Chapter 8: Sources of information on existing germplasm collections

E. Bettencourt
Genetic Resources, Ecophysiology and Plant Breeding Unit
Instituto Nacional de Recursos Biológicos, I.P. (INRB, I.P.)
Instituto Nacional de Investigação Agrária (INIA)
Oeiras, Portugal
(currently on leave of absence)
E-mail: eliseu.bettencourt@gmail.com

Abstract

This update of chapter 8 of the Technical Guidelines, on sources of information on existing germplasm collections, looks at new developments in information technology and available data, supplementing the information in the original chapter. It takes into account two types of data sources: data maintained in electronic storage-and-retrieval systems at the institutional, national, subregional, regional and global level; and data made available in the specialized scientific literature. This review and update does not pretend to be exhaustive but, rather, to identify examples that will give users an idea of the situation regarding information on plant genetic resources, from which they can extrapolate to identify and explore other ways of finding, identifying, analysing and making use of the information available.

Introduction

Information technology is a rapidly evolving field where news of technological innovations, new tools and breakthroughs has a permanent place in the headlines. Even if we don’t realise it, information technology is, for better or for worse, inextricably associated to our day-to-day life. For those who deal with large amounts of data and have the need to store, retrieve, analyse and use information, information technology is a blessing that cannot be praised highly enough.

The planning, organization and carrying out of missions for collecting germplasm started, in a scientific manner, in the 1920s. These missions have been based on an understanding of the importance of biodiversity in the production of more and better foods — and more recently, a realization of the important role that biodiversity plays in helping to address the challenges of climate change and constantly evolving pests and diseases, by providing a source of new materials and knowledge.

The total number of germplasm accessions maintained in ex situ collections worldwide is estimated as 7.4 million, but in spite of this apparently vast genepool, it is estimated that fewer than 30% (1.9 to 2.2 million) of this total are unique accessions (FAO 2010). The need for further diversity is a reality, and collecting must continue in order to fill in the diversity gap in existing germplasm collections.

This chapter is a synthesis of new knowledge, procedures, best practices and references for collecting plant diversity since the publication of the 1995 volume Collecting Plant Diversity; Technical Guidelines, edited by Luigi Guarino, V. Ramanatha Rao and Robert Reid, and published by CAB International on behalf of the International Plant Genetic Resources Institute (IPGRI) (now Bioversity International), the Food and Agriculture Organization of the United Nations (FAO), the World Conservation Union (IUCN) and the United Nations Environment Programme (UNEP). The original text for Chapter 8: Sources of Information on Existing Germplasm Collections, authored by M. C. Perry and E. Bettencourt, has been made available online courtesy of CABI. The 2011 update of the Technical Guidelines, edited by L. Guarino, V. Ramanatha Rao and E. Goldberg, has been made available courtesy of Bioversity International.
CHAPTER 8: SOURCES OF INFORMATION ON EXISTING GERmplASM COLLECTIONS

Since the 1995 publication of these Technical Guidelines, a great deal of germplasm has been collected: approximately 240,000 new accessions (FAO 2010). And new technologies for recording, storing, managing and retrieving data have become available.

This update of the chapter on sources of information on existing germplasm collections attempts to synthesize these new developments and to supplement the information in the original chapter.

Current status

Despite data being available in greater quantity and quality than ever before, it is not always recorded and maintained in a format that makes it easily, readily and universally available. However, “if there's a will, there's a way”, and provided that the data exists, there will be always a way to make the most of it.

In this update, two types of sources are considered:

1. data maintained in electronic storage-and-retrieval systems at the institutional, national, subregional, regional and global level
2. data made available in the scientific literature

Presently, many institutions maintaining germplasm collections have the information concerning their holdings online. If one wants to collate information about a particular geographical area or a specific species, it may be worth starting with a search of the holdings of either genebanks located in the target region or genebanks for which the target species is the focus of their research.

