

# **Innovation in Conservation, How Information Technology Tools Improve the Ex Situ Management of Plant Genetic Resources**

T.J.L. van Hintum  
Centre for Genetic Resources, The Netherlands (CGN)  
Wageningen University and Research Centre  
Wageningen  
The Netherlands

**Keywords:** plant genetic resources, documentation, information technology

## **Abstract**

**Many new technologies highly relevant to the PGR community have become available over the past years, especially in the fields of genomics and information technology. The effect of the second category of technologies on the ex situ management of plant genetic resources is explored. After a low initial level of standardization in genebank documentation systems, a strong increase of access and exchange of information could be observed. Important elements behind this increase are standardization - creating the possibility to interpret each others information - and web services - creating the possibilities for machines to access each others information via the internet. Some aspects of these developments such as the application of ontologies, persistent identifiers and anticipated developments such as the increased use of open source software, are presented. Also for the user of PGR and associated information many improvements can be foreseen. Ongoing developments include on-line querying and ordering facilities, and trait prediction. A final element that will be described is the development of virtual genebanks: thanks to IT tools, the management of PGR and the interfaces to these PGR can be decoupled. As a result, anyone can develop a website giving access to PGR, including the possibility to order material in one or more genebanks anywhere in the world.**

## **INTRODUCTION**

The community involved in the ex situ conservation of Plant Genetic Resources (PGR) is traditionally and by nature a conservative community. Conservation implies keeping what you have, preventing loss or change. However, it is the PGR that need to be conserved, and not the methodology to do so. In terms of innovations, the PGR conservation methodology often lags behind its scientific environment. Apart from conservatism, this is also due to the low level of investments made in PGR programs. As a result, PGR programs have difficulties showing their true value to the scientific and breeding community, and investments stay low.

The two main technologies providing options for innovation in the PGR community are the biotechnology and the information technology. Biotechnology is outside the scope of this paper. However, it is good to note that many of the issues of genomics are in nature information technological issues. Where high throughput sequencing becomes feasible for application in genebanks, the true challenges lie no longer in generating the data but in handling and analyzing them. Bioinformatics will be the key discipline in bringing innovations from genomics into genebanks.

The other technology that has and is changing society and also has a large impact on the PGR community is Information Technology. This is a very wide area: 'any technology that helps to produce, manipulate, store, communicate, and/or disseminate information', and comprises of the hardware, the networks, and the software including the user interfaces. The central question in the title of this paper 'how information technology tools improve the ex situ management of plant genetic resources' therefore offers a wide spectrum of answers, varying from the use of GPS during collecting trips, to the creation of virtual web-based genebanks. To describe this spectrum, a few components of it will be described in some detail: genebank documentation systems and the information supply to

users. These two elements will together lead to a new concept that will also be described briefly: the virtual genebank.

## **GENEBANK DOCUMENTATION SYSTEMS**

Documentation has played a vital role in genebank management since their establishment in the sixties. Genebanks have always needed to record at least what accessions they had and where they were stored, but also the observations on the material, the results of germination tests and the administration of where the material was sent formed natural parts of genebank documentation since the beginning. Initially this documentation was maintained on paper, later transferred to the computer in spreadsheets or text files, and in many cases later transferred into a proper database. This process of the development of a documentation system usually took place locally, without any standards or even examples of other similar systems. Obviously this resulted in very different systems, with different structures and coding systems.

When in the eighties and nineties the need to exchange data between genebanks grew, the need for standardization became obvious. Some standards developed and were generally adopted, such as the Multi Crop Passport Descriptor List (MCPD, FAO/IPGRI 2001) defining the data elements used for recording passport information, and the use of ISO 3166-1 alpha-3 codes for countries and FAO codes for institutions (WIEWS, 2010). However this remained limited to the domain of passport data, and does not include important data elements such as the taxonomic names, crop names, etc. This very limited standardization did however allow the creation in Europe of a central database with passport data of all European genebanks (EURISCO, 2010), which obviously also suffered from the lack of standardization of elements such as the taxonomy (Hintum and Knüpffer, 2010).

