THE DARWIN CORE EXTENSION FOR GENE BANKS OPENS UP NEW OPPORTUNITIES FOR SHARING GERMLASM DATA SETS

DAG T.F. ENDESEN*
Global Biodiversity Information Facility (GBIF), Copenhagen, Denmark

AND

HELMUT KNÜPFER
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

Abstract. – Darwin Core (DwC) defines a standard set of terms to describe the primary biodiversity data. Primary biodiversity data are data records derived from direct observation of species occurrences in nature or describing specimens in biological collections. The Darwin Core terms can be seen as an extension to the standard Dublin Core metadata terms. The new Darwin Core extension for gene banks declares the additional terms required for describing genebank data sets, and is based on established standards from the plant genetic resources community. The Global Biodiversity Information Facility (GBIF) provides an information infrastructure for biodiversity data including a suite of software tools for data publishing, distributed data access, and the capture of biodiversity data. The Darwin Core extension for gene banks is a key component that provides access for the genebanks and the plant genetic resources community to the GBIF informatics infrastructure including the new toolkits for data exchange. This paper provides one of the first examples and guidelines for how to create extensions to the Darwin Core standard.

Keywords. – Darwin Core; Darwin Core extension; GBIF; genebank collections; germplasm; plant genetic resources.

There are more than 1750 gene banks distributed all around the world, with more than 130 large and medium-size gene bank collections holding more than 10,000 accessions each (FAO, 2010). Each of these gene banks maintains living material of plant genetic resources. New accessions are added to the gene bank collections from collecting expeditions and from old cultivars obsolete to the commercial seed trade. The gene bank documentation systems are continuously being extended with new accessions and with updated information on existing accessions.

The International Treaty on Plant Genetic Resources for Food and Agriculture1 (ITPGRFA; FAO, 2009, page 29, Article 17.1) calls for building a global information system on plant genetic resources for food and agriculture. Such a system needs to be frequently refreshed with new and updated information from each gene bank collection (and other information sources such as inventories of crop wild relatives). With modern information technology, a distributed information system can be designed to allow extracting a snapshot of the decentralized gene bank data sets at any time. Moreover, a distributed germplasm information system will make updated germplasm information more easily accessible to plant breeders, crop scientists and other users, thus providing better access also to the plant material. The lack of easy access to germplasm information remains an important bottleneck for utilization of plant genetic resources material (Khoury et al., 2010). The Global Biodiversity Information Facility (GBIF)2 has developed a software toolkit for publishing decentralized biodiversity data sets, known as the “Integrated Publishing Toolkit” (GBIF IPT)3. It provides a successful example of a software tool designed for building a distributed network of biodiversity databases. The new Darwin Core extension for gene banks is required to enable the GBIF IPT to share a standard set of minimum terms for germplasm data sets.

GENEBANKS AND THEIR INFORMATION SYSTEMS

The First Genebank, The Vavilov Institute

The first genebanks for ex situ conservation of plant genetic resources were established more than one century ago, well before the advent of the digital computer. In 1894, Professor A.F. Batalin, Director of the Sankt Petersburg Botanical Garden, made the initiative to organize the Bureau of Applied Botany under the Scientific Committee of the Russian

---

* Corresponding author; email: dag.endresen@gmail.com.
1 http://www.planttreaty.org/.
2 http://www.gbif.org/.
Ministry of Agriculture. During 1901 and 1902, requests were distributed throughout the Russian provinces to collect and return seeds of local cultivars (landraces) of agricultural crops. In 1908 the institute organized a first expedition to collect Russian landraces (Regel, 1915; cf. Loskutov, 1999). Nikolai Ivanovich Vavilov (1887 –1943) joined the institute in 1910 and became its director in 1920. Under his leadership, the mandate of the institute was expanded to include the long-term conservation of plant genetic resources. This marks the advent of the modern genebanks, as we know them today (Loskutov, 1999). The Bureau of Applied Botany became the present N.I. Vavilov Research Institute of Plant Industry (VIR)⁴.

The First Genebank Information System, Index Seminum

The seed banks of botanical gardens and their seed exchange system can be seen as a predecessor to the present genebanks. Around 1543, the first botanical gardens in Europe were established in Italy (Stafleu, 1969; Stearn, 1971). The botanical gardens have traditionally published seed lists (Index Seminum) for the purpose of seed exchange (Heywood, 1964). However, the seed exchange of the botanical gardens has been criticized for problems with inaccurate classification, poor viability and the lack of information on the origin of the seeds (Thompson, 1970). The aforementioned Bureau of Applied Botany in Russia included, from its start in 1894, an information department with the task to provide information on the availability of seed from both cultivated (domesticated) and wild species (Loskutov, 1999). The last seed catalog (Delectus Seminum – list of selected seeds) of VIR was published in 1999 (Dragavtsev et al., 1999). The crop departments continued, however, to print more detailed crop-based “catalogues” after the last Delectus Seminum (see for example Loskutov and Ryabchenko, 2002).

The First Electronic Genebank Information Systems

The Fifth Yugoslav Symposium on Research in Wheat in 1966 included one of the first initiatives to develop international standards and mechanisms for sharing electronic documentation of crop genetic resources (Konzak and Sigurbjörnsson, 1966). A group of experts assembled by the Food and Agriculture Organization of the United Nations (FAO) ⁵ and the International Atomic Energy Agency (IAEA)⁶ in Vienna proposed to establish a distributed network with national crop information centers reporting national crop data to a central hub to be set up at FAO in Rome, Italy. The central file maintained and hosted at FAO would be published to become available for plant breeders, crop researchers and policy organs (Finlay and Konzak, 1970). When the International Board for Plant Genetic Resources (IBPGR) was established in 1974, one of its first tasks was to organize and coordinate such a distributed network of genebank information systems. Work was initiated on a distributed system under the name Genetic Resources Communication, Information and Documentation System (GR/CIDS; IBPGR, 1976, 1977).

The central file at FAO will fulfill two main roles. First, by accepting records from all holders of collections willing to exchange seed and, by adding information on new material, it will become a current record of stocks available throughout the world. Second, by also accumulating information about material which has severely restricted seed supply or for which seed is not available, the central file will become an archive for records of genetic variation (Finlay and Konzak, 1970:463–464).

When it is fully developed, GR/CIDS should encompass the whole of the information component, including documentation and the flow of information, of genetic resources work, from the initial collection of data about traditional materials in the field to the performance of improved varieties derived from them (IBPGR, 1976:5).