Sources of information at the accession/sample/species level

At the genebank level

The following genebanks have information on their holdings online:

- Centre for Genetic Resources, Plant Genetic Resources (CGN-PGR), the Netherlands (www.cgn.wur.nl/UK/CGN+Plant+Genetic+Resources)
- Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany (www.ipk-gatersleben.de/Internet)
- Centro de Recursos Fitogenéticos (C.R.F.), Spain (wwwx.inia.es/webcrf/CRFing/PaginaPrincipal.asp)
- N. I. Vavilov All-Russian Scientific Research Institute of Plant Industry (VIR), Russian Federation (www.vir.nw.ru/data/dbf.htm)
- Recursos Genéticos e Biotecnologia (EMBRAPA), Brazil (www.cenargen.embrapa.br)
- ISOPlexis, Centro de Estudos da Macaronésia, Centro de Competência das Ciências da Vida, Universidade da Madeira, Portugal (www3.uma.pt/cem/isoplexis/coleccao.html)
- Kew's Millennium Seed Bank, Wakehurst Place, UK (http://data.kew.org/sid/about.html)

The information systems of these genebanks allow the user to browse, search and view information on their germplasm holdings. Many offer the possibility to download the search results, and some include a function for requesting germplasm samples.

At the national level

Many countries have established national inventories on plant genetic resources that are freely and widely available online. This enables participating countries to contribute to their biodiversity and conservation obligations at the national level and to meet the requirements of international agreements such as the Convention on Biological Diversity (CBD), the Global Plan of Action for the Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture (GPA), the International Treaty on Plant Genetic Resources for Food and Agriculture (“the Treaty”) and the Global Biodiversity Information Facility (GBIF).
• Australian Plant Genetic Resource Information Service (AusPGRIS) (www2.dpi.qld.gov.au/eXtras/asp/auspgris)
• National Inventory of Plant Genetic Resources for Food and Agriculture, Austria (www.genbank.at)
• Genetic Resources Institute of the National Academy of Sciences, Azerbaijan (www.cac-biodiversity.org/aze/aze_institute.htm)
• Canadian Plant Germplasm System (PGRC) (http://pgrc3.agr.gc.ca/about-propos_e.html)
• Plant Genetic Resources Documentation (EVIGEZ), Czech Republic (http://genbank.vurv.cz/genetic/resources)
• Estonia (www.sordiaretnus.ee)
• Bureau des Ressources Génétiques (BRG), France (www.brg.prd.fr/index.php)
• Informationssystem Genetische Ressourcen (GENRES), Germany (www.genres.de/pgrdeu)
• Japan National Institute of Agrobiological Sciences (NIAS) Genebank (www.gene.affrc.go.jp/databases_en.php)
• Lithuanian Institute of Agriculture (www.lzi.lt)
• Centre for Genetic Resources, Plant Genetic Resources (CGN-PGR), the Netherlands (www.cgn.wur.nl/UK/CGN+Plant+Genetic+Resources)
• Nordic Countries (Denmark, Finland, Iceland, Norway, Sweden) (www.nordgen.org/ngb)
• Plant Breeding and Acclimatization Institute (IHAR), Poland (www.ihar.edu.pl/gene_bank)
• N. I. Vavilov All-Russian Scientific Research Institute of Plant Industry (VIR) (www.vir.nw.ru/data/dbf.htm)
• Agricultural Institute of Slovenia (www.kis.si/pls/kis/kis_web)
• Centro de Recursos Fitogenéticos (CRF), Spain (wwwx.inia.es/webcrf/CRFing/PaginaPrincipal.asp)
• USDA Agricultural Research Service, National Genetic Resources Program (NGRP), USA (www.ars-grin.gov)

These national inventories collate information from different data sources (germplasm collections) and constitute a network of centres and people dedicated to conserving the genetic diversity of crop plants and their wild relatives.

At the regional/subregional level

At the regional and subregional level, there are initiatives to collate and make available data on national holdings. These initiatives are normally undertaken to assist countries in meeting their national, regional and global commitments regarding documentation and information on their plant genetic resources.

An example of such a regional initiative is EURISCO (http://eurisco.ecpgr.org/home_page/home.php), a web-based catalogue that provides information about ex situ plant germplasm collections maintained in Europe. The EURISCO catalogue contains passport data on more than 1 million accessions of crop plants from 41 countries, representing 5,396 genera and 35,235 species (genus-species combinations including synonyms and spelling variants). EURISCO is hosted at and maintained by Biodiversity International on behalf of the Secretariat of the European Cooperative Programme for Plant Genetic Resources (ECPGR).

