

## Integrating genebanks into biodiversity information networks<sup>1</sup>

Helmut Knüpffer<sup>1</sup>, Dag Terje Filip Endresen<sup>2</sup>, Iva Faberová<sup>3</sup> and Samy Gaiji<sup>4</sup>

<sup>1</sup>Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Gatersleben, Germany; knupffer@ipk-gatersleben.de

<sup>2</sup>Nordic Gene Bank, POB 41, SE-230 53 Alnarp, Sweden

<sup>3</sup>Genebank, Crop Research Institute, Drnovská 507, CZ-161 06 Praha 6 – Ruzyně, Czech Republic

<sup>4</sup>Bioversity International, Via dei Tre Denari 472/a, I-00057 Maccarese, Rome, Italy

*“The efficiency of a genetic resources conservation network lies in the availability of materials and data. Increasingly, interest is being shown in wild relatives for disease resistance and qualitative characters. This gives new opportunities for crop specialists and botanists to work together. So what are the needs of genetic resources specialists with respect to botanical data?*

*First, they need up-to-date taxonomic and nomenclatural revisions at the genus or the tribe level. Second, they need the results of broad ecogeographical surveys to assist in preparing collecting missions.*

*You may think that only a limited set of plants is involved. But if the gene pool of a given crop may extend to an entire family then it looks different. If we add not only agricultural plants but fruits, vegetables, medicinal and ornamental plants, forest trees, and new biotechnology sources we see quickly that the whole plant kingdom is the resource.”*

M. Chauvet (1993)

### Abstract

Many genebanks have computerised information about the accessions they hold, but the database software and data models implemented may differ substantially. The information, however, is largely standardized, due to the wide adoption of germplasm data standards.

Despite such level of compatibility, today there is no single point of access to accession-level information from genebank collections worldwide. Some plant genetic resources (PGR) data portals such as EURISCO (European search portal), numerous Central Crop Databases, the Nordic Gene Bank (NGB), and the CGIAR's System-wide Information Network for Genetic Resources (SINGER), show that distributed data on genebank accessions can be accessed from global and regional as well as crop-specific data portals based on a conventional data warehouse model.

The Global Biodiversity Information Facility (GBIF) promotes the exchange of biodiversity related information using a new information technology called web services. This technology, deployed at the level of data providers, offers the opportunity to tap remotely into the “living” database. Germplasm collections are very similar information-wise to other biodiversity collections, such as botanical gardens or herbaria. Initiated by Bioversity International, GBIF data exchange technology was further developed to better suit the needs of the PGR community.

From 2004 on, several national genebanks as well as SINGER and EURISCO became GBIF data providers. Thus, with two million accessions, one-third of the world's germplasm holdings are currently searchable via GBIF.

---

<sup>1</sup> based on the presentation by H. Knüpffer, D. T. F. Endresen and S. Gaiji, and including material of the poster “Databases as tools for enhanced utilization of crop collections” by I. Faberová.

The wide adoption of GBIF technology for PGR information exchange will facilitate an alliance of distributed accession-level information systems and enable wide utilization of crop collections.

### **Keywords**

Crop genetic resources, GBIF, data providers, portals, web services

### **Introduction**

Genebanks are collections of living material of genetic resources of crop plants and their wild relatives, with the aim of collecting, preserving, evaluating and documenting this material, for utilisation by users such as breeders or researchers.

There are hundreds of genebanks worldwide with a total of more than six million *ex situ* germplasm accessions (FAO 1996). The database software and data models implemented in their information systems may differ substantially between genebanks.

As called by the International treaty on PGR for Food and Agriculture (Article 17), existing information systems should collaborate to form a global information network.

Present-day technological developments and the organisational frameworks developed in the plant genetic resources community and in biodiversity informatics provide the tools for the compilation of PGR information systems and their integration into wider biodiversity information networks, as will be shown below.

### **Genebanks in the context of biodiversity**

Genebanks preserve a variety of forms of agricultural and horticultural crop species, their wild and weedy forms, and other useful plants. Much of the material has been collected from farms, gardens, but also from wild and ruderal habitats such as forests, steppe, roadsides or field margins. Thus, they are biodiversity collections, like botanical gardens or herbaria (archive specimen collections). Genebank accessions share many properties and attributes with other biodiversity specimens. In terms of data exchange and information networks, the tasks and approaches of genebanks and other biodiversity collections have also much in common.

