

**Minutes of the EURISCO Advisory Group and the ECP/GR
Documentation and Information Network Coordinating Group for the
preparation of a PGR documentation project**

Joint Meeting, 5 March 2004, Wageningen, The Netherlands
F. Menting, *compiler*

Participants

EURISCO Advisory Group:

Frank Begemann (begemann@zadi.de)
Eliseu Bettencourt (e.bettencourt@meganet.pt)
Dag Terje Endresen (dagterje@ngb.se)
Iva Faberova (Faberova@vurv.cz)
Samy Gaiji (s.gaiji@cgiar.org)
Theo van Hintum (Theo.vanhintum@wur.nl)
Veronique Jamilloux (jamilou@inapg.inra.fr)

ECP/GR Documentation and Information Network Coordinating Group:

Frank Begemann
Dag Terje Endresen
Theo van Hintum
Helmut Knüpffer (knupffer@ipk-gatersleben.de)
Ahmet Tan (aari@egenet.com.tr)
Ian Thomas (ian.thomas@bbsrc.ac.uk)
Anne Zanetto (Zanetto@bordeaux.inra.fr)
Dražen Jelovac (djelovac@mrizp.co.yu) SEEDNET representative

ECP/GR Secretariat:

Lorenzo Maggioni (L.maggioni@cgiar.org)

Local facilitator:

Frank Menting (frank.menting@wur.nl)

Introduction

On January 27th, 2004 a meeting was held in Bonn to brainstorm about a project submission in the field of PGR documentation to the call for proposals for the new EU GENRES programme expected for spring 2004. The output of this brainstorming session was a list of items that could serve as building blocks for a PGR documentation project (appendix C).

The objective of this meeting was to create, with a wider group of actors (EURISCO Advisory Group and ECP/GR Documentation and Information Network Coordinating Group), a consensus on collaborative actions that would be the basis for a project proposal to the EU GENRES programme.

The meeting draft agenda (appendix A), list of participants with their contact details (appendix B) and the final report of the EPGRIS project were circulated.

Theo van Hintum opened the meeting and recommended that the main output of this meeting should be a 'blueprint' for a follow up project for EPGRIS in the context of the EU GENRES programme. It was suggested to use the minutes of the 'Bonn meeting' as a starting point for discussion.

Theo van Hintum was elected chairman and Frank Menting would second him in taking notes during the meeting.

All participants briefly introduced themselves.

Theo van Hintum mentioned that he became sub program leader of the third Challenge Program “Unlocking Genetic Resources in Crops for the Resource-Poor” of the CGIAR. He gave a short overview of this sub-program “Genetic resource, genomic and crop information systems” and pointed that many common areas of work could be found with the follow-up project on EURISCO (see: www.genesforcrops.com).

Lorenzo Maggioni gave a short overview on developments in PGR documentation in Europe and the recent re-structuring of the Documentation and Information Network. He mentioned that the EPGRIS project and EURISCO were referred to in the ‘Proposal for a Council Regulation’ and stressed that there were good chances to get a new PGR documentation project funded. The participants that were invited for this meeting, the ECP/GR Documentation and Information Network Coordinating Group and the EURISCO Advisory Group, are the most relevant bodies to reach a consensus on the type of proposal that might be important for the region. The project development should benefit the entire PGR documentation community in Europe. He also stressed the importance of the participants’ commitment toward EURISCO as well as the role they play as members of their respective committees. He also informed the group of the on-going selection process for the replacement of Brigitte Laliberté.

Dražen Jelovac on behalf of SEEDNET gave a short overview of latest developments. SEEDNET consists of 9 partners in South Eastern Europe and is supported by SIDA. A Regional Interim Committee has successfully been formed to coordinate activities on PGR documentation. It was noted that SEEDNET is in its early phase of implementation and more time will be required for the design of full project proposals. Dražen will ensure that SEEDNET is fully informed of the latest development in EURISCO and in particular on the outcomes of this meeting.

Frank Begemann briefly described the GENRES programme and the relevance to EURISCO. He also stressed that, due to limited funds for PGR projects it is necessary to aim at a feasible PGR inventory project in the order of 300,000-400,000 Euro EU contribution. Eligible partners to such projects are not only limited to EU countries but also to International organisations such as IPGRI as well as institutes from countries not yet members of the EU. Non-EU based institutions would however not be eligible for funding and would need to rely on alternative resources (possibly ECP/GR).

Setting the scope for this meeting: what do we expect?

Focused on ‘inventory’ in EU GENRES 1467 follow-up programme and using the ‘Bonn minutes’ as a starting point, a clear set-up for the proposal should be generated with:

- Workpackages and their main elements and responsible institutions/individuals
- A coordinator of the project

Possible elements of a PGR inventory project that are too expensive for the present proposal might be considered for inclusion in crop-specific GENRES projects.

