HISCOM

HERBARIUM INFORMATION SYSTEMS COMMITTEE

Summary of the

Herbarium Database/Information Technology Meeting

2 - 4 November 1995, National Herbarium of New South Wales

by

W.R. Barker & B.J. Conn

INTRODUCTION

In 1992, a workshop involving Information Technology staff from herbaria and museums discussed changes and future developments in database-design concepts (refer Conn, B.J. *et al.*, (1992). 'Selected papers from the Database Developments and Data Sharing Workshop', *Austral. Syst. Bot. Soc. Newsletter* **72:** 13-18). The current meeting was organised to enable Information Technology staff of Australasian herbaria to discuss similar important issues, but only as they relate to herbaria. Each session was introduced by a short talk, followed by discussion. Initially, the meeting was organised to try and develop:

- a framework for effective data interchange for exchange and loan specimens
- a format for a database of plant names.

However, several other important issues were discussed (e.g. HISPID, validation procedures, geocode calculation and precision, electronic loans management, Internet and email connection for all Australian herbaria), as outlined below. It was agreed that there was a need for an Information Technology Committee to co-ordinate and monitor IT issues as they relate to Australian herbaria. Such a committe could have a valuable role in co-ordinating the continued development of HISPID, with this workshop being the first meeting of the *Herbarium Information Systems Committee* (HISCOM). The workshop

proposed thirteen recommendations to CHAH, for their consideration, and five actions for HISCOM were agreed to (refer to Appendix for details).

The workshop provided an opportunity not only to share ideas, but to share solutions to technical difficulties.

 **PROGRAM**

Session 1 Welcome and Introduction Gwen Harden

Information Systems in Botany Jim Croft

Session 2 Specimen Databases Ken Hill

- structure

- special specimens eg. types

- supplementary collections eg. wet, photographic

- handling name changes

- problems

Session 3 Interchange of data Alan Brooks

- standards (HISPID)

- procedures / protocols

Session 4 Validation standards & procedures Alex Chapman

Session 5 Database of Plant Names Greg Whitbread

IOPI Karen Wilson

Session 6 Mapping Bill Barker

Discussion on ownership of data

Session 7 Electronic Loans Management Alan Brooks

Client server applications Peter Bostock

- using MS ACCESS <-> Oracle

Session 8 Discussion of Recommendations and Actions.

Session 9 Preparation of documents arising from the Meeting.

Demonstrations and further discussions.

List of Participants

AD: Bill Barker (billb@adam.com.au)

BRI: Peter Bostock (pbostock@ozemail.com.au)

CANB: Jim Croft (jrc@anbg.gov.au); Pennie Honen (pennie@pican.pi.csiro); Greg Whitbread (ghw@anbg.gov.au)

MEL: Geraldine Jones (entwisle@botany.unimelb.edu.au); Joan Thomas (entwisle@botany.unimelb.edu.au)

Museum of New Zealand: Joanna Newman (joannan@aotahi.govt.nz)

NSW: Alan Brooks (**Convenor**)(alan@rbg.syd.gov.au); Elizabeth Brown (elizabet@rbgsyd.gov.au); Barry Conn (barry@rbgsyd.gov.au); Dianne Godden (dianne@rbgsyd.gov.au); Ken Hill (ken@rbgsyd.gov.au); Karen Wilson (karen@rbgsyd.gov.au)

PERTH: Alex Chapman (alexc@calm.wa.gov.au)

UNE: Jeremy Bruhl (jbruhl@metz.une.edu.au); Gordon White

Summary of Presentations

Information Systems in Botany

Jim Croft

When this workshop was initially conceived, we had in mind a meeting of those people from botanical collections involved with the technical aspects of managing botanical databases and botanical information systems. We deliberately tried to exclude topics of a political and management flavour and, to some extent, even of a botanical flavour, to enable technologists to concentrate on achieving at the system, software, application, network and hardware level what we understand is required to make our task more efficient. That is, more efficient from the point of view of individuals, institutions and the botanical community as a whole. With this in mind, I will present a summary of the issues (negotiable, contentious, and intractable) which relate to botanical information systems.