Information on genebank collections belongs to different levels.

- Collection-level data – metadata about genebanks (collection-holding institutions) and the germplasm collections they hold (e.g., species coverage, country coverage). Such information is available from the FAO WIEWS<sup>2</sup> database (World Information and Early Warning System) and the “Bioversity Directory of Germplasm Collections”. In the field of biodiversity collections, similar directories exist, e.g., for herbaria.
- Accession-level data – information about individual accessions, especially passport data. Most PGR databases on various federation levels are based on accession-level data, such as individual genebanks (e.g., IPK’s GBIS, NGB’s SESTO, USDA-GRIN), central crop databases, regional or international “multi-crop” information systems (e.g., EURISCO, SINGER). In biodiversity informatics, this is called ‘unit-level data’.
- Taxon-level data are related to (plant) species (or other taxa) as a whole, e.g., distribution information, uses, taxonomic classification. Examples of such databases related to crop plants are the taxonomic core of GRIN (U.S. Germplasm Resources Information Network) and Mansfeld’s World Database of Agricultural and Horticultural Crops.

Here we will focus on accession-level data, especially on passport data.

---

<sup>2</sup> Links to websites of cited information systems and online documents are provided in the Appendix.

### **Information networking and standards in plant genetic resources**

Beginning from the 1950s, the FAO initiated discussions about a global inventory of plant genetic resources conserved in genebanks. In 1972, the Technical Advisory Committee (TAC) formed by a group of scientists of the FAO and the Consultative Group on International Agricultural Research (CGIAR, founded in 1971) recommended to set up an international network of centres for PGR conservation and requested the development of an international information system. The coordination of both activities belonged to the mandate of the International Board for Plant Genetic Resources (IBPGR) founded in 1974 (IBPGR 1979, Knüpffer 1983). In the aftermath, IBPGR and its successor, the International Plant Genetic Resources Institute (IPGRI), which recently became Bioversity International, developed and published descriptor lists for more than 90 crops (species or genera).

On request of the participants of a workshop on documentation in 1995 (van Hintum et al 1995, p 4; van Hintum and Jongen 1995), IPGRI took the lead in developing a descriptor list for passport data overarching all crops, which was first published in 1997 (Hazekamp 1997) after discussions at an ECP/GR<sup>3</sup> documentation workshop in Budapest in 1996 (Lipman et al 1997, p 5ff; Hazekamp et al 1997) and an intensive e-mail exchange. The revised version of the so-called Multi-Crop Passport Descriptors – MCPD (Alercia et al 2001) was designed to standardise passport data across crops. It was considered “a reference tool developed jointly by IPGRI and FAO to provide international standards to facilitate germplasm passport information exchange across crops. These descriptors aim to be compatible with IPGRI crop descriptor lists and with the descriptors used for the FAO World Information and Early Warning System (WIEWS) on plant genetic resources” (Bioversity International 2001). Many of the ‘Central Crop Databases’ (CCDBs, European or international centralised PGR databases for particular crops, mostly developed in the frame of ECPGR) soon adopted the MCPD. The European PGR Search Catalogue EURISCO hosted by Bioversity International is also based on the MCPD, adding a few descriptors, among them a National Inventory Code.

In the context of the Generation Challenge Programme (GCP), a research and capacity building network that uses plant genetic diversity to produce better crop varieties for resource-poor farmers, the GCP Passport data exchange schema was developed. It is being developed based on the MCPD and the ABCD data standard of the TDWG. While the MCPD was too limited in scope, the ABCD standard with more than 1200 concepts (descriptors) was considered too extensive. Furthermore, some of the ABCD concepts were renamed to better comply with the terminology familiar to the genetic resources community. The GCP Passport standard is included in the BioCASE data exchange software.