Its development and the sharing of phenotypic characterization and evaluation data remain a high priority for a rational utilization of plant genetic resources (FAO, 2010; Khoury et al., 2010; Ayling et al., 2012).

The first SoW [State of the World] report highlighted the poor documentation availability on most of the world’s ex situ PGR. This problem continues to be a substantial obstacle to the increased use of PGRFA in crop improvement and research. Where documentation and characterization data do exist, there are frequent problems in standardization and accessibility, even for basic passport information (FAO, 2010:77).

---

⁴ http://www.vir.nw.ru/
⁵ http://www.fao.org/
⁶ http://www.iaea.org/
The GENESYS Gateway to Genetic Resources was released in May 2011, including more than 2.3 million genebank accessions, which main data sets included EURISCO, SINGER and the USDA GRIN, (Figure 1) out of the estimated 7.4 million accessions existing worldwide (FAO, 2010). GENESYS provides the first global accession-based information system for genetic resources (Hershey, 2011). This was a major step forward, but important work still remains to build a global federated infrastructure for germplasm data sets to allow for the efficient and dynamic update of the GENESYS portal. The Darwin Core extension for genebanks is here proposed as a tool for building an efficient network of genebank data sets to provide updated information into the GENESYS portal.

**CROP DESCRIPTOR LISTS**

Bioversity International has developed and published more than 100 crop-specific descriptor lists since 1977 (Gotor et al., 2008). The first was for cultivated potato (IBPGR, 1977) followed by the descriptors for wheat and Aegilops (IBPGR, 1978). These crop descriptor lists provided a valuable standard for documentation of plant genetic resources and in particular for their phenotypic characterization and evaluation data (Gotor et al., 2008). The Multi-Crop Passport Descriptors (MCPD) were introduced in 1996 (Hazekamp et al., 1997), updated in 2001 (FAO/IPGRI 2001) and published in their current format in June 2012 (FAO/Bioversity, 2012). In the context of the EURISCO search portal for genebanks in Europe, a few amendments were made to the MCPD including descriptive names of institutes (in addition to standardized acronyms) and an URL for linking to additional accession-based information. The EURISCO amendments included also the status of the accession in the multilateral system (MLS) of the ITPGRFA and the status in the European Genebank Integrated System (AEGIS; EURISCO, 2012). Genebanks have tried to follow these standard crop descriptors in projects to describe genebank collections. This ensured good interoperability between the genebank data sets from different institutes and countries. The Darwin Core extension for genebanks is derived from the MCPD standard.

In former Eastern Bloc countries of the Council for Mutual Economic Assistance, (COMECON or CMEA, 1949–1991), “unified” crop descriptors ensured similar standardization and interoperability of germplasm data sets. The first COMECON crop descriptor lists were released in 1974 for Triticum (Bareš, 1974), Hordeum and Avena (for an overview, see Knüpffer, 1983). Their predecessors have been national crop descriptor lists of the USSR and Czechoslovakia since the 1960s (cf. Knüpffer, 1983). A first standard for passport data recording across genebanks has been proposed by the COMECON working-group for documentation of plant genetic resources (Rogalewicz, 1988). The COMECON passport descriptor list is similar in aim and coverage as the later MCPD. The crop descriptor lists from Bioversity International and the COMECON have contributed to acceptable data interoperability between the distributed genebank data sets across the world and make the present
development of an automatic data exchange mechanism for germplasm data easier.

**European Central Crop Databases**

During the 1980s and 1990s, a number of European Central Crop Databases (ECCDB) was developed as part of the European Cooperative Programme for Plant Genetic Resources (ECPGR)\(^{11}\) (Knüpffer, 1995; Lipman et al., 1997). The first such database was that of rye, developed by the Polish genebank at the Plant Breeding and Acclimatization Institute (IHAR). As the first of its kind (initiated in September 1981 at a joint meeting between the Polish gene bank and the Nordic Gene Bank), the rye catalogue comprised passport data of rye accessions maintained in 11 genetic resources centers (Serwiński and Konopka, 1984). This pioneer work was used as a reference as well as a model for other European databases (Podyma, 2001). The ECCDBs contributed to European collaboration and joint project activities for plant genetic resources. The ECCDBs have also contributed to the mobilization of some phenotypic characterization and evaluation data, but to a much smaller extent than presumed (Maggioni, 2007). The Central Crop Databases in Europe provide a distributed network of crop experts and publish an aggregated database with regular updates of data from the genebanks holding accessions of the respective crops. These crop networks would greatly benefit from a more standardized and automatic data exchange mechanism.

**EURISCO – European Search Catalogue for Plant Genetic Resources**

In September 2003, EURISCO\(^{12}\) was released as a searchable online database with passport data from many European genebank collections (EURISCO, 2003; IPGRI, 2003). EURISCO was developed during 2000–2003 with funding from the EU 5th framework programme (IPGRI, 2001; 2002). EURISCO is hosted by Bioversity International from the headquarters in Rome and is regularly updated by the designated national focal points representing almost all European countries (EURISCO, 2002). The EURISCO framework has been proposed as a model for other regions. One example is the presentation of the EURISCO infrastructure as a proposed model for the development of a distributed genebank network in Latin America (Gaiji et al., 2008). The many national and regional genebank institutions in Europe maintain numerous distributed genebank databases. Direct data exchange between these genebanks, the Central Crop Databases and the EURISCO Catalogue (via the respective National Inventories) makes for a complex information infrastructure. There is already today a substantial flow of data through this germplasm information web and many person-hours dedicated to keep the data pathways open (Dias et al., 2012). The development of standardized and automatic data exchange mechanisms for PGR in Europe has been on the agenda of the ECPGR Documentation and Information Network and its predecessors, e.g. the Internet Advisory Group, for many years (Maggioni, 2005; 2010).

**PGR Forum, CWRIS, CWRML**

The EU-funded European Crop Wild Relative Diversity Assessment and Conservation Forum (PGR Forum)\(^{13}\) project (2003–2005) produced a new information system for crop wild relatives (CWRIS)\(^{14}\) and a new XML (extensible markup language) based schema for exchange of data sets on crop wild relatives. The crop wild relative markup language (CWRML)\(^{15}\) was designed for compatibility with the Darwin Core (DwC) and the Access to Biological Collections Data (ABCD) data standards from Biodiversity Information Standards (TDWG)\(^{16}\). It was further envisioned that terms from the CWRML schema could be extracted to form an extension to the Darwin Core standard (Moore et al., 2008).

