

# ALICE

## software news



FIVE

## contents

<b>System 2.1</b> .....	2	Legumes of Northern Eurasia .....	12
Editions explained.....	3	Books from Kew.....	13
New Alice query engine .....	10	<b>Technical</b>	
New programs .....	15	Graphical data entry .....	6
<b>World-WideWeb</b> .....	5	Importing and exchanging data ...	14
An introduction .....	6	Behind the scenes .....	14
LegumeWeb service .....	7	<b>Miscellaneous</b>	
AliceWeb .....	8	Reduced fees .....	15
Comparing approaches .....	9	User documentation .....	11
<b>Projects</b>		Alice team news .....	16
Uses of <i>Araceae</i> .....	4	Alice homepage .....	16
<i>Allium</i> genetic resources .....	11		

## introduction

*Welcome to issue 5 of Alice Software News. This is a bumper issue thanks to contributions from Alice users and collaborators and because of the significant development of the Alice System during the past year.*

*We have tried to strike a balance between articles describing the software and articles describing how Alice is being used. Our preference is to minimise the technical content, but, inevitably, with the release of a new Alice System this has been difficult to avoid. We hope to reduce this 'techno-glut' in future issues with greater discussion of the use of Alice by projects and organisations.*

### So, what's in this issue?

With considerable pride and some relief we announce the formal release of Alice System 2.1. We describe a few key features of the new system. System 2.1 applications are bundled into 'editions'. We explain what this means and provide some guidance to help you decide which edition best meets your needs. We also announce reductions in the price of Alice System 2.0!

We summarise progress with software development and documenta-

tion, keep you up-to-date with the version numbers of System 2 programs and take a peek at some underlying improvements to the Alice System.

In the first of what we hope will be a regular feature focussing on new functionality, we describe improvements in the way in which Alice databases can be queried.

Publication of data on the world-wide web is a hot topic of general public interest. Richard White and Eduardo Dalcin describe and contrast

alternative ways of making Alice databases available on the web.

Three contributions look at the use of Alice. Kerry Taylor and Peter Boyce describe the "Economic uses of *Araceae*" project, David Astley and Niall Green describe the use of Alice to create an *Allium* genetic resources database and Yuri Roskov describes the creation of a Legumes of Eurasia database.

On the technical side, Michael Soloviev, who works on the Legumes of Eurasia project describes programs for displaying, editing and printing distribution maps from Alice databases.

As an antidote to an over dose of technology and to prove we are indeed human, we devote a column to some of the people who have made all of this possible.

We hope you find Alice Software News interesting and approve of the new look!

*The Editors*

# ALICE

## System 2.1

### DATA ENTRY

Recognising that data entry can be a time consuming process, even with well designed software, we have put considerable effort into improving the speed and convenience of data entry. The underlying data structures have been improved, which will have particular benefits when searching for taxa in large databases. System 2.1 contains a brand new data capture program *Ace* which replaces the *Alice*, *Atext* and *Nview* programs in System 2.0.

Completely rewritten from the bottom up, *Ace* has been given a user interface which is consistent with all other Alice programs.

- Pick lists and string searches can be used to minimise typing and improve consistency, particularly when entering and editing Latin names.
- There are new functions for copying and deleting data for taxa. Using a few key strokes, complex species descriptions can be copied to other taxa and then edited.
- There are improvements to the way data can be imported.
- Data validation has always been a strength of Alice. We have improved it further.

*We made it! Alice System 2.1 is here and, we believe, well worth the wait. We still haven't been able to add all the features we would have liked, but System 2.1 is a very significant advance on System 2.0. Further features are now being tested and will be added in the coming months.*

*What is new about System 2.1?*

*We make no attempt to describe all the changes here - there simply isn't space. Some advances were described in Alice Software News 4 and we repeat here only the most important additions and changes.*

- There is greater flexibility in exploring data and generating reports during data entry. You can, for example, find out which species have been recorded for a particular descriptor or state before deciding to edit that dictionary or descriptor definition.
- A new, simpler interface for searching for taxa either by scientific or common name. Greater control over which categories of taxa and class of name are included in searches. Searching can be done using a form or selecting from pick and point lists sequentially. Searching for taxa and creating a new one are now clearly separated.
- Includes a "citations manager" which is easier and more flexible to use and again allows the user to control program behaviour.

### MULTI-USER

The second significant change is that Alice databases can be placed on networked drives and several different people may either query data or enter data into the same database. This has been done without sacrificing ease of use for those users who work on their own. Alice is not completely multi-user yet, but we are purposefully moving towards a version which is, by supporting databases that can be accessed and used institution-wide. We intend to provide the best of both worlds; software that is easily set up and used on personal computers and yet which can be used to maintain shared institutional information assets. Forthcoming developments of *Sam* (see page 14) will make it easier to combine personal or small project databases so as to contribute to larger institutional databases. System 2.1 includes tools to help database administrators maintain shared databases and manage a group of Alice users and Alice databases at a networked site.

### MULTI-TASKING

System 2.1 is designed to be used in a multi-tasking environment. You can, for example, safely run the *Awrite* report writer at the same time as using the data capture program, *Ace*. This was possible with System 2.0 but wasn't recommended since this could lead to data corruption. We have tested System 2.1 under Windows 95, OS/2 and versions 3.51 and v4 of Windows NT.

### QUALITY

In System 2.1 we have fixed all known bugs in System 2.0.

# “ Editions ”

## E X P L A I N E D

*Alice System 2.1 opens up new possibilities for sharing data within projects or institutions. We have combined the programs comprising Alice System 2.1 into three editions, for Database author's, Database administrator's and End users, which reflect some of these possibilities. Below, to help you decide which edition is suitable for you, we examine four scenarios of increasing sophistication. The simplest, “A”, describes a user working alone. Scenario “B” presupposes a database author who wishes to distribute databases to others, possibly at different locations. Scenario “C” describes a situation in which the same database is to be searched and possibly even updated and modified by more than one person. The last scenario, “D”, describes how databases might be maintained on a shared network or at different sites and combined for subsequent management or use.*

### A) SINGLE USER

If you work on your own and do not wish to share your database with anyone else, other than those who use your computer, then you will need only purchase the *Database author's edition*. Many users currently working with Alice System 2.0 will fall into this category.

