THE MILLENIUM SEED BANK PARTNERSHIP
DATA WAREHOUSE

A NEW ONLINE DATA PORTAL FOR SEED ACCESSIONS, GERMINATION TEST INFORMATION AND IMAGES

DOCUMENTATION FOR USERS

Tim Pearce

Version 1.0  Feb 7th 2012
CONTENTS

INTRODUCTION 1
HOW TO USE THIS GUIDE 2
ACTIVE DEVELOPMENT AREAS 2
MENU PAGES EXPLAINED 3
  THE HOME PAGE 3
  THE SEARCH PAGE 3
  THE STATISTICS PAGE 4
  THE RESOURCES PAGE 5
  THE LOGIN PAGE 6
  THE HELP PAGE 6
AN OVERVIEW OF RESULTS PAGES 7
GRID VIEW FUNCTIONS 9
  TAGGING 9
  SORTING 9
SELECTING VISIBLE FIELDS 10
MEMO DATA AND LONG FIELDS 11
VIEW ALL DATA FOR SELECTED RECORD 12
SUMMARISE FIELD DATA 13
FILTERING 14
SHOW RECORDS WITH IMAGES 15
DOWNLOAD RESULTS 16
GRID VIEW LINKS TO EXTERNAL WEBSITES 17
REPORTS VIEW 19
PRINTING AND EXPORTING REPORTS 20
RESULTS PAGES EXPLAINED 21
  SUMMARY RESULTS PAGE 21
  FAMILY & GENUS RESULTS PAGES 21
  SPECIES RESULTS PAGE 22
  SEED ACCESSIONS RESULTS PAGE 24
  SEED TESTS RESULTS PAGE 26
  MAPPING RESULTS PAGE 28
  IMAGES RESULTS PAGE 29
SETTING YOUR SCREEN OPTIONS 30
INTRODUCTION

This Millennium Seed Bank Partnership Data Warehouse, one of many BRAHMS Online databases, is currently held at:

http://herbaria.plants.ox.ac.uk/bol/msbp

Information about the BRAHMS project, BRAHMS Online and further help files regarding the use of the BRAHMS software can be found at:

http://www.brahmsonline.com

In developing a use case for this database, the Millennium Seed Bank management team identified a need to establish an online data-sharing tool (The “Data Warehouse”). This would facilitate the tracking of key measures for the MSBP (e.g. numbers of species secure in seed banks across the MSBP), would aim to assist partners in setting collection targets, and also afford an opportunity to share data pertinent to germination tests, processing notes and images.

The purpose of the Data Warehouse is therefore to:

“....provide a tool for the MSBP Network to support and facilitate the prioritisation of seed collecting, display project progress and assist reporting by MSBP project managers (Kew) and other managers across the Partnership.”

The design of the database required that the Data Warehouse should:

“...support the compilation of seed accession, calculations and germination test data from a range of MSBP Partners. The data should be uploaded and refreshed remotely by Partners and the compiled data set will offer various tools for querying, analysis and download of data. Statistics of collection and species totals with properties will be offered through a “dashboard” of graphics.”

The data available through the Data Warehouse will include:

- All material thus far duplicated to the Millennium Seed Bank;
- All material collected via MSBP support but NOT duplicated to MSB; and
- Any other material potentially available from MSBP Partners which meet MSBP quality standards.

The data will fall into three broad categories:

- Geo-referenced passport seed accessions with seed counts;
- Germination test results where available; or
- Images of Tetrazolium cut tests and/or X-rays where available.

The complete format and scope of data that we are able to accept in the Data Warehouse is defined by the RBG Kew Seed Bank Data Exchange Schema which can be found as a downloadable .xsd file at:


The data schema is also available through the Resources tab in the Data Warehouse.
HOW TO USE THIS HELP GUIDE

This guide is largely pictorial and uses screen shots from the Data Warehouse. It aims to provide guidance to those wishing to use the Millennium Seed Bank Data Warehouse.

Firstly we will look at the Homepage and the different individual pages that can be accessed directly from there (SEARCH, STATS, RESOURCES, LOGIN and HELP).

A description of how to undertake searches of seed accessions is presented. Searches undertaken on the database will deliver a series of up to 8 Results Pages (SUMMARY, FAMILY, GENUS, SPECIES, SEED, SEED TESTS, MAP and IMAGES).

Each Results Page can be viewed either in Grid View (like a spreadsheet) or Reports View (text).

An overview of the different functions and tools available on each of these views is given. This is followed by an in-depth explanation of each of the Results Pages in turn.

Finally the guide gives some useful tips on how to correctly set your particular computer screen to best suit your use of the Data Warehouse.

