THE PALPARES RELATIONAL DATABASE:
AN INTEGRATED MODEL FOR LACEWING RESEARCH

MANSELL, M. W. and B. KENYON*

¹ARC – Plant Protection Research Institute, Private Bag X134 Pretoria, 0001 South Africa
Department of Zoology and Entomology, University of Pretoria, Pretoria
0002 Pretoria, South Africa; E-mail: vrehmwm@plant5.agric.za
²Interactive Medical Information and Ordering Systems, 325 Rosemary Road
Menlo Park, 0081 South Africa; E-mail: healthpr@mweb.co.za

The Palpares Relational Database has been designed for the Southern African Lacewing Monitoring Programme, to address the limitations of conventional databases and filing systems. The model has also been expanded to other collections and taxa. The application was developed in Microsoft Access®, and comprises Tables, Forms, Queries and Reports that accumulate and integrate data on Specimens, Taxonomy, Localities, Literature, Institutions, Type data, and Collectors and Authors. The model provides synchronised access to comprehensive data across multiple platforms, for potential use in many applications and products. These include publications, geographical information systems analysis, faunal inventories, environmental impact assessments, compilation of catalogues, bibliographies, and data on Neuroptera as biological control agents, as well as being an invaluable resource for taxonomic research.

Key words: Relational database, specimen data, taxonomy, localities, bibliography, institutions

INTRODUCTION

Vast amounts of specimen-related data are incarcerated in biological collections throughout the world, but it is difficult to access these records and to associate specimens with other pertinent information. Collections data are fundamental to many modern applications, research programmes, decision-making processes and products, all of which accentuate the relevance of biological holdings and taxonomic research. It is consequently imperative that these data and related information be easily accessible to the scientific community for wider application.

Labels usually provide details of locality, date and collector, but seldom reflect host associations, ecology or geo-references. Neither do specimen labels provide access to literature and taxonomy or details of type specimens and depositories. Specimen data in accession registers, card files, catalogues and taxonomic papers are merely lists that do not facilitate rapid retrieval or integration across multiple information sources. These conventional systems are also limited, as they cannot accommodate large data sets or the burgeoning accrual of biological information.

The advent of computerisation and electronic storage did not resolve these constraints as mainframe computers were not generally accessible, and required expert programmers to run even the most elementary procedures. Personal computers were initially unable to provide a solution either, as they were limited by the lack of suitable software programmes. Many products were developed and marketed as database programmes, but they were incapable of integrating large sets of data from numerous databases. These constraints have now been addressed by the relational database model that is based on concepts from a branch of mathematics known as relational algebra (CODD 1970). This model is a particular way of structuring and processing a database (KROENKE 1995) to effect maximum efficiency in data storage and retrieval. Many database management systems have consequently been developed, including Microsoft Access®, that enable scientists and business people to develop applications to accumulate, collate and manipulate data in an unprecedented manner.

Microsoft Access® has been used to design and produce database applications that electronically document information in biological collections. The model described here, the PALPARES RELATIONAL DATABASE, was specifically developed to gather and integrate data on specimens of Neuropteroidea, and to combine all pertinent information relating to Afrotropical Neuroptera into one source, especially with regard to the Southern African Lacewing Monitoring Programme (MANSELL 2002). This relational database application has also been applied to other regional faunas and widely disparate taxa, and is the model used to document the holdings of the South African National Collection of Insects in Pretoria.

RATIONALE

The relational database concept is based on a two-dimensional table (KROENKE 1995). Each row holds data that pertain to some tangible entity (*e.g.* specimen, locality, reference), and is referred to as a record. No records (rows) should be duplicated and each entity requires a separate table. Columns (also referred to as fields) contain related data pertaining to attributes of the entity (*e.g.* name, collection date, locality), and each column has a unique value and name. The resultant cells formed by the rows and columns must be single-valued, and the consequent relationship between the row, column and cell ensures that each cell in a relational table contains unique information. The relational tables are structured in such a way as to reflect as much information as possible about each entity, thereby providing a record with unique values. These unique values ensure maximum efficiency of data storage, integration and retrieval, and prevent duplication.