### B) DISTRIBUTING YOUR DATA TO OTHERS

The simplest and most flexible way of sharing your data with others is to supply them with copies of all, or part of, your Alice database. As database author, you will require a copy of the *Database author's edition*. Everyone else, unless they work on the same machine as you, requires an *End-user's edition*, which includes all the necessary tools for interrogating databases and writing reports. This is useful for database author who wish to ‘publish’ data in a flexible electronic format but do not wish

end users to add to or modify their data. We have structured our pricing to make electronic publication attractive. The more licences of the *End-user's edition* you require for your project or institute, the cheaper each copy becomes.

Should you need to distribute data for manipulation using other software, as a database author you still have the option to use *Alex* to export subsets of your database into a variety of formats.

### C) SHARING DATABASES ON A NETWORK

People who wish to work together on the same copy of an Alice database can take full advantage of the new features of Alice System 2.1. If you are to use Alice in this way then we strongly recommend that one person within your organisation is given the role of database administrator (DBA). A DBA is responsible for

organising and managing all the Alice databases, possibly configuring each work station, placing the various Alice databases on networked drives, ensuring all users have appropriate access privileges and resolving any problems that might arise. In such an environment at least one *Database administrator's edition* of Alice 2.1 is required. At small networked sites the roles of database author and database administrator can often be combined, in which case a separate licence for the *Database author's edition* is unnecessary. A separate licence for a *Database author's edition* is required, however, for each individual machine upon which data are entered or edited.

### D) SHARING DATA FROM MORE THAN ONE SOURCE

Perhaps the most sophisticated use of the Alice system is where several database authors work in parallel on their own separate databases and, at some agreed point, combine their efforts into a single database. Such ‘federated’ databases may be built by large multi-institutional projects, such as the “International Legume Database and Information Service” or “Plantas do Nordeste” or at larger institutions where the intention is to combine the work of several staff members into an institutional species database. Authors of the component Alice databases may commit all of their own data or just certain subsets of the taxa or datatypes to the combined database. The individual or team that provides central co-ordination of these databases will need a licence for the *Database administrator's edition* since this includes the program *Sam* which can be used to import data sets into Alice databases and to merge Alice databases. Use of *Sam* at any particular institute clearly does not preclude several authors working together much more closely on one database, as in scenario “C”, whilst others work more independently of one another.

Making the most of the available human and data resources in such scenarios requires considerable organisation and co-ordination. We are happy to advise those considering setting up such federated or institutional information management systems.

# Economic USES

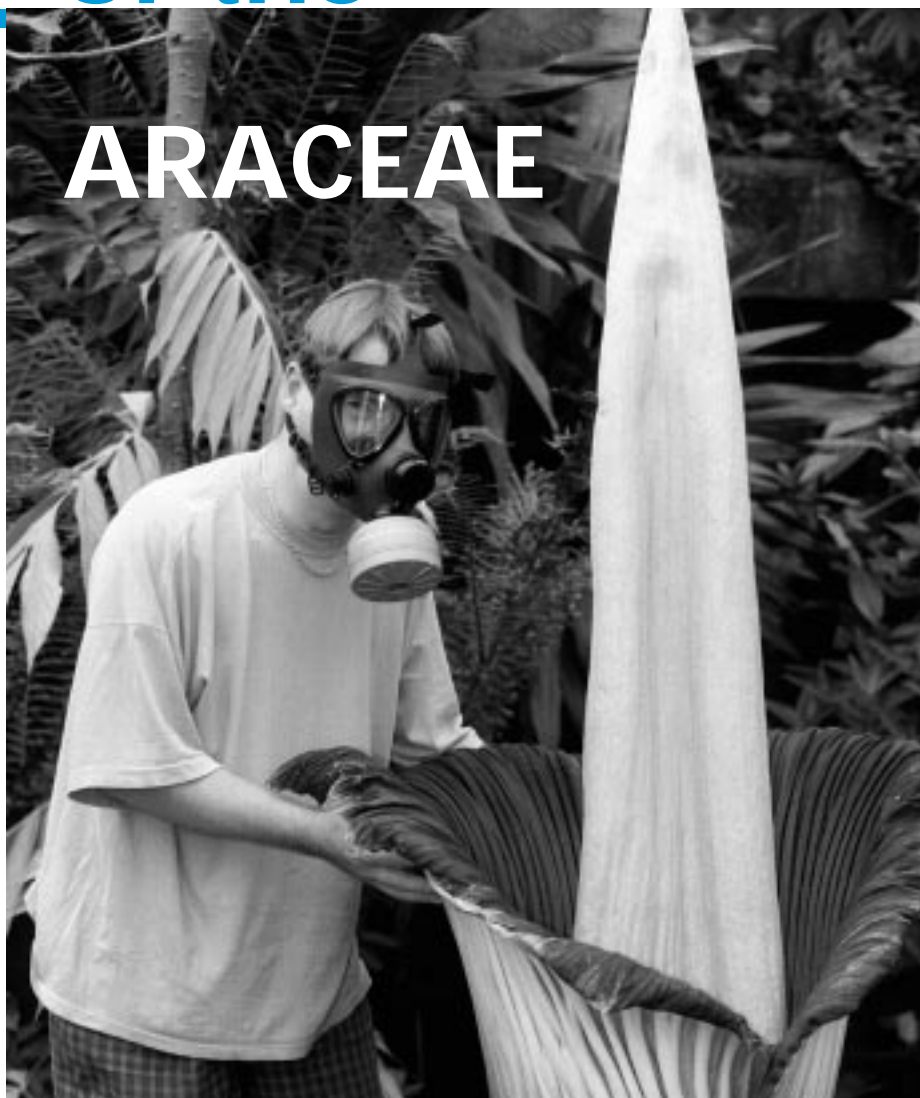
## of the

# ARACEAE

This is the most recent of the many species databases created using Alice in the Herbarium of the Royal Botanic Gardens, Kew. It differs from its larger monographic predecessors (e.g. for Palms, Orchids, Legumes and Cyperaceae) in its emphasis on information about the economic importance of the taxa described. The project was initiated by Peter Boyce and the collation and entry of information is being carried out by Kerry Taylor, a sandwich student from the University of the West of England, who is working at Kew for a year under Peter's supervision.

