12

How safe is your data?

Sue Gordon
(Winchester Museums Service, 75 Hyde Street, Winchester, SO23 7DW)

12.1 Introduction

The increasing technological resources invested in recording, analysing, and publishing archaeological data pose new storage and handling problems and should, perhaps, make us re-examine old ones.

The storage requirements of paper are well known and are, we hope, taken into account when important archaeological information is recorded and stored using this medium. But how well do we look after computer-held data; especially during the time between excavation and its deposition as an archive with a museum or other institution? This process can take years and during that time the records are at risk from a variety of hazards.

Small computer systems in particular face problems which may be easier to cope with in the highly controlled environment demanded by a mainframe computer. Security, for instance, although by no means perfect in most computer suites, is at least taken into consideration when planning, installing, running and maintaining a very expensive and large computer. Although a microcomputer system may not warrant or even need the extensive and often expensive security system required by a VAX the potential for damage caused by unauthorised persons gaining access to a small system should not be underestimated. After all a single microcomputer today can store as much data as many mainframe systems, in their air-conditioned, high security suites, did 10 or 15 years ago.

Microcomputers may be cheap in relation to their bigger brothers and demand less sophisticated working environments but this shouldn't lead microcomputer users to take any less care of their hardware and as a result, the data it holds. In the event of a loss, major or minor, can we at the very least recreate the data from primary records? This may require replacement of equipment as well as time to re-enter data and involve considerable expense. Maintenance contracts and insurance policies are often discounted on the grounds of cost, but how well have the options been thought out?

The results of a survey at the end of 1989 (section 12.6) indicated to me that some institutions have carefully considered the safety of computer held archaeological data but also that many do not attach any great significance to this aspect of their work. I suspect that this lack of interest is due, at least in part, to the attitude that 'it couldn't happen to us.' It may be likely that no loss of data will occur but as long as there is a possibility that a catastrophe could happen then I believe we should at least think about the consequences and if necessary act now rather than after the event when, of course, it will probably be too late. Simply by being aware of the hazards our data faces we should be able to plan and organise data storage and handling in ways that can help to safeguard it.

12.2 Computer-held archaeological records

There is an important distinction to be made between archaeological data entered directly onto computer based media from evidence which will subsequently be destroyed and data entered from primary paper records or artefacts which are to be stored indefinitely. Computer-held data which has been generated from material which no longer exists must be stored and handled with greater care than that for which the original material is still available. However, to say that data which *could* be recreated from the original paper records or artefactual evidence is not as valuable would be a mistake. Such data is always the result of may hours work in the form of the intital data entry and often the result of several stages of manipulation or interpretation. To recreate these secondary records at any stage of the post-excavation process would be time consuming and therefore costly even if all processes in its production had been well documented and the personnel involved were still available to do the work.

Historic documents and works of art are valued highly. Their storage and treatment reflects this. Data gathered from excavation is not so easily identified as a valuable commodity — it is disparate, often requires extensive analysis and interpretation and often does not look particularly interesting to anyone other than the specialist. It is, however, no less important as a record of the past. If we are responsible for recording information from archaeological remains which are subsequently destroyed, it follows that we also have an obligation to protect those records, whether on paper or some other media, in such a way as to enable the maximum retrieval of information from them in the future.

That computerised data is vulnerable seems obvious, but not everyone who uses a microcomputer may be aware of the variety of hazards awaiting the results of their hard work or that this lack of knowledge can be dangerous. By being aware of possible hazards and planning for loss we can reduce the risk of loosing any of this valuable information. Most people do not go through life expecting the worst to happen. To ask archaeologists or any other professional group to consider the hazards to which they, unwittingly, expose the results of their work, could be seen as a criticism of their professionalism. Nevertheless if we accept that the data is important then the risks must be assessed and all possible measures taken to reduce them.

Consider what happens to a piece of archaeological data, for example the record of a layer, between excavation and deposition in a permanent archive. A typical sequence of events might go something like this:

1. The layer is identified, given a unique number and information about it is recorded on paper.

2. The layer is removed to the spoil heap.
3. All or part of the paper record is transferred to computer.
4. All or part of the information held on computer about the layer is used to produce other computer records from which are produced analyses and reports, also held as files on the computer.
5. These reports and analyses may be used to produced a computer-held draft publication which is subsequently printed.
6. Cross-reference lists and indexes to the archive may also be produced on computer.
7. Finally all paper records, computer-held records and printouts are accessioned as part of the site archive by a museum or other institution.