**Generation Challenge Programme (GCP)**

The GCP\(^{17}\) is a time-bound (2003–2013) initiative of the Consultative Group of International Agricultural Research (CGIAR)\(^{18}\) to use genetic diversity for improving the crops for resource-poor farmers in developing countries. The GCP includes a theme for advanced research on crop information systems and bioinformatics (Bruskiewich et al., 2006; 2008). During 2005, the GCP Passport XML schema\(^{19}\) was harmonized for interoperability with the ABCD schema and processed for implementation with the Biological Collection Access Service for Europe (BioCASE)\(^{20}\) toolkit. The GCP Central Registry\(^{21,22}\) was designed to interact

---

\(^{11}\) http://www.ecpgr.cgiar.org/.
\(^{12}\) http://eurisco.ecpgr.org/.
\(^{13}\) http://parforum.org/.
\(^{14}\) http://www.pgrforum.org/cwris/.
\(^{15}\) http://www.pgrforum.org/CWRML.htm
\(^{16}\) http://www.tdwg.org/.
\(^{17}\) http://www.generationcp.org/.
\(^{18}\) http://www.cgiar.org/.
\(^{19}\) http://gcpcr.grinfo.net/include/webservices/schema-documentation.php.
\(^{20}\) http://www.biocase.org/.
\(^{21}\) http://gcpcr.grinfo.net/.
\(^{22}\) http://cropforge.org/projects/gcpcr/.
with research data sets published by the project partners using the BioCASE data publishing toolkit.

**GCP Crop Ontology**

The International Agricultural Research Centers (IARC) of the CGIAR have developed a Crop Ontology (CO) designed to provide a controlled vocabulary set for some of the economically important crops (Shrestha et al., 2010). The Crop Ontology is based on the Gene Ontology (GO) (Ashburner et al., 2000), Plant Ontology (PO) (Jaiswal et al., 2005), MIAME-Plant (Zimmermann et al., 2006), and the Multi-Crop Passport Descriptors (FAO/IPGRI, 2001).

**GENEBANKS IN THE CONTEXT OF BIODIVERSITY INFORMATION**

**The First Genebanks to Join GBIF**

The Global Biodiversity Information Facility (GBIF) was established in 2001 (GBIF, 2001) as an inter-governmental initiative to facilitate free and open access to biodiversity data online. During 2004, the Nordic Gene Bank (NGB, reorganized in 2008 as the Nordic Genetic Resource Center, NordGen), the Polish genebank in Radzików, and the German genebank in Gatersleben were the first genebanks to join the GBIF network (Knüpffer et al., 2004). GBIF is a distributed biodiversity information network based in part on the information standards defined by the Biodiversity Information Standards organization (TDWG; Taxonomic Databases Working Group, 1985–2006). At the time when the first genebanks joined GBIF, TDWG were developing two alternative biodiversity collection data exchange formats, both potentially suitable for genebank accessions. These were the Darwin Core (DwC, version 2) and the Access to Biological Collections Data (ABCD version 1.20) (Berendsohn, 2005). During 2005 and in collaboration with the ABCD task group, the standard genebank descriptors (MCPD) were mapped to the corresponding ABCD terms, or added as new descriptors to an updated version of the ABCD (version 2.06) (Berendsohn and Knüpffer, 2006). This new development of data interoperability between the crop data sets and the TDWG data-sharing standards opened the possibility for utilization of the GBIF data infrastructure by the PGR community for its own interoperability tasks. It was now possible to start the implementation of the data-sharing toolkits from the GBIF and the TDWG communities in the EURISCO network for Europe (Endresen et al., 2006) (Figure 2). The BioCASE (Berendsohn, 2002) data publishing toolkit was installed at 15 genebanks located in different parts of the world and a demo data portal was developed in 2006 to interact with these distributed web services established by the BioCASE installations.

**Dublin Core Metadata Initiative (DCMI)**

The DCMI was initiated at a joint workshop between the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA) on metadata semantics held in Dublin (Ohio, USA) in March 1995. The output from this workshop was called "Dublin Core metadata" based on the location of the workshop. The original target was to develop a
small, common set of metadata elements to describe Web content. The original Dublin Core elements (or terms) were: Subject, Title, Author, Publisher, OtherAgent, Date, ObjectType, Form, Identifier, Relation, Source, Language, and Coverage. The Dublin Core was designed to be extensible (Weibel et al., 1995).

**Darwin Core**

Natural history museums in the USA started early to develop information networks with distributed query systems using the Internet. The Species Analyst project was initiated in 1997 and coordinated from Kansas University (Viegilas et al., 1998; Peterson et al., 2003). The first version of the Darwin Core list of terms was developed in 1999 by the Species Analyst project (Stein and Wieczorek, 2004). The Mammal Networked Information System (MaNIS) was initiated in 1999 and established in June 2002 a distributed information network between 17 North American mammal natural history collections. The MaNIS network was developed in parallel with the Distributed Generic Information Retrieval (DiGIR) data publishing toolkit and contributed to the development of the next version of the Darwin Core (DwC version 1.21) (Stein and Wieczorek, 2004). The current version of the Darwin Core is more different from the earlier versions than the previous versions from each other, and was ratified and published by TDWG in October 2009. Darwin Core can be seen as an extension to the standard Dublin Core metadata terms and is designed to be extensible. Darwin Core provides stable semantic definitions of terms for sharing information on biological diversity (Darwin Core Task Group, 2009a; Wieczorek et al., 2012).

**Darwin Core Archive (DwC-A)**

The GBIF IPT Task Force introduced the Darwin Core Archive format (Döring et al., 2011). The DwC-A is based on the Darwin Core text guidelines with core entities linked in a one-to-many relationship to records in extensions. The core entities and the entities in each of the extensions are presented as fielded text such as comma separated values (CSV) or tab delimited values (TAB), with one file for the core and one file for each extension. Each record in an extension file is linked to one of the records in the core file. A metafile (meta.xml) describes the structure of the DwC-A including a mapping of the columns in the core and extension files to terms declared by published vocabularies. The DwC-A is created as a zip archive including the metafile, core and extension files. A resource metadata document (by default named EML.xml) can be included in the zipped archive or referenced with a link to an online metadata resource describing the dataset. GBIF recommends providing metadata using the GBIF Metadata Profile (Ó Tuama et al., 2011), which is based on the EML (ecological metadata language) (Michener et al., 1997; Fegraus et al., 2005). Sharing the entire dataset in this manner as a DwC-A allows for simpler and more efficient data transfer compared to web service interfaces provided by toolkits such as the BioCASE and DiGIR (Darwin Core task group, 2009b; Döring et al., 2011).

