

Chapter 11 A GLOBAL PORTAL ENABLING WORLDWIDE ACCESS TO INFORMATION ON CONSERVATION AND USE OF BIODIVERSITY FOR FOOD AND AGRICULTURE

Elizabeth Arnaud⁴⁶, Sónia Dias⁴⁷, Michael Mackay⁴⁸

Bioversity International, Biodiversity Informatics Project, "Understanding and Managing Biodiversity" Programme- Via dei Tre Denari 472/a, 00057 Maccarese (Rome), Italy -
tel: +39 0661181233, tel: +39 0661181204, e.arnaud@cgiar.org

Pete Cyr, Candice Gardner⁴⁹

USDA/ARS North Central Regional Plant Introduction Station, G212 Agronomy Hall, Iowa State University, Ames, IA, 50011, USA

Peter Bretting⁵⁰

USDA/ARS Office of National Programs, Room 4-2212, Mail Stop 5139 5601 Sunnyside Avenue, Beltsville, MD 20705-5139 USA, Peter.Bretting@ars.usda.gov

Gary Kinard

USDA/ARS National Germplasm Resources Laboratory, Bldg 004,
Beltsville Agricultural Research Center, Beltsville, MD, 20705, USA

Luigi Guarino

Global Crop Diversity Trust- Secretariat, FAO Headquarters, Viale delle Terme di Caracalla,
100153 Rome, Italy

Selim Louafi

ITPGRFA Secretariat, FAO Headquarters, Viale delle Terme di Caracalla, 10053 Rome, Italy

⁴⁶ SINGER Coordinator

⁴⁷ EURISCO Coordinator

⁴⁸ Principal Investigator, Global Germplasm Information Portal project

⁴⁹ Project Manager, GRIN-Global Project

⁵⁰ Principal Investigator, GRIN-Global Project

Abstract: *The genetic diversity of crop species is vital to ensuring more consistent, abundant and nutritious harvests, especially in developing countries. The agricultural community needs ready access to information that enables the most efficient and economical choice of appropriate diversity to address these challenges. Considerable amounts of diversity are stored in genebanks. However, information about this agricultural biodiversity is fragmented and scattered, and is particularly difficult to access in the resource-poor countries where it is most needed. It is often hard to find information about germplasm that can provide, for example, resistance to pests and diseases, or other stresses that reduce productivity and yields. The global germplasm information portal currently under development will specifically address information-retrieval obstacles faced by breeders, crop researchers and other users. The global information portal will use data standards and integrate data sets from existing information systems, or through a new scalable genebank information management system called GRIN-Global, for online publication within a global accession-level information platform. This global portal will be developed as a one-stop entry point to worldwide genebank information, enabling users to select and order samples online.*

Keywords: agricultural biodiversity; biodiversity informatics; conservation; use; global information portal; genebank information system; international standards; interoperability; international crop collections

1. INTRODUCTION

The conservation and use of the genetic diversity of domesticated species and their wild relatives is crucial for the well-being of future generations. This agricultural biodiversity is a fundamental source of solutions to food crises in developing countries. Mining the knowledge accumulated over decades by agricultural research institutions is essential for understanding crop diversity and using it effectively to help sustain abundant harvests and a healthy environment. Considerable amounts of diversity are stored worldwide as seeds and plantlets, depending on the species. Each individual sample conserved is known as an accession, and collectively they are known as germplasm. There are many germplasm collections, or genebanks, around the world, including those of the international research institutes of the Consultative Group on International Agricultural Research [CGIAR, 200]. However, information about germplasm is still much too fragmented and scattered, and its access is particularly difficult in resource-poor countries where it is most needed. Plant breeders and pre-breeders have assessed and utilized the genes and traits of traditional varieties to increase crops' resistance to stresses such as drought, floods, pests and diseases. But most of the data related to this fundamental research are still not easily accessible by other potential users. Genebank curators need feedback from

the users of diversity, mainly breeders and farmers, in order to optimize their collecting and conservation strategies.

