Available online at www.sciencedirect.com
ScienceDirect
Computers and Electronics in Agriculture 59 (2007) 90–96
Computers and electronics in agriculture
www.elsevier.com/locate/compag

Application note

Genebank Information Management System (GBIMS)

R.C. Agrawal *, Drona Behera, Sanjeev Saxena

National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi 110012, India

Received 22 April 2006; received in revised form 8 April 2007; accepted 10 April 2007

Abstract

Genebank Information Management System (GBIMS) software has been developed for catering to the needs of the genebank staff of the National Bureau of Plant Genetic Resources (NBPGR) of India. Presently, information for more than 320,000 accessions stored in the National Genebank of NBPGR is being maintained through this software. The GBIMS is developed using Visual Basic as the front end and Microsoft SQL Server as the relational database management system in the back end. This software is based on client–server technology in which many clients make requests from the same database. The design and description of GBIMS software in this paper may be used by various organizations involved in maintaining germplasm collections.

Keywords: Genebank; Database; Descriptors; Passport information; Viability; Crop; Germplasm collection; Seed bank

1. Introduction

Considerable worldwide efforts for *ex situ* conservation of plant genetic resources (PGR) have been undertaken at the national and international levels. Undoubtedly genebanks play an extremely important role in conservation of agro-biodiversity. The fact that in the majority of species, seeds are highly amenable for storage in genebanks at low temperatures, seed conservation is the most convenient and widely used method of *ex situ* conservation of species and genotype diversity for posterity. Based on the requirement of conservation programmes, seed storage may either be short, medium or long term. The favourable cool and dry conditions of the genebanks prolong the life spans of seeds by checking the deterioration during storage over a period of time and thereby enhancing the longevity of accessions. According to the Food and Agriculture Organization, there are about 6,000,000 accessions in 1308 genebanks (FAO, 1997). These genebanks range from very small to very large in size.

Documentation of information on PGR is imperative for planning and implementing activities related to their conservation, sustainable utilization and sharing of benefits accrued from their use. The need for countries to develop, maintain and exchange such information is specifically recognized in Articles 7d, 17 of the Convention of Biological Diversity (CBD, 1993) and the priority activities 17 and 18 of the Global Plan of Action (FAO, 1996). Genebanks with high-quality information about their resources are in better position to decide on current resource utilization, to reduce duplication of effort and to provide vision for future directions. Also full and proper documentation will help to authenticate information on the origin, development and maintenance of the germplasm accessions. Earlier, the information on PGR was documented through the collection tour reports, memoirs of collection expeditions and the

* Corresponding author. Tel.: +91 11 25846074; fax: +91 11 25842495.
E-mail addresses: rakesh@nbpgr.ernet.in, agrawal_rakesh_chandra@yahoo.com (R.C. Agrawal).

0168-1699/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2007.04.005

catalogues published after collection and/or evaluation of germplasm. The developments in computation technology made information management in PGR through databases very important at the national, regional and global levels using standard descriptors for effective exchange of information (Lipman et al., 1997). Presently, a good number of genebanks have developed their own information system, fitting their requirements, based on the availability of the computer system and tailored Database Management System software. The Germplasm Resources Information Network (GRIN) system is effectively being used to monitor information on the world's largest collection at the National Center for Genetic Resources Preservation, Fort Collins and the cooperating institutions within the United States Department of Agriculture (USDA) research system in USA (http://www.ars-grin.gov/). The Bioversity International (erstwhile IPGRI) and USDA have developed genebank management software known as pcGRIN, which is a stand-alone version of the GRIN database that runs on a personal computer and contains data for individual crops (http://www.bioversityinternational.org/Themes/Germplasm_Documentation/index.asp). The SINGER (System-wide Information Network for Genetic Resources), has been established as the genetic resources information exchange network for nearly 0.7 million accessions of the Future Harvest Centres of the Consultative Group on International Agricultural Research (CGIAR), and associated partners (http://singer.grinfo.net/). In addition, the International Agricultural Research Institutes have developed their own information systems for handling the germplasm in their respective mandate crops. The Federal Information System Genetic Resources (BIG), Germany, integrates databases on PGR *ex situ* (e.g. collections of botanical gardens and accessions of German gene banks) and *in situ*, including the wild flora of Germany (http://www.big-flora.de/).