**GBIF Knowledge Organization System (KOS)**

Recently GBIF has convened expert task groups for providing recommendations on implementing knowledge organization systems (KOS) (Catapano et al., 2011; Lapp et al., 2011), metadata standards (Jones et al., 2009), and persistent identifiers (Cryer et al., 2010; Richards et al., 2011) for biodiversity information resources. One of the recommendations from these task group reports was the implementation of persistent identifiers for each individual vocabulary term concept. When no previously established persistent identifier was available for a term, they recommended a new persistent identifier to be issued. An emphasis was also made on reusing existing terms and concepts wherever possible. These recent guideline principles align very well with the approach that was followed when the Darwin Core extension for genebanks was developed.

[Flat] vocabularies should be developed so that they are reusable as a terminological foundation for semantically richer vocabularies or ontologies (Catapano et al., 2011:3).

The World Wide Web Consortium (W3C) chartered a task force in 2004 for the development of guidelines for managing Resource Description Framework (RDF) vocabularies of terms (W3C, 2006). The task force recommended the following general principles: “(1) use URIs for naming; (2) provide readable documentation; (3) articulate maintenance policies; (4) identify versions; and (5) publish a formal schema” (Kendall et al., 2008). The W3C publishes two dedicated vocabularies with terms for the description of such vocabularies, namely the RDF vocabulary description language.

---

25 http://www.tdwg.org/standards/450/
26 http://rs.gbif.org/schema/eml-gbif-profile/1.0.1/.
(RDFS; Brickley et al., 2004) and the simple knowledge organization system (SKOS; Miles and Bechhofer, 2009). Terms from the germplasm vocabulary are described using a combination of properties from both RDFS and SKOS.

RESULTS

During 2008, work was started at GBIF for a major upgrade of the data publishing toolkit for the GBIF network. The new tool was named GBIF Integrated Publishing Toolkit (IPT) and is based on the Darwin Core (DwC). The Darwin Core was at the time under revision by TDWG for a new version scheduled to be ready in 2009. While the ABCD standard is very comprehensive with several thousand terms, the Darwin Core standard implements a more limited set of core terms with domain-specific terms organized in a number of published extensions. There was, as of 2008, no Darwin Core extension to ensure full interoperability with the genebank information requirements. During a Darwin Core workshop in Copenhagen (hosted by GBIF) in January 2009, work was initiated to develop an extension for germplasm to the new revised Darwin Core standard. The Darwin Core extension for genebanks (DwC-germplasm) is required for the rational use of the GBIF IPT in the genebank community (Endresen et al., 2009). The DwC-germplasm provides a comparable piece in the interoperability puzzle as the ABCD version 2.06 provided in 2005 (Berendsohn, 2005; Berendsohn and Knüpffer, 2006) to enable the rational use of the BioCASE toolkit in the genebank community (Endresen et al., 2006).

**Darwin Core Extension for Genebanks (DwC-germplasm)**

The first draft version of the DwC-germplasm was published for discussion at the EPGRIS3 wiki. The EPGRIS3 (Establishment of a European Plant Genetic Resources Information Infra-Structure, phase 3) is an initiative of the ECPGR Documentation and Information Network. The initial development of the DwC-germplasm at the EPGRIS3 wiki attracted feedback and suggestions from the ENSCONET (European Native Seed Conservation Network) project regarding additional terms for in situ conservation of genetic resources. The Millennium Seed Bank proposed additional terms to describe ex situ germplasm conservation management routines. After receiving further feedback from some other communities outside the European genebank community, a new DwC-germplasm project home page was established at Google Code. The Google Code site replaced the EPGRIS3 Wiki as the official home page for the DwC-germplasm vocabulary of terms. Future modifications and eventual additional terms to the DwC-germplasm will be discussed and agreed here before they will be passed on to be consolidated within the genebank community and eventually included in the official version of the extension. The official normative version of the Darwin Core extension for genebanks is published and maintained at “http://purl.org/germplasm/germplasmTerm.rdf”.

The terms included in the DwC-germplasm vocabulary are organized into 11 groups declared by the SKOS/RDF resource as “skos:Collection” resources. These groups can be seen as entity types and are declared as “rdfs:Class” resources. The germplasm terms describe properties with an intentional loose connection to these entities. The terms organized under dataset (1), taxon (2) and collecting event (3) were imported from Darwin Core and Dublin Core with no new terms declared by the germplasm extension. The description of the specimens (genebank accessions) (4) maintained in living collections was supplemented with terms related to the storage conditions and the “biological status of sample” defining the cultivation status ranging from wild plants via landraces and primitive crops to the modern so-called advanced cultivars. The terms for description of the breeding or domestication event (5) for crops are unique for the germplasm extension with no overlapping terms imported from the Darwin Core. The same orthogonality of terms in relation to Darwin Core, applies for terms organized to the acquisition event (6) describing the donation and sharing of living germplasm material between genebank collections, and also for the safety duplication event (7) describing the backup storage of germplasm material at multiple locations. The terms for describing international treaties and regulations (8) governing the access and ownership of germplasm material are also unique to the germplasm extension. The Darwin Core terms organized as “MeasurementOrFact” (9) were also supplemented and reorganized by the germplasm extension. The germplasm extension provides terms for linking to external resources describing the measurement method (trait descriptor) (10). External trait descriptions include measurement methods described and declared by ontologies such as the crop ontology (CO) (Shrestha et al., 2010), plant ontology (PO) (Jaiswal et al.,...
2005), the phenotypic quality ontology (PATO)\(^{33}\) and the plant trait ontology (TO)\(^ {34}\) (Jaiswal et al., 2002). Some terms were also added related to the description of the measurement experiment (11) including the time and location of the experiment. The goal was the maximum reuse of existing terms from existing vocabularies and ontologies and the minting of new terms only when no comparable terms were found.

