Future-proofing MELISR

Toward a robust and standards-compliant herbarium management system for the National Herbarium of Victoria

Executive summary

MELISR is the collections database of the National Herbarium of Victoria. MELISR has not undergone any major development in the last 8+ years, in order not to interrupt data entry for Australia's Virtual Herbarium (AVH). Consequently, there are now many outstanding issues that need a major redevelopment effort to resolve. These include unreliability of the back-up system, issues with data capture and data integrity, and issues with data delivery to AVH.

In order for MELISR to optimally fulfill the business needs of the Plant Sciences and Biodiversity Division (PS&B), MELISR also will need to incorporate the loans and exchange database and the MEL herbarium census, as well as efficiently interface with nomenclatural or taxonomic databases such as VicList, APNI and the Australian Plant Census (APC), and be able to communicate with mapping software such as ArcGIS. As not all these objectives can be met under the current database management system, KE Texpress, other database options have been investigated.

It is concluded that the Royal Botanic Gardens (RBG) does not currently have the expertise to develop its own herbarium management system. In any case, this option would take the greatest development effort and lead to the largest loss of productivity during development, while there are pre-built collections management systems available, some of them at little or no cost.

KE EMu is the successor of Texpress and still uses Texpress as its back-end database, and therefore would represent a logical upgrade path. However, it is very expensive and has a history of implementation problems at other herbaria. As EMu is designed for all kinds of museum collections, customisation cost also would be maximal.

BRAHMS is especially developed for herbarium management, is free, and would probably require the least customisation to accommodate the MELISR data structure. However, BRAHMS uses a soon to be phased out database management system, and it is unclear how BRAHMS will develop after that. Because of the rather weak back-end database there are issues with robustness, scalability and extensibility.

Specify is developed for natural history collections, including herbaria and natural history musea. Specify has a highly structured, standards-compliant data model and was found to be usable and robust. It is also the only system that is completely open source and hence can be adapted to our future needs. Implementation of Specify requires only minimal changes to existing Information Services (IS) infrastructure. It is recommended that the RBG adopts Specify as its new herbarium management system.

Converting to a new collections management system will involve significant commitment of time and resources to retrain existing staff who use MELISR. In order to minimise loss of productivity during the implementation phase it is recommended the new herbarium management system be implemented prior to initiating any further large-scale data capture projects, such as databasing the foreign collection or undertaking a specimen imaging program. On a broader scale, it is a good time to upgrade our database system as TDWG standards are now at a mature stage and are being adopted globally.

Background

The MELISR database supports the primary business of PS&B. MELISR has been running on the Texpress DBMS since its inception in the early 1990s. The Texpress DBMS (produced by KE software) is simple and efficient and at the time was adopted by most Australian herbaria. However, with increasing requirements for data sharing between herbaria and the emergence of global biodiversity data standards, the limitations of the Texpress DBMS are becoming apparent.

The MELISR database design was largely borrowed from other database systems in existence (namely CANB and NSW) leading to some parts being less than optimal for MEL's requirements. Despite these shortcomings, MELISR has not had any major development work done for the last 8+ years in order to minimise disruption to the Australia's Virtual Herbarium (AVH) databasing project. There are many outstanding issues with the database that will need a considerable development investment to resolve. It is therefore prudent to consider whether this effort would not be better directed toward upgrading MELISR to another database platform.

The MELISR database is a specimen database only; its simplistic structure cannot incorporate all the specimen-related or name-related information and curatorial tools used in the herbarium. In addition to MELISR, the herbarium maintains the Loans and Exchange database (an MS Access database), the Census database (MS Access), the VicList database (MS Access), a table of scheduled taxa (Texpress and MySQL) and an Authors of taxonomic names database (MS Access). The limited querying capability of Texpress, and the inability of Texpress to directly interface with the software required to present MEL's specimen data over the internet, have made it necessary to establish a duplicate of MELISR using the open source MySQL database. This is an inefficient solution that requires extra room on the server, resources to keep the two databases in synch and duplication of effort when implementing changes to MELISR.

