Computerisation of histopathology/cytology records—use of a commercial data storage system

SG SUBBUSWAMY, A McCORMICK, EE PETERS

From the Department of Histopathology, St Andrew's Hospital, Billericay, Essex

SUMMARY A system for computerising histopathology/cytology records using a commercial data processing system is described. It is emphasised that any system of computerising records should have some flexibility as no two laboratories have the same requirement and new applications and requirements may arise as one goes along. It is suggested on the basis of the authors' experience over the past two years that it might be worthwhile for histopathologists/cytopathologists interested in computerisation of records to consider the possibility of "user-friendly" software systems, which can be adapted to their particular requirements, rather than a tailor made system, which might be difficult to modify.

There has been considerable interest in the past few years in the use of computers and word processors in histopathology/cytopathology departments, both in generating reports and in data storage and retrieval. This is reflected in a number of reports in the *Journal of Clinical Pathology*¹⁻³ and the *Bulletin of the Royal College of Pathologists*.^{4,5} It is likely that there are pathologists who would like to introduce computers into their departments but are deterred by the problem of devising an adequate system. Departments differ widely in their requirements, and a system that works in one may not necessarily be of use in another. On the other hand, few departments can afford the luxury of a professional systems analyst.

One possible answer to this is the new breed of "user-friendly" software, consisting of systems which can be manipulated easily with relatively little computer expertise. We have been using such a system in the Histopathology Department of St Andrew's Hospital, Billericay, since July 1981. We handle about 9000 histology and 18 000 cytology specimens each year. We have a good manual system for storing records, but we believed that with the numbers we were handling we would soon reach a point where manual systems would become impractical. We have faced no real problem in generating reports and foresee none for some time. Our aim in computerising was therefore primarily to develop a system for data storage and retrieval.

We have used a commercial system called Data-trieve to create our own programs and have found that it offers a remarkable degree of flexibility. The system is a software package developed by the Digital Equipment Corporation operating under the IAS, RSTS/E, RSX-11 RSX-11MPLUS, and VMS systems. It uses simple commands and is designed for easy and rapid file creation, data storage, and retrieval. In a computer with the facility it is invoked by typing DTR. The machine responds with an identification message and a DTR prompt.

We have been using the PDP-11 computer belonging to the North East Thames Regional Health Authority. This is the same computer used by the Department of Morbid Anatomy at The London Hospital.^{1,3} Our system, however, is totally different and has nothing in common with The London Hospital system.

Methods

CREATION OF FILES

A file can be created in a few minutes. The program for our cytology file starts as follows:

```
DTR> DEFINE DOMAIN CYTOSA USING  
CYTREC ON STANDC;
```

```
DTR> DEFINE RECORD CYTREC USING
```

The computer now responds with a DFN prompt, and we can start defining fields and specifying their nature and maximum length.

```
DFN> 01 ALLINF.
```

```
DFN> 01 NAME PIC X (12).
```

DFN) 02 FORENAME PIC X (12).

DFN) 02 SERIALNO PIC 9(5).

We have now created three fields—Name, Forename, and Serialno—the first two being made up of letters or numbers up to a maximum of 12 characters and the last being made up of up to five numbers.

We proceed to define all the fields and finish with:

DFN) ;

This is followed by a DTR prompt. Now the nature of the file must be specified. (The name STANDC has been defined in the first line.)

DTR) DEFINE FILE FOR CYTOSA KEY = NAME (DUP), KEY = DOB (DUP).

The file STANDC has been created with all the fields specified. It is keyed on fields NAME and DOB (Date of birth), for each of which duplicate values are permitted.

OPENING A FILE

This is done by the command READY:

DTR) READY CYTOSA WRITE

This opens the corresponding file for any further operation. Several files can be opened at the same time so that data from one file can be compared with those from another (for example, to compare histology and cytology results).

INPUT

Having opened the file with a READY command one can store by a STORE command:

DTR) STORE CYTOSA

The computer responds with a prompt for each field:

Enter NAME:

Enter FORENAME:

And so on. Obviously a number of records will be entered at each session. A REPEAT command facilitates this.