**Darwin Core Archive (DWCA-A) Extension for Genebanks**

For the terms declared as part of the DwC-germplasm to be made available to the GBIF infrastructure software tools such as the GBIF IPT and to be included into data sets shared using the DwC-A data-publishing format (Döring et al., 2011), the relevant XML application lists were developed. The GBIF Resources Registry provides the formal specifications for these XML applications in the format of a XML schema\(^ {35}\). The GBIF Vocabulary Server\(^ {36}\) provided a software tool to assist in the development of the DwC-A extensions\(^ {37}\) for the DwC-germplasm terms. The final XML applications required for including the germplasm terms to DwC-A dataset resources were published at the GBIF Resources Registry\(^ {38}\).

**Deployment of DwC-germplasm in GBIF IPT**

In 2010, GBIF, NordGen and Bioversity International initiated a feasibility study to evaluate how the GBIF infrastructure can meet the needs of the European genebank community (Gaiji et al., 2010). This feasibility study was also coordinated with the ECPGR Documentation and Information Network for the genebank community in Europe (Maggioni, 2010). The prototype GBIF Integrated Publishing Toolkit (IPT version 1.0) was installed in five genebanks within the European plant genetic resources catalogue (EURISCO) using the DwC-germplasm extension. These were the national genebanks in the Russian Federation (Vavilov Institute, Sankt Petersburg), Germany (IPK Gatersleben), Czech Republic (Crop Research Institute, Prague), the Netherlands (Wageningen University and Research Center for Genetic Resources), and within the Nordic and Baltic countries (genebank database hosted from the Nordic Genetic Resources Center).

While the prototype version of the IPT software caused some problems of instability during installation (1), the mapping of the genebank data sets to Darwin Core, including the genebank extension (2) and finally the registration to the GBIF GBRDS (3) was completed satisfactorily. The hardware requirements and in particular the demands for internal memory were a major barrier encountered during most of the installations. The experiences from the genebank feasibility study provided the IPT development team at GBIF with feedback and suggested improvements leading from the different prototype versions to the new version 2. The hardware requirements for the new version of the IPT software (version 2) have been significantly reduced and solve all of the issues encountered and reported from the genebank feasibility project. The project resulted in a positive evaluation and the genebank community has started the initial plans for a second feasibility study to evaluate the IPT version 2 at other genebanks in Europe.

**Updates to the Vocabulary of Germplasm Terms**

Following the experiences from the first draft version of the DwC-germplasm (version 0.1) and the IPT feasibility study at some of the European genebanks, some updates to the germplasm terms were made. Based in part on the recommendations from the GBIF KOS task group (Catapano et al., 2011; Lapp et al., 2011) and the initial work by the Vocabulary Management Task Group\(^ {39}\) (Endresen et al., 2012a), the DwC-germplasm terms were declared as a RDF/SKOS vocabulary. The recommendations from the World Wide Web Consortium on best practices for management of RDF vocabularies (Kendall et al., 2008) provided other useful principles for the upgrade of the germplasm terms.

The updated vocabulary of germplasm terms\(^ {40}\) was moved to the PURL (persistent uniform resource locators) namespace to improve the commitment of long-term persistence. PURL is managed by the Online Computer Library Center (OCLC) to provide persistent and resolvable identifiers for online resources.

**RDF Vocabulary Maintenance Policy**

The DwC-germplasm vocabulary of terms is extensible and new terms can be added to the DwC-germplasm namespace in the future. Terms can evolve through refinement in response to deployment and testing. Refinements to the definition and description of a term will as far as

---


\(^{34}\) http://www.gramene.org/plant_ontology/.

\(^{35}\) http://rs.gbif.org/schema/extension.xsd.

\(^{36}\) http://vocularies.gbif.org/.

\(^{37}\) http://vocabularies.gbif.org/extensions/.

\(^{38}\) http://rs.gbif.org/extension/nordgen/.

\(^{39}\) http://community.gbif.org/pg/groups/21382/vocabulary-management/.

\(^{40}\) http://purl.org/germplasm/germplasmTerm.rdf.
possible maintain the semantic meaning of the term. In situations when the original semantic meaning of a term is jeopardized, the term will be deprecated and replaced by a new term. The previous versions of terms will be maintained in a separate vocabulary describing the history of the terms. All terms are described by a label (skos:prefLabel), a definition (skos:definition), some examples (skos:example) and a note (skos:note) explaining the scope and how to use the term. These natural language descriptions are only available in English. A separate vocabulary might be developed to provide descriptions of the terms expressed in other natural languages than English.

**DISCUSSION**

The Darwin Core standard is itself an extension of another standard, the Dublin Core Metadata Initiative (DCMI)\(^4\). The Dublin Core provides a bridge to ensure low-level interoperability between wide ranges of metadata standards. Implementing Darwin Core as an extension to the Dublin Core promotes the interoperability of biodiversity data with information from other domains. The Darwin Core set of core terms includes some terms from the Dublin Core ‘terminology’\(^4\). But more important than the shared terms, is the shared framework to describe and implement the terms in applied solutions. The same principles apply to the benefits of building a genebank extension based on the Darwin Core, or adapting other solutions from outside the germplasm community network. By following a few ways and guidelines for ‘best practices’, the genebanks can with few efforts adapt tools and principles developed in other communities for efficient use in their own information network (Knüpffer *et al*., 2007).

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

Figure 3: During the 2010 feasibility study for the European genebank community, the prototype GBIF Integrated Publishing Toolkit (IPT) was installed at the national genebanks in the Russian Federation, Germany, Czech Republic, the Netherlands, and at the Nordic Genetic Resource Center (hosting the genebank database for the Nordic and Baltic countries).
establish specific PGR information networks without creating its own technology (Knüpffer et al., 2007:7).

By finding a few common ways and guidelines for ‘best practices,’ genebanks can with fewer efforts adapt tools and principles developed in other parts of their ‘own’ community for efficient use across the entire germplasm information network.

**Automatic Data Exchange Mechanisms**

Many of the present data exchange mechanisms in use in the genebank community rely on laborious and repeated transformations of the original genebank data sets into the agreed standard formats. The genebanks in Europe regularly produce an updated subset from their information system complying with the EURISCO data exchange format (based on the Multi-Crop Passport Descriptors). Then the subset from each genebank in a country is combined into a so-called National Inventory and uploaded to the central EURISCO data portal (hosted by Bioversity International). The European Central Crop Databases (ECCDB) also request from each genebank to extract a similar subset from their information system.