We have used “call out boxes” and actual screen shots throughout.

ACTIVE DEVELOPMENT AREAS

As with many collaborative initiatives, we will always be looking for areas of development and new tools to assist users get the most out of the Data Warehouse. Here are some areas where we are actively developing the functionality of the Data Warehouse:

Search form:
We will be expanding the operators in the main search form based on user feedback. The search form is likely to be tabbed as the primary search option expand and include tree views. It will be available as a floating form from all data screens as another toolbar option.

Filtering on grid view:
Filtering is a powerful tool for analysing results. We are actively extending filter functions. Filtering and searching are potentially to be merged as they operate in a similar way.

Stats Page:
Extended statistics will be made available.

Login:
This function will ensure access to data is password protected, users registered and monitored by RBG Kew.

Images:
Images can be uploaded currently and searches will return these to the Images Results Page. However we are currently considering the best way to display these images on the Results Page.
MENU PAGES EXPLAINED

THE HOME PAGE

The Home page presents the opening, introductory page of the Data Warehouse. This page provides a very brief introduction to the MSB Partnership and the purpose of the Data Warehouse. It is not meant to duplicate the many excellent MSBP websites available but does contain hyperlinks to the www.kew.org/msbp website. These pages contain useful resources and information as well as providing access to the many partner websites available.

In addition, a quick taxonomic search (on genus + species, or parts thereof) of seed accessions and tests can be performed using the Quicksearch box in the bottom left hand corner.

Many of the images on the home page are available as high resolution versions. You can access these when the “hand pointer” appears on hovering over the image.

THE SEARCH PAGE

Searches on seed accessions in the Data Warehouse can be designed and performed using main search form is found through the Search tab. Perform searches of seed accessions using taxon, collector, country and major area. The matches in the right hand selection margin are generated using
auto-suggest. Longer, more complicated names can therefore be selected after entering a few letters into the search box rather than having to type the full search text. Type a few letters into the search box to bring up the corresponding list of names in the selection pane. For example, to find *Amphipterygium* simply type “amp” into the genus search box and select *Amphipterygium* from the selection pane.

Searches are compound, so clicking on the search button on the example below will retrieve all accessions of the genus *Acacia* which have been collected from Burkina Faso.

Select the delete icon “X” to remove the current entry in your search box. Alternatively use the “Clear All” button to clear all entries.

The “Restrict to..” tick boxes can be ignored when retrieving seed accessions.

Click on the “Search” button to perform the search.

![Search on any combination of taxon, collector name or geography options](image)

![Type a few letters into the country field and then select from the matched list](image)

![Remove entries by using this delete icon or the "Clear All" button below](image)

![The "restrict to" options can be disregarded when searching for seed accessions](image)

![After designing your search query, click on "Search" to deliver the results](image)

**THE STATISTICS PAGE**

One of the functions of the Data Warehouse is to undertake counts of all species that are effectively conserved in seed banks across the Millennium Seed Bank Partnership. Our intention is to use this page to report to funders on the progress against one of the headline targets for the MSBP, the effective conservation in MSBP Seed Banks of 25% of the world’s flora by 2020.

The statistics that are presented on the Statistics Page are calculated from all those collections currently shared through the Data Warehouse.

Complete BRAHMS Online databases potentially show separate tables for: TAXA, SEED, IMAGES, BOT RECORDS, SPECIMENS. Currently the Data Warehouse only contains Taxa, Seed data and some X-ray images so only three tables appear.
Use the “Show map” option to view all records currently held in the Data Warehouse on a map.

THE RESOURCES PAGE

Use this page to post documents and files of use to the Data Warehouse. Files can be uploaded by contributors to the Data Warehouse through the use of the BRAHMS WebConnect service.
Download files such as the Data Exchange Schema or this help file from the Resources Page.

THE LOGIN PAGE

The security of access to collections data is of paramount importance to RBG Kew. Despite the ever-growing plethora of data available online, we will initially permit access to this site through a password protected login process.

The Login function is currently being re-developed by the BRAHMS team in Oxford. Before the Data Warehouse is launched to an external audience, we will ensure that adequate function exists on the webpage to fully control access to data.

THE HELP PAGE

The Help Page may actually be removed from the site unless we convert this manual to a .chm Windows help file. This is currently still under discussion.
AN OVERVIEW OF RESULTS PAGES

This manual previously describes how to undertake a search on the database from the main search form. Searches display the results of the query in a number of Results Pages. These pages are accessible through tabs appearing under the main menu bar. Depending on the data category returned from the search, the tabs that appear will be any combination of:

- Summary
- Family
- Genus
- Species
- Botanical Records
- Specimens
- Seedbank Data
- Seed Germination Tests
- Mappable
- Images

The following tabs (or a subset of them) will appear at the top of each of the Results Pages. Use them to navigate through the different Results Pages.