The selection by scientists of germplasm that can be used to breed crops with beneficial characteristics such as enhanced productivity, nutrition or stress resistance requires easier access to the wealth of biological data scattered worldwide. A 'global accession-level information portal' is intended to serve as the online, one-stop entry point to the information managed by crop genebanks worldwide. Users will be able to order samples using an online 'shopping-cart' system that will be compliant with the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA – referred to here as 'the Treaty').

International cooperation and open exchange of genetic diversity are essential for food security. For the first time, ready access to plant genetic resources and the fair sharing of benefits arising from the use of these resources are being implemented at the international level through the Treaty and its Standard Material Transfer Agreement (SMTA). The Treaty aims to: recognize the enormous contribution of farmers to the diversity of crops that feed the world; establish a global system to provide farmers, plant breeders and scientists with access to plant genetic materials; and ensure that recipients share benefits they derive from the use of genetic materials with the countries that have provided those materials.

2. THE GLOBAL INFORMATION PORTAL ON GENETIC RESOURCES FOR FOOD AND AGRICULTURE

Bioversity International, the Global Crop Diversity Trust and the Secretariat of the Treaty are partnering in the development of the global information portal for agricultural biodiversity to facilitate the wider use of this diversity. It is often hard to find information about which accessions can provide, for example, resistance to pests and diseases, or other stresses that reduce agricultural productivity. This effort specifically addresses the barriers faced by breeders, crop researchers and others in locating specific types of genetic diversity within the world's genebanks. The project includes the development of: a) data standards for agrobiodiversity and data interchange protocols; b) GRIN-Global, a multilingual, scalable and flexible genebank information management system; and c) an online global germplasm information portal that will serve as the one-stop entry point to a wealth of data on the

characteristics and provenance of accessions held in genebanks worldwide, as well as a system for ordering samples of these materials.

These components will be integrated to provide a single point through which quality data will be accessible online, so that a wide audience can locate and order samples (see figure 11-1)

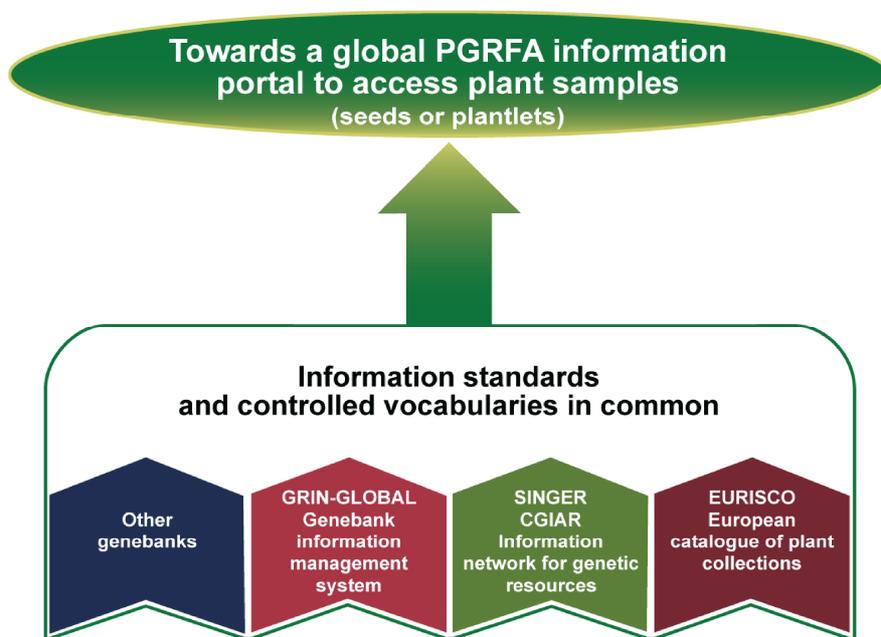


figure 11-1 THE GLOBAL INFORMATION PORTAL FOR *ex situ* COLLECTIONS OF PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE (PGRFA)

2.1 DATA STANDARDS

Bioversity has long been coordinating the development of internationally recognized crop descriptor lists to describe the key characteristics of genetic resources. The list of multi-crop passport descriptors (MCPD), developed by the UN Food and Agriculture Organization (FAO) and Bioversity, is a core standard for any information system on crop genetic resources because it contains the 'identity cards' for germplasm accessions. In the framework of the project for the global germplasm information portal, essential characterization and evaluation descriptors for traits that would have immediate impact on productivity are being identified for priority crops to enable a quick trait-based search and selection of important diversity. While

data generated by genebank curators are usually formatted as descriptors, information generated by breeders is more heterogeneous and sometimes requires an ontological approach for optimal access and utilization.