Many ECCDBs ask for data on selected descriptors from the Bioversity Crop Descriptor lists in addition to the MCPD. The ECCDBs are limited in scope each to a different crop species and ask thus for a different set of additional crop-specific descriptors. While the EURISCO has implemented an online data upload tool to receive the updated national inventories, the updated subsets for the ECCDBs are often exchanged as email attachments. The CGIAR genebanks share similar subsets from their information systems with the System-wide Information Network for Genetic Resources (SINGER)⁴⁳. The FAO WIEWS (World Information and Early Warning System on PGRFA)⁴⁴ also requests, on a regular basis, updated subsets from all genebanks worldwide. The requested format for these subsets is also roughly based on the MCPD standard. The record level data unit is however different, as WIEWS request metadata on stratified groups of genebank accessions, rather than the accession level data requested by EURISCO, ECCDBs and SINGER.

New data exchange mechanisms using web services have the potential to make all these aforementioned data exchange operations fully automatic. And with the new data provider toolkit software packages provided as an open source public

---

⁴³ http://singer.cgiar.org/.

---

ENDRESEN AND KNÜPFER – DARWIN CORE EXTENSION FOR GERMPLASM GENEBANKS

**GOOD**

good from the Global Biodiversity Information Facility (GBIF), the required efforts to establish and maintain such fully automatic multi-purpose data flow pathways with web services are getting less demanding and becoming more low-tech to implement. The GBIF Integrated Publishing Toolkit (IPT) is the latest and most user-friendly software package for sharing biodiversity data sets (such as the genebank data sets). The Darwin Core extension for genebanks (DwC-germplasm) provides a necessary ‘plug-in’ to make the new GBIF IPT available for rational use in the genebank community. The genebank community was one of the first biodiversity information networks to develop this type of plug-in to start using the GBIF IPT. It is expected that the experiences from the development and implementation of the DwC-germplasm for the genebanks can provide some examples for other biodiversity information networks to study. With the following section, we aim to describe the steps to follow in order to develop a similar Darwin Core extension in other biodiversity information networks.

**HOW TO Create a New Darwin Core Extension**

The development and implementation of the Darwin Core extension for germplasm can be used as an example for other biodiversity information communities to develop their own DwC extensions. The following steps have to be carried out:

1. The community needs to compile a consolidated list of terms to describe their data domain.
2. After finding agreement on the terms with the relevant stakeholders inside the relevant community, these terms should be harmonized and mapped to the standard Darwin Core terms⁴⁵. New terms should only be defined for an extension if they are not already included in the standard core terms. Some of the descriptor terms implemented in a community may be similar to one of the core terms, but with a different formatting or a slightly different semantic meaning. Whenever possible, it is recommended to try to convert the data content for a community descriptor term to follow the definition of one of the standard DwC terms. If a new community term is defined that could have been converted to one of the existing DwC terms, interoperability with biodiversity data sets from other communities will be broken.
3. We recommended declaring new terms for your Darwin Core extension using the simple knowledge organization system (SKOS) and

---

the resource description framework (RDF/RDFS) vocabulary.

(4) Darwin Core extensions and other community terminology vocabularies can be published at the GBIF Resources Registry.46

(5) The next step is to create appropriate XML lists including your terms following the GBIF XML schema specifications for Darwin Core Archive extensions. When designing new DwC-A extensions you may mix and match terms from many different term vocabularies. The GBIF Vocabulary Server (Harman et al., 2009) provides a software tool to assist you with defining a correctly formatted DwC-A extension.

(6) The final DwC-A extension must be loaded to the GBIF Resources Registry before it is available to software tools such as the GBIF Integrated Publishing Toolkit (IPT). You will (at least for now) need to contact the GBIF helpdesk (helpdesk@gbif.org) for assistance with loading your resources to the GBIF Resources Registry.

Please note that after publishing the DwC-A extension to the GBIF Resources Registry, any modifications to the extension (however minor) need to be released with a new version number. The recommendation for the release of new DwC-A extensions based on the DwC-germplasm terms is to include a postfix with the release date (e.g., “germplasm_20120710.xml”).

New terms and concepts should be developed in a collaborative manner allowing for feedback from your community. The Vocabulary Management Group (VoMaG)47 currently performs an evaluation of various software tools to support the collaborative development of new terms including the Semantic MediaWiki49, ISOcat50 and Drupal-based51 tools. The ratified version for each vocabulary of terms for the description of biodiversity information resources can be registered and deposited at the GBIF Resources Registry. Darwin Core Archive extensions and controlled value vocabularies should be designed to re-use terms from one of the ratified and published flat vocabularies (RDF/SKOS) or from a published ontology (OWL). We also recommend here as a best practice guideline to re-use terms declared by a flat vocabulary when developing new biodiversity ontologies (OWL resources).

Because the genebank community already had established information standards, the development of a draft extension to the Darwin Core (DwC-germplasm), and the subsequent testing of the new prototype information publishing toolkit from GBIF (GBIF IPT), progressed quickly and with relatively few problems.

**Evaluation of the Updated GBIF IPT Version 2**

The experiences so far from testing the updated version 2 of the IPT are mostly positive. All the graphical features for visualization of the data sets were removed in this version. This simplification was one of the recommendations reported by the genebank feasibility study. This focus of the IPT on data publishing rather than visualization has resulted in substantially improved performance of the toolkit. The new version 2 has also removed the embedded internal database. As a result of this modification, the web services providing various query interfaces to the data sets shared by IPT have also been removed. In particular the web service interfaces providing access with the TAPIR protocol (TDWG access protocol for information retrieval)52 (TDWG, 2010), and the OGC (open geospatial consortium) WFS (web feature service) were interesting APIs (application programming interfaces) to the underlying data. However, the simplicity of the IPT with a dedicated design for providing the Darwin Core Archive (DwC-A) data-sharing format makes the IPT a lightweight and efficient software application. There are also other similar data publishing toolkits such as the TapirLink53 and BioCASE54 that provide GBIF-compatible services for publishing biodiversity data sets.