An important application of MELISR data is the production of distribution maps. The inability of Texpress to interoperate with GIS software makes mapping MELISR data very laborious.

The label printing program within Texpress is primitive, and cannot be tailored to meet our needs without engaging programming assistance from KE Software. As there is often a conflict between the requirement to capture data and the requirement to print data, the inability to customise label printing in MELISR results in certain data (e.g. quarantine messages or acknowledgements of funding support) needing to be edited in to or out of records either before or after printing. A more sophisticated, and easily-customisable, printing system would allow the requirements for curatorial and specimen-related label data to be managed more consistently and more efficiently.

The reporting program in Texpress also leaves much to be desired. New reports within MELISR need to be individually programmed, and several curatorial reporting requirements are handled by other programs (such as the MySQL copy of MELISR, the Census database and the Loans and Exchange database) due to the deficiencies of the Texpress system. Loan information is only recorded in MELISR records for the duration of the loan, making it impossible to track the loan history of a specimen. The absence of loan history data for individual specimens means that the MEL collections cannot be accurately audited to meet reporting requirements (such as those of the bi-annual AVH Board report).

Our past experience with the Texpress system has demonstrated that it lacks robustness, making it susceptible to data loss. In March 2006, a low-level hardware error resulted in the loss of several records. The cause of the problem was difficult to pinpoint and remedy, and resulted in a three week disruption to data entry and retrieval at a time when ten staff were employed specifically to undertake data entry. As well as resulting in significant loss of productivity, this problem highlighted the inadequacy of the Texpress backup system; not all the lost records could be recovered and restored.

In addition to the technical deficiencies of the system, the current data structure is too simplistic to allow for accurate capture of taxonomic information, determination history and geographical information, which reduces the utility of our specimen database and compromises the integrity of our data. Texpress imposes constraints on the way that data can be structured, with all information for each object (or specimen) recorded in the one database record. This creates single records containing large numbers of fields, with the same information entered numerous times across the different records, resulting in inefficient data entry. As well as improving the way that specimen data is recorded, the database needs to be able to track data validation efforts and allow for better documentation of changes to records. Appendix A lists some of the improvements that need to be made to existing fields in MELISR, as well as suggested new fields that would improve the searchability and standards-compliance of the database. Many of these suggestions have been requested from staff and external clients. The key areas requiring improvement are expanded upon below.

One critical area that requires improvement is the way that taxonomic information is entered and stored in MELISR. The flat data structure of the Texpress system means that taxonomic names – which should ideally stand alone – cannot be separated from the specimen-related data (such as determination annotations and hybrid information) associated with individual records. Individual components of names are currently entered by the database operator from look-up tables. While this partially restricts the content of the taxonomic name fields, the tables are easily edited by any user with data entry privileges, so there is much scope for errors in data entry and inconsistencies in the application of names. The inadequacy of MELISR in dealing with uncertain determinations and cultivar, hybrid and informal names, combined with the inability to adequately restrict the content of the taxonomy fields, reduces the quality and reliability of the name data associated with our database records.

The current approach to recording taxonomic names makes it necessary to maintain a separate list of names that reflects the content of MEL's collection (the Census database). This represents a considerable duplication of effort, which could be eliminated by using a comprehensive herbarium management system based on an authoritative list of names, rather than a collection of separate databases each with their own name lists. A major benefit of having MELISR linked to such lists is that it would allow searching by synonyms, which is a powerful tool both for specimen curation and for data interrogation.

As well as improving the way that current names are recorded in MELISR, it would be valuable from both a taxonomic and curatorial point of view to record the determination history of a specimen. Currently, there is no facility in MELISR for recording original determinations and subsequent re-determinations. Although additional fields could be added to the new system, they would suffer the same shortcomings as the existing taxonomic name fields, thus determination histories would be better handled by a relational database with an underlying taxonomy table.

