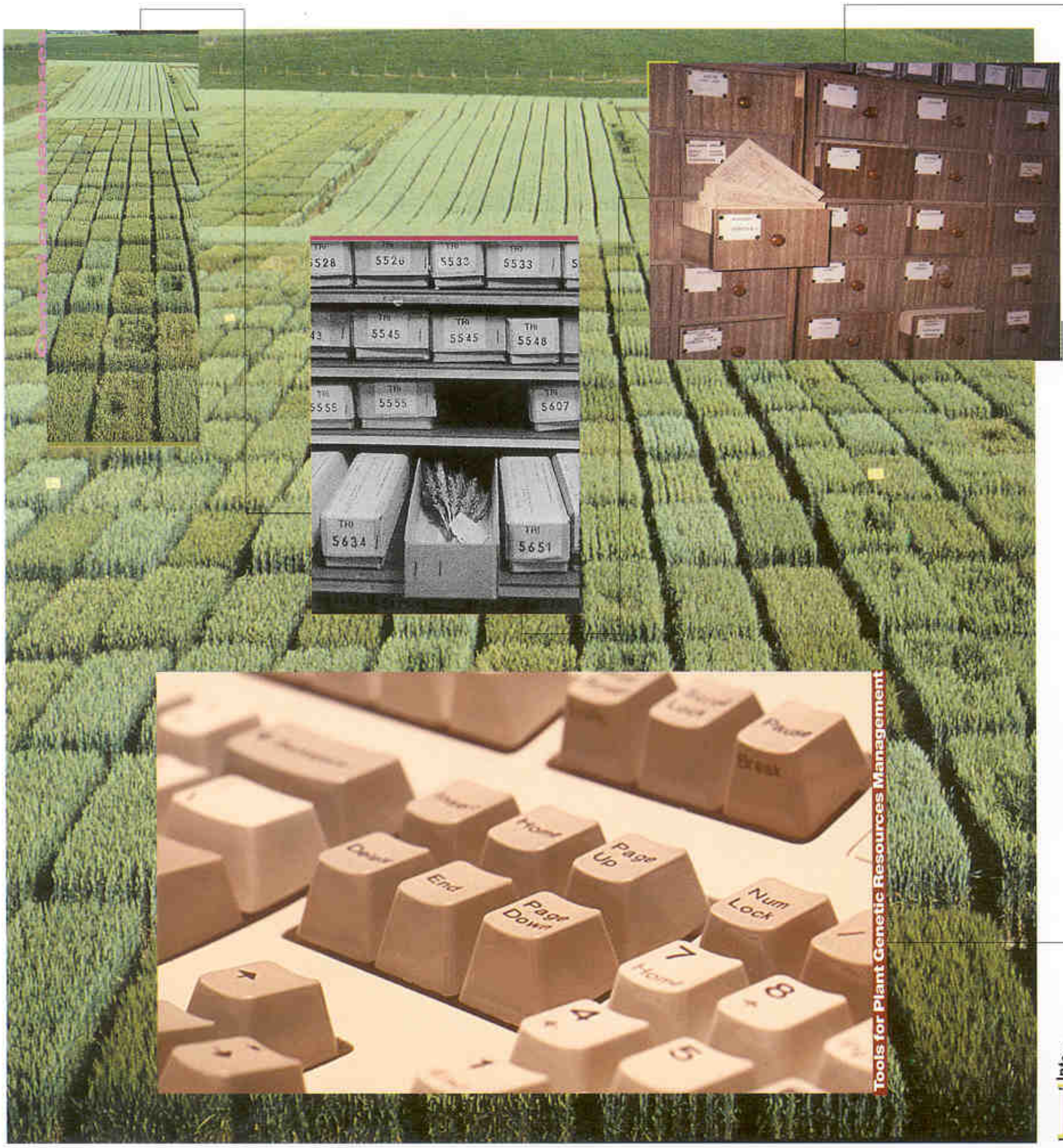
EUROPEAN COOPERATIVE PROGRAMME FOR CROP GENETIC RESOURCES NETWORKS (ECP/GR)

Central Crop Databases: Tools for Plant Genetic Resources Management

ECP/GR

Report of a Workshop, 13-16 October 1996, Budapest, Hungary

E. Lipman, M.W.M Jongen, Th.J.L. van Hintum, T. Gass *and* **L. Maggioni,** *compilers*



cpro - dlo
cgn

International Plant Genetic Resources Institute
IPGRI

The International Plant Genetic Resources Institute (IPGRI) is an autonomous international scientific organization operating under the aegis of the Consultative Group on International Agricultural Research (CGIAR). The international status of IPGRI is conferred under an Establishment Agreement which, by January 1997, had been signed by the Governments of Australia, Belgium, Benin, Bolivia, Brazil, Burkina Faso, Cameroon, Chile, China, Congo, Costa Rica, Côte d'Ivoire, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, Greece, Guinea, Hungary, India, Indonesia, Iran, Israel, Italy, Jordan, Kenya, Malaysia, Mauritania, Morocco, Pakistan, Panama, Peru, Poland, Portugal, Romania, Russia, Senegal, Slovak Republic, Sudan, Switzerland, Syria, Tunisia, Turkey, Uganda and Ukraine. IPGRI's mandate is to advance the conservation and use of plant genetic resources for the benefit of present and future generations. IPGRI works in partnership with other organizations, undertaking research, training and the provision of scientific and technical advice and information, and has a particularly strong programme link with the Food and Agriculture Organization of the United Nations. Financial support for the research agenda of IPGRI is provided by the Governments of Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, India, Italy, Japan, the Republic of Korea, Luxembourg, Mexico, the Netherlands, Norway, the Philippines, Spain, Sweden, Switzerland, the UK and the USA, and by the Asian Development Bank, CTA, European Union, IDRC, IFAD, Interamerican Development Bank, UNDP and the World Bank.

The European Cooperative Programme for Crop Genetic Resources Networks (ECP/GR) is a collaborative programme among most European countries aimed at ensuring the long-term conservation and facilitating the increased utilization of plant genetic resources in Europe. The Programme, which is entirely financed by the participating countries and is coordinated by IPGRI, is overseen by a Steering Committee (previously Technical Consultative Committee, TCC) composed of National Coordinators nominated by the participating countries and a number of relevant international bodies. The Programme operates through ten broadly focused networks in which activities are carried out through a number of permanent working groups or through ad hoc actions. The ECP/GR networks deal with either groups of crops (cereals, forages, vegetables, grain legumes, fruit, minor crops, industrial crops and potato) or general themes related to plant genetic resources (documentation and information, *in situ* and on-farm conservation, technical cooperation). Members of the working groups and other scientists from participating countries carry out an agreed workplan with their own resources as inputs in kind to the Programme.

The Centre for Genetic Resources, The Netherlands (CGN) is a part of the Centre for Plant Breeding and Reproduction Research (CPRO-DLO), with a separate budget and programme. CPRO-DLO is part of the research organization of the Netherlands Ministry of Agriculture, Nature Management and Fisheries (DLO).

Citation:
Lipman, E., M.W.M Jongen, Th.J.L. van Hintum, T. Gass and L. Maggioni, compilers. 1997. Central Crop Databases: Tools for Plant Genetic Resources Management. International Plant Genetic Resources Institute, Rome, Italy/CGN, Wageningen, The Netherlands.

ISBN 92-9043-320-5

IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy

Centre for Plant Breeding and Reproduction Research (CPRO-DLO)
Centre for Genetic Resources The Netherlands (CGN)
PO Box 16, 6700 AA Wageningen
The Netherlands

Contents

Part A. Discussions and Recommendations

Opening and welcome

The Workshop was opened by Dr László Holly, Director of the Institute of Agrobotany, Tápiószele who welcomed the participants on behalf of the Deputy Secretary of State for Agriculture, H.E. Lajos Buzássy.

Dr Holly emphasized the importance of this joint Workshop which is probably the largest one held in Europe on the documentation of plant genetic resources. He mentioned the value of the EGDS Project for the improvement of documentation systems in eastern Europe, not least for the Hungarian plant genetic resources programme.

Dr Thomas Gass, Director of IPGRI's Regional Office for Europe, welcomed the participants. He noted that the present Workshop is not a first initiative on plant genetic resources documentation but should be seen as the continuation of numerous previous initiatives.[1] Nevertheless, recent international events have given much visibility to plant genetic resources activities. On 23 June 1996 the world community adopted a Global Plan of Action on conservation and utilization of plant genetic resources. Effective documentation of these resources is a pillar of this Global Plan of Action. It is essential to ensure the coordination and rationalization of conservation activities in a well-informed way. It is also essential to the usefulness of plant genetic resources.

In recognition of this fact the Steering Committee of ECP/GR established a 'Documentation and Information Network' in September 1995. The present, first meeting within this network will allow ECP/GR to step into the next century. It is ultimately the European Crop Databases that will determine the usefulness and therefore the support given to plant genetic resources conservation.

Dr Theo van Hintum welcomed the participants and mentioned his pleasure in seeing that this third EGDS (Eastern European Germplasm Documentation Systems Project) Workshop will be very comprehensive. He stressed that in this meeting nearly all European plant genetic resources documentation specialists were present, showing the true integration of the EGDS network with the rest of Europe. He expressed his hope that during the meeting the participants would use the opportunity to establish links between Eastern and Western Europe on the personal level. The institutional level then follows automatically. Dr van Hintum then presented the agenda of the Workshop.

The session started with a report given by Ir Marcel Jongen, coordinator of the EGDS project, who briefly outlined the project history and achievements and gave

---

[1] Important milestones were:

- the ECP/GR Workshop on exchange of information, 23-25 Oct. 1984, Radzików, Poland;
- the EUCARPIA/IBPGR Symposium on Crop Networks, 3-6 Dec. 1990, Wageningen, The Netherlands;
- the Joint FAO/IPGRI Workshop on *ex situ* germplasm conservation, 7-9 Oct. 1993, Prague, Czech Republic;
- the Technical Meetings of Focal Points for Documentation in East European genebanks: First Meeting, 17-21 Oct. 1994, Prague, Czech Republic; Second Meeting, (Standardization in plant genetic resources documentation), 10-14 Oct. 1995, Radzików, Poland.

a summary of the working session held by the focal points of the project on 12 October 1996.

The FAO representative, Dr Jerzy Serwinski, then reported on the Technical meeting on the methodology of the FAO World Information and Early Warning System on plant genetic resources (WIEWS), held 30 September-3 October 1996 in Radzików, Poland.

Both these reports, and the presentations summarized below, are given in full in the section **Presentations and Other Contributions**.

Summary of the theoretical presentations

Introduction

Central Crop Databases – an overview

Theo van Hintum introduced the subject of the Workshop by presenting definitions and objectives of Central Crop Databases (CCDBs). He gave an overview of the current issues and of the foreseeable and desirable evolution of documentation systems for plant genetic resources in Europe. He identified the following causes for the presently limited use of the databases: low quality of the databases, low accessibility and lack of capacity for analyzing the databases. He raised fundamental issues such as the principles for the division of tasks between the partners involved in the establishment and management of CCDBs (e.g. should data be standardized before being sent to the database manager or standardized by the database manager?), and the 'snapshot versus monitor' options that can be chosen according to the needs and resources available for database updating.

Organizational aspects

The role of Central Crop Databases in ECP/GR

In his presentation, Thomas Gass emphasized the key role played by documentation systems in regional networks. He showed how, within the European Cooperative Programme for Crop Genetic Resources Networks (ECP/GR), Central Crop Databases had allowed the working groups to take informed decisions regarding the safety-duplication of unique material, the need for filling gaps through additional collecting activities or the implementation of joint projects such as the establishment of core collections or the collaborative characterization of genepools. Within ECP/GR, 32 Central Crop Databases are currently managed as inputs in kind by a total of 29 institutes from 18 countries. These cover a wide range of forages and cereals, fruit trees of the *Prunus* and *Malus* genera as well as the *Brassica* and *Allium* genepools. While a European *Pisum* database has existed since the early 1980s, CCDBs have recently been initiated for the other grain legumes of importance in Europe. A number of databases have widened their scope to record collections from other continents (e.g. the International Database for *Beta*).

Methodology of database compilation

Standardization of Central Crop Databases

Morten Hulden presented the experience of the Nordic Gene Bank (NGB) in the documentation of collections. The establishment of the centralized databases involved the compilation of a large number of databases received from the participating countries (over 150 individual files) and therefore standardization of data was an absolute necessity. M. Hulden explained how standardization can take place at different levels (descriptors, field structure, table structure, database structure, botanical nomenclature and taxonomy). He expanded on the interpretation made by NGB of the notion of culton, which allows, among others, the grouping of crops within taxa, the description of modern cultivars and the creation of summary records for groups of accessions.

Multicrop passport descriptors

The Second Technical Meeting for the Focal Points of the EGDS project held in Radzików, Poland, October 1995, had strongly recommended that FAO and IPGRI produce a standard list of multicrop passport descriptors to facilitate exchange of data between databases on crop genetic resources. Tom Hazekamp and Jerzy Serwinski welcomed the opportunity of this workshop to present the resulting draft list elaborated in common and to receive feedback from the workshop participants. The first section lists the most essential passport descriptors which are important for all crops, and a second section contains additional descriptors proposed for use with the FAO World Information and Early Warning System on plant genetic resources (WIEWS).

Methodology of database utilization

Handling evaluation data in Central Crop Databases

Lajos Horváth highlighted the difficulties encountered in the integration of evaluation data into the databases: complexity of evaluation data leading to difficult interpretation, existence of genotype x environment interactions, possible heterogeneity or redundancy in the populations evaluated, and last but not least the lack of standardization in documentation systems. Given the importance of these data to users, special efforts should be made to facilitate the integration and easy accessibility of evaluation data.

Monitoring the use of Central Crop Databases

Wieslaw Podyma stressed the need to better identify the users of CCDBs in order to meet their needs in an optimal way. He pointed out that very few are available so far and that monitoring the use of databases is much easier for on-line databases. The Internet offers specific tools which would greatly facilitate this function.

Access to Central Crop Databases

Elinor Lipman and Daniel Jiménez Krause addressed the issue of accessibility after analyzing further some of the results of the survey on European Central Crop Databases: updating, descriptors and softwares used. The technical aspects of accessibility to European CCDBs were summarized, with emphasis on the possibilities offered by new technologies and specifically by Internet. A proposal was made on behalf of ZADI, Germany to establish an Internet-based information system for plant genetic resources in Europe. In conclusion a synthetic diagram was presented, showing how the ECP/GR structures and its Documentation Network can play a key role in the implementation of such a system, with a strong support from selected Documentation Support Centers (DSCs) providing expertise and technical assistance. A demonstration of the new ECP/GR home page[2] was given, with prototype links to *Prunus* and *Secale* CCDBs illustrating the foreseen utilization of this home page as an information node.

Using Central Crop Databases: searching for duplicates and gaps

This presentation dealt with essential aspects of the potential usefulness of crop databases, related to the assessment of the value of a collection. Helmut Knüpffer reviewed the different types of duplicates which can be found in databases and the techniques available for their identification and verification. Lothar Frese focused on the further utilization of databases for the inventory of gaps in collections which can lead eventually to the planning of complementary collecting missions.

Recommendations of the Workshop

On 14 October, four working groups discussed in parallel the role of databases, the needs and opportunities for standardizing Central Crop Databases (CCDBs), the role of database managers, and opportunities to facilitate access to and use of the CCDBs. During a second session on 15 October additional working groups were asked to formulate specific recommendations concerning the FAO/IPGRI Multicrop Passport Descriptor List, the Internet access to CCDBs and the inclusion of evaluation data into CCDBs. The recommendations which were agreed upon further to these working groups sessions and to the discussions which followed in plenary are synthesized below.

The role and users of Central Crop Databases

The Workshop recognizes that the fundamental role of the CCDBs has not changed drastically in recent times. A principal role will remain the establishment of inventories of the plant genetic resources holdings at the regional level for specific crops.

The users of the CCDBs are considered to be:

1. **Primary users:** the curators and crop-specific working groups seeking to secure the conservation of a genepool in the most effective way. For these users, the CCDB will remain an essential management tool for the identification of duplicates and omissions in the collections, as well as a basic source of information for the development of collaborative activities such as the establishment of core collections, the planning of collecting missions, etc.
2. **End users:** including breeders, researchers, educational establishments and other users who wish to access the CCDBs to obtain specific information

---

[2] URL: http://www.cgiar.org/ecpgr

about the collections, frequently as an entry point to the collections themselves. The CCDBs should seek in the future to better take into account the needs of this group of users. This implies a shift of focus to include more evaluation-related data in the CCDBs.

The evolution and the increased availability of Internet offer to CCDBs the possibility to shift their role from being a depository of information to becoming an active provider of information. It is expected that if CCDBs take full advantage of the present technological opportunities, they will be able to better serve the needs of those potential users that have not, up to now, found sufficient convenience in approaching and utilizing the genebanks.

To fulfil the mission of CCDBs in the future, the Workshop recommends:

- increasing the quality and processing of data in CCDBs;
- facilitating the improved accessibility to the data included in the CCDBs;
- strengthening, when needed, the capacity of database analysis of institutions hosting CCDBs;
- including, as far as possible, evaluation data in the CCDBs to make the germplasm more meaningful/useful to users;
- moving towards a more regular updating of CCDBs, thereby offering a plant genetic resources monitoring function;
- ensuring the best possible coverage of relevant institutions and countries in Europe within CCDBs.

It was agreed that for the time being the CCDBs would not include data from *in situ* conservation activities.

Standardization of Central Crop Databases

It was recognized that a minimum level of standardization is essential to carry out tasks such as the identification of duplicates and omissions in collections. While the extent of standardization of characterization and evaluation data should be decided by the crop-specific Working Groups, passport data and environmental data should be standardized as far as possible for all crops.

IPGRI should continue to harmonize descriptors used for different crops, *inter alia* through coordination between working groups preparing these descriptors.

FAO/IPGRI Multicrop Passport Descriptor List

The Workshop agrees unanimously that a multicrop passport descriptor set is essential for efficient data exchange. The draft list proposed by IPGRI/FAO serves as a good basis for discussion.

The Workshop adopts the list, as amended below, as a standard format for data exchange. The Multicrop Passport Descriptor List should be used and promoted by IPGRI/FAO in all relevant activities.[3]

The draft descriptor list presented contains two components: multicrop passport descriptors and WIEWS descriptors. The Workshop **does not** considers the WIEWS descriptors to be part of the multicrop passport list. They are

---

[3] For the draft list presented at the workshop refer to Hazekamp *et al*., pp. 35-39. The final version is included as Appendix II. This final list is a reflection of the inputs received from a large group of individuals from all over the world. Although it was not always possible to implement changes exactly as proposed, it should be noted that the final list is fully compatible with the recommendations listed below.

optional, but recognized to be valuable tools for monitoring germplasm collections.

Standard address and country coding systems are considered an essential part of the data exchange format. Therefore the Workshop strongly recommends that IPGRI/FAO accept the responsibility to provide, maintain and promote the above standards.

IPGRI informed the Workshop that it is maintaining a country code list and agreed to send an updated version to the participants by the end of October 1996 (see Appendix III).

The FAO representative agreed to send by mid-November a printout of the INST.DBF records for each country to the respective ECP/GR National Coordinators for review. He agreed to incorporate updates and publish a new INST.DBF list by the end of February 1997. This will be available for distribution on Internet or diskettes on demand and will be announced in the IPGRI Newsletter for Europe.[4]

Amendments to the draft descriptors

N.B. The comments below refer to the draft list presented during the workshop (see Hazekamp *et al.*, this volume).

General comments:

- the suggested names/acronyms for the fields of this list were adopted by the Workshop as identifiers for the fields in data exchange;
- the length of the fields should not be restricted.

Multicrop passport descriptors

1. Institute codes

- Code and acronym lists should be maintained in parallel.
- Mechanisms for direct access will be established.
- For international institutes the three-letter code INT should be used instead of the country code (e.g. the Nordic Gene Bank would be ‘INTnnn’ instead of ‘SWE002’).

2. Accession number

The Workshop recognizes the value of using an institute identifier, preceding the accession number, wherever possible.

4. Scientific name

For the purpose of data exchange, it is recommended to divide the field into three:

- Genus <GENUS>
- Species <SPECIES>
- Subtaxa <SUBTAXA>, where subtaxa can be used to store any additional taxonomic identifier.

5. Cultivar name

Change to 'Accession Name' <ACCNAME>.

---

[4] For practical reasons, a list of European genebanks with addresses and codes will first be made accessible through the ECP/GR home page. Updating and enhancement of the list will be performed progressively.

6. Country of origin

- ISO codes to be extended to include current and old codes.
- Regional codes are considered not relevant in this context.

7. Location
It is recommended to place the town / site name first in the location statement.

8. Latitude and
9. Longitude
It is recognized that missing data have to be indicated by a standard character: hyphen (–).

12. Status of sample

3. Adopt 'Landrace' and remove 'Primitive cultivar'
6. 'Unknown' – to be removed
7. 'Other' – to be included in a memo field with appropriate identifier.

13. Collecting source
The Workshop agreed that the multilevel status is too detailed for a multicrop minimal descriptor list and recommends a re-evaluation of codes at the primary level only.
For 5. 'other' – same comment as in 12 above.

14. Donor institute code
Same comment as for 1. 'Institute code'

15. Donor number and
16. Other number(s) associated with the accession
Where accession numbers are used, refer to point 2. 'Accession number'

Other general comment: add a memo field.

FAO WIEWS Descriptors
17. 'Location of duplicates'
Change to 'Location of safety duplicates'
For the coding, refer to 1. 'Institute code'

18. Availability of passport data
Should be limited to two states (1 or 0)

19. Availability of characterization/evaluation data
Subdivide characterization and evaluation; each to be a 1 / 0 field as for 18 above

20. Acquisition type of the accession
Change to

1. collected / bred originally by institute
2. collected / bred originally by joint mission
3. received as a second repository

21. Acquisition date of the accession
Delete

22. Type of storage
Add 7. Other

Taxonomy

During the discussion the problem of inconsistencies relating to taxonomy was raised. It was agreed that whenever a widely recognized monograph exists for a given crop, it should be followed. In situations in which disagreement about the taxonomy prevails, the decision whether to take a pragmatic approach should be left to the crop-specific Working Group (e.g. the *Brassica* Working Group established an agreed list for internal use by the Group and the European *Brassica* Database).

Evaluation data

The Workshop acknowledges the great importance of evaluation data availability in CCDBs. Evaluation data play a key role in promoting the use of CCDBs and the utilization of collections. The following specific recommendations were made:

1. If staff availability is a limiting factor, recording of meta-data should be a first priority. Pointing to the source of information can be of high value when the information itself is not available.
2. The crop-specific Working Group should decide about type and number of descriptors. It is highly advisable to include resistance and quality characters if needed for a crop.
3. Curators along with evaluators should interpret the data. The most appropriate interpreter is the crop expert. Interpreted data should be sent to the CCDB Manager for incorporation.
4. The CCDB Managers should link the interpreted information to the source of the original data (donor of the information) or to the original data set.
5. Job-sharing between different institutes in cases of larger data sets is recommended.
6. It is recommended that, along with the interpreted data, data quality indicators should be provided. Crop experts should assess the quality / reliability of data.
7. It would be helpful to add information about the environment of the evaluation site, and the methods and standard lines used.
8. It is essential to produce well-described data sets in evaluation projects.
9. Further research on the methodology of combining heterogeneous data sets in a CCDB would be important (e.g. the German EVA project).[5] A case study for a cross-pollinating crop, analogous to EVA, would be useful.

Standardization of database software

The Workshop considers that standardization of database software is no longer necessary since commercial packages offer acceptable options for file exchange and Internet browsers are available to read Internet-based CCDBs.

---

[5] The EVA project, launched in May 1996 under the coordination of ZADI / IGR, and in which the major German institutions involved in plant genetic resources participate, aims to establish a central on-line information system which will include evaluation data (EVA: Informationssystem für Evaluierungsdaten pflanzengenetischer Ressourcen).

Role of the database manager

The database manager has the responsibility for the compilation and comprehensiveness of the database and facilitates its access through appropriate dissemination/publication channels. It is understood that by accepting to host a CCDB, an institution accepts the tasks involved in managing and analyzing the database and will offer the necessary link to the original data source, for the benefit of the user community.

Establishment and updating of the database

With regard to the establishment and updating of the database, the guiding principle should be that genebanks must not be overburdened with reformatting of the data requested by the CCDB. The Workshop notes that the adoption of the FAO/IPGRI Multicrop Passport Descriptors List as standard exchange protocol will reduce the standardization work required from the database manager.

While genebanks should be strongly encouraged to submit their data in this format, database managers should be prepared to accept differently formatted data as long as adequate coding tables are provided with the data.

Considering that CCDBs will be subjected to a higher scrutiny once they are published via Internet, it is recommended that the database managers endeavour to update the databases regularly, approximately once a year. It is expected that the increased visibility of genebank holdings given by an Internet-based genetic resources information system may also prompt curators to be more proactive in updating the CCDBs.

Mediation of information, advice and analysis

It is expected that despite technological changes that may in the future make CCDBs directly accessible to users, direct personal contact and advice will remain important. The database manager has easy access to, and generally a more thorough overview of, the holdings of the specific crop, and therefore plays an important role in analyzing the database and advising the crop-specific Working Group with regard to duplication or gaps in the collections or on collaborative activities such as the establishment of core collections, planning of collecting missions, etc.

Other activities

The Workshop considers that the monitoring of the collections or of the genebanks' response to seed requests is not part of the responsibilities of the CCDB manager.

Monitoring of information requests is considered useful to provide information about the degree of use and the needs of user groups. This task will be facilitated once CCDBs are accessed on-line via Internet.

The channelling of germplasm requests is a service which may be useful for users, and is recommended for CCDBs accessible via Internet.

Raising awareness about the existence of the CCDBs is an important responsibility of the database manager but should also be a primary concern of the entire crop-specific Working Group.

Facilitating access and use of the CCDBs

The Workshop considers that the quality and completeness of the information stored in the CCDBs will ultimately determine the degree of use of the CCDBs and of the crop genetic resources collections. Priority should therefore be given to monitoring data quality, regular updating of the CCDBs and to completing information gaps.

Recognizing the value of Internet communication to facilitate the access to the CCDBs, the Workshop recommends that an Internet-based information platform be established under the ECP/GR 'umbrella' linking, *inter alia*, to all the CCDBs.

This information platform should offer services at the following levels:

- on-line searches and requests via Internet;
- downloading of entire CCDBs, or subsets, from the Internet servers;
- ordering of CD-ROMs or diskettes;
- ordering of printouts or published catalogues.

The platform's content

The platform will consist of the following information subsystems: Crops, Institutions and Services.

The subsystem **Crops** will consist of a catalogue of crop databases presented as a list of crops with links to the respective European Crop Database and/or international crop database and/or local (institute) crop databases.

The subsystem **Institutions** will among others offer IPGRI/FAO's European Directory as a list of European plant genetic resources institutions together with data on addresses, telephone, fax, emails, experts, etc.

The subsystem **Services** is understood to be a communication forum for the European plant genetic resources community. Services that could be offered are: news, newsletters, Internet addresses of plant genetic resources institutions and experts, discussion lists, question and answer service, a depository of standardized information (FAO institutes' codes, IPGRI/FAO's multicrop passport descriptor list, etc.), links to other relevant plant genetic resources databases, job market, calendar of plant genetic resources activities, etc. The range of the services to be implemented will be defined by the advisory group (see below).

The platform will be implemented as a subpage of the ECP/GR home page so that a user visiting this site can start directly searching a crop database, institution data or using the services.

Minimum standardization for home pages

All home pages of European crop databases should be based on a common logic in order to build, together with the platform, a virtual network that will provide central access to all databases.

The recommended standard form for home pages of European crop databases foresees a home page with the following layout/features:

Header including:

- the ECP/GR logo at the top with a link to the ECP/GR Platform home page (a logo of the database manager's institute with a link to its home page can of course be added);
- the full name of the database;
- name, address, telephone and fax numbers, and email of the database manager.

Body with the following five links:

- contributors (link to a searchable list of contributing institutions including full contact details of curators/database managers - postal address, telephone and fax numbers, email address);
- database description (link to brief text providing information on objectives, scope, date of last comprehensive update, etc.)
- on-line - search (link to menu offering search possibilities, if available);
- off-line - download (link to menu offering downloading possibilities, if available);
- off-line - hardcopy/diskette (link to information on how to order, if available).