Summary  Family  Genus  Species  Seed  Seed Tests  Map  Images

Searches will always display the Summary Results Page first by default.

In this example, a search has been undertaken for Acacia seed collections made in Burkina Faso. The summary results page shows that there is 1 family and 1 genus, 12 species, no botanical records nor specimens, 14 seed bank accessions, 29 germination tests, 14 mappable points and no images.

You will see that the only Results Page tabs available are those for which there are data in the database for the search you executed. Any search for which there are no data, the appropriate tabs will not appear.

Results Pages can be viewed either as Grids (default) or as Reports. Use the toggle buttons in the top right of each page to switch between these different types of view.
Use these toggle buttons to view the returned data as either a "grid" (like a spreadsheet) or as a "report".
GRID VIEW FUNCTIONS

Grid views offer flexibility for analysing the search results online using a variety of useful functions.

Tagging

This allows for individual records or groups of records in a grid to be tagged.

In this example, a search on all seed accessions collected from Georgia has been performed.

Viewing the Species Results Page and extending the “Records per page” to 100, the first cell with Amaranthaceae has been made active. All the Amaranthaceae in the results have been tagged using “Tag all like selection” and for ease of viewing, the tagged records have been highlighted using “Highlight tagged records”. Tagged records can be filtered (see filtering below), or downloaded rather than complete search results (see downloading below).

Sorting

In these examples, a search of the database of the family Rosaceae is viewed and sorted by genus name. Field data can be sorted in three ways:

Select a field by activating any cell in that field. Select the relevant button on the task bar to sort records.
Alternatively, click on the column header (toggle A-Z, Z-A).

Lastly, use the drop down menu on the field header to select the relevant sort option.

Selecting visible results

The default list of available fields is determined and set by the database owner via the BRAHMS WebConnect service. Any user however can add and/or remove any data field that is available in the database to or from the default view. Select the function button in the Seed Results Page to open a separate floating table of available fields. Using the same search on all Rosaceae as before, this example below has the field “Donor Organisation” added to the default view. The “Donor Org.” Field has been lengthened to display the data.
These user-defined field views are not retained once you navigate away from a particular Results Page.

Memo data and long field entries

Restricted field lengths in the grid view mean that some data may appear shortened (truncated). These data can be viewed in two ways.

Using the Seed Tests Results Page for the previous query, select individual cells and use the “View text for current selected cell” function to display a separate text box with the full field text contents.
Alternatively, these extended data entries can be seen by enabling this “text-wrap” function.

![Example screenshot of data entry with text-wrap function enabled]

Text wrap is a useful method to view cell contents.

“Zooming-in” on a selected record

A separate 2-column table displays all the available fields for the record with any available data.

View all data for selected record

This is often referred to as the “Zoom” function as it “zooms in” on all the data for a particular record. Only those data currently selected in the view (see “Selecting visible results” above) will be displayed in the floating zoom window. Keep the zoom window open and move between records, the contents of the zoom window will update automatically. In this example we have zoomed in on a single seed test from the previous query using the Seed Test Results Page grid view.
Summarise field data

The “Summarise” function is used throughout the BRAHMS software and is an extremely useful analysis function. The search results for any active field can be summarised in a separate 2-column table (Column Summary table). Data in the Column Summary table can be downloaded and further processed in MS Excel; a very efficient way to produce charts, graphs etc.

In this example a search has been carried out for all seed collections from Georgia.

With the Column Summary table open, different fields can be activated and the data will honour the selected field.

In this example, the Family field has been summarised for the original search. Whilst this is interesting and useful in itself, the summary data can usefully be exported to Excel and processed further.
Here the Family summary table from the previous example has been exported to Excel and charted to produce a pie-chart of the relative abundance of collections in the “top ten” families for Georgia.

**Filtering**

The filter function is accessed via two routes. Carry out a simple filter on the contents of any selected cell by clicking the filter button on the function tool bar. Only one field can be used at a time in this function. In this example a query on the genus Acacia has been carried out.

Alternatively, filters can be performed on all fields by selecting the drop down menu on the field header and entering the required filter value manually.

In this example we can filter out the genus Vernonia and then perform another complex filter using the form and setting the value at “>=10000”, quickly showing the collections of Vernonia in the database with large seed counts.
Show records with images

Where images have been added to the database and attached to particular species (e.g. field images) or seed collections (e.g. X-ray images). Use this specific filter function to filter out the records in your result that have images associated with them.