What are we trying to achieve in the long term?

From the point of view of one who has to deal with the complexity of botanical information, I see that as botanical information managers, we have a huge task ahead of us, a task that is not made any the easier through a shortage of resources for information infrastructure and for botanical information systems development. There is a general feeling, with few dissenters, that botanists and institutions need to exchange specimens and data/information as part of their science and that databases of electronic records are the most effective way of achieving this. Furthermore, the general community wants to know basically six things:

- What are the names of plants?
- How can plants be identified?
- What impact or interest does this plant(s) have?
- Where do plants occur?

- What plants occur in this or that area?

and possibly

- How can I get or get rid of this plant?

At the very least, our information systems should cater for these half dozen questions. In particular, we need to take account of the following:

Accessibility

The data is not of much value to the botanical community and its clients if it is not accessible and not uniformly available from all institutions. Therefore, we are seeking wide and general agreement to making as much scientific data from as many institutions available as widely and as transparently as possible.

Reliability

The data is not of much value if it is incomplete or unreliable. It needs to be active, current, complete, accurate, precise and, above all, useful.

Usefulness

It not only needs to be useful from the point of view of herbarium botanists, it needs to be useful, and even essential, to a wider range of clients. The people and projects who are attracting money for biological information these days are not herbaria and museums, and if we can not deliver what they need they will go elsewhere for it.

Compatibility

We need to achieve compatibility between our databases and the data held in them so they can be used intelligently and efficiently by other data systems.

Transferability

We need an agreed set of protocols, packaging, standards, structure, format for the export of botanical data from databases so that it can accurately be imported by other users and databases.

Moveability

It is all very well to make data available, but it is not of much value if it can not be efficiently moved from one place to another. Therefore, we are looking for efficient mechanisms to package data elements and transfer them between remote sites, as quickly and cheaply as possible.

Efficiency

Because there are very limited resources available to the botanical community we need to be efficient in our capture and storage of data. Ideally, a piece of information should be inserted once and then be instantly available to others. Common authority files are a classic example, herbarium specimen duplicate data is another.

Ownership and Accountability

As far as possible, data and data sets should have a custodian whose responsibilities include many of the

previous points about currency, compatibility, standards, availability and so on. There needs to be a point of authority for the official version of a data set or a data element.

Distributed effort

The data needs to be widely available. There is a need for a distributed database where a wide range on institutions and individuals are responsible for various agreed and defined elements of the data. We want to do many things and we want to do them within our own organisations, within our own states, nationally and internationally. This would be easy if we all thought as one, but we do not. Furthermore, at each level there are different forces, different political factors, different legacies of existing systems, different ways of doing business and so on. All of these constraints hinder the establishment of a unifying botanical information system.

The creation of a sense of ownership by individuals and institutions of a larger, external botanical information system is a real challenge to achieving the above aims. There are some technological complexities, but in my opinion, they are minor compared to the socio-political ones.

What we want out of this workshop?

There are many important issues to be dealt with at this Workshop, namely:

Specimen databases

Interchange of data

Validation

Plant name databases

Mapping and GIS

Electronic management of loans

Exchanges

International efforts

Each one of these topics is a workshop in themselves. However, in spite of the limited time available, we should identify areas of conflict or impediments to efficient data handling and establish mechanisms for dealing with these. Even if we could come up with a solution to a single identified problem we will have made great progress. We do not want to reinvent anything that already exists or to start anything that is already being developed elsewhere. It follows that we should let each other know what we have, what we are developing, and agree to share this accumulated knowledge and expertise between institutions. I am pleased to note that the agenda offers the opportunity for this. Offering the tools is one thing, but we must acknowledge that some herbaria will have difficulty in installing or adapting them to their needs. It would be nice if we could establish a cooperative framework, where herbaria can provide reasonable assistance to other sites in installing and using products they have developed. At this point I would to acknowledge the assistance proved by Alan Brooks (NSW) to the Australian National Herbarium (CANB) in the installation of the NSW loans management system for evaluation.