Based on the success of EURISCO and the clear demand of users to get access to information about traits, the next domain of PGR data that will be exchanged will be that of the characterization and evaluation data. These data depend heavily on meta data for their interpretability. Just saying that a 'plant is short' has no meaning if it is not defined under what circumstances, when and compared to what that plant was short. Exchanging this type of data therefore is of a much higher complexity than plain passport data, and will require more complex standardization.

Another development that will have a large influence on the way data are shared between genebanks will be the introduction of web services, i.e., applications where computers communicate without interference of humans (Alonso et al., 2004). An example of an ongoing successful application of web services is the Global Biodiversity Information Facility (GBIF), who gives access to a variety of data sources allowing the user to search these sources simultaneously using only the GBIF interface. Obviously, this kind of applications is only possible with a high level of standardization, or at least the possibility to map information from one structure on the other. To allow for a proper adoption of web services, ontologies will have to be developed for the PGR data domain and persistent identifiers will have to be adopted (Page, 2008).

Yet another development in the field of genebank documentation systems will be the application of open-source documentation systems. To truly apply the advances in information technology it will become (or already is) too difficult and too expensive for each genebank to develop and maintain its own documentation system. The PGR community will need to share the software, use each others code, in a community of developers and users. The current initiative called 'GRIN Global' might be the start of such a community, provided that it will be sufficiently 'open source'. GRIN-Global aims at being a scalable genebank documentation system 'will be suitable for use by any interested genebank in the world' (GRIN-global, 2010). If the PGR community will be able to establish a shared open source application, this will solve many of the standardization issues since the open source software will automatically set the standard, allowing very efficient information exchange.

The result of the described developments will be a network of interlinked 'data-

sources' providing information about conserved germplasm. At that stage, in principle, anyone can build an interface giving access to a specific subset of the data in the network, serving the needs of a specific group of users, such as the lettuce breeders, the barley research community, the policy makers negotiating international treaties, etc. (Hintum et al., 2010). provided that the genebank community will also implement the possibility to order germplasm via web services, the virtual genebank is born. This will be discussed later.

### **INFORMATION SUPPLY TO USERS**

Another important perspective of the application of IT in genebanks is that of the user of the germplasm. Currently, the user usually sees a genebank website, often with a rather clumsy interface to the passport data of the material maintained by that genebank. Access to the important characterization and evaluation data is rarely provided, and the possibility to order germplasm on-line is even more rare. This will have to change rapidly.

First of all, the interfaces will improve and access will be given to a wider spectrum of data and functionalities. This spectrum will cover, amongst others, the possibility to select on the basis of traits, the on-line creation of core selections and most importantly on-line ordering, including handling of the contractual issues regarding the use of the material (the so called SMTA) in click wrap agreements. These changes can not be called innovations, since they have all been developed, they just need to be implemented in the various genebanks. This, however, will require investments. The interfaces need to be created, but more importantly, data of an acceptable quality have to be made available. This is not an obvious step. For example, to allow for searches on traits, first of all the characterization and evaluation data will have to be properly computerized. And since usually these data have been generated in different years, under different circumstances, and sometimes even using different scales and methods of measurement, they will have to be standardized and combined resulting in one score per trait per accession.

Functionalities of the interfaces can be further enhanced. Search algorithms can be improved towards a 'Google style' searching in genebank data, or image searches. But also the description of the material can be enhanced by implementing trait prediction based on analyses of molecular or geographic data in statistical or artificial intelligence algorithms (see for example Kaur et al., 2008).

No matter how far the user interface to a genebank's collection is improved, the user will still be restricted to the germplasm in that genebank only, which is only a fraction of the material available in the world. So the next logical step is to create simultaneous access to information about all germplasm conserved in genebanks irrespective their geographical location. If this access to information is extended to access to material, the virtual genebank starts to appear.

### **THE VIRTUAL GENE BANK**

Imagine a user. This user can log in at a website site, get access to germplasm he or she is interested in - irrespective of where it is maintained -, select accessions and order them on-line, including the click wrap agreement to the Standard Material Transfer Agreement (FAO, 2002). Soon after ordering them, this user will receive the accessions that were ordered, shipped from different locations around the world.

