Project:	EURISCO	Date:	2015-03-03

From:	Stephan Weise (IPK)	To:	Lorenzo Maggioni (ECPGR Secretary) Theo van Hintum (Documentation and Information Working Group)
CC:	Helmut Knüpffer (IPK), Markus Oppermann (IPK)		

This document describes the work packages to be processed in 2015 for hosting and maintaining the European search catalogue for plant genetic resources (EURISCO), and for coordinating the EURISCO network.

Content

1   Work plan for the EURISCO development

In 2014, EURISCO was transferred from Bioversity International, Italy, to its new host IPK Gatersleben, Germany. Due to technical and organisational issues, the existing solutions could not be transferred to the new host as is. Instead, a reengineering from scratch was performed. It was necessary to find a balance between developments which are visible (e.g., search interface) and those which completely took place in the background. For this reason, there are currently certain functional limitations of the new system, which will be further addressed in 2015. In addition, the work will focus on the extension of EURISCO for C&E data as well as on the improvement of the upload mechanism for National Inventories.

1.1 WP1: Extension of EURISCO for C&E data

1.1.1 Provision of exchange format

Although it was originally not planned for 2014, incremental updates were given a priority for implementation. Thus, the discussion about the C&E data exchange format was postponed to 2015.

The basis for the exchange format for C&E data are the respective proposals by Theo van Hintum (CGN)[1] and Jonas Nordling (NordGen)[2]. These proposals were screened and first ideas were developed. They will be used to prepare an exchange format, which will then be provided to the PGR community.

1.1.2 Implementation of extension

An extension for the EURISCO database schema will be designed and implemented, following a generic approach, such as the Entity-Attribute-Value concept or the Single-Observation concept. The extension will support the upload of pure metadata about experiments and will also allow attaching links to publications (at accession level). The EURISCO web application will be extended for searching, displaying and downloading of C&E data. Due to the expected amount of data, performance tuning tasks will be indispensable.

Standardisation of C&E data shall only be discussed after reaching a critical mass of data imported into EURISCO. In this context, the output of the ontology communities will be utilised. Furthermore, displaying of interpreted data should be a topic. For these tasks, a separate work package will be formulated in due time.

1.2 WP2: Extension of the upload mechanism for National Inventories

In 2014, a new data upload and integrity check mechanism was developed from scratch. For the user interface, the Oracle APEX technology was used. The data integrity checks were implemented as PL/SQL procedures. The upload mechanism will be further improved.

1.2.1 Direct import of data from MS Excel files

In order to improve the user-friendliness of the data import as well as to reduce encoding errors, which often occur during the preparation of tab-separated text files in EURISCO format, an import tool will be developed for importing MS Excel files (using SAX-like parsers with low resource requirements). The development of the MS Excel importer for passport data was started in 2014 and will be continued in 2015. The import functionality for C&E data will be developed after the implementation of the C&E data extension (see 1.1.2).

1.2.2 Improve data quality checks performed during import (to be continued in 2016)

For the new upload mechanism, quality checks, which are performed during the data import, will be improved. The following improvements are proposed and will be specified together with the NFPs:

- Collecting sites (geo-coordinate checks; geo-coding software)
- Institution code data (old codes vs. current ones; link with WIEWS data)
- Country codes (old & current codes need to be maintained)

---

[1] Inclusion of C&E data in EURISCO – analysis and options
[2] C&E data: the EURISCO standard

- Reporting mechanism (esp. latitude/longitude and country codes)
- Taxonomy (GRIN, Catalogue of Life, Mansfeld)
- Crop names (misspellings, typos, synonyms, duplicate names)

1.3 WP3: Extension of functionality of the public EURISCO application (permanent activity)

With the launching of the new public EURISCO application, the users were asked for feedback and for providing requirements for functionality extension. The following requirements were collected so far and will be implemented after checking of feasibility:

- Provide additional reports and download facilities
- Extend filter possibilities by additional fields
- Ignore MLS/AEGIS status during search
- Update NFP contact data automatically from ECPGR
- Show status codes (biological status, acquisition source etc.) in select lists
- Enable case insensitive search for taxa
- Provide a "free keyword search" over all alphanumeric fields

EURISCO will be continuously developed in the future, driven by the demands formulated by its user community.

