

X. LBASE
THE DATABASE SYSTEM OF THE
RJJKSHERBARIUM/HORTUS BOTANICUS

P. C. VAN WELZEN*

Rijksherbarium/Hortus Botanicus
 P.O. Box 9514, 2300 RA Leiden
 The Netherlands

R. P. VAN VALEN

Waaldijk 175
 2988 AW Ridderkerk
 The Netherlands

J. A. VALKENBURG

Tooroplaan 10
 3141 KP Maassluis
 The Netherlands

SUMMARY

The database system which is presently implemented in the Rijksherbarium/Hortus Botanicus, called LBASE, will also serve as the future databank for the Flora Malesiana project.

The system will contain information about specimens, persons, literature, taxa, and nomenclature. LBASE can serve many functions such as generating literature/nomenclature headings above descriptions, identification lists, listings of vernacular names, distribution maps, etc; but also administrative functions for library and collection maintenance.

Data capture will be from existing databases, current revisions, and new/loan material. Data from earlier revisions will only be added if additional funding becomes available.

In the future LBASE may grow into some kind of expert system with possibilities for automatic descriptions, key-construction, and (on-line) facilities for identification.

One of the major aims of the database system is to speed up the Flora Malesiana project by providing easily accessible information and by combining data and offering those ready for printing. Participants of the Flora Malesiana project will have free access to the data.

INTRODUCTION

The urgent need to computerize most of the taxonomic work (including curatorial and library tasks) in the herbarium of Leiden (L) became apparent for four reasons:

1. More efficient research and maintainance.
2. The necessity to speed up Flora Malesiana.
3. Development of international projects such as TROPICOS (Missouri Botanical Garden), IOPI (World checklist of plants), ILDIS (Legume database), in which participation is necessary/desirable.
4. Unification of already existing databases in L.

Presently most staff members have a stand-alone PC, which means that information cannot freely be accessed, exchanged, or combined. Also the danger exists of non-compatible data due to differences in data-capture formats. The conclusion is obvious, a local area network (LAN) is needed for exchange of information and a database system which combines scientific data with the curatorial and library information and which guarantees access to the data.

At first several existing database systems were reviewed for their suitability, but for very diverse reasons, varying from incompatible soft- and hardware surroundings to only partial fulfillment of the requirements, none of them was adopted. However, three systems pro-

*) To whom correspondence should be addressed.

vided an impressive amount of useful information: ALICE (Winfield & Allkin; ILDIS uses this system; Winfield et al., 1987), TROPICOS (Crosby & Magill, 1989), and especially the system used by the herbarium of Geneva (Zellweger et al., 1990).

As expert knowledge in database design was virtually absent in L two programmer/designers of the Polytechnic school of The Hague, Mr. Van Valen and Mr. Valkenburg, were temporarily employed to build a blueprint of LBASE. The results presented here are theirs. They adopted the design method called SDM (System Development Methodology, Turner et al., 1989). This method encompasses 7 phases: information planning, definition study, basis design, detailed design, realization, introduction, use, and maintenance. In addition to the SDM method different scheme techniques were used to picture the information.

SDM, although too elaborate for LBASE, ensures a sound database design. Researchers usually tend to start too detailed, for this reason attention will be given to the second, third and fourth phase of the SDM development of LBASE.

DEFINITION STUDY

In the definition study an analysis is made of the data flow in the institute and of its users. This may seem to be a very obvious and superfluous activity, but in this way no items are left out (Van Valen & Valkenburg, 1991a). Four major user groups were found: Scientists, curatorial staff, library, and botanical garden. The following major items of information could be discerned (more or less as in Zellweger et al., 1990, but elaborated with curatorial and library tasks):

Persons: Authors, collectors, identifiers, correspondents, subscribers, employees, lenders, borrowers, providers and checkers of information.

Literature: Books and articles in journals. Keywords referring to topics and to taxa are very important.

Specimens: Collector and number as unique combination for the specimen. Important information: identification, geography, ecology, uses, vernacular names, collector's notes.

Geography: Not only the exact positions of the collecting sites, but also indications of the distributions of the taxa.

Nomenclature: Authors, accepted names, synonyms, types, comments and references to publications.

Taxonomy: Classification of the taxa; may also include descriptions and identification keys.

Subscriptions: Administrative system of subscriptions on journals published by L.

Literature administration: Two tasks, bookkeeping of loans of books and journal issues, and bookkeeping of received journal issues.

Specimen administration: Bookkeeping of loans from and to L.