It is important that the header and body with the elements mentioned can be seen without scrolling the screen. It is also recommended to write name and email address of the person who created/updated the home page as well as the date of last updating.

ECP/GR will provide a model home page for the CCDBs and ZADI and NGB will jointly provide a list of commonly used terminology (see Appendix IV) and guidelines with the first steps for creating home pages or on-line searchable databases.

Organization and mechanisms

A two-step approach was proposed:

- First step: creation of a home page for each CCDB.
- Second step: to promote the databases going on-line (whether as on-line database and/or files to download via Internet). It is up to the database managers to decide whether a database is "ready" to publish on the Internet.

This approach raises the question of the location of the CCDB. The preferred solutions are:

- at an Internet server at the institution managing the CCDB;
- at a university or research centre within the country, or
- at a Documentation Support Centre (see definition below).

Other needs

Mirroring

Mirroring of the ECP/GR home pages on servers located in Europe may be necessary since it is currently located on the CGNET Server in the USA.[6] This could lead to bottlenecks in the transmission of information during times of intensive utilization.

Documentation Support Centres (DSC)

The Workshop expresses a clear need for the existence of DSCs. These would have a temporary role and their number is *a priori* not limited.

Services which are requested include:

---

[6] CGNET Service International was established to install and administer an electronic mail communication system for the Consultative Group on International Agricultural Research (CGIAR).

- training and/or technical advice in preparing home pages and making databases directly accessible via Internet;
- assistance in developing on-line databases;
- temporarily hosting CCDBs and home pages on the DSCs' Internet servers;
- mirroring the ECP/GR home page if needed.

Although this support will be in principle given in kind, it is recognized that additional funding may need to be sought.

Training

There is much need for training in the field of Internet publishing. Funding should be raised for this.

ZADI and ECP/GR offer to organize a training workshop in Bonn (Germany) in Spring 1997 on topics related to the Internet and on-line databases.

Advisory Group

The Workshop recommends the establishment of an advisory group, under the coordination of the ECP/GR Coordinator, to further develop the Internet-based information platform on plant genetic resources.[7]

---

[7] As of January 1997 the Advisory Group has been formally established. The members of the group are Pierre Campo (GEVES), Theo van Hintum (CGN), Morten Hulden (NGB), Daniel Jiménez Krause (ZADI), Kevin Painting (IPGRI) and Lorenzo Maggioni (ECP/GR Coordinator).

Other issues and closing remarks

The Workshop encourages European institutions to create CCDBs to increase the coverage of the system. Such services should be considered as input in kind to the European cooperation on plant genetic resources conservation. The Workshop further recommends to the ECP/GR Steering Committee that such databases be included in the ECP/GR system.

CCDBs established in Europe within the framework of other programmes are invited to make full use of the ECP/GR genetic resources documentation platform, to offer their services to users.

Next meeting

The Workshop recommends to the ECP/GR Steering Committee that a follow-up meeting on CCDBs be organized in 1998.

Closing remarks

The participants felt strongly that the Workshop was an important step towards implementing the Global Plan of Action adopted in June 1996 in Leipzig (Germany) and that the planned information platform would become a key tool for reaching the Global Plan of Action’s objectives in the European region. Dr László Holly and the staff of the Institute of Agrobotany of Tápiószele were thanked for their warm hospitality and the excellent organization of the Workshop logistics.

Part B. Presentations and Other Contributions

Summary of the working session for EGDS focal points

M.W.M. Jongen
Centre for Genetic Resources (CPRO-DLO), Wageningen, The Netherlands.

Introduction

On 12 October 1996 a working session was organized with the eight partners of the project ‘Technical support to east European genebanks to improve access of privatized plant breeding to germplasm collections’. The partners are the national genebanks in Bulgaria, Czech Republic, Hungary, Poland, Romania, Russia, Slovakia and Ukraine. This project, often referred to as EGDS (East European Germplasm Documentation System), was coordinated by the Centre for Genetic Resources, The Netherlands (CGN) and financed by the Dutch Ministry of Agriculture, Nature Management and Fisheries from January 1994 to December 1996.

Objectives

The network of focal points, i.e. the documentation specialists of the partner institutes, will:

- exchange information on the information stored in the documentation systems of the institutions;
- improve the availability of this information;
- improve the quality of this information.

Project activities

Consultancies were carried out to describe the documentation systems (Jongen and van Hintum, 1995) and to identify constraints in the field of documentation. With financial support some of the problems encountered could be solved, such as purchasing of hardware and software, training and hiring additional personnel for data input.

Yearly workshops were organized in October. These workshops support the formation and maintenance of a network of the focal points, and will improve the level of knowledge (Jongen and van Hintum 1994; van Hintum *et al.* 1995).

Meeting

During the working session for EGDS focal points progress reports were presented and discussed (Jongen and van Hintum 1997) and the project was discussed and evaluated. The main conclusions are mentioned below.

Conclusions

Some of the effects of the EGDS project are:

- intensification of documentation activities within institutions and contribution to the qualitative improvement of existing documentation systems of participating genebanks;

- visible improvement of equipment for documentation of genetic resources in all countries involved in the project;
- establishment of personal contacts between documentation specialists;
- improvement of fluency in English resulting in easier and better exchange of knowledge and contacts with colleagues;
- contribution to standardization of methods of documentation in genebanks and data dictionaries of central databases;
- many positive side effects.

References

Jongen, M.W.M. and Th.J.L. van Hintum (eds.). 1994. Report of the First Technical Meeting of Focal Points for Documentation in East European Genebanks. Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.

Jongen, M.W.M. and Th.J.L. van Hintum (eds.). 1995. Descriptions of national plant genetic resources documentation systems in eastern European countries. IPGRI, Rome, Italy.

Jongen, M.W.M and Th.J.L. van Hintum (eds.). 1997. Report of the Third Technical Meeting of Focal Points for Documentation in East European Genebanks. Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.

van Hintum, Th.J.L., M.W.M. Jongen and Th. Hazekamp (eds.). 1995. Standardization in Plant Genetic Resources Documentation. Report of the second technical meeting of focal points for documentation in East European genebanks. Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.

Report on the Technical Meeting on the Methodology of the World Information and Early Warning System on Plant Genetic Resources (WIEWS) held in Radzików, Poland,
30 September - 3 October 1996

J. Serwinski
Seed and Plant Genetic Resources Service, FAO, Rome, Italy.

The Meeting recommended the establishment, on a voluntary basis, of a global network of correspondents in order to effectively collect and elaborate information for WIEWS and cooperate in its development (FAO/IHAR 1997). It was requested to place the network under the auspices of FAO. The Meeting requested FAO to formulate a letter to invite all countries to identify a corresponding member from their national plant genetic resources programme to cooperate with WIEWS. The network members should be supported by governments and will serve as collaborators/correspondents for the WIEWS and provide information on plant genetic resources collections and organization of national programmes, as well as information on changes that could pose a threat to plant genetic resources in their countries.

Recognizing the potential role of existing networks for WIEWS development, the Meeting requested that FAO establish an inventory of existing crop networks on plant genetic resources and other networks involved in plant genetic resources activities and invite them to cooperate with WIEWS.

Reference

FAO/IHAR. 1997. Report on the Technical Meeting on the Methodology of the World Information and Early Warning System on Plant Genetic Resources held in Plant Breeding and Acclimatization Institute Radzikow, Poland, 30 September - 3 October 1996. FAO, Rome, Italy.

Central Crop Databases - an overview

Th. J.L. van Hintum
Centre for Genetic Resources, the Netherlands (CGN), Centre for Plant Breeding and Reproduction Research (CPRO-DLO), Wageningen, the Netherlands

Introduction

Central Crop Databases play a central role in the collaboration between genebanks in Europe. They are the most important output of the Crop Working Groups within the European Collaborative Programme for Crop Genetic Resources Networks (ECP/GR), and they form part of most of the projects in the current Genetic Resources Programme of the European Union.

A Central Crop Database (CCDB) can be defined as a centralized regional or international database of plant genetic resources accessions of a (group of) crop(s), held in several institutions in a region or worldwide (Knüpffer 1995). The main objective of a CCDB is to increase accessibility of information about accessions. This increased accessibility can benefit many applications, most importantly:

- improve accessibility of germplasm for users and curators; this will increase utilization of the material, avoid new unnecessary duplication and allow joint activities such as core collection formation and evaluation projects.
- identify omissions in the combined collections; this makes it possible to efficiently target collection expeditions and avoid redundant expeditions.
- identify probable duplication between collections; this makes it possible to set priorities for multiplication and safety-duplication, and to reduce the size of collections.

Achievements

In an extensive overview of CCDBs of Knüpffer (1995), 64 regional and global CCDBs were listed. It showed the variety in these databases in size, coverage, software and organizational framework.

The largest databases listed were the eastern European wheat database (Michalak *et al.* 1991) with 56 712 accessions, and the European barley database (Knüpffer 1988) with 55 369 accessions. In this barley database 26 countries were participating, whereas the largest number of genebanks (65) participated in the global cocoa database (End *et al.* 1990).

The software used for a database depends completely on the local circumstances such as the institute standards, the knowledge, experience and preference of the database manager, but also the size of the database and the requirements concerning number of users, integrity checks, etc. This is reflected in the range of software used for databases which includes commercial database management software packages such as dBase, Foxpro, Oracle, Access and ACEDB, but also a package developed for statistical analysis, SAS.

Another aspect in which CCDBs vary is the framework in which they were developed. Most CCDBs have been developed within the context of an ECP/GR Crop Working Group, some in an IPGRI International Crop Network, another Crop Network or as a local initiative.

Only in some cases were these databases actually used as the tools they were intended to be, for a number of reasons:

- low quality of the database, due to low quality of the data supplied by the sources and lack of data-processing prior to entering in the database;
- low accessibility of the database, due to lack of physical access, and a confusing diversity in software, logical structure and coding systems;
- lack of capacity for analysis of the database; genebanks have other priorities.

As a result, the databases were generally not used as often as intended, not maintained as appropriate, and often, in the end, did not meet the expectations. It can be expected, though, that this situation will improve considerably in the near future as a result of the availability of the Internet, coupled with an increasing availability of computers and an increasing computer affinity of the germplasm curators.

Current issues

Task of CCDB manager

Generally the workload of creating, maintaining and analyzing a CCDB is considerable, and the question 'who should do what' is a central one. In the phase of creating the database, a CCDB manager can decide only to accept data in a predefined format. This moves part of the workload from the CCDB manager to the persons responsible for supplying the data in the collaborating genebanks. An important undesired side effect is that the response to the request for data will decrease proportionally to the level of formatting required. This problem will become much smaller if a standard format for data exchange is accepted by the genebank community. This standard should include a standard data structure and a standard coding system.

The other tasks connected with a CCDB will remain highly time-consuming. These tasks, which determine the value of the database, include:

- data-processing to improve the quality of the data;
- analyzing the database, tracing duplicates and omissions;
- making the database accessible to the user, by distributing copies and/or putting it on the Internet;
- making the database widely known, via publications, posters and presentations.

Snapshot vs. monitor

In the ideal situation a CCBD should always be up to date, a monitor. This involves frequent updates which is only possible if a certain degree of standardization for the data exchange can be achieved. It is also important that a permanent link between the CCDB identifiers of the accessions and the genebank identifiers, i.e. accession numbers, is established. A monitor CCDB is a permanent activity and will therefore require structural funding which will generally only be possible if it is part of the regular genebank activity. There is a danger that in such a situation the limited capacity for the CCDB is fully spent on updating and that the important tasks – data-processing, analysis and making it accessible and well known – are neglected.

An alternative is to create the CCDB as a snapshot. For such an approach the database can be created, analyzed and the results of the analysis can be implemented in a predefined time span. This has the clear advantage that it can be managed as a project with a budget, milestones, etc. The obvious disadvantage is that the database will be out of date very quickly.

Scope

The scope of a CCDB is another matter that has to be decided. Aspects of the scope are:

- the data sources: European or global, genebanks only or all sources, etc.;
- the material: crop or species or genus;
- the type of data: passport data only or also evaluation data, minimal or full description.

The larger the scope, the bigger the job. A large scope is attractive since it will obviously give more information, but it will also be more difficult to manage because of the workload. A limited scope allows for more attention to the data and can thus result in a higher quality.

If the inclusion of evaluation results is considered, the problems that influence the interpretability should be solved, i.e. the genotype x environment interaction, the definition of descriptors and scores, the observation method, the within-accession variation and the variation in reliability of the data. An alternative to including actual data in the CCDB is to include so-called meta-data describing the datasets that are available. It is then left to the user to solve the interpretation problems.

Future

As mentioned earlier, the future for CCDBs looks bright thanks to the Internet. That the Internet will improve the physical access to the databases is already apparent. But it can also improve the functional access since it makes standardization of structure and coding more attractive and easier. Standardization of interfaces also will become more attractive. Internet also allows for new applications of CCDBs such as the facility to request material directly from the source via the CCDB.

Again, standardization is a keyword. Especially in the European system of plant genetic resources conservation, collaboration is dependent on information exchange, and that can only be organized efficiently if the community agrees to adopt some standards. The coordinating bodies in the field of plant genetic resources conservation, i.e. ECP/GR, IPGRI and FAO, should therefore also give high priority to supporting this standardization by developing, in collaboration with the users, standard systems that can easily be applied, maintaining these systems, and promoting their use in any possible way.

References

End, M.J., R.M. Wadsworth and P. Hadley. 1990. The primitive cocoa germplasm database. Cocoa Growers' Bull. 43:25-33.

Knüpffer, H. 1988. The European Barley Database of the ECP/GR: an introduction. Kulturpflanze 36:135-162.

Knüpffer, H. 1995. Central Crop Databases. Pp. 51-62 *in* Standardization in Plant Genetic Resources Documentation. Report of the Second Technical Meeting of Focal Points for Documentation in East European Genebanks (Th.J.L. van Hintum, M.W.M. Jongen and Th. Hazekamp, eds.). CGN, Wageningen, The Netherlands.

Michalak, M., P. Kolasinski and W. Podyma. 1991. International Wheat Database. *In* Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management

Management (Th.J.L. van Hintum, L. Frese and P. Perret, eds.). Papers of the EUCARPIA / IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

The role of Central Crop Databases in the European Cooperative Programme for Crop Genetic Resources Networks (ECP/GR)

T. Gass, E.Lipman and L. Maggioni
International Plant Genetic Resources Institute (IPGRI), Rome, Italy

Introduction

From its initiation in 1980, ECP/GR has focused on three principal objectives: the conservation of plant genetic resources, promoting the use of this germplasm and facilitating the international collaboration at the regional level. Documentation systems for crop genetic resources, and in particular central crop databases, play an essential role in reaching all three of these objectives. They provide a comprehensive overview of the collections maintained in Europe and thereby allow the crop-specific Working Groups of ECP/GR to take informed decisions regarding the need for safety-duplication, for characterization, the possibilities for rationalization of conservation activities, the filling of information gaps or the planning of further collecting activities.

The rapid development of technology in the areas of communication and electronic data storage and processing have recently brought within reach goals which seemed unrealistic 15 years ago. It has been a constant concern for the European genetic resources community to remain abreast of these developments despite the fact that documentation systems are generally a secondary activity within the institutions managing genebanks and consequently receive little financial support. The present Workshop should therefore be seen as the continuation of a series of workshops, meetings and symposia addressing the subject of crop genetic resources documentation in Europe (UNDP/IBPGR 1984; van Hintum *et al.* 1990; FAO/IPGRI 1993; CGN/IPGRI 1995)

This presentation provides a brief overview of the central crop databases established within ECP/GR, it shows the particularities of this region which make central crop databases so important, and discusses the challenges and opportunities facing European crop databases in the medium term.

European crop databases established within ECP/GR

Currently, 30 countries participate in ECP/GR (Table 1). The Programme, which is entirely financed by the participating countries and is coordinated by IPGRI, operates through ten broadly focused networks dealing with groups of crops or general themes related to plant genetic resources (Fig. 1).

Within these networks, 32 central crop databases have been established (Table 2). These are managed as ‘input in kind’ to European cooperation on crop genetic resources by a total of 29 institutes from 18 countries. The number of crops and crop groups covered by these databases reflects the broad scope of the Programme (Table 2).

Table 1. Countries participating in ECP/GR (status 1/10/1996).

Austria	France	Lithuania	Spain
Belgium	Germany	Malta	Sweden
Bulgaria	Greece	Netherlands	Switzerland
Croatia	Hungary	Norway	Turkey
Cyprus	Iceland	Poland	UK
Czech Republic	Ireland	Portugal	Fed. Rep. Yugoslavia (Serbia and

Denmark
Finland
Israel
Italy
Romania
Slovak Republic
Montenegro)

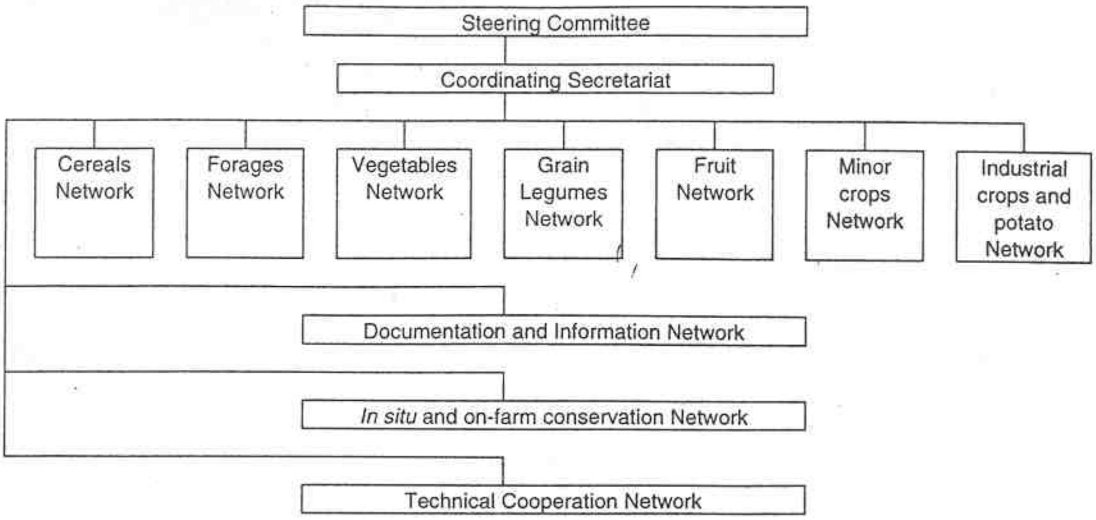


Fig. 1. Structure of ECP/GR.

The usefulness of a database depends strongly on the quality of the data recorded. This quality is related to factors such as data comprehensiveness, date of last update and accuracy of data. Currently the most comprehensive information in the databases is passport data of germplasm stored in the contributing genebanks. In the late 1980s the size of the collections and the restricted funds available to carry out characterization prompted ECP/GR Working Groups to establish lists of minimum descriptors so that all accessions would at least be characterized with regard to a restricted number of traits considered useful to distinguish putative duplicates or characters of particular importance to users. For a few databases it has been decided that any available characterization and evaluation data also could be included in the database, albeit not always in a standardized form.

A survey conducted among managers of the central crop databases before the present workshop provides an indication of the frequency of updating and the level of standardization with regard to descriptors and software. The validity of the survey is somewhat limited due to a level of response of 65% of the database managers. Also, many databases are under construction (e.g. Wheat, *Phaseolus*) or reconstruction (e.g. *Allium*) so that detailed information is still lacking or has only a temporary value.

While some of the first established databases have undergone several iterations of updating, a number of databases have only recently been established and are currently carrying out the initial data survey (Table 3). As 59% of databases have been updated or created in 1996, a high level of 'freshness' of data can be expected in the short term. With regard to updating frequency, no specific pattern exists. The variations in the frequency of updating are conditioned by the different sizes of the databases, by the level of priority given to this activity within the institution managing the database, financial constraints and the existence of a working group or a network utilizing the database.

Table 2. European Crop Databases established within ECP/GR.

ECP/GR Network	Crop or crop group	Wk Gp	Managing institution	Establishment date	Total no. of accessions
Cereals	*Avena*	√	FAL, Braunschweig, Germany	1984	17451 (in 1993) (*)
	Barley	√	IPK, Gatersleben, Germany	1984 to 1987	55370
	Secale		IHAR, Radzikow, Poland	1984; rebuilt 1995	9683
	Triticale	(√)	RAC, Nyon, Switzerland	under establ.	not applicable
	Wheat	(√)	RICP, Prague-Ruzyne, Czech Rep. and GEVES, Surgères, France	1996	250 000 (estim.)
	Maize		MRI, Belgrade-Zemun, F.R Yugoslavia	(Oct 89); ECP/GR 1996	(5437) (under establ.)
Vegetables	*Allium*	√	HRI, Wellesbourne, UK	1985 (rebuilding)	6209
	Brassica	√	CGN, Wageningen, The Netherlands	1992	12264
Grain legumes	*Phaseolus*	√	Fed. Office Agrobiology, Linz, Austria	under establ.	13854
	Vicia faba	√	INRA, Le Rheu, France	under establ.	not applicable
	Pisum	√	IPG, Poznan, Poland and JIC, Norwich, UK	1989	37103
	Lupinus	√	IPG, Poznan, Poland	May 1996	6782
	Cicer	√	ENMP, Elvas, Portugal	Nov 1996	4421
	Glycine	√	VIR St Petersburg, Russia	1992	6082
	Lens	√	AARI, Izmir, Turkey	under establ.	not applicable
Fruit trees	*Prunus*	√	INRA, Bordeaux, France	(1982); rebuilt 1996	15000
	Malus	√	NFC, Wye College, UK	Feb 1995	2500
Forages	*Lolium, Trifolium repens*	√	IGER, Aberystwyth, UK	Nov 1987	*Lolium* 8456l; *T. repens* 1247
	Lathyrus latif./ tub./heter./sylv.	√	IBEAS, Pau, France	1985	3734
	Vicia spp.	√	CNR, Bari, Italy	1992	5520
	Other Vicieae	√	University Southampton, UK		
	Agropyron	√	IPGR, Plovdiv, Bulgaria	under establ.	not applicable
	Arrhenat. elatius/ Triset. flavescens	√	OSEVA Pro, Zubri, Czech Rep.	1991	83 (67/16)
	Perennial *Medicago*	√	INRA-GEVES, Guyancourt, France	1987	1917
	Other perennial forage legumes	√	Inst. of Agrobotany (IA), Tápiószele, Hungary	under establ.	not applicable
	Poa	√	IPK, Malchow, Germany	1984	1963
	Bromus	√	IA, Tápiószele, Hungary	1984	430
	Trifol. pratense	√	IA, Tápiószele, Hungary	1984	1895
	T. subterran./ annual *Medicago*	√	INIA, Badajoz, Spain	1981	3613
	T. alexandrinum /T. resupinatum	√	ARO, Bet Dagan, Israel		*T. a.* 140; *T. r.* 100 (in 1995)(*)
	Dactylis/Festuca	√	IHAR, Bydgoszcz, Poland	1985	12918
	Phalaris/Agrostis /Phleum	√	NGB, Alnarp, Sweden	under establ. Oct 1996	not applicable *Phleum* 4268

(*) Data extracted from Working Group reports.

Other international databases of particular relevance to the European collaboration on plant genetic resources (not exhaustive).

Network	Crop	Managing institution	Establish. date	Total no. of accessions

Int. *Beta* Network	*Beta*	FAL, Braunschweig, Germany	Mar 1987	9168
APIC[†]	*Solanum*	CGN, Wageningen, The Netherlands	1993	11590
ESCORENA	Sunflower	Inst. for Field and Veg. Crops, Novi Sad, FR Yugosl.	1984	768
(not applic.)	*Vitis*	BAZ-Geilweilerhof, Siebeldingen, Germany	1984	15992

[†] Association of Potato Intergenebank Collaborators.

Table 3. Date of last update and frequency of updating the European central crop databases.

Date of last update	% of databases
before 1990	6
1990-1994	3
1995	13
1996	27
initial survey	13
being established	19
unknown	19

Updating frequency	No. of databases
twice a year	2
every year	4
every 2 years	1
less frequent than 2 years	5
irregular	3

Most national programmes have adapted their own system of botanical descriptors based on the needs of curators and breeders. These national descriptor lists are frequently based on existing international descriptor lists published by IBPGR/IPGRI. The widespread use in central crop databases of these international descriptor lists (Table 4) constitutes an important first step in the standardization and improved accessibility of central crop databases.

At the time of establishment of the central crop databases, dBase was the most frequently used software. More recently developed databases tend to use Windows applications and MS-Access is well represented. Smaller collections were computerized using software which is not strictly speaking a database management tool (listed under ‘others’). These types of software would prove to be unsuitable if a higher number of records were involved.

Considering the number of accessions, the most frequently used software is MS-Access and FoxPro. This is due to the large size of the Wheat database which is being established jointly by two institutes using, respectively, MS-Access and FoxPro.

Table 4. Descriptor systems and software used for the European central crop databases.

Descriptors used	% of databases
IPGRI list or adapted from IPGRI	58
National list of managing institute	15
Others	27

Software used	% of databases	% of no. of accessions
dBase	45	11
MS-Access	24	45
FoxPro	14	39
Oracle	7	
Others	10 (SAS, Excel, Word)	5

Central crop databases: a basic tool for European collaboration in PGR

In Europe, the role of central crop databases is emphasized by the diversity of national programmes in the region, the prevailing tendency of countries to maintain comprehensive collections of their own and the nature of crop genetic resources which require a long-term public commitment while yielding low short-term benefits.

National programmes for crop genetic resources in Europe range from centralized systems in which conservation work is mainly carried out by one institution; centrally coordinated systems in which a national board, commission or committee coordinates the activities of many, sometimes unrelated institutions; to completely decentralized systems in which no coordination mechanism has been established yet. Each country tends to store large collections of germplasm for those crops which its breeders are working with (Frison and Serwinski 1995), whereby the motivation for assembling comprehensive national collections is frequently related to lack of information about material available from another country rather than fear of losing access to the germplasm. Under such circumstances, the level of duplication of effort is potentially very high. Furthermore, a shift in priorities in agricultural research or production can cause one country to abandon a collection while a neighbouring country might be developing a collection for the same crop.

For the crop-specific working groups of ECP/GR, the central crop databases constitute, besides the reports of meetings, the principal interface with the germplasm user community. The establishment of a database prompts working groups to a thorough discussion on the standardization of characterization methods and frequently results in a list of priority descriptors for the identification of putative duplicates or for the separation of the genepool into subgroups for ease of reference by potential users. Following analysis, the database can yield important information regarding gaps in the collections, the need for safety-duplication, gaps in characterization, etc. In several groups the databases have been used in the development of concepts for core collections (e.g. *Lolium*, *Brassica*) or in priority-setting for collecting activities (e.g. World *Beta* Network).

Challenges and opportunities

Within the framework of ECP/GR no funding is *a priori* provided for the establishment, management or updating of databases. This has allowed overall programme costs to remain relatively low and ECP/GR to be viewed as a very cost-effective programme. This assessment, however, gives insufficient credit to the goodwill and personal time investment of the central crop database managers who have made these results possible and to those national programmes who are funding significant portions of staff time and have invested in modern equipment and software to provide this input in kind to European cooperation on crop genetic resources.

As a consequence of the above, the European central crop databases do not all evolve at the same speed. Updating frequencies and progress in expanding or in upgrading the databases to new software vary from case to case and reflect the financial constraints encountered by the institutions managing the database or the level of priority which these institutions give to the documentation activity.

Notwithstanding the above-mentioned challenges, a number of opportunities could currently be exploited. Rapid technological progress is contributing to making modern communication technology available to a far greater range of users. The conversion of files for exchange between different software programmes has been improved significantly in the past few years and many genetic resources institutes and genebanks possess highly qualified technical expertise in the area of documentation. Documentation systems for genetic resources no longer require the standardization of software or the merging of crop databases into a single regional or global database. This is illustrated very effectively by the recently established European Wheat Database which is managed jointly by two institutions (GEVES, France and RICP, Czech Republic) using an identical database structure but different database software. When completed, this database will be the Region's largest crop database, with approximately 250 000 records.

At the regional level a number of collaborative initiatives are currently putting valuable emphasis on the development of documentation systems for genetic resources collections. The European Union programme for genetic resources (EC1467/94), established in 1994 is providing financial support for the establishment/further development of a few central crop databases (e.g. *Allium*, *Prunus*, *Beta*, Potato). The Dutch-funded EGDS project[8] is making a very significant contribution to the development of national documentation systems for crop genetic resources in eastern Europe. In 1995, the Steering Committee of ECP/GR decided to modify the structure of the programme to give it more flexibility as a coordinating mechanism for crop genetic resources activities in Europe (IPGRI 1995). As illustrated in Figure 1, this new structure includes a Documentation and Information Network.