In the framework of the Generation Challenge Programme (GCP), a CGIAR programme aimed at using plant genetic diversity, advanced genomic science and comparative biology to develop tools and technologies for plant breeders in the developing world, a crop ontology – a predefined structure of terminology used for data searching and data curation – is being developed to annotate the information and underpin the information queries on the GCP platform and, in the future, on the global germplasm information portal. A standard for enabling the publishing of observation – or characterization and evaluation – data is also under development and will be deployed for the global portal.

2.2 EXISTING WORLDWIDE WEB CATALOGUES OF CROP COLLECTIONS

SINGER [SINGER, 2009], EURISCO and GRIN are online catalogues of plant collections that provide inventories of conserved agricultural diversity and offer primary access for identifying and localizing accessions. By linking these catalogues with other types of data such as information from breeders, the global germplasm information portal will strengthen and facilitate their role as gateways to the world's agricultural biodiversity.

2.2.1 SINGER, THE ONLINE CATALOGUE OF THE CGIAR INTERNATIONAL COLLECTIONS

The CGIAR System-wide Genetic Resources Programme [SGRP, 2008] unites several CGIAR research centres in a common effort to sustain biodiversity for current and future generations. Much of SGRP's efforts to date have focused on plant genetic resources; however attention is also being given to forest, animal and aquatic genetic resources, given the interdependence of all components of agricultural biodiversity. The CGIAR is committed to helping build a global information portal on genetic resources, and SGRP serves to bring together the CGIAR Centres in this common mission.

The germplasm collections of the CGIAR centres conserve a significant part of the diversity of the world's most important agricultural crops. The CGIAR centres have placed their collections under the inter-governmental authority of FAO, which means

that the diversity they conserve is held in the public domain and available for distribution in accordance with the Treaty. SINGER provides a single entry point into the inventories of the 11 CGIAR centres that have genebanks and the Asian Vegetable Research and Development Centre (AVRDC) (see figure 11-2). Information on almost 700,000 accessions from 77 collections is available. An online sample-ordering gateway has been added to SINGER so that anyone can now access material via a 'shopping-cart' function and send a request to the appropriate germplasm providers.



figure 11-2 SINGER homepage, (<http://singer.cgiar.org/>)

2.2.2 EURISCO, THE EUROPEAN PLANT GENETIC RESOURCES SEARCH CATALOGUE

The European Cooperative Programme for Plant Genetic Resources (ECPGR) aims to facilitate the long-term cooperative conservation and utilization of plant genetic resources in Europe. The Programme, which is entirely financed by participating countries and is coordinated by a secretariat at Bioversity International, operates through broadly focused networks dealing with groups of crops or general themes related to plant genetic resources.

The EURISCO Catalogue provides a web-based entry point for accessing information on European plant collections conserved in genebanks (see figure 11-3). EURISCO is a European network of 42 National Focal Points for the National Inventories that make European biodiversity data available worldwide, linking users to information on over 1.1 million accessions from 38 European countries. For the past four years, the EURISCO network has tackled a wide range of information needs in Europe and improved countries' capacities to make available biodiversity information. The

EURISCO Catalogue currently has information on 1,450 genera and 8,650 species. These samples of crop diversity represent more than half of the accessions maintained in Europe's genebanks and roughly 19% of total worldwide holdings.

EURISCO is designed to serve as the European information hub on plant genetic resources, providing access to data from all European National Inventories and all crops. It will contribute to implementing the Treaty by being part of the registration and reporting mechanism. Within the framework of the European Genebank Integrated System (AEGIS), plant accessions are registered in EURISCO by holding countries to be part of the collection of unique and important European accessions maintained for the long term.

The EURISCO search mechanisms, data and metadata standards, web services and other necessary components of the internet-based information infrastructure are constantly being improved with the feedback of European National Focal Points.

