

The Millennium Seed Bank Project's Specimen and Taxon Databases



Jill Bone

Seed Conservation Department, Royal Botanic Gardens, Kew, Wakehurst Place, Ardingly, West Sussex RH17 6TN, UK

Rob Turner

Information Services Department, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK

John Tweddle

Seed Conservation Department, Royal Botanic Gardens, Kew, Wakehurst Place, Ardingly, West Sussex RH17 6TN, UK

Summary

One of the largest international *ex situ* plant conservation projects ever undertaken, the Millennium Seed Bank Project (MSBP) aims to safeguard over 24,000 plant species world-wide against extinction and to secure the future of the UK's native flowering plants. The MSB is situated at Kew's satellite garden at Wakehurst Place in Sussex (about 50 km from Kew).

The MSBP has two main databases. The specimen database is the Seed Bank Database (SBD) that holds information about every seed sample held within the Millennium Seed Bank (MSB). The Seed Information Database (SID) is taxon-based and synthesises information contained in SBD with other seed research data (both from the Project and in the literature). These taxonomically-based summaries are available on-line and are of use to conservationists and to scientists in a wide range of other disciplines.

The Seed Bank Database

1. Technical Background

The Seed Bank Database is a client server system that holds collection, processing, storage, maintenance and seed distribution data on the MSB collections. In late July 2002, the SBD held 15,653 records relating to 7,139 wild species.

The client software runs on desktop personal computers (PCs) running the Microsoft Windows operating system. The data is stored in a Sybase database which runs on a Sun Solaris server at Kew. The Windows client was designed using object-oriented techniques and implemented using the Microsoft Visual Basic programming tool. The use of Visual Basic allowed rapid software development within the Windows environment. The data is accessed via TCP/IP (Transmission Control Protocol/Internet Protocol) using ODBC (Open Database Communication) installed on the client PCs. The code is written in Forms and Class modules.

Sybase is a relational database, which allows efficient storage of data. The database is queried and updated using stored procedures, which makes editing and querying the database efficient and minimises network traffic between the two sites.

The client software makes use of the Microsoft Windows visual interface. Dropdown lists are used for common fields and the system can be either keyboard or mouse driven.

2. Data and Functionality

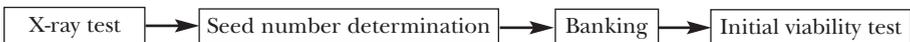
The data for each accession is organised under two main headings: donation data and processing data. Donation data records the field (collecting or 'passport') data plus the plant identification. The latter is the only element of crop genetic resources 'characterisation' noted on the MSB material. An attempt has been made to make the field data as objective as possible. Indeed, due to the potential longevity of the seed collections, it is essential that the related data will be meaningful to scientists in coming years. To this end, Geographical Positioning System (GPS) data accurately pinpoints the location of collection. Such accurately recorded locations immediately lend themselves to use within a Geographical Information System (GIS) thereby allowing linkage to climatic and other data. Processing data records the results of procedures carried out on the seed collection once it has arrived at Wakehurst Place. The sequential nature of this process makes it relatively easy to monitor where a collection is in the system. One function of the MSB is that, where agreements allow, samples are made available to *bona fide* research under a material supply agreement. Such seed orders are handled by the SBD.

2.1. Donation data

The system caters for a wide range of data recorded at the time of seed collection. This ranges from the essential fields of, for example, date collected, name of collector and geographical location to habitat and plant descriptors, sampling data and site details. All plant names are checked against a live version of 'Vascular Plant Families and Genera' (Brummitt, 1992) and, as an added precaution, against lists of species scheduled as having UK quarantine status or protection under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Accuracy of data entry is important at this point and particularly with respect to collectors' names, the plant names assigned in the field, and field collection numbers, as subsequent identification of a specific seed collection amongst the MSBP partners relies on this information. A six digit serial number (plus check digit) is given for the identification of every seed-lot arriving for storage in the MSB. The categories of donation data are noted in Table 18.1.

