

THE NEW GRONINGEN ^{14}C DATA BASE

JOHANNES VAN DER PLICHT

Center for Isotope Research, University of Groningen, Nijenborgh 4
9747 AG Groningen, The Netherlands

ABSTRACT. The Groningen Radiocarbon Data base has been operating on an Apple IIe PC until recently. Both software and hardware of this data base made exchange with other systems very difficult. Therefore, the data base has been transferred to an IBM AT compatible computer. A new program has been written in Turbo-Pascal, using the Turbo-Pascal Database Toolbox. The program features versatile search routines. The output can be directed to various date-list formats, such as the International Radiocarbon Data Base (IRDB). The program is written in such a way that new date lists can be added with minor adjustments to the source code.

INTRODUCTION

Since May 1952, the Isotope Physics Laboratory in Groningen has dated many thousands of ^{14}C samples; by May 1991, 17,500 GrN numbers have been issued, corresponding to 15,000 dates and 14,000 samples. In the beginning years, results were published in *RADIOCARBON* date lists. This practice was stopped because of the large amount of dates produced (at present around 1000 per year in Groningen), and because of the introduction of computerized information systems.

In 1985–1989, all data from the archives were entered in the first Groningen ^{14}C Data Base. This database system consisted of an Apple IIe microcomputer, equipped with 64 kb extra memory, a 10 Mb external disk and a tape streamer for backup. This setup was costly and sophisticated for the time the project was started. The database program was locally developed in both UCSD/Apple Pascal and Assembler programming languages. Standardized submission forms were introduced for the users of the laboratory (Engelsman, Taayke & Mook 1986).

The Apple database system operated successfully for several years in the Groningen laboratory. However, it had a few disadvantages:

1. The program is inaccessible to anyone outside the laboratory, making software modification close to impossible;
2. Converting the data base to other programs/computers/users is possible (Walker *et al.* 1990), but difficult;
3. The hard disk is limited to around 20,000 dates;
4. The hardware became unreliable and out of date.

For these reasons, we decided to change the Apple database system to a more modern and standard setup such as MS DOS. The changeover consisted of two steps: 1) transferring the data base from the Apple to the new PC; 2) adapting the database contents to a new software package. For the new PC, we have chosen an Olivetti M290 (AT compatible) with both a 40 Mb disk and 40 Mb tape streamer. The data base was transferred from the Apple to the M290 *via* the RS-232 interface, using specially developed software. At the moment of transfer, the data base contained 16,062 records. The transferring process took about 40 h. The Apple was finally decommissioned in the beginning of 1990. The possibilities for data retrieval are now enlarged (*e.g.*, search for ^{14}C -activity and $\delta^{13}\text{C}$). Also, space for textual information (such as references, comments, *etc.*) has been added.

THE DATABASE PROGRAM: SOFTWARE

The new Groningen database program is written in Turbo Pascal (Borland 1989), using the Turbo Pascal Database Toolbox package. The source code is written in a more-or-less standard programming language, which is easily accessible, and exchange of output with computers at other laboratories is straightforward.

The Database Toolbox can handle large data bases very efficiently – the maximum number of records is practically unlimited. The program is inexpensive, and Pascal programmers can make changes to the source codes relatively easily. Alternatives, at the time of writing the program, were large and/or expensive codes, such as ORACLE, INGRES and others. In the near future, the new database program will be connected to other programs used in Groningen, such as the calibration code (van der Plicht & Mook 1989) and the data collection software for our proportional counters (van der Plicht, Streurman & Schreuder 1992), which are also written in the same Turbo Pascal language.

The Turbo Pascal Database Toolbox (TAccess) uses the so-called “B+ tree,” a binary tree-like concept (Borland 1989; Bayer & McCreight 1972). With TAccess, the data file is just one big array of records. A separate file, called an Index File, is used to organize the data into the B+ tree structure. In the index file, each data record is abbreviated into a key and a data reference. A key is usually a portion of a data record that is sufficient to determine the ordering of the data records; for example, one or more fields in the data record. A data reference is a pointer that indicates where the complete data record is stored. The Toolbox provides procedures for accessing, inserting and deleting data records. Because it uses B+ trees, each operation involves a very small number of disk accesses and is very efficient.