Another major weakness of the current MELISR database is that specimen records from the core collection (the specimen component of the State Botanical Collection) cannot be distinguished from records from non-core collections such as the Victorian Reference Set, the Horticultural Reference Set and the Victorian Conservation Seedbank. It is important that these collections are kept separate on the database so that data-retrieval for loan enquiries, electronic data requests and specimen retrieval reflects the location and accessibility of the specimens. Database records from the Victorian Conservation Seedbank may not be associated with a vouchered specimen, which undermines the integrity of MELISR as a collections database based on verifiable specimens. While it makes sense to use the same database to record data for these distinct collections, the structure of the Texpress system is too simple to reflect the different purpose, location and accessibility issues associated with these collections.

The shortcomings with the current version of MELISR outlined above could be overcome by employing a comprehensive, relational herbarium management system that encompasses a specimen database, loans and exchange database, scheduled taxa database and herbarium census. The taxonomic tables that would be the basis of the herbarium management system could also encompass the VicList database, thus reducing duplication of effort in taxonomic name management at the RBG and streamlining curatorial practices associated with keeping VicList data up to date.

While many of the desired improvements to MELISR require a more sophisticated database management system, some improvements to the handling of primary collecting data could be implemented in the existing Texpress system. However, given the considerable shortcomings of the existing system and the effort required to replicate any changes in the duplicate version of MELISR, this development effort would be more judiciously applied to incorporating MELISR into a new herbarium management system, rather than adding further workarounds to the existing database.

Options

1. Keep existing Texpress system

The Texpress application currently used for MELISR is no longer widely used and is likely to become obsolete. The only technical support available is from KE software and this is unlikely to be available in the future.

The standard data exchange protocols used to link in to global biodiversity initiatives (such as BioCASE and TAPIR) cannot interface directly with Texpress. Keeping the existing Texpress system for our Collections database puts the RBG at risk of not being able to participate in emerging biodiversity information initiatives such as GBIF, ALA and EOL and poses a risk to the organisation's reputation as a quality data custodian.

The loss of corporate knowledge associated with using a near-obsolete system is a great risk to the organisation. The administration and maintenance of the Texpress system requires a large amount of specialised skills and experience that cannot be easily replaced. The archaic nature of Texpress means that skills developed in any other modern database system will not be transferable.

There may also be financial ramifications for the RBG if we persist with Texpress, given the potential for licensing and support costs to increase due to the small number of users. Texpress development costs are also very high, as has been experienced in the past, and these costs are likely to rise in the future. The RBG currently pays $6300 per annum for 25 user licenses. Although this cost could be reduced to $4800 for 15 licenses, the cost of purchasing new licenses when needed is much higher than the amount saved this way. Any redevelopment in Texpress will require purchasing the TexAPI package at a cost of $18,000 plus $2315 per annum for licenses.

2. Change to a custom-built collections management system

One option is to develop a custom collections management system. The greatest benefit of this option is that it would be specifically tailored to our needs. Along with this benefit come risks associated with using a system that is not widely used, and thus doesn't have a global community of users and developers. The development of such a system would be expensive in terms of staff time, and would require that staff time is taken away from other duties.

At the implementation stage, there is the risk of exceeding the anticipated development cost and the cost of data migration from Texpress to the new system. Also, the temporary loss of productivity during the changeover period would be greatest with this option. All future support and development would most likely have to come from within the organisation. The RBG do not have the expertise to develop a custom-built front end in-house.

3. Change to existing collections management system

The third option is to transfer MELISR to an existing collections management system. This is potentially an expensive option, but the cost of different collections management systems varies enormously. The major benefit of this option is that it will allow us to select a proven system that has a global community of users and developers. It will also be less expensive in terms of staff time than developing a custom-built system.

Risks associated with this option include high development/customisation, migration, and training costs. These costs can be minimised by choosing a system that most closely meets our needs, choosing a database management system we are already familiar with, and choosing a system with a user-friendly interface. There is also the risk of data loss during migration, which may be reduced by keeping good back-ups and by rigorous testing of the data model of the new application. We also need to minimise the risk that the new software will not be supported in the future and make sure the database can be modified and extended to meet future business needs of the RBG.