ECPGR also maintains a series of European central crop databases established through the initiative of individual institutes and of ECPGR working groups. The ECPGR Central Crops Databases (ECCDBs) (www.ecpgr.cgit.org/germplasm_databases/list_of_germplasm_databases/crop_databases.html#c5509) include the following:

• Cereals: Avena, Barley, Maize, Secale, Triticale, Wheat
• Forages: Dactylis, Festuca, Lolium, Phleum, Poa, Medicago (annual and perennial), Minor Forage Grasses (Agropyron, Agrostis, Alopecurus, Arrhenatherum, Bromus, Phalaris, Trisetum), Minor Forage Legumes (Astragalus, Anthyllis, Coronilla, Desmodium, Hedysarum, Lotus, Melilotus, Onobrychis, Ornithopus, Physanthyllis, Tetragonolobus, Vicia), Trifolium spp., Trifolium subterraneum, Vigna
• Oil and Protein Crops: Arachis, Cicer, Glycine, Lathyrus, Lens, Lupinus, Phaseolus, Pisum, Vicia faba, Vigna
• Sugar, Starch and Fibre Crops: Hemp, Beta, Flax, Potato (cultivated), Potato (wild bearing tubers)
• Temperate Fruits: Malus, Prunus, Pyrus, Ribes/Rubus, Vitis, Minor Fruit Trees
• Vegetables: Allium, Brassica, Chicory, Cucurbits, Cyphomandra, Physalis, Eggplant, Lactuca, Beta, Lettuce, Minor Leafy Vegetables, Pepino, Pepper, Potato (cultivated and wild), Spinach, Tomato, Umbellifer

The databases hold passport data and, to varying degrees, characterization and primary evaluation data of the major collections of the respective crops in Europe.

The Eastern Africa Plant Genetic Resources Network (EAPGREN) is a regional project of the national agricultural research systems of Burundi, Eritrea, Ethiopia, Kenya, Madagascar, Rwanda, Sudan and Uganda, aimed at strengthening collaboration, networking and linkages between the conservation and utilization of plant genetic resources at both the national and subregional level.

The objective of the EAPGREN data portal (www.nordgen.org/portal/index.php?scope=eapgren&PHPSESSID=4pa906im0ghl5l2i8a8015d1) is to publish passport data of all the documented accessions in the EAPGREN countries. This information is published according to the List of Multi-crop Passport Descriptors (MCPD). The data set contains 3,932 records searchable through a simple search form.

An example of a subregional network can be illustrated by the South East European Development Network on Plant Genetic Resources (SEEDNet) (www.seednet.nu), a network of 13 countries of the Balkan region. The available data (199,325 records) can be searched for accessions either originating or stored in the SEEDNet region.

There are a number of regional networks for plant genetic resources in all continents. Although they do not have a centralized online system with information about the participants’ germplasm holdings, they are an important entry point for finding information. Some examples are given below.

The Mesaoamerican Plant Genetic Resources Network (REMERFI), was established in the early 1990s by seven Mesoamerican countries (Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua and Panama), aiming at strengthening the national programmes for plant genetic resources of Mesoamerican countries through collaborative research and training. The network serves as a platform for countries to address key issues on the conservation and use of their plant genetic resources at the regional level. Information can be sought by contacting the regional coordinator at remerfi@es.com.sv (IPGRI 2001).

The Southern African Development Community (SADC) countries (Angola, Botswana, Democratic Republic of Congo, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, United Republic of Tanzania, Zambia and Zimbabwe) have pooled their resources and established the SADC Plant Genetic Resources Centre (SPGRC) (www.sadc.int/fanr/agricresearch/spgrc/index.php) in Lusaka, Zambia, where the base collection for long-term storage is maintained and activities for the region are coordinated. The individual national plant genetic resources centres (NPGRCs) maintain active collections for short-term storage for immediate use in crop improvement.

The genebank information for the SADC Plant Genetic Resources Network has been standardized through the development of the SPGRC Documentation and Information System (SDIS) at SPGRC and installed at all the NPGRCs. In addition, an assessment of existing collections that are held by plant breeders and in genebanks in the various countries has been carried out, together with herbaria surveys of important indigenous crops and their wild relatives. However, the information is not centralized and is maintained and managed by each of the individual member countries.

