STORING AND MANAGING DATA ON A COMPUTER

Nicholas Lee, Andrew Millman, Martin Osborne, John Cox

Advantages of an electronic database

- Highly compact data storage
- Fast and accurate searches
- Easy to amend stored data
- Easy data manipulation
- Multiuser access
- Printed reports
- Ease of backup

All readers will be familiar with the traditional card index which is used to store relatively simple information such as a list of books in a library. Database programs are the electronic equivalent of a card index but offer several advantages. Most important of these is the ability to search and analyse the stored data rapidly and reliably. In addition to this, large amounts of information can be stored in a fraction of the space that would be required for equivalent paper records; several people on a network or multiuser system can use the same database simultaneously; and the results of searches can easily be printed in the form of a report. Furthermore, any stored data can be used by other programs such as a word processor (for creating circular letters or mailing labels), a spreadsheet (for manipulation and charting of the data), or a statistical program (for detailed analysis).

Structure of databases

Simple flatfile database.

Information stored in a database is broken down into manageable portions which are stored in separate compartments called fields and records. For example, a surname would be entered into one field, and the first name or initial and date of birth into subsequent fields. A collection of fields relating to one person or event is called a record. A database can contain an unlimited number of records.

Flatfile databases

Flatfile databases are the most basic type and are analogous to a card index. They are easy to create with any database program and are ideal for storing simple information such as a list of names and addresses. They offer all the facilities that many people are likely to need.

Relational databases

In a relational database information is stored in several separate, cross referenced database files. Although more complex, this structure allows very powerful and efficient systems to be developed. A good example of this is a database containing basic information about a group of patients, including the name of the general practitioner with whom each patient is registered. In a flatfile database the name of the patient's general practitioner would have to be entered in every record. In a relational database you would enter the general practitioner's name just once into a separate small database. The patient's record is then designed to point to the correct doctor in the general practitioner database. The advantage of this approach is that the person entering patient data can be offered a list of general practitioners derived from the general practitioner database (called a pick list), which saves time and reduces the risk of error. Furthermore, if a doctor changes, it is necessary to alter only one record in the general practitioner database. The patients' records would then update automatically.

GLOUCESTER MEDICAL EYE CENTRE			
Surname	Smith	House Name	The Old Forge
Forename	Pauline	Street	High Street
Title	Mrs	Town	Painswick
Birthdate	15-07-38 age 56	Postcode	GL4 8JP
		Telephone	0123-456789
Is this patient exempt Y/N	N	GP Surname	Address 1
Recall this patient Y/N	Y	Address 2	
GP's name	Dr. A. Johnston	Dr. S. Higgs	The Health Centre
GP's address	Cotswold Surgery Hillside Walk Painswick GL4 8HP	Dr. A. Holland	The Surgery
		Dr. J. Bayland	The Surgery
		Dr. R. Burt-Williams	The Surgery
		Dr. K. Jacques	The Surgery
		Dr. B. James	Hillside
		Dr. M. Jones	The Surgery
		Dr. C. Jansen	The Surgery
		Dr. S. Johnson	Cotswold Surgery
		Dr. A. Johnston	Cotswold Surgery
		Dr. I. Kelsey	Barloughton Surgery
		Dr. B. Knight	The Surgery
		Dr. R. Lamb	Locking Hill Surgery
		Dr. E. Lewis	Locking Hill Surgery
		Dr. H. Lewis	The Surgery
		Dr. P. Losh	Barloughton Surgery
		Dr. A. Lynch	College Yard Surgery
		Dr. B. Mackay	The Surgery
		Dr. I. Macleod	The Surgery

GP Database

Example of a relational database.

Creating a database

1. Define the fields

2. Create a user friendly data entry screen

Database ready to use

Field Name	Type	Length	Format
Surname	C	25	Character
Firstname	C	25	Character
Initial	T	1	Character
Title	C	1	Character
Birthday	D	8	DD-MYY
Sex	T	1	Character
Address 1	C	28	Character
Address 2	C	28	Character
Address 3	C	28	Character
Postcode	C	8	Character
Telephone	D	11	Character
Registration	T	1	Character
Recall list	C	1	Character
Exempt	C	1	Character
Last entered	T	8	Character
GP's name	C	28	Character
GP's initial	C	1	Character
GP address 1	T	28	Character
GP address 2	T	28	Character
GP address 3	T	28	Character