From 1978 till 1990, the “Scientific-Technical Board of the COMECON<sup>4</sup> Member Countries for Collections of Wild and Cultivated Plant Species” coordinating the PGR activities of the East European countries, maintained a working group on PGR documentation which met every 2-3 years. This group was formed to coordinate the development of genebank documentation systems in the former COMECON countries that started already in the end of the 1960s in Czechoslovakia (Bareš and Škorpík 1969). During its last meeting in 1988, it established a standard list of passport descriptors (Rogalewicz et al 1988) in accordance with the passport section of IBPGR descriptor lists published in 1977-1988, to be used in the compilation of a planned “International PGR Database of the COMECON countries”. Like the MCPD (with the above-mentioned EURISCO extension), the COMECON list included additional descriptors needed in federated databases to distinguish between contributions from different countries and genebanks. Additionally, a few descriptors dealing with groups of

---

<sup>3</sup> at that time European Cooperative Programme for Crop Genetic Resources Networks

<sup>4</sup> COMECON – The Council for Mutual Economic Assistance (COMECON / CMEA), 1949 – 1991, was an economic organisation of communist states and a kind of Eastern European equivalent to the European Economic Community [after <http://en.wikipedia.org/wiki/Comecon>]

duplicate accessions were proposed, mostly based on earlier experiences with the CCDBs of rye (since 1980, Radzików, Poland) and barley (Gatersleben, Germany).

### **Information networking and standards in biodiversity collections**

Efforts to standardise data of biological collections, especially plant collections (e.g., herbaria and botanical gardens), started also before the 1970s, but almost independently from the parallel developments in the plant genetic resources community. An early landmark publication was the conference proceedings volume “Computers in Botanical Collections” (Brenan et al 1975), followed by volumes such as “Databases in Systematics” (Allkin and Bisby 1984). This community formed the Taxonomic Databases Working Group (TDWG) which met for the first time in 1985 and became a Commission of the International Union for Biological Sciences (IUBS) in 1988. The main aims were: to provide an international forum for biological data projects, to develop and promote the use of standards, and to facilitate data exchange. The ambitious goal of the 1990 annual TDWG conference was reflected in the title of its proceedings “Designs for a Global Plant Species Information System” (Bisby et al 1993). It was only in 1994 when TDWG’s scope and membership was widened to also represent geology, zoology and microbiology. After several name changes and reorganisations it is now called ‘Biodiversity Information Standards (TDWG)’.

For the exchange of unit-level data about biological collections, several data standards were developed, among them

- HISPID3 for herbaria (Conn 1996, first published 1989),
- ‘DarwinCore 2’ – a “flat” structure containing ca. 40 descriptors, initially designed for zoological collections (first published in 2002);
- and the extensive, hierarchically structured ‘Access to Biological Collections Data’ (ABCD version 2.06), “an evolving comprehensive standard for the access to and exchange of data about [biological] specimens and observations (i.e. primary biodiversity data)” with ca. 1,200 concepts (development started in 2000).

The Global Biodiversity Information Facility (GBIF), a Megascience project of the OECD (Organisation for Economic Co-operation and Development) was established in 2001. Its main aim is “to make the world’s primary data on biodiversity freely and universally available via the Internet”. In 2003-2004, GBIF made available its first tools for providing data from biodiversity collections to a biodiversity information network, using web services and based on data exchange standards developed within TDWG, such as DarwinCore and ABCD.

### **Biodiversity informatics data exchange tools**

The biodiversity information network established by GBIF is based on a distributed network of data providers and data portals for accessing their data. It is retrieving structured data from multiple, distributed, heterogeneous databases across the Internet. Data providers have to register themselves in a central registry, which can be queried by portals in order to find relevant datasets to be included in a search.

There are several unit-level data provider communication protocols, namely

- DiGIR (Distributed Generic Information Retrieval) to be used in connection with the DarwinCore 2 data exchange format,
- BioCASE (Biological Collection Access Services for Europe; Güntsch 2006; Güntsch et al 2006; Güntsch and Mergen in press) to be used with several data exchange formats such as ABCD, DarwinCore 2, SPICE, MCPD, GCP Passport, etc.,
- TAPIR – TDWG Access Protocol for Information Retrieval, an integration of the DiGIR and BioCASE data exchange protocols.

DiGIR and BioCASE are presently supported by GBIF. TAPIR is still under development. The new TAPIR data exchange protocol is currently supported by the TapirLink (based on DiGIR) and the PyWrapper version 3 (based on BioCASE).

There are several portals under development that can be used to query unit-level data.