3.1. KE EMu

KE EMu is a Windows-based collections management application, designed for all kinds of musea, that uses Texpress as the back-end database. While this is a powerful system and represents a logical upgrade path, it has serious limitations. It is very expensive and has a history of implementation problems with regard to herbaria and botanic gardens (e.g. NSW and BM). It is also closed source, meaning that any customisations will have to be performed by KE Software and will be costly.

Table 1. Features of KE EMu

Operating system	
Front-end	MS Windows
Back-end	Linux
Open source	KE EMu is completely closed source; any customisation will have to be performed by KE Software.
Back-end database	KE EMu uses KE Texpress as its back-end database. This is the same database MELISR currently uses, but with some extensions.
User-friendly interface	Yes
Customisability	KE EMu is fully customisable, but this customisability comes at a cost as it has to be carried out by KE Software. As EMu is designed for use by all kinds of musea customisation needed will be more than in the other applications considered.
Scalability	EMu scales well.
Extensibility	EMu is fairly self-contained; extension is possible, but will have to be custom-designed by KE.
Interoperability	Native interoperability is poor
Support	Support is probably good, but expensive. One would expect some support will come with the licenses. However, a large part of the problems that other herbaria have had with EMu is likely to have been caused by communication problems between herbarium people and application developers.
Startup costs	$130,000 (25 licenses plus initial customisation)

Ongoing costs	$120,000 pa (25 licenses)
url	http://www.kesoftware.com/content/view/512/356/lang,en/

3.2. BRAHMS

BRAHMS is a specialised herbarium collections management system developed by the Oxford University Herbaria and used by many herbaria all over the world, for instance the National Herbarium of The Netherlands (L, WAG). BRAHMS was found to be rich in features, but lacking in usability, robustness and extensibility. BRAHMS has just undergone a major redevelopment, but is still built around the same DBMS, Microsoft Visual FoxPro, which is not an enterprise database. BRAHMS will only run on Microsoft Windows platforms.

Table 2. Features of BRAHMS

Operating system	
Front-end	MS Windows
Back-end	MS Windows
Open source	BRAHMS is closed source
Back-end database	BRAHMS uses Microsoft Visual Foxpro. FoxPro is a legacy DBMS no longer actively developed by Microsoft and will not be supported after 2014.
User-friendly interface	No
Customisability	BRAHMS has limited customisability. However, the application is specifically designed for herbaria, so only little customisation will be necessary to accommodate the MELISR data model.
Scalability	BRAHMS scales poorly, mostly due to its rather weak back-end database.
Extensibility	BRAHMS is fairly self-contained. However, it is very feature-rich and is specifically designed for herbarium management, so the data model should be sufficient to accommodate all necessary fields at least for the near future.
Interoperability	BRAHMS has built-in operability with some other applications, such as ArcView and DIVA-GIS. The file format in which it saves its data (.DBF) can be read by some Windows applications. An extension for online publishing is available. The National Herbarium of the Netherlands is a member of EDIT and BioCASE and delivers data to GBIF, so dynamic data delivery through a BioCASE provider must be possible.
Support	Support for BRAHMS is provided by the BRAHMS Project at the Oxford University Herbaria. As herbarium taxonomists are involved in the project, there should be no communication problems. A support contract costs $US600 pa.
Startup costs	–
Ongoing costs	$US600 pa for support (optional)
url	http://dps.plants.ox.ac.uk/bol/home/default.aspx

3.3. Specify

Specify (University of Kansas) is designed for herbaria and zoological musea and was found to be usable and robust. The Specify software is currently being completely rewritten and will be released as open source. The new version, Specify6, is built in Java using Hibernate to abstract the database layer and will therefore run on many database management systems, including MySQL, PostgreSQL, and Microsoft SQL Server, which all already run on RBG servers (MySQL is used for the web interface of all botanical databases and for the AVH interface, as well as advanced querying, of MELISR; MySource Matrix, the Content Management System for the RBG website uses PostgreSQL; Hummingbird, the records keeping software uses MS SQL Server). Specify6 will be released 27 February 2009. Specify has a world-wide user community and is currently used by 112 institutes, including 34 herbaria. Development of Specify has been supported by the US National Science Foundation for the last twenty years. Judging from the proceedings of the 2008 TDWG annual conference, Specify is very much at the forefront of collections management systems.