Data are stored in the relational tables, each of which is a discrete database comprising related data. The data in a table are not repeated in other tables except for the fields that link the tables to each other to facilitate data integration and retrieval across multiple tables. Each table should have a key field that identifies the entity of that table (*e.g.* a genus and species combination defines the Species taxonomy table) and duplicates should not be permitted This facilitates efficient information retrieval and the integrity of data.

DESIGN

A relational database management system (RDBMS) provides tools for the development of applications, and the resultant model usually consists of Tables, Forms, Queries, Macros, Reports and Modules. The PALPARES RELATIONAL DATABASE application comprises the following tables: Specimen data, Species taxonomy, Genus taxonomy, Localities, Bibliography, Institutions and Persons, in

Fig. 1. The Localities form, to illustrate the use of the glide-down menu for Biomes

Fig. 2. The Specimen data form, to illustrate the various components of a data-entry form

Fig. 3. Opening page of the Palpares Relational Database, to illustrate the command buttons for opening the constituent database forms

Table 1. The fields and format of the Specimen data table of the Palpares Relational Database

Field Name	Data Type	Control Type (Field)	Description
Genus & Species	Text	Combo Box	Valid genus and species, *linked to Form: Species taxonomy*
Accession Number	Text	Text Box	Key Field - Unique Accession Number
Alternate Number	Text	Text Box	*e.g.* Field number, collection code, original accession
Determiner	Text	Combo Box	Most recent determiner, *linked to Form: Persons*
Determiner Initials	Text	Text Box	Initials of determiner, automatically updated
Year of Determination	Number	Text Box	Year in which specimen/s were determined
Institute	Text	Combo Box	Name of depository, *linked to Form: Institute*
Locality	Text	Combo Box	Name of locality, *linked to Form: Locality*
Begin Date	Date/Time	Text Box	Date on which specimen/s were collected
End Date	Date/Time	Text Box	Last date, if specimen/s were collected over a time-span
Males Pinned	Number	Text Box	Number of males pinned or micro-pinned
Females Pinned	Number	Text Box	Number of females pinned or micro-pinned
Unsexed Pinned	Number	Text Box	Number of unsexed specimens pinned or micro-pinned
Total Pinned	Number	Text Box	Total specimens pinned or micro-pinned
Larvae	Number	Text Box	Number of larvae
Cocoons	Number	Text Box	Number of Cocoons
Slides	Number	Text Box	Number of slides
Alcohol Material	Number	Text Box	Number of specimens in alcohol - not recommended !
Card Points	Number	Text Box	Number of specimens on card points - not recommended !
Capsules	Number	Text Box	Number of specimens in gelatine capsules
Photographs	Number	Text Box	Photograph/s of specimen/s- can be hyperlinked
Genitalia in capsules	Y/N	Check Box	Records whether dissected genitalia are associated
Collector1	Text	Combo Box	Collector of specimen/s, *linked to Form: Persons*

Table 1 (continued)

Field Name	Data Type	Control Type (Field)	Description
Collector1 Initials	Text	Text Box	Initials of collector, automatically updated
Collector2	Text	Combo Box	Second collector of specimen/s, *linked to Form: Persons*
Collector2 Initials	Text	Text Box	Initials of second collector, automatically updated
Collector3	Text	Text Box	Third collector of specimen/s, *linked to Form: Persons*
Collector3 Initials	Text	Text Box	Initials of third collector, automatically updated
Habitat	Text	Text Box	Habitat, *linked to Lookup Table: Habitats*
Collection Method	Text	Text Box	Method of collection, *linked to Lookup Table: Methods*
Publication	Text	Number	Publication pertinent to specimen/s, *linked to Form: Bibliography*
Date Entered	Date/Time	Text Box	Date on which entry was made, automatically updated
Recent Data Entered	Query	Command Button	Command button to run query on entries since specified date
Specimen Notes	Memo	Memo Field	Additional notes pertaining to specimens
Type Status	Y/N	Check Box	Indicates whether specimen/s have type status
Primary Type	Text	Combo Box	Indicates whether specimen/s include a primary type
Sex of PT		Toggle Button	Indicates sex of primary type
Secondary Type	Text	Combo Box	Indicates whether specimen/s include secondary types
Sex of ST		Toggle Button	Indicates sex and number of Secondary type/s
Original Name of Type	Text	Combo Box	Original name of type, *linked to Form: Species taxonomy*
Type Number	Text	Text Box	Catalogue number of type
Type Notes	Memo	Memo Field	Additional notes pertaining to type specimens

which incremental data are stored, edited and cross-referenced to other tables. As an example, Table 1 provides a list of the fields currently included in the Specimen data table, with an indication of the linkages to other tables (in italics). Specimen data is the core table that is supported by information from the other tables, although each of these is a functional database that, in turn, is linked to the other tables. There are also static tables that include only one or two fields to provide details for the glide-down menus in forms (Fig. 1).