DTR) REPEAT 100 STORE CYTOSA.

This enables up to 100 records to be entered at one command.

RETRIEVAL

This is done by opening the file, followed by a FIND command.

DTR) FIND CYTOSA WITH NAME = "SMITH".

The computer responds almost instantaneously with something like:

(203 records found).

The search can be narrowed by making the question more specific.

DTR) FIND CYTOSA WITH NAME = "SMITH" AND FORENAME = "JOSEPHINE" (2 records found).

DTR) PRINT ALL.

The system is highly versatile in retrieving records. Two features of particular value are: (i) Search by a given value or a range of values. For example, FIND CYTOSA WITH SERIALNO BT 1000 AND 1999. BT stands for between. FIND CYTOSA WITH FORENAME GE "W". This will pick out all records with forename beginning with W, X, Y, or Z. (ii) Any field or combination of fields can be specified. For example, FIND CYTOSA WITH NAME = "SMITH" AND RESULT GE 3 AND SERIALNO BT 5000 and 10000 AND SOURCE = "GP". (When making a composite request like this it is much quicker to start with a keyed field. A request phrased as above will take a few seconds, whereas if we had started with any of the other fields it might take a few minutes.)

MODIFICATION

To modify a value in a particular field (for example, if one has issued a RESULT of grade II and wants to change it to III), this is done by a FIND followed by a MODIFY statement.

DTR) FIND CYTOSA WITH NAME = "SMITH" AND SERIALNO = 5824

(1 record found).

DTR) PRINT ALL.

The computer prints the record so that one can check that it is the right record.

DTR) MODIFY ALL RESULT

Enter RESULT:

PROCEDURES

Any sequence of commands which is used regularly can be put in a procedure and can be set in motion by a single command. For instance, we usually put in 120 cytology records at one sitting. The process is set in motion by a single command.

DTR) : CYTENTRY.

The computer prompts for each record taking each field in turn. It does this 120 times. These are entered in the main file. All abnormal records (RESULT of 3, 4, and 5) are then automatically entered in a separate "abnormal" file. The computer then prints out all the entries (a breakdown by source, material, and result), the first and last serial numbers of the session, and the cases put in the "abnormal" file. It also prints out an error message if any serial number has been repeated.

Applications and experience

LIMITING FACTORS

Our use of the computer has been limited by two factors largely outside our control: (a) space—we have been arbitrarily allocated 7500 blocks (slightly less than 4 million characters) of storage space; (b)

NAME	FORENAME	DOB	SERIALNO	YEAR	SOURCE	MATERIAL	RESULT	RECALL	ENTRY
83/6073		000000	05075	83	00	GYN	2		0426
83/6077		281052	05160	83	00	GYN	1	783	0426
A	N	000000	05088	83	00	ANF	0		0426
A		240264	05254	83	GP	GYN	2		0426
A	CIA	190557	05187	83	GP	GYN	2		0426
A	R	240553	05243	83	GP	GYN	2		0426
A		170862	05100	83	FPC	GYN	2		0426
A	YN	151251	05080	83	FPC	GYN	2		0426
A		250517	05109	83	GP	GYN	2		0426
A		290748	05329	83	GP	GYN	2		0426
A		071122	05338	83	CS	GYN	2		0426
A		000000	05324	83	ONU	ANF	0		0426
B	CIA	220937	05124	83	GP	GYN	2		0426
B	A	000043	05236	83	QA	ANF	0		0426
B		091256	05291	83	GP	GYN	2	484	0426
B	ELINE	130859	05293	83	FPC	GYN	2		0426
B	AH	041160	05186	83	GP	GYN	2		0426
B	N	120751	05262	83	GP	INF	1		0426
B		030639	05200	83	GP	GYN	2		0426
B		240417	05106	83	GP	GYN	2		0426
B	Y	030157	05276	83	GP	GYN	2		0426
B		000000	05234	83	OB	ANF	0		0426
B	RET	221141	05110	83	GP	GYN	2		0426
B		151037	05211	83	OB	AIA	0		0426
B		120241	05129	83	GP	GYN	1	083	0426
B		040821	05337	83	CS	GYN	2		0426
B		210850	05152	83	GP	GYN	2		0426
B	CE	210104	05118	83	OB	GYN	1		0426
B	EY	291158	05145	83	FPC	GYN	2		0426
B	BETH	080253	05076	83	FPC	GYN	3	783	0426
B	E	220148	05117	83	OB	GYN	2		0426
B	A	240541	05149	83	GP	GYN	2		0426
B		040264	05098	83	FPC	GYN	2		0426