How do we look after this mass of paper and computer records? If we lose one or more links in the information chain can we recreate the others? If the layer is destroyed before it is recorded or the paper record is lost after the layer is destroyed the data is lost forever — once the material is on the spoil heap it is gone for good. If the initial recording is directly onto computer and is lost or corrupted this also means irrecoverable loss unless backup or paper printout exists. At some point in this sequence the original paper records may be photocopied or microfiched for security but unless this is done at an early stage it could be months or even years before the data has some form of security copy.

If *all* the information from the primary paper record is entered onto computer media, at least two copies of the data will exist. However, it is probable that even if the intention is to put the information onto computer in full, in fact some records will only be entered in part either due to lack of space in the computer record template or because of the unsuitability of the information; for example if it is in the form of a sketch. Inevitably the data on computer will not be an exact copy of the original paper record and cannot therefore be regarded as a security copy of it. If the original paper record were lost, reconstruction of it from the computer-held data would be difficult if not impossible. Conversely, if all data except the original paper records were lost, the draft publication could still be arrived at — but how many months or years work would this entail, how much would it cost and who would pay for the work to be done again?

The point is that paper and computer records deriving from the same original data *are* interrelated but not necessarily interchangable. Do not assume that your paper record will back up your computer record or *vice versa*. Consider the security of each separately according to your ability to replace the data in question.

12.3 Hazards

In common with any other information held on computer media, archaeological data is vulnerable to corruption or destruction from a wide variety of hazards. Accidental or malicious damage[1] or destruction can result from any of the following:

12.3.1 Fire

Statistically fire is not a frequent occurrence caused by computer hardware, but computer equipment and media is vulnerable to even a relatively small rise in temperature and in these terms paper offers much safer long-term storage than any magnetic media or microfiche (Fig. 12.1). Equally important, magnetic media will be damaged by the products of combustion, especially those involving plastics which usually form a large part if most computer media and equipment. Also the gases in Halon fire extinguishers, until recently recommended for computer areas, are probably corrosive to computer equipment.

12.3.2 Water damage

Think about the location of your computer equipment and floppy disk storage. What is on the floor above it — toilets or perhaps a roof space with unlagged water pipes which have a habit of freezing in winter? Of course, if you have a fire, water will most likely be involved in puting it out, particularly by the fire brigade. Water sprinkler systems, however, are very effective at stopping fires escalating.

12.3.3 Theft

Microcomputers, particularly portable ones, and their associated equipment, are surprisingly easy to pick up and walk away with, and even today when such equipment is relatively cheap, are still a target for theft. It makes sense to have a good file cataloguing system for floppy disks so that you at least know what you've lost should the odd one go missing as well as being essential in the event of a disaster.

If unauthorised personnel have access to the area where the computer is kept, daily backup is especially advisable. Remember that with time and money you may be able to recover data from a damaged micro but if your computer is stolen the data on it is probably gone for good. A fire and waterproof cabinet in which to store security copies is well worth considering despite the expense and has the added advantage of perhaps deterring theft as well.

12.3.4 Magnetic interference

Close proximity to poorly shielded high voltage electrical equipment can cause disruption of data on magnetic media. Telephones no longer pose a threat, but it is still advisable to think before putting floppy disks on or next to any electrical equipment including computer equipment. If you have to send data through the post keep at least two copies until you're sure it has got there in a readable state.

12.3.5 System or operator error

Computer viruses are widely publicised and details of the various types and their effects can be found in many computer magazine and books. As with most hazards, good security procedures are probably the best defence. However, even the best backup systems can be defeated by those viruses that do not manifest themselves for several months.

[1] A third of all commercial losses in the UK caused by fire are the result of arson.