Data are entered into tables through specially designed forms that function as an interface between data input personnel and the tables. Each form includes the fields from its underlying table, and consists of an electronic page that can comprise one or more tab controls (Fig. 2). The forms, and consequently their underlying tables, are linked to other forms by command buttons and glide-down menus, making it possible to move between forms by simply clicking a button.

Queries are structured to question the data in the tables, and are programmed to retrieve and correlate data from single or multiple tables to answer specific questions. The structure of the queries is fundamental to the design of a database, and consideration should continually be given to the nature of the questions that the database would be expected to provide. Queries can also be used to facilitate the manipulation of data, and in the compilation of the database.

Reports are drawn from the data and structured into a variety of publishable formats, for example, lists of publications. Macros and modules are used in the design process, and are essentially the programming languages of the database.

The PALPARES RELATIONAL DATABASE is designed to provide a user-friendly interface that is self-describing and requires minimal computer skills to operate. This is to ensure that large amounts of data can be entered without prerequisite knowledge of a specific taxon, or a background in database modelling and philosophy. Because of the nature of relational databases, potentially repetitive information need only be entered once. It is then accessed by data input personnel through glide-down menus and controls (fields) on forms that automatically suggest words after a few keystrokes. This facility protects the integrity (correctness) of data and prevents duplication, unnecessary keystrokes and variant spelling.

PROCEDURE

The database is opened via an icon on the screen desktop of a personal computer. This accesses the opening form (Fig. 3) that provides access to all forms through a series of command buttons arranged as a switchboard. By clicking the button displaying the name of a form, the form will be opened and data input can

commence. This stage is reached by only two clicks of a mouse button, ensuring maximum efficiency.

Input usually commences by opening the Specimen data form (Fig. 2) and entering information about a specimen (or batch of identical specimens) into the relevant controls (fields). If data (*e.g.* species names, localities) are already available in the other forms (Species taxonomy, Localities), these details are accessed and entered through the glide-down menus activated by clicking the down arrow on the control. Details of that entry can then also be viewed and edited by clicking the command button associated with that specific control. This opens the relevant form. For example: the genus and species can be entered into a blank field by typing the first few letters of the name. If the name is already in the Species taxonomy table, it will be automatically entered from that linked source. Alternatively, the down arrow on the control can be clicked to activate the glide-down list of species names, and the relevant name can be entered by highlighting it and double-clicking a mouse button. The name is then inserted in its correct form. All taxonomic details of that species can be viewed and edited by clicking the command button that opens the form.

Table 1 provides a list of fields in the Specimen data table, together with data and field types, and an indication of which fields are linked to a common field in another form. The linkage is activated through a combo box (combining data from another table) or a command button. Each of the other tables is similarly structured with a key field that links it in a one-to-many relationship with the other tables.

The forms can also include sub-forms that are placed under the tab controls to provide additional related information. The sub-forms are usually based on a query that is activated by clicking on the tab of the tab control. For example: the Localities database includes sub-forms that, when activated, provide a list of all species and specimens recorded from a specific locality. This is based on a query that filters all records of specimens in the Specimen data table from that locality.

The Bibliography table (and form) includes the Neuropterida database compiled by OSWALD (1998), and currently comprises 9143 literature records that are linked, through a unique number, to other forms.

Similarly, data in other tables and forms are derived from currently available sources. Data in the Localities table is derived from existing gazetteers such as the National Imagery and Mapping Agency (NIMA) database (www.nima.mil), 1:250 topographical maps, regional gazetteers, atlases and geographical positioning systems (GPS). Information in the Species and Genera taxonomy tables is derived from published taxonomic literature.

Additional features include command buttons on forms that run queries associated with that form. For example, a command button on the Persons form runs a

query that will list the publications by specified authors that are in the Bibliography. Likewise, it is possible to view all recent entries in a table by activating a query through a command button and entering a start date.