Meeting the demand of the next decade

Through its networks and particularly through its documentation activities, ECP/GR has successfully fostered a climate of trust and cooperation among countries in the area of crop genetic resources. In coming years, central crop databases will continue to be a key tool for the management of collections by the crop-specific working groups or individual curators. There is a need, however, to focus the activities and development of these databases more towards the needs of the end-users (breeders, research institutions, education establishments and others).

This can be achieved through:

- a higher level of standardization of data contained in the databases to facilitate analysis and the searching for specific information;
- making use of modern data-processing and communication technology to provide a wider and more user-friendly access to the databases;
- including in the central crop databases more information of direct interest to the end-users (e.g. summaries of evaluation results), and foremost
- by ensuring that the data made available through the database are recent and accurate.

By setting the direction to follow in the development of the documentation of European crop genetic resources collections, the present Workshop can make a significant impact on the efficiency of conservation activities, the level of utilization of the collections and the commitment of governments towards these collections.

References

CGN/IPGRI. 1995. Descriptions of plant genetic resources documentation systems in eastern European countries. CGN/IPGRI, Rome, Italy.

FAO/IPGRI. 1993. Proceeding of a Joint FAO/IPGRI Workshop on *ex situ* Germplasm Conservation. 7-9 October 1993, Prague, Czech Republic. FAO/IPGRI, Rome, Italy.

Frison, E.A. and J. Serwinski (eds.). 1995. Directory of European Institutions Holding Crop Genetic Resources Collections (Fourth Edition) - Vol. 1. Holdings - Vol. 2. Index. FAO/IPGRI, Rome, Italy

---

[8] Eastern European Germplasm Documentation Systems Project.

IPGRI. 1995. Report of the Technical Consultative Committee (Sixth meeting), 21-23 September 1995, Nitra, Slovakia. European Cooperative Programme for Crop Genetic Resources Networks. IPGRI, Rome, Italy.

UNDP / IBPGR. 1984. Report of a Workshop on Exchange of Information (23-25 October 1984 in Radzików, Poland). *In* Report of the Second Meeting of the Technical Consultative Committee, 3-5 December 1984, Oeiras, Portugal. UNDP / IBPGR European Cooperative Programme for Conservation and Exchange of Crop Genetic Resources, Rome, Italy.

van Hintum, Th.J.L., L. Frese and P.M. Perret (eds.). 1990. Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management. Papers of the EUCARPIA / IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

Standardization of Central Crop Databases

M. Hulden
Nordic Gene Bank, Alnarp, Sweden

Introduction

A decade ago or so, a **database** on a personal computer referred to a single file containing **records** with **fields** for different types of data. The concept gradually changed after more advanced versions of database programmes became available, particularly programmes capable of handling multiple files in **relations**. A database now denotes any number of files, referred to as **tables**, which are interlinked in relations as part of an information system.

The term **central**, meaning ‘at one place’, has also been augmented by modern technique. A ‘virtual centre’ on a global computer network appears to be ‘at one place’ to the users, but may in fact be composed of parts that are physically distant from each other.

A **central crop database** thus can be a single file, and also a combination of individual files containing data on crop plants, but the individual parts of the database can, depending on means of distribution and access, be physically placed at several different computers, and utilize global networks like Internet, in different parts of the world.

Background

The following presentation is based on experience in database handling at the Nordic Genebank. Since 1979, when NGB was founded, a large number of databases, compiled by national focal points in Denmark, Finland, Iceland, Norway and Sweden, have been sent to NGB for inclusion in a Nordic information system on plant genetic resources. The original databases include passport data (accompanying material sent to NGB’s seed store), variety inventories and pedigree information, characterization and evaluation data (produced subsequently in working group projects). The staff at NGB has compiled with these data secondary databases (produced through normalization and fusion of the original databases), and also administrative databases containing information on contact persons, institutes, literature references and correspondence, and databases on amounts and status of the material stored. At present the number of individual files exceeds 150, and more data are added each year, mostly in the form of new evaluation data resulting from the continuous documentation work administered by crop-specific working groups within NGB.

It is hoped that the opinions and solutions to common problems in the topic of central database standardization in this presentation will contribute to the work of constructing a plant genetic resources information system for the whole of Europe, within the frame of the ECP/GR. The task is similar to that performed for the Nordic region, but also different in that the number of mandate crops and varieties is much higher, as well as the number of contributors and institutes involved.

Objectives of standardization

Aspects of what parts of the information system need standardization:

- choice of descriptors;

- format of descriptor fields;
- structure of individual tables;
- botanical nomenclature and taxonomy;
- database structure and table relations;
- compilation of primary data tables;
- coordination of crop-specific tables;
- data publication and access to databases.

Choice of descriptors

The choice of descriptors to be included in a crop database depends on the type of table in question but also on the type of crop plant to be described. A division of descriptors into classes for **passport, environment, management, characterization** and **evaluation** data seems to have become a *de facto* standard, though selection of descriptors to be used differ between species and also between institutes holding material of plant genetic resources. Understandably, characterization and evaluation descriptors must be different for different crops, while greater uniformity can be achieved for passport, environment and management descriptors.

Nevertheless, **all** of the descriptor lists are often worked out by crop-specific working groups, which may result in some variation between the databases regarding the choice and format of descriptors. Better standardization can be achieved by more coordination between the working groups. Combining data for different crops into a flat (single-table) database requires at least some standardization between these descriptors in the original tables.

Format of descriptor fields

Most database programmes provide the user with a choice of several field formats. Character, memo, numeric, date and logic data are traditional types, but lately extended field formats for graphics, sound or binary data in general have been added in new programme versions. The choice between field types usually presents no problem. Conversions of one field type to another and adjustment of field lengths can be used when files to be merged have different field formats.

Table structure

Not too long ago, even the most advanced database programmes for personal computers could only handle single files. As a result, databases either contained a large number of fields, many of which were empty as they were relevant only for a subset of the records in the file, or there were several databases, each with a smaller number of fields, but with a high level of data redundancy between files. The files could not be used simultaneously, i.e. each file was an isolated database system on its own. On the other hand, report generation and data export from a 'flat file' were easy to perform. Nowadays, with the availability of more efficient, relational database programmes, information systems are built on several files, with a lower level of data redundancy and additional flexibility. But as a result these systems often rely on special applications for browsing, editing and report generation.

In local information systems, passport, environment, management, characterization and evaluation data are usually kept in different files. In the construction of a central database, it is often desirable that similar files for different crops can be merged as far as possible, to allow easy distribution of single files for those users that do not have access to on-line databases via global networks. For characterization and evaluation data, merging is not possible, but passport and environment data, and to some extent also management data, could be combined in tables common to all taxa.

A central crop database can thus be compiled as a single table, if characterization and evaluation data are excluded. The flat-file approach is well suited for distribution of easy-to-handle databases on removable storage media, or for downloading. The descriptor list for a ‘multicrop’ database, to be presented by others during this seminar, may become an important first step towards a more sophisticated central crop database.

But since characterization and evaluation often are the most interesting and desired part of the information, an efficient central database system must be constructed as a relational database, to allow inclusion of the crop-specific tables.

A multi-table central database becomes a necessity also if data from other main branches of the information tree should be included, i.e. one-to-many relations to literature references, contact persons and institutes.

Multi-table databases may still be distributed as stand-alone, but a more user-friendly approach is to distribute the database together with a software that can be used to access the data. For databases published on the Internet, universal browsers used by the clients provide enough functionality to make special applications obsolete. The responsibility for a user-friendly interface in these cases relies more on the structure and scripts at the server side.

Botanical nomenclature and taxonomy

One of the most noticeable inconsistencies between databases for crops collected from different sources is their variation in usage of scientific names. The concept of the correct botanical names in each instance often differs among different persons, institutes and countries. Without any standardization in this respect, a central crop database would become very inconvenient to use, and reports or statistics generated from the unmoderated version would suffer both in readability and credibility.

One solution to this problem is to ask the contributors to revise their usage of botanical names, i.e. to pick the names from a provided, precompiled list of accepted names. But in many cases it is not possible to go back to the original sources for a revision, and so it is often up to curators and central database managers to find an accepted synonym for inadequate names in the database.

Publishing central crop databases with contemporary taxonomic information is extremely important in a global medium like the Internet. Many specialist groups working with botanical taxonomy are publishing projects on Internet, and it would be very unfortunate not to be able match them, just because of obsolete or inconsistent nomenclature. A long-term goal should be to link to, or use a common taxon database with, some of the botanical taxonomy projects on the Internet.

Database structure and table relations

The most suitable method of combining biological data with computerized information technology is not always apparent. Branching and formation of hierarchies seem to be common phenomena in both nature and technology. A hierarchical approach has been used here to illustrate some advantages. See Figure 1 for database structure and table relations.

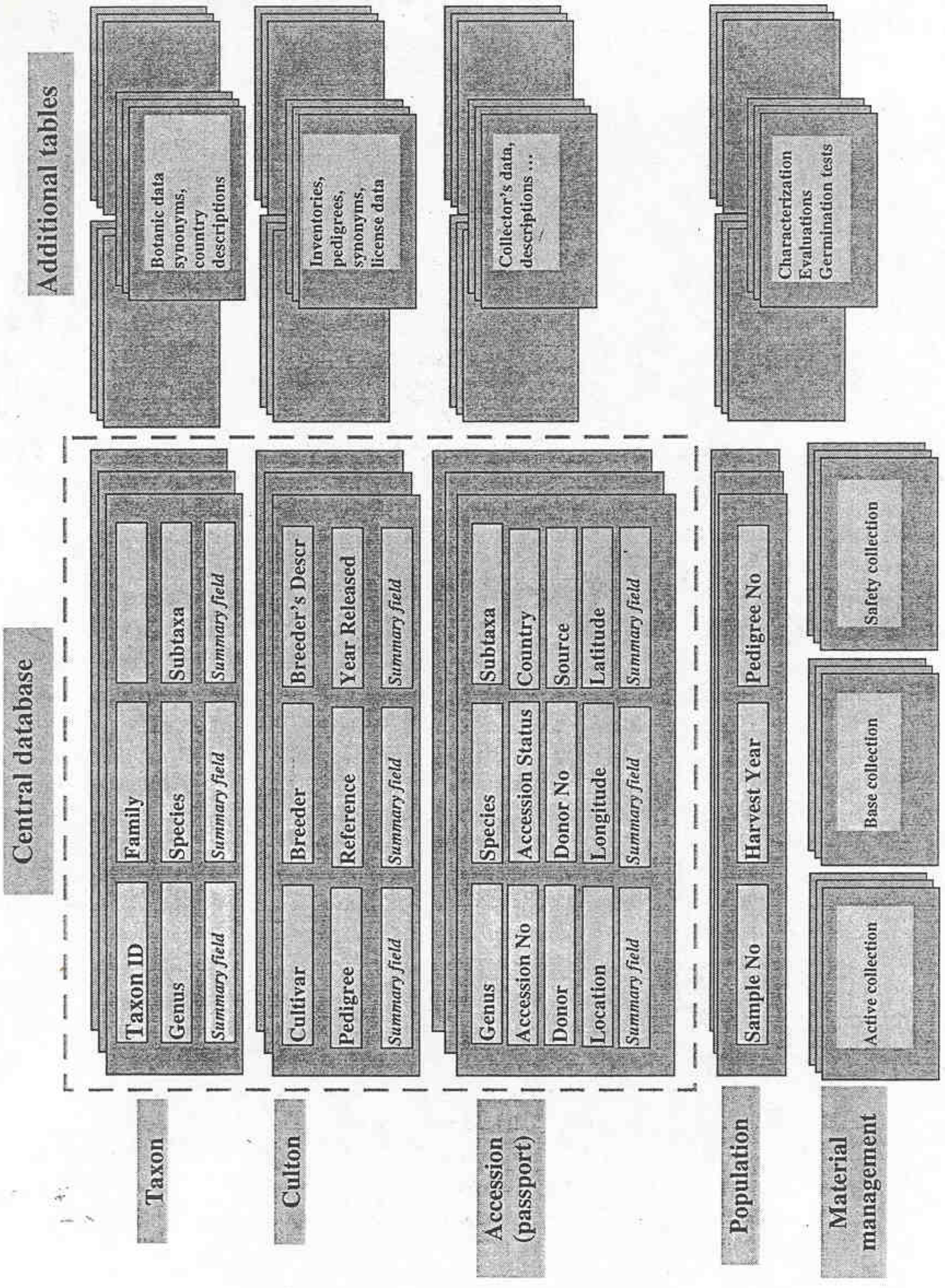


Fig. 1. Database structure and table relations.

The taxon

To change data in original files would be unacceptable. At NGB, a special taxon database has been constructed to preserve all original information in the primary databases. The taxon database contains taxonomic and other botanical information on mandate taxa. The nomenclature information in this file is used in place of the originally registered names in report generation, statistics and data export. A field for taxon number has been added to the passport files, and is used to build a relation between the taxon and passport tables, but also to link horizontally to additional files, e.g. tables with synonyms, country-specific data, botanical data and crop usage. This system allows data to be viewed and exported with up-to-date scientific names and authorities, and at the same time preserves all the original data in the primary databases.

The culton

The botanical nomenclature of cultivated plants is notoriously variable and disputable. Often only older crop plants, with an evolution spanning several centuries or millennia, may be described with acceptable botanical names similar to those used for wild plants, and in accordance with the International Code of Botanical Nomenclature (ICBN). Attempts to use this code, or even the augmented International Code for Botanical Nomenclature for Cultivated Plants (ICBC), for modern varieties with complex pedigrees have resulted in unsatisfactory and ridiculously long and often disputable names. An interesting approach to solve the disputes has been presented by Hetterscheid and Brandenburg (1995). Parts of the taxon names could be replaced with a culton name, which would make no claims on formal taxonomic status or attempt to describe the pedigree of the crop, but still would work as a descriptive and clearly defined name for a crop group.

With the restriction that culta are non-overlapping between taxa, NGB has used an interpretation of the culton concept to facilitate:

- **grouping within taxa.** Crop groups within the taxon can be resolved at the culton level to allow grouping of intraspecific crop groups (convars or analogue groups within *Beta*, *Brassica*, *Pisum*, row types with *Hordeum*), without contradiction to the nomenclature used at the taxon level;
- **description of modern cultivars.** Cultivars of special interest are included as individual records, and fields for cultivar name, breeder, year of release, breeding methods, pedigree, etc. have been included. Cultivars for which such information is not available may be grouped together into summary records (see below);
- **providing summary records for various groups of accessions.** Landraces, breeding material (different types), and any group of cultivars where an individual record is not considered necessary (e.g. at NGB: non-Nordic cultivars).

Other advantages of splitting the cultivars from the passport database into the culton level are:

- cultivar-specific fields from the accession level can be elevated from the passport data table to the culton level, avoiding cluttering of the accession table with fields that are mostly empty. Fields like breeding company, year of release, pedigree, etc., if included in passport table, are empty for all records for which there is no relevance, e.g. landraces;

- additional data from cultivar inventory and pedigree tables can conveniently be linked to one record in the culton level, and not to *n* records at the accession level (resolves many '<–>many' relations into many '<one->many' relations);
- finding duplicates within the accession file is easier since the culton table would contain the preferred (originally registered) cultivar names, with horizontal relations to tables containing synonym names, names licensed in other countries, etc. There would be no need to change the accession name in the passport table, since relations are formed using a special field for culton identification, which is also present in the accession table;
- avoids the conflict between the country of origin and breeding country, sometimes seen when a cultivar bred in one country is collected in another. The breeding country belongs to the culton level, while the country of origin in the passport table always refers to the country where the accession was collected, regardless of where it was bred;
- the culton table could include known cultivars for which no accessions have yet been found (cultivars with no corresponding records at the accession level), making statistics on preservation programme efficiency easier and providing hints for collecting missions and active searches for missing cultivars. Cultivars not yet free from licence and patents can also be included in the culton table, although no material will be available.

The accession

The passport table and the environmental table should have the potential to be merged for all crops, and could form either one large table or two smaller tables in a one-to-one relationship. In practice, for distribution a smaller stand-alone table includes the most important descriptors, while other tables could be linked in on-line systems. A list of the multicrop passport descriptors will be discussed during this conference and may result in a recommendation of a minimal list.

Crop databases compiled for distribution and publishing usually do not contain data from management tables, or other tables below the accession level. In local information systems the management data are essential, but in central crop databases, summary fields may be sufficient. The availability of material and additional data may be indicated by yes/no descriptors, or logical fields. For evaluation data, mean values and standard deviations are more appropriate, integrating evaluation results from different years, environments and sample populations.

The accession database contains both taxon and culton identifications, but in separate fields. This makes is easy to use either or both of the databases at the higher levels, depending on purpose. For example, to list all material available for a given taxon, the culton table need not to be included in the relation.

Descriptors, other than summary fields, are usually not included from below the accession level. A summary of the population level is included here, in order to, at least, note its existence.

The population

A seed sample, or batch of seeds, coming into a seed storage as a new accession, is not a sample of the population that the collector described. It is a sample of the offspring of that population. Changes in the genetic composition, compared with the parent population, have occurred during reproduction, and will continue to do so during storage and subsequent regeneration. The influence of evolutionary forces may be minimized, but cannot be avoided completely (Holden *et al.* 1993). In genebanks, the originally registered accession name and number do not change with the regeneration, but usually some other system is used to distinguish between the different population samples resulting from regeneration. The harvest year or a special field for batch number can be used. The NGB uses a running number starting from 1. A value of zero is only valid for clonal material. One further field is used to allow tracing back any population to the original sample: the batch number of the parent population.

If a selection within the material is done on purpose, e.g. a landrace is split into a number of distinguishable lines, the lines of course get new accession numbers (and are subsequently classified as 'breeding material').

Tables with evaluation results should not only have a field for the accession number, but also one for the batch number. For averaging of evaluation values into summary fields on the next higher level the batch numbers are not needed, but for monitoring and analyzing the genetic changes over time and generations they become increasingly important.

The batch table can be linked horizontally to evaluation tables, most of which are crop-specific. The germination tests are one of the rare (and very simple) evaluations that can be registered in a crop-independent table. Management tables, containing information on seed amounts and storage locations (for clones and *in situ* preservations: storage sites), are analogous to evaluation tables, in respect to their position to the batch table.

Calculated fields

Each of the tables in the taxon-culton-accession hierarchy can contain summary fields from tables at a lower level or in horizontal relations. As an example, the taxon tables can contain fields for the number of known cultivars from different countries (calculated from the culton table), and the culton table can have summary fields for the number of accessions for each culton (calculated from the accession table), and the accession table can have fields for averages and standard deviation (calculated from evaluation tables) and logical fields showing the availability of material.

In on-line central databases, local or on Internet, the calculations can be performed as real-time processes, whenever the information is asked for. But considering speed, the calculations are better performed in advance, or at regular intervals, to allow instant retrieval of the information. Separate tables containing crop-specific summary fields can be used in conditional relations to the accession table, in order to include evaluation data in an on-line central crop database.

Conclusions and ECP/GR considerations

The decision on whether a central database should be constructed as a single file, or as a multi-file information system that resembles the systems used for internal database handling at institutes involved, depends very much on the purpose of the central database, and what user groups it is intended for.

A single-file database containing a limited number of passport descriptors common to all crops could be the first step in constructing a central crop database. A single file is easy to distribute and can easily be accessed without special application software. On the other hand, crop-specific data are difficult to include.

A multi-table database system allows greater flexibility, with better possibilities to include crop-specific information. Access to multi-table databases could be based on special application software, distributed with the data files. On-line versions (on Internet) could use standard browser software on the client's side, with scripts on the server side to achieve the same or better functionality.

Multi-table databases could be a more distant, second goal for a central crop database, with the aim to give more complete access to the information available from characterizations and evaluations.

On-line databases

Internet provides a medium for on-line databases. The most popular protocol in use today on Internet is the HTTP (HyperText Transfer Protocol), which is the basis for the WWW (World Wide Web) network of servers. But other protocols exist, some of them especially planned for access to databases. WAIS (Wide World Information System) servers, with a protocol especially designed to access local and remote databases, form a network of database servers on Internet, similar to the WWW.

WWW pages can be connected to WAIS databases through scripts connected with search and result pages. The two server networks used together provide a very powerful tool to build and maintain on-line central crop databases.

The NGB is running an experimental on-line system of WAIS databases linked to WWW pages at the URL address http://www.ngb.se.

Compilation of primary data tables

An agreement on a standard format and choice of passport descriptors for a central crop database would considerably relieve the task of the primary data provider, compared with the present situation in ECP/GR, where working groups are asking for a different format for every crop. This would also mean that the transfer of data from internal to external (export) databases could be done once for all crops, resulting in shorter average response time for participating genebanks.

In single-table context, the data could be provided in two phases: a minimum table containing a limited set of descriptors, and a larger table containing the remaining crop-independent passport and environmental descriptors.

In a multi-table context, the data providers would also have to provide a table containing information on the cultivars bred in the country or region in question (the culton table), as well as summary tables for characterizations and evaluations.

As Internet becomes more readily available to all ECP participators, file transfer and updating in general become faster and more efficient, and can to some extent be automated.

Coordination of crop-specific tables

The responsibility of the central database managers would include checking the integrity of the files received from the data providers, combining tables from different contributors and relaying the resulting crop table to a data publisher (if the central database manager is also a data publisher: publish the crop table on Internet).

For a multi-table system a central database manager should also process the culton table received from the contributors, and the summary tables for evaluations.

Data publication and access to databases

Publishing central crop databases on Internet means establishing a new duty with the ECP/GR: the data publisher. A data publisher could be any institute with an Internet server, which has agreed to provide one or more of the following database services within the ECP/GR.

FTP service: publishing raw ECP/GR databases for downloading. The raw databases are provided by the central database managers. They are crop-specific, but conform to the same structure and descriptor list. A primary FTP-publisher would be a server where the central database manager agrees to place the table. Secondary FTP servers mirror the file at their own location.

WWW service: publishing ECP/GR home pages, pages with http-links and search pages linked to the databases. Primary WWW-servers would be sites maintaining the pages, secondary WWW-server sites mirroring the primary WWW-servers (and their scripts). Search and result pages can be mirrored if WAIS databases are used, since these can be searched remotely and do not have to be copied with the WWW pages.

WAIS service: publishing searchable databases. This service can be automated as long as the structure of the crop databases is unaltered: when databases at the primary FTP-servers are changed, the new database is downloaded, converted to WAIS format and re-indexed. A primary WAIS publisher is a server who has agreed to be a primary WAIS server for a specific crop. Mirror WAIS-services are possible, but every search page script should not search more than one WAIS server for each crop, to avoid duplicate results.

In multi-tables context the data publishers should coordinate their taxon and culton tables, and others, like the institute code tables and the scripts for displaying the crop-specific tables (evaluations).

Extensive and detailed coordination between data publishers is also needed to agree on primary and secondary services for each type of service and crop.

References

Hetterscheid, L.A. and W.A. Brandenburg. 1995. Culton versus taxon: conceptual issues in cultivated plant systematics. Taxon 44.

Holden, J., J. Peacock and T. Williams. 1993. Genes, Crops and the Environment, pp 89-92. Cambridge University Press. Cambridge.

Multicrop passport descriptors

***Th. Hazekamp*[1], *J. Serwinski*[2] *and A. Alercia*[1]**

[1] International Plant Genetic Resources Institute, Rome, Italy
[2] Food and Agriculture Organization of the United Nations, Rome, Italy

Introduction

To facilitate the management and use of germplasm accessions it is essential that basic identification and characterization/evaluation data be available.

Since 1977 IPGRI, then IBPGR, has been involved in the production of crop descriptor lists. These crop descriptor lists provide internationally accepted definitions for descriptors for different crops and aim to promote the standardized documentation of the crops. So far, 78 descriptor lists covering major and minor crops have been produced (see Table 1). Over the years, the concept and format of the descriptor lists have been subject to change. Initially the crop descriptor lists aimed to provide a minimum set of descriptors. However, this approach left many known descriptors without internationally accepted definitions. Therefore the concept was revised in 1988; since then, comprehensive lists of descriptors are presented. It is by no means intended or assumed that users of the descriptor lists will use all descriptors listed.

Table 1. IPGRI crop descriptor lists (1977-1996).

- Almond (revised) (1985)
- Apple (1982)
- Apricot (1984)
- Avocado (1995)
- Bambara groundnut (1987)
- Banana (1984, 1996)
- Barley (1994)
- *Beta* (1991)
- Black pepper (1995)
- *Brassica* and *Raphanus* (1990)
- *Brassica campestris* L. (1987)
- Buckwheat (1994)
- *Capsicum* (1995)
- *Cardamom* (1994)
- Cashew (1986)
- Cherry (1985)
- Chickpea (1993)
- Citrus (1988)
- Coconut (1992)
- Coffee (1996)
- *Colocasia* (1980)
- Cotton (Revised) (1985)
- Cowpea (1983)
- Cultivated potato (1977)
- *Echinochloa* millet (1983)
- Eggplant (1990)
- Faba bean (1985)
- Finger millet (1985)
- Forage grass (1985)
- Forage legumes (1984)
- Grape (1983)
- Groundnut (1992)
- Kodo millet (1983)
- Lentil (1985)
- Lima bean (1982)
- Lupin/Lupinos (1981)
- Maize (1991)
- Mango (1989)
- *Medicago* (Annual) (1991)
- Mung bean (1980)
- Oat (1985)
- Oca (1982)
- Oil palm (1989)
- *Panicum miliaceum* and *P. sumatrense* (1985)
- Papaya (1988)
- Peach (1985)
- Pear (1983)
- Pearl millet (1993)
- *Phaseolus acutifolius* (1985)
- *Phaseolus coccineus* (1983)
- *Phaseolus vulgaris* (1982)
- Pigeonpea (1993)
- Pineapple (1991)
- Plum (1985)
- Potato variety (1985)
- Quinua (1981)
- Rice (1980)
- Rye and Triticale (1985)
- Safflower (1983)
- Sesame (1981)
- *Setaria italica* and *S. pumilia* (1985)
- Sorghum (1993)
- Soyabean (1984)
- Strawberry (1986)
- Sunflower (1985)
- Sweet potato (1991)
- Tomato (1996)
- Tropical fruit (1980)
- *Vigna aconitifolia* and *V. trilobata* (1985)
- *Vigna mungo* and *V. radiata* (Revised) (1985)
- Walnut (1994)
- Wheat (Revised) (1985)
- Wheat and *Aegilops* (1978)
- White Clover (1992)
- Winged Bean (1979)
- *Xanthosoma* (1989)
- Yams (1980)

The choice of descriptors is up to the individual users who decide which ones will facilitate the management and use of their collections. To assist in the process of selecting appropriate descriptors, a minimum list of highly discriminating descriptors is clearly indicated in the published descriptor lists. Further changes to facilitate the use of the lists have involved the reclassification and reformatting of descriptors. Also, additional site environment descriptors have been included. The format and concept of descriptor lists is constantly under review by IPGRI in collaboration with crop experts to implement improvements whenever possible. Over the years the dynamic nature of crop descriptor list development has led to a situation where some common descriptors are defined in different ways. Especially in cases where genebanks are dealing with multiple crop species, this complicates the documentation. It is often required to modify coding schemes provided in the descriptor lists in such a way that they are fit for use with multiple crops. This has prompted IPGRI to impose more stringent rules for the format in which new crop descriptor lists are developed.

Multicrop passport descriptors

In 1995 a number of documentation specialists in Europe requested that IPGRI, in collaboration with FAO, develop a core list of passport descriptors that would be standard for all crops. This multicrop passport descriptor list would serve several purposes:

- the descriptors in this list could be used in all future IPGRI crop descriptor lists;
- the list would assist in standardizing data exchange for these descriptors.

To further expand the use of the list, it was recommended that it should be compatible for use with the FAO World Information and Early Warning System on Plant Genetic Resources (WIEWS).

In the consequent development of the multicrop passport descriptor list a number of guiding principles were developed.

Initially the list will contain only a limited number of passport descriptors that are important for all crops. Descriptors dealing with the use of germplasm are currently not included, but their suitability for inclusion at the multicrop level will be investigated. Also descriptors such as pedigree that are pertinent to a limited number of crops are not included. If the current proposed concept proves successful and practical, the multicrop descriptor list could be expanded in the future and even more crop group specific annexes could be developed.