The Genetic Resources Network for West and Central Africa (GRENEWECA) was created in 1998 and comprises the national programmes for the genetic resources of West and Central Africa, regional or international institutions of research, and nongovernmental organizations in the subregion. One of the main duties of the network is to “facilitate the circulation and exchange of information within as well as among...
member countries and those outside the network”. More information can be obtained from the network coordinator: R. S. Vodouhe, Bioversity International (r.vodouhe@cgiar.org).

ECPGR maintains a list of PGR regional networks, which can be consulted and accessed through the following link: (www.ecpgr.cgiar.org/networks/inter_regional_coop/pgr_regional_nw_coordinators.html).

At the global level

As its germplasm information-exchange network, the Consultative Group on International Agricultural Research (CGIAR) and its partners implemented the System-wide Information Network for Genetic Resources (SINGER) (http://singer.cgiar.org), which provides easy access to information about the diversity maintained by SINGER members, covering more than half a million accessions of crops and wild relatives.

As a result of the collaboration between Bioversity International on behalf of the CGIAR, the Global Crop Diversity Trust and the Secretariat of the Treaty, a global portal to information about plant genetic resources has been established, bringing together data from SINGER, EURISCO and the US Germplasm Resources Information Network (GRIN). The portal, GENESYS (www.genesys-pgr.org), is a gateway through which germplasm accessions from genebanks around the world can be found and ordered directly through the web interface. In addition to passport data, the system provides access to characterization and evaluation data as well as to environmental information associated with the accessions’ collecting sites.

From 1974 to 2003, Bioversity International (under its previous names of IBPGR and IPGRI) supported germplasm collecting missions around the world. The IBPGR/IPGRI Supported Missions Database (http://singer.cgiar.org/index.jsp?page=coll-sample-data) gives access, at a sample level, to passport data for about 130,000 samples, as well as access to 27,000 original reports, collecting forms and other documents in PDF format.

The Global Biodiversity Information Facility (GBIF) (http://data.gbif.org/welcome.jsp;sessionid=37820DBEE26CAA62FAB10ADD1EE5B67D) is focused on making biodiversity data available online for scientific research, conservation and sustainable development. Enabling access to 293,485,946 data records, the GBIF information infrastructure is an internet-based index of a globally distributed network of interoperable databases that contain primary biodiversity data: information on museum specimens, field observations of plants and animals in nature, and results from experiments. The search interface also supports searches about the occurrence of species at particular times and places.

Sources of information at a metadata level

The FAO World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture (WIEWS) (http://apps3.fao.org/wiews/wiews.jsp) is both a dynamic worldwide mechanism to foster information exchange and an instrument for the periodic assessment of the state of the world’s plant genetic resources for food and agriculture.

WIEWS maintains information on country profiles, including the structure of 190 national programmes and activities for plant genetic resources, as well as information on worldwide ex situ collections, containing summary records (metadata) of germplasm holdings on 7,184,418 accessions of 53,109 species, reported by more than 1,500 national, regional or international genebanks.

There are over 2,500 botanic gardens worldwide, maintaining 80,000 plant species, representing nearly one-third of all known plant species (FAO 2010). Botanic Gardens Conservation International (BGCI) was established in 1987 as a small secretariat under the auspices of the World Conservation Union (IUCN). Today, BGCI is an international organization with over 700 members and other partners from 118 countries worldwide and has the mission “to ensure the world-wide conservation of threatened plants, the continued existence of which are intrinsically linked to global issues including poverty, human well-being and climate change”. BGCI maintains a database, PlantSearch (www.bgci.org/plant_search.php), which can locate rare and threatened plant species in cultivation around the world. The database, presently including over
575,000 records, is compiled from lists of living collections submitted to BGCI by the world's botanic gardens. It also maintains a database, GardenSearch (www.bgci.org/garden_search.php), with over 2,826 records, that can find a botanic garden (member or not) anywhere in the world.