Each field in a table and, consequently a form, is self-describing in that the domain (description of format) of the field's attributes is specified when compiling the table. Furthermore, each field is labelled, and includes a control tip – a line of text that indicates how data should be entered. This feature is activated when the mouse pointer is used to indicate a particular control.

In compiling a database, it is important for a specialist to provide the taxonomy, localities, institutions and persons data, to ensure its accuracy and prevent errors by inexperienced data input personnel.

ADVANTAGES OF THE RELATIONAL DATABASE MODEL

The PALPARES RELATIONAL DATABASE and similar models provide numerous advantages over conventional systems, including:

– The application ensures complete integration of data and simultaneous access to multiple sets of data.
– It provides flexibility, in that the design can be modified and expanded at any time without recourse to database programmers.
– The database co-ordinator has complete control over the system at all times.
– Microsoft Access® is the state-of-the-art relational database management system, and any future developments will have to accommodate applications derived from it. Data will consequently not become inaccessible in an obsolete system.
– The development software will accommodate upgrades.
– Time consuming entry of duplicate data is eliminated. Potentially repetitive information is only entered once, and many records can be updated (*e.g.* in the event of name changes or errors) by modifying a single entry.
– Data are logically consistent as a result of the referential integrity facility.
– The model can operate on personal computers as well as in multi-user network environments.

FUNCTIONS OF THE PALPARES RELATIONAL DATABASE APPLICATION

Although this application had been specifically developed for Neuropteroidea, it can be adapted to any group of biological organisms. It is designed to accomplish, among others, the following functions:

– Accumulation, electronic storage and integration of all information pertaining to Neuropteroidea;

– Compilation of electronic archives of specimen data to provide a virtual duplicate resource as insurance against loss of specimens or destruction of the collection;

– Rapid access to specimen data via internet and other electronic pathways;

– Cross-referencing the specimen data to information on taxonomy, type specimens and depositories, literature, localities, distribution, phenology, collectors, authors and economic applications;

– Design of queries based on the above data-sets provide rapid retrieval of information (*e.g.* specimens and species per locality, depository or collector; compilation of faunal lists, bibliographies, type depositories; updating data-sets with external data by append queries; modification of data-sets with update queries).

– Incorporation of graphics affords illustrations of specimens, type specimens, localities, persons and museums;

– Instant access to information in linked objects through command buttons, glide-down menus, tab controls and sub-forms;

– Large amounts of related data are immediately accessible through these facilities.

CONCLUSIONS

Modern computer and software technology have placed powerful archiving and analytical tools in the hands of biologists. Biological data can now be accrued, co-ordinated, disseminated and applied as never before. Electronic relational database management systems have provided a vehicle for the development of increasingly significant products, that emphasise the infinite value and relevance of biological collections and taxonomic research. For the first time in over 250 years of taxonomic research biologists are able to gather and collate data from over two centuries, and from multiple sources, into single access and retrieval applications.

Instead of lying fallow in countless collections and outmoded filing systems because of their inaccessibility, biological data now have the potential to become a living resource. This will, however, require a revision of established mind-sets where, even now, such data are zealously guarded. In future, any collections whose data are not electronically accessible, and shared, will rapidly become irrelevant. Unprocessed specimen data in collections is not intellectual property, but only become such once value has been added to them through their inclusion in scientific applications, and in the generation of demand-driven products.

*

REFERENCES

CODD, E. F. (1970) A relational model of data for large shared databanks. *Communications of the ACM. June* **1970**: 377–387.

KROENKE, D. M. (1995) *Database Processing. Fundamentals, Design, and Implementation.* 5th ed. Prentice-Hall International, Inc., Engelwood Cliffs, New Jersey.

MANSELL, M. W. (2002) Monitoring lacewings (Insecta: Neuroptera) in southern Africa. *Acta zool. hung.* **48**(Suppl. 2): 165–173.

OSWALD, J. D. (1998) *Bibliography of the Neuropterida*. A working bibliography of the literature on extant and fossil Neuroptera, Megaloptera, and Raphidioptera (Insecta: Neuropterida) of the world. http//:entowww.tamu.edu/research/neuropterida/ndi-home.html