When codes have to be used for specific descriptors, numerical coding schemes are proposed. Compared with abbreviations, for example, numerical coding schemes are language neutral.

The draft list below (Table 2) contains two sections. The first section contains the actual multicrop passport descriptors, while the second section contains additional descriptors that are proposed for use with the FAO-WIEWS.

The multicrop passport descriptor list contains the descriptor, its definition and a suggested field name and field length (in brackets) for data exchange purposes.

Concluding remarks

The list presented below is still under development and is not to be considered the definitive one. This first draft has been circulated to various groups outside IPGRI and FAO to obtain feedback. We look forward to receiving more feedback during this meeting to further improve the list and prepare it for final release.

Table 2. Draft Multicrop Passport Descriptors (as presented at the workshop).[9]

MULTICROP PASSPORT DESCRIPTORS
1. Institute code (INSTCODE: C 12) Code of the institute where the accession is maintained. The codes consist of 3-letter ISO country code plus number as specified in the Institute database that will be made available by FAO. Preliminary codes (i.e. codes not yet incorporated in the FAO Institute database) consist of a 3-letter ISO country code and an acronym.
2. Accession number (ACCNUM: C 12) This number serves as a unique identifier for accessions and is assigned when an accession is entered into the collection. Once assigned this number should never be reassigned to another accession in the collection. Even if an accession is lost, its assigned number is still not available for reuse. Letters should be used before the number to identify the genebank or national system (e.g. IDG indicates an accession that comes from the genebank at Bari, Italy; CGN indicates an accession from the genebank at Wageningen, The Netherlands; PI indicates an accession within the USA system).
3. Collecting number (COLLNUM: C 20) Original number assigned by the collector(s) of the sample, normally composed of the name or initials of the collector(s) followed by a number. This item is essential for identifying duplicates held in different collections. It should be unique and always accompany subsamples wherever they are sent.
4. Scientific name (BOTNAME: C 180) Full botanical name with authority. Following abbreviations are allowed: sp.; subsp.; var.; convar. Genus Species Subspecies Botanical variety
5. Cultivar name (CULTNAME: C 80) Either a registered or other formal cultivar designation given to the accession.
6. Country of origin (ORIGCTY: C 3) Name of the country in which the sample was originally collected or bred. Use the ISO 3166 extended codes (i.e. current and old ISO codes).
7. Location of collecting site (COLLSITE: C 254) Subdivision below the country level and other locality information that describes where the accession was collected or the distance in kilometers and direction from the nearest town, village or map grid reference point (e.g. CURITIBA 7S means 7 km south of Curitiba).
8. Latitude of collecting site (LATITUDE: C 5) Degrees and minutes followed by N (North) or S (South) (e.g. 1030S).
9. Longitude of collecting site (LONGITUD: C 6) Degrees and minutes followed by E (East) or W (West) (e.g. 07625W).
10. Elevation of collecting site [m asl] (ELEVAT: C 4)
11. Collecting date of original sample (COLLDATE: C 8) Collecting data of original sample in format [DDMMYYYY] where DD is 2 digits to represent the day, MM is 2 digits to present the month and YYYY is 4

---

[9] The final version of this list is given in Appendix II.

digits to represent the year.

12. Status of sample (SAMPSTAT: C 1)

1 Wild
2 Weedy
3 Primitive cultivar / Landrace
4 Breeders line
5 Advanced cultivar
6 Unknown
7 Other (elaborate in memo or comment field)

13. Collecting source (COLLSRC: C 3)

1. Wild habitat
    1.1 Forest / woodland
    1.2 Shrubland
    1.3 Grasslands
    1.4 Desert / tundra
2. Farm
    2.1 Field
    2.2 Orchard
    2.3 Garden
    2.4 Fallow
    2.5 Pasture
    2.6 Store
3. Market
    3.1 Town
    3.2 Village
    3.3 Urban area (around city)
    3.4 Other exchange system
4. Institute / Research organization
5. Other (elaborate in memo or comment field)

14. Donor institute (DONCODE: C 12)

Code for the donor institute. The codes consist of 3-letter ISO country code plus number as specified in the Institute database that will be made available by FAO. Preliminary codes (i.e. codes not yet incorporated in the FAO Institute database) consist of a 3-letter ISO country code and an acronym.

15. Donor number (DONNUM: C 12)

Number assigned to an accession by the donor.

16. Other number(s) associated with the accession (OTHERNUM: C 12)

Any other identification number known to exist in other collections for this accession (other than Collecting number). Other numbers can be added.

FAO WIEWS DESCRIPTORS

17. Location of duplicates (DUPLSITE: C 12)

Code of the institute where a sample has been duplicated. The codes consist of 3-letter ISO country code plus number as specified in the Institute database that will be made available by FAO. Preliminary codes (i.e. codes not yet incorporated in the FAO Institute database) consist of a 3-letter ISO country code and an acronym.

18. Availability of additional passport data (PASSAVL: C 1)

(i.e. in addition to what has been provided)

0 Not available and not possible to trace back
1 Not available, but may be added in the future
2 Less than 10 additional descriptors available
3 More than 10 additional descriptors available

19. Availability of characterization/evaluation data (CHARAVL: C 1)

0 Not available and will not be added in the future
1 Not available, but may be added in the future
2 Less than 10 descriptors available
3 More than 10 descriptors available

20. Acquisition type of the accession (ACQTYPE: C1)

1 Collected originally by the institute
2 Received as a second repository

21. Acquisition date of the accession (ACQDATE: C8)

Acquisition date of the accession in format [DDMMYYYY] where DD is 2 digits to represent the day, MM is 2 digits to present the month and YYYY is 4 digits to present the year

22. Type of storage (STORTYPE: C 5)

Maintenance type of germplasm. If germplasm is maintained under different types of storage, multiple choices are allowed, separated by a semicolon (e.g. 2;3). (Refer to FAO/IPGRI Genebank Standards, 1994).

1 Short-term
2 Medium-term
3 Long-term
4 *In vitro* collection
5 Field genebank collection
6 Cryopreserved

Handling evaluation data in Central Crop Databases

L. Horváth and G. Szabó
Institute for Agrobotany, Tápiószele, Hungary

Introduction

The rapid advances in computer technology and its increased availability have enabled the use of relatively powerful microcomputers in many remote areas of the world. This facilitated the integration of data on germplasm held in different centres into one computer database, with numerous advantages. General concepts behind the compilation of Central Crop Databases (CCDBs) were developed early by IBPGR. The databases should be organized separately for each crop and they should be located in an internationally recognized 'centre of excellence' for the crop. The latter point helps to ensure dedicated support for the database and makes available expertise for the verification of the data (Perry 1992).

The types of data compiled in a CCDB are identical with the general genebank descriptor categories (Table 1), but an agreement on the role of a CCDB curator in the management of the various types of data is not yet fully reached. Two data types – preliminary and further evaluation data – are of greater interest in this paper. They belong to the fairly permeable characterization-evaluation complex. The important objective of characterization and evaluation is to get more insight into the structure of genetic diversity, which will eventually enable genebanks to better guide plant breeding by more methodical germplasm selection from collections (Frese and van Hintum 1989). Until a collection has been evaluated and something is known about the material it contains, it has little practical use (Chapman 1989).

Interpretability of evaluation data

As verified by the complexity of the evaluation data, their collecting and handling are the most delicate tasks for a genebank curator, not only at the time of their generation but also at the interpretation stage after compiling them in a single or central crop database.

The problems of interpretation of evaluation data are due to the diversity of the genebank's activities:

- the obvious differences among the working systems of the individual genebanks and also the differences between the main purposes and possibilities of the genebank and plant breeding;
- the unavoidable 'site-effect';
- the specific features of the evaluating activity of the genebank, characterized mainly by the great number of collected accessions;
- the various levels of heterogeneity of genebank materials;
- the lack of standardization of the genebank documentation systems.

While the characterization is of common interest to genebanks, the evaluation work depends on individual circumstances and on the perceived needs of breeders. Evaluation needs may well change in response to changes in agricultural practices, farm economics, breeding strategies, and pest and disease races (Chapman 1989). Undoubtedly, the breeders have a justified need for well-established 'prebreeding information'.

Table 1. Crop genebank descriptor categories.

Category	Description	Generating activity	User
Passport	Basic information for the general management of accession: **registration, other passport data**	Introduction Collecting	Genebank Breeders
Management	Basic information for the management of accession in the genebank: **storage, viability (germination), multiplication, regeneration, circulation**	Genebank management	Genebank
Environment and site	Environment and site-specific parameters for origins and trials interpretation: **regeneration, field trials, collecting site description**	Introduction Collecting Regeneration Breeding	Genebank Collector Breeders
Characterization	Qualitative, invariant highly heritable, environment-stable, mono-oligogenic characters, with low G x E interaction: **identification (morphological, cytological, biophysical)**	Genebank trials: multiplication characterization Breeder trials	Genebank Breeders
Preliminary evaluation	Quantitative, environmentally less influenced, oligo-polygenic, easily accessible and measurable characters with medium G x E interaction: **breeding value, identification**	Simple genebank and breeding evaluation trials	Breeders Genebank
Further evaluation	Quantitative, environmentally strongly influenced, oligo-polygenic, not so easily manipulated characters with strong G x E interaction: **breeding value, yield, agronomic performance, stress susceptibility, disease and pest resistance, biochemical traits**	Replicated trials breeding programmes	Breeders

Genotype x Environment interaction

As an additional problem, deriving here from the broad geographical range of genebank activities, the Genotype x Environment (G x E) interactions are certainly a hindrance to full exploitation of evaluation data in different environments. This has the practical consequence of significantly changing the rank order of accessions when it occurs. In practice, it seems that little can be done, since the number of accessions processed in most genetic resources work is extensive and any single accession is likely to be evaluated only a few times, perhaps only once, for any single trait.

Although multi-site or multi-year trials of well-organized networks would be necessary to meet this need, some specific features of the usual genebank practice strongly impede the solution, besides financial, staff and other management constraints. The main intention of a genebank curator is to multiply an accession while maintaining its genetic integrity. Therefore preliminary evaluation work is frequently carried out in the course of the first multiplication of an accession. Sometimes the needs of the two procedures conflict, e.g. insect-pollinated material may need to be grown in cages to prevent contamination and, hence, in a highly atypical environment. Where such a conflict arises, the needs for multiplication must take precedence. A well-maintained accession can always be evaluated in another growing season (Chapman 1989).

The adoption of reference varieties within a network, which are known for their general behaviour or for their behaviour on specific characters in different environments, is of help (Perret 1990). For this reason a number of check lines need to be included as standards for comparison, but the choice of check lines depends very much on circumstances and very often it is not easy to choose reliable check lines because of the various interests of breeders and of the broad geographical range of the planned trials. For instance in the case of disease-resistance evaluations the main impeding factors can be the merits of natural infestation, as well as the occurrence of diverse races of diseases in different regions. In other cases the photo- or thermosensitive reaction of the check line candidates complicates the approach of G x E interaction (IBPGR 1991).

Heterogeneity in genebanks

Many of the genebank accessions, and probably the majority in any collection, will be genetically heterogeneous. This is inevitable for outbreeding species, but also occurs in landraces and wild population of inbreeding species. Only clonal material or ‘advanced’ inbreeding material is likely to be uniform (Chapman 1989). The genetic diversity can be observed at various levels: within field populations, among populations, among regions and as species mixtures.

The heterogeneity of populations complicates evaluation. In subsequent evaluation and selection, it may be necessary to quantify intrapopulation variation, to determine breeding potentials of the best-performing populations (Elings 1992).

It does not follow that the heterogeneous accessions will show variability for all traits; in many respects they may appear uniform and a single score will adequately describe the whole. Where between-plant variation is apparent, it is desirable to take it into account. But a completely satisfactory approach does not appear to exist, because the use of some solutions can result in a much larger collection or database, while the others do not always provide sufficient information.

Another possible approach of the accession heterogeneity-identity issue is the use of biochemical monitoring methods, such as seed protein and isozyme work, RFLPs, RAPHDs, etc. The data obtained this way do not always have true breeding value in themselves, but they can be used as marking tools of the accession identity or heterogeneity. The application of these methods is fairly restricted in some genebanks, because of technical and skills insufficiencies.

Redundancy in the collections

The great amount of accessions (or sometimes their redundancy) is a major problem in the genebank collections. Such numbers may be too large for a collection manager to evaluate in depth, because the evaluation is costly and site-dependent.

To solve the problem within the framework of international crop networks, it is recommended that a limited number of accessions be selected and that the characterization and evaluation work be shared among institutes, but also the 'hot-spot' evaluation concept can be taken into account. The difficulties which can occur here are similar to those mentioned for the 'check lines'. Additional financial resources would also be necessary for such studies.

Another response to this matter is the 'core collection' concept which has recently been developed further. The objective was to identify a set of accessions which would represent and cover the major kinds of genetic diversity known to be present in the crop and its wild relatives. Both the theoretical and practical aspects of developing core collections raise a number of problems. Information on the genetic diversity present in a collection is required, the size of the core needs to be determined and sampling methods have to be developed which maximize the genetic variation present in the core (Hodgkin 1991).

Diverse documentation systems

The major bottleneck for compiling and interpreting a CCDB is the lack of standardization among the participating documentation systems.

One of the central activities of genebanks is to document the accessions in their collections. More and more information on plant genetic resources is made available in electronic form. Many different documentation systems are used to generate and manage these data. A major problem related to data interchange is the actual interpretation of data by the receiving party. There is no doubt that to further improve international collaboration and to achieve better access to information on plant genetic resources on a worldwide scale, the implementation of standards in plant genetic resources documentation is beneficial.

Nevertheless, the widespread implementation is by no means easy to accomplish. Plans for full-scale standardization, including computer hard- and software and the data themselves, have been developed in the past, but it rapidly became clear that this was not practical. Factors that determine the availability of information on species are technology, technical skills and local users' needs and preferences.

Through its programme activities IPGRI aims to provide a framework to facilitate the implementation of such standards and, since 1977, it has produced 78 crop descriptor lists. Each list provides users with guidelines for the uniform description of accessions for a particular crop. The descriptor lists include descriptors, descriptor states and coding schemes (Hazekamp 1995). Although it would be difficult to assess the direct positive impact of this action, there is no doubt that it has already facilitated the standardization of description states and coding systems.

Gaps in compatibility

The first issue of genebank documentation strongly related to the concept of central crop databases is compatibility. Data have to be exchanged and incorporated in a single central system. Hardware compatibility is hardly a problem anymore since conversion procedures are available. The only problem on this level are data that have not been computerized yet, but this problem will be solved with time (Hazekamp and van Hintum 1990).

Although the physical compatibility problems have been largely solved, the logical compatibility problems remain. Genebanks differ, as well as the available hardware, software and computer science expertise. Their information systems are, therefore, also different. Conversion of the data into another structure is labour-intensive, and may, in addition, cause loss of information. A special compatibility problem is that of the language. National documentation systems tend to use the national language and/or English (Knüpffer 1995).

Corrective devices

Mandatory characters

The first step to improve the data interpretation may be the proposal of a set of mandatory characters. They would be of primary importance and would be selected by national curators in collaboration with breeders. It may be questioned again, how the list of these selected descriptors could be of common interest and how the G x E factor would work in the many countries evaluating germplasm for the same characters.

On a next level of standardization the germplasm documentation would be reorganized and new data types can be used.

File dictionary

The data files often lack adequate description; this prevents unambiguous interpretation by others. One way of dealing with this is to accompany the data files with a separate data dictionary. The data dictionary describes the data fields and gives a full description of any codes that are used (Hazekamp 1995).

Meta-data

A solution chosen by several ECP/GR Working Groups to resolve these problems is to include so-called ‘meta-data’ on characterization and evaluation in their databases. Meta-data help the user to find out where characterization and evaluation are available, indicating which genebank holds data on a given trait. With the help of the database manager the user can then obtain the data sets on this species from the holding genebank and use his own expertise to interpret them (Faberová and van Hintum 1995).

Site data

As previously stressed, the obvious problem of evaluation data, and especially quantitative data, is the influence of environment on the phenotype. Extensive data sets containing detailed site data (weather and soil characterization) of location in various agro-ecological regions should be available at the genebank. As is recommended, two evaluation methods can be followed, depending on the significance of the relationship between the observed plant character and environmental characteristics of the collection site. If such relationships are significant and casual, then a preliminary selection of germplasm can be based upon collection site characteristics. Whereas only a limited part of a large germplasm collection undergoes multi-location and multi-seasonal evaluation, crop growth analysis can be applied to an entire collection (Elings 1992). The increasing availability of evaluation data for the same material in different environments (and recorded with the same descriptors) through the crop network documentation system will stimulate research on G x E interactions. Results of these researches, if they will not allow the prediction of the exact value of a character from an environment to another, will at least provide useful clues for the selection of preliminary material by users in different parts of the world (Perret 1990).

Simulation models

Another possibility to help solving difficulties of interpretation is the use of modern, computerized simulation models for the analysis of data of various origins. Appropriate computer hardware and software is required for this and training in modelling and simulation will be necessary for curators who are unacquainted with these methods (Elings 1992).

For instance a statistical method is recommended to solve the number of problems encountered in attempting to use data collected over a number of years (Giles 1989). After evaluating the recommended models it was admitted that, if the theoretical basis for such analysis comparing different sets of germplasm accessions evaluated at different locations and different years was rather thin, the practical value for researchers was high due to the large amount of information available (IBPGR 1989).

Another model (RoDMod: Rate of Development Model) provides a method of transforming plant germplasm evaluation data into characterization data. The descriptors which comprise evaluation data include those genes which control crop duration; this character is often critical to crop adaptation, particularly in the more marginal and disadvantaged environments for agriculture. The authors have developed three simple linear equations which, when combined, describe numerically the effect of environment on duration from sowing to flowering (Watkinson *et al.* 1994). Further research would be necessary to determine whether this model can be successfully used in other aspects of transforming site-affected evaluation data.

New documentation systems

As a nearly final step of standardization the entire genebank documentation might be replaced by a more universal and updated documentation system. The Genebank Management System (GMS) released by IPGRI could be a possible device for this, but, undoubtedly, it was prepared for the ‘late-comers’, and it is obvious that the genebanks with advanced documentation systems are not too interested in converting their full data set. However the new version, supported by the dBase-GMS converting software, makes it easier to transfer at least DBF data into the Paradox system used in GMS.

The status of evaluation data in CCDBs

To summarize the current status of the issue of evaluation data in CCDBs, it must be emphasized that evaluation data are the most important to plant breeders and other users. However, this category of data is also the most problematic while the usefulness of diverse evaluation data in an international crop database remains questionable. Two types of arguments are prevailing: the first one is based on biological considerations, mainly on the G x E interaction, whereas the second type of arguments deals mainly with the reluctance of national breeding programmes to share data which could be exploited through the use of relevant material without any return to them (Perret 1990). Sometimes the inclusion of evaluation data into the international documentation system was considered, at least at this stage, as unrealistic. However, it is outlined that easy accessibility of passport data, together with information on a number of accessions evaluated for specific descriptors, was the best way to stimulate exchange of evaluation data. The establishment of cooperative evaluation projects would be primarily a task for breeders. Consequently it should not be a priority for a CCDB.

The range of descriptors included in a central database differs. The most important part, the core of such a database, consists of passport data. Some central databases, especially the smaller ones, also include characterization and evaluation data. Larger databases did not include such data, because of the difficult logistics involved in keeping such databases up to date, especially some years ago, when computerized communication was not as easy as it is nowadays (Knüpffer 1995). According to the report of the European *Avena* Database (EADB), up to the reporting time only six partners of EADB had provided evaluation data and the number of descriptors from the contributing institutions varied between 2 and 60 (Schittenhelm and Seidewitz 1993). Combining scores in one database can be very complicated, even if the scores are comparable. It might be better to store the data as they are, and leave the interpretation, standardization and transformation to the user, who should be the expert (Faberová and van Hintum 1995).

Conclusion

It may be concluded that the smaller the geographical range and the intraspecific diversity of the given species, the better the possibility to handle evaluation data in a CCDB. Nevertheless, the database manager cannot supply such dedicated support for the verification of these data as he can do for other data types, or as can be done by the other participants involved in a compiling activity (breeders, crop or national curators).

Today more and more genebanks have direct access to international computer networks and this number will certainly increase in the near future. An alternative is to have data circulated via an international network and left to the interested parties to interpret. However, the basic concept of the CCDB holds many advantages, which would be lost if the compiling of evaluation data was fully neglected. Consequently, it seems to be more recommendable to build up a ‘fit for use’ central database with the aid of a proper software which makes it possible to handle evaluation data ‘as they are’, and also makes transformation or interpretation by the user possible.

References

Chapman, C. 1989. Principles of Germplasm Evaluation. Pp. 55-64 *in* Scientific Management of Germplasm: Characterisation, Evaluation and Enhancement (H.T. Stalker and C. Chapman, eds.). IBPGR, Rome, Italy.

Elings, A. 1992. Evaluation Methods for Large Germplasm Collections; Distribution, Variation and Evaluation of Syrian Durum Wheat Landraces. Wageningen, The Netherlands.

Faberová, I. and Th.J.L. van Hintum, 1995. Standardization in plant genetic resources documentation VI: Characterization and evaluation data. Pp. 109-113 *in* Standardization in Plant Genetic Resources Documentation. Report of the Second Technical Meeting of Focal Points for Documentation in East European Genebanks (Th.J.L. van Hintum, M.W.M. Jongen and Th. Hazekamp, eds.). Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.

Frese, L. and Th.J.L. van Hintum. 1989. The International Data Base for *Beta*. Pp. 17-45 *in* Report of an International Beta Genetic Resources Workshop held at the CGN, Wageningen, the Netherlands, 7-10 February 1989. IBPGR, Rome, Italy.

Giles, R. 1989. The handling of evaluation data in the AFRC Barley collection. Pp. 44-51 *in* Report of a working group on Barley (third meeting) held at the Zentralinstitute für Genetik und Kulturpflanzenforschung (ZIGuK), Gatersleben, GDR, 18-20 April 1989. IBPGR, Rome, Italy.

Hazekamp, Th. 1995. IPGRI's role in the standardization of plant genetic resources documentation. Pp. 115-118 *in* Standardization in Plant Genetic Resources Documentation. Report of the Second Technical Meeting of Focal Points for Documentation in East European Genebanks (Th.J.L. van Hintum, M.W.M. Jongen and Th. Hazekamp, eds.). Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.

Hazekamp, Th. and Th.J.L. van Hintum. 1990. Central crop databases in collaborative genetic resources management. Pp. 37-42 *in* Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management Management (Th.J.L. van Hintum, L. Frese and P. Perret, eds.). Papers of the EUCARPIA / IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

Hodgkin, T. 1991. The core collection concept. Pp. 43-48 *in* Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management Management (Th.J.L. van Hintum, L. Frese and P. Perret, eds.). Papers of the EUCARPIA / IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

IBPGR. 1989. Report of a working group on Barley (third meeting) held at the Zentralinstitute für Genetik und Kulturpflanzenforschung (ZIGuK), Gatersleben, GDR, 18-20 April 1989. Pp. 1-11. IBPGR, Rome, Italy.

IBPGR. 1991. Report of a Working Group on Sunflower (fourth meeting) held at Facoltá di Agraria, Universitá degli Studi di Pisa, Pisa, Italy, 11-12 May 1991. ECP / GR-IBPGR, Rome, Italy.

Knüpffer, H. 1995. Central crop databases. Pp. 51-62 *in* Standardization in Plant Genetic Resources Documentation. Report of the Second Technical Meeting of Focal Points for Documentation in East European Genebanks (Th.J.L. van Hintum, M.W.M. Jongen and Th. Hazekamp, eds.). Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.

Perret, P.M. 1990. Actual and future concepts for collaboration in crop genetic resources. Pp. 13-22 *in* Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management Management (Th.J.L. van Hintum, L. Frese and P. Perret, eds.). Papers of the EUCARPIA / IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

Perry, M. 1992. Documentation and Databases for Germplasm Collections. Pp. 157-162 *in* Report of the first meeting of the International Network for Cassava Genetic Resources held at CIAT, Cali, Colombia, 18-23 August, 1992. IPGRI, Rome, Italy.

Schittenhelm, S. and L. Seidewitz. 1993. Progress on the European *Avena* database (EADB). Pp. 37-49 *in* Report of a working group on *Avena* (fourth meeting) held at the Agricultural University, Gödöllõ, Hungary, 26 to 28 May 1993 (E.A. Frison, H. Koening and S. Schittenhelm, eds.). IBPGR, Rome, Italy.

Watkinson, A.R., R.J. Lawn, R.H. Ellis, A. Qi and R.J. Summerfield. 1994. RoDMod. A computer program for characterising genotypic variation in flowering responses to photoperiod and temperature. CSIRO / IPGRI / University of Reading, England.

Monitoring the use of Central Crop Databases

W. Podyma
Plant Breeding and Acclimatization Institute, Radzików, Poland

Summary

Central Crop Databases play a key role in the management of collections within the framework of the European Cooperative Programme for Crop Genetic Resources Networks (ECP/GR). The information system needs to be improved and the increasing use of the Internet may help to facilitate access to information and communication between curators and users of plant genetic resources. The assessment and continuous monitoring of users' needs are crucial to the development of the system.

Introduction

Central Crop Databases (CCDBs) play a key role in the management of collections by the crop-specific working group. Overview of and access to European crop genetic resources are facilitated by the large number of crop-specific databases. Some of the central crop databases have been functioning for many years, e.g. the European Barley Database (Knüpffer 1991) or the International Data Base for *Beta* (Frese and van Hintum 1989), but many of them are being newly developed within the framework of the European Cooperative Programme for Crop Genetic Resources Networks (ECP/GR). All these databases build a wide information system on plant genetic resources.

When building an information system, proper attention to all its components is crucial for the information system design. The information system design should cover information system development methodologies, with special emphasis on the conduct of users' needs assessments, the use of database development techniques, and issues relating to the framework within which the system is developed. The aim is to establish an effective network of individuals and institutions possessing collections of plant genetic resources and to facilitate communication between the owners and the users of these collections. Many aspects of the development of central crop databases were discussed before (Hazekamp and van Hintum 1991) but the issue of users seems to be partially neglected. The process leading to the production of the functional specification of CCDBs should also consider some key outputs, i.e. a clear definition of intended users, and precise information needs of these users.

The users of Central Crop Databases

The users of the CCDBs, whatever their function, can be described as:

- the people who provide originally collected data;
- the people who develop and maintain the crop databases;
- the people who create and disseminate information, and
- the people who receive the information.

The users can be divided into two groups; primary users and end users. The first two categories are **primary users:** they are database managers, curators and members of crop-specific working groups seeking to secure the conservation of a genepool in the most effective way. To these users, the CCBD remains an essential management tool for the identification of duplicates and omissions in the collections, as well as a basic source of information for the development of collaborative activities such as the establishment of core collections, the planning of collecting missions, etc.

End users include breeders, researchers, educational establishments and other users who wish to access the CCDBs to obtain specific information about the collections. End users not only have different expectations from the databases, they also have practically a lower level of accessibility to data. This is because the data are often stored in obsolete or inaccessible forms, or are poorly standardized and documented. End users are seldom able to use raw, unprocessed data. They require data relevant to some particular issue to be selectively extracted, integrated with other data. The database manager should also bear responsibility for being a source of advice on those modelling and analytical tools that are most appropriate for the data concerned.

Database managers are the main users of CCDBs. They perform very important work on standardization of data, identification of duplicates, etc. Data providers also are strongly involved by the provision of data to CCDBs (Fig. 1). Data from CCDBs return to genebanks in the form of databases or printouts. The links between primary users are strengthened by activities carried out in the framework of working groups or *ad hoc* initiatives. Currently, the lowest flow of information is between CCDBs and end users. This situation is due to different reasons: limitation of accessibility to CCDBs is the main problem because most databases are still under development. The structures of the databases are different, not standardized. CCDBs maintainers do not register systematically the use of the databases.

