture and source code.
- From the point of view of Genesys, the crucial role of EURISCO lays in its networking activity, which is providing capacity building, advice and harmonization across the European genebank information systems. This results in a constant improvement of the data quality, which is the essential asset of any aggregated catalogue. The provision of clean data from all of Europe to EURISCO is a key service for Genesys.

A discussion followed on the expectations of the AC for the development of EURISCO during the next 10 years.

The AC agreed that the most important role of EURISCO will continue to be focused on helping genebanks to provide increasingly high quality and standardized data. It was generally agreed that the priority should remain with the data themselves rather than developing data processing tools, as the value of the catalogue is dependent on reliability and accuracy of the data. On the other hand, when the data are good, it is easy to aggregate, display and analyse them with existing tools.

The main assets of EURISCO in the near future were therefore expected to remain the networking function, as well as the knowledge of specific crops, which is also an aggregated

[1] Concurrent Versions System (https://en.wikipedia.org/wiki/Concurrent_Versions_System)

expertise provided by the ECPGR Working Groups. This expertise can be directed to improve data quality, data standards, as well as, in the near future, to analyse phenotypic data and provide annotated information about the collections documented in EURISCO.

In conclusion, EURISCO's role as data platform may become less essential in the future, since different options to harvest and display data will likely emerge. Currently EURISCO is expected to remain as a network of data suppliers, providing education, training and standards. In this process, the genebank management system GRIN-Global[2] might also have a role and EURISCO might acquire an additional role as GRIN-Global advisor. Feedback of crop experts is also still important for curation and data aggregation.

It was also reaffirmed that the aims of Genesys and of ECPGR/EURISCO are the same, with complementary roles and that it would be useful to jointly address, whenever relevant, any issue related to filling data gaps or improving data quality and data acquisition in Europe.

2. Software development (frontend, backend, support tools for NFPs)

Based on the presentation of the current status of the EURISCO software, the AC discussed what priorities should be considered for further development (improvements of existing functionalities or new additions).

A proposal had been received from José Iriondo to link the EURISCO database to a set of environmental layers for Europe (climate, soil and topographic data). In this way, it would become possible to search the database by imposing thresholds to e.g. annual rainfall or minimum temperature and thus identifying which accessions originate from areas with given climatic parameters.

The AC expressed some doubts about the applicability of environmental layers to the scale of Europe. It was also considered that Genesys already includes this functionality and therefore it would not be advisable to duplicate the same effort in EURISCO. Linking EURISCO also to other types of databases may facilitate the user to analyse the data with various software tools, but the main asset (and therefore priority for action) of EURISCO remains the quality of the data itself, rather than the development of analytical tools.

The addition of the date of last update of passport data to the web interface would be important for the user and the AC agreed that this feature should be implemented in EURISCO.

Similarly, in order to increase the user-friendliness of the catalogue with respect to access to the material, *the AC recommended that the EURISCO Coordinator provide on the web interface clear instructions for users wishing to access material from the holding genebanks*.

The possibility to publish data from the National Inventories offering an open access API (Application Programming Interface) was not considered a priority as this technology is not yet mastered by most of the genebanks and collection holders.

In summary, apart from the two improvements indicated above ('last update' and 'instructions for access'), no other obvious priorities for software or frontend development were identified at the moment. However, *the EURISCO Coordinator was expected to take into account feedback from users' interviews and eventually prepare a prioritized list of actions.*

[2] www.grin-global.org

3. Future workshops (regional approach, teaching materials, collaborations)

In the last two years of ECPGR Phase IX (2017 and 2018), funds were allocated to organize EURISCO training workshops every year. However, a new strategy might be created to involve all the NFPs, to improve the curriculum, to establish longer-term collaborations in these workshops. For example, writing/compilation of an extensive technical manual for genebank documentation and EURISCO uploading could be an idea for submission to the Grant Scheme.

The AC agreed that it would be useful to organize a workshop involving all NFPs at the same time in order to raise the sense of network community as well as to maintain a good level of collaboration towards an increasingly improved EURISCO. Moreover, in such a workshop, a demonstration of GRIN-Global could be organized also with involvement of the Trust. As the budget of ECPGR currently would not be able to cover this meeting, the idea should remain floating and be picked up for implementation when possible.

In the short term and in support of the next training workshops, it was recommended that S. Weise write a short document with instructions for uploading data onto EURISCO, and prepare a set of practicals to test the tools and to find errors in the data set.