Specify has a strongly structured data model that is DarwinCore compliant (GBIF uses DarwinCore) and therefore most likely also ABCD compliant. Specify6 contains 138 tables and 1658 fields. While some fields in MELISR that are specific to MEL will not be already in the Specify data model, there are several blank fields which can be used for these. Given that the database layer is abstracted from the front end and the database management systems that can serve as back end are very powerful, we expect excellent scalability and extensibility, as well as interoperability with other applications (e.g. GIS, electronic flora, image storage).

Specify optionally comes with a fully customisable (using CSS only) web interface. The Specify front end, which includes all the forms and reports (including labels), is fully customisable, without requiring programming. If in future we want to make changes to the data model, the associated changes in the front end would require programming in Java. However, given the growing user community we expect that changes in the data model necessitated by outside factors would be taken care of in new minor versions of Specify.

Table 3. Features of Specify

Operating system	
Front-end	MS Windows, Mac OS 10 or Linux
Back-end	MS Windows, Mac OS 10 or Linux
Open source	Specify is open source. Source code for everything except the label printing application can be obtained from the developers. Specify6 is entirely written in Java, which means that in future we will actually be able to make changes to the source code if necessary.
Back-end database	Because the front-end user interface is separate from the back-end database, Specify6 can use most of the major database management systems, such as MySQL, PostgreSQL, MS SQL Server or Oracle. Specify6 by default uses MySQL.
User-friendly interface	Yes
Customisability	Specify 6's Graphic User Interface is entirely customisable, with the possibility to choose fields, change the format or type of fields and even change field names (similar to forms in MS Access).
Scalability	Because of the very powerful back-end database systems Specify6 will scale very well.
Extensibility	Specify has good extensibility. While extension of the data model at the back end is easy and only requires knowledge of SQL, the associated changes in the front end require more knowledge of Java than is currently available at the RBG. However, the Specify data model is very rich, with many customisable fields, so should easily be able to accommodate all MELISR fields at least in the near future.
Interoperability	Specify comes with a web-interface (which we may not use) and a DiGIR provider (which we definitely will not use). MySQL interoperates very well with PHP for dynamic web applications and, through the MySQL ODBC, with MS Access and ArcGIS.
Support	Specify is free and offers free support to registered users. While priority support is given to US institutes, Specify is happy to provide support to non-US institutes as resources allow. Part of the support is migration of data into Specify. With Specify5.2 there was a waiting time of 2–3 months between registration and migration. We expect waiting times to be longer once Specify6 is released, as existing Specify users will need to have their data migrated as well.
Startup costs	–
Ongoing costs	–
url	http://www.specifysoftware.org/Specify; http://specify6.specifysoftware.org/ (temporary Specify6 website)

Recommendation

Given the Texpress system is nearing obsolescence and changing to a custom-built system is an inefficient and expensive option, we recommend changing to an existing collections management system. Based on our comparison of existing collections management systems we recommend upgrading to Specify. We finally recommend a new collections database management system be implemented prior to initiating any further large-scale data capture projects, such as databasing the foreign collection or imaging of herbarium specimens.

Of the three collections management systems considered KE EMu was discarded as an option early, because of its high implementation, customisation and licensing cost and because of the problems other herbaria have experienced with it. BRAHMS was found to be feature-rich and, because it was designed especially for herbaria, BRAHMS' data model is currently probably the most compatible with the structure of MELISR. However, we are concerned about the back-end database management system BRAHMS uses, and that because of its limited scalability and extensibility, we will not be able to adapt BRAHMS to meet the RBG's future needs.