About 500 species of aroids have economic or ethnobotanical importance but the information about these species and their use has never been brought together. Data are being assembled from a wide range of sources. Apart from scanning the printed literature available in the various libraries at Kew, Index Kewensis and systematic searches of Kew's Herbarium collection (approximately 30,000 sheets, including 1,200 type specimens), much use is also being made of the Internet as a source of information about commercial usage. Yet further information comes by letter or email from colleagues around the world - particularly from members of the International Aroid Society whose headquarters is in Florida.

The first step to creating a database was to download into an empty Alice database all the available nomenclatural data from the Index Kewensis database at Kew. This has since been edited to provide a modern and coherent view of the taxonomy of the family and to date the database contains 524 taxa of which 498 are species and 26 are infraspecific taxa. For these taxa the database contains a total of 1,160 different scientific names including synonyms, suspected syn-



Express Newspapers

*Peter Boyce gets to grips with an *Amorphophallus titanum*, an aroid of surprising economic importance to Kew Gardens. Hundreds of visitors were attracted by its scent last summer.*

onyms, orthographic variants and misapplied names. 450 taxa share 433 different common or vernacular names.

Most of the information about plant use is stored as 61 separate Alice user-defined descriptors. 1,263 separate data observations have so far been recorded for the 524 taxa using these descriptors and much additional information stored as structured notes or as free text using the *Atext* program. 855 different publications have been referred so far as well as 1,148 other types of citation.

The project started in August 1996 and will finish in the summer of 1997. The data then available will be used with the *Awrite* program to prepare a checklist of economic and ethnobotanically important Araceae which we hope to publish at Kew. Once published we hope to make all the information available on Kew's World-wide Web site using *AliceWeb*.

For further information please contact: Kerry Taylor (k.taylor@rbgkew.org.uk) or Peter Boyce (p.boyce@rbgkew.org.uk) at the Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK.



# Putting Alice

## databases

### on the World-wide Web

Four separate articles on the following pages describe how you can make the information contained in your Alice database available to colleagues via the World-wide Web or your institutional intranet.

Richard White first introduces the World-wide Web. In a second article, Richard describes the "LegumeWeb" service by which the ILDIS project is making its Alice database available by enabling remote users to search the database using scientific names. *Araneus*, a program developed by Richard, generates HTML pages in response to these questions.

Eduardo Dalcin describes the *AliceWeb* program which he has developed and is available as a component of Alice System 2.1. *AliceWeb* creates a set of linked HTML pages from any subset of your database. Other people can then view or search your data using descriptors, uses, habitats, higher taxa and vernacular or scientific names (including synonyms). Two projects that have already made good use of *AliceWeb* have been Eduardo's own project in Rio de Janeiro (see Alice Software News No. 4) and a global database of palms by Dr. John Dransfield at the Royal Botanic Gardens, Kew.



Finally, Richard and Eduardo compare notes about the relative merits of their alternative approaches and point to possible future collaboration.

#### PLEASE TAKE A LOOK AT SOME ALICE DERIVED HTML PAGES

Links are being added to Alice Software's main (home) page which lead to HTML pages generated from Alice demonstration databases using the *AliceWeb* program

[<http://dspace.dial.pipex.com/alice>]. Alternatively you can view pages with information about Legume species generated by both *AliceWeb* and *Araneus* programs, by following links from the ILDIS home page [<http://www.ildis.org>].

You can contact the authors of the following articles at these addresses:

**Eduardo Dalcin,**  
**Programa Mata Atlântica,**  
**Jardim Botânico do Rio de Janeiro,**  
**Rua Pacheco Leão 915,**  
**Rio de Janeiro, 22.460,**  
**Brazil.**

**Email: [e.dalcin@biosys.alternex.com.br](mailto:e.dalcin@biosys.alternex.com.br)**

**Dr Richard J. White,**  
**Biological Sciences,**  
**The University,**  
**Southampton**  
**SO16 7PX U.K.**

**Email: [rjwhite@soton.ac.uk](mailto:rjwhite@soton.ac.uk)**

# WORLD WIDE WEB

for its astronomic growth lies in the interlinking of pages. The author of a page can make particular words, phrases or pictures into links which, when clicked on with the mouse, lead to another page which is automatically fetched and displayed in place of the present one. To follow the pages we require no special skills, only to be able to point at what we want!

This works whether the individual page we want is on a computer in Adelaide, Delhi or Edinburgh.

It's a new kind of publishing system. With the Web, many more people can create

pages which can be seen by millions of others around the world.

The Web can be a (slow) way to advertise products and services but is also a delivery medium for them. It is an easily learnt standard way for ordinary people to use many kinds of Internet facilities.

## WHAT ABOUT ALICE DATABASES?

Many of us use Alice to collect and organise biodiversity information, at least some of which we then wish to make available to others. How can we use the Web to do this? I have tried to answer some of these questions and learnt by trial and error rather than reading theoretical treatises. The situation is developing rapidly and sharing of ideas should help us all to move forward.

Present versions of Alice run as DOS applications under all versions of Windows, including Windows NT but are unable to communicate with the Web directly. The good news is that there are at least two ways already being used to make Alice databases available on the Web as the articles on these pages describe further.

*Richard White*

## AN INTRODUCTION

I assume that nearly everyone by now has heard of the World-wide Web. Briefly, people all over the world are making pages of information available on computers connected to the Internet, the world-wide network of computers. The pages are stored in a standard format called HTML. The beauty of the system, and the reason

## Graphical data entry and mapping of distribution records

*The "Legumes of Northern Eurasia" project (see page 12) has developed four memory resident programs for use with the Alice system. The first of these provides for the graphical entry of distribution records for species in our database while the other three display species distributions as recorded in any Alice database.*

*Areas* allows users to select geographical areas from a map displayed on the screen when entering species distribution records from local floristic accounts.

*Regview* displays a screen image of the species distribution of the current taxon using the Northern Eurasian region of our own local Alice database.

*Worldmap* displays a screen image of

the species distribution of the current taxon on a map of the World when working with Alice databases that contain world-wide distribution records.

*Mapscan* generates a \*.BMP format file containing an individual species distribution map that subsequently can be inserted into reports generated by *Awrite*.

These programs operate as intermediaries between the user and Alice software. To be used, each program must first be loaded into the computer's memory (RAM) before you run Alice. Once in memory, each program can be invoked from the keyboard using particular keys, whilst you are using any of the Alice applications. The programs either capture Alice normal text output and display it as a map or allow users to select places upon a map and then emulate data entry to the Alice database.