- GBIF's own portal allows searching for taxa (scientific names) and countries. The new GBIF portal under development has four main entry points allowing to explore the dynamic data index via taxonomy (1), countries (2), datasets (3) and occurrences (4) with a wide variety of search options based on all concepts indexed. Both portals can display georeferenced units on interactive maps.
- A generic, individually configurable portal for BioCASE providers is being developed by the Botanical Garden and Botanical Museum Berlin-Dahlem (BGBM).
- In the frame of the Generation Challenge Programme, a PGR-specific central registry for BioCASE data providers was developed, providing access to the source datasets.
- Bioversity International and the Nordic Genebank are also developing a generic portal for PGR data providers.

### **Making Biodiversity and PGR data exchange standards compatible**

Although it was early recognised that plant genetic resources collections form part of the global biodiversity collections (cf. Knüpffer 1983), it was rather rare that members of the PGR community articulated, at biodiversity informatics meetings, the need for cooperation and coordination of work (e.g., Chauvet 1993; Cubero 1993; Hazekamp and Guarino 1999), or vice versa (e.g., Walter 1995).

When GBIF became operational, and more and more international biodiversity collections became GBIF data providers with millions or thousands of individual unit-level records (either using the simple DarwinCore 2 with the DiGIR protocol, or ABCD with the BioCASE protocol), this was an impetus for genebanks to utilise this new technology for their own purposes (Weise et al 2003), but also to make their collections visible within a broader biodiversity context. For this purpose, the MCPD was completely mapped to ABCD in 2004-2005<sup>5</sup>, after extending ABCD by a PGR-specific sub-unit<sup>6</sup>.

### **Genebanks as GBIF data providers**

So far, most of the federated PGR databases – the ECPGR central crop databases as well as EURISCO and SINGER – have been created and updated by data either sent from the contributors to the central database manager, or via the National Inventories through an online upload tool (as in the case of EURISCO). All these processes involve repeated transformations of data into a target format (cf. Knüpffer 1995). Data providers contributing to several such central databases may need to transform their data into different formats. Even standardised formats such as the MCPD may be interpreted differently. This situation led to a high workload on the side of the providers, but especially on the side of the compilers of centralised databases.

Using web services technology and in particular the tools promoted by GBIF, these tasks can be made easier. Each contributor first needs to establish a data provider software. The local database schema then has to be mapped to agreed standards or schemas only once, and the managers of central databases can harvest data from all providers using the same technology.

---

<sup>5</sup> see <http://www.bgbm.org/TDWG/CODATA/Schema/Mappings/EURISCO-2-ABCD.pdf>

<sup>6</sup> The newly included PGR-specific ABCD concepts from MCPD were: National Inventory Code, Breeding Institute Code, Biological Status of Accession, Ancestral Data, Collecting/Acquisition Source, Other Identification (Number) associated with the accession, Location of Safety Duplicates, Type of Germplasm Storage, Decoded Breeding Institute, Decoded Donor Institute, and Decoded Safety Duplication Location.

At the same time, data sets updated by individual database managers are immediately available to users worldwide.

In 2004, three European genebanks became GBIF data providers (Knüpffer et al. 2004; Endresen et al 2006): the Nordic Genebank, the Polish genebank (IHAR) and IPK (cf. Table 1). This served as a demonstration project for wider application of GBIF's web service technology for genebanks.

In May 2005, Bioversity International decided to support the application and further development of this technology, and started to develop customized tools for PGR collections in cooperation with NGB. Following a demonstration and training workshop in June 2005, most CGIAR genebanks and several national genebanks, including the USDA-GRIN data portal, installed the BioCASE data provider software, and the technology was adopted by the Generation Challenge Programme. Training workshops in the use of the GBIF technology combined with a series of on-site visits permitted in less than a year to link more than 15 genebank providers and information networks. In 2006, both EURISCO and SINGER were included in the network by Bioversity International. This development is considered a great step toward the implementation of a global accession level information system under the leadership of Bioversity.

### **Global Unique Identifiers**

In distributed information networks on genebank accessions or other objects, a system to uniquely identify these objects is of critical importance. EURISCO and SINGER are presently using a combination of the fields Accession Number, Holding Institute and Collection Code (sometimes Genus name) for this purpose. When combining or sharing unit-level data across different thematic networks and domains of biodiversity – each domain with different protocols for unique identification of the units –, a more generic, global standard for uniquely identifying the data units becomes crucial. Furthermore, in dynamic distributed networks built on web services, the protocol for unique identification of the data units needs to be adopted for automatic discovery by these web service protocols. A global, generic, and friendly protocol for unique identification of data units or accessions is thus needed.