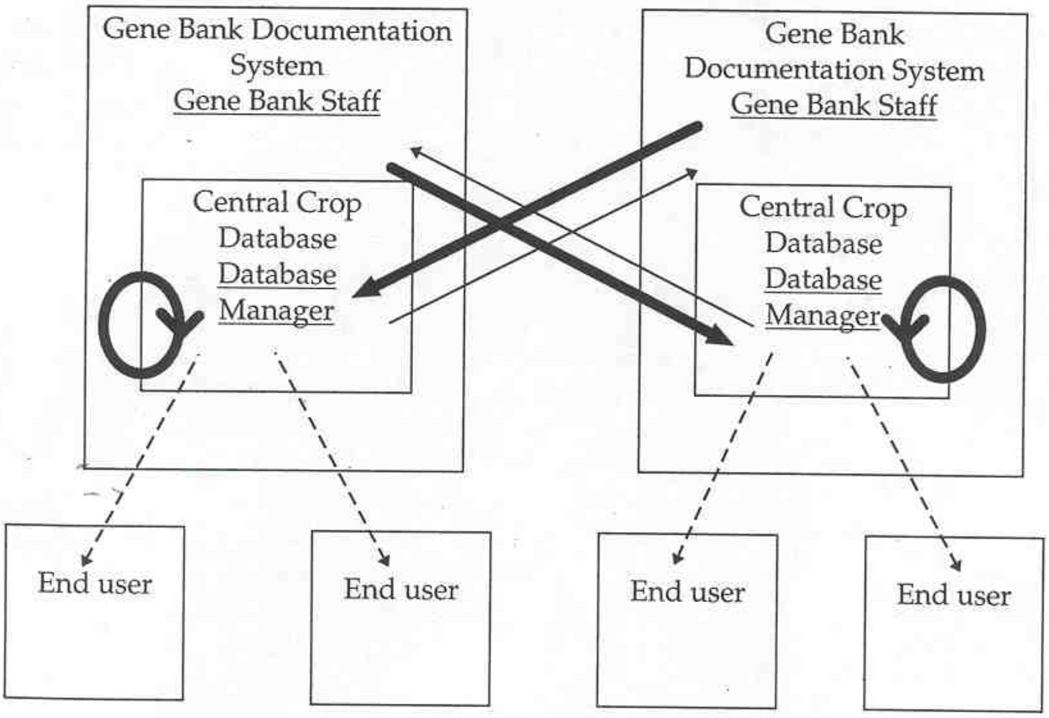


Fig.1. Data flow between different groups of users.

Methods of users registration

Most of the existing central databases are off-line systems. The databases can be accessed by contact with the database manager, or compiled databases are provided to potential users in printed or electronic form.

Registration of users of off-line databases:

- no system of registration;
- correspondence databases;
- register of users.

Most off-line CCDBs do not have any system of user registration, very few use correspondence databases or a register of users to monitor the use of data. Genebanks have developed methods of registration of seed exchange, but systems for recording data exchange are rarely used.

An alternative to the off-line systems is an interactive on-line system.

Registration of users of on-line databases:

- log file on server;
- WWW server statistics.

Managers of on-line databases provided via Internet have at their disposal useful tools to analyze access to data. Log files record the usage of servers. Statistical programmes on the server make it possible to summarize the use of the servers. Different types of services are available. Sometimes the results of these analyses are provided on the home pages themselves (e.g. GENRES, GRIN). ZADI's server automatically produces in eleven tables weekly statistics of the use of its Internet pages.

The database manager using available shareware programmes[10] can generate such statistics. He needs access to the server and has to know the path where the log file is located. The log file can also be copied by file-transfer protocol (FTP) to the local disk and analyzed there. Of course the information stored in the log file is only part of the information provided to users by the database manager; in everyday practice, different communication media are used.

Assessment of users' needs

The assessment of users' needs is crucial for the information system design. Without a proper analysis of users' needs, time can be wasted providing the wrong kinds of services, and systems can be designed which fail to address key problems. Methods of assessing users' needs can be different: consultation between developers and prospective users, modelling or prototyping. But the most effective way is to analyze the use of databases. Now we possess very fragmentary data on the use of CCDBs. Searches in the database mostly concern specific accessions. Sometimes specific characters or groups of origin are requested. Researchers additionally search for information on molecular markers or biosystematics.

Conclusion

Several milestones were reached in the process of building the CCDBs. Now we have a concrete, workable architecture supporting the desired links. Much of the future work should involve analysis of human factors. The work should be

---

[10] (ftp://ftp.boutell /pub/boutell/wusage)

directed towards the establishment of mechanisms for the monitoring of users' access to the databases and their needs.

References

Frese, L. and Th.J.L.van Hintum. 1989. The International Data Base for *Beta*. Pp. 17-45 *in* International Crop Networks Series. 3. Report of an International Workshop on *Beta* Genetic Resources. IBPGR, Rome.

Hazekamp, Th. and Th.J.L. van Hintum. 1991. Central Crop Databases in Collaborative Genetic Resources Management. Pp. 37-42 *in* Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management Management (Th.J.L. van Hintum, L. Frese and P. Perret, eds.). Papers of the EUCARPIA/IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

Knüpffer, H. 1991. The European Barley Database. Pp. 93-100 *in* Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management Management (Th.J.L. van Hintum, L. Frese and P. Perret, eds.). Papers of the EUCARPIA/IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

Access to Central Crop Databases

D. Jiménez Krause[1] *and E. Lipman*[2]
[1] ZADI/IGR, Bonn, Germany
[2] IPGRI, Rome, Italy

Introduction

During the last 15 years, great efforts were put into the compilation of data for the establishment of Central Crop Databases (CCDBs). It is essential that this work be well valorized, so that the CCDBs fulfil their goal of being useful 'management tools' for germplasm collections throughout the region, and not only an accumulation of data. Until recently, the only way of accessing such a database was to ask the database manager either to perform a query on the database, or to send an electronic copy of the data, usually on a diskette, if the user had a compatible software allowing the importation of files on her/his computer and execution of the search locally. Answering these information requests can be very time-consuming, and given the often very broad spectrum of responsibilities of the database manager, the whole operation sometimes encounters serious bottlenecks. Making the data available in such a way that the user can find the requested information without resorting to the database manager is becoming a priority. The increasingly available 'new information technologies' and specifically the Internet offer interesting possibilities for the improvement of access to CCDBs, which will be discussed below.

Different types of access to CCDBs

The advantages and inconvenience of different types of access to the existing CCDBs are summarized in the table below.

Table 1. Ways of accessing CCBDs: pros and cons.

Type of access	Database	Pros	Cons
on-line	on-line database on Internet	• optimal access • optimal searchability	• limited availability (Internet connection needed)
off-line/ electronic	off-line database on Internet (for downloading)	• very affordable • optimal access • good searchability	• limited availability (Internet connection needed)
off-line/ electronic	database on diskette or CD-ROM	• available upon request • good searchability	• high cost of distribution software, if any • high cost of updating
off-line/ hardcopy	database content in printed form	• available upon request	• high cost of updating • limited searchability

The constraint of limited availability of on-line and off-line Internet-based databases is decreasing as connection to Internet is becoming a standard in most European institutions dealing with plant genetic resources.

Setting up on-line databases on the Internet requires special software and hardware, a permanent, expensive Internet connection and technical expertise.

On the other hand, on-line access has proven to be most efficient since it allows interactive and real-time database searching. Establishing on-line Internet CCDBs can greatly improve the access to plant genetic resources information and thus the use of plant genetic resources. The technical barriers can be overcome by the use of synergetic effects and by interinstitutional cooperation.

Off-line CCDBs on Internet are a pragmatic and affordable way of making plant genetic resources data available. A user can download the whole database via Internet to a hard disk and perform an off-line search using standard database software. In this case an excellent description of the database (descriptor and decoding tables) is essential in order to guarantee meaningful use of the data.

Databases on diskette and/or compact disk require either a thorough description of the database (when it is to be searched with standard database software), or a special distribution software such as PC-GRIN.[11] The distribution software highly facilitates the access to the data, but its development and maintenance are quite expensive.

The printed form of a database does not allow such flexibility in searching the data as does the electronic product, but it still is and will remain a useful complement to electronically distributed data.

Improving access to CCDBs means first of all making the databases accessible in either one of the ways described above, the effectiveness of the information transfer increasing from the hardcopy to the on-line Internet database. An Internet-based platform to provide central access to existing CCDBs would highly increase their use. The characteristics of such a platform are detailed below.

Principles for the establishment of an Internet-based European information system on plant genetic resources

The need for a region-wide information system on plant genetic resources arises from the low accessibility to databases, the partially low quality of their contents, and an insufficient level of communication between the different institutions involved in plant genetic resources activities.

The establishment of an Internet-based European information system on plant genetic resources is proposed to help resolve these constraints. This information system would:

- provide central access to the CCDBs;
- support the CCDB managers;
- serve as a communication platform;
- serve as a tool in the coordination of activities, organization of workshops, etc.;
- provide general plant genetic resources information;
- provide references to specialized plant genetic resources information.

The system could be developed by ECP/GR, formally identified as the agreed platform for plant genetic resources activities in Europe.[12] The establishment of the information system would be based on feasibility, voluntary participation, self-dynamism (implying a bottom-up approach, the use of synergies, with natural development) and should involve minimal or no extra costs for the members of the system.

---

[11] Data Query system for the PC, Germplasm Resources Information Network (United States Department of Agriculture).

[12] At the Regional preparatory meeting for the IVth International Technical Conference on plant genetic resources for food and agriculture, Nitra, Slovak Republic, September 1995.

Objectives

The main objectives of the system would be:

- to increase the quality and accessibility of information on plant genetic resources;
- to improve communication between all plant genetic resources institutions;
- to increase the utilization of genebank collections.

The proposed system would therefore stimulate a better coordination of plant genetic resources conservation strategies, leading to increased effectiveness of activities such as location of duplicates, setting up core collections and plant breeding.

Structural issues

The backbone of the system would be the Internet. The traditional means of communication (personal contacts, publications, telephone and telefax, regular mail, floppy diskettes) would not be excluded but rather enhanced by the use of this new technology.

The Internet provides a standardized communication platform and allows low-cost networking. Its major advantages are its increasingly availability, its simplicity of use and its interactivity. The World Wide Web (WWW) service of the Internet makes hypertext documents, pictures, on-line databases and references available. Email and mailing lists enhance the communication between partners. Other Internet services, like ftp (file transfer protocols) for data transfer, can also prove very useful in the networking context.

The proposed information system would be based on the WWW and would comprise the following main modules:

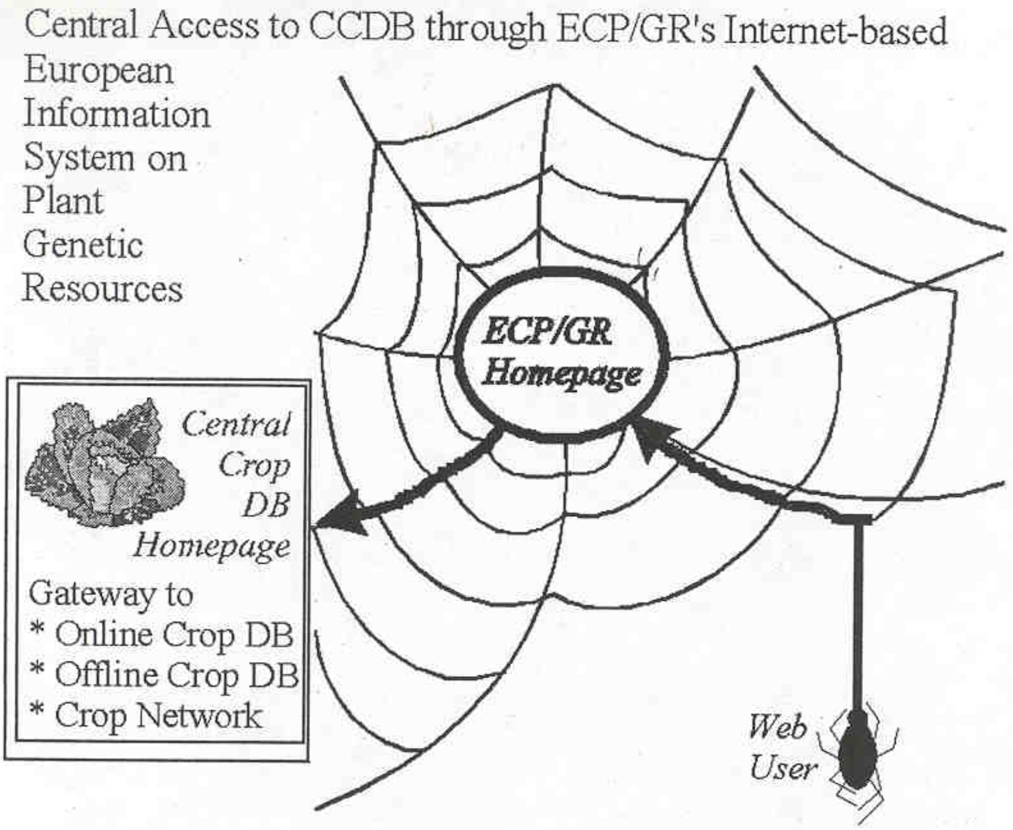
- a central access point to CCDBs (see Fig. 1) by creating and updating a reference catalogue of CCDBs;
- a communication platform providing lists of partners of all European plant genetic resources institutions (addresses, phone/fax numbers, and emails); a ‘Net-publishing’ medium support (e.g. newsletter); general information related to the field of plant genetic resources: job offer list; activities schedules; discussion lists and access to related newsgroups; etc.
- a gateway to other relevant plant genetic resources information, e.g. by providing references to genome and DNA-sequence data, taxonomy databases, etc.

Besides the establishment of the WWW system, the system would also encourage and support the creation of CCDB home pages and support the CCDB managers in setting up on-line access to their databases.

Challenges

The system should provide:

- a minimal logic for the CCDB home pages;
- data structure standards;
- data transfer protocols;
- updating strategies, and
- a seed-requesting mechanism allowing communication between the Internet user and the institution holding the desired germplasm.



Technical support

The support tasks would be carried out by Documentation Support Centres (DSCs), which are plant genetic resources institutions with human and technical resources, general expertise in Internet (e.g. experience in using HTML, ftp), general expertise in database management, including on-line and off-line databases on the Internet. So far CGN, IPGRI, NGB and ZADI have offered to act as Documentation Support Centres. Furthermore, NGB and ZADI already support other institutions in creating their own Internet information systems.

The system would become a virtual network with the establishment of CCDB home pages and the central reference catalogue to CCDBs. The stepwise and bottom-up approach for the implementation of the system would be supported by the institutions with more expertise in the field of databases and Internet. This system would greatly improve the access to plant genetic resources information in Europe and consequently improve the use of genebank material.

Implementation and possible *modus operandi* of the system

The ECP/GR Documentation and Information Network

The high concern of ECP/GR for issues related to documentation of plant genetic resources was illustrated by the establishment of the 'Documentation and Information Network' by the ECP/GR Steering Committee meeting in Nitra, Slovak Republic, in September 1995.

The activities of this network are launched with the present workshop, which is also expected to set priorities and establish a workplan. This ECP/GR Documentation and Information Network can become an effective and integrating tool for the implementation of the above-mentioned principles.

At present, the components of the system can be listed as follows:

1. European Crop Databases established within the framework of ECP/GR. The network is also open to other relevant databases (Sunflower, *Beta*, Potatoes);
2. national documentation/information systems on plant genetic resources for countries who wish to include these in the network;
3. a number of Documentation Support Centres (DSCs) as mentioned above (e.g. CGN, IPGRI, NGB, ZADI).

These components already have a high level of interaction through classical communication media and carry out joint or collaborative activities which include support to database managers, organization of symposia and workshops, training activities, publications, etc.

The fast development of the Internet must be fully exploited to improve the access to information on collections and therefore access to germplasm itself – it is not unrealistic to think that within the next 5 years most providers and users of plant genetic resources information in Europe (CCDBs, genebanks, agricultural research institutes…) will have access to Internet. Figure 2 shows how the existing crop databases could be interlinked via Internet, so that the potential users would be offered access to a wide range of information about the collections maintained in the genebanks.

Taking into account the current projects underway in some European countries, it seems likely that a number of European Crop Databases will have their own specific site in the near future (e.g. the *Prunus*, Wheat, *Malus* databases, a number of forages databases, etc.). However, many of the institutes who are presently hosting the European central crop databases will probably never have an independent server, because of institutional or national policies, technical reasons or most frequently financial constraints. For these institutes, the alternative already mentioned would be to load the central database for which they have taken responsibility onto the server of a DSC, as is already the case for a number of national databases hosted by ZADI (Germany) and NGB (Baltic States).

Access point to the system

The recently established ECP/GR home page[13] would provide an excellent node for the organization and redistribution of information on plant genetic resources activities in Europe. Most items listed as the 'main modules' of the information system presented in the previous section are already present. The home page currently offers the following information:

- objectives and structure of the ECP/GR;
- contact details of National Coordinators, Focal Points and Working Groups members;
- list of Central Crop Databases including addresses of host institutes and names and contact details of their managers;
- full text of the *Newsletter for Europe* published by IPGRI;

---

[13] URL: http://www.cgiar.org/ecpgr/

- information on upcoming meetings, vacancy announcements, list of publications, etc.

landscape
BUD_FIG2.doc

Fig. 2. Role of Internet in improving access to central crop databases (CCDBs); DSC=Documentation Support Centre.

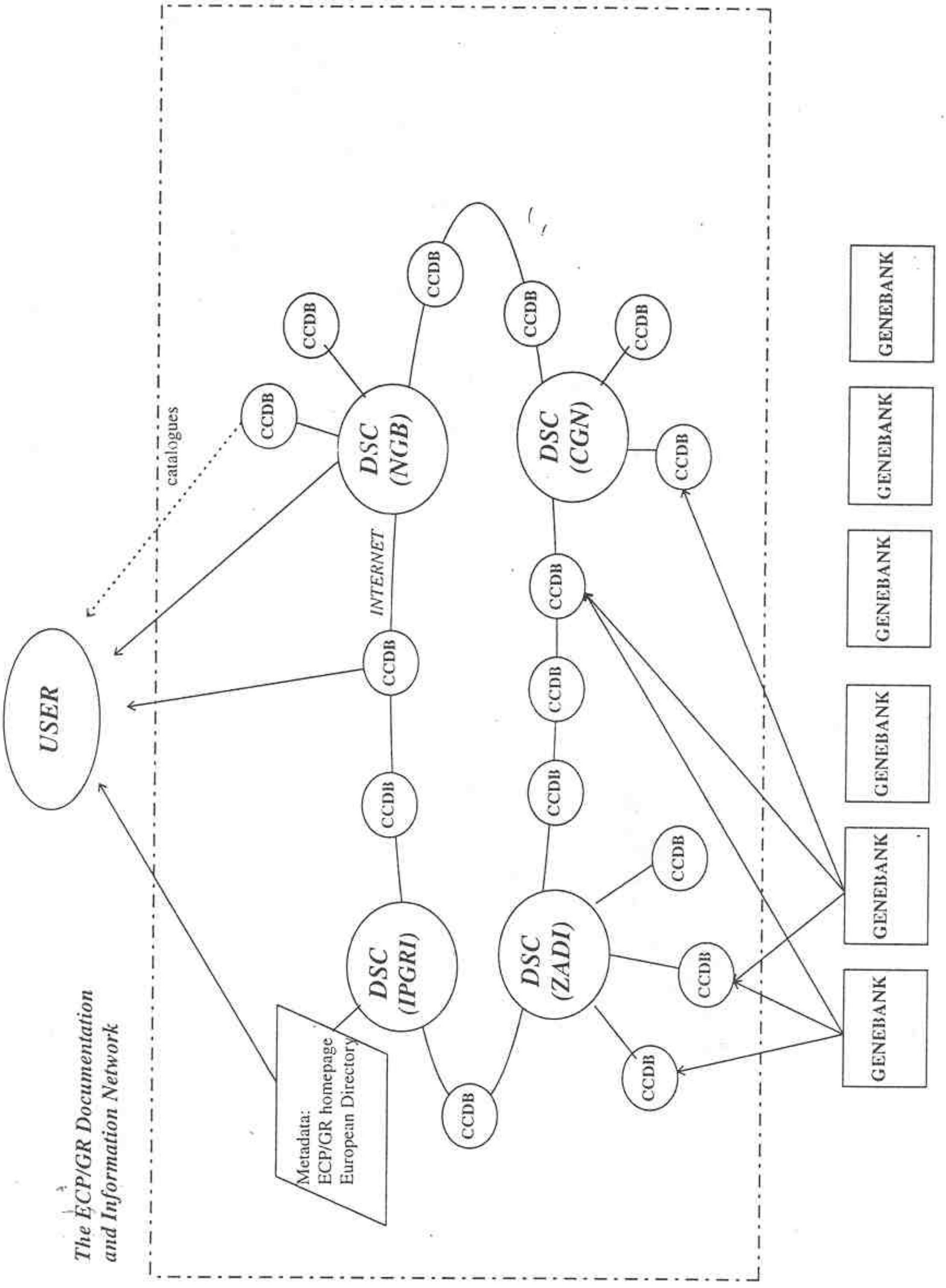


Fig. 2. Role of Internet in improving access to central crop databases (CCDBs); DSC=Documentation Support Centre.

For our present concern, an essential item of the home page is the list of Central Crop Databases, seen as the 'reference catalogue to CCDBs' mentioned above. Links could be established directly from this list of databases to other Web sites where the data would be actually available. To visualize this concretely, prototype links have been established on the new version of the ECP/GR home page, leading to the opening menus of the European *Prunus* and *Secale* databases (on-line demonstrations given at the Workshop).[14]

The Internet-based European information system on plant genetic resources described above could use the ECP/GR home page as the point of entry from which appropriate links would be offered to the relevant information sources:

- on-line central crop databases, either loaded on their own site or hosted on the server of a Documentation Support Center;
- Web sites of other plant genetic resources-related institutions;
- directories, and any other relevant information sources.

Conclusion

The system described above could be implemented very quickly, building on the existing components, and could be developed in a flexible way each time a new element becomes available. The advantages of this formula are several:

- establishing a unique entry point would greatly increase the comfort of the user in search of information on European plant genetic resources;
- through enhanced, user-friendly access to information about collections the utilization of germplasm from genebanks would be increased;
- using the ECP/GR home page would improve the visibility of activities carried out by ECP/GR and by its partners in member countries, and by the programme as a whole in the field of plant genetic resources information, and would contribute to strengthening the cohesion of the plant genetic resources community in Europe.

Another positive side effect would be expected in terms of data quality, since providing access to a central database on Internet and thus making its data potentially accessible to millions of users would be an excellent incentive to the database manager to ensure the best possible quality of the data provided.

---

[14] At the time of publication of this report the *Phleum* database maintained by NGB can already be accessed directly via the ECP/GR home page.

Using Central Crop Databases: searching for duplicates and gaps

***H. Knüpffer*[1]*, L. Frese*[2] *and M.W.M. Jongen*[3]**

[1] Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Genbank, Gatersleben, Germany
[2] Federal Centre for Breeding Research on Cultivated Plants (BAZ), Gene Bank, Braunschweig, Germany
[3] Centre for Genetic Resources The Netherlands (CGN), CPRO-DLO, Wageningen, The Netherlands

Introduction

Genebank collections in the world comprise about 6.1 million accessions (FAO 1996). About half of these accessions have to be considered as threatened. The world holdings for some crops are very large, for example, in wheat (784 000 accessions), barley (485 000), rice (420 000) or *Phaseolus* (268 500). In 1984 it was estimated that at least 50% of the germplasm accessions kept in genebanks worldwide are duplicates (Lyman 1984).

Duplicates between genebanks occur for various reasons. An obvious cause is exchange of accessions between genebanks, or acquisition of the same accession (e.g. a cultivar) by several genebanks. Often material obtained for comparison in national variety testing trials was also deposited in the national genebank. Another reason for duplication, safety-duplication in this case, is the common practice of dividing a collected sample between the institutions participating in the collecting expedition and a genebank in the country of collection.

Duplicates within genebanks occur when by mistake a variety or an accession has been incorporated in the collection more than once. It might even be possible that the next generation of an accession within a genebank receives a new, unique genebank identification number and thus could be considered as a new, but duplicated, accession.

At present the political barriers between countries are fading, and the need for self-sufficient genebanks decreases. Since rejuvenation and multiplication are costly, it becomes also economically interesting to avoid duplication of work when it is rather easy to request these seeds again from another genebank. In view of the problems with respect to safety of germplasm maintained in genebanks, as reported in 'The state of the world’s plant genetic resources for food and agriculture' (FAO 1996), it is necessary to identify duplicates and to concentrate work internationally on unique accessions and their safety duplicates. Therefore, genebanks should make efforts to share responsibilities for certain crops or accessions.

The cooperation between Germany and the Netherlands[15] may serve as an example. These two countries combined their collections of *Beta* and *Solanum* species. At this moment the *Beta* collection is managed in Germany and the *Solanum* collection is managed in the Netherlands. Of course both countries still have their own “working collections” of these crops. They can also be considered as safety-duplication collections.

---

[15] This cooperation started several years before the reunification of Germany. The inclusion of IPK’s potato collection in Gross Lüsewitz and the *Beta* collection in Gatersleben are still under discussion.

To be able to make agreements on sharing responsibilities, it is important to build up crop networks and to create a reliable overview of the collections involved. For many important crops in Europe, such networks exist, and Central Crop Databases (CCDBs) were created. These “are centralized regional or international databases of plant genetic resources accessions of a particular crop or group of crops held in several or many institutions in a region or worldwide” (Knüpffer 1995). Usually the following tasks can be found among the main objectives of such CCDBs:

- to identify duplicates in order to reduce the workload for the maintenance and characterization of the material in cooperation between genebanks;
- to identify unique material in need of safety-duplication;
- to identify gaps in existing collections in order to elaborate strategies for further collecting (cf. Knüpffer 1995).

These tasks are closely connected with each other. Only after completing the identification of duplicates is it possible to state which accessions are unique, i.e. those which do not have duplicates. The search for geographical gaps and under-represented areas in genebank collections will also yield the best possible results once duplicates between genebanks have been identified and removed from the analysis of the geographical origin of the accessions.

Undetected duplication is undesirable and unnecessary. Some of the efforts wasted to maintain and evaluate duplicated material could be better spent in filling gaps in collections.

In this paper we describe approaches to searching for duplicates and gaps on the basis of passport data accumulated in central crop databases.

Types of duplicates

Van Hintum and Knüpffer (1995) proposed a terminology to distinguish different types of duplication (Fig. 1). **Identical duplication** refers to genetically identical genebank accessions. This definition is applicable only for material which has not been rejuvenated, or which is completely homogeneous and homozygous or vegetatively propagated (Hazekamp and van Hintum 1991). **Common duplicates** are derived from the same original population and share all alleles, but possibly with different allele frequencies. They are the most common case of duplication in genebanks. **Partial duplicates** are two accessions selected from the same original population, but having only a part of the alleles or genotypes in common. A special case of partial duplication is **compound duplication**, where one accession is a selection from the other one.

Problems in identifying duplicates

Passport data in CCDBs are the starting point to search for duplicates. However, identical passport data do not necessarily imply that the corresponding accessions are identical (Sahu 1989). In the contrary, there are examples of accessions with identical names (e.g. cultivar names) but different genetic background. Owing to the preliminary nature of the results based on passport data only, the term **probable duplicate** is being used for such findings.[16]

Knüpffer (1988) and Frese and van Hintum (1989) showed some of the complications of searching for duplicates in a CCDB for barley and *Beta*,

---

[16] Other terms proposed by members of the ECP/GR Barley Working Group in the past were **ostensible duplicates** or **potential duplicates.**

respectively. Other CCDBs also reported about approaches to identification of duplicates, e.g. Schittenhelm and Seidewitz (1993) for *Avena*.

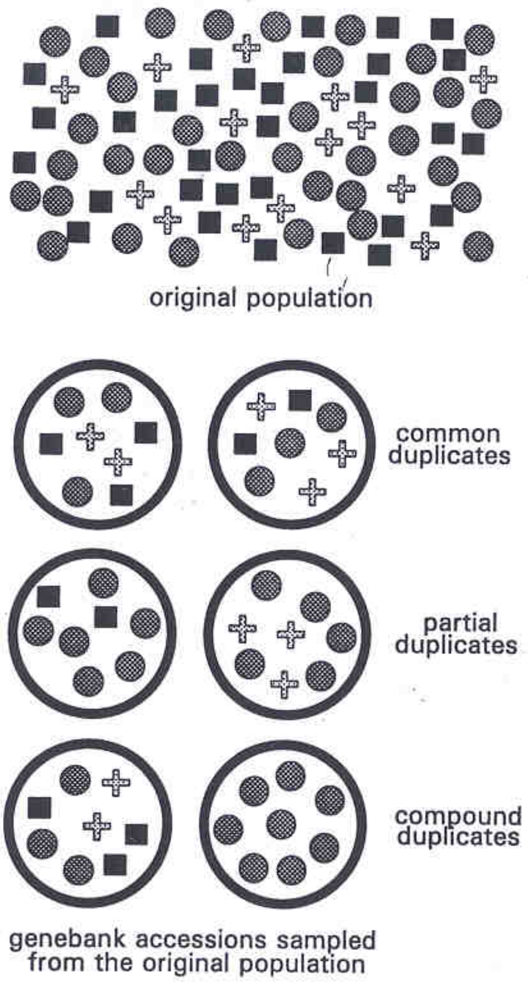


Fig. 1. Different types of duplicates.