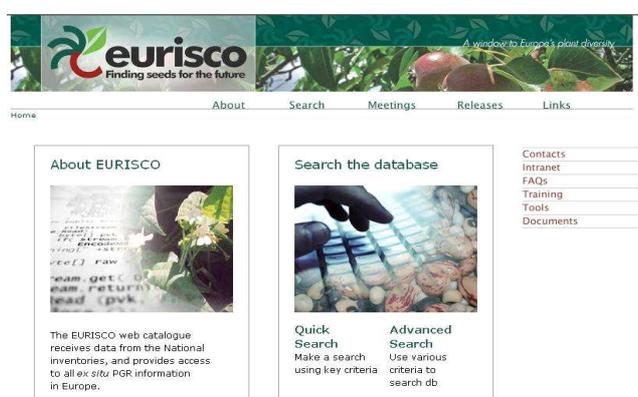


figure 11-3 EURISCO homepage (<http://eurisco.ecpgr.org>)

2.2.3 THE GERMLASM RESOURCES INFORMATION NETWORK (GRIN) OF USDA-ARS

The United States National Plant Germplasm System (NPGS), administrated by the US Department of Agriculture-Agricultural Research Service (USDA-ARS), is a cooperative effort by US public and private organizations to preserve the genetic diversity of agriculturally important plants and encourage their use. The NPGS aids the global agricultural community by acquiring, preserving, evaluating, documenting, and distributing crop germplasm in order to improve the quality and productivity of crops.

GRIN provides online germplasm information about the collections of plants, animals, microbes and invertebrates managed by USDA-ARS (figure 11-4). The GRIN database already houses information on more than 530,000 distinct accessions of plants in the NPGS. The Canadian national agricultural genebank system has adopted GRIN and customized it for its own needs as GRIN-Canada. USDA-ARS is committed to supporting and enhancing GRIN, which was first developed more than 20 years ago. Now, thanks to the partnership among the Global Crop Diversity Trust, USDA-ARS and Bioversity, redesign and software upgrades will enable the new GRIN-Global system to be used by genebanks of all sizes all over the world, making more information about plant genetic resources available to researchers worldwide.



figure 11-4 GRIN homepage (<http://www.ars-grin.gov/npgs/>)

2.2.4 COLLABORATION WITH THE GLOBAL BIODIVERSITY INFORMATION FACILITY (GBIF)

In 2006, a collaborative agreement was signed between Bioversity International and GBIF to publish SINGER and EURISCO data on the GBIF portal. GBIF protocols were deployed in some partner genebanks but the facility to maintain the technology on-site was lacking. GBIF is working to offer alternative solutions for providers with limited internet connectivity, particularly in developing countries. However, GBIF does not provide access to crop trait data or the ability to cross-reference records of SINGER, EURISCO, and GRIN, which would make searches much more efficient and relevant.

2.3 THE GLOBAL GERMLASM INFORMATION PORTAL, A ONE-STOP ENTRY POINT TO ORDER SAMPLES FROM GENE BANKS WORLDWIDE

The global accession level-information portal will be *the* entry point for users to mine genetic variation by using combinations of data on the characteristics, environments and other aspects of genetic diversity in order to identify accessions of interest and order them online. The data underlying this platform will initially come from SINGER, EURISCO, and GRIN, which are in the public domain. The online ordering toolkit currently in development will be available for genebanks that need to implement a system compliant with the recommendations of the Treaty. The global information portal will enable users to access information using a login from their own web pages, storing items such as their favorite search criteria or their previous germplasm selections. Feedback and comments on the quality of the information and use of the plant material will be elicited from users, with blogs, quality-scoring and other fora. The global germplasm information portal will also enable access to complementary knowledge such as bibliographic material, best practices and a directory of crop collections.