The B+ tree (index file) based search procedures can only be used for numerical fields. Search procedures on alpha-numeric fields are sequential, *i.e.*, the complete data base will be searched record-by-record; the index-file concept cannot be used in this case.

Actually, the database program consists of two independent data bases: 1) the Groningen Radiocarbon Data Base; 2) the Institute Data Base. The latter consists of names and addresses of the submitters. This avoids unnecessary duplication of these data (and saves about 2 Mb of disk space).

The 40 Mb hard disk of the computer is divided into three disk drives – C, D and E as follows:

C	10 Mb	Programs, menus, <i>etc.</i>
D	20 Mb	Radiocarbon data file
E	10 Mb	Institute data and index files

The size of the files are: 82 kb for the (executable) radiocarbon database program, 49 kb for the (executable) institute database program, 14 Mb for the Radiocarbon Data Base and 235 kb for the Institute Data Base. The index files take about 500 kb each. One record comprises 795 bytes. For 16,062 records in the data base, a search time of 25 s was noted for index files. For sequential searches, the search time will be around 2 min. This is largely determined by the disk-access time of the hardware, and is 40 ms for our setup.

THE DATABASE PROGRAM: DESCRIPTION

Upon starting the computer, a menu offers the user choice of the Radiocarbon Data Base, Institute Data Base, Tape Streamer programs or other (not directly related to the data base). The database programs are safeguarded by a password. If no password is entered, the data base can only be

searched/viewed. When the correct password is entered, the data base may be edited or records may be added.

Each record of the Radiocarbon Data Base contains 60 fields. The fields are listed in Table 1. The field length (no. of characters) is also given, as well as a code. Code numbers 1–24 correspond to Integer variables in the program, 25–99 to Real variables, and ≥ 100 to other entries (such as character strings). These code numbers determine the searching procedures (such as Index or Sequential) and are necessary information for programmers searching the source code. The fields are organized into five screens. An example of screen 1 is shown in Figure 1. Screen 1 contains information on the sample context; screen 2 – location; screen 3 – the submitting institute; screen 4 – the results; screen 5 – dates of measurement and counter. Key figures (such as GrN number, sample name and final age) are shown on top of each screen.

The data base can be searched in two ways: *View* and *Find*. The *View* option allows searching for a particular GrN number or Institute code. The *Find* options are the most extensive. They are listed

TABLE 1. Field Names and Codes for Records in the Radiocarbon Data Base

Field names	Length	Code	Field names	Length	Code
Record status	*	5	Expected age 1	8	6
GrN number	5	3	Expected age 2	8	7
Sample number	8	1	Final age 1	8	2
Sample name	30	100	Final age 2	8	8
Submitter	30	108	Final age 3	160	9
Collector	30	109	Remarks	4	105
Communication date	8	110	Institute code	8	4
Town/region	20	106	Geol. coord. longitude	8	34
Province/district	20	107	Geol. coord. latitude	8	35
Country	20	104	Local coord. X	8	36
Depth one (from)	8	27	Local coord. Y	8	37
Depth two (to)	8	28	Measurement date 1	8	118
Altitude one (from)	8	40	Measurement date 2	8	119
Altitude two (to)	8	41	Measurement date 3	8	120
Material 1	20	101	Measurement date 4	8	121
Material 2	20	111	Measurement date 5	8	122
Pre-association	15	102	Measurement date 6	8	123
Investigation type	15	112	Measurement date 7	8	124
Context of sample	15	113	Measurement date 8	8	125
Layer	15	114	Counter meas. 1	2	10
Field of study	20	103	Counter meas. 2	2	11
Chemical treatment	20	115	Counter meas. 3	2	12
Dated material	15	116	Counter meas. 4	2	13
¹³ δ (C14)	6	25	Counter meas. 5	2	14
¹³ δ (carbonate)	6	29	Counter meas. 6	2	15
¹⁸ δ (carbonate)	6	30	Counter meas. 7	2	16
Measured activity	7	26	Counter meas. 8	2	17
Sigma meas. activity	6	31	Percentage C	5	38
Final activity	7	32	Percentage C sample	5	39
Sigma final activity	6	33	Security	1	117