In 2006, the TDWG (Bioversity Informatics Standards) dedicated two workshops to this question, and decided to explore the suitability of the LSID (Life Science Identifier; Clark et al 2004) protocol as the recommended GUID (Global Unique Identifier) for biodiversity informatics. The Generation Challenge Programme had made a similar decision already in 2005.

The LSID framework has been developed by the Interoperable Informatics Infrastructure Consortium (I3C) and Object Management Group/Life Sciences Research (OMG/LSR) and was first implemented by a research team at IBM in 2003 (Atev and Szekely, 2003). The LSID framework provides more than just a protocol for uniquely naming biodiversity resources: the LSID protocol defines a simple, common way to identify and not least also to access the data unit itself. This LSID resolution service provides a protocol to access metadata about, as well as methods to retrieve the data unit itself. LSIDs are location-independent and actionable. LSID conforms to the standard URN (Uniform Resource Name) specification of the Internet Engineering Task Force (IETF). Every LSID has up to five parts separated by a colon: the network identifier ("LSID"), the root DNS name of the issuing authority, a namespace chosen by the issuing authority, an object identifier unique in that namespace, and an optional revision identifier. Examples of the LSID format are:

- [urn:lsid:ncbi.nlm.nih.gov:pubmed:12571434] (identifier for a PubMed article),
- [urn:lsid:ubio.org:namebank:11815] (identifier for a taxon name concept).

TDWG has developed a useful web resolver for LSIDs, <http://lsid.tdwg.org/> (example of usage: <http://lsid.tdwg.org/urn:lsid:authority:namespace:object>).

## Conclusions and outlook

The achieved compatibility of data standards between PGR and biodiversity collections allows integrating the worldwide germplasm collections into biodiversity information networks. Using GBIF technology (and contributing to its development), the PGR community can easily establish specific PGR information networks without creating its own technology.

Global, regional or crop-specific PGR information systems will be developed on the basis of GBIF-compatible web services, thus avoiding manual physical data transfer from providers to the central data index or portal. Data providers will establish web services; a central registry will contain technical metadata about these web service end points; a central data portal will index metadata about provided datasets and collections (e.g., contact information, lists of taxa and countries covered) and often also a cached index of selected attributes for the individual data records provided (accession numbers, occurrence data, etc.). General or specialised portals will be developed for searching information simultaneously from heterogeneous sources and displaying results in a unified manner.

To achieve these aims, training and capacity building, as well as a helpdesk will be required for data providers. There are GBIF training courses (check out for “How to become a GBIF data provider” on the GBIF website).

In addition, the data provider software needs to be customised to the requirements of the PGR community. For example, when mapping a provider’s database to the ABCD standard with its >1,200 concepts, only a small, PGR-specific proportion of ABCD should be displayed for PGR collections. For example, the GCP Passport standard uses a terminology familiar to the PGR community in defining a PGR subset of ABCD. Another possible solution is foreseen in TAPIR, where concept identifiers can be abbreviated by using Concept Name Servers (CNS). The CNS allows to develop centrally maintained concept lists to be connected with concise, well established (in the context of a particular community) alias names<sup>7</sup>. Thus, a potential provider from a particular biodiversity sub-community would not need to know the complete ABCD schema, but just map his concepts to the sub-schema relevant to his community.

Users from the biodiversity community (who may not be aware of the existence of relevant material in genebanks) will find in GBIF genebank material of, e.g., crop wild relatives, along with data about the same species from herbaria, botanical gardens and floristic observations.

Using GBIF technology, the PGR community can reach various project goals such as:

- improvement of data flow and accessibility for centralised PGR information systems, e.g., EURISCO (and its National Inventories), SINGER, CCDBs;
- establishment of eight “Global Crop Registries” for major crops of the CG and other genebanks worldwide in the frame of the GPG2 (Global Public Goods) project of the CG centres; 13 more crops to follow;
- establishment of a “One-stop shop” for requesting germplasm from CG centres’ genebanks and from other genebank partners;
- establishment of SMTA (Standard Material Transfer Agreement) reporting mechanisms linked to a global accession level information system in support of the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA).

Recently, Bioversity International has undertaken a feasibility study for a global accession-level information system by linking these online resources. It was demonstrated that information on more than 2.3 million accessions can be compiled into a dynamic index and made available through a central portal.