Surname: Smythe

Firstname: Pauline

Title: Mrs Initial P

Birthday: 15-07-38 Age: 56

Is this patient exempt Y/N: N

Recall this patient Y/N: Y

GP's name: Dr R Johnston

GP's address: Cotswold Surgery, Hillside Walk, Painswick, GL4 9HF

Tel No.: 0197 454321

House Name: The Old Forge

Street: High Street

Town: Painswick

Postcode: GL4 8JF

Telephone: 0129 456789

LAST TEST ON: 17-04-95

Test again in 24 months

Recall for test Apr 97

Steps for creating a database

Unlike with other programs such as word processors, which can be used immediately, you have to create the database structure before you can enter any data. Although designing flatfile databases is relatively straightforward and well within most people's capability, creating a relational database is more difficult.

Traditionally, it has been necessary to employ a professional programmer or to master a database programming language such as dBase. Handholding techniques such as Microsoft's Wizards have now made it easier to create quite complex databases without using a programming language. Nevertheless, unless you are already familiar with a database your time would probably be better spent explaining your needs to a programmer. It is very easy to invest considerable time and effort trying to write the program yourself but remain disappointed with the end result.

Validation of data and error trapping

ADD RECORD

Col:1

Surname: Smythe

Firstname: Pauline

Title: Mrs Initial P

Birthday: 15-07-38 Age: 56

Is this patient exempt Y/N: N

Recall this patient Y/N: Y

GP's name: Dr R Johnston

GP's address: Cotswold Surgery, Hillside Walk, Painswick, GL4 9HF

Tel No.: 0197 454321

House Name: The Old Forge

Street: High Street

Town: Painswick

Postcode: GL4 8JF

Telephone: 0129 456789

LAST TEST ON: 17-04-95

Test again in 24 months

Recall for test Apr 97

Invalid Date Entered

Some programs can be set up to detect errors automatically

It is important to ensure that information is entered into the database consistently and accurately. Errors made during data entry often result in records being missed in a search. To reduce the possibility of error, many programs offer error trapping facilities that can be included in data entry screens. These could, for example, be used to define a valid age or height range so that bizarre mistakes such as men aged 164 years or women 4 m tall are not accepted.

Running the database

Advantages of compiled databases

Run faster

Can be distributed freely

No need for the main database program

Database structure cannot be inadvertently corrupted

Once you have created and tested the database structure it can be run using the main database program. It is better, however, to use a compiling program to convert the database structure into a stand alone program. Alternatively use a "run-time" module, which is a small program capable of running your database without the need for the main program and which can usually be distributed freely to other users. Both of these methods prevent accidental damage to your database structure. Compiled databases have the additional advantage of running more quickly. Both techniques allow the database to be safely distributed to other users and are used by companies who market "ready to use" databases for specific purposes such as general practice systems.

BMJ VOLUME 310 26 AUGUST 1995

To retrieve information from the database you need to search one or more fields. A simple search on a single field might, for example, be used to list all the patients with a particular surname or living in a given area. More complex searches of several fields using terms such as "and," "or," "if," and "not" could be used to identify a group such as all men aged under 60 who do not smoke but have coronary heart disease.

Most databases offer facilities to search for fragments of a name—for example, "Jo" could be used to search for Jones, Johnson, or Joseph. Other programs offer an option to expand the search to include similar sounding names spelt differently—for example, Smith and Smythe.

The quality of reports is entirely dependent on the accuracy of the data entered.

"Rubbish in, rubbish out"

Creating a report

PATIENTS FOR RECALL			
RECALL MONTH: Nov 95 List Generated on: 17-09-95			
First Name	Surname	Address	Telephone
Sharon	Davies	Pilkington Court	01452-504167
Anne-Marie	Fellows	Rocking Road	01452-615478
John	Hartley	Riverside Close	01452-619730
Sandra	Lassiter	Sandhurst Lane	01452-601875
Mark	Leatham	Harrington Street	01452-601815
Safwabibi	Mustafa	456 Falkner Street	01452-951753
Caroline	Peters	52 Barlett Street	01452-601508

Total Number of Patients due for recall in Nov 95 = 7

Recall list which can be used to generate a standard letter.