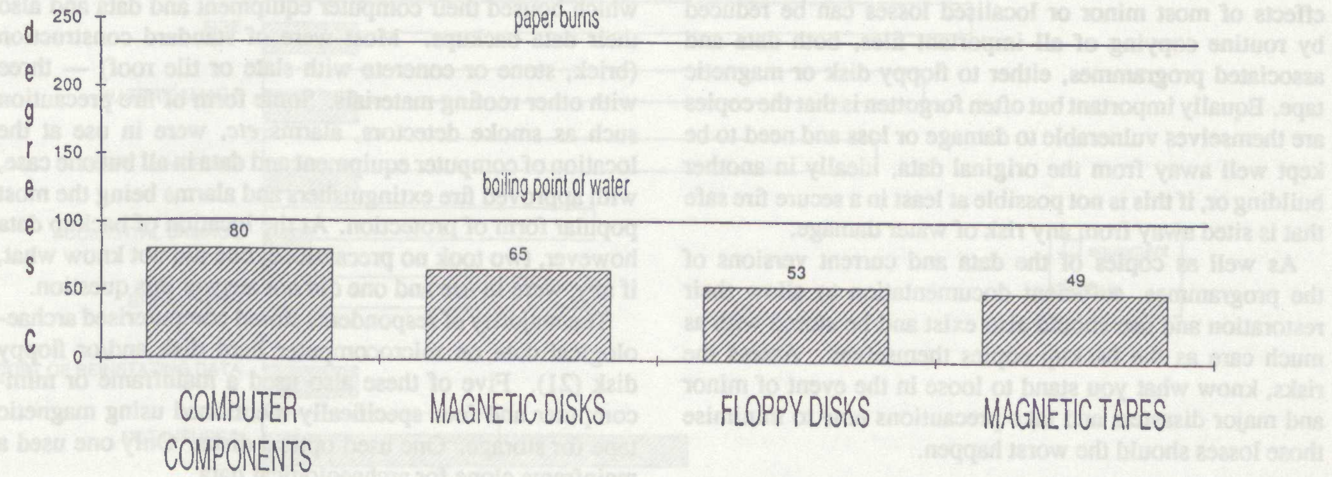

Figure 12.1: Temperature levels at which damage occurs

Even if you know precisely when the virus was introduced into the system and have a security copy prior to that date you will probably still loose many month's work. If you want to avoid the risk of introducing a computer virus into your system, only use software from a reputable source; don't let unathorised persons use the computer and make sure staff know about computer viruses.

Corrupted data can also be caused by system breakdown; for example sudden power failures; and of course, operator error. The second of these is probably the most likely cause of data loss. Accidental erasure of files is unfortunately all too easy on most microcomputer systems but there are several utility programmes that can be used to recover erased files and the cost of these is minimal compared with the charges computer bureaux make to do the same job.

12.3.6 Natural aging

All magnetic media has a limited life beyond which it is unreliable. As the length of time between excavation and final deposition of data in a permanent archive can be many years, long term storage of computerised data *is* a problem facing many archaeological organisations. Although opinions vary regarding actual time periods, magnetic media probably has the shortest expectancy of *reliable* life, followed by CD-ROM, and then laser disks. It is worth noting that silver halide microfiche and archival quality paper both have a proven record of reliable long term data storage.

12.3.7 Obsolescence

Retrieving data from magnetic media requires specialised equipment which one day will probably become obsolete. This is not usually a problem with short term storage but bear in mind that data held on old or unusual equipment may become unusable if the equipment is damaged, lost or wears out. Any data requiring special programmes to read it should be adequately documented and the programmes stored and backed up in the same way as the data.

12.4 Planning for loss — documentation

The question of documentation is an important one, not only because in the event of a disaster the seldom, if ever, used procedures for restoring the data will have to be put into action by staff who may not be familiar with the data and programmes but also because there may be a tendency, especially for organisations with small computer installations, to rely on one person to run the system. What happens if a member of staff with sole responsibility for a computer system, and all the information it contains, is off work for many months — or leaves suddenly? Has he or she left adequate documentation? Have other staff members sufficient knowledge, not only of the microcomputer, but the software and the way the data is organised? A list of names, addresses and phone numbers of anyone you may need to contact in the event of a disaster, or even a simple breakdown is essential and should be stored as hard copy right away from your computer equipment.

The process of drawing up a plan of action to be used in the event of a loss is a worthwhile exercise. It is bound to show up gaps in any existing security procedures and should prompt you to formalise and document such procedures where this has not been done already.