Globally, there are approximately 3,400 herbaria with approximately 10,000 associated curators and biodiversity specialists. These herbaria, collectively, maintain an estimated 350,000,000 specimens. The Index Herbariorum, a Global Directory of Public Herbaria and Associated Staff (http://sciweb.nybg.org/science2/IndexHerbariorum.asp), keeps records for each herbarium on street and web address, number and type of specimens and history as well as names, contact information and areas of expertise of associated staff.

Sources of information in the scientific literature

Although not specifically addressing plant genetic resources, there are other tools that can help to unearth data from web-available sources.

Zanran (www.zanran.com) is a good example of such tools. Zanran is a search engine that unearths data in charts, graphs and tables. It indexes and maps the numerical content on the web, finding “semi-structured” data, which could be anything from a graph in a PDF report, a table in an Excel spreadsheet or a bar chart shown as an image on an HTML page. In the near future, it will also process PowerPoint and Word documents. The system examines millions of images and determines whether they are a graph, chart or table and if they have numerical content.

A test search made using the indicator “germplasm” yielded 1,753 matching results. Hovering the mouse over the PDF icon on the left column opens a document-preview window, facilitating the identification and selection of desired documents.

The test search results varied from data on germplasm characterization and germplasm duplication to germplasm distribution and the holdings of germplasm collections.

Besides source data housed in electronic storage and retrieval systems, there are enormous amounts of valuable data made available in the scientific literature as well as through “grey literature” like the reports of collecting missions and field notebooks. In spite of its value and importance, such information is not in a format that makes it easily, readily and universally available. The need for recording and maintaining data in a way that is easily, readily and universally available has prompted initiatives to digitalize and make available original collecting reports.

In addition, many accounts of germplasm collecting trips have been published in specialized journals like the FAO/Bioversity International Plant Genetic Resources Newsletter. An index of articles from issues 25 to 100 of the Newsletter and from issues 1 to 24 of the FAO Plant Introduction Newsletter (the previous title of the publication) has been published, facilitating the search and selection of information, although they are not available online at this time.

ECPGR publishes the Newsletter for Europe (until recently on hard copy and now in PDF format) and Bioversity International publishes a series of regional or network-related newsletters, which, in many issues, carry articles and notices on the results of germplasm collecting missions.

Bioversity’s website (www.bioversityinternational.org), under the label “publications”, offers a search facility that provides access to a number of these newsletters in PDF format.

Other sources of information

As pointed out by Lister (2011), natural history collections comprise not only the products of opportunistic collecting but are also repositories of major surveys, particularly in the major national museums and institutions. Coupled with the availability of accurate provenance data on the material, they can provide a rich source of data.
Museum specimen labels and registers ideally indicate the place and date of collection. However, even when such information is available, considerable work may be required to make it accessible for research, for example, by georeferencing obscure place names and entering all records onto an electronic database.

Although data on material held in natural history collections is sometimes not in a format that is easily, readily and universally accessible, this information represents an invaluable source of data and can prove to be extremely useful in collating information on past plant surveying and collecting as well as being an important repository of environmental and ecological information.

University museums contain some of the richest and most extraordinary collections in the world, and historic collections provide historical snapshots of past biodiversity. They are a rich source of material for studies in phylogenetics, population dynamics and conservation biology (MacDonald and Ashby 2011). Again, in many cases, the existing data may not be in a readily usable format but, nevertheless, it is of great importance and worth consulting and analysing.

**Future Challenges/Needs/Gaps**

It is estimated that a considerable amount of information still lies on genebank shelves in the form of field notebooks, reports of collecting missions and collectors’ notes. It is certainly a challenge but also necessary to have that valuable data analysed and recorded in a way that facilitates access to it, thus promoting the use of the germplasm with which this information is associated.

The existing and available information systems on plant genetic resources are priceless sources of information, although data on characterization and evaluation tend to be poor or absent. Closing this gap by adding this information, thus fostering the use of germplasm, would represent a giant step towards addressing the threats these resources and agriculture are facing, from changes in climate and constantly evolving pests and diseases.

**Conclusions**

During the time since the publication of the Technical Guidelines in 1995, the dissemination of information through the internet has increased tremendously in both quantity and quality.

Today, there is more information available than ever before and, despite the fact that it is not always recorded and maintained in a format that makes it easily, readily and universally available, the user has access to an enormous number of data sources.