This meeting should be the starting point of a process in order to generate a project proposal under the EU GENRES programme. It was felt that the participants to the meeting are representing fairly the views and opinion in the region. It was also recognized that additional partners could be involved at a later stage. Therefore the project should be developed in a transparent process in communication with colleagues outside the present group.

Following a brief discussion on the functionalities of EURISCO, it was also felt that there were some necessary improvements to the current functioning of EURISCO. Eliseu Bettencourt recommended to address such requests for improvement to the existing EURISCO committee and that a team in SINGER is ready to further improve EURISCO. It was also suggested not to focus on the existing EURISCO and the necessary improvement but probably to focus on the new areas of developments.

The following areas of necessary/desirable developments were identified:

- Improvement of data flow within countries between the original data providers and the National Inventories
- Furthering the integration between EURISCO and ECCDBs
- Sustaining the funding support to EURISCO and related activities
- Improved definition of roles and responsibilities with a particular emphasis on ECCDB and EURISCO Focal points
- Improved data standard for EURISCO
- Upgrading of existing technologies
- Improved downloading functionality from EURISCO

The group agreed that the priority of the discussion should focus on the design of sound information flow strategy that builds on the existing EURISCO model. The National inventories nodes were recognized as the most important element. In consequence, it was agreed that all data from a national level would require, at least initially, to transit through the National Inventory acting as the responsible unit. It was also recommended to establish a more flexible data flow mechanism that would handle not only passport data but also characterization as well as evaluation data relevant to the ECCDB. Such approach would allow the ECCDB to strengthen their updating mechanism through the existing EURISCO infrastructure. Such cost-effective approach was also endorsed by ECCDB coordinators present at this meeting.

Such architecture integrating both ECCDB and EURISCO into a coherent data flow architecture would require:

- Commitment of data providers
- Reinforcement of the ECCDB managers (crop experts) and National Focal Persons in charge of the National Inventories
- Improvement of existing Data standards in particular to meet the additional needs of the ECCDB

There was consensus that (at least for the time being) the dataflow would go via the NIs, but that the flow might change in terms of the format and type of data (beside the current fixed format also allow XML based data that could include characterization and evaluation data). It was agreed that for the project proposal the data flow would follow the existing routes through the NI. Adjustments could be made in the future.

Detailed discussion of the 'Bonn minutes'

Consensus

1: Incorporation of in situ and on farm data.

Since the EURISCO descriptors are not suitable yet to handle in situ data, the latter should not be included at this stage. With the creation of suitable XML schemas or extended MCPD for EURISCO new opportunities will be explored. It was mentioned that PGR Forum is in the process of developing an information system covering *in situ* data and that EURISCO could, in the future, have a link to that information system.

2: Better integration between EURISCO – ECCDB.

This has been discussed earlier (Setting the scope for this meeting: what do we expect). It is expected that the project under the EU GENRES would lay the basis for such integration and complementarity.

3: Molecular markers.

The group felt that at this stage of knowledge, the focus should be directed toward the upgrade of EURISCO together with the ECCDB. Molecular data will be considered at a later stage. Other projects such as Germinate, and Challenge Programme are already engaged in this direction and the further collaboration with them would allow EURISCO to learn more about the optimum opportunities.

Explored project ideas

Starting the discussion on the XML schemas it became clear that there was some disagreement on the logical order on the project ideas in the Bonn minutes and how to fit them in a project. The division of the task into workpackages would require a more logical approach.

It was decided to create four 'phases' as the building blocks of the proposal. The explored project ideas of the Bonn minutes were reviewed and reorganized into this model (Table 1).

The four phases were:

- Data standardization
Multi crop and crop specific descriptors for Passport and C+E data, an extension of the limited EURISCO descriptors
- Data Schemas
XML: Increasing the flexibility and efficiency of data transfer between data source, NI, EURISCO, ECCDB, and data user.
Should incorporate established standards from GML and other standards
- Infrastructure upgrade
It was recognized that the existing EURISCO infrastructure is a key starting point. Such technical backbone to allow for data standardization and the use of schemas. Installing upgrading and testing of wrappers, web servers etc.
- Capacity building
Once developed, the technology should reach the PGR user community, via training to partners.

Within these four phases, it appeared possible to position most elements of the Bonn meeting outcomes in only four areas: (1) data exchange, (2) EURISCO/CCDB, (3) visualization, data quality and taxonomy and –(4) pilot crop portals (see table 1). By putting the ideas of the Bonn meeting in this schema, eight new elements could be identified. Below is the list of Bonn minutes items, indicating how they were regrouped into these elements.