2.4 GRIN-GLOBAL, AN EASY-TO-USE INTERNET-BASED INFORMATION MANAGEMENT SYSTEM FOR THE WORLD'S PLANT GENE BANKS.

GRIN-Global, being developed by the Agricultural Research Service of the US Department of Agriculture (USDA-ARS) (see figure 11-5) in collaboration with Bioversity and the Global Crop Diversity Trust, is composed of scalable data-management software suitable for any genebank, and is organized into three tiers. Tier 1 (the presentation tier) can be a web browser or full desktop client; Tier 2 (the business tier) responds to data requests from Tier 1 clients via web services technology; and Tier 3 (the data tier) is a database server (either MySQL, Oracle, or Microsoft SQL Server) that provides data to Tier 2. GRIN-Global is multilingual by design to facilitate adoption by genebanks throughout the world (initially the GRIN-Global project will support French, Spanish, Russian, Arabic, and English). Bioversity will assist in deploying the software in genebanks that request it. Through this system, the genebanks will produce formatted data sets that can be exchanged and published online on any platform, including the global accession-level information portal. GRIN-Global will enable genebanks to conserve and use genetic resources

more effectively, and also help researchers, farmers and producers make the best possible use of information.

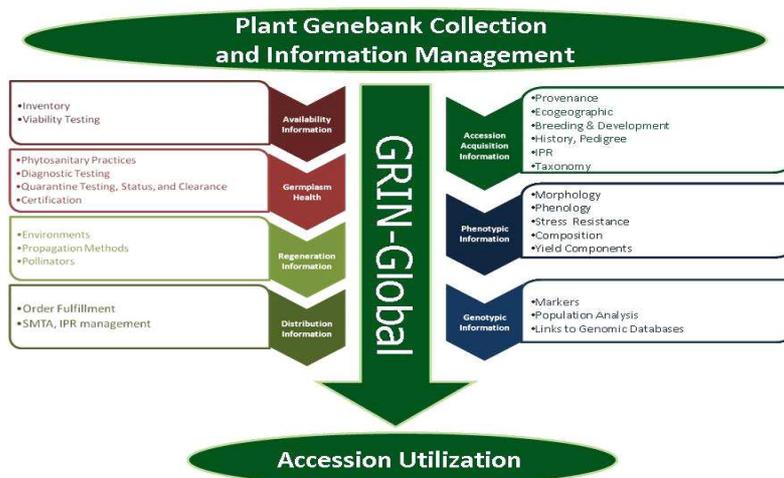


figure 11-5 Concept of GRIN-Global universal genebank data management system

3. ISSUES AND CHALLENGES IN ACCESSING INFORMATION ON AGRICULTURAL DIVERSITY

For any information system and, particularly for global platforms that aspire to serve as reliable sources of knowledge, the most important goal is to publish quality data. Online and offline quality-checking tools enable data providers to check their data sets prior to publication – for example taxonomy and georeference checkers. EURISCO already provides such a model of online data-quality checking with quality reports.

Data providers require peer review of their data sets for quality checking and scientific validation. Various technological solutions are available to facilitate communication within communities of practice, including: online fora, D-groups, blogs, scratchpads, and wikis. The challenge of peer review is to establish an adequately moderated process that involves appropriate experts and provides useful feedback to data providers. Metadata on the quality, method and source of the data needs to be included in order to facilitate secondary use of data sets in diversity analysis.

3.1 LIMITING FACTORS TO DEPLOYMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES (ICT)

Despite the willingness to adopt the most effective technologies for managing data on crop collections, several factors constrain the deployment of adequate ICT in agricultural institutions, particularly in developing countries. For example, the funds allocated by research institutions to conservation and documentation of crop diversity are often insufficient. The Global Crop Diversity Trust is now addressing this problem with long-term funding for priority crop collections worldwide, including activities on documentation and data exchange. The most important factor limiting the flow of quality data is the lack of proper training of genebank staff in best practices for managing genebank information and using the latest informatics tools. As a result, these data sets often do not fully meet international quality standards for data exchange, which limits their value. And because many data providers continue to use off-line replication methods, the most up-to-date data are not readily available to researchers in a timely manner – a situation that is often exacerbated by the need for human interaction to replicate these off-line data sets.

4. CONCLUSIONS

The global germplasm information portal for managing information associated with plant genetic resources for food and agriculture will eventually extend beyond SINGER, EURISCO and GRIN to include many other materials. Through the deployment of GRIN-Global, genebanks in developing countries will be the primary focal points for training in the use of data standards and best practices for information management and sharing. A wide range of ICT solutions for interoperability with large data sets must be made available to address the needs of data providers with unequal capacities. The greatest challenge will be to keep the system simple, rapid and nimble, and to offer customizable data-analysis tools to end users.