---

<sup>7</sup> for a detailed explanation, see

[http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/TAPIRSpecification\\_2007-02-07.html#toc21](http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/TAPIRSpecification_2007-02-07.html#toc21)

The recently started development of a “Bioversity global accession level information system” by a grant from the Global Crop Diversity Trust (mid April 2007) will also highly benefit from the developments described above. This system will entirely be based on GBIF technology. It aims to cover 80% of the worldwide genebank accessions. Bioversity International intends to connect 60 more genebanks with GBIF technology within the next four years.

### Acknowledgements

The mapping of MCPD to ABCD was carried out in collaboration with W. Berendsohn and J. de la Torre and their colleagues from the Botanical Garden and Botanical Museum Berlin-Dahlem, Germany. We would like to thank J. Holetschek (Berlin-Dahlem), M. Oppermann, R. Narang and W. Schölch (Gatersleben) for useful comments to improve the manuscript. The Generation Challenge Programme supported the work of (DTFE and SG and colleagues at) Bioversity International to develop and install the BioCASE web service protocol for the CGIAR centres. The contribution of IF was supported by the Research Project number 0002700602, Ministry of Agriculture of the Czech Republic.

### Appendix. Web links to documents, databases and information systems<sup>8</sup>

ABCD – Access to Biological Collections Data. <http://www.bgbm.org/TDWG/CODATA/Schema/>;  
<http://wiki.tdwg.org/twiki/bin/view/ABCD/WebHome>  
 BioCASE – Biological Collection Access Services. <http://www.biocase.org/>  
 BioCASE portal. [http://www.biocase.org/products/portals/biocase\\_2/index.shtml](http://www.biocase.org/products/portals/biocase_2/index.shtml)  
 Bioversity International. Bioversity Directory of Germplasm Collections.  
[http://www.bioversityinternational.org/Information\\_Sources/Germplasm\\_Databases/Germplasm\\_Collection\\_Directory/index.asp](http://www.bioversityinternational.org/Information_Sources/Germplasm_Databases/Germplasm_Collection_Directory/index.asp)  
 DarwinCore <http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome>  
 DiGIR – Distributed Generic Information Retrieval. <http://digir.net>  
 EURISCO Multi-crop passport descriptors.  
[http://www.ecpgr.cgiar.org/epgris/Tech\\_papers/EURISCO\\_Descriptors.doc](http://www.ecpgr.cgiar.org/epgris/Tech_papers/EURISCO_Descriptors.doc)  
 EURISCO – European search catalogue for plant genetic resources. <http://eurisco.ecpgr.org/>  
 FAO WIEWS – World Information and Early Warning System on PGRFA. <http://apps3.fao.org/wiews/>  
 GBIF – Global Biodiversity Information Facility. <http://www.gbif.org>  
 GBIF data portals: <http://www.gbif.net>; <http://newportal.gbif.org/> (new portal, under development)  
 GBIS – Genebank Information System of IPK Gatersleben. <http://gbis.ipk-gatersleben.de>  
 Generation Challenge Programme (GCP) “Cultivating Plant Diversity for the Resource Poor”, Bioinformatics Subprogramme. <http://www.generationcp.org/subprogramme4.php>  
 GRIN – Germplasm Resources Information Network (US). Taxonomic core.  
<http://www.ars-grin.gov/cgi-bin/npgs/html/queries.pl>  
 GUID – Global Unique Identifier – The GUID working group of the TDWG.  
<http://wiki.tdwg.org/twiki/bin/view/GUID/WebHome>  
 IETF – Internet Engineering Task Force. <http://www.ietf.org/>  
 LSID – Life Science Identifier. The LSID resolution project. <http://lsid.sourceforge.net/>;  
<http://www.ibm.com/developerworks/opensource/library/os-lsid/>; <http://xml.coverpages.org/lsid.html>  
 LSID web resolver, http proxy to retrieve metadata or data from a LSID string. <http://lsid.tdwg.org/>  
 Mansfeld’s World Database of Agricultural and Horticultural Crops. <http://mansfeld.ipk-gatersleben.de>  
 MCPD – Multicrop passport descriptors.  
[http://www.bioversityinternational.org/publications/pubfile.asp?id\\_pub=124](http://www.bioversityinternational.org/publications/pubfile.asp?id_pub=124)  
 SESTO, Nordic Gene Bank regional data portal for the Northern Europe. <http://www.nordgen.org/sesto/>  
 SINGER – System-wide Information Network for Genetic Resources. <http://singer.grinfo.net/>  
 TAPIR – TDWG Access Protocol for Information Retrieval. <http://www.tdwg.org/activities/tapir/>  
 TDWG – Biodiversity Information Standards. <http://www.tdwg.org>

### References

Alercia A, Diulgheroff S, Metz T (2001) FAO/IPGRI Multi-Crop Passport Descriptors.