If we could come up with an agreement to **always** supply specimen data with loans and exchanges, **and always actually do it**, we would have made a monumental advance. We have agreed to do this in the past, but it does not actually happen, other than sporadically.

This means we have to actually agree on acceptable media for data exchange, mutually acceptable data file formats, agreed data content standards and so on. Not an easy task, but one only this group, or one like it, can do

What is available now?

For a given task, there are a confusing number of tools and applications to choose from. This seems to be the by product of the market economy and the socio-political factors mentioned above. At the level of hardware and software, it is not a major issue and there are industry-standard means that can be used. At the level of the application, we have attempted to work around this by developing mutually acceptable data standards to facilitate exchange of information. The recent compilation of a single **Eucalypt dataset** for the continent demonstrated that this approach is possible, albeit with difficulty.

For a variety of reasons, each institution has an idiosyncratic database design and idiosyncratic data conventions, reflecting different ways of doing business, differing preferences of database managers and designers. This duplication of intellectual and technological effort was perhaps unavoidable, and will probably continue for many years. But there is a need to minimise these differences and to make our limited collective resources go further.

At the international level, I often despair. Most botanists agree on the need for an internationally consistent approach to botanical information, and indeed, **several** international organisations have been set up to do this. Several have been established at national and regional levels, sometimes out of frustration with the perceived shortcomings and lack of progress of other efforts. It is becoming quite difficult to keep track of all these efforts and their current status.

At this point, I will inject a major caveat. I have a major predilection for on-line information, especially through the internet and the World Wide Web. This tool with its enhancements and gateways to underlying applications and data has the potential to meet almost all the requirements of a unifying global botanical information system once it makes its way onto every botanists and prospective client's desk. However, it is quite possible that the very popularity of this technology is the thing that could kill it. The demands on the network infrastructure are increasing so rapidly that within days of a major transpacific upgrade, the link becomes saturated and the downloading or viewing of botanical information becomes frustratingly tedious. Having cast this sombre shadow on exciting opportunities, we can look at, access and in many cases download the following types of information from the internet and elsewhere:

Botanical name files (mostly national or regional)

Authority files

Herbarium specimen information

Type specimen registers

Type specimen photographs

Botanical photograph indices

Botanical photographs

Botanic gardens records

Bibliographic data

Species checklists

Botanical illustrations

Indices of workers and work in progress

Plant distributions and distributions

Geographically based queries

Taxon descriptive data, taxon descriptions

Regional floras

Data standards

Database design specifications

Database models

Interactive identification software

Interactive identification data sets

GIS software

GIS applications

GIS background layers

GIS point locality data

Some of this is free, some of it is restricted, some of it is partial or dithered, some of it is commercial, some of it is cheap, whereas some of it is expensive. Some of it, sometimes vital parts of it, are tied up in CDROM or worse still on paper. But perhaps it might be more informative to look at what is **not available** to the wider botanical community?

A single accepted database design

A single accepted data exchange mechanism

An accepted global set of botanical database standards

A global list of plant names and typification

A complete set of accepted authority files, even though there are accepted botanical database standards

A complete specimen database of any significant plant group

Complete distributional information for any significant plant group

Index Kewensis

TL2

Brummit's list of genera

Kew list of author abbreviations

Brummit's list of regions and codes

What is potentially achievable in Australia?

To a large extent, this depends on groups like us and workshops such as this. This will depend largely upon the resources and expertise that we have available for each project. What is potentially achievable Internationally is largely unknown.

What are the limitations?

The limitations to achieving great things in botanical information systems vary from institution to institution, depending upon:

technology available

management attitudes in institutions

technology in participating institutions

staff resources and expertise in institutions

available authority files

validation routines

available standards

information models

socio-political factors

Specimen Databases

Ken Hill

Australian herbaria joined the Information Technology `bandwagon' initially by developing specimen databases, with the aims of recording their complete holdings electronically. It has since become evident that, while this may have many benefits, there are shortcomings to be addressed. Databases of specimen records do not adequately address many of the needs of herbarium users (as pointed out by Jim Croft in the introduction), and data sets with much wider scope are required. There is also a great need for effective data exchange between specimen databases. Different views on the importance of database structure were voiced, although there was agreement that it was important that the ability to send and receive data in an agreed format was essential.