Seed collection: Publication of a seedlist and exchange of seeds.

The file 'Persons' is the centre of the whole database as persons provide, produce, type, and check all information and demand all activities.

BASIS DESIGN

The information/data flow has been formalized into a blueprint (Van Valen & Valkenburg, 1991b). Not all details are available as it is still a basis design. Figure 1 summarizes the groundplan. Of course there should be more connective lines going from person to the other rectangles as it are persons for instance who subscribe or order seed. Figure 1 roughly

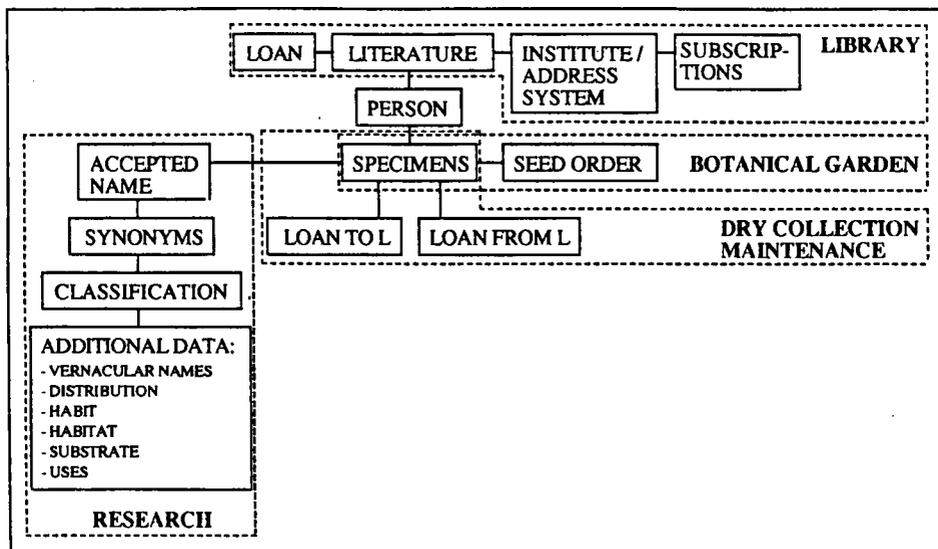


Figure 1. Summary of LBASE groundplan.

indicates which section may be responsible for the data, but this may change in the future. Access to the data is mainly unlimited, but changing, editing or adding data often will be restricted to authorized persons only.

Figure 2 provides details of the very complex system which is designed; only the part of figure 1, research and dry collection maintenance is presented in figure 2. The basal part on the left of figure 2 contains loans from and to L. In the centre 'Specimen' can be found glued to 'Specimen sheet'. Above 'Specimen' '(Infra)species' with on top of it its nomenclature and higher classification. Left of 'Species' indications of habit, habitat, substrate, and use are shown; to the right of 'Species' information about hybrids, vernacular names, and geography can be found. Finally, everything marked with an asterisk can be referred to the origin of the information: persons, literature, specimens, or combinations; this is shown in the right basal corner.

In the system most of the rectangles and some ovals will be datafiles (in dBase language databases). The ovals are data connecting files, links between data. This system ensures that every item of information is only stored once and reduces typing errors and saves storage space on discs. If one looks to 'Specimen' then specimens are mounted on sheets (go downwards), 'Specimen sheet', every specimen can be mounted on 1 to many (n) sheets, while the other way round (going upwards), every sheet will only contain one and only one specimen; this explains the use of 0, 1, and n around the linking files (ovals). (N.B.: if more than one specimen is present on a sheet then these are referred to under a different combination of collector and number, also if it concerns mixtures.) Likewise, every 'Specimen' has one and only one 'Collecting locality', while every 'Collecting locality' can have one to many 'Specimens' (localities where no collections are made are not entered into the database).