<sup>8</sup> All links listed here have been checked in June 2007.



- Allkin R, Bisby FA (1984) Databases in Systematics. Acad. Press London etc. 329 pp
- Atev S, Szekely B (2003) Build an LSID Resolution Service using the Java language. A Java-based Life Sciences Identifier authority consolidates biological data resources. From IBM developerWorks. May 27, 2003, updated March 03, 2004. Available from <http://www-106.ibm.com/developerworks/opensource/library/os-lsid/> (accessed June 2007)
- Bareš I, Škorpík M (1969) Příspěvek k možnosti využití samočinného počítače Minsk 22 ke třídění a hodnocení kolekce odrůd kulturních rostlin. Vedecké práce VÚRV Praha-Ruzyně 14, 125-132
- Bioversity International (2001) Multicrop passport descriptors. Available from [http://www.biodiversityinternational.org/publications/pubfile.asp?id\\_pub=124](http://www.biodiversityinternational.org/publications/pubfile.asp?id_pub=124) (accessed June 2007)
- Bisby FA, Russell GF, Pankhurst RJ (eds) (1993) Designs for a Global Plant Species Information System. Clarendon Press. 346 pp
- Brenan JPM, Ross R, Williams JT (eds) (1975) Computers in Botanical Collections. Plenum Press, London & New York, 204 pp
- Chauvet M (1993) The need for information on genetic resources. In: Bisby FA, Russell GF, Pankhurst RJ (eds) Designs for a Global Plant Species Information System. Clarendon Press. pp. 55-61
- Clark T, Martin S, Liefeld T (2004) Globally distributed object identification for biological knowledgebases. Briefings in Bioinformatics 5 (1), 59-70
- Conn BJ (ed) (1996) Herbarium Information Standards and Protocols for Interchange of Data, version 3. Available from <http://plantnet.rbgsyd.nsw.gov.au/HISCOM/HISPID/HISPID3/H3.html> (Accessed June 2007)
- Cubero JI (1993) Plant breeding and resource information. In: Bisby FA, Russell GF, Pankhurst RJ (eds) Designs for a Global Plant Species Information System. Clarendon Press. pp. 61-68
- Endresen DTF, Bäckman J, Knüpfner H, Gaiji S (2006) Exchange of Germplasm Datasets with PyWrapper/BioCASE. In: Proceedings of TDWG 2006. TDWG annual conference, Missouri Botanical Garden, St. Louis, USA, October 15 to 22, 2006. [<http://www.tdwg.org/proceedings/article/view/64>]
- FAO (1996) Report on the state of the world's plant genetic resources for food and agriculture. Food and Agriculture Organization of the United Nations, Rome
- Güntsch A (2006) Integration of heterogeneous biodiversity data sources using BioCASE and ABCD. NoE Marine Genomics Europe – Exploratory Workshop: Marine Genomics Meets Marine Diversity, Bremen 2006, Abstract
- Güntsch A, Döring M., Berendsohn WG (2006) Mobilisierung von primären Biodiversitätsdaten: Das BioCASE Protokoll und seine Anwendung in internationalen Netzwerken. In: Knetsch G (ed) Umweltdatenbanken und Netzwerke. Dessau, Germany. pp 129-138
- Güntsch A, Mergen P (in press) The BioCASE Project – a biological collections access service for Europe. 1<sup>st</sup> International Recorder Conference, Luxembourg 2005
- Hazekamp T (1997) Appendix II. Multicrop passport descriptors (final version). In: Lipman E, Jongen MWM, van Hintum TJL, Gass T, Maggioni L (compilers) Central crop databases: tools for plant genetic resources management. International Plant Genetic Resources Institute, Rome, Italy/CGN, Wageningen, The Netherlands, pp 75-78
- Hazekamp T, Guarino L (1999) Taxonomy and the conservation and use of plant genetic resources. In: Andrews S, Leslie A, Alexander C (eds) Taxonomy of Cultivated Plants. Royal Bot Gardens Kew. pp 11-17
- Hazekamp T, Serwinski J, Alercia A (1997) Multicrop passport descriptors. In: Lipman E, Jongen MWM, van Hintum TJL, Gass T, Maggioni L (compilers) Central crop databases: tools for plant genetic resources management. International Plant Genetic Resources Institute, Rome, Italy/CGN, Wageningen, The Netherlands, pp 35-39
- IBPGR 1979: A review of policies and activities 1974-78 and of the prospects for the future. IBPGR Secretariat, Rome, AGPE: IBPGR/78/17, 56 pp
- Knüpfner H (1983) Computer in Genbanken — eine Übersicht. Kulturpflanze 31, 77-143
- Knüpfner H (1995) Central crop databases. In: van Hintum TJL, Jongen MWM and Hazekamp T (eds) Standardization in plant genetic resources documentation. Centre for Genetic Resources The Netherlands, Wageningen. pp 51-62
- Knüpfner H, Biermann N, Endresen DT, Kolasinski P, Podyma W, de la Torre J (2004) Genebanks as GBIF data providers – the first experiences. In proceedings of TDWG 2004. Taxonomic Databases Working Group (TDWG) annual conference, Christchurch, New Zealand, October 10 to 17, 2004. [[http://www.nhm.ac.uk/hosted\\_sites/tdwg/2004meet/paperabstracts/TDWG\\_2004\\_Papers\\_Knupffer\\_1.htm](http://www.nhm.ac.uk/hosted_sites/tdwg/2004meet/paperabstracts/TDWG_2004_Papers_Knupffer_1.htm)]
- Lipman E, Jongen MWM, van Hintum TJL, Gass T, Maggioni L (compilers) (1997) Central crop databases: tools for plant genetic resources management. International Plant Genetic Resources Institute, Rome, Italy/CGN, Wageningen, The Netherlands