12.5 Summary

I have deliberately avoided the phrase 'disaster planning' as in computer terms it is usually associated with large installations and is geared towards reducing financial loss rather than loss of data. Our aim should be to minimise loss of irreplaceable data through careful planning rather than fostering the attitude that if enough funds are available the *status quo* can be restored after a disaster, especially as, in the archaeological world, it is not common practice to solve problems by throwing large sums of money at them. The effects of most minor or localised losses can be reduced by routine copying of all important files, both data and associated programmes, either to floppy disk or magnetic tape. Equally important but often forgotten is that the copies are themselves vulnerable to damage or loss and need to be kept well away from the original data, ideally in another building or, if this is not possible at least in a secure fire safe that is sited away from any risk of water damage.

As well as copies of the data and current versions of the programmes, sufficient documentation to allow their restoration and use should also exist and be stored with as much care as the backup copies themselves. Assess the risks, know what you stand to loose in the event of minor and major disasters and take precautions *now* to minimise those losses should the worst happen.

12.6 Appendix A: results of a survey, December 1989/January 1990

12.6.1 Introduction

In the course of preparing this paper I felt it would be helpful to conduct a survey enquiring into the security precautions currently taken by archaeological organisations to protect the computerised archaeological data in their care. The questions asked were related to storage conditions, backup procedures and insurance cover.

Of 500 questionnaires sent out by Southampton University with CAA90 and other mailings only 24 were returned (two from overseas).

The IFA survey *Computer Usage in British Archaeology* in 1986 (Richards 1986) includes information from over 100 organisations using computers for archaeological data and, assuming the 'HOW SAFE IS YOUR DATA' questionnaire reached at least half of these organisations a maximum response of 48% can be assumed.

This apparent lack of interest may be due to several factors; the information asked for was too time consuming to search out: the questions asked were ambiguous or otherwise unclear; archaeological organisations or those employees that were able to answer the questions did not receive a questionnaire; the subject was not thought important enough to spend time answering the questionnaire. This last may be significant in showing that a large number of archaeological organisations are not concerned with the safety of computer-held archaeological data.

12.6.2 The results

The questionnaire asked respondents to identify their organisation's type of parent body. Most were universities (7) or local authorities (7). Four independent trusts and two museums also replied. Of the two who did not fit into any of the above categories one reply came from a commission and one from an individual. The two replies from outside the UK have not been included in any of the figures quoted in these results.

The respondents were asked about the kind of buildings which housed their computer equipment and data and also their data backups. Most were of standard construction (brick, stone or concrete with slate or tile roof) — three with other roofing materials. Some form of fire precaution such as smoke detectors, alarms *etc*, were in use at the location of computer equipment and data in all but one case, with approved fire extinguishers and alarms being the most popular form of protection. At the location of backup data however, two took no precautions, one did not know what, if any, were in use and one did not answer this question.

The majority of respondents stored computerised archaeological data on microcomputer hard disk and/or floppy disk (21). Five of these also used a mainframe or minicomputer and four specifically mentioned using magnetic tape for storage. One used optical disks. Only one used a mainframe alone for archaeological data.

Access security procedures to the computer system were in use at system level at thirteen sites although two indicated that this level of security only applied to mainframe data. One had only partial access security and one had it at application level only. Nine had no access security procedures at all apart from locked rooms mentioned by one respondent.

The quantity of computerised archaeological data held by half of the respondents was between 100 and 500mb. The rest held either less than 100mb (8) or more than 500mb (3).

Questions relating to backup procedures for archaeological data were divided into three parts: backup (computer or non-computer) for primary paper records; backup for computerised primary data and backup for computerised secondary data[2].

All but three had some backup for primary paper records but four of these only had backups for selected record groups. Most had a combination of paper, microfiche/film and magnetic media copies. Most had copies of both primary and secondary data, usually on paper and magnetic media but microfiche/film was used, either instead of magnetic media or as well as, in six cases. Two did not answer the question and one said they only had some copies of primary data.

When asked where the copies of computerised records were kept, seven said they were at the same *and* different locations (presumably two copies are kept, one at each location). A further ten kept their copies at a different location only. Five kept copies at the same location but four kept them in a fire safe.

The majority of respondents backed up their data either daily or weekly or both (18) although one of these backed up their mainframe data daily but did not back up micro data at all! One did not answer the question and the rest said frequency of backup varied. Procedures to restore backed up data were tested on a regular basis by only eight respondents. Just over half (12) tested them sometimes and three never tested restore procedures. One did not answer the question.