Fig. 1 Top portion of the daily cytology printout showing the various fields (parts of the names and forenames have been blanked out).

distance—we are situated about 30 miles from the computer, so that telephone bills have to be taken into consideration.

We have therefore had to restrict ourselves to using the computer for 2 h a day and to store material on disk for a limited period (after which the information is stored on magtape). In the following paragraphs we shall describe the limited use we have made of the computer, as well as other possible applications for which we have devised programs but which are not practicable at the moment.

PRESENT APPLICATIONS

Data storage

Cytology. Our procedure for cytology record storage has been described above. The daily print out (Fig. 1–3) is quickly checked. Any appreciable error is easily identified and can be corrected. Cytology records are keyed on name and date of birth. A RECALL code is included. This indicates any further action advised. Thus a RECALL code of “883” indicates a patient from whom we should have received a repeat smear by August 1983.

Histology. These are entered in a one line format with SNOP codes. Up to three SNOP entries can be made on one line. If necessary an extra line can be used for more SNOP entries. Histology records are keyed on the patient's last name (the rationale of keying particular fields is discussed later).

Regular retrieval

(a) Monthly statistics in cytology. We use the computer regularly for this exercise, which takes about 20 min (as opposed to about 8 h manually).

(b) Annual histopathology/cytology correlation. As we have been on the computer for only two years, we have done this exercise only once (though we did a trial run using 6 months' material). This was completed in about 4 h, as opposed to several days of checking and rechecking with a manual system.

(c) Monthly recall. We have a procedure by which the computer prints out the details of all patients with a particular RECALL code and then checks and prints out details of all patients with the corresponding date of birth who have had cytology done in the previous few months. This whole process takes

SOURCE		
CS	10	
FPC	28	
GP	114	
IB	11	
IO	18	
QA	2	
QB	55	
QNU	2	
QC	30	
		270
MATERIAL		
AIA	9	
ANF	12	
BCF	8	
GYN	176	
INF	9	
PLF	3	
SPT	39	
UR	1	
VAS	15	
		270
RECALL		
	248	
484	4	
683	1	
783	10	
C08	3	
083	4	
		270

Fig. 2 Part of the daily cytology printout showing breakdown by source, material and recall.

up about 10 min of computer time; another 15 min of manual checking gives us a list of patients who have not had a repeat smear, enabling us to send a reminder to the general practitioner. Comparison of the present system with our previous totally manual system has shown that it takes about 15 min of human time to do what used to be at least a whole afternoon's work and that our results are now far more accurate.

Random retrieval

(a) Individual patients. Looking up previous material from a specific patient using name and forename. At present an inquiry like this takes 1½–2 min, out of which the computer search takes only a few seconds, the rest of the time being that taken to log in and out.

(b) Groups of patients. For example, all thyroids removed in a particular period, broken down by disease; all cytology smears seen in one year, broken down by age, result, and source (general practitioner, family planning clinic, and so on). These are two genuine inquiries made in the last few months. The former took about 15 min while the latter, which needed a little number crunching, took about 1½ h.