4. Strategy concerning characterization and evaluation (C&E) data (targeted expansion, role of Crop Working Groups)

A mechanism to upload phenotypic data to EURISCO has been created. It will now be critical to obtain such data from the National Inventories for EURISCO. Additionally, the questions would be: what needs to be done to make such data digestible to the users and should we create and promote the use of ontologies? Should we try to process the phenotypic data in EURISCO in order to get to 'standardized data sets' that combine separate datasets into 'the best estimate' per trait/accession/combination?

Having in mind the point of view of plant breeders, Paul Olson recommended not to create additional layers that may or may not have value. It would be more useful if crop experts could offer some curation of the data and be able to separate the accessions into groups (such as resistant, intermediate, susceptible for a particular disease or environmental stress, or make note of relative maturity or unusual phenotypes for the species). An authority would be needed to take the responsibility to recommend a list of accessions as particularly valuable, owing to a peculiar trait or a specific extreme behaviour. In fact, users are unlikely to spend their time attempting to interpret the scores, but would be very interested to find accessions marked as, e.g., susceptible vs. resistant.

The AC agreed that the Working Groups could be the suitable fora to curate data sets in EURISCO and to propose lists of elite material or diversity panels for various purposes.

It was also agreed that the current focus should however remain on getting the phenotypic data into EURISCO as a starting point.

5. Strategy concerning x-omics data

X-omics data are being produced at accelerating pace. The approach of EURISCO vis-à-vis x-omic data should be discussed. One option could be simply to write a statement regarding the management and sharing of this type of data, or we could start thinking about ways of linking EURISCO data with x-omics datasets, or even try to create an interface allowing queries from EURISCO combined with x-omics data.

The ongoing genotyping work on barley at IPK will offer a testing ground. New approaches will have to be developed at IPK to visualize these data. Additionally, EURISCO staff is involved in an ERA-CAPS project proposal where, if successful, they will try to learn

and link to data through an interface so that valuable specificities of given accessions would be highlighted.

Apart from genetic data, also phenomic, transcriptomic, metabolomic etc. data are being generated and the genebank community should remain alert about developments, although a role for EURISCO in this respect is for the moment not realistic.

A critical issue is the possibility to firmly link data to specific accessions. In this respect, the use of DOI (Digital Object Identifier) codes should allow precise linkages between genotypic data and specific accessions that may be stored in genebanks.

It will therefore be important to strongly promote and support the adoption of DOIs by the genebanks.

6. Improving the coverage of material

A discussion on the coverage of material suitable to be included in EURISCO was mainly triggered by the recent addition of a very large dataset of *Arabidopsis* accessions by the UK.

Questions were related to the bias determined to the EURISCO statistics by this non-food or feed subset of accessions. Also the risk of overloading the database and slowing down the searches was considered. However, it was appreciated that *Arabidopsis* is a genuine gene resource and a gene discovery tool model species. The importance for EURISCO to be inclusive rather than exclusive about the choices made by individual National Inventories was also brought up. The EURISCO Coordinator confirmed that a very large data set slows down the database a bit, but this is not a big problem. It was also noted that the EURISCO Data Sharing Agreement provides sufficient clarity as to the type of material that is expected to be included in EURISCO and research material is contemplated.

The AC was eventually convinced that Arabidopsis data are a legitimate set of accessions for EURISCO and M. Obreza announced that these data will also be included into Genesys, from where they were so far excluded since the dataset was not updated on Genesys after inclusion of Arabidopis data.

Regarding other sets of data that are missing and should be included in EURISCO, it was noted that the French data are in the process of being prepared for inclusion. Other national datasets such as from Malta and an extension of the coverage, such as for the Russian Federation (include all data available at VIR's own website) or Italy (inclusion of Bari genebank), would be welcome. It was recommended that the EURISCO Coordinator approach NFPs to find out whether relevant data were missing from the various National Inventories, in order to compile a list of priority datasets to be encouraged for inclusion.

7. Strategy concerning inclusion of *in situ* data

At the last meeting of the ECPGR Doc&Info Working Group in Prague (May 2014), it was decided that data of *in situ* crop wild relative (CWR) populations could be included in EURISCO, provided that a manager could be clearly identified, who could be approached in case someone would want access. Also the ECPGR Steering Committee during its last meeting in June 2016 reiterated the recommendation that the Doc&Info and the Wild species WGs, together with other selected Crop WGs should develop an agreed minimum data exchange format on the basis of existing *in situ* CWR descriptor lists.