Scope of specimen databases was questioned when a number of special cases were discussed. These fell into three main categories:

1. Special specimens

- types

- historical specimens

- vouchers for other studies

2. Supplementary collections

- spirit collections

- outsize & carpological collections

- special preparations (slides, pollen) - photographs

3. Non-vouchered records

- supplementary collections with no herbarium voucher - photographs

- photographs of type and other specimens held elsewhere

- observational records

It was generally accepted that all of the above classes were useful and relevant, and that an adequate specimen database should have the capacity to accommodate all of these by using various flagging techniques.

Interchange of data

Alan Brooks

The databasing of herbarium collections received as part of the specimen exchange program results in a duplication of effort and a reduction of resources available for other data-capture programs. To ensure the interchange of electronic herbarium-label information between Australian herbaria, it is necessary for this meeting to re-affirm our commitment to the HISPID standard. Furthermore, a continuation of the meetings reviewing HISPID are necessary. These meetings should be held on a regular basis with definite recommendations proposed at each. In the past, many of the verbal agreements reached at HISPID meetings have not been implemented.

If the interchange of electronic data between herbaria is to become a reality, then procedures and protocols will need to be developed. Although it is recognised that there are difficulties in transferring large data sets, it is proposed to evaluate the effectiveness of transferring these data via email. NSW has developed a protocol for effecting this interchange of data, using the ISO ASN.1 transfer format. Since the ASN.1 standard requires SHORTNAMES as field descriptors, these should be the same as those being developed for the *International Transfer Format for Botanic Garden Plant Records* (ITF) Version 02.00.

Validation standards & procedures

Alex R. Chapman

Herbaria manage datasets which fall into four fundamental categories: **names**, **specimen**, **spatial** and **descriptive** data.

There are some notable international and national projects attempting to coordinate and present valid

nomenclatural information, such as IOPI and APNI. In Australia at least, the natural custodial level for authoritatively maintaining plant name information is at the State level.

Specimen data standardisation and spatial data validation are parallel processes required in any herbarium information system in order to ensure data quality. The procedures now in place at the Western Australian Herbarium (PERTH) use the ERIN data validation process as a model, but refine it by allowing semi-automated validation of geocodes at the specimen level via the use of a customised GIS application.

Is there a role for ERIN as a data validation service bureau for all Australian collections-based agencies? They already provide recognised validation procedures, the ability to interact with their data sets to produce reports and BIOCLIM maps and a sophisticated WWW user interface. A modification to existing tools to enable submission of datasets for validation via a WWW from would provide a significant and standardised service to such agencies, especially those without existing resources for major spatial validation projects.

Descriptive data is largely a synthesis of the other three fundamental data types and as such has the largest potential audience, especially as methods evolve to automate its presentation via tools such as the WWW. Sources will include DELTA-based treatments, floras or monographic works. As these online publications develop, methods for standardising character and character state definitions will be required. Custodianship, procedures and standards for such a distributed descriptive database will become issues as projects cross state and national boundaries.

Database of Plant Names

(The future of APNI)

Greg Whitbread

Greg Whitbread covered the history of the development by the Australian Centre for Plant Biodiversity Research of an upgraded computerised *Australian Plant Name Index*, incorporating the *Census of Australian Vascular Plants*.

The original database had full name strings (combining scientific names, authorship and infraspecific names) in the one field, references in single strings in the one field, etc. The process of 'normalisation' (breaking into the appropriate constituents) via filter programmes and checking has been completed.

The aim was to achieve an IOPI data model standard, but this was very complex. The new APNI, now available on-line, includes no taxonomic judgements, synonymies and a combination with CAVP. Most of the users of the on-line APNI have proved to be non-taxonomists. The new APNI would cater for adding any name and for different points of views on the status of a name.