In India the National Genebank at National Bureau of Plant Genetic Resources (NBPGR) acts as a nodal organization for maintaining the *ex situ* collection of PGR on long-term basis. The National Genebank operates in a network mode to have an effective linkage with other national and international organizations engaged in PGR management and is entrusted to maintain the base collection. The major partners in the network include 10 regional stations, spread over different phyto-geographical zones in the country and 57 National Active Germplasm Sites located in various Indian Council of Agricultural Research (ICAR) Institutes, National Research Centres and All India Coordinated Crop Improvement Projects and State Agriculture Universities (SAUs).

In order to cater to the requirements of the National Genebank in terms of large number of accessions and species conserved, the Genebank Information Management System (GBIMS) has been developed. The GBIMS is being used at the National Genebank that manages the large amount of data on germplasm in various crop species. The database presently contains data related to more than 320,000 accessions conserved in the National Genebank belonging to 1187 species. It was designed specifically for data capture in an organized, structured format and provision for data retrieval with facility for data processing, data analysis and data exchange.

2. Material and methods

The GBIMS has been developed as a relational database and the basic architecture consists of client-side interface layer in Visual Basic, an event driven programming language from Microsoft, front end and database layer on Microsoft SQL server, a relational database management system. In client–server technology, many clients/applications can access the same database. As data entry can be done only from designated computers (having client application software), this technology provides better safeguard and information fidelity. In addition, for making simple queries on the intranet, a web-based application has been incorporated for the general user. The regular software updates for the client application are provided by the sharing of executable files.

2.1. Seed processing procedure in the National Genebank

Seeds received in the genebank are accessioned by providing the national identity number. These numbers are provided in a serial order and have a prefix IC, for indigenous collections and EC, for exotic collections (received through exchange from other countries). The accessions are physically screened to remove under-sized, shrivelled, diseased or immature seeds and clean and healthy seeds are subjected to the seed germination test. Accessions having more than 85% viability are transferred into muslin cloth bags and are allowed to equilibrate at 15 °C temperature and 15% relative humidity in the seed dryer to attain a moisture content in the range of 3–7%. The dried seeds are hermetically sealed in a tri-layered aluminium foil pouch for conservation as base collections in one of the 10 modules maintained at −20 °C. The seed genebank aims to store good quality seed

