

## DATABASE OF A REFERENCE COLLECTION - THE ADVANTAGE OF HAVING SPECIFIC NAMES AS MAIN ENTRY

By Hakon Regnar Jalk<sup>1</sup> & Anuwat Nateewathana<sup>2</sup>

1. Department of Ecology & Genetics, Institute of Biological Sciences, Building 540, Aarhus University, 8000 Aarhus C., Denmark.

2. Phuket Marine Biological Center, P.O. Box 60, Phuket 83000, Thailand

### ABSTRACT

We suggest that a database system should be designed at an early stage of the establishment of a reference collection. Since the genus and family levels are frequently changed, we recommend that the specific part of the binomen constitutes a keyword in the database. This part is very unlikely to be changed as a result of taxonomic revisions.

#### Function of a reference collection

A reference collection can be considered as a "library of plant and/or animal species" which serves as a reliable tool for the working scientist to obtain and secure correct identification of the species (UNESCO 1979). Thus, collections are just as necessary as books in a library, which are not in continuous use, yet must be available when needed (Mayr & Ashlock, 1991). Like in a library, it is necessary to have a system so data can be retrieved, fast and reliably.

The aim of this paper is to discuss the use and structure of a database system, in connection with a reference collection, in order to ease the work for the scientist who wants to get information about the collection.

#### Data of the base

A database is defined as a collection of related data (Elmasri & Navathe 1994). Today databases have become the standard technique in managing large amounts of data and information (McFadden & Hoffer 1994), and the development of personal computers at a relatively low cost has facilitated this development. At Phuket Marine Biological Center we plan to make a database of the Reference Collection and we have discussed how to make a base with a structure, which is easy to use and maintain.

Obviously, the registration of specimens in the collection must be standardised from the beginning, starting with notes on the locality, the date, the collector, and identification of the specimen. This primary information must always be kept together with the specimen. At the Phuket Marine Biological Center, a registration

