Stuttgarter Beitr. Naturk.	Ser. A	Nr. 456	131–137	Stuttgart, 31. 12. 1990

2. Data Storage and Analysis by Computer

Floristic and faunistic Databases on Personal Computers

By Paul Diederich, Luxembourg

With 4 figures

1. Insufficiencies of a distribution atlas – necessity of a database

A distribution map gives inaccurate and incomplete data; a database is necessary to know the precise location, date, ecology, references, etc. Furthermore it is difficult to correct errors on a distribution map without knowing the precise references. Finally a database allows a lot of other applications.

2. Mainframe versus personal computer

The following table gives the main characteristics of, and differences between, mainframes and personal computers. Both systems are nowadays very quick and offer sufficient memory capacities for large floristic and faunistic databases containing more than 10 million data. The success of personal computers may be derived from their easy accessibility, especially for amateur scientists. They have the disadvantage that simultaneous use of the same database on several computers leads to coding difficulties.

	Mainframe	Personal computer
Accessibility	Easy for institutions, difficult or impossible for private scientists	Easy access for institutions, private scientists and amateurs
Memory capacity	Sufficient for every database	Sufficient for more than 10 million records (on one hard disk or optical disk)
Speed	Generally high, but repeated access to memory by telephone line or simultaneous activity on many terminals may slow down the speed	Generally high, but access to disk is slow. Good database programs using indexes can be very fast
User comfort	Bad	Generally good (e. g. graphic possibilities)
Use by several persons	Easy	Coding difficulties

3. Coding difficulties on personal computers

3.1. Difficulties

On mainframe computers several users work with the same database. If, for instance, a new name of a taxon is introduced in the database, this name will be accessible to every user.

On personal computers, however, there are as many physically different and separated databases as users, working each on his own PC. By adding a new taxon to the database, each user must use the same code for the same species. There is the risk that two users will assign the same code to several different taxa, or that two users will assign different codes to the same taxon. The same problem arises with the coding of publications, observers, etc. From time to time all the data introduced into the different private systems must be merged to form the complete database.

3.2. Solutions

Method 1

One rapid PC with a large memory capacity functions as a server, i. e. it is linked by modem to a telephone line, and every user uses his PC as a terminal.

Result: only 1 database, as with a mainframe.

Method 2

To begin with, rather a complete list of taxa, publications, etc. should be available. If new taxa or publications must be added, one person should be responsible for the attribution of new codes. To code publications, a particular coded list is established for each country, and for each group of organisms: one supervisor is needed for every country.

For practical reasons, it must be accepted that persons from different countries use different programs and a different database structure lacking any compatibility. The flux of data between several countries can then be achieved by sending printed listings by post (e. g. list of the existing European specimens of one species; list of UTM squares concerning one species) or by transmitting a subset of data understandable by every database (containing for instance the country, locality, UTM square, date and species) by telephone line or on a floppy disk.

Merging different databases

Once a year the supervisor of the database receives the new data from other users (on a floppy disk or by a telephone line) and merges them with the existing data. Each user can get the complete database afterwards.

4. Coding of taxa

4.1. Different ways of coding taxa

In databases on mainframe computers it may be conceivable and even useful that a proper code is automatically assigned to each of the different epithets, and that the synonymies are added by a specialist later on.

On PCs, however, an almost complete list of epithets must be available to begin with, and only one supervisor is allowed to add further taxa or synonyms.

To add data to the database, a special interface may allow the user to designate the species by their full name, an abbreviation or a code. If we want a comfortable and easy management of the data, the machine should, however, internally use a suitable code for dealing with problems like synonymy or hierarchy of taxa. Three different coding methods are often used:

(a) Abbreviation

E. g. RamFar *Ramalina farinacea*

Disadvantages: Instability of nomenclature.
Taxonomical changes.
Identical abbreviation for 2 taxa.
Different abbreviations for 1 taxon.

(b) Hierarchical coding

E. g. 003.02.2 *Amanita citrina* var. *citrina* (cf. Standaardlijst van Nederlandse Macrofungi; 003 denotes the genus, 02 the species, and 2 the variety).

Disadvantages: Taxonomical changes.