Now was the time to get APNI up and running on the World Wide Web. This was envisaged to include:

- a series of permission levels (0, 1, 2) for different types of users.
- editing by custodians over the Web. Custodians could be specialists of particular plant groups, herbaria for their State plant censuses.

a structure which allowed every change to every record, everything being logged, and previous records (versions) saved. This was feasible, with 70,000 names at about 4-5 Kb per record; even 10 Gb of data would not be a problem.

Discussion

The meeting discussed the method of permitting only selected authorities, e.g. via a CHAH register of authorities, giving extra importance to the register of taxonomic work on Australian plants. It also considered the scanning of ABLO responses to requests to avoid duplication which inevitably takes place. Storage was an issue, but not impossible. Photocopies of literature requested from the ABLO or from poorly accessible locations could be scanned and incorporated (within the regulations of copyright).

As a result, it was recommended that CHAH urge all herbaria to gain access to the Internet and to make every effort to having taxonomists participate in using APNI as a tool to show their individual developments in classification.

After discussion, a diagram illustrating data flow and custodianship in APNI was developed (refer Figure below). Under this scheme, it will be possible for specialists with conflicting taxonomic views of a particular plant group to present these details in APNI. These differing views will be displayed with clear indication of the ownership of each concept. It would be possible to show the current State plant censuses and the way in which they differ in their concept of the Australian flora. Potentially, any name-linked data could be included under the 'umbrella' of APNI (e.g. descriptions in DELTA or standard formats, images, keys, types and associated information, BRI's register of type photographs, distribution maps, uses, aboriginal and other vernacular names, etc.).

To make APNI easy and attractive for specialist taxonomists to participate in, it was important to provide them with easy methods to construct and edit their classifications (conspectuses). The Fauna section of ABRS has been developing a software tool for use on PC's which might be adaptable to the needs of taxonomic botanists.

As a result of these discussions, various recommendations were made to CHAH relating to:

- its continued support for APNI as a dataset owned and maintained by the Australian plant taxonomic community;
- provision of an 'update' form on the Internet;
- attachment of National, State and regional conservation statuses to names to enable regulation of access to specific location data;
- interchanging information with BRI's Australian Register of Type Photographs;
- extension of APNI to include and display alternative taxonomies where agreed by their custodians;
- development of an application for specialists to develop and maintain their alternative taxonomies.

IOPI

Karen Wilson

Recent IOPI meetings in Madrid reconsidered the start-up procedure for the Global Plant Checklist, given the lack of significant progress in 1995 and no likelihood of extra funding being found without having some preliminary product for demonstration purposes. The aim now is to start a simplified initial Checklist with minimal extra funding, make it available on the WorldWide Web via the IOPI Home Page, and then to seek funds for the fully relational database version described in the Checklist Project Plan.

For each species (or infraspecific taxon if present), there will be three levels of data present in the initial Checklist:

Level 0 - the data for each taxon name as submitted by a contributor, with acknowledgement of source

added; as data for each taxon is added from other sources, this entry will consist of multiple paragraphs, one paragraph per source.

Level 1 - the data partially edited to Checklist format, so that the paragraphs for each taxon are more or less united into one paragraph/statement about that particular taxon (with editor's name added).

Level 2 - the entry is fully botanically edited, i.e. the taxonomic and geographic data have all been checked for that taxon name, and the entry consists of a single paragraph/statement (again with editor's name added).

The basic data to be included will depend, initially, on what is available from each source. However, we would aim at:

- genus + species (+ infraspecific epithet if relevant)
- author and protologue bibliographic details
- type details
- geographic distribution
- family name
- acknowledgement of source of data

The technicalities of implementing this on the Home Page will be discussed with David Green by Berendsohn, Pankhurst and Wilson. Once that has been done, a timetable can be drawn up for implementation and estimates made of the funding needed.

It was agreed to start the initial Checklist with the same three datasets as proposed for the start of the 'full' Checklist, namely those for Australia, Peru and Europe. Other datasets can be added as soon as we have our procedures in place and are sure that we can handle the work-load.

The initial Checklist will be made available on floppy disk as well as on the Home Page, both for users and also for specialists who are botanically editing the Checklist.

Mapping in the herbarium context

Bill Barker

Issues related to providing latitudes and longitudes for collection localities and producing distribution maps.