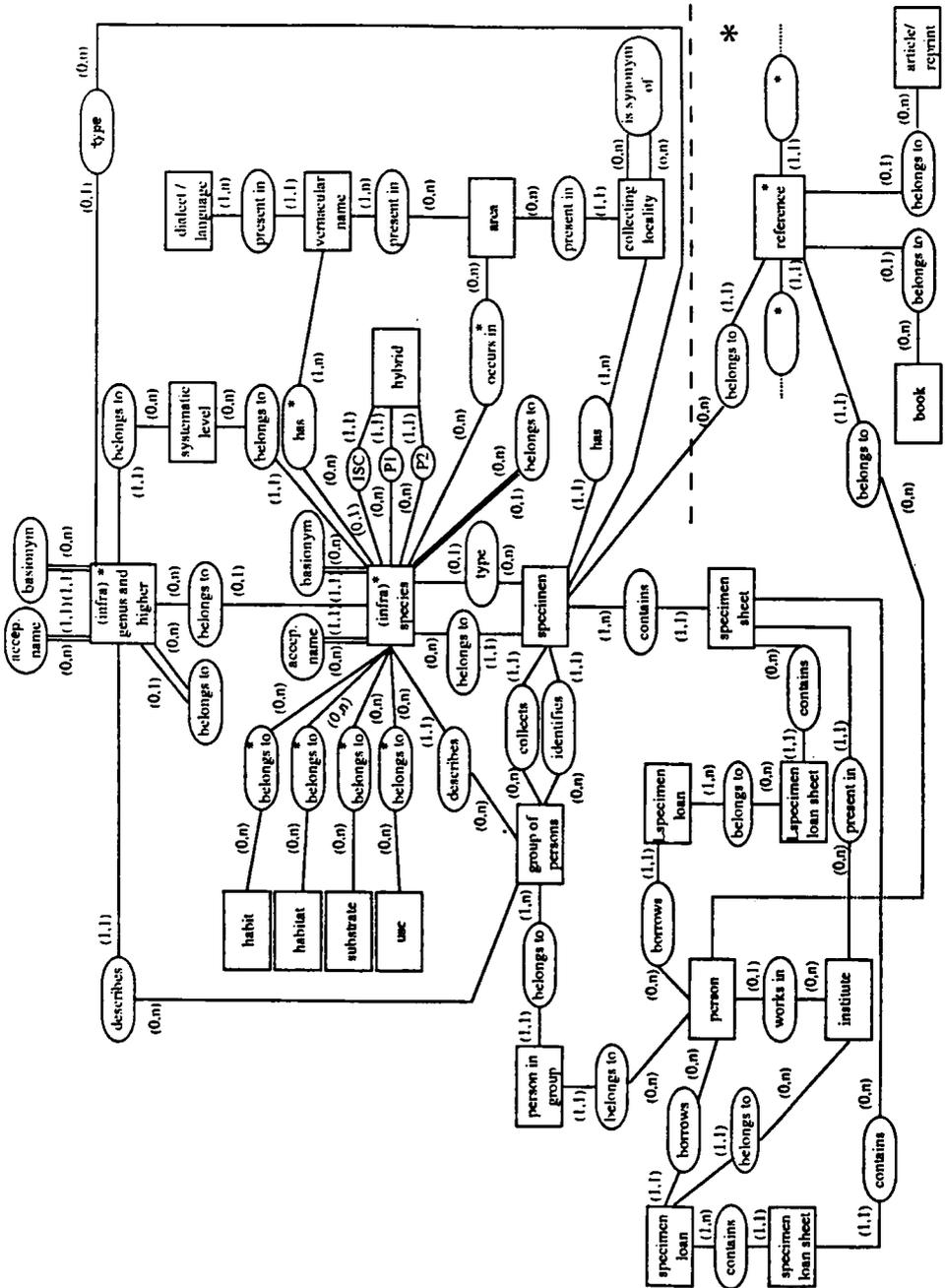


Figure 2. Groundplan of LBASE for research and dry collection maintenance.

Some parts of interest to taxonomists are *habit*, *habitat*, *substrate*, and *uses*. Besides a precise description in a free text field, they include several keywords which act as indicators. The keywords allow sorting and listings of taxa per keyword, like all the species growing on ultrabasic soil (which can be combined with a certain area like New Caledonia for instance). Each taxon can have several keywords. The keywords for the uses are mainly the commodity groups of the PROSEA project (Jansen et al., 1991). This will for instance allow for listings of all taxa with medical use, whereby the free text field provides the exact use. It can also provide a heading 'Uses' under the species descriptions for Flora Malesiana, with all uses listed, not only the medical ones.

Headings below species descriptions which can be generated automatically will be: *collector's notes* (only containing the habit), *distribution* (see below), *ecology* (with data on substrate, habitat and altitude only), *uses*, and *vernacular names*.

The *geography* is noted in two ways, by two- or three-level keyword areas (e. g. Indonesia, Java, Jawa Barat or Australia, New Queensland; in fact one level is added for collection maintenance which more or less corresponds to the different continents) and by the precise coordinates of the collecting locality. The areas will be on species level ('Area' in figure 2), the collecting localities on specimen level ('Collecting localities' in figure 2). As a result dotted distribution maps can be obtained or checklists of taxa per area (country, island, or province).