- 1) XML schemas for EURISCO, DS, NI, ECCDB (shared under phase 'Schemas' in column 'Data exchange')
An XML schema is necessary to allow for efficient data exchange of all PGR passport descriptors and C+E data. An inventory must be made on activities in this field to avoid double work (ABCD schema, Challenge Programme). Such a schema must be based on new data standards of EURISCO and is therefore dependent on the outcome of this phase. These new data standards should be developed with involvement of all relevant actors, probably coordinated by IPGRI (similar to the development of the MCPD list).
The actual creation of schemas can be carried out by a project participant or outsourced.
The technical part of this issue (implementation of the schemas and their interfaces) was shared under 'Infrastructure', column 'Data exchange'.
Revision and extension of MCPD: 1 to N links, timestamp and unique identifier for each data source became part of phase 'Standardization', column 'Data exchange' (Passport MCPD).
- 2) Transfer of common data EURISCO-ECCDB
These issues were categorized in the phase 'Infrastructure'. Note: the term 'atomicity' of data was incorrectly used in the 'Bonn minutes'. Atomisation of data is the splitting of values to their smallest unit, resulting in "atomicity". E.g. a taxon into the fields genus, species and sub species, and separate fields for their authors.

The timestamp issue is transferred to the phase 'Standardization', column 'Data exchange' (Passport MCPD). The timestamp is given by the data source when uploading to EURISCO.

- 3) Metadata of characterization and evaluation data
The issues on the structure and description of the meta- and actual data were categorized in the phase 'Standardization', C+E.
The technical implementation and creation of schemas for exchange of C+E data were placed under 'Infrastructure' and 'Standardization' respectively.
In the development of new standards the already available IPGRI crop descriptor lists will play a critical role.
- 4) Taxonomic backbone
This issue was merged with bullets 6 (Visualization tools) and 7 (Improvement of data quality) to 'Visualization, data quality and taxonomy'.
Standards for taxonomy like GRIN-tax or Mansfeld do not cover all species in EURISCO. The taxonomic reference list can be the starting point for the creation of a taxonomic name service for PGR. Spelling and synonymy of scientific names should be checked using the IPGRI Taxonomy Nomenclature Checker (based alternatively on names from GRIN-tax or Mansfeld). In order to explore possibilities to create a taxonomic reference list for EURISCO, and to define functions and structures of an appropriate web service, a meeting was proposed.
- 5) Prototype of crop portals
Part of the phase 'Infrastructure'. The most important characteristic of a crop portal is the focus on the crop-specific user demands. This differs per crop thus a standard model for a crop portal is probably not possible. However, one of the main tools of each crop portal is access to relevant germplasm databases, for which generic, crop-independent (or "multi-crop") solutions can be developed. CCDB managers, with their crop knowledge will play an important role in creating crop portals. CCDB managers will be stimulated to build their own crop portal (capacity building). Although it was suggested to use the AEGIS crops for the pilot project, no decision was made yet.
In the formulation of the actual project proposal this activity should be described as the final link between the inventories and the user community (for reasons of eligibility).
- 6) Visualization tools
This issue was merged with bullets 4 (Taxonomic backbone) and 7 (Improvement of data quality) to 'Visualization, data quality and taxonomy'.
Visualization requires standardization. Geographic data are standardized with the fields latitude and longitude in MCPD. Methods for standardization and visualization of ancestral data or gene pool structure of collections must be explored. Visualization tools can be used as validation tools, e.g. for geographic data. SINGER already created Web-GIS as a visualization tool for EURISCO geographic data. This expertise could also be used to create visualisation tools for ancestral data or collection structure.
- 7) Improvement of data quality
This issue was merged with bullets 4 (Taxonomic backbone) and 6 (Visualization tools) to 'Visualization, data quality and taxonomy'.
A tool for validation of taxonomic data (IPGRI/GRIN Taxonomic Nomenclature Checker) is already available. Ways how to make this tool compatible with a taxonomic reference list and a taxonomic name service for EURISCO must be explored.
Gazetteer based tools for validation of geographic data are available, both as freeware as on the commercial market. A survey must be made to assess the utility and quality of these gazetteers.
Feedback mechanisms to ensure that data improvements suggested by EURISCO and/or CCDBs are reported to the original data provider must become part of the infrastructure.
- 8) Interfaces of EURISCO with other international information systems and developments.
There are many ways to provide interfaces with other PGR documentation systems. Possibilities in this field must be explored.

Making EURISCO a GBIF data source was considered important, and the linking of EURISCO to GBIF should be considered as a part of the project. However if possible, EURISCO should be made a GBIF data source as soon as possible (relatively small activity to be initiated by the technical EURISCO management at IPGRI)¹.

Coordination

Among the participants, nobody was volunteering to coordinate the project proposal. It was agreed that Ian Thomas would check if he would get support for writing and coordinating the project. If not, Frank Begemann would check in Germany for a coordinator, and Ann would find out if INRA could supply someone to accept this responsibility².