Practices of genebank work may lead to changes in the genetic composition of accessions. For example, splitting populations (e.g. collected landraces) into morphologically distinct lines produces genetically different accessions with virtually identical passport data (except for the accession number and sometimes the botanical classification on the infraspecific level). Genetic drift during multiplication of small populations, natural selection due to adaptation to growing conditions in the place of multiplication, or unintentional selection by man during multiplication and seed handling are other reasons for such changes. Finally, cross-pollination, mixing or confusion of seed lots may affect the genetic identity of an accession (van Hintum and Knüpffer 1995). Repeated sampling of wild plant or landrace populations by different collecting missions can also result in duplication.

The quality and completeness of documentation is another factor influencing the identification of duplicates. Common problems include incomplete data (e.g. country of origin lacking), information placed in the wrong data field (e.g. collection site or collection number under accession name), homonyms and synonyms of cultivar names, different transcription or transliteration rules applied for names from languages written in non-Latin alphabets. For a more detailed discussion, see Knüpffer (1989) and van Hintum and Knüpffer (1995). Accessions for names from languages written in non-Latin alphabets. For a more detailed discussion, see Knüpffer (1989) and van Hintum and Knüpffer (1995). Accessions without clear documentation will never be identified as duplicates. Consequently, they have to be considered as unique accessions.

A study of duplication in garden beet germplasm (*Beta vulgaris* subsp. *vulgaris* Garden Beet Group) (Frese and van Hintum 1989) gives an interesting example for the complications connected with probable duplicates. If the term 'probable' is defined widely, more accessions can be identified as probable duplicates than with a narrow definition. A search for the string '%gypt%' in the field containing the vernacular names of the International Database for *Beta* (IDBB) resulted in 50 probable duplicates of the garden beet variety 'Egyptian'. According to Holland's (1956) description, accessions within that duplicate group are not at all identical. Using morphological differences, the 'Egyptian' can be grouped into two major classes and a total of 15 subclasses. With a narrower definition of 'probable', allowing for accessions with the name 'Egyptian' only, the number of resulting probable duplicates is much lower, i.e. 3 accessions.

Probable geographical duplicates form a particular problem. Using the geographical coordinates for mapping of collection sites for wild beets, Hazekamp and Frese (1992) showed, for example, that, in comparison with the size of the Greek island Chios, very much material exists in germplasm collections. Taking into account that wild beet is a ruderal outcrossing species, without doubt, Chios has to be regarded as a very densely sampled area.

Identification of duplicates on the basis of passport data

Searching for duplicates in a CCDBis usually carried out on two levels.

The first level is comparing records of a central database by single fields with a certain probability of uniqueness of field contents. This refers to accession names, collector's numbers (in connection with a unique identification of the collector or collecting expedition), or accession numbers and donor numbers (sometimes referred to as **parallel numbers**, for example, by the Centre for Genetic Resources of the Netherlands). If identical values in such fields are found, it is quite likely that the corresponding accessions are duplicates.

At the next level it is necessary to compare accessions with similar, but not identical, passport data for these descriptors. This means that different ways of presentation of the same information should be identified, such as different spellings of names, slight typing errors, transpositions of parts of names, hyphens vs. blanks in accession numbers, etc. More sophisticated techniques have to be applied to find out similarities between passport data. Such techniques have been developed and used for different central crop databases in the past. Probable duplicates identified this way have to be checked by specialists even more carefully than those resulting from the first level.

Techniques for identification of probable duplicates

Some commonly applied computer-supported techniques have to be mentioned here.

The KWIC index (keyword in context) technique was originally developed for bibliographical databases and can be found in abstract services such as Biological Abstracts. Within the field of plant genetic resources databases it was first applied for the European Barley Database (Knüpffer 1988). This method makes it possible to detect accessions with matching or similar words or numbers, for example, as parts of accession names, even if the names differ. To create the KWIC index, names consisting of several words or numbers (sequences of digits) are split into these components, and each component becomes an entry in the index, which is subsequently sorted alphabetically.

For example, the Swedish barley variety ‘Britta’ can be found in the European Barley Database as ‘Weibulls Britta’, ‘Britta Weibull’ or ‘Britta’. This name would be found in the KWIC index three times under ‘Britta’, and once under each of ‘Weibull’ and ‘Weibulls’. For a more detailed description with many examples, see Knüpffer (1989).

The Soundex function was designed to group words with similar 'sounding' and is available as a function in database management systems such as FoxPro and Oracle. It preserves the first letter and converts the following three (groups of) consonants into three digits. The phonetic similarity reflects the pronunciation of English words. The applicability for plant genetic resources databases with information originating from many different languages is, therefore, limited. However, if applied carefully, it can be used to find probable duplicates which would not be found either in an alphabetically sorted list or by using the KWIC index alone.

The Soundex function would group, for example, the following barley accession names: ‘Bavaria’, ‘Bibior’, ‘Bobro’ and ‘Buhobori’ which obviously do not form a group of duplicate accessions. However, the following group contains some probable duplicates which would neither be found by using an alphabetically sorted list or complete matching, nor by using the KWIC index: ‘Closess IV’ (misspelling), ‘Colcess’ (misspelling), ‘Colcess IV’, ‘Colchicum’, ‘Colses’ (misspelling) and ‘Colsess’.

The SoundKWIC procedure proposed by Maschka (1995) consists of a combination of the KWIC index and the Soundex function. It was applied for searching duplicates between the three German rye collections in Gatersleben, Braunschweig and Gülzow and yielded four more findings of duplicates than the KWIC index alone.

Verification of duplication

Probable duplicates identified in a passport database have to be verified. The first step would consist of checking the passport data of all tentative groups of duplicate accessions by specialists for the particular group of accessions (e.g. barley varieties from a certain country). Often a specialist will find out that a group of probable duplicates consists of two cultivars with the same or very similar names. Unless it is very clear from the data that two accessions have been derived from the same original accession (e.g. in the case of a subsample sent from one genebank to another one), it is necessary to confirm such duplicates. A possible approach is to grow such accessions side by side and to compare them, based on morphological characters and/or evaluation descriptors. If they differ, it is clear that they are not duplicates. However, if they do not seem to differ, other methods of comparison can be applied, for example, molecular characterization techniques. Even molecular techniques will not give an ultimate answer, if no differences are found, because only a small part of the DNA is used for comparison. In any case, the costs of carrying out the necessary analyses have to be compared with the costs of continued maintenance of two accessions which might be duplicates.

Costa *et al.* (1989) carried out multivariate analyses on 22 evaluation descriptors for 400 apomictic accessions of *Panicum maximum* collected in East Africa. Duplication was also studied by Pillai *et al.* (1991) for coconut accessions and by Gironella (1991) for rice, based on characterization and evaluation data.

Molecular techniques for identifying and verifying duplicates

Recently, some more reports have been published about the use of molecular and biochemical techniques to find or verify duplicates. These are often the same methods as those used for diversity studies in plant genetic resources or for the creation or investigation of core collections. A discussion of such methods in practical genebank work can be found in Lux and Hammer (1994).

For example, He *et al.* (1995) used DNA amplification fingerprinting to identify duplicates among 72 sweet potato (*Ipomoea batatas*) accessions.

Margale *et al.* (1995) identified possible duplicate accessions among French cultivars of cauliflower, cabbage and kale, using RAPD markers. Some cases of misclassification could be detected. RAPD analysis was also applied to probable duplicates in a large rice collection (Virk *et al.* 1995). Duplicates found using passport data were in a first step analyzed using two or three RAPD primers, and only indistinguishable genotypes were subjected to a full RAPD analysis. Waycott and Ford (1994) used RAPD markers integrated with phenotypic measures for separating closely related accessions of butterhead lettuce (*Lactuca sativa*).

Biotechnological characterization was used by Golmirzaie and Ghislain (1995) to study duplication in a collection of Andean root and tuber crops.

Van Hintum and Visser (1995) applied isoenzyme methods to verify suspected duplicates of barley accessions found in four larger European barley collections. Isoenzyme studies also were used to find duplicates in the International Cocoa Genebank (Sirju-Charran *et al.* 1991).

SDS-PAGE of seed proteins was applied by Forde and Gardiner (1991) to look for redundant duplicates in some forage legume and grass species collections. Electrophoretic characterization of about 15 000 potato accessions led to a reduction of the collection to about 5000 accessions representing about 3500 different genotypes (Huamán and Stegemann 1989; Huamán 1991).

The review on the use of biochemical gene markers in maize (Orman and Smith 1988) also considers the detection of duplicates.

Handling of duplication and unique accessions

Duplicates within a genebank collection are usually unintentional and undesired, and when detected either they will be put together into one accession, or one of the two accessions will be discarded. There are differences in the handling of duplicates or very similar accessions in the case of self- and cross-pollinating species. Examples are known for *Allium* and *Brassica*, where very similar selections from old landraces were put together to build up a smaller number of genebank accessions (Boukema and de Groot 1992; van Hintum *et al.* 1996). As van Hintum *et al.* (1996) state, this has to be done very thoughtfully. The similarity of the components of bulked accessions was confirmed by isoenzyme analysis.

Duplication between genebanks may be undesired or intentional. When the identification and verification of duplicates is finished, the accessions in a central crop database will be divided into unique accessions (those for which no duplicate was found) and groups of duplicates. Within each duplicate group, the genebanks involved would define a particular accession as the “most original” one, usually the one maintained in a genebank in the country of origin of the accession, and designate one or two others as safety duplicates. If there are more than two or three accessions in such a duplicate group, it will be left up to the holding genebanks whether they keep these accessions (perhaps in their working collections), or whether they discard them. In the case of unique accessions it will be necessary to designate a responsible genebank for safety-duplication, and to send a seed sample of this material to the designated genebank.

Inventory of gaps in collections – ecogeographical approach

Without identification of duplicates and unique accessions, any statistical summaries about the number of accessions per country of origin or the density of collecting for a given geographical area will be biased because of hidden duplication. A reduced data set, containing information on unique accessions only, is a good basis for such analyses. Based on a summary of the countries of origin, compared with the known or expected distribution of wild species or of the cultivation areas of the particular crop, sound conclusions can be drawn about the necessity of further collecting.

A detailed analysis was carried out using geographical coordinates of collecting sites in the International Database for *Beta* (IDBB) (Hazekamp and Frese 1992; see also Frese 1996). Data sets were plotted on maps to visualize the distribution of collecting sites. The distribution area of a crop and its related wild species is usually known from previous explorations or from botanical literature. By comparison of this area with the geographic origin of the material in hand, gaps become visible. Gaps within the distribution area can have two reasons: either the species does not occur in parts of the area because of unsuitable climatic, topographic or edaphic factors, or the area has not been visited by germplasm collectors yet.

Genetic differences between populations can be linked to the geographic distances between them. Large geographic gaps in collections are closed by collectors assuming that the additional material would add new genetic diversity to the available germplasm. When planning a collecting mission, investigation into the possible reasons of geographic gaps is necessary. The wild beet (*Beta vulgaris* subsp. *maritima)* is mostly encountered along the sea coast on a narrow band between high tide and a few metres above sea level. Along the northwestern coast of Portugal, up to 70 km long intercepts of monotonous sandy beaches occur which are an unsuitable habitat for wild beets. Such edaphic factors are shown on soil maps. Soil maps should be consulted to assess the chance of finding material in a particular area. It was known from previous collecting missions that wild '*maritima*' beet populations with a medium to high effective population size very seldom occur on sandy beaches or at the bottom of high cliffs. Therefore, Frese *et al.* (1990) skipped such areas and searched successfully in estuaries in coastal regions where the loamy soil type suits *Beta vulgaris* subsp. *maritima* much better.

The manager of a central database can also work on improvement of old passport data. Frese (1996) gives an example. Using US flight navigation maps, topographic or Michelin maps, he searched for location names recorded by collectors and added the geographic coordinates to the central database. Some locations such as those in Iran before the revolution could only be traced using old maps dating back to the 1930s. This exercise was done to visualize sites visited by collectors of the USDA/ARS in the middle of this century. So far, the Near East and West Asia have been considered as a less important area just because the distribution of Swiss chards (*Beta vulgaris* subsp. *vulgaris*, Leaf Beet) and other cultivated types in that region was not very well known. The maps generated by the IDBB show that landraces do exist in Asian countries. As a conclusion, it could be recommended to collect them since a large region from Eastern Europe to East China has not been systematically sampled in the past.

Reid *et al.* (1989) reported on the identification of major geographical and diversity gaps for major forage crops, using passport data from genebanks and herbaria, among other sources. Ecogeographical approaches for identifying gaps and under-represented areas in germplasm collections are also described and used by Engels *et al.* (1995), Guarino (1995), Maxted *et al.* (1995), Perry and McIntosh (1991) and Wilkes (1992).

Conclusion

Despite the modern techniques available, searching for duplicates and gaps in collections is still complicated work. Although we can find groups of probable duplicates, decisions should be made on how to use this information in a positive way. Intensive cooperation between the genebanks will be indispensable for this process.

Acknowledgements

We are particularly grateful to K. Hammer and Th. van Hintum for their comments on the manuscript.

References

Boukema, I.W. and L. de Groot. 1992. Reduction of redundancy in the CGN *Allium* collection. P. 19 *in* The Genus *Allium* - Taxonomic Problems and Genetic Resources (P. Hanelt, K. Hammer and H. Knüpffer, eds.). IPK, Gatersleben, Germany.

Costa, J.C.G., Y.H. Savidan, L. Jank and L.H.R. Castro. 1989. Morphological studies as a tool for the evaluation of wide tropical forage grass germplasms. Pp. 277-278 *in* Proc. XVI Internat. Grassland Congr. Assoc. Franç. Product. Fourrag., Centre National de Rech. Agron. Versailles, France.

Engels, J.M.M., R.K. Arora and L. Guarino. 1995. An introduction to plant germplasm exploration and collecting: planning, methods and procedures, follow-up. Pp. 31-63 *in* Collecting Plant Genetic Diversity. Technical Guidelines (L. Guarino, V.R. Rao and R. Reid, eds.). CAB International, Wallingford, UK.

FAO. 1996. The state of the world's plant genetic resources for food and agriculture. Background documentation prepared for the International Technical Conference on Plant Genetic Resources, Leipzig, Germany, 17-23 June, 1996. FAO, Rome.

Forde, M.B. and S.E. Gardiner. 1991. Electrophoretic seed protein banding patterns as genetic resource descriptors. Pp. 95-108 *in* Forest Res. Inst. and Malays. Nat. Comm. on Plant Genetic Resources (A.H. Zakri, M.N. Normah, A.G.A. Karim and M.T. Senawi, eds.), Kuala Lumpur, Malaysia.

Frese, L. 1996. Uses of the International Database for *Beta* (IDBB). *In* Genebank Documentation, the Users Perspective. Proc. Eucarpia Genet. Res. Sect. Meet., 17-20 Oct. 1996, Budapest. Submitted.

Frese, L. and Th. J. L. van Hintum. 1989. The International Data Base for *Beta*. Pp. 17-45 *in* Report of an International Workshop on *Beta* Genetic Resources. Internat. Crop Networks Series No. 3. IBPGR, Rome.

Frese, L., E. de Meijer and J. Letschert. 1990. New wild beet genetic resources from Portugal and Spain. Zuckerind. 115:950-955.

Gironella, A. 1991. Documentation and data management for rice germplasm banks. Pp. 57-58 *in* Rice Germplasm. Collecting, Preservation, Use. IRRI, Manila, Philippines.

Golmirzaie, A.M. and M. Ghislain. 1995. Application of biotechnology for conservation and utilization of Andean root and tuber crops. Plant Breed. Abstr. 65:469-470.

Guarino, L. 1995. Mapping the ecogeographic distribution of biodiversity. Pp. 287-314 *in* Collecting Plant Genetic Diversity. Technical Guidelines (L. Guarino, V.R. Rao and R. Reid, eds.). CAB International, Wallingford, UK.

Hazekamp, Th. and L. Frese. 1992. Application of mapping systems for the analysis of geographical origin of collected material. Pp. 69-70 *in* International *Beta* Genetic Resources Network (L. Frese, ed.). Internat. Crop Network Series No. 7. IBPGR, Rome.

Hazekamp, Th. and Th. J. L. van Hintum. 1991. Central crop databases in collaborative genetic resources management. Pp. 37-42 *in* Crop Networks: Searching for New Concepts for Collaborative Genetic Resources Management Management (Th.J.L. van Hintum, L. Frese and P. Perret, eds.). Papers of the EUCARPIA / IBPGR symposium held at Wageningen, The Netherlands, 3-6 December 1990. International Crop Network Series No. 4. IBPGR, Rome, Italy.

He Guo Hao, C.S. Prakash and R.L. Jarret. 1995. Analysis of genetic diversity in a sweet potato (*Ipomoea batatas*) germplasm collection using DNA amplification fingerprinting. Genome 38:938-945.

Holland, H. 1956. Classification and performance of varieties of red beets. Pp. 16-42 *in* Rep. Nat. Veg. Res. Stn. for 1956, UK.

Huamán, Z. 1991. Electrophoretic characterization of potato germplasm at CIP. Pp. 129-134 *in* Report of the planning conference on application of molecular techniques to potato germplasm enhancement. CIP, Lima, Peru.

Huamán, Z. and H. Stegemann. 1989. Use of electrophoretic analyses to verify morphologically identical clones in a potato collection. Plant Varieties and Seeds 2:155-161.

Knüpffer, H. 1988. The European Barley Database of the ECP / GR: An introduction. Kulturpflanze 36:135-162.

Knüpffer, H. 1989. Identification of duplicates in the European Barley Database. Pp.22-43 *in* Report of a Working Group on Barley (Third Meeting). IBPGR, Rome.

Knüpffer, H. 1995. Central crop databases. Pp. 51-62 *in* Standardization in Plant Genetic Resources Documentation (Th.J.L. van Hintum, M.W.M. Jongen and Th. Hazekamp, eds.). CGN, Wageningen, The Netherlands.

Lux, H. and K. Hammer. 1994. Molecular markers and genetic diversity - some experience from the genebank. Pp. 49-53 *in* Evaluation and Exploitation of Genetic Resources: Pre-breeding (F. Balfourier and M.R. Perretant, eds.). INRA, Versailles, France.

Lyman, J.M. 1984. Progress and planning for germplasm conservation of major food crops. FAO / IBPGR Plant Genet. Resour. Newsl. 60:3-21.

Margale, E., Y. Hervé, H. Hu and C.F. Quiros. 1995. Determination of genetic variability by RAPD markers in cauliflower, cabbage and kale local cultivars from France. Genet. Resour. and Crop Evol. 42:281-289.

Maschka, R. 1995. Die Identifizierung von Duplikaten in Genbankkollektionen im Rahmen des Informationssystems Genetischer Ressourcen in Deutschland. Diplomarbeit, Univ. Kassel Gesamthochschule.

Maxted, N., M.W. van Slageren and J.R. Rihan. 1995. Ecogeographic surveys. Pp. 255-285 *in* Collecting Plant Genetic Diversity. Technical Guidelines (L. Guarino, V.R. Rao and R. Reid, eds.). CAB International, Wallingford, UK.

Orman, B.A. and J.S.C. Smith. 1988. Use of biochemical gene markers for measuring maize genetic diversity. Pp. 118-129 *in* Recent Advances in the Conservation and Utilization of Genetic Resources. CIMMYT, Mexico.

Perry, M.C. and M.S. McIntosh. 1991. Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. Crop Sci. 31:1350-1355.

Pillai, R.V., E.V.V.B. Rao and P.M. Kumaran. 1991. Characterization of coconut cultivars. Pp. 75-82 *in* Coconut Breeding and Management (E.G. Silas, M. Aravindakshan and A.I. Jose, eds.). Kerala Agric. Univ., Vellanikkara, India.

Reid, R., J. Konopka, P.M. Perret and L. Guarino. 1989. Forage germplasm resources for the Mediterranean and adjacent arid/semi-arid areas. Pp. 289-290 *in* Proc. XVI Internat. Grassland Congr., Nice, France.

Sahu, R.K. 1989. Screening for duplicates in the germplasm collection. Int. Rice Res. Newsl. 14:4.

Schittenhelm, S. and L. Seidewitz. 1993. Progress on the European *Avena* Database. Pp. 37-49 *in* Report of a working group on *Avena* (fourth meeting) (E.A. Frison, J. Koenig and S. Schittenhelm, eds.). IBPGR, Rome.

Sirju-Charran, G., E. Johnson and J. M. Warren. 1991. Isozymes and the description of cocoa germplasm in Trinidad. Cocoa Growers' Bull. 44:25-28.

van Hintum, Th.J.L., I.W. Boukema and D.L. Visser. 1996. Reduction of duplication in a *Brassica oleracea* germplasm collection. Genet. Resources and Crop Evol. 43:343-349.

van Hintum, Th.J.L. and H. Knüpffer. 1995. Duplication within and between germplasm collections. I. Identifying duplication on the basis of passport data. Genet. Resour. and Crop Evol. 42:127-133.

van Hintum, Th.J.L. and D.L. Visser. 1995. Duplication within and between germplasm collections. II. Duplication in four European Barley collections. Genet. Resources and Crop Evol. 42:135-145.

Virk, P.S., H.J. Newbury, M.T. Jackson and B.V. Ford-Lloyd. 1995. The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor. Appl. Genet. 90:1049-1055.

Waycott, W. and S.B. Ford. 1994. Differentiation of nearly identical germplasm accessions by a combination of molecular and morphological analyses. Genome 37:577-583.

Wilkes, G. 1992. Strategies for sustaining crop germplasm preservation, enhancement, and use. Issues in Agric. No. 5. CGIAR, Washington, USA.

Other contributions

A case study on merging evaluation data from different genebanks: the Inter-genebank Potato Database

***R. Hoekstra*[1], *J.B. Bamberg*[2] *and Z. Huamán*[3]**

[1] Centre for Genetic Resources (CGN, CPRO-DLO), Wageningen, The Netherlands
[2] NRSP-6, USARS, Wisconsin Experimental Station, 4312 Highway 42, Sturgeon Bay, WI 54235, USA
[3] Centro Internacional de la Papa, Lima 100, Peru

Summary

Within the development of a central database containing passport and evaluation data obtained from different genebanks, the comparison of evaluation data from different sources is difficult and in many cases nearly impossible. A pragmatic approach for the establishment of a central evaluation database for potato genetic resources containing data obtained from the most important potato genebanks in the Americas and Europe is discussed. This database has been made available as a component of an Inter-genebank Potato Database (IPD).

Introduction

Many central crop databases only include passport data. Because of the lack of evaluation data these databases have a limited value to plant breeders. The problem for the database managers is that comparing evaluation data from different sources is very difficult, due to differences in the evaluation methods used and the environments where the material was screened. One needs to find the most effective way to mesh the data and be honest about the limitations of the applications which one can prudently make. Therefore pragmatic approaches are needed to progress in this field.

The Inter-genebank Potato Database is one of the few central databases that includes evaluation data. The most important potato genebanks in the world founded an ‘Association of Potato Inter-genebank Collaborators’ (APIC) in 1990 (Bamberg 1990; Bamberg *et al.* 1995). One of the main goals of APIC was the development of an Inter-genebank Potato Database (IPD), a global database of potato germplasm. The first step was to concentrate on the development of the passport database. This was achieved by merging the geographical data of the wild potato accessions maintained by the different potato genebanks (Huamán *et al.* 1996a, 1996b). The key field was the collector’s number, since this is the one internationally invariable germplasm identifier. Starting in 1993, similar efforts were made to develop the IPD evaluation database.

Methods

For the purpose of merging evaluation data from the participant APIC potato genebanks, there was a need to agree on standard descriptors and descriptor states for pests and diseases as well as for other traits. For wild potato species, other characterization data such as plant height or other morphological data can be derived from taxonomic species descriptions and therefore are not included as part of the evaluation database. The descriptor states that were agreed upon to describe the reaction of the plants to stresses (diseases and pests) as well as to quality traits

Table 1. Descriptor states of the evaluation data in the Inter-genebank Potato Database.

Resistance to pests/diseases		Quality traits	
VR	Very Resistant	VH	Very High
R	Resistant	H	High
M	interMediate	M	interMediate
S	Susceptible	L	Low
VS	Very Susceptible	VL	Very Low
T	Tolerant		
H	Hypersensitive	numeric values	
I	Immune		

are shown in Table 1. Letter codes were adopted instead of numbers as most users are familiar with categories such as Resistant, Susceptible, etc. The use of letter codes also prevents confusion on the meaning of the descriptor states that may occur when a numerical scale is used, considering for example the reverse numerical scoring conventions of IPGRI and UPOV.

Since most evaluation data on wild potatoes are the results of screening populations and the data included in the database were obtained from different sources, plant breeders will have to use the information only as a general guide for selecting germplasm. The quality of the data must be validated by the users. They will need to request seeds of those accessions where others have found valuable traits and screen the material under their own conditions to select the best individual plants to be used in their breeding programmes.

In some cases, other useful additional information to the evaluation results is available in the genebanks. Therefore, the IPD evaluation database includes two extra digits next to the two characters representing the reaction (VR, R, M, S, VS):

- a number (1-9) indicating the percentage (1=10% or less and 9=90% or more) of the population with that reaction;
- a symbol (e.g. an asterisk ' * ' meaning consistent reaction, or a question mark ' ? ' for conflicting results among evaluators).

Examples:

S7	means that about 70% of the progeny reacted susceptible.
VR3*	means that about 30% of the sample was very resistant and that this result was consistent in replicated trials.
R4?	means that 40% of the sample was resistant, but the result may be less reliable, owing to, for example, poor root development of the plants (as indicated by the evaluating person).

Another convention adopted was that when very few good sources for resistance are available or the valuable trait is rare, for example resistance in a minor portion of the population, then the reaction will be coded for example as R1 instead of S9. It is important that the genebank curator in cooperation with the experts who carried out the evaluations make this decision and transform the data into the common format.

Results

Crucial for the development of the IPD evaluation table was the standardization of the IPD passport table on the basis of the collector’s number. It allowed the determination of the different identification numbers assigned to the same collector number (i.e. the unique IPD accession number) by each potato genebank.

This made it possible to access the total pooled evaluation data for any given accession.

Data on more than 36 000 evaluations of wild potato accessions maintained in four potato genebanks (Huamán *et al.* 1996b) were included in the IPD evaluation table. Evaluation data on the same accession maintained and evaluated in different genebanks are stored in different records. The table size comprises almost 5000 records with a total of 62 fields for different traits. Only a few evaluations show additional information related to the percentage of the population with a given reaction and reliability of the data. The IPD database clearly shows the breeding potential of the wild potato genetic resources in *ex situ* conservation.

Conclusion

The experience of developing the IPD evaluation database showed that it is essential to use descriptor states that can be accommodated to data generated by different researchers, in different environments, and with different evaluation methods. As much as possible there should be access to the descriptions of the screening method and environment where the material was screened. The descriptor states should be more or less of general use and preferably should include those that are more or less self-explanatory (i.e. resistant, susceptible, etc.).

References

Bamberg, J.B. 1990. Potato intergenebank collaboration. Diversity 6(3&4):6-7.

Bamberg, J.B., Z. Huamán and R. Hoekstra. 1995. International cooperation in potato germplasm. Pp. 177-182 *in* International Germplasm Transfer: Past and Present (R.R. Duncan, D.M. Kral and M.K. Viney, eds.). Crop Science Society of America, Special Publication No. 23. CSSA, Madison, WI, USA.

Huamán, Z., J.B. Bamberg and R. Hoekstra. 1996a. The Inter-genebank Potato Database and the dimensions of wild potato germplasm available in *ex situ* conservation worldwide. In preparation.

Huamán, Z., R. Hoekstra and J.B. Bamberg. 1996b. The Intergenebank Potato Database. P. 315 *in* Abstracts of Conference Papers, Posters and Demonstrations of the 13th Triennial Conference of the EAPR, 14-19/7/1996, Veldhoven, The Netherlands.