It would be most convenient for our users if they were able to use these programmes from within Alice applications rather than having to pre-load them into memory first. I think that it would be a good idea if Alice programs were provided with a "macro command option" through which separate function such as these could be invoked.

Michael Soloviov  
Komarov Botanical Institute,  
St.Petersburg, 97376 Russia

# THE ILDIS "LEGUMEWEB" SERVICE

I work at Southampton University, where I help the ILDIS Co-ordinating Centre. The International Legume Database and Information Service, ILDIS, is one of the most advanced species database projects of global scope for any group of organisms. The ILDIS database is built, edited and queried using Alice System 2.0 and recently I have been experimenting with making the ILDIS database available on the World-wide Web at: <http://www.ildis.org>

A trial version of the ILDIS Legume Database, which we plan to update every ten weeks, can be found on the Web at:

<http://www.ildis.org/LegumeWeb>

A simple name entry screen will be displayed allowing you to search the database for any of the 19,000 taxa using any of the 36,000 scientific names so far entered. The query form is simple. In the name entry boxes enter either a genus name, species name or both. You can use an asterisk (\*) to widen the search or if you are uncertain of a spelling. The asterisk will match as much as it can, or nothing at all. For example, "canade\*" will match "canadensis", "canadense", "canadenopsis" etc. The asterisk can be used at the beginning, at the end, in the middle of a word or instead of a complete word.

When you enter a name the database is searched for matches. If sev-



eral are found, you are offered a list so that you can click on the name you wish to proceed with. If the chosen name is a synonym, a brief explanation is given before ILDIS's preferred name is shown, followed by a page of information about that species. Near the end are buttons that point to other databases which may contain information about the species; at present these are Tropicos at Missouri Botanical Garden, the Australian Plant Name Index and the Riken database of root-nodule bacteria in Japan.

## LIMITATIONS

There are some present limitations to the *Araneus* program which provides the LegumeWeb service. It is specific to the requirements of ILDIS and its particular version of an Alice database. Bibliographic reference numbers are currently shown, but lead nowhere. In the next version, the numbers will be links to entries in the bibliography as in *Awrite* reports. ILDIS has decided that while the entire ILDIS Phase 1 database can be made available on the Web, data should be available to users only one species page at a time so that users cannot download the whole database at once.

The present version is slow which is due to the data structure used for searching. The advantage was that I could get something working relatively quickly which can be modified as we get comments and our plans become clearer. I intend to replace the slow bits with, well, something faster!

## FUTURE DEVELOPMENTS

It should be easy for users to select just those kinds of information they want to see, such as nomenclature and distribution, instead of getting the whole lot. It would also be possible to search the database on criteria other than the name of a species, such as by country, or use, but this is not currently planned by ILDIS. Links from the ILDIS home page lead to experimental mock-up pages showing what might be available in the future, including images, which are not currently available in the live database.

*Araneus* is fairly simple and would be relatively easy for me to customise. In principle, therefore, it can be made available to anyone with an Alice database. It does require nomenclatural files in LegumeWeb format, however, which I create from *Awrite* reports. This format may change as the database access software is optimised.



Richard White

# "Surfing" your data with AliceWeb: a new program for Alice users

Since 1989, the "Programa Mata Atlântica" (PMA) in the Rio de Janeiro Botanical Garden has undertaken phytosociological inventories of areas of "Mata Atlântica" - remnants of the forests once found along much of the Atlantic coast of Brazil. In 1993 we converted our species databases to use Alice (see Alice News No. 4) and currently have a central database of about 1,800 taxa.

More recently we needed to make our data available on our local network and our solution was to create *AliceWeb*. This program is a general tool for creating HTML pages directly from Alice databases which I hope will be useful to many Alice users. I developed the program during 1996, with the knowledge and collaboration of Alice Software, and describe here my approach and its significance to our own project.

The 'Programa Mata Atlântica' has:

1. many research staff who contribute to or use, our central species databases.
2. staff in different buildings all linked via a local area network.
3. an Information Services Centre with qualified staff that is dedicated to the administration and curation of the project's databases.
4. a regular influx of new information from our field surveys. Some data is passed into our species database directly from a collections database but where taxa are entered for the first time we need to encourage research staff to collate and enter morphological, ecological, phenological and other data about these taxa.

## THE NEED

To date we have been using Alice System 2.0 which does not permit simultaneous access by more than one user to our central database (see *articles in this issue describing the benefits to data sharing of Systems 2.1 and 3.0 - Ed.*). This has complicated data entry by research staff and led to some

duplication of effort. Our goal was to provide an interface between the research staff in their offices and the centrally held databases. Research staff needed to be able to

1. query the data easily.
2. access the data simultaneously throughout the Garden, limited only by the access restrictions imposed centrally.
3. update, edit or add to the data in the database via the *PMA database administrator*. We did NOT wish to allow all research staff to be able to make changes to the database itself without the prior knowledge or control of the DBA.

## ALICEWEB

We wanted a system capable of mirroring the data within our Alice database in a format that would be available within our multi-user environment whilst ensuring that the database administrator has control of data. HTML pages were an obvious medium. Programs to display them were already available cheaply (Web browsers) and we wouldn't have to train our scientists (most are already familiar with use of the WWW). All that was needed was a tool to migrate data from our Alice databases into this new format.

I developed *AliceWeb*, using Foxpro Version 2.6. It translates Alice data sets into a suite of interlinked HTML pages. Rather than read data directly from the Alice database itself, *AliceWeb*, reads data from ATF (Alice Transfer Format). This gave us several important advantages:

1. ATF format is well documented

and relatively stable - as the developer of *AliceWeb* I wanted to be protected from changes to the internal file structure of Alice databases as the system evolves.

2. By making use of the flexibility already provided by the *Alex* program in generating ATF data sets, users of *AliceWeb* can select exactly that subset of their database which they wish to make available as HTML pages.
3. Authors of other database systems, desirous of making *AliceWeb* available to their users may also provide an option to export data into ATF (as the *Delta and Brahms* systems already do - Ed.)

## BENEFITS TO THE "PROGRAMA MATA ATLÂNTICA"

The most important result for us has been a dramatic improvement in data quality. Establishing a direct link and daily use of the databases by our research staff has meant that they criticised the data on view, often energetically, and began to assume an ownership and responsibility for those data which they had previously tended to leave to the staff in the Information centre. This led to an amazing increase in the volume of corrections and additions to our database but everyone is now more motivated as their efforts are recognised and valued by other members of the group.