Closing of the meeting

The chairman thanked ECP/GR for the support in organizing the meeting, and the participants for participating, and closed the meeting.

¹ Helmut approached the EURISCO Steering Committee for permission to assist EURISCO in becoming a GBIF data provider using the ABCD schema and the BioCASE protocol. For time reasons, this topic was not discussed. The MoU of EURISCO allows to freely distribute EURISCO data. Helmut will send another mail to the EURISCO Steering Committee for negotiation.

² In the meantime, shortly after this meeting IGER has agreed to co-ordinate the further work and the project itself.

Table 1: structure of the proposal.

	Data exchange				EURISCO- CCDB	Visualization, data quality and taxonomy	Pilot crop portals
Standardization	passport		C+E		Time stamp	-Reference list for taxonomy -Standards in order to visualize geographic, ancestral and collection structure information	
	MCPD	Crop spec.	Meta	Crop spec.			
Schemas	Create XML schemas for data exchange based on standards.					-GML schema for visualization tools	
Infrastructure	Testing and implementation of schemas and interfaces Mechanism for feedback on data quality to DS				EURISCO download	-Name service for taxonomy -Visualization tools in EURISCO	Create a pilot crop portal
Capacity building	Agree on standards				NI CCDB managers??		Stimulating CCDB managers

Table 2: Workpackages extracted from the structure of the proposal with their possible candidate workpackage leaders.

Workpackage	Description	Candidates for WP leaders
WP1	Data standardization of passport data	IPK, IGER, (IPGRI)
WP2	Data standardization of C+E data	INRA, BAZ (IPGRI)
WP3	XML schema and infrastructure	IPGRI, NGB (BRG)
WP4	Capacity building	IPGRI, NGB
WP5	Pilot Crop Portals	CGN, RICP (BRG)
WP6 ³	Visualization, Data quality and Taxonomy	ZADI, IPK, CGN, IPGRI (BRG)
WP7	Project Coordination	IGER (see comment above)
WP7a	External linkages	CGN, IPGRI

³ The possibility of splitting the heterogeneous WP6 into separate (sub) WPs on Geographical data (WP6a), Taxonomy (WP6b) and Collection structure (WP6c) with separate WP leaders was discussed and also considered feasible.

APPENDIX A: Draft agenda of the meeting.

**Meeting of the EURISCO Advisory Group and the ECP/GR Documentation
and Information Network Coordinating Group for the preparation of a PGR
documentation project
Joint Meeting, 5 March 2004, Wageningen, The Netherlands**

Agenda

Friday March 5th 9:00 - 17:30

<i>Time</i>	<i>Topics</i>
9:00 – 10:45	<ul style="list-style-type: none">• <i>Opening of the meeting, welcome, election chairperson</i>• <i>Introduction participants</i>• <i>Discussion expected outcome of the meeting</i><ul style="list-style-type: none">• <i>elements project(s)</i>• <i>consortium</i>• <i>project preparation</i>• <i>Brief introduction 'Minutes informal Bonn Meeting'</i>
10:45 – 11:15	Coffee break
11:15 – 12:30	<ul style="list-style-type: none">• <i>Detailed discussion of the 'Minutes informal Bonn Meeting'</i><ul style="list-style-type: none">• <i>General set up of project</i>• <i>Elements to be added a-priori</i>• <i>Discussion individual elements</i>
12:30 – 13:30	Lunch (together with some CGN colleagues)
14:00 – 15:20	<ul style="list-style-type: none">• <i>Detailed discussion of the 'Minutes informal Bonn Meeting' (continued)</i>
15:20 – 15:50	Coffee break
15:50 – 17:00	<ul style="list-style-type: none">• <i>Detailed discussion of the 'Minutes informal Bonn Meeting' (continued)</i>• <i>Next steps towards 1467 documentation project(s)</i>• <i>Closing of the meeting</i>
19:00 onwards	Informal dinner

APPENDIX B: List of participants with their contact details.

Frank Begemann

ZADI/IBV
Information Center for Biological Diversity
Villichgasse 17
53177 Bonn
Germany
Tel: (49-228) 9548202
Fax: (49-228) 9548220
Email: begemann@zadi.de

Eliseu Bettencourt

Dept. of Genetic Resources and Breeding
Estação Agronómica Nacional EAN/INIAP
Quinta do Marquês
2784-505 Oeiras
Portugal
Tel: (351-21) 4403688
Fax: (351-21) 4416011
Email: e.bettencourt@meganet.pt

Dag Terje Endresen

Nordic Gene Bank
PO Box 41
23053 Alnarp
Sweden
Tel: (46-40) 536654
Fax: (46-40) 536650
Email: dagterje@ngb.se