International portals are now maintained with the objective of increasing exposure to and accessibility of the data sets submitted by their providers. However, this is only possible if data providers can assess the value these portals are adding to their data sets. Providers expect portal publishers to monitor the use of their data and elicit user feedback, while users of the portals wish to download comprehensive and meaningful data sets that fit their needs.

Enabling public access to scientific datasets presupposes that scientists accept the concept of data as '*global public goods*', and support keeping their data in the public domain. Data sets are usually made available once results have been published through a peer-review process.

Finally, the effective deployment of information and communication technologies and tools always relies on adequate institutional capacity and training of data providers, as well as an effective mechanism for the technology transfer and sustained use.

5. FUTURE WORK

The long-term goal of the global germplasm information portal will be to collaborate and combine forces in order to enable online access to information and germplasm from all crop collections worldwide through an easy-to-use ordering system. In addition to the already established collaboration among the Secretariat of the Treaty, the Global Crop Diversity Trust, USDA-ARS, ECPGR and CGIAR-SGRP, Bioversity International is seeking to collaborate with communities and scientific societies such as Biodiversity Information Standards (TDWG) and the Generation Challenge Programme in order to expand the standards for data formatting and exchange to increase the portal's usability. ICT-Ensure is facilitating the networking of teams involved in the application of ICT for agriculture and biodiversity information at the European level. This kind of initiative, if duplicated in other regions, could have a significant impact on global ICT in support of agricultural biodiversity systems.

Partners from developing countries and European genebanks will be invited to participate in testing and deploying the global germplasm information portal, its accompanying data standards and GRIN-Global, and the features of these tools will be translated into as many languages as possible. Once the final English version of GRIN-Global is released, it will be translated into Arabic, French, Spanish, and Russian, and deployed in genebanks. In the meantime, a core set of partners is being selected to install and test early prototypes of this software.

In the on-going development of a crop ontology, it will be necessary to assess novel approaches to defining standards based on concept occurrences and text mining. Methodologies will be established to facilitate a community-based infrastructure for data quality checking and expert advice.

Development of the online 'shopping-cart' ordering system will continue within the global information platform to address the varying capacities of data providers. A persistent identifier (PID) server is also being developed by the Secretariat of the Treaty to uniquely identify germplasm requesters and providers, keep records of transactions and enable consecutive logins to the portal. SINGER will serve as one of the sources of data for the first test of the PID application.

6. REFERENCES

[ECPGR, 2009] ECPGR. 2009. A Strategic Framework for the Implementation of a European Genebank Integrated System (AEGIS). A Policy Guide. European Cooperative Programme for Plant Genetic Resources (ECPGR). Bioversity International, Rome, Italy.

[SGRP, 2008] About SGRP, United for Diversity. 2008. SGRP web site:
<http://www.sgrp.cgiar.org/?q=node/35> (last accessed 10 August 2009).

[GRIN, 2009] The Germplasm Resources Information Network – USDA/ARS web site:
<http://www.ars.usda.gov/main/main.htm> (last accessed 10 August 2009).

[Kaplan, 2008] Kaplan K., 2008. The International Partnership to Develop a Global Plant Genebank Information System. ARS press release, 11 February 2008:
<http://www.ars.usda.gov/is/pr/2008/080211.2.htm> (last accessed 10 August 2009).

[Dias, 2008] Dias, S. 2008. EURISCO Coordinator Report 2004 – 2008 - ECPGR Phase VII:
http://www.ecpgr.cgiar.org/SteeringCommittee/SC11/Docs/EURISCO_ProgressReport_SD.pdf

[EURISCO, 2009] The European Plant Genetic Resources Search Catalogue -EURISCO Website, March 2009. <http://eurisco.ecpgr.org/> (last accessed 10 August 2009).

[ITPGRFA, 2009] International Treaty on Plant Genetic Resources for Food and Agriculture
<http://www.planttreaty.org/> (last accessed 10 August 2009).

[SINGER, 2009] The System-wide Information Network for Genetic Resources
<http://singer.cgiar.org/> (last accessed 10 August 2009).