From this beginning we have generated a hypertext information system for the whole research programme in which our Alice database is but one part. We have bibliographic information, an interactive calendar of collecting trips, a contacts directory and our collections database (including the generation of herbarium labels) all "integrated" and available via one interface using the Web Browser. This has helped bring many more research staff into daily contact with our data and a growing confidence in our data as a result.

## LIMITATIONS

The initial creation of the suite of linked HTML pages is relatively slow and involves two steps: use of the *Alex* program to create an appropriate ATF data set and then use of *AliceWeb*. The HTML pages for a large database obviously take up a large amount of disk space but this, happily, has not been a limitation in



our case and more than offset by the speed at which the user can access the data as a result.

One limitation of *AliceWeb* is that it is not possible, as it is with *Aquery* or *Awrite*, to generate lists of taxa by building complex queries using more than one question. Our emphasis was on the display and maintenance of the data rather than to duplicate the effective tools for data analysis that the Alice system already offers.

### THE FUTURE

The HTML format, with its increasing number of extensions, could quickly become an efficient common interface for accessing biological databases (for example, the *Species 2000 project* is planning for this - Ed.). We anticipate that *AliceWeb* will increasingly be useful to many Alice users and are developing the program further. Future possible enhancements include:

- automatic incorporation of images within the species HTML pages created (*This has now been*



*Eduardo Dalcin (second from right) and other members of the computing support group in the Rio de Janeiro botanical garden*

*implemented. Ed.).*

- organisation of pages into subdirectories to facilitate management of Web sites.
- a single nomenclatural index to include all scientific names rather than separate lists of synonyms and preferred names.
- "Edit" buttons on species pages

that will allow end-users to generate pre-filled data editing sheets to facilitate generation of "data update requests" to be filed, electronically, with the database administrator(s).

- password control to limit access to different levels or types of data.

**Eduardo Dalcin**

## Comparing Araneus with AliceWeb

*Araneus*, unlike *AliceWeb*, generates and displays an HTML page for a species as the end user asks for it; the page is not stored. *AliceWeb*, on the other hand, is run by the author of a database (in Windows on his local machine) to generate all of the linked pages that might be needed in advance of any user search. The information contained will be all of that which the author selected to export from their Alice database in ATF format. You can then transfer this set of linked pages to the Web server computer.

### Advantages of AliceWeb are that:

- it creates numerous indexes which permit you to "surf" your data finding those taxa which have a particular attribute or use or occur in a particular country for example. By consulting vernacular names and then clicking on "abobrinha-do-mato", for example, you may discover two species which share this name. Selecting one of them, you can go on to discover, after two more clicks, which other species share the same geographical distribution and from which bibliographical references this information derives. LegumeWeb only allows users to search the database by scientific name.
- the species pages load faster when end-users make enquiries than do the pages generated dynamically by *Araneus*,
- proper links are included to bibliographic references or other data sources.
- it is available for Alice users now and is not restricted to any particular Alice database.

### The advantages of *Araneus*, the first two of which arise from its dynamic generation of pages, are that:

- you do not need lots of disk space to hold the pre-formed HTML pages.
- possibilities exist for future customisation to allow general complex searches
- you can undertake searches of scientific names using "\*" wild characters.

### Combining the two approaches

Our two approaches are complementary. While *AliceWeb* has no pattern matching capability, it provides indexes to allow searches using any datatype. It is possible that we will be able to combine our approach, using *AliceWeb* to generate all the index pages (for descriptor states, uses etc.) and using *Araneus* to generate species pages from the database when requested. We are currently discussing this.

Eduardo Dalcin and Richard White

# New Functionality: a new Alice Query Engine

*This is the first of a regular "New functionality" slot. Each issue of ASN will contain the description of one particular area of new functionality. This time we describe the new Alice query engine: that part of our programs that allows users to ask questions about the species in their databases using <AND>, <OR> and <NOT> operators.*

The query engine in System 2.0 was first implemented in *Aquery* and subsequently improved upon in releases of *Awrite* and *Alex*. Some of our long standing (dare we say "die-hard") users, are so fond of this interface that some have threatened to sulk, or worse still, should we ever dare to change it! Times move on, however, and a fresh approach became inevitable.

The new query engine, available in Systems 2.1 and 3.0, has been completely rewritten. This has improved performance and provides a framework which makes it simpler to make further enhancements. This new version has lost none of the flexibility and ease of use of the existing version.

## THE ADVANTAGES:

- 1 **Faster performance:** particularly noticeable with larger data sets and more complex queries.
- 2 **Consistency** between all our programs. The minor differences between *Aquery*, *Awrite* and *Alex* have disappeared as a result of the reorganisation and documentation of the Alice programming libraries undertaken during the last eighteen months.
- 3 **Greater precision and clarity when searching nomenclatural data.** Some users experienced difficulty when asking questions involving higher taxa or genera. The new interface makes

things much clearer - offering only accepted generic names for example, while the enriched nomenclatural data model in System 3.0 enables much more precise queries to be defined.

- 4 **New search operators.** The System 2.0 query engine had only two search operators: "Equals to" or "Not Equals to". The new version will offer the following four operators for **all** data types:
  - **Equals:** use this to find all those taxa recorded for the chosen data value (e.g. all species having "red" flowers).
  - **Not Equals:** use this to find those taxa that have NOT been recorded as having the selected value (e.g. "red").
  - **Unique:** use this to find those taxa that have been recorded as having the chosen data value alone (e.g. "red"): no other data value has been recorded for that data type. This operator will be useful, for example, to list those taxa that are endemic to a particular place.
  - **Missing:** use this to find those taxa for which no data values have been recorded for the chosen data type (e.g. "habit").
- 5 **Comparative search operators.** The new query engine also offers the following four operators for use with User Defined Descriptors that have been pre-defined as "Ordered" (e.g. height, month, number etc.)

- **Greater than**
- **Greater than or equal to**
- **Less than**
- **Less than or equal to**

- 6 **A Missing Data Filter** will let you decide, for each part of your query, whether you want to include or exclude those taxa which have no data at all recorded. For example, when you search a database for all plants recorded as trees the result will **always** include those taxa for which "habit" is recorded as being set to "tree" but you **may** also decide to include those species for which data about "habit" is missing altogether and therefore **might** be trees. The filter will be available for use together with any of the above search operators except for "Missing".
- 7 **A Query History Manager** will enable users to store, and name, any complex queries that they may wish to use again at a later date. Such queries will be stored in and recalled from a catalogue of useful existing queries.
- 8 **A Query History Editor** will allow users to edit any query that they have stored previously in the catalogue. You can add, delete or modify the individual components of a complex query and save the changes before executing that query again.