*Used by the Database Toolbox

```

1  2  3  4  5  Next  Previous  Update  Exit
14C DATABASE

GrN: 8350      Nr Sample: 4610      Screen : 1
Sample Name: Hazendonk 29e (57-100/102)
Final age BP: 4755      35

Material 1: peat      Sample Context: sediment
Material 2:      Investigation: excavation

Study: archaeology      Pre Association: Neolithic

Expected Age BP: 4800      - 5000

Remarks: Base of peat layer, overlying habitation
layer (Hazendonk 3 pottery).
Lit. (subm., Rhine/Meuse Delta. Thesis L
eiden 1974).

```

Fig. 1. Example of screen 1 for the new Groningen Radiocarbon Data Base

in a window on the screen (Fig. 2A). The search for GrN number, sample number, age BP, delta ^{13}C and ^{14}a (measured activity) uses the index files (they are integer and real variables), and is very fast. The user has to enter lower and upper limits for the values to be searched.

The other entries listed (sample name through study) are character strings, so that the search process is sequential (and consequently, relatively slow). The program asks the user to enter (part of the) string. In exceptional cases, one may wish to search for a field that is not in a standard format: this is made possible through the "Own Choice" option. One must enter a code (listed in Table 1), so that the program knows what field to search and whether it is an indexed or sequential search. After the search process is completed, the program produces a *List* menu.

The user can select the type of date list (described below), select output (screen, printer or file on disk) or go through another search with *Find*. In principle, one can continue the search process until the last record is found. Multiple simultaneous string searches are not possible. The second (and higher) search processes are always sequential; the list of records found after every search can be inspected as one of the date lists.

```

Find  View  Add  Quit
14C DATABASE      17587 records
Nr Sample:      Screen : 1

Sample Context:
Investigation:
Pre Association:

A GrN
B Sample nr
C Age BP
D Delta C13
E Meas. Act. C14
F Samplename
G Material 1
H Pre Association
I Country
J Province
K Town
L Remarks
M Study
N Own Choice Index
O Own Choice Seq.

```

Fig. 2A. Find options for the Radiocarbon Data Base

For the Institute Database program, the View/Find options are similar to the Radiocarbon Data Base described above. With *View*, one can only search for the Institute code; with *Find*, for the institute code, name, address, city and country (Fig. 2B).

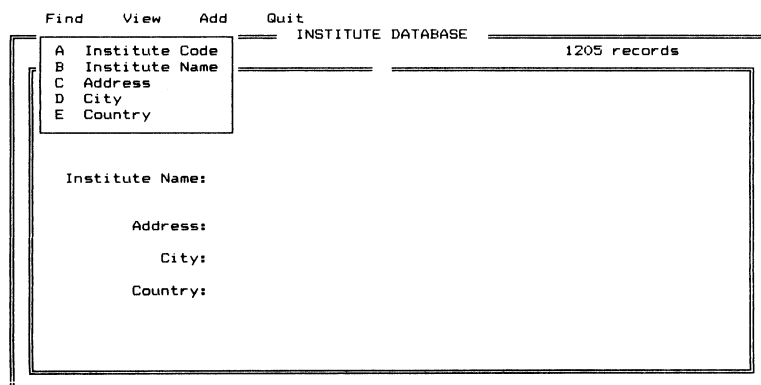


Fig. 2B. Find options for the Institute Data Base.

DATE LISTS

Several date-list formats can be selected as output, and routed to screen, printer or file on disk. At present, four date lists can be selected.

Date list 1 is the radiocarbon date list, which has been used over the years in the publications of *Radiocarbon*.

Date list 2 is called HLF (High Level Format), and has been used for data-exchange programs between Groningen and Harwell/Oxford in the UK (Walker *et al.* 1990).

Date list 3 is a one-line summary consisting of only GrN number, sample name and the age BP.

Date list 4 contains the fields of the IRDB format (Kra 1989).

