Minutes of a meeting of the ECP/GR NCG database managers

Sugar, Starch and Fibre Crops Network

Held at the Federal Centre for Breeding Research on Cultivated Plants (BAZ) at Braunschweig, Germany

27-29 July 2005, Braunschweig, Germany

Provisional agenda

Wednesday 27 July
Arrival of participants

Thursday, 28 July
08:30 Transport from the guesthouse to the meeting place

TOP 1	*Introduction* *Opening of the meeting, welcome, organisational matters, acceptance of the agenda* *Introduction of the participants*
TOP 2	*Aims of the meeting*
TOP 3	*Flax*
	Coffee break
TOP 4	*Hemp*
TOP 5	*Potato*
TOP 6	*Oat / Beet*
TOP 7	*Discussion and preparation of the afternoon session*
	Lunch
TOP 8	*Prospects for joint developments*
	Coffee break
TOP 9	*Drafting of a project proposal and the report*

17:00 Transport to the guesthouse

19:00 Transport to the city, joint dinner

Friday, 29 July
Departure

Participants

A. Carboni, Italy (excused), L. Frese, Germany, C. Germeier, Germany, R. Hoekstra, The Netherlands, M. Pavelek, Czech Republic, J. Nozkova (guest), Slovak Republic

Minutes

TOP 1 Introduction
L. Frese, chair of the meeting, welcomed the participants also on behalf of Lorenzo Maggioni and outlined the task of the group. He explained that representatives from SASA had been invited on short notice but were unable to attend. The agenda was accepted. R. Hoekstra and C. Germeier agreed to take notes for the meeting report. The chair agreed to draft the report under the condition that the participants provide their powerpoint presentations. The group agreed to distribute the powerpoint files as report attachments. The participants briefly introduced their functions and interests.

TOP 2 Aims of the meeting
C. Germeier explained the technical aim of the meeting (see attachment 1).

TOP 3 Flax database
M. Pavelek described the historical development of the International Flax Data Base (IFDB), its information content and the structure of flax collections (see attachment 2).
J. Nazkova presented the results of a project called GENOTYPDATA which is a universal system for recording, classification, cataloguing, study, selection, and evaluation of genetic resources of many (currently 10) species (see attachment 3). Amongst others the Agricultural University, Institute of Biodiversity Conservation and Biosafety at Nitra, developed the Linum application of GENOTYPDATA. The system contains passport and evaluation data on 121 genotypes and is able to present the data on a specific accession along with a collection of close-up photos of the described plant parts in a very user-friendly and attractive manner.

The Czech and Slovakian participant noted that intensive working relationships exist between the Institute at Nitra and the company AGRITEC at Sumperk in the field of flax germplasm documentation. In addition the Czech Genebank, in particular I. Faberova, is informed on all activities such as the meeting today.

The presentations were followed by a discussion leading to the following conclusions:
- The MCPD structure is implemented in the IFDB.
- A duplicate search in the IFDB using identical or similar sounding accessions names resulted in 9 duplicate groups. The number of unique accessions is 2 965.
- The germplasm collections are being characterised and evaluated but the evaluation as well as passport data cannot be accessed on-line. An image analyse system (software LUCIA 4.61) is used by Agrictec for the characterisation of accessions based on standardised photos of leaves, flowers etc.
- EURISCO holds passport data on 16 596 accessions while 27,000 accessions are known to exist in European holdings. Since the IFDB contains passport data on approximately 8,000 accessions, information on approximately 19,000 accessions need to be added.
- Adding and updating these 27,000 accessions should be combined with a software assisted duplicate search and classification.
- It was noted that important holdings of flax germplasm exist in Canada and the USA.
- Support (expertise and staff) is required to improve the IFDB (internet access, modules for C&E data) and to update the data base via EURISCO and/or through the individual national holdings.
- It should be investigated whether the European Flax Database and the GENOTYPDATA Linum can be used to develop as complete system holding passport and C&E data as well.

TOP 4 Hemp
Due to unforeseeable reasons, the hemp database could not be presented.

TOP 5 Potato
R. Hoekstra introduced the European Wild Potato Database (EWPD)developed and operated by the CGN and the European Cultivated Potato Database (ECPD) at SASA (see attachment 4).

The latter comprises 11,560 accessions of approximately 4,000 different varieties and 1,400 breeding lines. SASA is implementing new software with improved internet query functions using a commercial product. A search for duplicates is conducted by the database manager, but due to a rest uncertainty in the search results and the often unknown virus infections status of the most original and duplicate samples no attempt was made to rationalise the European collection of cultivated potatoes.

The EWPD contains passport and C&E data on approximately 12,000 accessions of wild and Andean cultivated potato species. He noted that the database includes some additional non-standard descriptors not present in the MCPD format. They facilitate an easy search for duplicates using ‘old’ genebank numbers. Furthermore, the locality information was divided over three fields, following the format of the IPD database (developed by the APIC group) to remain compatible with the IPD. It also facilitates data extraction for GIS software like DIVA. A separate field, primary based on the collector number, indicates redundancy within and between collections.