The Query History Manager and the Query History Editor are available in the System 2.1 applications *Aquery*, *Awrite* and *Alex* as released. The other features described above are currently undergoing tests and will probably become available in the next release of System 2.1.

# A PLANT GENETIC RESOURCE DATABASE FOR ALLIUM

The Genetic Resources Unit of the Horticultural Research Institute (a gene bank devoted to cultivars of crop vegetables) and the vegetable section of the Scottish Agricultural Science Agency (the registration centre responsible for National List and Plant Breeder's Rights testing of vegetable species in the UK), are collaborating to develop an *Allium* germplasm database. The database will be built using Alice System 3.0. The purpose is to create a shared resource of genetic information about cultivated and wild *Allium*, which will be used by the gene bank for identifying genotypes of interest to plant breeders and researchers, and to assist with requests for seed. The registration centre will use the database to

select cultivars to be grown in field trials for further evaluation or as an aid to characterising other taxa.

The exact circumscription of the database has yet to be agreed but is anticipated to cover 8 to 9 crops including some 1600 - 1700 taxa, the majority of which will be cultivars. The database will use a common descriptor set including origin and up to 400 descriptors covering passport information, morphology, and economic importance. The morphological descriptors will be based on those used by the International Plant Genetic Resources Institute (IPGRI) and the Union for the Protection of Plant Variety Rights (UPOV).

This is the first time in Europe that a gene bank and a registration centre have collaborated to create a shared information resource. Gene banks hold seed collections but are often have poor characterisation data for the seed in their collection. Registration centres collect good quality descriptive data, often over many years of field trials and a major goal of our collaboration is to make this information available to the gene bank in a useful way.

D. Astley, Genetic Resources Unit, HRI, Wellesbourne, Warwickshire.

F.N. Green, SASA, East Craigs, Edinburgh, EH12 8NJ. Email: green@sasa.gov.uk

## User

## Documentation

### USER GUIDES FOR SYSTEM 2.0

We continue to employ Writelines Ltd. (a UK based technical documentation company) to prepare and produce User Guides to our software. In the last twelve months User Guides for *Awrite* and *Alex* have been completed and distributed to all registered users. If you didn't receive yours then please let us know.

Portuguese versions of the *Aquery* Guide have been distributed to users in Brazil and a Portuguese version of the *Awrite* Guide is complete and being printed. The *Alex* User Guide is to be translated.

### USER GUIDES FOR SYSTEM 2.1 AND 3.0

Most effort on documentation is being focused on production of User Guides for the program *Ace* for Systems 2.1 and 3.0. After consultation with users we felt that although a User Guide for such a large and flexible program as *Ace* was necessary for reference purposes, there was still much merit in the tutorial approach which we had adopted with the documentation for the *Alice* program for System 2.0. Consequently, a tutorial for *Ace* is to be written at the same time as the User Guide. Versions of this tutorial will be available for both Systems.

*Some recent Alice Manuals*



# LEGUMES

## of Northern Eurasia

*A database of more than 2,250 species of Leguminosae from Northern Eurasia has been compiled at the Komarov Botanical Institute (St.Petersburg) using Alice System 2.0. Taxonomists Andrey Sytin and Yuri Roskov have been building the database since 1992 working under the scientific supervision of Prof. G.P.Yakovlev. Our team participates in the International Legume Database & Information Service Project (ILDIS).*

The database currently contains information about more than 2,300 native, adventive and cultivated legume taxa recorded from the former USSR and Mongolia. In August 1991, we started compiling the database using a version of Alice that was restricted to ILDIS's needs. Our database contents then reflected exactly those data categories chosen by the ILDIS project as part of its "Phase One" implementation. Later, however, we decided to include more information in our own database so as to provide a more detailed regional synthesis than the minimum data content required by ILDIS. We moved to the normal (non-restricted) version of Alice therefore, and began adding new descriptors of our own definition. Five new descriptors (see panel) were entered and have since been recorded for all species.

Further modifications were made to the original ILDIS data definitions. The ILDIS descriptor "Climbing / Not Climbing", for example, was replaced by the seven state descriptor "Stem orientation" (see panel) and a far more detailed geographical scheme of Northern Eurasia was defined to include 6 macroregions, 19 countries (including 4 artificial parts of large territory of Russia) and 224 areas within those countries. We also slightly modified the classes of economic importance to

### SPECIES DESCRIPTORS

*Defined at the Komarov Institute additional to those used by ILDIS*

Chromosome number	with 41 states
Stem orientation	with 7 states
Vegetation Type	with 43 states
Substrate	with 26 states
Aspect	with 10 states

better match an existing Russian data standard. Finally we decided to include a complete list of vernacular names for each species including names from all the different FSU languages accompanied by an abbreviation version of the name of the language.

By the end of 1995, a full report was generated from our database using the *Awrite* program and which we then processed using Word 6.0 for Windows and Alice Word macros. The final result was published in 1996 by the Royal Botanic Gardens Kew (see panel opposite). Since then further data entry work has gone on and our database will soon be merged into the central ILDIS database which is currently housed at Southampton University.





As part of our project we were keen to see mapping facilities linked into our database. A computer programmer Michael Soloviev (see separate article page 6) developed four memory resident programs that facilitate both compiling and displaying geographical distribution records in Alice databases.

#### Database statistics

Species 2,262  
 Subspecies & vars. 46  
 Provisional taxa 28

Synonyms 1,650  
 Misapplied names 32  
 Orthographic variants 56  
 Nomenclatural Citations 7,207  
 Vernacular names 3,002

Project descriptors 10  
 Descriptive observations for taxa 26,791

Geographical places 262  
 Distribution records for taxa 28,558  
 Categories of use 21  
 Uses recorded for taxa 1,224

Text notes 5,567  
 Data sources 733  
 Literature pointers for taxa 9,881

To conclude, I am happy to say that during our five year project, during which time we have used almost all of the System 2.0 programs, we have been very satisfied with the reliability and flexibility of Alice software. There has been no loss of data, the software does an excellent job of ensuring data consistency and is sufficiently easy to use that others in the herbarium can query our database without troubling us compilers. Of particular importance to us has been the ability to add descriptors and modify them as the project progressed and to be able to design our own reports exactly as we wanted them to be. Also important have been the flexible ways in which data can be used once stored and the security offered us by the data quality enforced by Alice and the data export facilities.