The results of a search can be used in several ways. You might, for example, want to view the results immediately on the screen, either one at a time or with all the results listed in a table. Alternatively, you could print the results in a report complete with subtotals and totals. Results can also be merged with a template to produce a personalised letter addressed to each person who meets the search criteria (as in call and recall systems). Once you have created a search strategy, it can be saved and used again as often as necessary.

Most databases allow you to export information in different file formats so that the data can be used by other programs. It is therefore important to ensure that the database is compatible with your chosen program.

Bibliographic software

Bibliographic software helps organise your references.

References from Medline are automatically converted into the format used by the bibliographic software.

Biomedical literature has expanded rapidly in recent years. Over 300 000 new articles are added to the Medline database each year, and this is just one of many computerised databases. Doctors doing literature searches with these databases commonly find more than 100 references on a specific subject. The challenge is to manage this flow of information efficiently so that key references can easily be stored on your personal computer and located some months later when you need to use them. Several bibliographic reference databases have been developed to meet this challenge. They allow large numbers of references to be stored, indexed, and retrieved easily.

Creating a reference database

References are managed in bibliographic software packages by storing them in groups of related subjects. The software automatically takes care of the structuring necessary to accommodate a range of different reference types—for example, journal articles, book chapters, abstracts.

Importing data or references

Although you can type references into the database using the menus and prompts provided, it is far easier to import them electronically from other sources. Most libraries offer facilities which allow you to copy references from CD ROMs, or you could dial into a remote database such as Medline in the BMA Library. When importing data electronically the bibliographic software automatically converts references from the native format into that required by the program. This eliminates the tedium of typing and reduces the risk of transcription errors.

Reference Manager will show you the style required by the main medical journals.

References can be converted to the required style simply by clicking a button.

Fast and flexible retrieval of specific references is essential, especially as the database grows. The software lets you search for references by the name of an author, a word in the title of an article, or the name of a journal. Different combinations of terms can be searched by linking them through the “and,” “or,” and “not” operators. Thus a search for “kidney and transplantation” will find only those references in which both terms occur.

Managing reprints

Bibliographic software can also help you keep track of any reprints that you might have obtained. Some packages include a field in which it is possible to indicate if a reprint is held for a given reference and, if so, where the reprint is filed.

Generation of a bibliography

For anyone who publishes papers regularly the most valued feature of bibliographic software is the facility to create a list of references which can be appended to a manuscript. The publishers of medical journals and books all have their own specifications about the way in which references should be formatted—for example, Vancouver or Harvard styles. Most of the software packages will automatically produce the reference list in several different styles to meet the requirements of the major journals. Furthermore, the software will also place appropriate reference numbers in the text of your manuscript. These features take much of the drudgery out of preparing a paper.

Features of bibliographic software

- Allows creation of groups of related articles—for example, one for AIDS and one for tuberculosis
- Automatic incorporation of data or references from other sources such as Medline
- Easy retrieval of data or references
- Management of reprints
- Generates lists of references in all standard formats for inclusion in paper

Nicholas Lee is a consultant ophthalmologist, Western Eye and Hillingdon Hospitals, London; Andrew Millman is an occupational physician, Gloucestershire Royal Hospital, Gloucester; Martin Osborne is a senior registrar, Chase Farm Hospital, Enfield, Middlesex; and John Cox is head of information services, Wellcome Centre for Medical Science, London.

The ABC of Medical Computing is edited by Nicholas Lee (nicklee@online.nildram.co.uk) and Andrew Millman (af89@dial.pipex.com).

Database programs

Flatfile databases:

Cardbox-Plus	0171 925 0636
Claris Filemaker Pro	0181 756 0101
Button file (Shareware)	01297 24088
File Express (Shareware)	01297 24088

Relational databases:

Microsoft Access	Windows	01734 270000
Paradox	DOS and Windows	0800 177277
dBase	DOS and Windows	01734 320022
Superbase	Windows	01344 867100
Lotus Approach	Windows	01784 445808
Alpha Four	DOS and Windows	01752 606881
FoxPro	Windows	01734 270000
DataEase	DOS and Windows	0181 5540582

Bibliographic software:

Reference Manager—Costly but aimed specifically at medical market	01895 813544
EndNote—Strong on generating bibliographies, but aimed at humanities rather than medicine	01865 784800
Pro-Cite—Very powerful but expensive	01865 326612
Papyrus—Good value for money, but onus is on user to develop routines for importing data	01960 351532