Iva Faberová

Research Institute of Crop Production
(RICP)
Drnovska 507
16106 Praha 6 – Ruzyne
Czech Republic
Tel: (420) 233022478
Fax: (420) 233022286
Email: Faberova@vurv.cz

Samy Gaiji

CGIAR System-wide Genetic Resources
Programme (SGRP)
International Plant Genetic Resources
Institute (IPGRI)
Via dei Tre Denari 472/a
00057 Maccarese (Fiumicino)
Italy
Tel: (39) 06 6118 235
Fax: (39) 06 61979661
Email: s.gaiji@cgiar.org

Theo J.L. van Hintum

Centre for Genetic Resources
The Netherlands (CGN)
P.O. Box 16
67001 AA Wageningen
The Netherlands
Tel: (31-317) 477078
Fax: (31-317) 418094
Email: Theo.vanhintum@wur.nl

Veronique Jamilloux

Bureau des Ressources Génétiques
16 Rue Cl. Bernard
75231 Paris cedex
France
Tel: (33) 1 44087264
Fax: (33) 1 44087263
Email: jamillou@inapg.inra.fr

Dražen Jelovac

Maize Research Institute “Zemun Polje”
Slobodana Bajica, 1
11080 Zemun
Serbia and Montenegro
Tel: (381-11) 3756704
Fax: (381-11) 3756707
Email: djelovac@mrizp.co.yu

Helmut Knüpfner

Institut für Pflanzengenetik und
Kulturpflanzenforschung - Genbank
Corrensstrasse 3
06466 Gatersleben
Germany
Tel: (49-39482) 5283
Fax: (49-39482) 5155
Email: knuepfer@ipk-gatersleben.de

Lorenzo Maggioni

ECP/GR Coordinator
Regional Office for Europe
International Plant Genetic Resources
Institute (IPGRI)
Via dei Tre Denari 472/a
00057 Maccarese (Fiumicino)
Italy
Tel: (39) 06 6118 231
Fax: (39) 06 61979661
Email: l.maggioni@cgiar.org

Frank Menting

Centre for Genetic Resources
The Netherlands (CGN)
P.O. Box 16
67001 AA Wageningen
The Netherlands
Tel: (31-317) 477083
Fax: (31-317) 418094
Email: frank.menting@wur.nl

Ahmet Semsettin Tan

Aegean Agricultural Research Institute,
AARI
P.O.Box 9
Menemen
35661 Izmir
Turkey
Tel: (90-232) 8461331
Fax: (90-232) 8461107
Email: aari@egenet.com.tr

Ian D. Thomas

Genetic Resources Unit
Institute of Grassland and
Environmental Research (IGER)
Plas Gogerddan
Aberystwyth, Ceredigion SY23 3EB
United Kingdom
Tel: (44-1970) 828255/2131
Fax: (44-1970) 828357
Email: ian.thomas@bbsrc.ac.uk

Anne Zanetto

INRA, Unité de Recherches sur les
Espèces Fruitières et la Vigne
BP 81
33883 Villenave d'Ornon
France
Tel: (33) 5 57122445
Fax: (33) 5 57122439
Email: Zanetto@bordeaux.inra.fr

APPENDIX C: Minutes of the Bonn meeting.

Tools and Services to Improve the Existing PGR Inventory Infrastructure in Europe

Minutes of an informal meeting for exploring project ideas for EU 1467

Date: 27 January, 2004
Place: ZADI, Bonn, Germany

Participants:

Dag Terje Endresen, NGB, Alnarp, SE (dagterje@ngb.se)
Daniel Jimenez-Krause, ZADI, Bonn, DE (krause@zadi.de)
Frank Begemann, ZADI, Bonn, DE (begemann@zadi.de)
Helmut Knüpffer, IPK, Gatersleben, DE (knupffer@ipk-gatersleben.de)
Lothar Frese, BAZ, Braunschweig, DE (lothar.frese@fal.de)
Sabine Roscher, ZADI, Bonn, DE (roscher@zadi.de)
Siegfried Harrer, ZADI, Bonn, DE (harrer@zadi.de)
Theo van Hintum, PRI/CGN, Wageningen, NL (theo.vanhintum@wur.nl)

Introduction

The aim of the meeting was brainstorming for a possible European project submission in the field of PGR documentation to the call for proposals for the new EU GENRES programme (see <http://www.ecpgr.cgiar.org>), expected for spring 2004. Under the heading "Inventory", a project could be submitted on the basis of EURISCO and the European documentation infrastructure developed under EPGRIS.

Frank gave a brief overview to the frame of the EU 1467 follow-up programme (simply referred to as EU 1467 in this document), with emphasis on the fact that there will be relatively little money available and thus any project proposal must be based on realistic and achievable goals.

Theo gave a brief overview on existing international efforts in the area of PGR informatics like GeneMine, Germinate and the Challenge Program 'Unlocking Genetic Diversity in Crops for the Resource-Poor'

Agreement: today's brainstorming focuses on the inventory context of EU 1467.