Yuri Roskov  
 Herbarium,  
 Komarov Botanical Institute,  
 St.Petersburg 197376, Russia,  
 Email: [roskov@herb.bin.ras.spb.ru](mailto:roskov@herb.bin.ras.spb.ru)

## Books published by the Royal Botanic Gardens, Kew originating from Alice databases

In 1989 "Legumes of Africa" was the first book to be published by Kew Gardens directly from an Alice database. The latter was built by Mike Lock as part of Kew's contribution to the International Legume Database and Information Service (ILDIS). Since then Kew has published a number of books whose contents have been generated as *Awrite* reports from Alice databases. Several other Kew staff are now preparing to publish their databases either as books or on the World-wide Web.

	No. of taxa	Price (£)
<b>Legumes of Africa</b>		
J.M.Lock (1989) 619 pp.	5825	15.00
<b>Orchids of Vanuatu</b>		
B.Lewis and P.Cribb (1989) 171 pp.	158	12.00
<b>Legumes of West Asia</b>		
J.M.Lock and K.Simpson (1991) 263 pp.	2015	15.00
<b>Legumes of Indo-China</b>		
J.M.Lock and J.Heald (1994) 164 pp.	994	9.00
<b>CITES Orchid Checklist</b>		
J.A.Roberts, C.R.Beale, J.C.Benseler, H.N.McGough and D.C.Zappi (1995) 136 pp.	1349	7.50
<b>Legumes of Northern Eurasia</b>		
G.P.Yakovlev, A.K.Sytin and Yu.R.Roskov (1996) 724 pp.	2260	30.00

Any of these books maybe purchased by writing to the Mail Order Department, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK.

All prices quoted are in pound sterling (£). Please add 15% for Packing and Surface Mail Postage. Major Credit Cards are accepted. Please send number, expiry date, and registered address of card (i.e. the address to which the account is sent).



# Importing data and exchanging data between projects

One of the strengths of the Alice system is the high degree of internal consistency and logical integrity that it imposes upon your data. Very few programs available to biologists, however, are as strict in this regard. Before data from other sources can be imported into Alice databases, therefore, they must be rigorously tested and this is one of the uses for which *Sam* is intended. Interestingly you could also use *Sam* to test the quality of other people's data sets by importing them into an empty Alice database, even if you had no subsequent intention to use Alice programs to manage those data.

Alice users, whether researchers, institutes or projects are increasingly interested in being able to exchange data. A whole new set of data integrity issues arise when exchanging data between projects. Inconsistencies between the data sets, such as differences of opinion, alternative spelling or nomenclatural changes, may mean that two databases, even when built with the same software and stored in the same format, cannot simply be stuck together end-to-end. Careful comparison is needed when two sets of data are merged.

*Sam*, when importing data from whatever source, tests each data item, bit by bit, against the target database for incompatibility or logical conflict. Only data that do NOT

*Sam* is a sophisticated tool for:

- importing into an Alice database information that has been generated with other programs or databases
- reviewing the logical consistency of data sets generated with other programs
- merging two or more Alice databases

conflict are added to the target database. Conflicting data are filtered out and reported to the user who can then edit or reject them. Such checking is essential if the resulting database is to retain the high degree of internal data integrity expected of an Alice database. The greatest complexity comes in dealing with potential "conflicts" within the scientific nomenclature. *Sam* must detect and respond intelligently to many possible situations in which names, homonyms, misapplied names and synonyms may be used in different ways by different authors to refer to the same taxa.

As reported in our last newsletter, funding was obtained from the Reuters Foundation for the further development of *Sam* for use by the 'Plantas do Nordeste' project, a collaborative initiative in Northeast

Brazil involving 20 different institutes most of whom are building Alice databases. Low level routines in an earlier version of *Sam* have been modified and rewritten. They have also been used to provide a "copy data" function in the *Ace* program. Significantly, we have had the resources to take a long look at the fundamental design of the nomenclatural conflict manager which sits at the core of *Sam*. The design for this module, originally based on ideas from the literature and experiences of the ILDIS project (see separate article by Richard White), proved to have a number of shortcomings. We now have a new design and a detailed work programme which breaks down further development into a number of phases. Implementation of the next phase is due to start early in 1997.

## BEHIND THE SCENES

For those with a technical interest we describe here some of the changes to Alice that underpin the new functionality available in System 2.1 and System 3.0.

### Improved data structures

As reported in previous newsletters we are developing the data structures underlying the Alice system. The fruits of these changes are most obviously evident in Alice System 3.0 which supports a richer set of taxonomic ranks below the species level. These are of particular interest, for example, to horticulturists, agronomists and gene bank curators (please refer to the article describing the "Allium plant genetic resource". We have made a number of changes that will allow a greater variety of name, taxon and place categories to be supported within System 2.1 including some that require longer field lengths than those in System 2.0. The ground work has also been laid for supporting classes of descriptors,

dependencies between descriptors and for the integration of images and documents. Another advantage to both Systems of these structural changes is better performance, particularly of nomenclatural searches.

### New tool for upgrading your database

A key development has been *Dbupdate*, a tool for upgrading databases between different versions of the Alice System. *Dbupdate* has been successfully tested by upgrading a large number of System 2.0 databases to System 2.1 or 3.0. We are confident, having tested the program, that users without technical knowledge will be able to upgrade their databases quite easily.

### Improved Code Libraries

We have completely reorganised the Alice code-base which underlies all Alice programs. This is arranged as libraries of small program modules or functions and has been reorganised, considerably reworked and code re-use maximised. These changes have three important effects: we have removed inconsistencies between programs, made transition toward a graphical user interface easier and improved independence from the underlying database engine. We are looking to the future and listening to our users, some of whom speak of integrating their Alice databases with institutional systems for example. We are designing the Alice of the future such that it will be possible to build Alice databases in Oracle, Informix or any other relational database management system.

# New programs for the Alice system

All System 2.0 programs are available in System 2.1, which will in addition contain three completely new programs and three new database management tools.