- Rogalewicz V (ed), Knüpfner H, Korneychuk VA, Lozanov IG, Plotnikov DB, Serwiński J, Shvytov IA (1988) Passport descriptors of the COMECON International Database of Genetic Resources (Russian, English). Výzkumný ústav rostlinné výroby, Praha-Ruzyně. 26 pp.
- van Hintum TJL, Jongen MWM (1995) Standardization in plant genetic resources documentation II: Passport data. In: van Hintum TJL, Jongen MWM, Hazekamp T (eds) Standardization in plant genetic resources documentation. Centre for Genetic Resources The Netherlands, Wageningen. pp 77-88.
- van Hintum TJL, Jongen MWM, Hazekamp T (eds) (1995) Standardization in plant genetic resources documentation. Centre for Genetic Resources The Netherlands, Wageningen
- Walter K (1995) Standardization of documentation in botanic gardens. In: van Hintum TJL, Jongen MWM, Hazekamp T (eds) Standardization in plant genetic resources documentation. Centre for Genetic Resources The Netherlands, Wageningen. pp 119-124
- Weise S, Biermann N, Flemming S, Vorwald J, van Hintum TJL, Knüpfner H, Grosse I. (2003) Use of XML for exchange of data on plant genetic resources. Poster, Annual Meeting of TDWG, Oeiras, Portugal, 21-26 October 2003

Table 1. Genebanks and information systems as GBIF data providers (as of October 2007). All genebanks are BioCASE providers, except for Poland (DiGIR provider). Records from the four European genebanks are also provided via EURISCO.

<b>Genebank, provider</b>	<b>records</b>	<b>georeferenced</b>	<b>year<sup>1</sup></b>
EURISCO <sup>2</sup>	1,304,245	89,579	2006
SINGER <sup>2</sup>	553,842	178,184	2006
US National Plant Germplasm System	283,815	24,103	2005
IPK Gatersleben, Germany	143,669	7,420	2004 <sup>3</sup>
National Centre for Plant Genetic Resources, Poland	42,980	0	2004
Nordic Gene Bank, Alnarp, Sweden	31,274	3,890	2004
Centre for Genetic Resources, Wageningen, The Netherlands	23,700	5,409	2006

<sup>1</sup> year when the provider was initially connected to GBIF

<sup>2</sup> via Bioversity International

<sup>3</sup> updated June 2007