Upon consideration that genetic reserves are starting to be created, also as a result of the ECPGR *in situ* crop wild relatives' conservation concept, it seems that the time is right to allow inclusion of the first "*in situ* data" into EURISCO.

Th. van Hintum, Chair of the Doc&Info WG, confirmed that he would approach Nigel Maxted, Chair of the Wild species WG and come up with a proposal for a suitable data exchange format, to be considered for adoption by the ECPGR community. Genesys staff will also be included in the discussion.

8. Adoption of the latest version of the MCPD

EURISCO is currently based on the first version of the MCPD, completed in 2001. Since then, various updates were released by Bioversity International and FAO, the last being MCPD v.2.1 (2015). Updated versions are the results of global consultation of experts and include important improvements or additional descriptor fields, such as specifications of the geographic coordinate method used and, most recently, the Permanent Unique Identifier (PUID).

The AC agreed that updated MCPD should be implemented as soon as possible in EURISCO. However, this might require adaptation of database schemata and user interfaces and revised import/check procedures.

At the same time, the AC agreed that the release by FAO of a new version of the MCPD (v.2.2) should be solicited, since there was agreement to include information regarding CWRs conserved in genetic reserves (see above) and about "historic accessions". In fact, it is important for the user to track historic accessions, which have been used or available in the past, but are no longer actively maintained or existing or alive and therefore they are no longer available.

The Chair of the Doc&Info WG agreed to liaise with FAO to propose the issuing of MCPD v.2.2, including 'historic' and 'in situ' *descriptors (as soon as agreed by the EURISCO AC). The EURISCO Coordinator agreed to implement the MCPD v.2.2 (or equivalent if not yet formalized by FAO) within the first half of 2017. Genesys will also be kept in the loop for the drafting of the new descriptors.*

Regarding a proposed descriptor on "availability" of the accessions, the AC took note of a discussion held during the Angers Training Workshop, whereby no general consensus was expressed for the introduction of such descriptor, owing to the difficulty to keep it up-to-date and the consequent risk to communicate misleading information. It was however noted that the descriptor "Part of AEGIS" indicates those accessions that are expected to be promptly available, owing to the formal agreements concluded by the AEGIS member countries.

9. Taxonomy issues regarding data in EURISCO

The quality of taxonomic data is still a weak point in EURISCO. There are still many errors and it would be desirable, apart from checks prior to including the data in EURISCO and feedbacks on errors sent to the NFPs, also to include a 'preferred taxon name' or something similar (see the GRIN solution). In the end the user must be able to find also the accessions that have been included with spelling errors in their taxonomic name or with synonyms.

An idea would be to internally map the data provided by NFPs (which may not be modified by EURISCO) onto a preferred taxonomic nomenclature, e.g. GRIN taxonomy, and to use this information for searches too. In this way, the original data would not be altered, but users would be able to search all relevant accessions, even if misspelled or named differently from the 'preferred taxonomy'.

It was noted that there might be funds provided by Germany during 2017, for the involvement of a Post-Doc (to be clarified).

The AC agreed that a 'taxonomy project' to improve taxonomic quality of the data in EURISCO would be the most appropriate priority in case funding would be made available by a donor.

It was also noted that the EURISCO approach to resolve the taxonomy issues would be an encouragement to Genesys to consider doing the same.

10. Procedure for NFP approval for upload of C&E data

Currently only NFPs can reach the EURISCO intranet, so they are the only ones who can upload passport or phenotypic data. However, it was proposed that also others should be able to do so, provided that NFPs maintain the ultimate authority and responsibility to approve uploaded datasets before they are published. Phenotypic data would likely be uploaded in bulks including sets of accessions held by different countries. It should therefore be clarified who should have the authority to approve each upload (i.e. the NFP from the "uploading country" or all the NFPs corresponding to the "uploaded data").

It was agreed that the EURISCO Coordinator should provide draft instructions related to the upload procedure, including authorization steps. This draft would then be circulated to the NFPs, National Coordinators and Bioversity International legal office for approval.

11. Other issues

Ian Thomas, on behalf of the Forages WG, requested the possibility to add (two or three) additional crop-specific descriptors into EURISCO. It was agreed that initially a quick solution would be the inclusion of crop-specific information in the 'Remarks' field with specific codes for remarks.

I. Thomas and S. Weise agreed to work together to find a pragmatic solution for the time being. Frequent and consistent use of remark codes might subsequently evolve into the decision to adopt additional crop-specific descriptors in EURISCO.