The MAFF CORE-STORE for *Malus* Relational Database – Update

R. Janes
Wye College, University of London, National Fruit Collections, Faversham, Kent, UK

The MAFF CORE-STORE database is being developed by Horticulture Research International (HRI), Wellesbourne, and Wye College, University of London under contract to the UK Ministry of Agriculture, Fisheries and Food (MAFF). The major objective of the project is to design and implement a fully relational database and user application to cater for the specific requirements of the National Fruit Collection. Earlier versions of the database have been demonstrated elsewhere (van Hintum *et al.* 1995; Case 1996). Implementation has remained the same for the database, utilizing Microsoft Access version 2 for Windows, and it runs on IBM-PC 486 and compatible processors. However, many refinements have been carried out to improve performance by eliminating 'bugs' and upgrading the 'user-friendly' aspects of the database. The project is now in its final stages, with the version presented here closely representing the final product.

A demonstration of the database was given by asking some of the queries which are expected to be regularly requested by end users. For example, if the question posed is, "Are there any synonyms/homonyms of this cultivar?", through comparison of traits measured for this cultivar, a list of possible synonyms/homonyms is suggested. Or alternatively, when the question, "What is the identity of the tree located at position X in the collection?" is formulated, the Passport Data of this accession will be displayed.

User-friendly aspects

To arrive at conclusions such as those given above, the user navigates through the database by utilizing switchboards, forms which contain buttons. In addition, the database includes Hypertext, which enables 'short-cuts' between forms to be made by the user, improving ease of usage. Password-enforced browse or edit security levels can be introduced using standard Microsoft Access procedures.

Importing and exporting data

There is a wide range of options within Microsoft Access version 2 for Windows for importing and exporting data to and from text files, spreadsheets and other databases. In MAFF CORE-STORE, new data can be entered directly via forms while HRI is developing a Bulk Loader to facilitate entry of existing electronically held data.

Standardization and compatibility

Data held in the database have been fully normalized. Utilization of standard formats, for example the IPGRI list of Fruit Descriptors and the ISO Country Codes, has been made wherever possible. As new standard formats are agreed upon, these will be adopted into MAFF CORE-STORE.

References

Case, H.J., ed. 1996. European *Malus* Germplasm. Proceedings of a workshop, 21-24 June 1995, Wye College, University of London. IPGRI, Rome, Italy.

van Hintum, Th.J.L., M.W.M. Jongen and Th. Hazekamp, eds.). 1995. Standardization in Plant Genetic Resources Documentation. Report of the Second Technical Meeting of Focal Points for Documentation in East European Genebanks. Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.

Appendix I. Agenda

Central Crop Databases
Tools In Plant Genetic Resources Management
A Joint EGDS - ECP/GR workshop held in Budapest, Hungary
12 - 16 October 1996

Saturday 12 October ***Arrival and EGDS day***
Arrival of participants
9:30 Third technical meeting of focal points for documentation in East European genebanks (session for EGDS focal points only)

Sunday 13 October ***Opening and theory day***
09:00 Opening and welcome (László Holly, Thomas Gass, Theo van Hintum)
09:30 Summary of the extra session for EGDS focal points (Marcel Jongen)
09:45 Summary of the Technical meeting on the methodology of the FAO World Information and Early Warning System on PGR, 30 September - 3 October 1996 in Radzików, Poland (Jerzy Serwinski)

Introduction
10:15 Central crop databases, an introduction (Theo van Hintum)
10:45 Coffee

Organisational aspects
11:15 ECP/GR and central crop databases (Thomas Gass)

Methodology of database compilation
11:45 Standardisation of central crop databases (Morten Hulden)
12:30 Multicrop standard passport data descriptor lists (Tom Hazekamp, Jerzy Serwinski)
13:00 Lunch

Methodology of using the database
14:00 Handling of evaluation data (Lajos Horváth, Gusztav Szabó)
14:30 Inventory of users of central crop databases (Wieslaw Podyma)
15:00 Tea
15:30 Access to central crop databases (Daniel Jiménez Krause, Elinor Lipman)
16:15 Searching for duplicates and inventory of omissions (Marcel Jongen, Helmut Knüpffer, Lothar Frese)
17:00 Closing first day

Monday 14 October ***Demonstrations and discussions I***
Demonstrations I
09:00 Plenary introduction of databases to be demonstrated, group formation (Marcel Jongen)
10:00 First session of demonstrations and discussion of Central Crop Databases and related information sources (in six parallel groups)

11:00	Coffee
11:30	Second session of demonstrations and discussion of Central Crop Databases and related information sources (in six parallel groups)
12:30	Lunch

Working group discussions I

13:30	Introduction and group formation (Thomas Gass)
14:30	First working group session (in four parallel groups)
15:30	Tea
16:00	Continuation first working group session
17:00	Closing second day

Tuesday 15 October *Demonstrations and discussions II*

Working group presentation I

09:00	Plenary presentations of conclusions of first working group session

Demonstrations II

10:00	Third session of demonstrations and discussion of Central Crop Databases and related information sources (in six parallel groups)
11:00	Coffee
11:30	Fourth session of demonstrations and discussion of Central Crop Databases and related information sources (in six parallel groups)
12:30	Lunch
13:30	Fifth session of demonstrations and discussion of Central Crop Databases and related information sources (in six parallel groups)

Working group discussions II

14:30	Second working group session (in four parallel groups)
15:30	Tea
16:00	Continuation second working group session
17:00	Closing third day

Wednesday 16 October *Closing and excursion*

Working group presentation II

09:00	Plenary presentations of conclusions of second working group session
10:00	Introduction and discussion of future plans and possibilities (chair: Frank Begemann)
11:00	Coffee
11:30	Discussion (continued) and adoption of recommendations
12:30	Closing
13:00	Lunch
14:00	Excursion to the genebank at Tápiószele
19:00	Workshop dinner.

Appendix II. Multicrop passport descriptors (final version)

The list of multicrop passport descriptors is developed jointly by IPGRI and FAO to provide consistent coding schemes for common passport descriptors across crops. These descriptors aim to be compatible with future IPGRI crop descriptor lists and with the descriptors to be used for the FAO World Information and Early Warning System (WIEWS) on PGR.

The list should **NOT** be regarded as a minimum descriptor list, since many additional passport descriptors are essential for the description of crops and need to be recorded. This document lists an initial set of common passport descriptors at the multicrop level. At a later stage the list could be expanded with additional multicrop descriptors; for example, descriptors dealing with the use of germplasm are currently not included, but their suitability for inclusion at the multicrop level will be investigated. Future expansion could even result in the development of more specialized lists of common descriptors at the crop group level.

Attached you will find the final version of the list which contains two sections of which the latter one (FAO WIEWS DESCRIPTORS) lists a number of optional descriptors used in the FAO WIEWS. The list provides descriptions of content and coding schemes, but also provides **suggested** fieldnames (in parentheses) that can assist in the computerized exchange of this type of data.

21 February, 1997

IIa. MULTICROP PASSPORT DESCRIPTORS

1. Institute code (INSTCODE)

Code of the institute where the accession is maintained. The codes consist of the 3-letter ISO 3166 country code of the country where the institute is located plus number or an acronym as specified in the Institute database that will be made available by FAO. Preliminary codes (i.e. codes not yet incorporated in the FAO Institute database) start with an asterisk followed by a 3-letter ISO 3166 country code and an acronym.

2. Accession number (ACCENUMB)

This number serves as a unique identifier for accessions and is assigned when an accession is entered into the collection. Once assigned this number should never be reassigned to another accession in the collection. Even if an accession is lost, its assigned number should never be reused. Letters should be used before the number to identify the genebank or national system (e.g. IDG indicates an accession that comes from the genebank at Bari, Italy; CGN indicates an accession from the genebank at Wageningen, The Netherlands; PI indicates an accession within the USA system)

3. Collecting number (COLLNUMB)

Original number assigned by the collector(s) of the sample, normally composed of the name or initials of the collector(s) followed by a number. This item is essential for identifying duplicates held in different collections. It should be unique and always accompany subsamples wherever they are sent.

4. Genus (GENUS)

Genus name for taxon. Initial Uppercase letter required.

5. Species (SPECIES)

Specific epithet portion of the scientific name in lowercase letters plus authority (only provided at the most detailed taxonomic level). Following abbreviation is allowed: “sp.”

6. Subtaxa (SUBTAXA)

Subtaxa can be used to store any additional taxonomic identifier plus authority (only provided at the most detailed taxonomic level). Following abbreviations are allowed: “ssp.” (for subspecies); “var.” (for variety); “convar.” (for convariety); “f.” (for form).

7. Accession name (ACCNAME)

Either a registered or other formal designation given to the accession. First letter uppercase. Multiple names separated with semicolon.

8. Country of origin (ORIGCTY)

Name of the country in which the sample was originally collected or derived. Use the ISO 3166 extended codes, (i.e. current and old 3 letter ISO 3166 country codes)

9. Location of collecting site (COLLSITE)

Location information below the country level that describes where the accession was collected starting with the most detailed information. Might include the distance in kilometers and direction from the nearest town, village or map grid reference point, (e.g. CURITIBA 7S, PARANA means 7 km south of Curitiba in the state of Parana)

10. Latitude of collecting site (LATITUDE)

Degrees and minutes followed by N (North) or S (South) (e.g. 1030S). Missing data (minutes) should be indicated with hyphen (e.g. 10-S).

11. Longitude of collecting site (LONGITUDE)

Degrees and minutes followed by E (East) or W (West) (e.g. 07625W). Missing data (minutes) should be indicated with hyphen (e.g. 076-W).

12. Elevation of collecting site [m asl] (ELEVATION)

Elevation of collecting site expressed in meters above sea level. Negative values allowed.

13. Collecting date of original sample [YYYYMMDD] (COLLDATE)

Collecting date of the original sample where YYYY is the year, MM is the month and DD is the day.

14. Status of sample (SAMPSTAT)

1 Wild
2 Weedy
3 Traditional cultivar/Landrace
4 Breeder’s line
5 Advanced cultivar

0 Unknown

99 Other (Elaborate in REMARKS field)

15. Collecting source (COLLSRC)

The coding scheme proposed can be used at 2 different levels of detail: Either by using the global codes such as 1, 2, 3, 4 or by using the more detailed coding such as 1.1, 1.2, 1.3, etc.

1 Wild habitat
1.1 Forest/woodland
1.2 Shrubland
1.3 Grassland
1.4 Desert/tundra

2 Farm
2.1 Field
2.2 Orchard
2.3 Garden
2.4 Fallow
2.5 Pasture
2.6 Store

3 Market
3.1 Town
3.2 Village
3.3 Urban
3.4 Other exchange system

4 Institute/Research organization

0 Unknown

99 Other (Elaborate in REMARKS field)

16. Donor institute code (DONORCODE)

Code for the donor institute. The codes consist of the 3-letter ISO 3166 country code of the country where the institute is located plus number or an acronym as specified in the Institute database that will be made available by FAO. Preliminary codes (i.e. codes not yet incorporated in the FAO Institute database) start with an asterisk followed by a 3-letter ISO 3166 country code and an acronym.

17. Donor number (DONORNUMB)

Number assigned to an accession by the donor. Letters should be used before the number to identify the genebank or national system (e.g. IDG indicates an accession that comes from the genebank at Bari, Italy; CGN indicates an accession from the genebank at Wageningen, The Netherlands; PI indicates an accession within the USA system)

18. Other number(s) associated with the accession (OTHERNUMB)

Any other identification number known to exist in other collections for this accession. Letters should be used before the number to identify the genebank or national system (e.g. IDG indicates an accession that comes from the genebank at Bari, Italy; CGN indicates an accession from the genebank at Wageningen, The Netherlands; PI indicates an accession within the USA system). Multiple numbers can be added and should be separated with a semicolon

19. Remarks (REMARKS)

The remarks field is used to add notes or to elaborate on descriptors with value “99”(=Other). Prefix remarks with the field name they refer to and a colon (e.g. COLLSRC:roadside). Separate remarks referring to different fields are separated by semicolons.

IIb. FAO WIEWS DESCRIPTORS

1. Location of safety duplicates (DUPLSITE)

Code of the institute where a safety duplicate of the accession is maintained. The codes consist of 3-letter ISO 3166 country code of the country where the institute is located plus number or an acronym as specified in the Institute database that will be made available by FAO. Preliminary codes (i.e. codes not yet incorporated in the FAO Institute database) start with an asterisk followed by a 3-letter ISO 3166 country code and an acronym. Multiple numbers can be added and should be separated with a semicolon.

2. Availability add. Passport data (PASSAVAIL)
(i.e. in addition to what has been provided)

0 Not available
1 Available

3. Availability of characterization data (CHARAVAIL)

0 Not available
1 Available

4. Availability of evaluation data (EVALAVAIL)

0 Not available
1 Available

5. Acquisition type of the accession (ACQTYPE)

1 Collected / bred originally by the institute
2 Collected / bred originally by joint mission / institution
3 Received as a secondary repository

6. Type of storage (STORTYPE)

Maintenance type of germplasm. If germplasm is maintained under different types of storage, multiple choices are allowed, separated by a semicolon (e.g. 2;3). (Refer to FAO/IPGRI Genebank Standards 1994 for details on storage type)

1 Short-term
2 Medium-term
3 Long-term
4 *In vitro* collection
5 Field genebank collection
6 Cryopreserved
99 Other (elaborate in REMARKS field)

We would welcome your feedback on the use of this list.
Please forward your feedback to:

Tom Hazekamp, Germplasm Documentation Officer
International Plant Genetic Resources Institute
Via delle Sette Chiese 142
00145 Rome, Italy
email: t.hazekamp@cgnet.com. Fax: +39-6-5750309.

Appendix III. Standard country codes (extended list)

This list of codes (© ISO) is to be used for genetic resources documentation purposes only and, subject to the agreement between ISO/DIN and IPGRI, redistribution of the list of codes by others is not allowed.

* Asterisk indicates obsolete country name/code.

Extended list of ISO country codes (ordered by country code)

Code	Country	Code	Country
ABW	Aruba	CHN	China
AFG	Afghanistan	CIV	Cote d'Ivoire
AGO	Angola	CMR	Cameroon
AIA	Anguilla	COG	Congo
ALB	Albania	COK	Cook Islands
AND	Andorra	COL	Colombia
ANT	Netherlands Antilles	COM	Comoros
ARE	United Arab Emirates	CPV	Cape Verde
ARG	Argentina	CRI	Costa Rica
ARM	Armenia	CSK	*Czechoslovakia
ASM	American Samoa	CUB	Cuba
ATA	Antarctica	CXR	Christmas Island
ATF	French Southern Territories	CYM	Cayman Islands
ATG	Antigua and Barbuda	CYP	Cyprus
AUS	Australia	CZE	Czech Republic
AUT	Austria	DDR	*German Democratic Republic
AZE	Azerbaijan	DEU	Germany
BDI	Burundi	DJI	Djibouti
BEL	Belgium	DMA	Dominica
BEN	Benin	DNK	Denmark
BFA	Burkina Faso	DOM	Dominican Republic
BGD	Bangladesh	DZA	Algeria
BGR	Bulgaria	ECU	Ecuador
BHR	Bahrain	EGY	Egypt
BHS	Bahamas	ERI	Eritrea
BIH	Bosnia and Herzegowina	ESH	Western Sahara
BLR	Belarus	ESP	Spain
BLZ	Belize	EST	Estonia
BMU	Bermuda	ETH	Ethiopia
BOL	Bolivia	FIN	Finland
BRA	Brazil	FJI	Fiji
BRB	Barbados	FLK	Falkland Islands (Malvinas)
BRN	Brunei	FRA	France
BTN	Bhutan	FRO	Faroe Islands
BUR	*Burma	FSM	Micronesia, Federal States of
BVT	Bouvet Island	FXX	France, Metropolitan
BWA	Botswana	GAB	Gabon
BYS	*Byelorusian SSR	GBR	United Kingdom
CAF	Central African Republic	GEO	Georgia
CAN	Canada	GHA	Ghana
CCK	Cocos (Keeling) Islands	GIB	Gibraltar
CHE	Switzerland	GIN	Guinea
CHL	Chile	GLP	Guadeloupe

GMB	Gambia
GNB	Guinea-Bissau
GNQ	Equatorial Guinea
GRC	Greece
GRD	Grenada
GRL	Greenland
GTM	Guatemala
GUF	French Guiana
GUM	Guam
GUY	Guyana
HKG	Hong Kong
HMD	Heard Island and McDonald Islands
HND	Honduras
HRV	Croatia
HTI	Haiti
HUN	Hungary
IDN	Indonesia
IND	India
IOT	British Indian Ocean Territory
IRL	Ireland
IRN	Iran
IRQ	Iraq
ISL	Iceland
ISR	Israel
ITA	Italy
JAM	Jamaica
JOR	Jordan
JPN	Japan
KAZ	Kazakhstan
KEN	Kenya
KGZ	Kyrgyzstan
KHM	Cambodia
KIR	Kiribati
KNA	Saint Kitts and Nevis
KOR	Korea, Republic of
KWT	Kuwait
LAO	Lao People's Democratic Rep.
LBN	Lebanon
LBR	Liberia
LBY	Libyan Arab Jamahiriya
LCA	Saint Lucia
LIE	Liechtenstein
LKA	Sri Lanka
LSO	Lesotho
LTU	Lithuania
LUX	Luxembourg
LVA	Latvia
MAC	Macau
MAR	Morocco
MCO	Monaco
MDA	Moldova, Republic of
MDG	Madagascar
MDV	Maldives
MEX	Mexico
MHL	Marshall Islands
MKD	Macedonia, Former Yugoslav Republic of
MLI	Mali
MLT	Malta
MMR	Myanmar
MNG	Mongolia
MNP	Northern Mariana Islands
MOZ	Mozambique
MRT	Mauritania
MSR	Montserrat
MTQ	Martinique
MUS	Mauritius
MWI	Malawi
MYS	Malaysia
MYT	Mayotte
NAM	Namibia
NCL	New Caledonia
NER	Niger
NFK	Norfolk Island
NGA	Nigeria
NIC	Nicaragua
NIU	Niue
NLD	Netherlands
NOR	Norway
NPL	Nepal
NRU	Nauru
NTZ	*Neutral Zone
NZL	New Zealand
OMN	Oman
PAK	Pakistan
PAN	Panama
PCN	Pitcairn
PER	Peru
PHL	Philippines
PLW	Palau
PNG	Papua New Guinea
POL	Poland
PRI	Puerto Rico
PRK	Democratic People's Rep. of Korea
PRT	Portugal
PRY	Paraguay
QAT	Qatar
REU	Reunion
ROM	Romania
RUS	Russian Federation
RWA	Rwanda
SAU	Saudi Arabia
SDN	Sudan
SEN	Senegal
SGP	Singapore
SGS	South Georgia & South Sandwich Is.
SHN	Saint Helena
SJM	Svalbard and Jan Mayen
SLB	Solomon Islands
SLE	Sierra Leone
SLV	El Salvador

SMR	San Marino	TWN	Taiwan, Province of China
SOM	Somalia	TZA	Tanzania
SPM	Saint Pierre and Miquelon	UGA	Uganda
STP	Sao Tome and Principe	UKR	Ukraine
SUN	*Union of Soviet Socialist	UMI	United States Minor Outlying
SUR	Suriname	URY	Uruguay
SVK	Slovakia	USA	United States of America
SVN	Slovenia	UZB	Uzbekistan
SWE	Sweden	VAT	Vatican City State
SWZ	Swaziland	VCT	Saint Vincent and the Grenadines
SYC	Seychelles	VEN	Venezuela
SYR	Syrian Arab Republic	VGB	Virgin Islands (British)
TCA	Turks and Caicos Islands	VIR	Virgin Islands (US)
TCD	Chad	VNM	Viet Nam
TGO	Togo	VUT	Vanuatu
THA	Thailand	WLF	Wallis and Futuna Islands
TJK	Tajikistan	WSM	Samoa
TKL	Tokelau	YEM	Yemen, Republic of
TKM	Turkmenistan	YMD	*Yemen, Democratic
TMP	East Timor	YUG	Yugoslavia
TON	Tonga	ZAF	South Africa
TTO	Trinidad and Tobago	ZAR	Zaire
TUN	Tunisia	ZMB	Zambia
TUR	Turkey	ZWE	Zimbabwe
TUV	Tuvalu		

Extended list of ISO country codes (ordered by country name)

Country name	Code	Country name	Code
Afghanistan	AFG	Bosnia and Herzegowina	BIH
Albania	ALB	Botswana	BWA
Algeria	DZA	Bouvet Island	BVT
American Samoa	ASM	Brazil	BRA
Andorra	AND	British Indian Ocean Territory	IOT
Angola	AGO	Brunei	BRN
Anguilla	AIA	Bulgaria	BGR
Antarctica	ATA	Burkina Faso	BFA
Antigua and Barbuda	ATG	*Burma	BUR
Argentina	ARG	Burundi	BDI
Armenia	ARM	*Byelorusian SSR	BYS
Aruba	ABW	Cambodia	KHM
Australia	AUS	Cameroon	CMR
Austria	AUT	Canada	CAN
Azerbaijan	AZE	Cape Verde	CPV
Bahamas	BHS	Cayman Islands	CYM
Bahrain	BHR	Central African Republic	CAF
Bangladesh	BGD	Chad	TCD
Barbados	BRB	Chile	CHL
Belarus	BLR	China	CHN
Belgium	BEL	Christmas Island	CXR
Belize	BLZ	Cocos (Keeling) Islands	CCK
Benin	BEN	Colombia	COL
Bermuda	BMU	Comoros	COM
Bhutan	BTN	Congo	COG
Bolivia	BOL	Cook Islands	COK

Costa Rica	CRI
Cote d'Ivoire	CIV
Croatia	HRV
Cuba	CUB
Cyprus	CYP
Czech Republic	CZE
*Czechoslovakia	CSK
Democratic People's Republic of Korea	PRK
Denmark	DNK
Djibouti	DJI
Dominica	DMA
Dominican Republic	DOM
East Timor	TMP
Ecuador	ECU
Egypt	EGY
El Salvador	SLV
Equatorial Guinea	GNQ
Eritrea	ERI
Estonia	EST
Ethiopia	ETH
Falkland Islands (Malvinas)	FLK
Faroe Islands	FRO
Fiji	FJI
Finland	FIN
France	FRA
France, Metropolitan	FXX
French Guiana	GUF
French Polynesia	PYF
French Southern Territories	ATF
Gabon	GAB
Gambia	GMB
Georgia	GEO
*German Democratic Rep.	DDR
Germany	DEU
Ghana	GHA
Gibraltar	GIB
Greece	GRC
Greenland	GRL
Grenada	GRD
Guadeloupe	GLP
Guam	GUM
Guatemala	GTM
Guinea	GIN
Guinea-Bissau	GNB
Guyana	GUY
Haiti	HTI
Heard Island and McDonald Islands	HMD
Honduras	HND
Hong Kong	HKG
Hungary	HUN
Iceland	ISL
India	IND
Indonesia	IDN
Iran	IRN
Iraq	IRQ
Ireland	IRL
Israel	ISR
Italy	ITA
Jamaica	JAM
Japan	JPN
Jordan	JOR
Kazakhstan	KAZ
Kenya	KEN
Kiribati	KIR
Korea, Republic of	KOR
Kuwait	KWT
Kyrgyzstan	KGZ
Lao People's Democratic Republic	LAO
Latvia	LVA
Lebanon	LBN
Lesotho	LSO
Liberia	LBR
Libyan Arab Jamahiriya	LBY
Liechtenstein	LIE
Lithuania	LTU
Luxembourg	LUX
Macau	MAC
Macedonia, Former Yugoslav Republic of	MKD
Madagascar	MDG
Malawi	MWI
Malaysia	MYS
Maldives	MDV
Mali	MLI
Malta	MLT
Marshall Islands	MHL
Martinique	MTQ
Mauritania	MRT
Mauritius	MUS
Mayotte	MYT
Mexico	MEX
Micronesia, Federal States of	FSM
Moldova, Republic of	MDA
Monaco	MCO
Mongolia	MNG
Montserrat	MSR
Morocco	MAR
Mozambique	MOZ
Myanmar	MMR
Namibia	NAM
Nauru	NRU
Nepal	NPL
Netherlands	NLD
Netherlands Antilles	ANT
*Neutral Zone	NTZ
New Caledonia	NCL
New Zealand	NZL
Nicaragua	NIC
Niger	NER
Nigeria	NGA
Niue	NIU

Norfolk Island	NFK
Northern Mariana Islands	MNP
Norway	NOR
Oman	OMN
Pakistan	PAK
Palau	PLW
Panama	PAN
Papua New Guinea	PNG
Paraguay	PRY
Peru	PER
Philippines	PHL
Pitcairn	PCN
Poland	POL
Portugal	PRT
Puerto Rico	PRI
Qatar	QAT
Reunion	REU
Romania	ROM
Russian Federation	RUS
Rwanda	RWA
Saint Helena	SHN
Saint Kitts and Nevis	KNA
Saint Lucia	LCA
Saint Pierre and Miquelon	SPM
Saint Vincent and the Grenadines	VCT
Samoa	WSM
San Marino	SMR
Sao Tome and Principe	STP
Saudi Arabia	SAU
Senegal	SEN
Seychelles	SYC
Sierra Leone	SLE
Singapore	SGP
Slovakia	SVK
Slovenia	SVN
Solomon Islands	SLB
Somalia	SOM
South Africa	ZAF
South Georgia and the South Sandwich Is.	SGS
Spain	ESP
Sri Lanka	LKA
Sudan	SDN
Suriname	SUR
Svalbard and Jan Mayen	SJM
Swaziland	SWZ
Sweden	SWE
Switzerland	CHE
Syrian Arab Republic	SYR
Taiwan, Province of China	TWN
Tajikistan	TJK
Tanzania	TZA
Thailand	THA
Togo	TGO
Tokelau	TKL
Tonga	TON
Trinidad and Tobago	TTO
Tunisia	TUN
Turkey	TUR
Turkmenistan	TKM
Turks and Caicos Islands	TCA
Tuvalu	TUV
Uganda	UGA
Ukraine	UKR
*Union of Soviet Socialist Republics	SUN
United Arab Emirates	ARE
United Kingdom	GBR
United States Minor Outlying Islands	UMI
United States of America	USA
Uruguay	URY
Uzbekistan	UZB
Vanuatu	VUT
Vatican City State	VAT
Venezuela	VEN
Viet Nam	VNM
Virgin Islands (British)	VGB
Virgin Islands (US)	VIR
Wallis and Futuna Islands	WLF
Western Sahara	ESH
*Yemen, Democratic	YMD
Yemen, Republic of	YEM
Yugoslavia	YUG
Zaire	ZAR
Zambia	ZMB
Zimbabwe	ZWE

Appendix IV. Internet glossary

This glossary is based mainly on 'A Beginner's Web Glossary' (URL: http://www.cwru.edu/help/webglossary.html) by the Case Western Reserve University, Cleveland, Ohio, USA. It was reedited, enhanced and compacted by Morten Hulden, Nordic Gene Bank, Alnarp, Sweden and Daniel Jiménez Krause, ZADI, Bonn, Germany.

Anchor
Synonymous with **hyperlinks**, anchor refers to non-linear links among documents. More simply, it is the word or phrase that can be selected to connect to another page or resource.

Anchor Colour
You guessed it – the colour on screen that represents the anchors. The reason so many are blue is that it is often the default colour. This colour can be changed to any combination of red, green and blue.

Archie
Derived from the word 'archive', Archie is a Net-based service that allows you to locate files that can be downloaded via **FTP.**

ASCII
(pronounced 'Ask-ee') An acronym for American Standard Code for Information Exchange, ASCII is an international standard in which numbers, letters, punctuation marks, symbols and control codes are assigned numbers from 0 to 127. Easily transferred over networks, ASCII is a plain, unadorned text without style or font specifications.

Bandwidth
The range of transmission frequencies a network can use. The greater the bandwidth the more information that can be transferred over that network at one time. The term bandwidth also broadly includes throughput, meaning the amount of data sent.

Baud
A unit of speed in data transmission, or the maximum speed at which data can be sent down a channel. Baud is often equivalent to bits per second. Named after J.M. E. Baudot (died 1903).

Bit
A contraction of binary digit, a bit is the smallest unit of information that a computer can hold (in the form of "zeros and ones"). Eight bits is equivalent to a byte. The speed at which bits are transmitted or bit rate is usually expressed as bits per second or bps.

Bitmap
A data file or structure which corresponds bit for bit with an image displayed on a screen.

BMP
Microsoft Windows **Bitmap** format.

