
The Use of Computer Databases During the Foreign Exploration Phase of A Biological Control Programme

W.A. Palmer

North American Field Station, Queensland Department of Lands, 2801 Arrowhead Circle,
Temple, TX 76502 USA

A considerable amount of data can be generated during the foreign exploration phase of a biological control programme. These data include collection records of faunistic surveys, opinions of entomologists familiar with the identified insects, literature searches, formal biology and host-specificity studies and information gleaned from museum specimens. During the past 10 yrs personal computers have become widely available. Laptop computers can now even be carried in the field to record the primary entry of data. With the concurrent improvement in suitable software it is now possible, and desirable, to record data once only. Since 1982, the North American Field Station has undertaken faunistic surveys and associated activities of a number of weed species such as *Baccharis*, *Parthenium*, *Xanthium*, *Ambrosia* and *Lantana*, and 3 databases have been constructed to manage the resultant data. Initially, a flat file database was constructed for the *Baccharis* data. Later, improved computers and software allowed the use of the relational databases, *Reflex Plus*® and *4th Dimension*®. The latter programme is an exceptionally sophisticated programme able to meet all requirements including making insect labels and exporting data directly into word processing software. The capability of *4th Dimension*® is demonstrated with reference to the database gathered for *Lantana camara*. In this case, 4 surveys conducted independently over 90 yrs have lead to the recording of over 600 phytophagous insect species. Various analyses performed on the data are also demonstrated.

Introduction

The North American Field Station is 1 of 3 exploratory units of the Queensland Department of Lands. It was set up in 1982 in Temple, Texas, to find suitable potential agents for a number of weeds of North American origin. Because of the sub-tropical to tropical nature of the Queensland climate an unusually large proportion of Queensland's weeds come from the North American region. In the past 10 yrs some 8 weed species have been investigated by staff of this station. These species are *Baccharis halimifolia* L. (Asteraceae), *Ambrosia artemisiifolia* L. (Asteraceae), *Parthenium hysterophorus* L. (Asteraceae), *Xanthium strumarium* L. (*sensu lato*) (Asteraceae), *Parkinsonia aculeata* L. (Caesalpinaceae), *Prosopis glandulosa* Torr. (Mimosaceae), *Senna obtusifolia* L. (Caesalpinaceae) and *Lantana*

camara L. (Verbenaceae).

The operation of this field station perhaps differed from other exploratory units in that almost all of the target weeds had already been investigated by other workers. The fact that the faunas on these plants were not unknown before 1982 is both a function of the importance of these weeds to Queensland and elsewhere, and also the high degree of entomological knowledge available on this continent. The present project to find further agents for *L. camara* provides a typical example. A survey of the insects associated with *Lantana* in Mexico was first conducted in 1902 by Albert Koebele (Perkins and Swezey 1924). Broader surveys in Mexico were conducted by John Mann of the Queensland Department of Lands and Noel Krauss of the Hawaiian Department of Agriculture (Mann 1954). Thus the present

effort which began in 1988 is actually the fourth major survey in the neotropics of North America.

The question addressed by this paper is how to handle efficiently the quite considerable data generated by a field station such as this. Three basic parts to this question exist. Firstly, how to record the data associated with each field collection of insects? Secondly, how to incorporate any other knowledge that exists about an identified species? Thirdly, how to report and analyse the results of the project? In the case of the North American Field Station, efforts related to these 8 weeds have now resulted in nearly 4,000 collection records and information being accumulated on some 1,500 insect species. In addition to reporting on these projects we hope to use the data to provide useful information contributing to the theory behind this phase of a biological control programme. For example, we are ideally situated to throw some light on the question of whether second or third surveys are likely to be profitable and on the question of the reproducibility of survey results.

Computer databases appear to be the best way of handling the large quantities of data generated. In the past 10 yrs there has been an almost incomprehensible advancement in computer power available to the research worker. At the Brisbane *Biological Control of Weeds Symposium*, Spencer *et al.* (1981) proposed a data recording system for collection records. Today the computing power is many times greater and one can now even record data directly onto a computer in the field. Parallel to this improvement in computer hardware has been a commensurate development in computer software so that today very sophisticated database programs are available on personal computers.