The input data

Taking latitudes and longitudes from more than one source enabled error checking and, if taken in conjunction with precision ratings, the automatic selection of the more accurate alternative on a case by case basis. In the case of a collector's latitude and longitude being superseded, in AD both the collector's and the more precise ouputs are shown on a specimen label.

By minimising the use of inexact procedures, even if these may seem far too accurate for the data concerned, it is possible to keep the error level of the ouput data as close as possible to that of the input data.

It was important when recording latitudes and longitudes by a GPS to make it clear if the GPS output is in degrees and decimal minutes or degrees, minutes and seconds.

Grid references changed in 1966 to the 10,000 metre Australian Map Grid from a previous 10,000 yard

Transverse Mercator grid (Australian series). In about the year 2000 the World Universal Transverse Mercator grid will be adopted for Australia. Localities in some gazetteers (e.g. the South Australian) have been developed from grid references; it will be more accurate to use these instead of the latitudes and longitudes provided to the nearest minute if a grid to spheroid conversion is available.

I have developed a highly accurate program to calculate latitudes and longitudes from a distance and direction from a locality whose grid or latitude and longitude position was drawn in from a Gazetteer database. This program was developed in TITAN and is for use at data entry stage. (Note: This program and a conversion from grid reference to latitude and longitude can be provided in TITAN language for anyone who requires it).

Discussion

It was agreed that accuracy of the output locality depended not only on that of the gazetted locality used, but also on the way in which the distance and direction were presented. Alex Chapman described the WA Herbarium's more accurate (and expensive) alternative of tracing a distance along a road on a digitised 1:250 000 map and calculating the latitude and longitude electronically. Precision ratings: Herbaria utilising the Adelaide precision ratings were CANB, MEL, PERTH. It was recognised that this was not core data, but there was probably a need in the HISPID standards for a flag to indicate the rating system used to obtain the HISPID distance value. It might be useful to introduce a code of 1.1 for GPS values at the exact point of collection and 1.5 for those in the vicinity (e.g. GPS in the car).It may be possible to develop automation of estimation of precision ratings using feature codes in gazetteers, and the value of the distance and direction and the way they are presented. One could also envisage the development of precision ratings of collectors to account for the varied quality of their calculation of latitudes and longitudes.

Use of data

Currently mapping is largely utilised in presenting simple distribution maps. Software was now available to link with herbarium specimen databases to provide point distribution maps such as traditionally used in taxonomic publications. Ken Hill demonstrated the production of maps from the NSW specimen database (*nswdata*) using a package developed by Kevin Thiele (CANB) using *MapInfo*. The package also provides the small maps suitable for publication in the *Flora of Australia* project.

It was agreed that, in obtaining mapping software, it was important to consider the ease with which it could handle data sets and map images from various sources. *MapInfo* and *ArcInfo* (*ArcView*) were examples of such versatile software, *Viridans* an example of a purpose-built package which was difficult to adapt to other needs.

Bill suggested that it was in the area of mapping that we could create a demand for herbarium data by *value-adding*.

One example was in the development of cluster analysis of point distribution records of taxa. Taken across a whole flora or region, it had the potential to indicate priorities in land reservation in relation to high frequencies of plant 'isolates' (disjunct sets of population records). Such analyses would also be useful in directing survey priorities to address the issues of whether gaps in distribution were real or not and the varying coverage of existing data in time and space.

Ownership of data

(General discussion)

Concern was expressed about the lack of control over the use of data once it had been passed to a second party. It was agreed that there was a need for the source of data to be attached to the data file. Although several other issues were raised, the discussion was brought to a close because most concerns centred on the 'cost of data', which was regarded more as a matter of policy than an IT issue. It was agreed that it was not appropriate for this Workshop to addresse such policy matters.

Electronic Loans Management

Alan Brooks

An electronic *Loans Management System* is being developed at the National Herbarium of New South Wales (NSW). This package will enable the comprehensive management of loans of herbarium specimens. The system provides information on the collections, as well as information on the movement of loan material from its preparation for loan through to its return from loan. Statistics and monitoring abilities are incorporated into the system. All information relating to a loan may be accessed via a central menu facility. The application was developed under the KE Texpress DBMS. The *Outgoing* and *return of Outgoing loan module* is functional. It is planned to complete the *Incoming* and *return of Incoming loan module* as soon as possible.