Date list 5 is reserved for a "Standard Letter," which is a standard communication letter with results forwarded to customers. This option will be used in the new Groningen Laboratory, when the data base can be connected to other administrative programs through a network. Also under consideration is connecting the data base to a spreadsheet.

Finally, there is a date list that contains all information from the records selected. Illustrative examples of Date lists 1–4 are presented in Table 2.

The database program is written so that other date lists are relatively easily implemented with only a limited knowledge of Turbo Pascal.

TABLE 2. Examples of Date Lists

Date list 1, Radiocarbon:

GrN-15113	Bandung basin E4-3C	27,060 ± 260 BP
	($\delta^{13}\text{C} = -26.01$)	
	Bandung basin / Java / Indonesia	
Peat from exposure, dating sediment		
Treatment: acid		
Taken from upper section of layer		
Dated fraction: organic residue		
Submitted by W. Roeleveld (AWA)		

TABLE 2. Continued

Date list 2, HLF:

*LABCODE	GrN-15124 GRONINGEN
*SITENAME	Hengelo A1 P7
*SUBM	By S. Bohncke & J. v. Huissteden
*COLL	By
*TOWN	Hengelo
*REGION	Overijssel
*COUNTRY	Netherlands
*SAMMATERIAL	Peat organic residue
*PERIOD	Pleniglacial
*DISCIPLINE	Geology
*TREATMENT	Acid / alkali / acid
*DC13	-26.98
*PERMOD	2.038 ± 0.155
*AGEC	31,250 ± 600
*SITETYP	Taq ice wedges
*SUBMADDR1	Instituut voor Aardwetenschappen Vrije
*SUBMADDR2	Universiteit

Date list 3, Summary:

GrN-15113	Bandung basin E4-3C	27,060 ± 260 BP
GrN-15124	Hengelo A1 P7	31,250 ± 600 BP
GrN-15123	Hengelo A1 P4	33,100 ± 600 BP
GrN-15114	Bandung basin E17-10C	34,700 ± 1000 BP
GrN-15119	Bandung basin F73-14C	35,700 ± 900 BP

Date list 4, IRDB:

Lab code	GrN-15113 GRONINGEN
Lab no.	11222
Sample name	Bandung basin E4-3C
¹⁴ C determination (BP)	27,060
¹³ C value	-26.01
Sample material	Peat
Country	Indonesia
Lat	107°32'29"E
Long	6°59'35"S
Sitename	Bandung basin
Submitter	W. Roeleveld
Discipline	Geology
Association	
References	
Comments	

ACKNOWLEDGMENTS

The new database program was written by L. H. Bijlsma, J. Flapper and H. Reitsma. H. G. Jansen made possible the transfer from the Apple to the new computer. E. Taayke has been responsible for maintaining the data base.

REFERENCES

- Bayer, R. and McCreight, E. 1972 Organization and maintenance of large ordered matrices. *Acta Informatica* 1: 173–189.
- Borland 1989 Turbo Pascal. Scotts Valley, California.
- Engelsman, F. M. R., Taayke, E. and Mook, W. G. 1986 Groningen ¹⁴C data base. In Stuiver, M. and Kra, R. S., eds., Proceedings of the 12th International ¹⁴C Conference. *Radiocarbon* 28(2A): 788–796
- Kra, R. 1989 The International Radiocarbon Data Base: a progress report. In Long, A. and Kra, R. S., eds., Proceedings of the 13th International ¹⁴C Conference. *Radiocarbon* 31(3): 1067–1076.
- van der Plicht, J. and Mook, W. G. 1989 Calibration of radiocarbon ages by computer. In Long, A. and Kra, R. S., eds., Proceedings of the 13th International ¹⁴C Conference. *Radiocarbon* 31(3): 805–816.
- van der Plicht, J., Streurman, H. J. and Schreuder, G. R. 1992 A new data acquisition system for the Groningen counters. *Radiocarbon*, this issue.
- Walker, A. J., Olet, R. L., Housley, R. A. and van der Plicht, J. 1990 Operation of the Harwell UK ¹⁴C data base and its expansion through data exchange with other laboratories. *Radiocarbon* 32(1): 31–36.