The virtual genebank that this user uses is just one of very many, each group of users can have its own separate genebank, presenting those accessions and that information that this group of users is interested in. If the user is a barley scientist, it will show data of barley accessions all over the world, links from barley publications to germplasm, information about barley mapping-, NIL- and RIL-populations, etc. If the user is a policy maker it will show live data about origin and usage of germplasm, changes in the genebank holdings, overviews of the Multi Lateral System, etc. These websites will present an interface based on web services, that searches the databases of all relevant genebanks on-line and simultaneously, based on an agreed protocol. Selected material can be ordered from all genebanks maintaining a part of the selection, and the

procedure for distribution of seeds is started in all genebanks the moment the user completes his or her request. The only thing the user will notice of the fact that (s)he ordered material from a virtual genebank is that the postage will deliver the selection in several packages, one from each genebank that stored part of the selection.

This scenario does not require any new technology, but it does require a lot of standardization, normalization and quality improvement. This relates to the information in the genebanks, but also the management of the genebanks and their collections. The information provided by the nodes has to be complete and reliable, the germplasm has to be authentic and available, and the service has to be rapid and reliable. For most genebanks this requires a major quality improvement. In Europe an initiative called AEGIS in the framework of the European Collaborative Program on Plant Genetic Resources (ECPGR) is working towards this ideal of a rational and effective European virtual genebank (ECPGR, 2008).

## **CONCLUDING REMARKS**

There are many more opportunities to innovate using IT in the PGR community than those presented, think of improvements in the logistic of the genebank using bar codes or RFIDs, or the possibilities to phenotype collections using image analysis. Also the role of IT in coping with the DNA data deluge that is expected soon with the advent of affordable high throughput sequencing techniques could have been discussed. Instead, the paper concentrated on collaboration between genebanks and service to users. Standardization and quality improvement are the key words in this context, both in terms of genebank data and of genebank procedures.

Standardization and quality improvement will allow genebanks to better collaborate and thus improve the efficiency and to better use the limited financial resources available to the PGR community. It will also allow genebanks to improve their service for their users by improving access to the material and data. And the user will certainly need better access to PGR to counteract the enormous challenges that mankind will be confronted with in the coming decades.

## **Literature Cited**

- Alonso, G., Casati, F., Kuno, H. and Machiraju, V. 2004. Web Services: concepts, architectures and applications. Springer, 354p.
- ECPGR. 2008. A strategic framework for the implementation of a European genebank integrated system (AEGIS). Discussion paper. European Cooperative Programme for Plant Genetic Resources (ECPGR), Bioversity International.
- EURISCO. 2010. Web Catalogue providing access to all ex situ PGR information in Europe. URL: eurisco.ecpgr.org (accessed 18 August 2010).
- FAO. 2002. The International Treaty on Plant Genetic Resources for Food and Agriculture. Food and Agriculture Organization of the United Nations, Rome, Italy.
- FAO/IPGR. 2001. FAO/IPGRI Multi-Crop Passport Descriptors December 2001, FAO, Rome, 4p.
- GRIN-global. 2010. Website of GRIN-Global. URL: www.grin-global.org (accessed 19 August 2010).
- Kaur, N., Street, K., Mackay, M., Yahiaoui, N. and Keller, B. 2008. Molecular approaches for characterization and use of natural disease resistance in wheat. *Eur. J. Plant Pathol.* 121:387-397.
- Page, R.D.M. 2008. Biodiversity informatics: the challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics* 9:345-354.
- van Hintum, T.J.L. and Knüpffer, H. 2010. Current taxonomic composition of European genebank material documented in EURISCO. *Plant Genetic Resources: Characterisation and Utilisation* 8:182-188.
- van Hintum, T.J.L., Begemann, F. and Maggioni, L. 2010. The European Ex Situ PGR Information Landscape. p.155-171. In: L. Maurer and K. Tochtermann (eds.), *Information and Communication Technologies for Biodiversity Conservation and*

Agriculture. Shaker Verlag, Aachen.  
WIEWS. 2010. World Information and Early Warning System on PGRFA. FAO, Rome.  
URL: [apps3.fao.org/wiews](http://apps3.fao.org/wiews) (accessed 19 August 2010).