2.2. Processing data

The processing stages (see Terry *et al.*, 2003 – Chapter 17) that yield key processing data are as follows:



At each of these stages of processing of the seed collection, data is recorded on the system. These include: the conditions used and results of the X-ray; the sample seed weights used in quantity determination; the bank location and type and number of containers; and all of the germination conditions used in the viability tests and the results obtained. In addition, the confirmation of

Table 18.1 Donation data field categories

Categories	Notes
Material Agreements	Details the Access & Benefit Sharing Agreement (ABSA) associated with the accession.
Donor and Collectors	
Distribution Policy	Records what restrictions may apply to the distribution of the seeds from the MSB.
Geographical Data	Includes country, province, specific location, latitude, longitude and altitude (see Figure 18.1).
Environmental Data	Includes habitat type and associated species.
Site Notes	Notes on landform, aspect, geology and soil.
Plant Name	Identification by the collector, followed by any further verification details and plant descriptions.
Sampling Data	Number of plants sampled/found in the population and the approximate area sampled.

Millennium Seed Bank - Donation data for seeds related to serial number 0144027

File General Donor/Colls. **Location** Natural Habitat Identification Sample Data

Geographical Data

Location:

Local Situation:

Lat. and Long.

	Orient	Deg.	Min.	Secs	Accuracy
Latitude	N	33	50	48	Seconds
Longitude	E	035	55	24	Seconds

Meth. Det:

GPS Datum:

GPS Error (Metres):

Location Precision Diameter (km.):

Altitude

Altitude (metres):

Method Altitude Determined:

Map Reference

Publisher:

Series:

Scale: Date:

Map Co-ords:

Data Checked

Figure 18.1 A sample processing data screen from the SBD.

Table 18.2 Processing data field categories

Categories	Notes
X-ray data	Conditions used and number of seeds full/empty/infested (see Figure 18.2).
Seed weight	Number of seeds per sample weight and standard error (see Figure 18.2).
Seed quantity	Original quantity (usually derived from seed weight data) and current quantity (see Figure 18.2).
Germination test data	Date tested, conditions used and result (see Figure 18.3).
Moisture content test data	Date tested, conditions used and result.
Tetrazolium test data	Date tested, conditions used and result.
Duplicate locations	Location (where collection has been split).
Banking data	Date banked, container type and number and location in bank ¹ .
Voucher data	Wild or cultivated pressed specimen, herbarium notes and location of specimen(s).
Glasshouse data	Number of seedlings and date sent to glasshouse, reason for growing, seed harvesting details and stock check (see Figure 18.4).

¹ A stylised representation of the MSB cold-rooms allows individual drawers within each room to be allocated to certain types of container depending upon current need. The system then assigns locations to the collections awaiting banking. It is possible, therefore, to see how much space is left in any given drawer, stack or room and plan further commissioning of rooms and stock purchase.

identification of the collection is recorded, as are the results of moisture content and tetrazolium tests. There are also fields for noting the location and details of herbarium specimens and the purpose and outcomes of sending plants to be grown on.

The processing data can be used to evaluate the status of a collection or batch of collections through the various stages and the results employed to produce a quality assessment of the collections. The processing data field categories are summarised in Table 18.2.

2.3. Seed order management

One of the functions of the MSB is to distribute seed samples (from collections that meet certain criteria) for *bona fide* research purposes. The SBD maintains a list of ‘customers’ and, through the seed ordering programme, produces all the information and printouts required to fulfil this commitment. It is possible to keep records on what a customer has ordered and which customer has ordered a certain accession.

Millennium Seed Bank - Processing Data for Serial number 0144027

File General **Calculations** Testing Bank/Distribution Growing/Progeny Notes

X-Ray Data

Date X-rayed: 12/03/2001 Kv: mA: Seconds: Cut Test:

No. Okay: 100 + No. Infested: 0 + No. Empty: 0 = No. Used: 100 No. Part Filled: X-Ray Data Checked

Weight Information

Sample Size: 50 Precision: 7

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Remainder	
Seed Weight (g)	0.4345	0.4481	0.4382	0.4321	0.4677	349.82	
Average Seed Weight	0.44412		Std. Error	0.0538995		1000 Seed Weight	8.8824000

Seed Quantity

Method Seed Number Determined: By hand Seed Quantity Data Checked

Original seed Quantity: 29496 Current Seed Quantity: 29196 Adjusted Seed Quantity: 29196

Average Number of Seeds/Fruit: Quantity sent to backup: 150

Date Counted:

Data Checked

Figure 18.2 A sample processing data screen from the SBD.

3. Searching the SBD: Use of the Data Analysis and Reporting Tool (DART)

Sybase data can be linked to tables in a Microsoft Access database. This data can then be analysed using Access queries and easily exported to other Office packages such as Excel and Word. The DART is used for a wide range of purposes. A few examples are: the production of statistics over the entire database; quality assessment of collections from a specific country; planning of certain work projects; counting the number of species; and tracking down a particular accession or species.