Vernacular names together with their dialect and language can be given, just as the area where the name is used. This will not only reveal the names used per taxon, but also whether or not a local name is used for more than one taxon. Also listings with all local names per region or per language or dialect can be printed.

The most difficult part to implement is the *classification* and the *nomenclature*. The present basis scheme allows only for one classification to be opted, but reference can be made to alternative classifications. The system can handle innumerable synonyms and taxonomic levels. As a result the literature headings above descriptions can be automatically produced, including accepted names, basionyms, synonyms, literature, types, comments (like nomen nudum, etc.), all arranged chronologically, starting with the accepted name, with the names grouped per type specimen.

Literature information can always be linked to taxon names and lists can be obtained of literature references per taxon name (if wished for, including those of the synonyms). Literature can be marked for a chapter 'References' in a publication, like at the end of this article, or one can give as many topic-related keywords as wished for. Lists of literature per keyword can be printed.

DETAILED DESIGN

This phase (Van Valen & Valkenburg, 1991c) is not yet completely finished and will end in a final version of the blueprint, including a definition of the data fields, the functions the system has to perform, menu structures, data integrity checks, and authorizations for the users.

The fields are defined, just as a detailed list of all the functions LBASE will perform in standard configuration (this already includes the fields which will have to be linked). Figure 3 shows these functions. A start is made with the definition of the menus and of the different data integrity checks.

Person data a, f, c, d	Book data a, f, c, d
Institute data a, f, c, d	List of L books
Section of L data a, f, c, d	Journal data a, f, c, d
List of L employees	List of L journals
Telephone list of L employees	Journal issue data a, f, c, d
Persons of whom bibliographic data in L	Letter journal issue not received
Address on label or letter heading	Article data a, f, c, d
Kew index down-loading	Reprint data a, f, c, d
Specimen loan from L a, f, c, d	List of L reprints
List of L specimens on loan	Keywords a, f, c, d
Specimen loan to L a, f, c, d	List of keywords
(Pre)identified specimen data a, f, c, d	List of articles per keyword
Specimen label data a, f, c, d	Marking literature for reference list
Specimen label printing	Reference list
Identification list per collector	Literature loan data a, f, c, d
Identification list per taxon	Reminder letter for loan
Literature list per taxon	Yearly acquisition list
Distribution maps	Order data L publications a, f, c, d
Check-lists per area	Subscriptions L journals a, f, c, d
Nomenclature/literature heading	L journal order data a, f, c, d
Ecological data heading	Bills and postal giro
Vernacular names heading	Overview subscribers
Uses heading	Payment reminder
Habitus heading	Garden specimen data a, f, c, d
List of plantfamilies and L number	List of Garden specimens
Synonymy data (incl. types) a, f, c, d	Seed data a, f, c, d
Taxonomy data a, f, c, d	Marking and numbering taxa for seedlist
Identification labels per taxon	Printing and deleting of seedlist
	Seed order data a, f, c, d
	Overview of seed orders
	Plant label engraving

a = adding, f = finding
c = changing, d = deleting

Figure 3. Standard functions of LBASE.

FUTURE AND INTERNATIONAL PROJECTS

The detailed design will be finished and then the final phases will start with the implementation of the system and the checks necessary to test the accuracy and performance of the system. After this phase, hopefully at the end of 1992, the data can be entered.

Cooperation in other internationally important projects besides Flora Malesiana will be ensured by entering the data as atomic as possible. This allows for the arrangement of the data in different formats as necessary for each project. A simple example: The name of one of the authors, P. C. van Welzen, will be placed in 3 fields, one for the initials (P. C.), one for the middle name (van), and one for the surname (Welzen); the name can now be cited in different arrangements: Welzen, P. C. van, or less correctly, Van Welzen, P. C.

Cooperation with the following projects is envisaged:

- IOPI (World checklist of vascular plant species)
- TROPICOS (Chinese flora and other fields of interest)
- ALICE (SE Asian Legumes)
- GENEVA (fields of interest)
- ETI (Expert Center for Taxonomic Identification; SE Asian plants)
- PROSEA (Uses of SE Asian plants)

Involvement in the ETI project will mean an elaboration of the tasks of LBASE with possibilities for automatically written descriptions and keys (e.g. DELTA package) and scanning facilities for the capture of images.