[2] Primary data — data entered directly from the primary evidence in the trench or from the small finds; data recorded from the primary paper record. Secondary data — data produced from the manipulation of primary data.

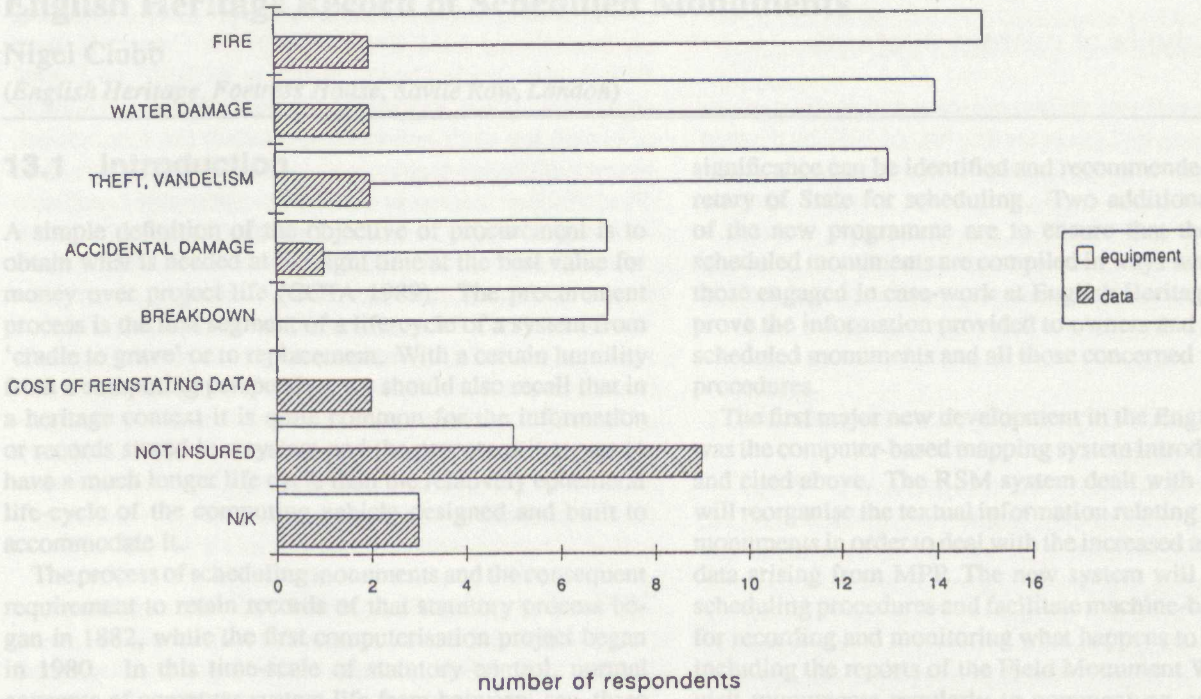

Figure 12.2: Respondents to the questionnaire with computer equipment and data insurance

The majority of respondents had insured their equipment against fire, water damage and theft/vandalism but very few insured their data at all (3) (Fig. 12.2). Only eight had insurance cover for breakdown although all but six had some kind of maintenance contract which may have included breakdown cover.

Acknowledgements

I am grateful for the help and information provided by the following in the course of preparing this paper:

Alexander Stenhouse UK Ltd, Hampshire Archives Trust, Hampshire Fire Brigade, Winchester Museums Service.

Bibliography

GRANT, S. 1986. "Summary and recommendations", *in* Richards, J. D., (ed.), *Computer Usage in British Archaeology*. Institute of Field Archaeologists. Occasional Paper 1.

RICHARDS, J. D., (ed.) 1986. *Computer Usage in British Archaeology*. Institute of Field Archaeologists. Occasional Paper 1.

THORP, J. 1988. "Fire, flood and pestilence: disaster planning", *in* Thorp, J., (ed.), *In Safe Keeping: The preservation of Rare Books and Archives*. South-Western Branch of the Library Association and the Hampshire Archives Trust.

WILLIAMS, B. 1988. "The newer information media and conversation", *in* Thorp, J., (ed.), *In Safe Keeping: The preservation of Rare Books and Archives*. South-Western Branch of the Library Association and the Hampshire Archives Trust.