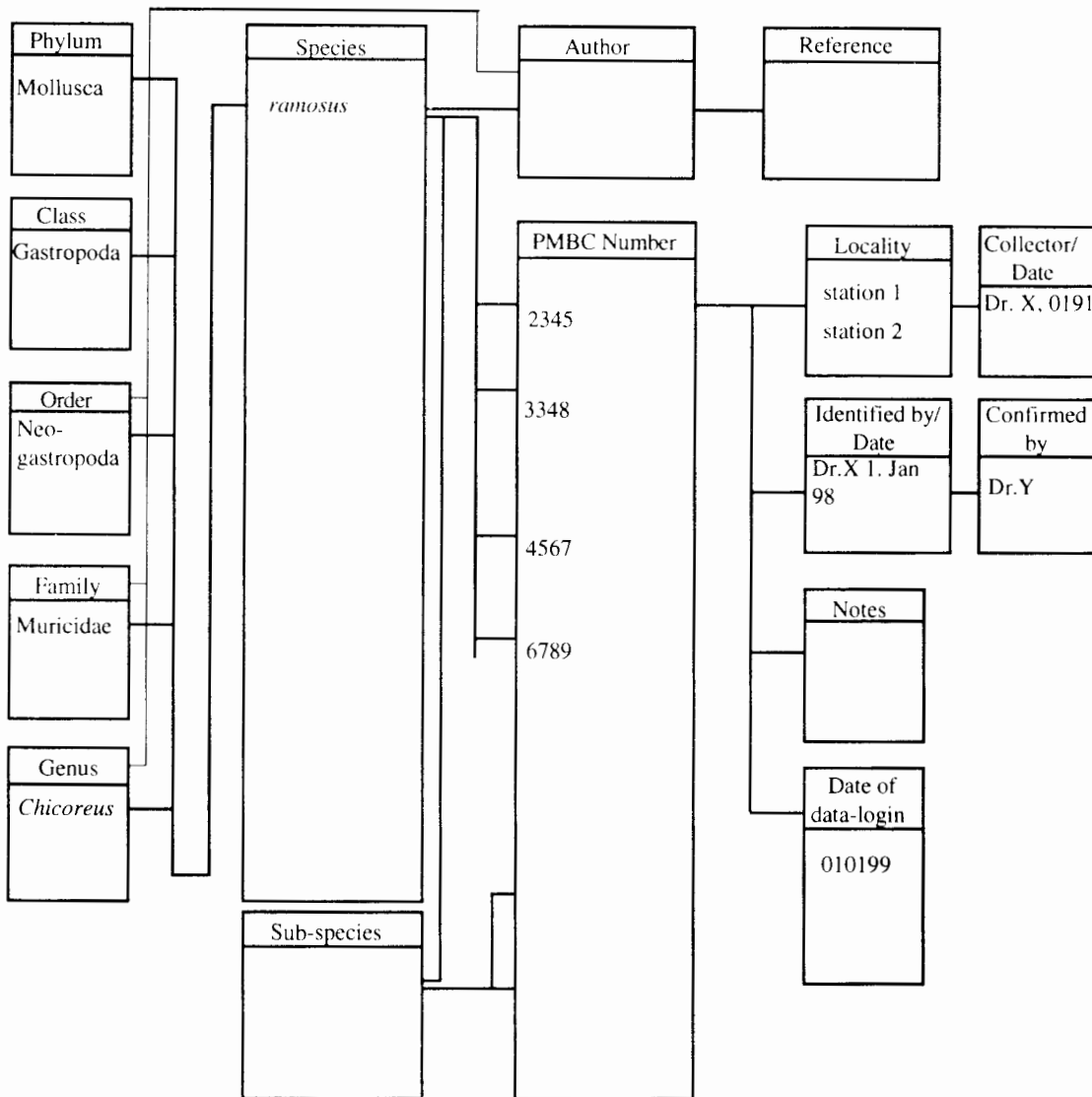
card is presently used to file this information. In addition, each lot of specimens is given a registration number.

#### Hardware and Software

Ideally, the hardware should be powerful, the processor should be fast enough to handle large amounts of data, and an extra harddisk should be used as a back up device. The machine used for the database should be exclusively for this purpose. Also it should be possible to connect the database to international data networks, e.g., Internet. This will allow it to communicate with databases of other Reference Collections. The type of software chosen for the database depends upon the size and amount of data the base should contain. For small size databases it can be of advantage to use the pre-designed programmes like dBase, FoxPro, Excel, but for big databases it is necessary to programme the database by using programmes like Turbo Pascal or Fourth Dimension (the latter has been chosen by Phuket Marine Biological Center). The technical support from the company producing the software is important if problems should arise.

#### Structure of the database

The scope of this paper is to discuss the data we want the database to contain. First of all it is necessary to define whether the family, genus or the species name should be the essential keyword/main entry from which the other data can be obtained. Figure 1 illustrates how we imagine it should be done at Phuket Marine Biological Center.



**Figure 1.** The proposed structure of the relation database at the Reference Collection at Phuket Marine Biological Center.

In working out this structure our concern has been to make the main entry as stable as possible. Looking at the “boxes” (Fig. 1) representing the different records of the Database, only a few of them will be subject of changes. The “boxes” can be divided into 3 categories: 1. constant ( PMBC No., locality, collector/date,

identified by/date, confirmed by, date of data-logging, reference, authorship and year, and notes). 2. rarely changed (phylum, class, order, species). 3. often changed (family, genus, subgenus, subspecies).

Without doubt the PMBC number will always be the most stable data type, irrespective of taxonomic

changes. The idea is to let all the information concerning the actual species be cross referred from this number. The number though, is not good as a main entry to the base if one wishes to search for information about the taxa.