(c) Numerical coding

E. g. 1410 *Thelotrema lepadinum* (cf. British list for lichen mapping).
Disadvantages: None.

4.2. Principles of numerical coding of taxa

(a) A numerical rank is attributed to every taxonomical rank

E. g.:

4	Regnum	75	Species s.l.
12	Division	76	Species
20	Class	77	Species s.s.
28	Order	78	
36	Family	79	Subspecies s.l.
44	Tribe	80	Subspecies
52	Genus	81	Subspecies s.s.
60	Section		
68	Series		
76	Species		
84	Variety		
92	Form		

(b) Taxa of different ranks are coded in the same way

Code	Taxon	Rank
1407	Graphidales	28
1408	Thelotremataceae	36
1409	*Thelotrema*	52
1410	*Thelotrema lepadinum*	76
1411	*Thelotrema monosporum*	76
0795	*Lecidella elaeochroma* s.l. (incl. *L. achristotera*)	75
0796	*Lecidella elaeochroma* s.s. (excl. *L. achristotera*)	77
0797	*Lecidella elaeochroma* var. *elaeochroma*	84
0798	*Lecidella elaeochroma* var. *soralifera*	84

(c) Taxonomical synonyms must have different codes

Code	Taxon	Rank	Synonym
1234	*Ramalina farinacea*	76	−1237
1237	*Ramalina reagens*	76	1241
1241	*Ramalina subfarinacea*	76	1234

If one author considers that *Ramalina farinacea, R. reagens* and *R. subfarinacea* are synonyms, he will nevertheless code epiphytic continental specimens as 1234 *R. farinacea,* and saxicolous maritime specimens as 1241 *R. subfarinacea.* If these taxa are treated as different species later on, no information will be lost.

In the chart above, each taxon points to the synonym following, and the last synonym points to the first one. The currently used epithet is marked by a minus sign.

Note: Nomenclatural synonyms may have the same code.

(d) Coding of the hierarchy using 2 links

Code	Taxon	Rank	Syn	Link1	Link2
0300	Gen1	52		0301	
0301	Sp1	76			0302
0302	Sp2	76		0304	0303
0303	Sp3	76			0300
0304	Sp2Var1	84			0305
0305	Sp2Var2	84			0302

Each taxon points to the first taxon of a lower rank through Link1. Link2 points to the taxon, following, of the same rank, or, if no taxon of the same rank remains, to the taxon of a higher rank (Fig. 1).

5. Coding of doubtful or inaccurate data

5.1. Doubtful determination

Thelotrema sp. is coded as
1409 *Thelotrema*
Thelotrema cf. *lepadinum* is coded as
1410 *Thelotrema lepadinum*
and the record will be marked as doubtful.

5.2. Inaccurate date

Different possibilities:

27. 9. 1929
9. 1929
1929
about 1929
1929 ± 4
1925 – 1933

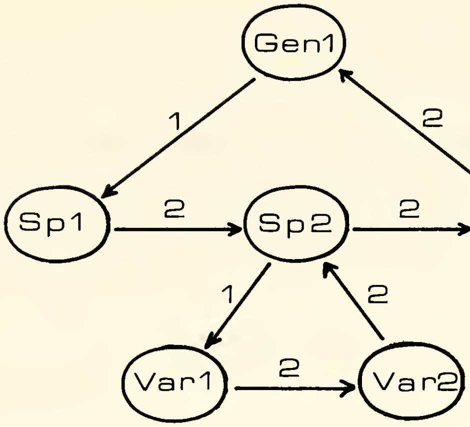


Fig. 1. Coding of the hierarchy using two links, 1 and 2.

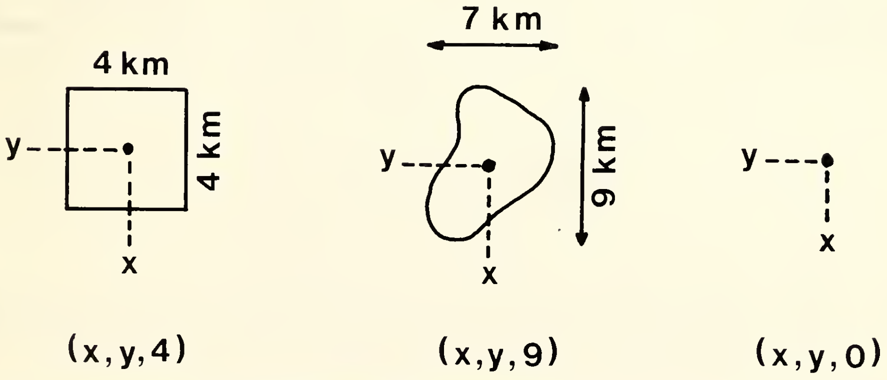
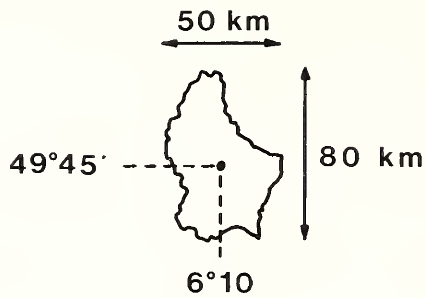


Fig. 2. Coordinates (x, y, d) in three different situations.