Specify has the ability to employ a very powerful database management system and can therefore make use of the DBMS' back-up and security facilities. It has a highly structured data model that can include most MELISR fields as is, and all fields after some modification. Specify comes with a very user-friendly, fully customisable front end and with extensions such as a web interface and DiGIR provider. The worldwide diverse user and development community guarantees that Specify will adapt to future needs better than the other systems. Also from an infrastructure perspective Specify fits best, as all required infrastructure is already in place at the RBG (nevertheless a more detailed analysis of infrastructure requirements will be part of the project planning).

We would like to emphasise that while upgrading to a new collection management system is urgent in order to safeguard the quality and integrity of our collections data and the RBG's reputation as a quality data custodian, the process is not going to be painless. The implementation of a new collection management system will require a large time commitment, especially from the Programmer, Information Services and Collections Information Officer, and will affect all MELISR users. In order to ensure data integrity it will not be possible to run the old and new systems concurrently and therefore MELISR will not be accessible for a period of time during the implementation phase. Also the loans and exchange administration system will not be accessible during this period as it will be included in the new collections management system.

The temporary loss of productivity during the implementation period will be more than made up for by improved efficiency and increased productivity once the new herbarium management system has been implemented. On a broader scale, this is a good time to upgrade our collections database system, as international data standards have come of age and only minor changes are expected in the near future.

Appendix B describes an implementation roadmap that aims to ensure the implementation period is as short as possible and to minimise loss of productivity during this period.

Table 5. Summary of features of the different herbarium management systems considered. The current KE Texpress collections management system is included for comparison.

	KE Texpress	KE EMu	BRAHMS	Specify
Operating system				
Front-end	Linux shell emulator	Windows	Windows	Windows, Linux or Mac OS X
Back-end	Linux	Linux	Windows	Windows, Linux or Unix
Open source	No	No	No	Yes
Back-end database	Texpress	Texpress	Microsoft Visual Foxpro	MySQL, PostgreSQL, MS SQL Server, Oracle or any other server-side DBMS
User-friendly interface	No	Yes	No	Yes
Customisability	Poor	Good, but very expensive	Good	Good
Scalability	Poor	Good	Poor	Excellent
Extensibility	Poor	Good, but very expensive	Poor	Good
Interoperability	Poor	Poor, but with ample inbuilt functionality	Good	Good
Support	Very limited[1]	Good	Good	Good
Startup costs	N/A + $18,000 (TexAPI)	$130,000	–	–
Ongoing costs	$6,300 pa + $2,315 (TexAPI)	$102,000 pa	$US600 pa[2]	

[1] Support for Texpress is very limited as Texpress as a stand-alone application is being phased out and replaced by EMu.

[2] $US600 pa is for support, there are no licensing costs.

Appendix A — Summary of suggested improvements to MELISR

Field/s	Suggested improvements
Taxonomy fields	• Link specimen names to an authoritative table of taxonomic names. Information related to individual specimens (determination annotations, qualifiers and hybrid information) to be stored with individual records, rather than with names. • Improve handling of higher level taxonomies, particularly for fungi and algae
Determinations	• Include original determination and subsequent determination history.
Collector/Add. coll.	• Create separate fields for recording verbatim label data and standardised data, which would link to an authoritative list of collectors.
Collecting date	• Add a memo date field to record non-standard collecting dates, e.g. 13-17 June; late March; Spring; Christmas etc.
Additional collector	• Add new fields to record collecting numbers of additional collectors.
Geocode	• Allow for recording of geocode as originally provided (DMS or decimal). • Add new fields for recording AMG references, and enable autoconversion of AMG to geocode. • Add a new field for recording error measure when provided by collector. • Improve handling of geocode source data.
Cultivated data	• Improve handling of locality data for cultivated records. Currently, provenance data is entered in the Notes field, and cultivating locality details are entered in the locality fields (minus geocode). Need to record both cultivated and provenance locality data in a way that allows them to be queried and mapped (or excluded from queries or mapping) on request. • Add new fields to cater for Plant Occurence and Status Scheme values.
Unit relationship field	• Add new fields to record the range of relationships between herbarium sheets. Currently, the only way of recording a relationship between one or more herbarium sheets is to multisheet them. Need to convert the multisheet field to a unit relationship field that allows for other types of relationships to be recorded (e.g. cultivated seedling and wild-collected parent plant; uncertain links between foreign specimens).
Duplicates/Specimen Received from/Original herbarium (for images)	• Apply a restricted vocabulary to these fields to prevent the inclusion of non-standard entries.
Protologue	• Separate the publication title and the page and date citation into two distinct fields so that publication title can be linked to an authoritative list of names (e.g. BPH/TL-2) to avoid incorrect and inconsistent entries.
Precision	• Check that our precision code values are sensible and add a built-in guide to help data entry personnel to use them correctly.
Depth	• Stop decimal places from being automatically appended to values in this

