DINTERIA NO. 23:55-59 - WINDHOEK, NAMIBIA - OCTOBER 1992

Documentation and use of data in genebank management

T.J. Ruredzo*

Introduction

The utilization of plant genetic resources can only be achieved if plant breeders and other plant scientists are aware of the germplasm that is available from genebanks. Genebanks collect a lot of information about their samples which is invaluable to plant breeders and for decision making and management. The genebanks should maintain efficient documentation and information systems in order to ensure that the end users know about the material that is available.

Before 1980, genebank data could only be managed manually or on mainframe and minicomputers using custom-made programmes. With the advances in computer technology, detailed information about genebank samples can now be managed on microcomputers using a variety of database management systems. This is less costly and requires basic computer programming skills. Emphasis has shifted from the development of computer programmes for handling genebank data to developing efficient information systems using readily available database management systems. Plant scientists can now manage data in genebanks, reinforcing the relationship between the data manager and the end users of the information.

Data management systems which can relate different databases through a common field are recommended for the management of genetic resources data. Databases containing specific information can be managed separately and then linked when necessary to combine information from different databases.

This paper explores collection, documentation and use of data in genebank management.

* Documentation Officer, SRGB, P/Bag CH6, ZA-153 02, Lusaka, Zambia

The End User

The most important consideration in genebank data management is the end user of the data. In plant genetic resources, the users of the genebank data are primarily plant breeders, agricultural scientists, genebank staff and other plant scientists. Breeders are mainly interested in highly heritable characters such as number of rows of grain per inflorescence and head size which can determine yield; head shape, hairiness and presence or absence of awns which can determine pest tolerance and colour whichcan indicate palatability or desirability. Agronomists are interested in matching agro-climatic conditions. Data such as latitude, rainfall patterns and soil (type, colour and pH) are important. These and other scientists are also interested in the usage of the plant (food, feed, medicine, fuelwood, timber etc.), its local name, history and diseases.

Genebank staff are interested in the history of each sample, the management of the sample, the requesters of the sample, germplasm collections carried out in their mandate areas and the production of informative catalogues to solicit requesters. The data managers of the genebanks need to know the data they are keeping in the different databases and the meaning of the descriptors they are using.

The requirements of breeders, agronomists and other plant scientists determine the type of data that is collected and recorded in a genebank. It is therefore important to consult the users of the data and other data managers to ensure that a comprehensive information system is developed for genebank data management.

Data collection

Genebank data is collected by various personnel in and outside the genebank. The data is collected as objectively as possible. Where possible, standard equipment and aids e.g. colour charts for soil and seed colour, portable pH kits for soil pH, global positioning system (GPS) units for coordinates etc., are used to ensure objectivity. Rulers, balances, timers, thermometers etc. are also used for accurate measurements.

The first set of data is usually provided by the collector or donor who provides the passport data of the sample. The passport data is recorded at the collection site and provides information about the sample. This includes the collection site, the agricultural practices where it was grown, its uses, local name and any other information provided by local people and that observed by the collector. The collector fills in a comprehensive collection form at the collection site. Most genebanks use the IBPGR collection form with additional information collected to suit their specific needs or environment. End users and genebank staff are consulted for possible additional information.

Information about a sample also originates from the genebank. Management data is provided and used by genebank staff in the management of each sample. This includes information on various genebank activities e.g. germination tests, distribution, storage, multiplication, rejuvenation etc. This data is recorded on forms or in field books which are used for future reference.

Genebank and other scientists also supply characterization and evaluation data which gives highly heritable characters and environment dependent agronomic characters respectively. This information is increased by 'feedback' information provided by requesters who characterize, evaluate or use genebank samples in research. Data managers keep adequate information on requesters and periodically ask for any information about the samples which might be useful to the genebank and its users.

Data about other genebanks keeping duplicate samples and any publications involving a particular sample is useful to the genebank and its users. The data manager also keeps comprehensive information about the databases in which the information is electronically stored and any explanatory notes to the databases.

Data Management

Genebank data is usually recorded on paper before it is entered into an electronic database. This is encouraged since it provides the original record which is kept safely for future reference. Where electronic data collection units are used, the recorder's data is always printed out, verified by the recorder and filed systematically.

Database management systems which relate different databases are encouraged for data management in genebanks. In these management systems, several databases are created each holding data which is available at the same time from one or two activities. For example, the management data mentioned above will contain a lot of information some of which will not be available for all the samples in the genebank. Management data is logically divided into databases containing information on: active collection, base collection, duplicate collection, germination tests, multiplication, requests etc. The databases are related by including the accession number field in all of them. This permits the linking of two or more databases to retrieve or process data.

Related small databases have advantages over one large database. Small databases are easier to sort and faster to retrieve than larger databases. New databases can be created when new sets of information become available. Since the data required in each database is available at the same time, there will be very few empty fields in the databases. For example, in a single large database, the absence of multiplication data for some accessions will result in several empty fields which cause problems during data processing. In the relational database system the accessions without multiplication data will not be recorded in the multiplication database. Each database is there-

fore a selfcontained module in a data management system which can access and link several databases through a common field such as the accession number.

Most database software requires the identification of different fields according to the types of data expected. Fields assigned character data types will accept all entries as characters; those assigned numeric will only accept number; those assigned logical will only accept T or F or Y or N for true, false, yes or no respectively. This greatly enhances the chances of entering the correct data into the fields. Most databases also allow limits to be set for the type of data expected. For example, in the field for 1000 seed weight in a finger millet descriptor database a limit can be set to make sure that 1000 g will not be accepted. These facilities should be fully utilized to make sure that incorrect data is checked and novel characters identified at an early stage.

Once the necessary databases and their fields are established, the data management system can be developed for automatic processing of frequently required data and ease of use by genebank personnel and other users.

Use of data

Genebank data is mostly used to aid decision making and efficient running of activities in the genebank. A good data management system facilitates an efficient information system, registration and exchange of data and allows rapid accessioning of new samples and location of samples in storage. It also allows monitoring of quantity and quality of samples, therefore guiding regeneration and distribution activities.

The genebank uses its data to identify duplicate samples and gaps in the collection and to plan collection activities. Genebanks also use their data to rapidly match their collections with the requirements of requesters and to monitor genebank efficiency.

In order to promote utilization of germplasm, genebanks produce comprehensive catalogues of the material available to requesters. Samples are also accompanied by relevant information e.g. origin and characteristics of the accession and technical data on germination procedure etc. Requests and feedback from germplasm recipients are monitored using genebank data. Evaluation and other data from the genebank is also provided to other genebanks holding duplicate samples and information about the samples is sought from them.

Concluding remarks

Efficient data managment is a prerequisite to the utilization of plant genetic resources. The end user should be involved in determining the data that is required. Data should be collected objectively using internationally recognized descriptors and the

original forms should be filed for reference. Data management systems should be easy to use and should be able to accommodate new categories of information.

Further reading

BLIXT, S. & WILLIAMS, J.T. (*eds*). 1982. *Documentation of Genetic Resources: A Model.* Rome: IBPGR Secretariat.

KONOPKA, J. & HANSON, J. (*eds*). 1985. *Documentation of Genetic Resources: Information Handling Systems For Genebank Management.* Rome: IBPGR Secretariat.

FORD-LLOYD, B. & JACKSON, M. 1986. *Plant Genetic Resources: an introduction to their conservation and use.* London: Edward Arnold.