## **Ace**

A program, which replaces *Alice*, *Atext* and *Nview*, to capture and edit your data. *Ace* represents a fantastic leap forward for those entering or editing data.

## **Aview**

A program to view and explore Alice databases without being able to make any changes to it. This is ideal for use by students or other research team members without authority to edit or enter data.

## **AliceWeb**

A program to generate HTML pages automatically from your Alice database.

## **Dbupdate**

A tool to migrate your data from a System 2.0 database to a System 2.1 or 3.0 database.

## **Dbadmin**

A tool for the nominated Alice database administrator on a networked site to manage use of and access to multiple Alice databases or applications by individual Alice users.

## **Dbunlock**

A tool for managing access to Alice databases and applications for individual Alice users on a networked site or for those Alice users working on stand-alone PCs running DOS or Windows.

## Reduced Registration for Alice System 2.1

System 2.1 is now available. Those users of Alice System 2.0 who have current maintenance contracts can now upgrade to System 2.1 for a reduced fee. This offer will close on the 1st of December 1997. All prices are given in pound sterling. For details of the programs included with each type of Registration please see the article "*Editions*" explained.

Registration type	Normal Registration Fee	Upgrade fee for current users
<b>Database author's edition</b>	£ 1,750.00	£ 650.00
<b>End-user's edition</b>	£ 550.00	£ 250.00

## New releases for Alice System 2.0

This year we have distributed new versions and maintenance releases of all programs in System 2.0. All users that have maintenance contracts should have received their copies by now. Please check that you are using the most recent versions.

Application	Function	Version
<b>Alice</b>	Data capture & editing	2.6.1
<b>Amie</b>	Database maintenance	2.4.1
<b>Aquery</b>	Search the database	2.7.1
<b>Awrite</b>	Design & write your own reports	2.8.1
<b>Alex</b>	Export data to other programs	2.10.1
<b>Atext</b>	Manage free text	1.4.1
<b>Nview</b>	Nomenclatural exploration	3.3.1

These will be the last versions of these programs to be released since all development of System 2.0 has ceased as we concentrate our efforts on Systems 2.1 and 3.0.

## Reduced prices for Alice System 2.0

Now that System 2.1 is available we have decided that we will continue to distribute System 2.0 at significantly reduced prices. All prices are given in pound sterling.

	Price with Documentation	Price without Documentation
<b>Author's edition</b> ( <i>Alice</i> , <i>Atext</i> , <i>Nview</i> , <i>Amie</i> , <i>Aquery</i> , <i>Awrite</i> and <i>Alex</i> )	£ 550	£ 200
<b>End-user's edition</b> ( <i>Nview</i> , <i>Aquery</i> and <i>Awrite</i> )	£ 200	£ 75

Those interested in purchasing copies of System 2.0 should note that:

- all development of System 2.0 has ceased as we concentrate on current systems
- *Alice* is a sophisticated tool and first time users will need documentation
- those purchasing System 2.0 after the 1st December 1996 will not be able to upgrade to System 2.1 for the reduced fee available to existing users.

# Alice team news

## HOW TO CONTACT US

Please: !!! We urge all our readers to correspond with Alice Software via e-mail and cannot guarantee a rapid response to letters or faxes.

Our email address is: [alice@dial.pipex.com](mailto:alice@dial.pipex.com).

The latest information about the Alice Systems can be found on our World-wide web site (please note change of address):

<http://dspace.dial.pipex.com/alice>

If you wish to write to us by ordinary mail then please write to Bob Allkin, clearly labelling the envelope "for attention of Alice Software".

**Dr. Bob Allkin (Alice Software)**  
Royal Botanic Gardens Kew  
Richmond, Surrey TW9 3AB, UK.  
Fax: +44 - 181 - 332 - 5740

## Alice Homepage has moved

The Alice Software Home (main) page on the World-wide Web has changed address. If you have a World-wide Web browser, then point to:

<http://dspace.dial.pipex.com/alice>

You will find general information about the Alice System, up-to-date information about developments and releases and information about Alice Software Partnership. One new feature being added to our site as the newsletter goes to press are sets of HTML pages generated from Alice databases.

## Alice software news

### EDITORIAL AND DISTRIBUTION POLICY

Alice Software News (ASN) is edited by Bob Allkin and Peter Winfield. It is distributed free of charge to users of the Alice System and others that wish to be informed about developments of the System. To reduce our costs we request that you use the enclosed form to request your copy of ASN electronically.

Anyone may send articles or letters to be considered for publication in ASN. We do reserve the right to edit and reduce contributions but any changes to text will be agreed with the author before publication.

1996 was a year for celebrations for several members of the Alice team. In April, Adriana Menezes gave birth to Felipe - bouncy, bonny and Brazilian. Mother, child and father, Naerceo, are all doing well. We wish Naerceo well in the final stage of his Ph.D. In July, there was also an addition to the Wymer's Dublin household when Eddie and Margaret had their second child, a daughter named Olwyn. Orthographic variants are apparently proving troublesome - even Eddie had to look up the preferred name in a database! Bob and Pete are saving up their holidays for a spell in Dublin getting to know Olwyn and drinking lots of Guinness. Congratulations to Daniela Zappi, who has just translated a second User Guide into Portuguese, and her husband, Nigel, who are expecting their second child. Daniela's happy news confirms a previously observed trend, and Peter and Bob were reported to be consulting specialists (Uri Geller et al) for an explanation of the remarkably invigorating effects that writing Alice manuals have on their colleagues!

We welcome Clive Beale as the most recent addition to the Alice team. Clive is a biology graduate from Liverpool with considerable experience using Alice. In a previous life Clive was a sandwich student in the Conventions and Policy section of the Royal Botanic Gardens, Kew and had then built and maintained various CITES species databases, created an orchid database importing a large data set using *Sam* and provided general computing support. Clive has upgraded the Alice installation programs for use with Systems 2.1 and 3.0, helped prepare technical documentation for Alice System 2.1 and is also to write a Tutorial for Ace.

## NEXT ISSUE

In the next issue of ASN we will bring you up to date with the progress with System 3.0 and hope to include articles about projects using the World-wide Web to publish their Alice databases and descriptions of some larger collaborative projects that use Alice. Anyone wishing to include articles in the next issue of should submit copy (on a disk or by email) to reach us *no later than the 1st of July 1997 please*.