HARD- AND SOFTWARE

A few technical details are provided in this chapter. The system will run with a VAX workstation as file-server, VMS as operating system, and PCSA as network software on a network of mainly thin ethernet. MS-DOS PC's, preferably 386's, will function as terminals with the opportunity to work stand-alone.

LBASE will use RDB of DEC as Data Base Management System (DBMS) and UNIFACE as fourth generation developing language. UNIFACE uses the 3-scheme architecture, where-by the above mentioned blueprint of the database system will serve as the middle, conceptual scheme. The screen menus and the print-outs will be the external scheme, and the driver for RDB the internal scheme. UNIFACE offers many benefits, to mention a few: Ease of development, user friendliness with pull-down and pop-up menus, independence of hardware/DBMS (another driver solves a system switch), possibility to address at the same time different systems using different operating systems, compression of data and quick transport via bridge-links to other buildings, opportunity to make individual one-time queries. The latter allows for instance for very complicated (and therefore not often used and not standard present) queries; an example: A list of all Dutch plant names used in Belgium linked to their scientific names with all synonyms plus all references to the vernacular names.

DATA CAPTURE

Data will come from several sources, but there is no desire to computerize the whole collection or to enter the data published in the Flora Malesiana volumes. Moreover, Flora Malesiana is not the only project of L, other projects also have to be included in LBASE, like the Dutch flora, and not only data on Angiosperms and Pteridophytes will be entered, but also information on Algae, Fungi, and Bryophytes.

Data will come from the following sources:

- Already existing databases in L; e.g. specimen oriented databases on Magnoliaceae, Symplocaceae, Sapindaceae, and Orchidaceae, literature data (books and journals in L, and those articles with a plant name in the title), spirit and carpological collection, living plants in the botanical garden.
- Commercially available databases as for instance the Kew Index.
- Entries by researchers revising a group, data from new material or material sent on loan. Recorded specimens will be marked with a bar code.
- If additional funding becomes available, parts of the collections and Flora Malesiana data can be digitized. Priority will be given to the type specimens and the specimens of historical interest.

These plans mean that during the start of the system's use as many people as possible should enter data. The more data present the more time saving LBASE becomes.

Researchers from outside L, participating in Flora Malesiana, can require listings or can perhaps retrieve information on-line as a connection with international networks will be obtained. While visiting L they can of course have free access to the database.

ACKNOWLEDGEMENTS

We are grateful to the network committee of the Rijksherbarium/Hortus Botanicus for their advice and guidance. We also would like to thank Catherine Zellweger, her colleagues and Prof. Spichiger for their hospitality and invaluable advice during the visit of Mr. A.J. de Lange and the first author. We also would like to thank Bob Alkin and Bill Loader (Kew Herbarium) and the staff of the Missouri Botanical Garden for discussions and demonstrations.

REFERENCES

- Crosby, M.R., & R.E. Magill. 1989. TROPICOS, a botanical database system at the Missouri Botanical Garden. 'The Booklet'. Missouri Botanical Garden.
- Jansen, P.C.M., R.H.M.J. Lemmens & L.P.A. Oyen (eds.). 1991. PROSEA basic list of species and commodity grouping. Final version. Pudoc, Wageningen.
- Turner, W.S., R.P. Langerhorst, G.F. Hice, H.B. Eilers & A.A. Uytenbroek. 1989. SDM (System Development Methodology), Dutch version. Pandata, Rijswijk.
- Valen, R.P. van, & J.A. Valkenburg. 1991a. Definitiestudie automatisering Rijksherbarium/Hortus Botanicus. In Dutch, internal report in L.
- Valen, R.P. van, & J.A. Valkenburg. 1991b. Basisontwerp automatisering Rijksherbarium/Hortus Botanicus. In Dutch, internal report in L.
- Valen, R.P. van, & J.A. Valkenburg. 1991c. Functioneel deel detailontwerp automatisering Rijksherbarium/Hortus Botanicus. In Dutch, internal report in L.
- Winfield, P.J., R. Alkin & F.A. Bisby. 1987. ALICE: Biological checklist and species diversity database system. Agricultural Scientific Services, Edinburgh.
- Zellweger, C., R. Palese, P. Perret, L. Ramella & R. Spichiger. 1990. Concept and use of an integrated database system for the Chaco. Application to a preliminary checklist. Contribution to the study of the flora and vegetation of the Chaco. II. *Candollea* 45: 681–690.