First carry out a search on all seed accessions from Kenya. View the results in the Seed Results Page grid view.
Download results

Grid view records can be downloaded to an MS Excel spreadsheet.

Download either all the records, or only the tagged records from any grid. If you have used a filter on the results grid then only those records filtered will be downloaded.

You can also download data appearing in any analysis window that you generate.

Using the query for all seed accessions from Kenya as before, open the Seed Results Page in grid view. Download all the records to an Excel table.
In the example below we have created a Family summary by using the “column summary” function. After downloading the results of the summary table to Excel we can easily produce a pie-chart of the relative abundance of collections in the “top ten” families in terms of the Kenya collection counts.

GRID VIEW LINKS TO EXTERNAL WEBSITES

When viewing results in a grid view that contains species names (Species, Seed, Seed Tests), you can search alternative websites for additional data, images and collections relevant to the active record using the shortcut links.

These weblink are accessed via a drop down menu:
The external websites currently available are:

The Plant List
http://theplantlist.org

Centro de Referência em Informação Ambiental (CRIA)
http://names.cria.org.br/index

The International Plant Names Index (IPNI)
http://www.ipni.org/ipni/advPlantNameSearch.do

The Royal Botanic Garden Edinburgh Herbarium Catalogue
http://elmer.rbge.org.uk/bgbase/vherb/bgbasevherb.php

Global Biodiversity Informatics Facility (GBIF)
http://data.gbif.org/

Harvard University Herbarium specimen search
http://kiki.huh.harvard.edu/databases/specimen_index.html

Harvard University Herbarium collector names search
http://kiki.huh.harvard.edu/databases/botanist_index.html

RBG Kew Electronic Plant Information Centre (ePIC)
http://epic.kew.org/searchepic/editsearch.do

RBG Kew Seed Information Database (SID)
http://data.kew.org/sid/SidServlet

Missouri Botanic Gardens, Tropicos Database
http://www.tropicos.org

The New York Botanical Garden C.V. Starr Virtual Herbarium
http://sciweb.nybg.org/science2/hcol/all/vasc/index.asp

Oxford University Department of Plant Sciences Virtual Field Herbarium
http://herbaria.plants.ox.ac.uk/vfh/image/

The African Plants Database
REPORTS VIEW

Whilst grid views offer enormous flexibility for online data analysis and export, viewing the results screens in a report view provides options for designing, saving and printing reports directly from an online query.

In this example, collections from the genus Galanthus from Georgia have been queried. In the Seed Tests Results Page you can select 1 or more standard report format through a drop down list.

In addition, some other Results Pages provide the added flexibility of user-defined report designs. In the example below, the seed accessions of Galanthus from Georgia are presented. There are a set of list options which allow users to include and remove particular fields from the report. Here we have removed the major area and the species authorities from each record.

Another way of looking at these same data would be to refer to the Species Results Page. If sufficient species related data are uploaded to the Data Warehouse, then some useful species profiles can be designed and printed/exported. Here is an example from the Conifers database compiled by Dr. Aljos Farjon at RBG Kew (http://herbaria.plants.ox.ac.uk/bol/conifers/).
In this example we have queried collections of the genus Afrocarpus and the Species Results Page is viewed in Reports view. On this page, options to include species level data such as TDWG global distribution, IUCN ratings, Description, can all be included in a species profile type report.

**PRINTING AND EXPORTING REPORTS**

Once designed, online reports can either be printed directly or downloaded to a view file. Currently, you cannot download to a MS Word document.

- Species Results Pages offer significant flexibility to design reports on uploaded species data.
- Species Profiles can be designed through the Species Results Page using the Reports View.
- Select printing options from the web browser menu.
- Reports can either be printed or, when relevant software is installed, saved to document writers such as Adobe (.pdf) or Windows (.xps).
RESULTS PAGES EXPLAINED

Summary Results Page

After undertaking any search, the summary results page will be displayed by default. On this Results page, the search parameters are displayed in the left hand table and the results that matched the search displayed in the right hand table.

The direct URL for this Summary Results Page is displayed in green as a hyperlink and also as text in the “copy & paste” box at the foot of the page. This is a useful tool when sharing information with colleagues.

A drop down list will generate the appropriate URL (in green) and corresponding text in the “cut & paste” text box depending on which Results Page is selected.

Family and Genus Results Pages

Lists of families and genera represented in the searched collections are presented by default in grid view. This facilitates name checking, distribution, seed behaviour etc. via the external website links. Family/Genus Authors and Citations will be presented if they have been added to the database.