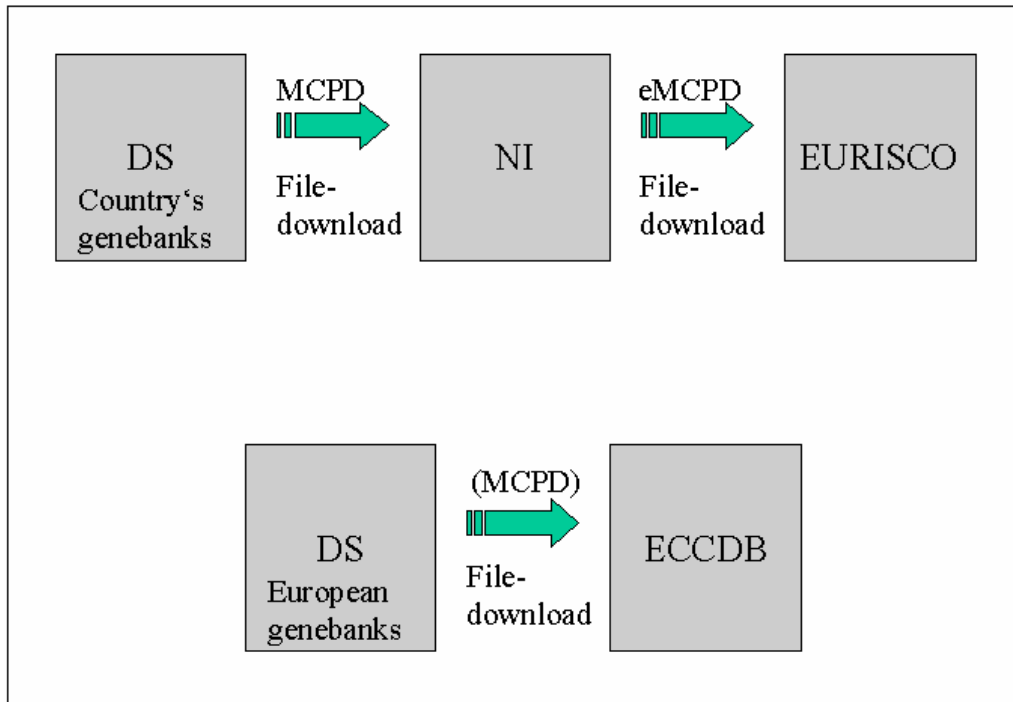
Helmut mentioned the recent call for GBIF project proposals (www.gbif.org) (seed money for projects aimed at making biodiversity collection and name-based data freely accessible to the GBIF network) and suggested that EU and GBIF funds might be considered as complementing each other.

Agreement: it is too early to incorporate in-situ/on farm issues (data/functionality) to the existing PGR inventories, thus no respective project ideas could be made in this meeting. The developments of PGR Forum shall be tracked for potential fields of action to be grasped for the 2nd call for proposals of EU 1467.

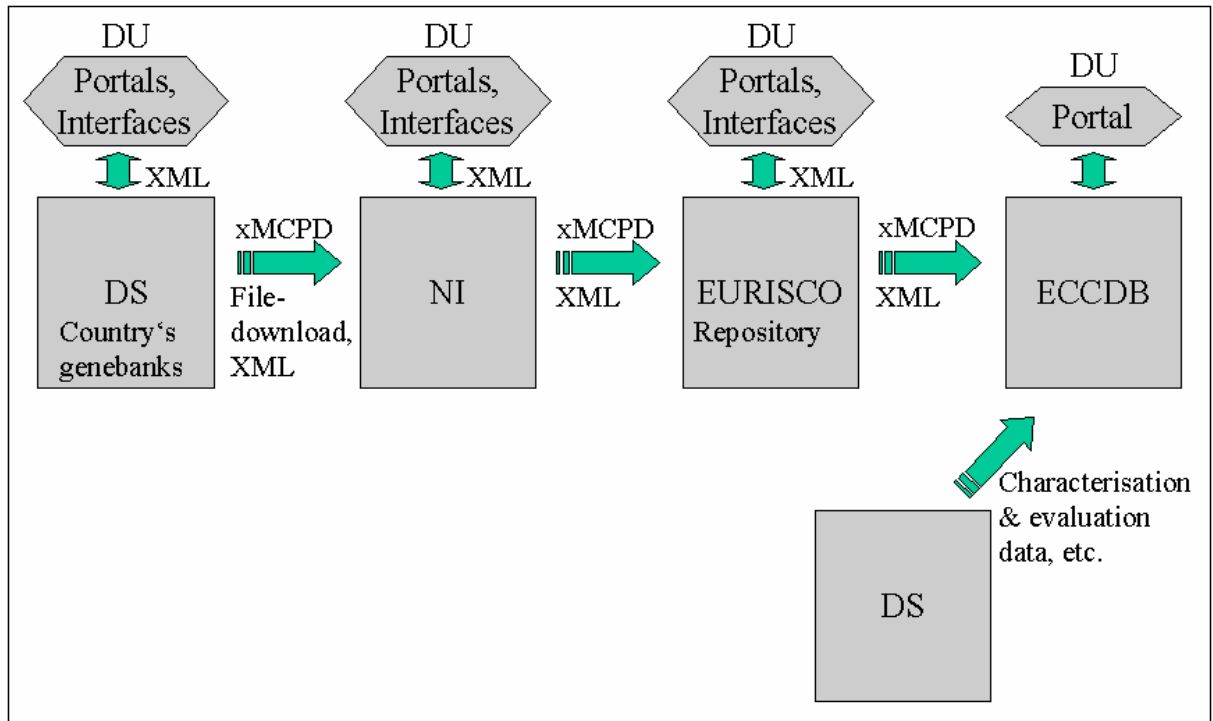
Agreement: there is an overlap between EURISCO - ECCDB. Further development of both pillars of PGR documentation in Europe should therefore seek and exploit integration and synergies.