	field as it gives an inaccurate impression of the precision of the measurement.
Notes	• Divide this field into two (or more) categories to allow the collector's notes to be recorded separately from annotations on the specimen and curatorial or explanatory notes added by the database operator.
Managed habitats	• Improve plant occurrence fields to better document specimens collected from managed habitats, e.g. those that have been self-established in a Botanic gardens context, and should not be treated either as wild-collected or as cultivated.
Validation level	• Add new fields to represent the level of validation of a database record (includes identification, geocode, distribution, predictive distribution).
Images	• Differentiate between images of the sheet and images of plant in its habitat etc. Also need to be able to record when we have produced a digital image of a specimen to send to another institution. • Add a field to record file paths or URLs for digital images.
Vic. Ref. Set	• Improve flagging of Vic. Ref. Set specimens. Vic. Ref. Set specimens are currently listed as duplicates, despite the Vic. Ref. Set not being an official, accessible collection. It would be better to flag these records in a different way than duplicate specimens are flagged.
Type status determination	• Move type status determination data. This information is currently stored in Extra Info., but would be better stored as part of the determination history, with type status of the determination recorded.
Verbatim label field	• Add a new field to record verbatim label data for foreign-language labels. • Allow for unicode characters to be captured so that foreign-language data can be recorded more accurately (not possible in Texpress).
Original language field	• Add a new field to record the original language that a label is written in (if non-English). This will allow ease of searching in the event that we want to query for language (e.g. for batch translation of labels).
Global gazetteer	• Link to a global gazetteer for ease of geocoding foreign collections. e.g. GEONet Names Server files
Library catalogue no.	• Add fields to enter call numbers for additional information stored in the library, e.g. letters, photos, colour transparencies etc.
Quarantine notes	• Add a new field to enter quarantine notes that are not printed on any labels or exported, and are only used by curation staff. Currently, these messages must be deleted prior to labels being printed, then re-entered into the record.
Destructive sampling	• Add a new field to record when material has been removed for destructive sampling.
Ethnobotanical information	• Add a new field to flag the presence of ethnobotanical data associated with a record.
Indigenous name	• Add a new field to record indigenous plant names when provided by the collector.

Appendix B — Implementation roadmap

The major stages required to implement this project are outlined below:

Project planning

- Registration with Specify;

- Development of a detailed implementation plan, by September 2009.

Comprehensive needs analysis

- Consultation with MELISR users regarding the current MELISR database structure to determine what fields need changing, and what new elements are required;

- Mapping of fields in MELISR against the HISPID5 (ABCD) standard to ensure compliance;

- Preparation of draft MELISR data entry manual.

Data preparation

- Mapping of MELISR fields against Specify data model;

- Performance of major quality assurance work on non-compliant fields in MELISR to make data migration to Specify as smooth as possible.

Implementation and testing

- Installation of Specify on MEL server and work stations;

- Data migration;

- Comprehensive testing of the new system and revision of the MELISR data entry manual.

Training

- Training of MELISR users in the use of the new system. This will need to be undertaken incrementally, starting with those staff whose work is most reliant on the database.

Configuration of provider software

- Configuration of TAPIR and BioCASE providers for data delivery to AVH, the ALA and the GBIF.