Browser
A software that allows you to navigate (i.e. to read information) in the **World Wide Web**; examples are Netscape Navigator, NCSA Mosaic and Microsoft Explorer.

Byte
The number of bits used to represent a character (8 bits).

CD-ROM
Compact Disk-Read Only Memory; an optical disk from which information may be read but not written. A CD-ROM has a large storage capability (about 650 MB). In order to read a CD-ROM you need a CD-ROM drive.

Client
A computer that has access to services over a computer network. The computer providing the services is a **server.**

Client-Server Architecture
An information-passing scheme that works as follows: a **client** programme, such as Netscape, sends a request to a **server**. The server takes the request, disconnects from the client and processes the request. When the request is processed, the server reconnects to the client programme and the information is transferred to the client.

Configuration
This is a general-purpose computer term that can refer to the way you have your computer set up. It is also used to describe the total combination of hardware components that make up a computer system and the software settings that allow various hardware components of a computer system to communicate with one another.

Cyberspace
A term coined by William Gibson in his novel "Neuromancer" to refer to a near-future computer network where users mentally travel through matrices of data. The term is now used to describe the Internet and the other computer networks.

Dial-up Connection
The most popular form of Net connection for the home user, this is a connection from your computer to a host computer over standard telephone lines (using a modem). Dial-up connections are nowadays inexpensive.

Direct Connection

A permanent connection between your computer system and the Internet. This is sometimes referred to as a leased-line connection because the line is leased from the telephone company. If you want to be a **server** you'll need a direct connection in order to make your information accessible 24 hours a day. Direct connections are (very) expensive.

Document
When used in reference to the **World Wide Web**, a document is any file containing text, media or **hyperlinks** that can be transferred from an **HTTP server** to a **client** programme.

Document Window
This is the **Web browser**'s scrollable window in which **HTML documents** can be viewed.

Download
To transfer to your computer a copy of a file that resides on another computer.

Email
Electronic mail; one of the oldest Internet services - and still the most popular. An email server provides its users with the possibility to send and receive text messages (and small files of any type, which are then called "attachments") to other users on Internet.

FAQ
This is the acronym for Frequently Asked Questions. A common feature on the Internet, FAQs are files of answers to commonly asked questions. Read FAQs before asking obvious questions.

Firewall
This term refers to security measures designed to protect a networked system from unauthorized or unwelcome access.

FTP
File Transfer Protocol is an Internet **protocol** that allows the transfer of files from one computer to another. FTP is also the verb used to describe the act of transferring files from one computer to another. Most **browsers** today include an FTP **client** for downloading files from FTP **servers**.

GIF
This acronym stands for Graphic Interchange Format, a commonly used file compression format developed by CompuServe for transferring graphics files to and from on-line services.

Gopher
Internet's menu-oriented, text-only predecessor to **WWW**. Most **browsers** today include a gopher **client** for connections to gopher **servers**.

Home Page
A Home Page is the first document you come to at a **Web site.**

Hotlists
Lists of frequently used Web locations and **URL**s (Uniform Resource Locators).

Host
A computer acting as an information or communications **server.**

HTML
An acronym for HyperText Markup Language, HTML is the language used to tag various parts of a **Web document** so browsing software will know how to display that document's links, text, graphics and attached media.

HTML Document
A document written in HyperText Markup Language.

HTTP
The abbreviation for Hypertext Transfer Protocol, HTTP is used to link and transfer **hypertext** documents.

Hyperlink
A reference (link) from some point in one **hypertext** document to (some point in) another document or another place in the same document. A **browser** usually displays a hyperlink in some distinguishing way, e.g. in a different colour, font or style. When the user activates the link (e.g. by clicking on it with the mouse) the browser will display the target of the link.

Hypermedia
The hypertext concept extended to include linked multiple media.

Hypertext
This term describes the system that allows documents to be cross- linked in such a way that the reader can explore related documents by clicking on a highlighted word or symbol.

Inline Images
These are the graphics contained within a **Web document**.

Internet Access Provider (IAP)
A company or other organisation which provides access to the Internet to businesses and/or consumers. An IAP purchases an Internet link from another company that has a direct link to the Internet and resells portions of that **bandwidth** to the general public.

Internet Service Provider (ISP)
A company which provides other companies or individuals with access to, or presence on, the Internet. Most ISPs are also **Internet Access Providers**; extra services include help with design, creation and administration of **World-Wide Web sites**, training, etc.

IP
The abbreviation for Internet Protocol, IP refers to the set of communication standards that control communications activity on the Internet. An IP address is the number assigned to any Internet-connected computer.

ISDN
The abbreviation for Integrated Services Digital Network, ISDN is a telecommunications standard that uses digital transmission technology to support voice, video and data communications applications over regular telephone lines.

JPEG
The acronym for Joint Photographic Experts Group, JPEG is an image compression format used to transfer colour photographs and images over computer networks. Along with **GIF**, it is one of the most common ways photos are moved over the Web (the file extension is usually JPG).

Links
These are the **hypertext** connections between Web pages. This is a synonym for **anchor**, **hotlinks** or **hyperlinks**.

Mosaic
This is the common name of a **World Wide Web** multimedia **browser** programme developed at the National Center for Supercomputing Applications in Urbana-Champaign, Ill. The official, copyrighted name of the programme is NCSA Mosaic(tm).

Modem
An electronic device for converting between serial data (typically RS-232) from a computer and an audio signal suitable for transmission over telephone lines. In one scheme the audio signal is composed of silence (no data) or one of two frequencies representing 0 and 1. Modems are distinguished primarily by the maximum **baud** rate they support. Baud rates can range from 75 baud up to 28800 and beyond.

MPEG
The acronym for Moving Pictures Expert Group, MPEG is an international standard for video compression and desktop movie presentation. A special viewing application is needed to run MPEG files on your computer.

Node
A device attached to a network. A node uses the network as a means of communication and has an address on the network.

Protocol
A set of standards that define how traffic and communications are handled by a computer or network routers.

Protocol stack
A layered set of **protocols** which work together to provide a set of network functions. Each intermediate layer uses the layer below it to provide a service to

the layer above. ISO's seven layer model (Open Systems Interconnect) is an attempt to provide a standard framework within which to describe protocol stacks.

Router
A communications device designed to transmit signals via the most efficient route possible.

Search Engine
This term refers to a programme that helps users find information in text-oriented databases or in the Web. There are a number of search engines in the Web, like Altavista (http://www.altavista.digital.com) or Yahoo (http://www.yahoo.com).

Server
A computer system that manages and delivers information for **client** computers. In the Web, a server is an information provider, where clients are information consumers.

SGML
The abbreviation for Standard Generalized Markup Language, SGML is an international standard for the publication and delivery of electronic information.

Shareware
This term refers to software that is available on public networks. Users are asked to remit a small amount to the software developer.

Tags
These are formatting codes used in **HTML documents**. Tags indicate how parts of a document will appear when displayed by browsing software.

TCP-IP
The basic protocols controlling applications on the Internet; it stands for "transmission control protocol/Internet protocol."

TELNET
A protocol that allows users (TELNET-**clients**) to log into a remote computer (a TELNET-**server**) and establish a terminal connection with it. The user can then run local programmes installed only on the remote computer. In most cases a user account with password on the remote computer is required.

TIFF
This is the acronym for Tagged Image File Format, a graphic file format developed by Aldus and Microsoft. Mosaic supports the viewing of TIFF images.

Trumpet Winsock
A popular **TCP/IP protocol stack** for Windows.

URL
This is the abbreviation for Uniform Resource Locator, the addressing system used in the **World Wide Web** and other Internet resources. The URL contains information about the method of access, the **server** to be accessed and the path of any file to be accessed.

Viewer
A programme used for presenting graphics, audio and video files. Programmes that allow the viewing of graphic files (e.g. files with the extension **GIF, JPG, BMP, TIF**) and the hearing of audio files fall into this category.

WAIS
The abbreviation for Wide Area Information Service, WAIS is a Net-wide system for looking up specific information in Internet databases.

Web Document
An **HTML document** that is browsable on the Web.

Webmaster
This term refers to the person in charge of administrating a **World Wide Web site.**

Web server
A **server** process running at a **web site** which sends out **web pages** in response to **HTTP** requests from remote **browsers**

Web site
Any computer on the Internet running a **World Wide Web server** process. A particular web site is identified by the hostname part of the **URL.**

Web Page
See **Web Document**.

Webspace
This term refers to the space created by the World Wide Web. The totality of Web documents in the Internet.

World Wide Web - WWW
Also known as WWW, Web or W3, the World Wide Web is a **hypertext-** based Internet service used for browsing Internet resources.

Appendix V. Participants

Nevin Açikgöz
Aegean Agricultural Research
Institute
PO Box 9
35661 Menemen, Izmir
Turkey
Tel: +90 232 8461331
Fax: +90 232 8461107

Mike J. Ambrose
Department of Applied Genetics
John Innes Institute
Norwich Research Park, Colney Lane
Norwich NR4 7UH
United Kingdom
Tel: +44 1603 452571
Fax: +44 1603 456844
email: michael.ambrose@bbsrc.ac.uk

Siyka Angelova
Institute of Plant Genetic Resources
4122 Sadovo
Bulgaria
Tel: +359 32 267625
Fax: +359 32 270270

János Apostol
Research Institute for Fruitgrowing
and Ornamentals
Park u. 2
1223 Budapest
Hungary
Tel: +36 1 2268035
Fax: +36 1 2268037

Dave Astley
Genetic Resources Unit
Horticulture Research International
Wellesbourne, Warwick CV35 9EF
United Kingdom
Tel: +44 1789 470382
Fax: +44 1789 472023
email: dave.astley@hri.ac.uk

Frank Begemann
ZADI/IGR
Villichgasse 17
53177 Bonn
Germany
Tel: +49 228 9548212/210
Fax: +49 228 9548149
email: begemann@zadi.de

Valeri Bourdous
Belarus Research Institute
of Arable Farming & Fodders
Timiryazev St. 1
Zhodino, Minsk Region 222160
Republic of Belarus
Tel: +375 1775 32087
Fax: +375 1775 33840

Pierre Campo
GEVES - Domaine du Magneraud
17700 Surgères
France
Tel: +33 5 46683093
Fax: +33 5 46693024
email: p-campo@calvanet.calvacom.fr

Daniel Combes
IBEAS
Université de Pau et des Pays de
l'Adour
Avenue de l'Université
64000 Pau
France
Tel: +33 5 59923147
Fax: +33 5 59841696
email: combdelb@crisv1.univ-pau.fr

Neculai Cristea
Gene Bank Suceava
Boulevard 1 Decembrie 1918 nr. 17
5800 Suceava
Romania
Tel: +40 30 227087
Fax: +40 30 227087

Branislav Dozet
Oil Crops Dept.
Institute of Field and Vegetable Crops
Maksima Gorkog 30
21000 Novi Sad
F.R. Yugoslavia
Tel: +381 21 411888 / 413833
Fax: +381 21 413833

Marjan Dremelj
Agricultural Institute of Slovenia
Hacquetova 2
1000 Ljubljana
Slovenia
Tel: +386 61 1375375
Fax: +386 61 1375413

Iva Faberová
Research Institute of Crop Production
Drnovská 507
16106 Prague - Ruzyne
Czech Republic
Tel: +420 2 360851
Fax: +420 2 365228
email: faberova@genbank.vurv.cz

Lothar Frese
BAZ Gene Bank
Bundesallee 50
38116 Braunschweig
Germany
Tel: +49 531 596617
Fax: +49 531 596365
email: frese@pf.fal.de

Thomas Gass
IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy
Tel: +39 6 51892221
Fax: +39 6 5750309
email: t.gass@cgnet.com

Vincent Gensollen
GEVES - Domaine du Magneraud
17700 Surgères
France
Tel: +33 5 46683020
Fax: +33 5 46683024

Siegfried Harrer
ZADI / IGR
Villichgasse 17
53177 Bonn
Germany
Tel: +49 228 9548204 or ..210
Fax: +49 228 9548149
email: harrer@zadi.de

Tom Hazekamp
IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy
Tel: +39 6 51892236
Fax: +39 6 5750309
email: t.hazekamp@cgnet.com

Theo van Hintum
Centre for Genetic Resources, The Netherlands (CGN, CPRO-DLO)
PO Box 16
6700 AA Wageningen
The Netherlands
Tel: +31 317 477078
Fax: +31 317 418094
email: th.j.l.vanhintum@cpro.dlo.nl

Roel Hoekstra
Centre for Genetic Resources, The Netherlands (CGN, CPRO-DLO)
PO Box 16
6700 AA Wageningen
The Netherlands
Tel: +31 317 477077
Fax: +31 317 418094
email: r.hoekstra@cpro.dlo.nl

László Holly
Institute for Agrobotany
Külsõmezõ 15
2766 Tápiószele
Hungary
Tel: +36 53 380016
Fax: +36 53 380072

Lajos Horváth
Institute for Agrobotany
Külsõmezõ 15
2766 Tápiószele
Hungary
Tel: +36 53 380070
Fax: +36 53 380072

Morten Hulden
Nordic Gene Bank
PO Box 41
23053 Alnarp
Sweden
Tel: +46 40 461790
Fax: +46 40 462188
email: morten@ngb.se

Rachel Janes
National Fruit Collections
Wye College
Brogdale Road
Faversham, Kent, ME13 8XZ
United Kingdom
Tel: +44 1795 590272
Fax: +44 1795 590272
email: nfc@wye.ac.uk

Drazen Jelovac
Maize Research Institute 'Zemun Polje'
Slobodana Bajica 1
11080 Beograd-Zemun
F.R. Yugoslavia
Tel: +381 11 617434
Fax: +381 11 197890
email: djelovac@eunet.yu

Daniel Jiménez Krause
ZADI/IGR
Villichgasse 17
53177 Bonn
Germany
Tel: +49 228 9548132
Fax: +49 228 9548149
email: krause@zadi.de

Marcel Jongen
Centre for Genetic Resources,
The Netherlands (CGN, CPRO-DLO)
PO Box 16
6700 AA Wageningen
The Netherlands
Tel: +31 317 477078
Fax: +31 317 418094
email: m.w.m.jongen@cpro.dlo.nl

Wolfgang Kainz
Federal Office of Agrobiology
Wieningerstr. 8
4020 Linz
Austria
Tel: +43 732 81 261 / 260
Fax: +43 732 385482
email: genbank@agrobio.gv.at

Helmut Knüpffer
Institut für Pflanzengenetik und
Kulturpflanzenforschung - Genbank
Corrensstraße 3
06466 Gatersleben
Germany
Tel: +49 39482 5283
Fax: +49 39482 5155
email: knupffer@ipk-gatersleben.de

Vahur Kukk
Jógeva Plant Breeding Institute
2350 Jógeva
Estonia
Tel: +372 77 22565
Fax: +372 77 22443
email: vahur@sort.sai.ee

Annick Le Blanc
GEVES - Domaine du Magneraud
17700 Surgères
France
Tel: +33 5 46683093
Fax: +33 5 46683024
email: joelle@calvanet.calvacom.fr

Elinor Lipman
IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy
Tel: +39 6 51892251
Fax: +39 6 5750309
email: e.lipman@cgnet.com

Ivan Lozanov
Institute of Plant Genetic Resources
4122 Sadovo
Bulgaria
Tel: +359 2 4930026
Fax: +359 32 270270
email: 0037@main.infotel.bg

Lorenzo Maggioni
IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy
Tel: +39 6 51892231
Fax: +39 6 5750309
email: l.maggioni@cgnet.com

Isaura Martin
(representing F. Gonzalez Lopez)
Centro de Recursos Fitogeneticos, INIA
Autovia de Aragon Km 36, PO Box 45
28800 Alcala de Henares, Madrid
Spain
Tel: +34 1 8819261 or 8819286
Fax: +34 1 8819287

Petter Marum
The Norwegian Crop Research Institute, Løken Research Station
2940 Heggenes
Norway
Tel: +47 613 40205
Fax: +47 613 40665
email: petter.marum@planteforsk.nlh.no

Andrei Omeltchenko
N.I. Vavilov Institute of Plant Industry
Bolshaya Morskaya Street 42
190000 St Petersburg
Russia
Tel: +7 812 3116117
Fax: +7 812 3118762
email: vir@glas.apc.org

Pietro Perrino
Germplasm Institute - CNR
Via G. Amendola 165/A
70126 Bari
Italy
Tel: +39 80 5583608
Fax: +39 80 5587566
email: germpp04@area.ba.cnr.it

Wieslaw Podyma
Plant Breeding and Acclimatization Institute (IHAR)
Radzików
05-870 Blonie
Poland
Tel: +48 22 7252611
Fax: +48 22 7254714
email: w.podyma@ihar.edu.pl

Isaak Rashal
Institute of Biology
3 Miera street
2169 Salaspils
Latvia
Tel: +371 2 945435
Fax: +371 9 345412
email: izaks@genet.edu.lv

Genrikh Razorenov
N.I. Vavilov Institute of Plant Industry
Bolshaya Morskaya Street 42
190000 St Petersburg
Russia
Tel: +7 812 3144727
Fax: +7 812 3118762
email: vir@glas.apc.org

Hrvoje Rukavina
Dept. Plant Breeding, Genetics and
Biometrics - Faculty of Agiculture
Svetosimunska st. 25
10000 Zagreb
Croatia
Tel: +385 1 2335 777
Fax: +385 1 215 300
email: ikolak@magr.agr.hr

Victor K. Ryabchoun
Institute of Plant Production
Moskovskyi pr 142
310060 Kharkov
Ukraine
Tel: +380 572 921033
Fax: +380 572 920354

Jerzy Serwinski
FAO / AGPS
Via delle Terme di Caracalla
00100 Rome
Italy
Tel: +39 6 522555767
Fax: +39 6 522556347
email: jerzy.serwinski@fao.org

Magdalena Sevcíková
Grassland Research Institute
756 54 Zubrí
Czech Republic
Tel: +420 651 583195 or ..6
Fax: +420 651 583197

Istvàn Sulyok
Institute for Agrobotany
Külsõmezõ 15
2766 Tápiószele
Hungary
Tel: +36 53 380070
Fax: +36 53 380071

Wojciech K. Swiecicki
Institute of Plant Genetics
Streszynska 34
60-479 Poznan
Poland
Tel: +48 61 233511
Fax: +48 61 233671
email: wswi@igr.poznan.pl

Gusztáv Szabó
Institute for Agrobotany
Külsõmezõ 15
2766 Tápiószele
Hungary
Tel: +36 53 380070
Fax: +36 53 380072
email: gusztav.szabo@ella.hu

Kostutis Tamosiunas
Lithuanian Institute of Agriculture
5051 Dotnuva - Akademija
Lithuania
Tel: +370 57 37145
Fax: +370 57 56996

Manuel Tavares de Sousa
National Plant Breeding Station
PO Box 6
7351 Elvas
Portugal
Tel: +351 68 622844 / 47
Fax: +351 68 629295

Ian D. Thomas
IGER
Aberystwyth, Wales SY23 3EB
United Kingdom
Tel: +44 1970 828255
Fax: +44 1970 828357
email: thomasi@bbsrc.ac.uk

David Turnbull
Potato section
Scottish Agricultural Science Agency
East Craigs, Edinburgh EH128NJ
United Kingdom
Tel: +44 131 2448938
Fax: +44 131 2448940
email: turnbull@sasa.gov.uk

Mária Záková
Research Institute of Plant
Production
Bratislavska 122
92168 Piestany
Slovakia
Tel: +421 838 722311 or ..12
Fax: +421 838 26306

email: vurv@bb.sanet.sk

Anne Zanetto
INRA, Unité de recherches sur les
Espèces Fruitières et la Vigne
BP 81
71 av. Edouard Bourleaux
33883 Villenave d'Ornon cedex
France
Tel: +33 5 56843089
Fax: +33 5 56843083
email: zanetto@bordeaux.inra.fr

Grzegorz Zurek
Botanical Gardens, IHAR
Ul. Jedziecka 5
85-687 Bydgoszcz
Poland
Tel: +48 52 721407
Fax: +48 52 224454
email: obihar@bydg.pdi.net

Appendix VI. Central Crop Databases and their managers

European Crop Databases established under ECP/GR

For easier reference the databases are listed by crop categories. The list also includes the details of the database managers who could not attend the Workshop.

Crop/crop group	Managing institution	Database manager	Contact details
ECP/GR Network: Cereals			
Avena	FAL, Braunschweig, Germany	Lothar Frese	Tel: +49-521 596617 Fax: +49-531 596365 email: frese@pf.fal.de
Barley	IPK, Gatersleben, Germany	Helmut Knüpffer	Tel: +49-39482 5283 Fax: +49-39482 5155 email: knupffer@ipk-gatersleben.de
Secale	IHAR, Radzikow, Poland	Wieslaw Podyma	Tel: +48-22 7252611/7253611 Fax: +48-22 7254715 email: w.podyma@ihar.edu.pl
Triticale	RAC, Nyon, Switzerland	Gert Kleijer	Tel: +41-22 363 4722 Fax: +41-22 362 1325 email: geert.kleijer@rac.admin.ch
Wheat	GEVES, Surgères, France	Annick Le Blanc (data from 15 EU countries)	Tel: +33-5 46683093 Fax: +33-5 46683024 email: joelle@calvanet.calvacom.fr
	RICP, Prague Ruzyne, Czech Rep.	Iva Faberová (data from all other countries)	Tel: +420-2 360 851 Fax: +420-2 365 228 email: faberova@genbank.vurv.cz
Maize	MRI, Belgrade-Zemun, F.R. Yugoslavia	Drazen Jelovac	Tel: +381-11-617 434 Fax: +381-11-197 890 email: djelovac@eunet.yu
ECP/GR Network: Grain legumes			
Phaseolus	Federal Office for Agrobiology, Linz, Austria	Wolfgang Kainz	Tel: +43 732 81 261/ 260 Fax: +43 732 385482 email: genbank@agrobio.gv.at
Vicia faba	INRA, Le Rheu, France	Gérard Morin	Tel: +33-2 99285144/5127 Fax: +33-2 99285120
Pisum	IPG, Poznan, Poland	Wojciech Swiecicki	Tel: +48-61 233511 Fax: +48-61 233671 email: wswi@igr.poznan.pl
	JIC, Norwich, UK	Mike J. Ambrose	Tel: +44 1603 452571 Fax: +44 1603 456844 email: michael.ambrose@bbsrc.ac.uk
Lupinus	IPG, Poznan, Poland	Wojciech Swiecicki	Tel: +48-61 233511 Fax: +48-61 233671 email: wswi@igr.poznan.pl
Cicer	ENMP, Elvas, Portugal	Manuel Tavares de Sousa	Tel: +351-68 622844/47 Fax: +351-68 629295
Glycine	VIR, St Petersburg, Russia	(to be confirmed)	Tel: +7-812 314 4732 Fax: +7-812 311 8762 email: vir@glas.apc.org
Lens	AARI, Izmir, Turkey	Nevin Açikgöz	Tel: +90-232 8461331

Fax: +90-232 8461107

Crop/crop group	Managing institution	Database manager	Contact details
ECP/GR Network: Forages			
Lolium *Trifolium repens*	IGER, Aberystwyth, UK	Ian D. Thomas	Tel: +44 1970 828255 Fax: +44 1970 828357 email: thomasi@bbsrc.ac.uk
Lathyrus latifolius, tuberosus, heterophyllus, sylvestris	IBEAS, Pau, France	Daniel Combes	Tel: +33-5 59923147 Fax: +33-5 59841696 email: combdelb@crisv1.univ-pau.fr
Vicia spp.	CNR, Bari, Italy	Pietro Perrino	Tel: +39 80 5583608 Fax: +39 80 5587566 email: germpp04@area.ba.cnr.it
other Vicieae'	University of Southampton, UK	Frank A. Bisby	Tel: +44-1703-595000 Fax: +44-1703-594269
Agropyron	IPGR, Plovdiv, Bulgaria	Siyka Angelova	Tel: +359-32 267625 Fax: +359-32-270270
Arrhenaterum elatius, *Trisetum flavescens*	OSEVA Pro, Zubri, Czech Rep.	Magdalena Sevcíková	Tel: +420-651 583195/6 Fax: +420-651 583197
Perennial *Medicago*	INRA-GEVES, Guyancourt, France	Vincent Gensollen	Tel: +33-5 4668 3020 Fax: +33-5 4668 3087
Other perennial forage legumes	Inst. of Agrobotany, Tápiószele, Hungary	Lajos Horvath	Tel: +36-53-380070/71 Fax: +36-53-315 844
Poa	IPK, Malchow, Germany	Evelin Willner	Tel: +49-38425 20316 Fax: +49-38425 20316 email: e.willner@SO.HS-Wismar.de
Bromus	Inst. of Agrobotany, Tápiószele, Hungary	Lajos Horvath	Tel: +36-53-380070/71 Fax: +36-53-315 844
Trifolium pratense	Inst. of Agrobotany, Tápiószele, Hungary	Lajos Horvath	Tel: +36-53-380070/71 Fax: +36-53-315 844
T. subterraneum, annual *Medicago*	SIA, Badajoz, Spain	F. Gonzalez Lopez	Tel: +34-24 449761 Fax: +34-24 449748
T. alexandrinum, *T. resupinatum*	ARO, Bet Dagan, Israel	Avi Perevolotsky	Tel: +972-3 968 3389 Fax: +972-3 966 9642
Dactylis/Festuca	IHAR, Bydgoszcz, Poland	Grzegorz Zurek	Tel: +48-52-721407 Fax: +48-52-224454 email: obihar@bydg.pdi.net
Phalaris/Agrostis *Phleum*	NGB, Alnarp, Sweden	Merja Vetelainen	Tel: +46-40-461790 Fax: +46-40-462188 email: merja@ngb.se
ECP/GR Network: Fruit trees			
Prunus	INRA, Bordeaux, France	Anne Zanetto	Tel: +33 5 56843089 Fax: +33 5 56843083 email: zanetto@bordeaux.inra.fr
Malus	NFC, Wye College, UK	Rachel Janes	Tel: +44 1795 590272 Fax: +44 1795 590272 email: nfc@wye.ac.uk

ECP/GR Network: Vegetables			
Allium	HRI, Wellesbourne, UK	Dave Astley	Tel: +44 1789 470382 Fax: +44 1789 472023 email: dave.astley@hri.ac.uk
Brassica	CGN, Wageningen, The Netherlands	Ietje W. Boukema	Tel: +31-317 477077 Fax: +31-317 418094 email: i.w.boukema@cpro.dlo.nl

Other international databases of particular relevance to the European collaboration on plant genetic resources (not exhaustive)

Crop	Managing institution	Database manager	Contact details
Network: International *Beta* Network			
Beta	FAL, Braunschweig, Germany	Lothar Frese	Tel: +49-521 596617 Fax: +49-531 596365 email: frese@pf.fal.de
Network: APIC (Assoc. of Potato Intergenebank Collaborators)			
Solanum	CGN, Wageningen, The Netherlands	Roel Hoekstra	Tel: +31 317 477077 Fax: +31 317 418094 email:r.hoekstra@cpro.dlo.nl
Network: ESCORENA			
Sunflower	Inst. for Field and Veg. Crops, Novi Sad, F.R. Yugoslavia	Branislav Dozet	Tel: +381 21 411888/413833 Fax: +381 21 413833
Network: (not applicable)			
Vitis	BAZ-Geilweilerhof, Siebeldingen, Germany	Erika Dettweiler	Tel: +49-6345 410 Fax: +49-6345 41177

Appendix VII. Abbreviations and acronyms used in the text

CCDB	Central Crop Database
DSC	Documentation Support Center
ECP/GR	European Cooperative Programme for Crop Genetic Resources Networks
EGDS	Eastern European Germplasm Documentation Systems Project
ESCORENA	European System of Cooperative Research Networks in Agriculture
GENRES	Informationssystem Genetische Ressourcen (Information System for Genetic Resources) (ZADI)
GEVES	Groupe de Contrôle et d'Etude des Variétés et des Semences, France
HRI	Horticulture Research International, Wellesbourne, UK
MAFF	Ministry of Agriculture, Fisheries and Food of the UK
NGB	Nordic Gene Bank, Alnarp, Sweden
RICP	Research Institute of Crop Production, Prague, Czech Republic
UPOV	Union pour la Protection des Obtentions Végétales, Geneva, Switzerland
WIEWS	World Information and Early Warning System on plant genetic resources (FAO)
ZADI/IGR	Zentralstelle für Agrardokumentation und -information / Informationszentrum für Genetische Ressourcen, Bonn, Germany (Centre for Agricultural Documentation and Information / Information Centre for Genetic Resources)