An initial search carried out on all Chinese collections will enable the Family/Genus analysis of the collections using the Genus Results Page in grid view. In this example the resulting analysis has been exported to Excel and a graphic showing the families that hold 50% of the generic diversity in the Chinese collection is produced.
Display the data in a reports view to create complex checklist-style reports can be created depending on the amount of data that has been shared through the Data Warehouse. In the following example we are using the Conifers database courtesy of Dr. Aljos Farjon, RBG Kew (http://herbaria.plants.ox.ac.uk/bol/conifers).

Species Results Page

When in grid view, the Species Results Page facilitates name and distribution checking via the external website links. Results will open in a separate window. This example uses an initial search of all collections from Chile.
Some further taxonomic analyses can be undertaken. Here, using the same data as above, the grid view is used with the summary function for Family to generate a species count across the families collected from Chile.

- Filter the search results by the taxon you wish to carry out name checks against.
- The active record will be used to perform a name check search using e.g. plantlist.org.
- Create a summary analysis of the Family field.
- The results are displayed in a floating 2-column summary table which can be downloaded to Excel.
When viewed in the reports view, the Species Results Page offers a powerful function for designing checklists and species profiles depending on the quantity of additional species-linked data shared through the Data Warehouse.

Seed Accessions Results Page

This Results Page displays what is often referred to as the “Passport field” data; essentially the “who, what, when & where” of an individual collection. In addition, any unique accession numbers assigned by the seed banks which are associated with the collection are often available.

In this example we use the same query as before and view all the seed collections from Chile on the Seeds Results Page in grid view.

If provided to the database, the Original quantity collected and the Adjusted quantity based on estimated viability are presented. Also any corresponding X-ray data that have been provided are also available.
In reports view there are a number of items that can be edited into the report. In this example the original query was filtered in grid view (above) to show only collections in the family Alstroemeriaceae.

The banner has been hidden and the no. of records per page increased.

Explore the use of the summary functions. This example has analysed the field “Species”

The resulting analysis can be sorted &/or sent to MS excel for further processing.

Selecting extra, non-standard fields – in this instance the field “Loc Notes” has been included.

Memo fields can be viewed in full using “text-wrap”
This filter remained in place when viewed as a report.

In this example an initial search was performed on the Genus Rosa. The Seed Tests Results Page in grid view displays the 40 germination test results currently held in the database.

Now, the filter form has been enabled in the Results field and “>74” has been entered. This results in a grid containing all seed tests carried out on collections of the species of Rosa that have attained a result of 75% or more, collected by MSBP partners.
By enabling the text wrap function and sorting the tests in decreasing order, we can review the successful test conditions used.

In Reports View, the Seed Test Results Page has three default options for reports selected from a drop down menu.
On the Mapping Results Page, a Google map of all the returned results is presented in the left hand pane. The right hand pane contains the summary of each record represented on the map. If any filters have been applied in the Seed Results Page grid view, only those filtered records will be mapped.

Here is a map of all Compositae thus far collected and shared through the Data Warehouse.

Currently, three default report views exist for the Seed Test Results Page.

Here, the results Report has been saved to a MS XPS Document (.xps)
The symbols first appear as “Collection Heat Indices” (Red, Orange & Blue) representing groups of collections from a particular grid cell. The Indices also indicates the actual number of collections associated with that grid. Click on any of these indices or zoom in on the map using the magnifier slide bar to gradually uncover the individual mapped points.

Select individual records on the map to reveal the summary data for that record as a call out box on the map. Hovering over the point will temporarily highlight the record in the right hand pane and in the bottom pane.

Images Results Page

Currently only restricted image data has been uploaded to the Data Warehouse. The functionality of the Image Results Page is also being developed further and will provide versatile image management and viewer options.
SETTING YOUR SCREEN OPTIONS

“But I can’t see that on my screen....” is a common complaint! What might look good on my screen doesn’t necessarily mean it’ll look good on yours. The reality is that there are many ways to optimise screen resolutions, allowing us to view a page without the need for scrolling. This is often important for web pages that display data where we frequently want to compare rows and scan data without scrolling.

Use the Ctrl and “+” or “-” keys to increase and decrease the zoom function of the browser.

Here are some more useful tips (in this instance using FireFox web browser) to optimise your screen view when using the Data Warehouse.

1. Right click on browser header to minimise the visible options
2. Turn the website banner on or off using these toggle buttons.
3. Resize the fields by selecting the far right hand edge of the field and dragging to the right or left
4. Increase the number of records displayed per page (a scroll bar appears)
5. Move fields by clicking on the header and dragging to a new position

End.