Figure 18.3 A sample germination test data screen from the SBD.

4. Management of the SBD

A Data Supervisor manages the SBD. Amendment of some parts of the database, such as the pick-lists, needs to be carefully controlled. Consequently, the supervisor controls access levels to different parts of the database by staff. As part of the training of staff, they also check the inputted data. This is facilitated by an approval system prior to incorporation of the input data into the database. Furthermore, the supervisor can make certain records confidential, thereby hiding data on screen and preventing it from being printed out. Finally, a key role of this post is to ensure the maintenance of the linkage between the conditions within ABSAs pertaining to each collection and that collection's data record. This is especially important with respect to seed distribution.

Millennium Seed Bank - Glasshouse Data for serial number 0142610

General Data | Harvesting | Stockcheck | Outcomes

General Information

Where grown: RBG Kew, Wakehurst Place - Glasshouse 1 [Select where grown]

Isolation tech.: None. Assume open pollinated.

Chemical analysis? [Plant Lists...]

Dates Sent to glasshouse

First date: 12/09/2000 | Last date: 12/09/2000 | Total no. seedlings: 10

Purpose(s)

Purpose: [] [Add] [Remove] | Voucher: two sites

[Update] [Cancel] [Close]

Figure 18.4 A sample glasshouse data screen from the SBD.

The Seed Information Database

As a product of the scale of its activities, the MSB Project has a unique opportunity to obtain and store large amounts of seed biology information, both from collections and published sources. While the primary intention is to analyse these data for predictive patterns that support seed conservation operations, it is likely that a wide variety of users will find the information valuable for many purposes. To aid both in-house analysis and dissemination to outside users, a Seed Information Database (SID) is being developed. It is one of RBG Kew's strategic databases and will ultimately form part of the institute's electronic Plant Information Centre (ePIC) project, in which species name will be a common field linking a number of other taxonomic, bibliographic and specimen databases (e.g., International Plant Names Index, Angiosperm DNA C-Values).

The SID is a taxon-based database and is available on-line via the RBG Kew website (<http://www.rbgekew.org.uk/data/sid/>). The database is currently updated twice-yearly and although in its early stages already contains storage behaviour data for >8,500 species, seed weights for >10,600 species and germination protocols and seed dispersal mechanisms for *ca* 3,800 and 3,000 species, respectively (Tweddle *et al.*, 2002). At present, these data represent a total of 283 APG families (APG, 1998), and coverage will become more complete over time. The SID uses Java Servlet technology to access a Sybase Adaptive Server relational database via Java Database Connectivity (JDBC). Searching of the database can be done at a variety of taxonomic levels, from clade down to species. It is also possible to choose which subsets of the data are to be queried (e.g., germination data only, germination and seed weight data, all datasets), and searches can be limited to a particular seed storage behaviour type. It is hoped to allow boolean-type searching combining several seed characters in future releases. Upon searching, the initial results screen displays summary information for all taxa fitting the search parameters. Selecting a name from this list then displays a species dossier which details all of the information that the SID contains for that taxon (including higher taxonomy) in a fully referenced format.

When fully developed it is envisaged that the SID will contain a wide range of seed biological characteristics in addition to the above, including anatomical and morphological features of selected seeds, dormancy types, and chemical constituents (e.g., oil and protein contents). Features of adult plant ecology that facilitate the interpretation of the seed biological data, including life-form and habitat preferences, will also be incorporated. Whilst the database is thus still very much in its developmental stages, it is intended that the product will become a valuable tool for seed biological research and enquiry, and will greatly aid future seed conservation activities. For further information on the project, visit the website noted above.

References

- APG (Angiosperm Phylogeny Group) (1998). An ordinal classification for the flowering plants. *Annals of the Missouri Botanical Garden* **85**: 531–553.
- Brummitt, R.K. (1992). *Vascular plant families and genera*. Royal Botanic Gardens, Kew, UK.
- Terry, J., Probert, R.J. and Linington, S.H. (2003). Processing and maintenance of the Millennium Seed Bank Collections, pp. 307–325. In: R.D. Smith, J.B. Dickie, S.H. Linington, H.W. Pritchard and R.J. Probert (eds). *Seed Conservation: turning science into practice*. Royal Botanic Gardens, Kew, UK.
- Tweddle, J.C., Turner, R.M., and Dickie, J.B. (2002). *Seed Information Database* (release 3.0, Jul. 2002). Available online: <http://www.rbgekew.org.uk/data/sid> (01/08/2002).