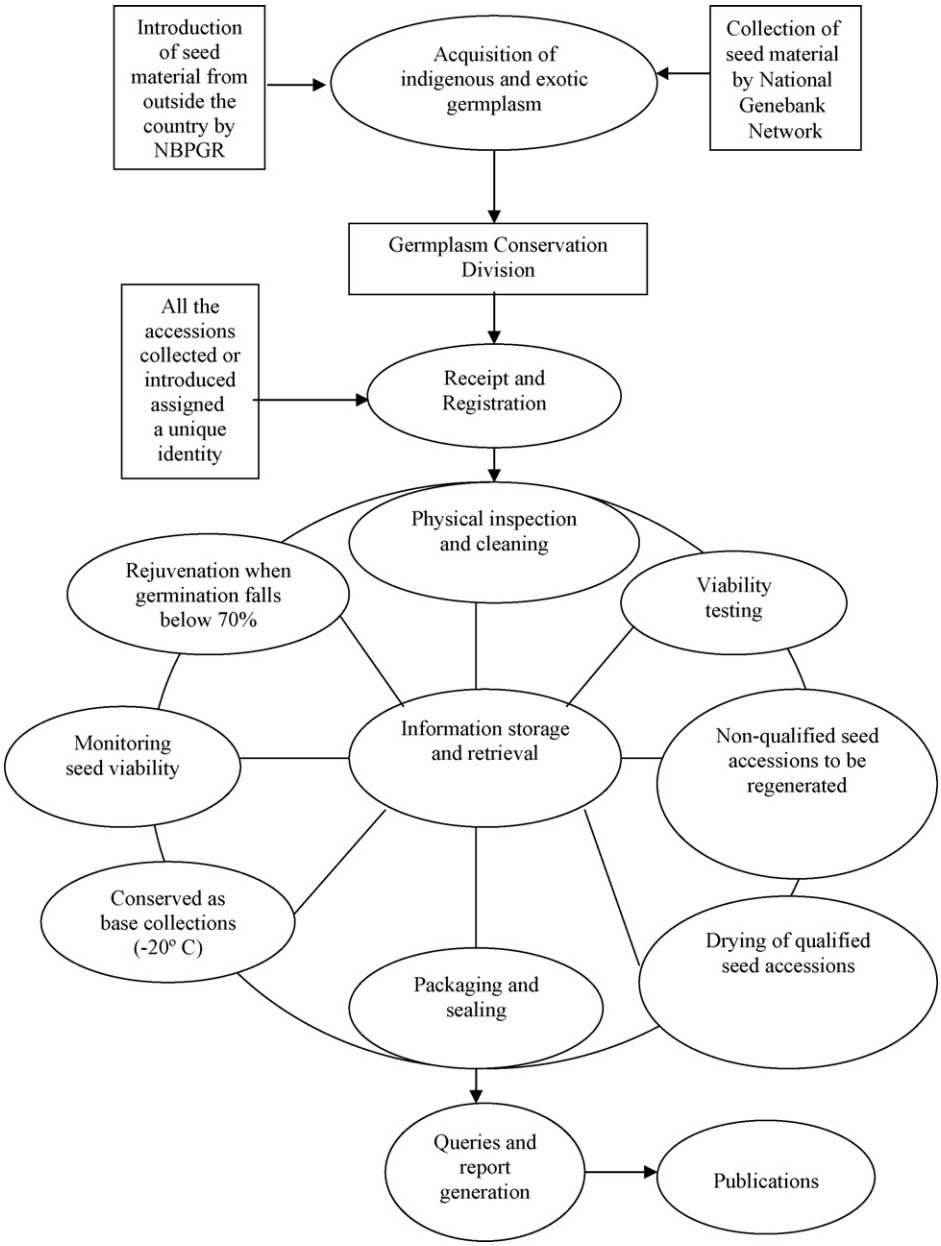


Fig. 1. Flow of seed material in National Genebank from acquisition to storage.

and maintain viability of the accessions above 85%. Therefore, approximately 10% of the accessions kept in the long-term storage are randomly monitored periodically, after every 10 years. If on test it is found that viability has fallen below 85% a request is sent to the appropriate site to regenerate the accession for replacement in the base collection (Dhillon and Saxena, 2003). Fig. 1 depicts the complete flow of seed material from acquisition to storage.

2.2. Data classification

Seed bank data have been classified into four major categories—genebank management data, passport data, taxonomy data and address data.

2.2.1. Genebank management data

The processing of accessions received for storage in the genebank generates the management information and the parameters on which the information is recorded are the management descriptors. A unique batch number is automatically generated for germplasm received from a particular stakeholder that specifies the total number of records to be updated which may belong to single or multiple crops. Information on date of receipt, number of accessions and the person responsible for the processing is also included. Once the crop name is recorded the corresponding taxonomic details are automatically selected from the linked Taxonomic table (discussed below). Details of each accession are recorded including the national identity, donor/collector identity, moisture content and the viability test value with provision for recurrent recording after each monitoring cycle. The biological status of each accession such as wild, weedy, traditional cultivar/landrace, genetic stock, breeder's line, mutant, advanced/released cultivar, or other (with elaboration in remarks field), is also recorded. After seed processing, if the accession conforms to genebank standards, the location for storage in the genebank indicating module (from 1 to 10), rack (from A to Z), shelf (from 1 to 8) and basket numbers (in continuation) is recorded. The GBIMS also maintains an inventory of rejected accessions along with the reason for rejection (low viability, possible infection and low seed quantity). The GBIMS can also print labels for the packets to be kept in the genebank. The entity relationship (ER) diagram of GBIMS is shown in Fig. 2. In the ER diagram, rectangles denote an entity, the diamond shape shows a relationship and the ellipse depicts attributes. Further, the letter "N" denotes many and "1" denotes one, in a one-to-many relationship between different entities. For example, as "institute" (entity) having main attributes institute name, district and city "belongs to" (relationship) many seed collectors (entity).

2.2.2. Passport data

The information accompanying an accession is called the passport information. This pertains to the complete collection details such as place (district, state, village), name of the donor and other location specific details. The NBPGR has already developed a separate database for the management of passport information and GBIMS has been linked to this information system.