This review and update of the chapter 8, on sources of information on existing germplasm collections, does not pretend to be exhaustive but, rather, to identify examples that can give the user an idea of the actual situation in the field of information on plant genetic resources.

The user will be able to extrapolate from these examples, to go on to identify and explore other ways of finding, identifying, analysing and making use of information on past germplasm collecting missions and to build on this information in the planning and implementation of future explorations and collecting missions.

**References**


CHAPTER 8: SOURCES OF INFORMATION ON EXISTING GERMPLASM COLLECTIONS


Internet resources

Agricultural Institute of Slovenia: www.kis.si/pls/kis//kis.web
Botanic Gardens Conservation International (BGCI), GardenSearch: www.bgci.org/garden_search.php
Bureau des Ressources Génétiques (BRG), France: www.brg.prd.fr/index.php
Canadian Plant Germplasm System (PGRC): http://pgrc3.agr.gc.ca/about-propos_e.html
Centre for Genetic Resources. Plant Genetic Resources (CGN-PGR), the Netherlands: www.cgn.wur.nl/UK/CGN+Plant+Genetic+Resources
Centro de Recursos Fitogenéticos (C.R.F.), Spain: www.x.inia.es/webcrf/CRFing/PaginaPrincipal.asp
CGIAR System-wide Information Network for Genetic Resources (SINGER): http://singer.cgiar.org
EAPGREN data portal: www.nordgen.org/portal/index.php?scope=eapgren&PHPSESSID=4pa906im0gfl5li2il8u80l5d1
ECPGR Central Crops Databases (ECCDBs): www.ecpgr.cgiar.org/germplasm_databases/list_of_germplasm_databases/crop_databases.html#c5509
ECPGR list of PGR regional networks: www.ecpgr.cgiar.org/networks/inter_regional_coop/pgr_regional_nw_coordinators.html
Estonia: www.sordiaretus.ee
EURISCO: http://eurisco.ecpgr.org/home_page/home.php
FAO World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture (WIEWS): http://apps3.fao.org/wiews/wiews.jsp
GENESYS: www.genesys-pgr.org
Genetic Resources Institute of the National Academy of Sciences, Azerbaijan: www.cac-biodiversity.org/aze/aze_instgen.htm
Global Biodiversity Information Facility (GBIF): http://data.gbif.org/welcome.htm?sessionid=37820DBEE26CAA62FAB10ADD1EE5B67D
IBPGR/IPGRI Supported Missions Database: http://singer.cgiar.org/index.jsp?page=coll-sample-data
Index Herbariorum, a Global Directory of Public Herbaria and Associated Staff: http://sciweb.nyb.org/science2/IndexHerbariorum.asp
Informationssystem Genetische Ressourcen (GENRES), Germany: www.genres.de/pgrdeu
ISOPlexis, Universidade da Madeira, Portugal: www3.uma.pt/isoplexis/index_eng.html
Kew’s Millennium Seed Bank, Wakehurst Place, UK: http://data.kew.org/sid/about.html
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany: www.ipk-gatersleben.de/Internet
Lithuanian Institute of Agriculture: www.lzi.lt
N. I. Vavilov All-Russian Scientific Research Institute of Plant Industry (VIR), Russian Federation: www.vir.nw.ru/data/dbf.htm
National Genetic Resources Program (NGRP), USA: www.ars-grin.gov
National Inventory of Plant Genetic Resources for Food and Agriculture, Austria: www.genbank.at
Nordic Countries (Denmark, Finland, Iceland, Norway, Sweden): www.nordgen.org/ngb
Plant Breeding and Acclimatization Institute (IHAR), Poland: www.ihar.edu.pl/gene_bank
Plant Genetic Resources Documentation (EVIGEZ), Czech Republic: http://genbank.vurv.cz/genetic/resources
Recursos Genéticos e Biotecnologia (EMBRAPA), Brazil: www.cenargen.embrapa.br
SADC Plant Genetic Resources Centre (SPGRC): www.sadc.int/fanr/agricresearch/spgrc/index.php
South East European Development Network on Plant Genetic Resources (SEEDNet): www.seednet.nu