TOP 6 Oat / Beet
Since the database model of the European Avena (EADB) and International Beta Data Base (IDBB) are very similar, C.Germeier presented modules and function of both (see attachment 5).

TOP 7 Discussion and preparation of the afternoon session
The chairman remembered the participants that the ECP/GR has allocated 83,125 Euro to the SS&F crops network for funding of network activities in the period from 2004 to 2008. All planned activities will cost approximately 44,300 Euro. The NCG has discussed how to best use the remaining funds and finally suggested to use it for a small database development / improvement project. It was stressed that the NCG is responsible for

4 crops. The chair outlined two basically different project aims: a) the development of a tool that serves all five crop data base managers and b) a flax specific approach as this crop is currently requiring assistance most urgently. Funding of a specific technical project through the NCG budget would be new within the ECP/GR and probably raise controversial discussion within the Steering Committee. In view of the large gap between the many and often reiterated recommendations of network meetings and their implementation in practice the NCG sees currently more need for technical support than for meetings. The chair noted that the Steering Committee will probably take a decision based on political considerations and technical aspects. It is therefore very important to submit a proposal of a scientifically and technically convincing project that clearly supports EURISCO.

TOP 8 Prospects for joint developments

C. Germeier said that the option a) and b) do not reflect conflicting aims. A flax database project can also yield tools for other crop databases. The group agreed to submit a flax data base project and then discussed the financial frame. The budget would suffice to employ an IT-expert for example in the Czech Republic for one year. If tools developed by C. Germeier for automatic duplicate search or for uploading C&E data from Excel tables into the database are to be integrated in the IFDB, such decision would cause travel costs and costs for a stay of the IT expert at the BAZ, Braunschweig. Some of the European genebanks holding flax collections still need to computerise flax data. Potential data providers may need financial support (like the recently successfully completed VIR-BAZ oat project funded by the ECP/GR). The group agreed that the ECP/GR budget should be used for pump-priming and that additional funds can perhaps be acquired from the EU (for example Marie-Curie Foundation, if adequate calls are open) or other sources.

The group, considering the limited budget and time, also agreed to prioritise work on passport data, duplicate search and the establishment of an on-line searchable IFDB. It was suggested and accepted by the group to link, not integrate, C&E data with the IFDB for the time being. C&E data provided by the different data donors will be offered as downloadable Excel files containing the original data together with a detailed description of the experimental conditions and the applied methods. This approach allows integration at a later point of C&E data according the object model used by the CGN, the BAZ and several more institutions.

The group noticed that the GENOTYPDATA concept is very valuable, in particular for querying and displaying C&E data. During the project the data models of the GENOTYPDATA and EADB/IDBB should be compared and analysed with the idea to use the best problem solutions for the IFDB improvement. J. Nazkova supported to this suggestion in principle, but said that she would need feedback from the Institute.

TOP 9 Drafting of a project proposal and the report

The group finally developed elements of a project proposal.

A Passport data

A1 Compare passport data models (flax, hemp, oat/beet). Implement in mySQL as reference data model.

A2 Improve/adjust model for flax

A3 Update/add data to the IFDB and search for duplicates. Test search data base module already available at the BAZ, Braunschweig. The NCG was established with the understanding that co-operation between the crop networks would generate synergies. Testing of the module would allow debugging of the programme code and its further development as well as allow the flax database manager to identify duplicates in a more efficient way. This is synergy.

A4 Apply the module to the first 8,000 accessions (already in the database) with reference to already done manual duplicate search.

A5 Start further uploading accessions from 8,000 to 16,000 … 27,000. Considering possibility for uploading from EURISCO (16,000 accessions available in EURISCO). This work would generate experiences in using EURISCO as the future main data source for CCDBs. The ECP/GR information network as well as the Steering Committee will be provided with the results and recommendations.

B Characterisation and evaluation data

B1 Provide available open source solutions for C+E data documentation

B2 Collect data from donors in Excel or Access format. Describe within each experimental data set columns as table of contents and relate it to descriptor lists (IPGRI, UPOV, etc.). This keeps the option of future integration of the original, measured and/ or scaled data into the IFDB open.

B2 Analyse GENOTYPDATA as a model for C&E data. Currently the front end is programmed in VB and the database backend in Access (two tables in two databases: genotypes, descriptors)

Time schedule

- End of December 2005: Proposal is drafted by M. Pavelek and commented by all including the ECP/GR secretariat. Final version is ready.
- January 2006 or earlier: L. Frese informs the group how the proposal needs to be submitted and on the decision making process of the ECP/GR. It may be necessary to submit the proposal before the next Steering Committee meeting (March 2006?)
- Provided that the ECP/GR approves the proposal the project could start in September 2006 and end one year later in 2007 just before the end of Phase VII of the ECP/GR programme.

The meeting ended at 17:00 h. The chair thanked the participants for joining the group, their presentations and the constructive discussions.

L. Frese
Braunschweig, 08 August 2005