Agreement: the topic of molecular marker data is intentionally left out of the discussion, due to unresolved questions (standardisation, comparability), and the expected high costs involved.

Based on those agreements the discussion started. In order to facilitate the exploration process, the data flows from and to EURISCO and the ECCDBs were visualised.



Where we stay. Today's data flows towards EURISCO and ECCDBs



Where we want to go. The drawing above represents a realistic and sound data flow scenario for EURISCO and the ECCDBs for the next future (3 or 4 years).

Abbreviations used in this document:

- DS** Data Source (in the context of EURISCO, European genebanks are meant)
- NI** National Inventory (EURISCO)
- DU** Data User (can be an end user, an application or another information facility...)
- MCPD** Multi-Crop Passport Descriptor List
- eMCPD** EURISCO's present extension of the Multi-Crop Passport Descriptor List
- xMCPD** Extended future Multi-Crop Passport Descriptor List

Summary of explored project ideas

1) XML Schemas for EURISCO, DS, NI, ECCDBs

Communication via XML aims at increasing the efficiency and speed of data transfer (e.g. automated download and actualisation between DS, NI, EURISCO, ECCDB, DU...).

Issue: Development of XML schemas and respective interfaces (e.g. wrappers). It needs more detailed specification which concrete schemas and respective interfaces shall be implemented in the next future. A few examples were discussed:

- Implementation of GML-compliant interface to EURISCO in order to allow the visualisation of EURISCO data with GIS applications,
- Implementation of MCPD XML schema and respective interfaces for future data transfer between the NIs and EURISCO, between the DS and NI (or directly from DS to EURISCO), between EURISCO and the ECCDBs. Questions:
 - implementation of an ABCD⁴-compliant schema for compatibility with GBIF (avoiding the double work of having a separate schema/interface developed),

⁴ ABCD = Access to Biological Collections Data (<http://www.bgbm.org/TDWG/CODATA/Schema/>), one of the two GBIF standards for exchange of unit-level data (biodiversity collections, species occurrence records). Helmut mentioned that he is developing a mapping of the EURISCO descriptors to ABCD, together with Walter Berendsohn from Berlin Botanical Garden, and an extension of ABCD to include PGR-specific data elements will be released soon.

- Revision and extension of MCPD will be needed. Two concrete potential/desired extensions were discussed:
 - allow 1-to-N links respective description to an accession record (Today's schema allow only 1 link, namely to the accession's details at the data source).
 - an obligatory unique identifier for each data source in the framework of EURISCO and ECCDB, as well as a timestamp field, in order to unequivocally identify an accession record and its date of creation/actualisation. This will ensure that the data collecting instances (NIs, EURISCO, ECCDBs) only have one (i.e. the most current) version of each accession record. In the future, an accession record shall be edited only at the data source (atomicity⁵ of accession data).
- Implementation of XML schema and respective interfaces for characterisation and evaluation (meta-)data (see Top 3),
- Implementation of XML schemas and respective interfaces for connectivity with international PGR information systems (see Top 8),
- Improvement of data validation mechanisms and their efficiency.

Based on existing knowledge, existing schemas and structures (e.g. MCPD) and well-known standards and methods (XML, GML, wrappers).

2) Transfer of common data EURISCO-ECCDB

Issues:

- A mechanism be established to import passport data (xMCPD) for the ECCDBs from EURISCO, in order to avoid inconsistent overlapping of data (passport data in EURISCO and in the ECCDBs obtained by different ways, different scopes, at different times).
- In order to achieve this, the atomicity of each accession data record that leaves a DS must be warranted. This can be handled by extending MCPD by corresponding IDs and a timestamp (see Top 1).
- ECCDB update mechanisms have to ensure that additional value (e.g. linking with C+E data, determination of groups of duplicates, etc.) will not be lost when an institution's or national inventory's data set is being replaced by an update.