2.2.3. Taxonomy data

The taxonomic data provides the overall organization for germplasm accessions in the GBIMS. The purpose of the taxonomic tables is to provide correct scientific names for the crops conserved in the national genebank. The taxonomic table includes scientific names for 831 generic and 1187 species or infraspecies corresponding with common names of the crops and is automatically updated if a new species is conserved. The scientific names are verified, in accordance with the international rules of botanical nomenclature, using available taxonomic literature and consultations with taxonomic specialists. Accepted taxonomic records are searchable at the level of family, genus or species.

2.2.4. Address data

The addresses of national and international organizations with complete details are stored with a unique code, which is used for storing the information about the donors of the germplasm. This information also includes the details of the contact person and the other concerned employees in a particular organization. The organizations have been further categorized as ICAR, SAUs, Governmental, Non-Governmental, Private, International and CGIAR.

3. Results and discussions

The GBIMS is a user-friendly system with no specific training requirement and is being effectively used by the curators at National Genebank for day-to-day management. It helps in effective data capture and its retrieval in the desired format. It has been customized so that it can not only respond to specific requirements in terms of querying, sharing and reporting, but also has the capacity to be expanded in future when additional needs arise.

Due to the reason of involvement of composite and overlapping keys in GBIMS, after completing the activity of identifying attributes, their aggregations and associations, an intuitive normalization of relationships to third normal form was used (Elmasri and Navathe, 2006). Provisions have been made for built-in database level constraints and client level validation of attributes and the master tables for information such as taxonomic identity, addresses of the collaborating organizations, biological status. Therefore, in mandatory fields null values are not allowed and only