Client server applications

using MS ACCESS <-> Oracle

Peter Bostock

Queensland Herbarium runs the HERBRECS database as an Oracle relational database on a DEC Alpha server. This server is accessible via a peer-group network (Windows for Workgroups 3.11) within the Herbarium building. Oracle supports the Open Database Connectivity methodology (ODBC) which allows more or less transparent connection between Windows programs and the Oracle database. This connectivity is enabled by software drivers which overlay the TCP/IP network protocols. In effect, the ODBC connection allow programs such as Word for Windows, Excel (version 5), Visual Basic and Access version 2, for example, to open Oracle tables or views, or to query the Oracle database directly.

Microsoft Access version 2 allows views and tables from ODBC sources (which include Btrieve, FoxPro, MS SQL Server and others) to be opened directly, attached (database is remote, but a definition of an individual table is stored locally), or queried via Pass-through SQL. This latter technique directs queries from Access, written in Oracle SQL, to the server for execution by Oracle itself. The results of such queries are then passed back to Access (or Word/Excel etc). ODBC allows people with little or no Unix or SQL experience to connect to the Orace server, and extract data. Users of Word or Excel can perform similar operations via the program MSQUERY which is part of the Excel suite in Microsoft Office.

APPENDIX

Herbarium Database/Information Technology Meeting

2 - 4 November 1995, National Herbarium of New South Wales

OUTCOMES

A. FOR CONSIDERATION BY CHAH

For recommendations 2-4 concerning the *Australian Plant Name Index* refer to the appended flow diagram

- **FORMALISING CHAH IT SUBCOMMITTEE:**

Recommendation 1:

That CHAH formally constitutes an advisory working group or subcommittee to make recommendations to it on Information Technology and databasing, with membership at least of one representative from each Australian State herbarium, to meet in advance of CHAH meetings.

Contact: B.J. Conn (NSW)

- **APNI (and flow diagram):**

Recommendation 2:

That CHAH continues its support for the *Australian Plant Name Index* and encourages its development as a data set owned and maintained by the Australian plant taxonomic community.

Contact: J.R. Croft, G. Whitbread (CANB)

Recommendation 3:

That the Centre for Plant Biodiversity Research be urged by CHAH to include an "update" form on the internet version of the *Australian Plant Name Index* (APNI) so that corrections and new names can be readily entered by registered users.

Contacts: J.R. Croft, G. Whitbread (CANB)

Recommendation 4:

In line with CHAH's previous decision on publication in Australian house journals of precise location data relating to rare taxa,
that the Centre for Plant Biodiversity Research be urged by CHAH to have plants of conservation significance (as in the States' Conservation Listings) added to the *Australian Plant Name Index* (APNI) which will enable regulation of access to specific location data in all herbarium databases.

Contacts: J.R. Croft (CANB), A. Chapman (PERTH)

Recommendation 5:

That CHAH urges ABRS to develop an application for specialists to develop and maintain their alternative taxonomies utilising the *Australian Plant Name Index* (along the lines of that being developed for the *Zoological Catalogue*).

Contact: J.R. Croft, G. Whitbread (CANB)

Recommendation 6:

That CHAH supports the extension of the *Australian Plant Name* Index to include and display alternative taxonomies where agreed by their custodians (in the same way as being developed internationally by IOPI).

Contact: J.R. Croft, G. Whitbread (CANB)

Recommendation 7:

That the Queensland Herbarium be urged by CHAH to modify and expand the type photograph information currently held in the Australian Register of Type Photographs in a form compatible with the *Australian Plant Name Index* (APNI).

Contact: P. Bostock (BRI)

Recommendation 8:

That the Centre for Plant Biodiversity Research be urged by CHAH to collaborate with BRI and add a 'mirror' copy of the Type photograph information currently held in the Australian Register of Type Photographs to the internet version of the *Australian Plant Name Index* (APNI), with a suitable "update" form so that corrections and new entries can be readily entered by all users,
and that the Centre for Plant Biodiversity Research contribute updates of the database in BRI as requested.