3) Metadata of characterisation and evaluation data

Issues:

- Extension of EURISCO's scope on characterisation and evaluation (meta-) data.
- How to structure descriptive data, what to describe, who/where describes.
- Whether meta-data extracted from raw data or separately collected.
- How to structure the actual data (raw data, summarised or standardised per experiment)

Based on existing expertise (like e.g. at the TDWG working group for Structure of Descriptive Data SDD, <http://160.45.63.11/Projects/TDWG-SDD/>) :

- definition of a metadata XML schema for characterisation and evaluation data,
- implementation of a corresponding upload interface for EURISCO,
- experiment with XML schema and implementation for actual data.

Agreement: although all participants see an urgent need in the development of efficient methods for storage, management, analysis and visualisation of characterisation and evaluation data, and although there is much theoretical and even some practical work done in this area, the activities in this area (in the "inventory" frame of EU 1467) should be limited to developing pilots/prototypes. Nevertheless, further ideas need to be discussed in order to grasp this crucial issue in the "crop" frame or as a concerted action of EU 1467.

⁵ Atomic data in this context meaning that an accession data record shall never be edited nor changed in a site other than the data source. Unequivocal identification of accession data records respective their timeliness will thus play a major role in a data flow model ensuring data atomicity.

4) Taxonomic backbone

Issues:

- Implementation of a taxonomic reference list, to be used in EURISCO, and possibly also for standardisation purposes
- Implementation of a taxonomic name service for EURISCO (matching, synonyms, common names in European languages).

Based on existing international standards for taxonomy like GRIN and/or Mansfeld, and experiences with name services like e.g. in BIGTax. A taxonomic reference list for EURISCO would need to be extended, since not all species represented in European genebanks are covered by the mentioned taxonomic databases.

5) Prototype of crop portals

Issues:

- Development of some pilot portals for selected crops. Aiming at improving the usability and attractiveness of ECCDB data, and providing the PGR user access to all relevant information and services, obviously not limited to Europe
- Testing the interaction with EURISCO such as the on-line extraction of the needed passport data.

Which crops to be proposed is still to be discussed: e.g. AEGIS crops and/or crops with a particular European importance which are not covered by major CGIAR.

Based on experiences and results from well known crop portals like that for rice at IRRI (<http://www.irri.org/>).

6) Visualisation tools

Issues:

Development of visualisation tools and implementation in EURISCO and the ECCDBs for

- Accessing the hierarchical structure of the gene pools queried (e.g. <http://www.cs.umd.edu/hcil/treemap/> as well as the gene pool structuring experience with the *Lactuca* collection at CGN)
- GIS tools.

In the development of these tools the relevance to the user (end-user or curator) will have the central focus.

7) Improvement of data quality

Issues:

- Development/adaptation of tools to help DS to improve their data quality :
 - Tool for validation of taxonomic data (spell checker). Further development of the already existing IPGRI/GRIN Taxonomic Nomenclature Checker <http://pgrdoc.ipgri.cgiar.org/taxcheck/grin/> to make it compatible with the taxonomic backbone(s).
 - Gazetteer-based tool for validation of geographic data (e.g. longitude and latitude), to e.g. prove the validity of coordinates and to allow the capture of coordinates from existing country and location data.
 - Implementation of feedback mechanisms that ensure that improvements proposed by ECCDBs and EURISCO be reported back to the original data provider (possibly via the Nis).

Based on existing tools and experiences in this field.

8) Interfaces of EURISCO with other international information systems and developments

Issues:

- To provide interfaces to SINGER and other CGIAR information systems, GRIN and other PGR information systems in order to ensure connectivity for a future global PGR information system
- To provide an ABCD-compliant XML interface of the MCPD and become so the main European PGR data source for GBIF

Based on appropriate experiences at SINGER etc, and on existing expertise in mapping the MCPD to ABCD (IPK Gatersleben).

Conclusions

The group felt that this list of items provides a good basis for the development of a PGR documentation project. For this project the items could be transformed into workpackages after adding and rearranging them. Therefore this report can be seen as the basis of a discussion in a wider group of actors within the ECP/GR documentation and information network. Since the issues all circle around EURISCO and since the EPGRIS team (now EURISCO Steering Committee) proved to be able to create a viable proposal, it is suggested that the next step would be to discuss this paper in this forum.

Task that cannot be included into an EU 1467 proposal remain on the to-do list of the European PGR documentation community.

Theo indicated he would be willing to host a meeting early March in Wageningen.