5.3. Inaccurate location

The localities are easily coded using geographical coordinates expressed either in a „natural" system (e. g. longitude – latitude) or in an „artificial" (national or international) grid system (e. g. UTM, MTB in Germany, IFBL in Belgium and Luxembourg, etc.). Many locations given by authors of the 19th century are inaccurate, and a correct designation in any system of (x, y) coordinates is difficult or even impossible. For treating these data by computer in an efficient way (without loss of information), coordinates (x, y, d), where (x, y) represents the centre of the studied area, and d represents the diameter of the same area should be used. Three different situations may occur, either squares (or rectangles), areas of an arbitrary shape or a precise location. The diameter (in km) is chosen as the largest among the vertical and horizontal dimension of the given area (Fig. 2).

The inaccurate location „Luxembourg" (precise locality unknown) (Fig. 3) will be coded as (6°10′, 49°45′, 80).



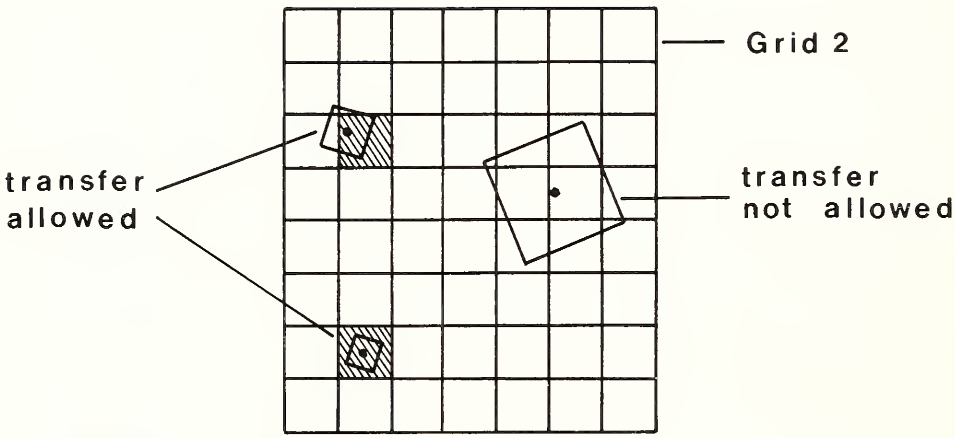


Fig. 3. (above) Coding of the inaccurate location „Luxembourg".
Fig. 4. (below) Transfer of data from grid system 1 (three different squares) to grid system 2. The transfer is allowed for the two small squares, but it is not allowed for the large square whose diameter is larger than the diameter of the squares of grid 2.

6. Compatibility of different mapping systems

6.1. Principles

(a) A distribution map need not necessarily contain precise data (which can be found in the database).

(b) Errors occurring on a distribution map must be rare (e. g. less than 25% of the data) and small (e. g. less than 30% of the diameter of the squares).

6.2. Method

The transfer of data from a square S1 (x1, y1, d1) to a square S2 (x2, y2, d2) is allowed if:
(a) the centre (x1, y1) is situated in the square S2;
(b) $d1 \leq d2$.

Otherwise the transfer is not allowed, and data of system 1 cannot be used for a distribution map in system 2 (Fig. 4).

7. Applications of a database

(a) Atlas of distribution maps.

(b) List of species (from a locality, from a country [checklist], from a period, growing on *Quercus*, etc.).

(c) Data on one species (concerning herbarium material, literature reports, field observations; investigations about the ecology, frequency, etc.).

(d) Complex requests (e. g. find a list of all the localities where *Lecanora argentata* or *L. carpinea* was found in the 19th century on *Fagus*, and where none was found after 1980).

(e) Prepare reports to be included in a manuscript by a word processor (e. g. a list of the specimens of one species; a list of the species from one locality).

(f) Herbarium management.

8. Acknowledgements

I wish to thank Mr. Claude Meisch for revising the manuscript and Mr. Jean Krier for reading the English text.

Author's address:

Paul Diederich, Musée National d'Histoire Naturelle, Marché-aux-Poissons, L-2345 Luxembourg, G. D. Luxembourg.