Contact: P. Bostock (BRI)

- **QUALITY DATA THROUGH VOUCHERED SPECIMENS:**

Recommendation 9:

Since funding is currently being directed towards non-vouchered databases rather than vouchered ones, CHAH is urged to inform appropriate organisations (granting agencies, data purchasers) of the immediate and longer term limitations and dangers inherent in non-vouchered data and to emphasise the advantages to be gained by investing in vouchered data (such as held in herbaria).

[*Addendum*: The similar potential threat to data quality are voucher specimens of poor quality (as those collected in drought or outside of flowering and fruiting season), often caused by late receipt of funding and inability to carry forward the funding into another year.]

Contact: W.R. Barker (AD)

- **DATA EXCHANGE (HISPID):**

Recommendation 10:

In view of the treatment of HISPID as an unpublished document and its use without adequate citation or acknowledgment in the development of several other data standards, it is recommended that CHAH publish (with ISBN number) the forthcoming edition of HISPID.

Contact: B.J. Conn (NSW)

- **DELTA FORMATTED DESCRIPTIVE DATA SETS:**

Recommendation 11:

That CHAH encourage the continued development and use of DELTA standards in descriptive data sets and associated tools, such as interactive keys.

Contact: A. Chapman (PERTH)

- **DATA VALIDATION:**

Recommendation 12:

That CHAH encourage ERIN to place its data validation service in the public domain to enable its internal use by herbaria and other Australian collections-based agencies.

Contact: J.R. Croft, G. Whitbread (CANB)

- **BARCODING AND EFFICIENT SPECIMEN AND DATA HANDLING:**

Recommendation 13:

That CHAH actively promotes barcoding of specimen record numbers in all Australian herbaria. This process has been successfully adopted in the Western Australian, Canberra, New South Wales and Northern Territory herbaria. Great gains in efficiency can be made in the arena of incoming loans with relatively little outlay.

Contact: A. Brooks (NSW)

B. Agreements and Actions for Workshop Participants

AGREED

The workshop participants confirmed the acceptance by the previous HISPID workshop in 1992 of a standard transfer format (ASN.1) for interchange of electronic herbarium data. It was agreed that field identifiers in HISPID should be abbreviated and that these will be the same as the abbreviated field identifiers used in the *International Transfer Format for Botanic Garden Plant Records* (ITF) Version 02.00 (currently in preparation). Herbaria would participate in a process of evaluation involving data exchange when they had developed their interchange procedures.

ACTION 1

That Barry Conn develop the required abbreviated list of field identifiers for HISPID. The completed list should be circulated to workshop participants by the end of November, 1995.

ACTION 2:

That Alan Brooks and Barry Conn (i) refine the NSW protocol for specimen electronic data interchange; (ii) test it in conjunction with the Centre for Plant Biodiversity Research; and (iii) circulate the resultant protocol for evaluation by each herbarium by end of December, 1995, with the intent of adopting it as the standard for data interchange.

AGREED:

To produce a published version of HISPID

ACTION 1:

That Jim Croft and Barry Conn compare and evaluate the completeness and accuracy of the latest

versions of the HISPID standard as held by CANB and NSW (to be completed by the end of November, 1995).

ACTION 2:

That Jim Croft and Barry Conn coordinate the preparation of HISPID for publication in consultation with workshop participants and other State Herbaria.

ACTION 3:

That Alan Brooks and Barry Conn outline the procedures to be developed in order to participate in the interchange programme, including examples of the output which will be generated, for comment and then inclusion in an appendix to the HISPID publication.

APPENDIX: The new APNI.

Data sets related to Australian plant names.
Collation and distribution of information relating to Australian plants compiled from data in the Australian herbaria for use by taxonomists and general users. (A computerised system being developed by the Centre of Plant Biodiversity Research, Canberra, in collaboration with the Australian State Herbaria.)

17 November 1995