Development of Database

Our efforts to develop a database have also paralleled these developments in hardware and software. In 1983, a database which was very similar to that of Spencer *et al.* (1981) was developed on a 48 K *Tandy Model 3* computer. This database consisted of only one file and was what is known a flat file database. Information common to each record had to be reentered

every time for each record. It was extremely limited in the number of records that it held and in its searching and sorting capabilities.

Some years later we switched to an *Apple Macintosh* computer and a software programme called *Reflex Plus*. This programme was a "low end" relational database. It was selected because it was relatively inexpensive, because it had only limited programming capability and thus did not need professional expertise to design the database, and because it was thought that the data did not require the more sophisticated functions of the "top of the line" databases. A third switch was necessitated when the company producing *Reflex Plus* decided to discontinue its line of *Macintosh* products. Without the necessary upgrades the programme would not run on later versions of the *Macintosh* disk operating system. This is, of course, one of the hazards of any selection of software; that it will no longer be supported by its manufacturer.

The third database selected is probably the most sophisticated programme available to the *Macintosh* and may be the equal of those programmes available to the *MS-DOS* machines; *4th Dimension* manufactured by *Acius*. This is a fully relational, fully programmable software package. It is well supported by documentation so that one can design one's own database without undue difficulty. However, without special training in its use, one is unlikely to be able to exploit fully the capacity of this programme. We have therefore enlisted the services of a professional developer in customizing the database for our needs and we would highly recommend others do the same. The programme is now meeting our every expectation. Indeed, we are still finding new uses for its capability. In retrospect it would have been a better decision to select this package in the first place.

Relational Databases

A modern database consists of any number of files each of which consists of records. Each record consists of a number of pieces of information called fields. Thus, for example, the date on which a collection was made becomes a

field within a collecting record which is stored in a file of all the collection records.

A sophisticated database program can draw information from any of the individual files within it. Perhaps the most significant advancement in database design for the personal computer over the past decade has been the development of relational databases.

When files within a database are related, data stored in one file can be accessed from another file. For example, if one were to collect 10 insect species at a particular site, information common to all 10 insects (e.g., location, host, weather, etc.) could be stored *once* in one file and then "related" to the data about each of the 10 insect species which would be stored in another file. Similarly, scientific papers and general entomological texts need be recorded once and then related to any number of pertinent insect species. Related databases therefore allow for the more efficient storage of information, the ability to update data in one place and have it reflected everywhere that the data is used and the ability to search and sort one file based on data in another.

A second major criterion for an efficient database is that it be internally programmable. By that is meant that the software contains its own programming language (not dissimilar to conventional languages such as *PASCAL* and *FORTTRAN*) which can be used to perform very sophisticated tasks with relative simplicity.

Database Uses

The primary function of a database is simply to store information in an efficient manner and to have it available for future recall. In this database a programme was written that allows an information profile to be drawn up about each insect species. This information includes the data on its collection by the NAFS (date, host, stage, location, etc.), other known hosts, references, pest status, host-specificity status, previous collectors, etc.

A second very useful function is that lists of insects can be very easily drawn up. These can be "exported" to a word processor application and with very little further formatting can be put in final form for publication. Thus a table of a faunal list of 100-200 insect species for a

publication such as that of Palmer (1987), might take a half hour to prepare. A similar list prepared by typewriter might well take a day or more to prepare and would then require considerable proof reading to eliminate inevitable typographic errors.

A third use to which our database is put is to prepare insect labels for pinned specimens. With just a few key strokes, labels in 5-point type can be prepared on a laser printer.

Lastly, the database is ideal to perform various analyses of the data. Examples of such analyses might be comparisons of the efficiency of different collectors, comparisons of the number of species on different plants or regions, or a look at the phenology of an insect of significance. Some of these analyses can even be performed in a graphical form. The database gathered on the insect fauna of *Lantana* in North America is used to give some idea how the present database is used.

The *Lantana* Database

Lantana has received considerable attention over the years and there have now been 4 major surveys of the insect fauna found on *Lantana* spp. in Mexico and surrounding countries.

One of the first tasks at the beginning of this project was to "clean up" the existing data, which in this case meant updating many of the names to the currently accepted name. Quite often the same insect was reported under more than one name. Old names were assigned to a subfile of the database so that they could still be used to select records if necessary.