FUTURE DEVELOPMENTS

As mentioned above our present use is limited to about 2 h each day. We intend to get a direct line to the computer, which will mean that we will have continuous access at an economical rate. We plan to enter specimens as they are received and keep track

SERIALNO									
THE FIRST SERIALNO= 05071									
SERIALNO									
THE LAST SERIALNO= 05340									
THE FOLLOWING HAVE BEEN ENTERED IN DOMAIN C345									
NAME	FORENAME	DOB	SERIALNO	YEAR	SOURCE	MATERIAL	RESULT	RECALL	ENTRY
C		180156	05072	83	GP	GYN	3	683	0426
B	BETH	090253	05076	83	FPC	GYN	3	783	0426
S		111127	05128	83	GP	GYN	5	C08	0426
H		081158	05143	83	FPC	GYN	4	783	0426
P		000000	05188	83	IO	GYN	4	C08	0426
R	NE	010437	05256	83	GP	GYN	3	083	0426
S	E	290156	05287	83	UB	GYN	4	C08	0426
D									

Fig. 3 Final portion of the daily cytology printout showing records which have been entered in the "abnormal" file.

of them as they are processed. We have developed programs for this in anticipation.

There are other possible applications, which, however, we do not intend to use for reasons indicated.

(a) *Generating reports.* Our present manual system is satisfactory and we see no advantage in computerisation.

(b) *Creating daybooks.* At first sight this seemed to be a good way of using the computer, but when we discussed the matter both technical and secretarial staff expressed misgivings. These were not based on the performance of the computer but on the pattern of receipt of specimens and the dispatch of reports in the department. It appeared that this use of the computer might lead to a delay of a day in sending out some reports and so we decided against it.

We have proved that both these uses are theoretically possible, by trying out programs.

Discussion

The three major questions raised by any new system are: Has it proved of any practical use?, has it justified the cost incurred?, and how does it compare with other systems?

USEFULNESS OF THE SYSTEM

Time saving

(1) *Cytology.* Input of information takes 150 h a year. This is balanced by the time saved in monthly statistics, monthly recall, and annual histopathology/cytology correlation. The time saved in answering random queries is a bonus; in fact, once we have continuous access to the computer the time taken for each query can be further lowered (from 2 min to a few seconds).

(2) *Histology.* Input occupies 75 h a year. This is balanced, however, by the time that would be used for manual SNOP indexing and alphabetical indexing. Thus any time that is saved in answering random queries and looking up previous material can be regarded as time gained. Again, as with cytology, continuous access would mean a considerable further saving in time. One exercise that our secretaries find particularly frustrating and time consuming at the moment is to receive a call regarding a specimen sent within the previous three days. We anticipate that continuous access to the computer with a program to keep track of specimens within the department should be of great value.

Some comment appears necessary regarding keying of fields. When a file is created, specific fields can be keyed. The advantage of keying a field is that an inquiry for a given value in a keyed field is answered almost instantaneously, while a similar inquiry in an

unkeyed field may take a few minutes. The disadvantage of keying a field is that each keyed field takes up extra space. Thus a compromise has to be reached, depending on which fields one is likely to use for retrieving a record. Our histology inquiries are mainly for material from a specific patient, and therefore the only field keyed is the name. The other common inquiry in histology is for all patients with a particular lesion. We could key SNOP codes and get this information in a few seconds; an inquiry of this sort, however, is for a research project or a lecture and does not need instantaneous retrieval. We have therefore not keyed this field.

With cytology we have a somewhat different situation. The number of records is larger, patients get married and change their names, and material is received from special clinics identified by code number only. Boddington⁴ uses NHS numbers as a field. We cannot do this as very few of our request forms have the NHS number filled in. Hospital numbers are not applicable either. We have found the best way of accessing a report is firstly by date of birth and, if more than one record is found, by checking the forename and name. With a file of about 15 000 records we found the highest number of patients with identical date of birth (day, month, and year) was five. There were only two pairs of patients with identical date of birth and name, one pair with identical date of birth and forename, and none with identical date of birth, name, and forename; we have therefore keyed cytology records on name and date of birth. The possibility that when one has accumulated several years' records one might come across two patients with identical names, forenames, and dates of birth cannot be excluded. We do not take any major step such as advising a cone biopsy without looking at copies of previous reports with all the information given on them.