The common approach of catalogues has been to use the family or genus level as entries, but continued revisions by specialists make these categories unstable (Hylleberg 1993). Relatively often the database must be updated and rearranged. The idea of letting the specific names, ordered alphabetically, be the main entry/ approach to catalogues is more actual than ever (Kabat & Boss 1992) and must be the way to do it in the future (Hylleberg 1994). Subspecies do not make a stable entry to the database since the subspecies is an expression of a "geographic race" (Mayr & Ashlock 1991) which are frequently raised to the specific level or synonymized. The subspecies will be connected closely to the species, since the trinomial subspecies name always will be given from the binomial species name. Thus from the point of view that the base should be as stable as possible the species name, followed by the catalogue number given by the reference collection, must be the most stable set-up as main entry.

#### Catalogue

The database must be updated regularly and the contents published as a book, *e.g.*, every five years. The

design of the database allows the listed animals to be printed with the species names in alphabetic order followed by the genus name. The "box" in the database called "Data of data-logging" makes it possible to print the newcoming species since the last catalogue. The box "notes" makes it possible to show if the species is a type species (holotype/paratype). This can also be done by adding a \* to the PMBC No. for the species.

#### Daily work

When the species name has been logged one time it is not necessary to do it again, because all additional specimen of the same species will have a new number. Thus a reference from the species to the number is all that has to be done. Looking at the example given in Fig. 1, additional specimens of the species *ramosus* will be given a number and details about the new lot will be connected to that number. Thus changes and updates will be relatively easy to do. With this structure it is also easy to add new categories since it is based on cross references.

The structure of this database differs from a database where each specimen has a "card" or "file". In the latter case, if changes are going to be made, it is necessary to find every "card" of the species and change the information. Furthermore, a database of the proposed structure is capable of keeping more information together than a base based on individual "cards".

## REFERENCES

- Elmasri, R. & S. B. Navathe. 1994. *Fundamentals of Database Systems*. Second edition. The Benjamin/Cummings Publishing Company, INC. Redwood City, California, USA.
- Hylleberg, J. 1993. Indo-Pacific Cockles (Bivalvia:Cardiidae). Part 1. Generic diagnoses and an overview of species with mention of taxonomic problems encountered in Thailand. - Phuket marine biological Center Special Publications **13**: 113-136.
- Hylleberg, J. 1994. Phylum Sipuncula. - Part 1. A detailed catalogue of valid genera, species, synonyms & erroneous interpretations of sipunculans from the world, with special reference to the Indian Ocean and Thailand. - Phuket marine biological Center Research Bulletin **58**: 1-88.
- Kabat, A. R. & K. J. Boss. 1992. An Indexed Catalogue of Publications on Molluscan Type Specimens. - Occasional Papers on Mollusks. The department of Mollusks, Museum of Comparative Zoology, Harvard University Cambridge, Massachusetts. **69**(5): 157-336.
- Mayr, E. & P. D. Ashlock. 1991. *Principles of Systematic Zoology*. Second edition. McGraw-Hill, INC. Singapore.
- McFadden, F. R. & J. A. Hoffer. 1994. *Modern Database Management*. Fourth edition. The Benjamin/ Cummings Publishing Company, INC. Redwood City, California, USA.
- UNESCO. 1979. Organization of marine biological reference collections in the Mediterranean Arab countries. UNESCO reports in marine science 6. 30 pp.

## ANNEX

The basic Registration Card from Phuket Marine Biological Center, and the Card used together with the primary note, at Phuket Marine Biological Center.

PMBC

**REGISTRATION CARD**  
FOR REFERENCE COLLECTION

Phylum	Institution Donated by	Coll. Number Ph-Number Photograph Number
Class	Order	Family
Species		
Locality		
Collector	Date	Notes
Identifier	Date	Notes
Confirmed by	Date	Notes

\* the back side of the card is used to write "field data recorded by collector".

PHUKET MARINE BIOLOGICAL CENTER Dept. of Fisheries, Phuket, Thailand.	
Family	Ph. No.
Scientific name	
Locality	Date
Identified by	Date