1.4 WP4: Documentation and planning of the next steps

In parallel to the implementation processes, all developments will be documented.

Furthermore, the next development steps (WP1-3) will be specified in detail.

2 Schedule

	Work package	2014			
		Q1	Q2	Q3	Q4
1	**Extension for C&E data**				
	Provision of exchange format	▓			
	Implementation of extension		▓	▓	▓
2	**Upload mechanism extension**				
	Implement direct import of data from MS Excel files	▓	▓		
	Improve data quality checks (to be continued 2016)			▓	▓
3	**Extension of functionality of public application**	▓	▓	▓	▓
4	**Documentation and planning**	▓	▓	▓	▓

3 Coordination tasks

3.1 Coordination task 1: EURISCO network maintenance

3.1.1 EURISCO stakeholders, outreach and dissemination

The contact to (potential) EURISCO stakeholders will be intensified. The aim is to raise the awareness of EURISCO for breeders, researchers and also on political level by demonstrating the potentials of this common European approach.

In this context, the EURISCO fact sheets will be updated. A first paper describing the current status and content of EURISCO was submitted for publication end of December 2014.

3.1.2 Reactivation of the network of National Focal Points and EURISCO newsletter

After finishing the EURISCO transfer, a large group of potential users (all National Focal Points and National Coordinators as well as all members of the ECPGR working groups) was contacted by email. A short report about the status of the EURISCO transfer was given and all recipients were asked for their input for the future development of EURISCO. Thereby, 45 new subscribers to the EURISCO newsletter were gained. We will continue to get in contact with this user community, and will try to acquire new groups of users (see 3.1.1). This will be complemented by direct, personal communication with the users, e.g., for supporting updates, provisioning of specific database queries and special data export formats, discussion about future developments etc.

The regular issuing of the EURISCO newsletter (e-bulletins) will be indispensable to provide feedback to the EURISCO users. We aim at issuing a newsletter twice a year (June/December).

3.1.3 Training of EURISCO data providers

Regular training for data providers will especially be critical for C&E data and also in case of staff changes. Smaller regional meetings will be organised regularly, depending on ECPGR funding. A proposal for financing a first regional training workshop for South-East Europe was submitted to the ECPGR Activity Grant Scheme (Phase IX) in September, and accepted end of December 2014. This workshop will take place at the Agricultural University of Tirana, Albania, in May 2015. All necessary organisational tasks will be performed. Training materials will be developed.

3.2 Coordination task 2: Coordination of the EURISCO development

The coordination of the future EURISCO development mainly comprises (i) the definition of new services, (ii) the advancement of current standards and (iii) the discussion and definition of new standards. Currently, the following steps are planned:

- Definition of new services:

  In order to improve the data quality, additional services will be defined, especially improved error checking and taxonomic validation during the data import. The implementation will be performed in the frame of the EURISCO development (see 1.2.2).

- Discussion and definition of new standards:

  As a prerequisite for extending EURISCO for C&E data, a data exchange format will be defined until end of March 2015 (see 1.1.1).

As a continuous task, at least two structured interviews per year with EURISCO users will be performed (personally or by telephone) in order to get feedback regarding both existing features of the web application and new requirements. The results of the interviews will be documented and will form a basis for further improvements.

In order to improve the coverage of EURISCO, we will focus on bilateral communication with the respective National Focal Points / National Coordinators. The basis will be the analysis by Theo van Hintum (CGN)[3,4]. This will be a permanent activity.

3.3 Coordination task 3: Participation in project proposals

An indispensable task of the EURISCO coordination is to participate in the preparation of project proposals in order to acquire additional funding for the future development of EURISCO.

---

[3] The documentation of Plant Genetic Resources in Europe
[4] EURISCO and CCDBs - coverage and overlap