Versatility

With our system we can access records by specific values or a range of values in any combination of fields. The mode of presentation is also of use. We can illustrate this by the two examples cited earlier.

(a) All thyroids surgically removed in a given period. This could be done quite efficiently with our earlier manual indexing system, but it would give a list of accession numbers only; to get any further details we would have to delve into old files, get copies of reports (with the risk that some copies might get misplaced), and analyse the information. Anyone who has attempted this exercise would know how laborious a collection of even 100 records could be. The computer, on the other hand, could not only collect the information but also print it in

various ways (for example, sorting by lesion or age, or both) in about 15 minutes.

(b) Cytology smears seen in a year, sorted by age, source, and result. This is an exercise which we would never even have attempted without the computer.

COST EFFECTIVENESS

Our system is not intended to make secretarial or technical staff redundant. We believe that cost effectiveness lies in being able to access information and sort it more efficiently. In addition, the rapid and regular increase in workload that we have been experiencing indicate that without computerisation we would soon need either to expand our staff or cut down on services. Taking all these into consideration we believe that the money spent on computerisation is justified.

COMPARISON WITH OTHER SYSTEMS

It is impossible to compare the cost effectiveness of different systems. What has to be remembered is that a system which is considered cost effective by one department may be a total waste for another department, whose requirements may be different.

Comparison of the results obtained by different authors is also difficult. We cannot compare our results with those of Briggs *et al*² as their aim has been generating reports and ours is data storage and retrieval. Swettenham *et al*³ have used the same computer as we have for generating reports as well as for storage in a one line format. Having tried their histology programs for a short time we are in a position to say that in terms of time taken to enter or retrieve a record and of space taken by each record our system is at least as efficient as theirs. It would also appear that Swettenham *et al* are not concerned with histopathology-cytology correlation, which we consider to be of great importance.

The point we must emphasise, however, is that our system actually developed as we went along. This is because the Datatrieve system was designed for commerce and industry and therefore had to be flexible to be viable. So despite our inexperience in computer programming we could create our own data base and modify it as the need arose. We found that Datatrieve has the following striking features:

- (1) Ease with which files can be created and modified.
- (2) Rapid sorting facilities.
- (3) Rapid retrieval by one to several fields.
- (4) Ease with which records from different files can be compared.

The workload and requirements of each laboratory are different, and hence no system can be totally transplanted from one laboratory to another. Also, new applications and requirements arise as one goes along, and one should therefore have a system which can be modified as and when necessary.

Our experience indicates that it might be a good idea for any histopathologist/cytologist planning to computerise records to try a good commercial data storage system and develop it for his or her own use, rather than get a tailor made program, which may be difficult to modify. This is a time consuming process and the benefits are not seen for some months. When one considers the long range benefits of a well set up system, however, the time and trouble taken in the initial stages appear worthwhile.

Addendum

Some of the developments suggested in the text have now been implemented. Cytology and histology records are now entered on the day they are received and the computer is used to keep track of the material within the laboratory. This has proved of great value within a very short time.

References

- ¹ Swettenham KV, Nickols C, Berry CL. Computer programs in histopathology record keeping. *J Clin Pathol* 1982;**35**:40-4.
- ² Briggs JC, Ibrahim NBN, Mackintosh I, Norris D. Practical use of a word processor in a histopathology laboratory. *J Clin Pathol* 1982;**35**:151-8.
- ³ Swettenham KV, Nickols C, Berry CL. Computer programs in cytology reporting and record keeping. *J Clin Pathol* 1983;**36**:856-9.
- ⁴ Boddington MM. Computers in histopathology. *Bull Royal Coll Path* 1983;**44**:5.
- ⁵ Byers PD. Computers in histopathology. *Bull Royal Coll Path* 1983;**44**:5.

Requests for reprints to: Dr SG Subbuswamy, Consultant Pathologist, St Andrew's Hospital, Stock Road, Billericay, Essex, England.