Appendix C — Glossary

ABCD	**Access to Biological Collections Data** – a comprehensive TDWG standard for access to and exchange of primary biodiversity data
ALA	**Atlas of Living Australia** – a project funded under the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) to develop a biodiversity data management system for Australia's biological knowledge
ArcGIS	a group of geographic information system (GIS) software product lines produced by ESRI
AVH	**Australia's Virtual Herbarium** – an online botanical information resource that provides access to data associated with scientific plant specimens in Australia's major herbaria
BioCASE	**Biological Collections Access Service for Europe** – a transnational network of European biological data providers
BioCASE provider	A data exchange protocol developed by BioCASE. The BioCASE provider abstracts data from a database and turns it into standard format, such as ABCD or DarwinCore. MEL and AD (Adelaide) use the BioCASE provider to deliver data to AVH. Other similar protocols are TAPIR and DiGIR.
BM	Acronym of **The Natural History Museum, London** (British Museum)
BRAHMS	**Botanical Research And Herbarium Management System**
CSS	**Cascading Style Sheets** – a stylesheet language used to apply formatting to web pages
customisable	able to be customised to our particular needs, without the need for extensive programming
DarwinCore	a standard designed to facilitate the exchange of information about the geographic occurrence of species and the existence of specimens in collections
DBMS	**database management system** – examples of database management systems are: MS Access, MySQL, MS SQL Server
DiGIR	**Distributed Generic Information Retrieval** – a client/server protocol for retrieving information from distributed resources
EMu	**Electronic Museum** – collections management software developed by KE Software
EoL	**Encylopedia of Life** – a project to create an online reference source and database for the 1.8 million named and known species on earth
extensible	able to be extended and adapted: an extensible database allows for expansion of the data structure, i.e. additional tables and fields
future proofing	the selection of physical media and data formats that best ensure the continued accessibility of data into the future. This process involves anticipating future developments and ensuring that only well-documented formats, standards and specifications are used to store and describe data.
GBIF	**Global Biodiversity Information Facility** – an international organisation that focuses on making scientific data on biodiversity available via the Internet using web services. The data are provided by many institutions from around the world; GBIF's information architecture makes these data accessible and

	searchable through a single portal. Data available through the GBIF portal are primarily distribution data on plants, animals, fungi, and microbes for the world, and scientific names data.
GIS	**geographic information system** – captures, stores, analyses, manages, and presents data that refers to or is linked to location. In a more generic sense, GIS applications are tools that allow users to create interactive queries (user created searches), analyse spatial information, edit data and maps, and present the results of all these operations.
HISPID	**Herbarium Information Standards and Protocols for the Interchange of Data** – a standard format for the interchange of electronic herbarium specimen information, initially developed by the Australian herbaria and later adopted as a TDWG standard. The current version – HISPID5 – is ABCD compliant.
Java	an object-oriented programming language
MEL	acronym of **The National Herbarium of Victoria**
MELISR	**MEL Information System Register** – the National Herbarium of Victoria's specimen database
Microsoft SQL Server	a relational database management system, used as the back-end database of Specify5
MySource Matrix	an open source content management system (CMS) written in PHP, used for the new RBG website
MySQL	a powerful, open source, relational database management system
NSW	**The National Herbarium of New South Wales**
PostgreSQL	on object-relational database management system
PS&B	**Plant Sciences and Biodiversity Division**, Royal Botanic Gardens (Melbourne)
RBG	**Royal Botanic Gardens**
roadmap	a plan for change execution
robust	able to withstand pressures or changes in procedure or circumstance
scalable	able to handle growth without having to replace the existing platform or architecture
Specify	research software application, database and network interface for biological collections information
TAPIR	**TDWG Access Protocol for Information Retrieval** – a computer protocol designed for the discovery, search and retrieval of distributed data over the internet
TDWG	**Taxonomic Database Working Group** (also referred to as Biodiversity Information Standards)
Texpress	an object-oriented multi-user database management system developed by KE Software
usability	the efficiency with which a user can perform tasks in a given application
VicList	*Census of the Vascular Plants of Victoria*, an up-to-date list of the species and infraspecific taxa of vascular plants occurring in Victoria