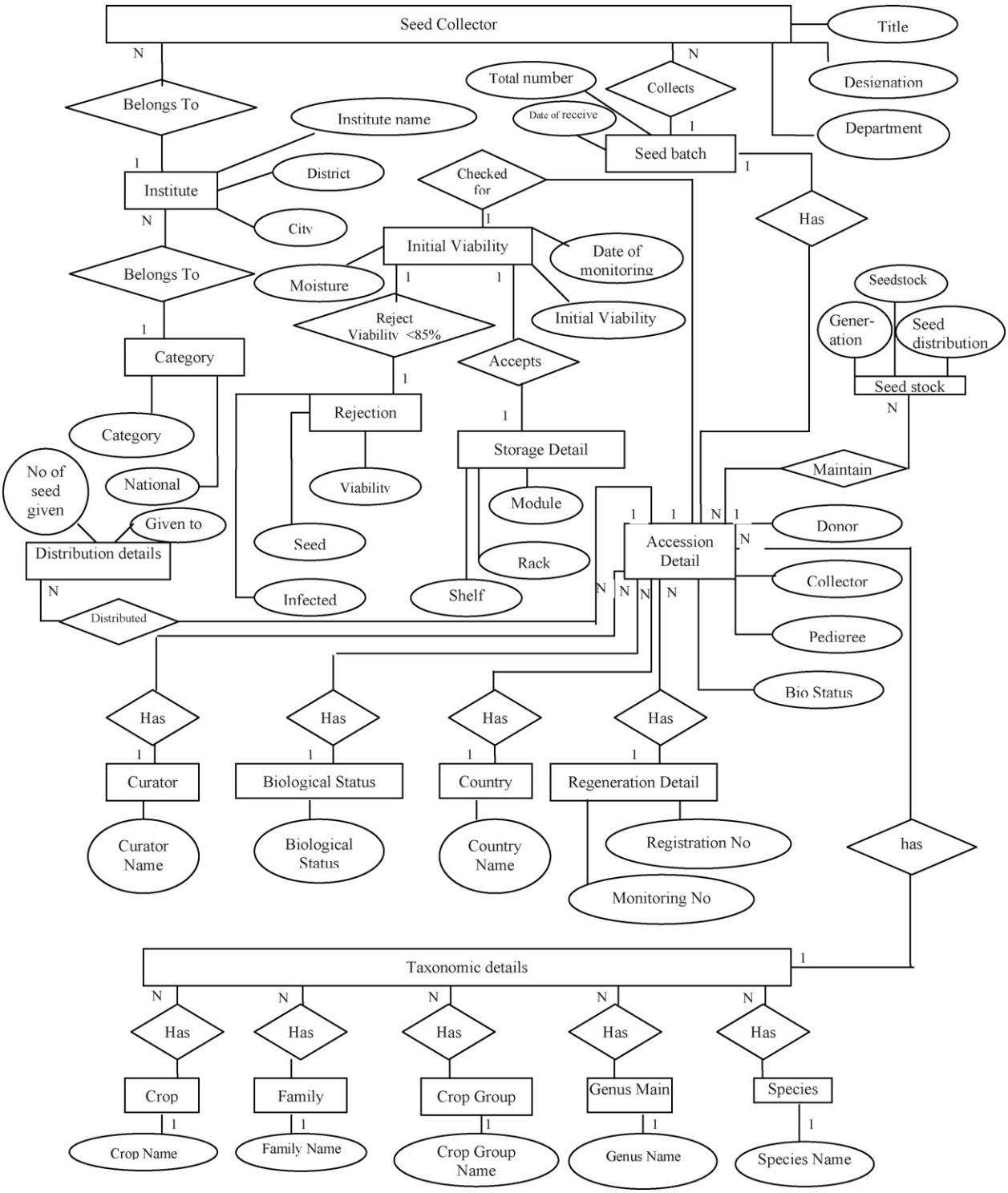
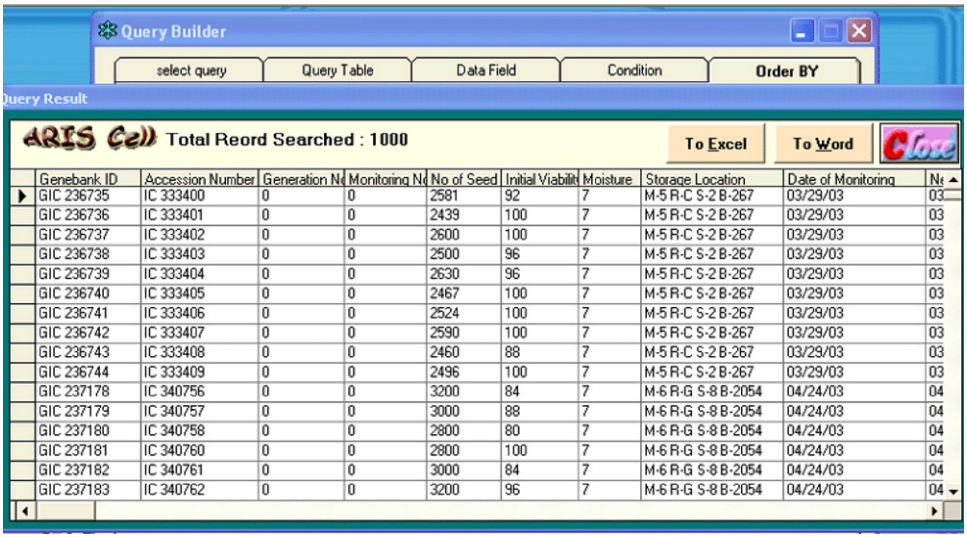


Fig. 2. Entity-relation diagram.



Genebank ID	Accession Number	Generation No	Monitoring No	No of Seed	Initial Viability	Moisture	Storage Location	Date of Monitoring	No
GIC 236735	IC 333400	0	0	2581	92	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236736	IC 333401	0	0	2439	100	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236737	IC 333402	0	0	2600	100	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236738	IC 333403	0	0	2500	96	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236739	IC 333404	0	0	2630	96	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236740	IC 333405	0	0	2467	100	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236741	IC 333406	0	0	2524	100	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236742	IC 333407	0	0	2590	100	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236743	IC 333408	0	0	2460	88	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 236744	IC 333409	0	0	2496	100	7	M-5 R-C S-2 B-267	03/29/03	03
GIC 237178	IC 340756	0	0	3200	84	7	M-6 R-G S-8 B-2054	04/24/03	04
GIC 237179	IC 340757	0	0	3000	88	7	M-6 R-G S-8 B-2054	04/24/03	04
GIC 237180	IC 340758	0	0	2800	80	7	M-6 R-G S-8 B-2054	04/24/03	04
GIC 237181	IC 340760	0	0	2800	100	7	M-6 R-G S-8 B-2054	04/24/03	04
GIC 237182	IC 340761	0	0	3000	84	7	M-6 R-G S-8 B-2054	04/24/03	04
GIC 237183	IC 340762	0	0	3200	96	7	M-6 R-G S-8 B-2054	04/24/03	04

Fig. 3. File export facility to MS-Excel.

unique values are permitted in key fields such as the national identity at the time of data entry. This helps in maintaining fidelity of information and minimizes the time and effort required for validation.