**Efficient Access to Distributed Germplasm Data Sets Stimulates Novel Uses**

Research integrating genebank passport data (georeferenced occurrence data for the original collecting site) with phenotypic measurements (characterization and evaluation data) and with ecolimatic layers has opened new possibilities for a rational utilization of genebank materials (Bhullar et al., 2009; Endresen, 2010; Endresen et al., 2011; Bari et al., 2011; Endresen et al., 2012b) using the Focused Identification of Germplasm Strategy (FIGS) approach (Mackay and Street, 2004). The efficient application of the FIGS approach depends on the availability of germplasm passport and trait evaluation data. The analysis of gaps in the
genebank collections to guide the planning of rational germplasm collection expeditions to complement the genebank collections with novel and insufficiently sampled genetic diversity is also dependent on the availability of genebank passport data (Jarvis et al., 2003; 2005; 2009; Ramirez-Villegas et al., 2010). Such data analysis experiments to identify the ecological environment linked to a target trait property, or the genetic gaps of the genebank collections, will of course benefit from occurrence data on crop wild relatives provided from other communities. The value of external data from outside the genebank community, in such studies, strengthens the argument for the development of common semantic data standards (like the Darwin Core) and standardized data exchange protocols (such as the Darwin Core Archive format). Limited access to genebank accession-level information is a bottleneck to the efficient use of genebank material (FAO, 2010), as well as to the development of novel uses for the associated data. The authors of this manuscript propose the Darwin Core extension for genebanks and its implementation in the GBIF Integrated Publishing Toolkit (IPT) as a contribution for an upgrade of the current data exchange mechanism for genebank data sets.

Future Work

After the first experiences with the deployment of the Darwin Core extension for genebanks, a useful next step will be to seek ratification of the extension as a TDWG standard. The genebank community has long and successful experience with the development and maintenance of descriptor standards, in particular through the work at Bioversity International (Bioversity International, 2007; Gotor et al., 2008). However, as discussed above, one of the major achievements with the DwC-germplasm is the interoperability with other biodiversity information standards and communities outside the genebank community. The ratification of genebank standards like the DwC-germplasm in TDWG will contribute to improved information interoperability.

The first version of the DwC-germplasm included the proposed EPGRIS3 descriptors for evaluation and characterization data. The sharing of trait data sets for germplasm has received renewed attention with the second report on the state of the world’s plant genetic resources for food and agriculture (FAO, 2010). These descriptors need further work after the first experiences with the sharing of germplasm trait data sets.

The implementation during the last years of new international regulations for the sharing of benefits for the use of plant genetic resources prescribes the reporting of the distribution of seed samples (defined by the ITPGRFA, Annex 1). If the terms to describe and report these seed distributions are developed and included to the DwC-germplasm, then the GBIF IPT could be used to report seed distributions to the Governing Body of the ITPGRFA.

When germplasm data sets are published, the entities are almost exclusively identified using local identifiers. Often the institutes sharing these data sets are identified using institute codes from the FAO WIEWS (world information and early warning system)\textsuperscript{55}. The combination of the WIEWS institute code and the local identifier for entities (such as genebank accessions) are generally sufficient to ensure unique identification. One major concern with this practice is that the institution codes are not designed to be globally unique and persistent identifiers (PIPs). For a description of globally unique and persistent identifiers see Coyle (2006), Campbell (2007) and Richards et al. (2011). The required combination of the institute identifier and the local identifiers for distinct identification of entities is another major bottleneck for efficient use of information from the germplasm data sets published online today. The GBIF \textit{data publishing framework task group} recommends the publication of biodiversity data sets as citable “data papers”\textsuperscript{56} and that each dataset is identified by a PID for consistent data citation (Moritz et al., 2011). The Dryad data repository provides a similar service where the data sets supporting published peer-review papers can be archived and provided with a DOI (digital object identifier)\textsuperscript{56} for consistent citation and persistent data availability (Greenberg et al., 2009; Michener et al., 2011). The persistent identification of data sets could become a major step towards consistent identification of entities in germplasm collections. However, we recommend that also the entities inside the germplasm data sets will be identified using PIDs. It is our opinion that the full potential of using the DwC-germplasm terms when publishing germplasm data sets will be limited without the use of PIDs for consistent identification of entities inside the data sets such as the genebank accessions. Efforts should further be made for the reuse of existing PIDs for such entities.

The Darwin Core extension for genebanks was developed as a flat vocabulary of basic terms. This vocabulary is declared using RDF/SKOS and

\textsuperscript{55}http://apps3.fao.org/wiews/institute_query.htm?i_l=EN
\textsuperscript{56}http://www.doi.org/
CONCLUSIONS

The Darwin Core germplasm extension provides access to the GBIF bioinformatics infrastructure, including the GBIF Integrated Publishing Toolkit (IPT). Using the GBIF IPT and the Darwin Core germplasm extension, genebanks can now share germplasm data sets with each other. This new data exchange mechanism will make the development of distributed germplasm information networks easier. The DwC germplasm extension also provides a frame for implementing a standardized process of data exchange. Implementation of general biodiversity information standards and toolkits will ensure the interoperability of genebank data sets with other biodiversity data sets.

ACKNOWLEDGMENTS

Thanks for help and support from colleagues of the GBIF secretariat, GBIF network participants, BioCASE team, DiGIR team, TDWG community, Bioversity International, Nordic Genetic Resource Center, EURISCO participants, EPGIS3 members, and the ECPGR Documentation and Information Network Coordinating Group. The mapping of MCPD to ABCD was carried out in collaboration with Walter Berendsohn and Javier de la Torre and their colleagues from the Botanical Garden and Botanical Museum Berlin-Dahlem, Germany. John Wieczorek and the other members of the Darwin Core task force provided advice and support during the development of the genebank extension to Darwin Core. Tim Robertson, Markus Döring, José Cuadra, Vishwas Chavan and Samy Gaiji at the GBIF secretariat provided support and assistance with the deployment of the DwC-germplasm in the GBIF IPT software. Kehan Harman provided valuable feedback regarding the GBIF Vocabulary Server. Theo van Hintum was the driving force with the development of the trait descriptors for the EPGIS3 activity and provided important feedback and discussion leading to the DwC-germplasm vocabulary. The EURISCO secretariat and colleagues at the Nordic Genetic Resources Center including Jonas Nordling and at Bioversity International including Elizabeth Arnaud, Sónia Dias, Milko Škofič, and Michael Mackay provided assistance with the development of the DwC-germplasm and with the feasibility study to deploy the GBIF IPT for the first European genebanks. Thanks to Markus Oppermann (IPK Gatersleben) for providing feedback to the manuscript. Many thanks for valuable suggestions and corrections provided by the reviewers.

REFERENCES


57 http://www.bgmb.org/TDWG/C0DATA/default.htm.


EURISCO (2012). Descriptors for uploading information to EURISCO, 70.


EURISCO (2012). Descriptors for uploading information from National Inventories to EURISCO70.