Altogether it now appears that over 600 phytophagous species have been collected by at least one of the surveys. Once these data have been entered on the computer it is then possible to conduct an unlimited variety of analyses on them by searching for records that meet various criteria. A simple search involves selecting records that satisfy the "search condition" for one field. For example, simple searches revealed that Koebele, Mann, Krauss and Palmer collected 214, 120, 227 and 202 phytophagous species, respectively. Compound searches in which the record is only selected if it meets the criteria for 2 or more

fields (or 2 or more conditions of 1 field) can also be conducted. Compound searches showed the additive effect of additional collectors. For example, Koebele found 214 phytophagous species of which 9 were stenophagous (i.e., potential biological control agents). Koebele and Mann between them found 301 phytophages including 13 stenophagous species. The addition of Krauss to the search criteria selected 491 insect records of which 17 were stenophagous. Finally the present effort (Palmer) raised the number to 609 phytophages and 25 stenophagous insects. Thus the database can be used to assess the contributions of additional surveys. One of the most interesting analyses indicated that of these 609 species only 8 were collected in each of the 4 surveys.

Similarly the database can be used to compile statistics about host associations. In the *Lantana* survey 4 closely related species have been searched in the present project. They are *L. camara* L., *L. urticifolia* Mill., *L. hispida* HBK. and *L. urticoides*. Simple searches show that 97, 59, 42 and 28 species, respectively, were collected on each of these 4 plant species. Of the 97 species collected from *L. camara*, 30% have been found on at least 1 other *Lantana* sp. but only 7 of the 25 stenophagous species have been found exclusively on *L. camara*.

Discussion

This paper has attempted to show some of the uses to which a properly designed database might be put. We are getting to the stage where data obtained about the insect, be it about its collection, identification, literature search, etc., are entered into our records once only. We are also attempting to maximize the use to which the data is put be it drawing up faunal lists, making insect labels, preparing graphs or other analyses.

Furthermore it is hoped that the data may be in such a form so as to be of use to other researchers in the future. Theoretical studies such as those of Strong *et al.* (1984) rely on the analyses of sound datasets. A computerized database is an efficient way of making "raw" data available to one or more researchers so

that they can then conduct their own analyses. Faunistic surveys undertaken in foreign countries are now done at considerable cost so that it is important that any resulting information be utilized to its maximum extent.

The database now in use at the North American Field Station might be of use, either in its present form or with minor modification, to others performing similar tasks and indeed could also be utilized by other insect collectors whether amateur or professional. It would be an easy task to install the present design on any suitable computer system and it, and accompanying data sheets, are available from the author to all interested colleagues.

Acknowledgments

The contribution of Mr. Rich Gay, of Foresight Technology, Fort Worth in designing and programming the *4th Dimension* database is acknowledged with appreciation. Mr. Kim Pullen contributed to the design of the collection and rearing record sheets.

References

- Mann, J. 1954. *Lantana* investigations in Mexico from July 1953 to January 1954 and a brief summary of the investigations by N. L. H. Krauss in Florida, Mexico and Central America from April 1953 to January 1954. Queensland Department of Lands.
- Palmer, W.A. 1987. The phytophagous insect fauna associated with *Baccharis halimifolia* L. and *B. neglecta* Britton in Texas, Louisiana and northern Mexico. *Proceedings of the Entomological Society of Washington* **89**:185-99.
- Perkins, R.C. L. and O.H. Swezey. 1924. The introduction into Hawaii of insects that attack *Lantana*. *Bulletin of the Experimental Station of the Hawaiian Sugar Planter's Association* **16**:1-83.
- Spencer, N.R., S.G. Rosenthal and N. Hostettler. 1981. A computer-assisted method for the storage, retrieval and analysis of biological field data. *Proceedings of the V International Symposium on Biological Control of Weeds*, Brisbane, Australia. Delfosse, E.S. (ed.). CSIRO, Melbourne, pp. 627-33.
- Strong, D.R., J.H. Lawton and R. Southwood. 1984. *Insects on Plants*. Harvard University Press, Cambridge, Mass., 313 p.