The batch-wise log generated helps in monitoring the day-to-day status of data entry and processing of germplasm received. The reports on the accepted and rejected accessions help in keeping a complete record of material received. Options to export these results in various file formats such as MS-Excel (Fig. 3) and MS-Word have also been added and the system supports activities such as acknowledgment for donor/collecting institute and automatic label printing with specific information for pasting on packets to be stored in the genebank.

The GBIMS helps the curator in proper management of accessions. A curator is able to maintain the viability record of each accession over a period of time. For periodic monitoring of the accessions the database prompts the curator 6 months in advance and maintains a record of the result of each monitoring. The curator can also generate a list of accessions in need of regeneration, that is, where the germination percentage and or seed quantity has been recorded below acceptable limit. The linkage of the management table with passport information helps the curator to take a correct decision for regeneration of the sample in the appropriate agro-climate region.

A separate menu option for querying has been included in GBIMS. The curators and research managers with access permissions can draft customized queries on any of the chosen attributes. Information can be generated on various fields such as indigenous or exotic accessions in the collection, accessions in a particular crop stored between specific periods and accessions received from a particular institution. Analysis of information on collections conserved in the national genebank and its linkage with passport information table helps in the planning of future explorations and acquisition of germplasm to fill identified gaps in the collection. The GBIMS thus assists policy makers in deciding on current resource utilization, in reducing the duplication of effort and in developing vision for future directions.

For security and reliability of data in the GBIMS database, three levels of access have been provided. The highest level of access is provided to systems administrators with permission to perform any operation on information and database structure. The second level of access is provided to the data entry operators and the curators who can enter data and update specified fields. More than one curator can update records in the same batch but not the same record (for a particular accession) simultaneously, thus ensuring that conflicting data entry in the same record is avoided. The third level of access has been provided to the research managers and general user scientists at NBPGR who can read information and conduct customized queries.

4. Conclusions

The exact aim of this software was to design and develop a management information system for genebank data at NBPGR to reduce paper record and filing work. The system does manage to achieve quick and efficient processing

of data, provision for multi-user data update/entry, generation of quick and well-formatted reports, proper storage of data, etc. and this enhances the performance of the curators. The web-based query system has been provided on the NBPGR intranet through which an user can access specific information on particular accession(s) conserved in the National Genebank. The design and description mentioned above may be adopted by various organizations working in the area of PGR.

Acknowledgements

The authors would like to thank Dr. B.S. Dhillon, Dr. A.K. Singh, Mr. Rajeev Gambhir and Ms. Nirmala Dabral, at National Genebank, NBPGR, New Delhi for their critical input, and discussion throughout the development of this GBIMS software.

References

CBD 1993. Convention on Biological Diversity. Full text available at http://www.biodiv.org/convention/articles.asp.

Dhillon, B.S., Saxena, S., 2003. Plant Genetic Resources in SAARC Countries, SAIC Publication, India. Director, SAARC Agricultural Information Centre Bangladesh, pp. 241–296.

Elmasri, R., Navathe, B.S., 2006. Fundamentals of Database Systems. Addison-Wesley Publishing Company, pp. 335–379.

FAO, 1996. Global for the conservation and sustainable utilization of plant genetic resources for food and agriculture: leipzig declaration. In: Proceedings of the International Technical Conference on Plant Genetic Resources, Leipzig, Germany, June 17–23. Food and Agriculture Organization, Rome, Italy, p. 63.

FAO, 1997. The State of the World's Plant Genetic Resources for Food and Agriculture. Food and Agriculture Organization of the United Nations, Rome, Italy, p. 511.

Lipman, E., Jongen, M.W.M., Hintum van, Th.J.L., Gass, T., Maggioni, L., 1997. Central Crop Databases: Tools for Plant Genetic Resources Management. International Plant Genetic Resources Institute/CGN, Rome, Italy/Wageningen, the Netherlands.