---

68 http://community.gbif.org/lg/file/read/21582/


76 http://www.gbif.org/governance/governing-board/governing-board-meetings/gb1/.


26
IBPGR (1976). First report of the advisory committee on the genetic resources communication, information and documentation system (GR/CIDS). International Board for Plant Genetic Resources (IBPGR), Rome, Italy. AGP 62:IBPGR/76/7.
IBPGR (1977). Descriptors for the cultivated potato and for the maintenance and distribution of germplasm collections. International Board for Plant Genetic Resources (IBPGR), Rome, Italy.
IBPGR (1978). Descriptors for wheat and Aegilops: a minimum list. International Board for Plant Genetic Resources (IBPGR), Rome, Italy.
IPGRI (2002). EPGRIS entering its third and final year. EURISCO and central crop databases. Newsletter for Europe 25: 5. 90

http://www.bioversityinternational.org/nc/publications/publicatio
n/issue/newsletter_for_europe-21.html
http://www.bioversityinternational.org/nc/publications/publicatio
n/issue/newsletter_for_europe-19.html
83 http://imsgbif.gbif.org/CM5_NEW/get_file.php?FILE=2d85d0e8c76
40f1230046c90a07236
84 http://www.w3.org/2006/07/SWD/Vocab/principles.
Mackay, M.C., and K. Street (2004). Focused
databases: Tools for plant genetic resources
management. International Plant Genetic Resources
Institute, Rome, Italy and Centre for Genetic
Resources (CGN), Wageningen, Netherlands. ISBN
92-9043-320-5.

of the world collection of plant genetic resources in
Russia. International Plant Genetic Resources
Institute (IPGRI), Rome, Italy. ISBN: 92-9043-412-
0.

Catalogue of the VIR world collection, edition 735,
Oats. The N.I. Vavilov All-Russian Scientific
Research Institute of Plant Industry, St. Petersburg,
Russia.

Mackay, M.C., and K. Street (2004). Focused
Identification of Germplasm Strategy – FIGS.
Cereals, pp. 138–141. In: Black, C.K., J.F. Panozzo,
and G.J. Rebetzke (eds). Proceedings of the
Australian Cereal Chemistry Conf., 54th, and the
Wheat Breeders’ Assembly, 11th, Canberra, ACT. 21–
24 September 2004. Royal Australian Chemical
Institute, Melbourne, Australia.

Coordinating Group on Documentation and
Information and the EURISCO Advisory Group.
International Plant Genetic Resources Institute
(IPGRI), Rome, Italy.86.

Maggioni, L. (ed.) (2007). Minutes of a joint meeting of
the documentation and information network
coordinating group and the EURISCO advisory
group. Planning for the continuation of EPGRIS, 2–3

documentation and information network coordinating
group, forth meeting, 17–18 February 2010,
Maccarese, Rome, Italy. Bioversity International,
Rome, Italy.88.

Michener, W.K., J.W. Brunt, J.J. Helly, T.B. Kirchner,
and S.G. Stafford (1997). Nongeospatial metadata for
the ecological sciences. Ecological Applications 7(1):
330–342. DOI: 10.1046/j.eaconf.2005.08.004

Michener, W., D. Vieglaís, T. Vision, J. Kunze, P. Cruse,
and G. Janee (2011). DataONE: Data observation
network for earth – preserving data and enabling
innovation in the biological and environmental
sciences. D-Lib Magazine 17(1/2), 12 pp. DOI:
10.1045/january2011-michener

Miles, A. and S. Bechohofer (2009) SKOS Simple
Knowledge Organization System: Reference. W3C
recommendation 18 August 2009.89

Moore, J.D., S.P. Kell, J.M. Iriondo, B.V. Ford-Lloyd,
and N. Maxted (2008). CWRML: Representing crop
wild relative conservation and use data in XML.
BMC Bioinformatics 9: 116. DOI: 10.1186/1471-
2105-9-116

Moritz, T., S. Kirshnan, D. Roberts, P. Ingwersen, D.
Agosti, L. Penev, M. Cockerhill, and V. Chavan
(2011). Towards mainstreaming of biodiversity data
publishing: recommendations of the GBIF data
publishing framework task group. BMC
Bioinformatics 12(Suppl 15) S1. DOI: 10.1186/1471-
2105-12-S15-S1

Ó Tuama, É., K. Braak, and D. Remsen (2011). GBIF
metadata profile, how-to guide. Global Biodiversity
Information Facility (GBIF), Copenhagen, Denmark.90

Peterson, A.T., D.A. Vieglais, A.G. Navarro Sigienza,
and M. Silvia (2003). A global distributed
biodiversity information network: building the world
museum. Bulletin of the British Ornithologists’ Club
123A: 186–196.

Report of a network coordinating group on cereals,
ad hoc meeting, 7–8 July 2000, Radzików, Poland.
International Plant Genetic Resources Institute
(IPGRI), Rome, Italy.

Ramirez-Villegas, J., C. Khoury, A. Jarvis, D.G.
Debouck, and L. Guarino (2010). A gap analysis
methodology for collecting crop genepools: A case
study with Phaseolus beans. PLoS ONE 5(10):
e13497. DOI: 10.1371/journal.pone.0013497

Regel R.E. (1915). Organizatsiya i devyat’ nos’ Byuro
po prikladnoy Botanike za pervoe dvadtsatiletie ego
sushchestvovaniya [Organization and activity of the
Bureau of Applied Botany for the first twenty years
of its existence]. Bulletin of the Bureau of Applied

A beginner’s guide to persistent identifiers, version
1.0. Released on 9 February 2011. Global
Biodiversity Information Facility (GBIF),
Copenhagen, Denmark.91

Rogalewicz, V. (ed.), in collaboration with H. Knüpffer,
V.A. Korneychuk, I.G. Lozanov, D.B. Plotnikov, J.
deskitory otitel’nyh byuro mezhduunarodnyh baz
sushchestvovaniya [passport descriptors of the
COMECON International Database of
Genetic Resources]. Výzkumný ústav rostlinné
výroby, Praha-Ruzyně, Czechoslovakia. 26 pp.
of genus Secale L. First edition. European
Cooperative Programme for the Conservation and
Exchange of Crop Genetic Resources, International
Board for Plant Genetic Resources (IBPGR), Rome,
Italy.

90 http://www.gbif.org/orc/?doc_id=2821
91 http://www.gbif.org/orc/?doc_id=2428


93 http://www.w3.org/2003/12/swa/swbpd-charter.