

CONVERTING SITE NOTES INTO A DATABASE

Philip Crummy

Colchester Archaeological Trust, Camulodunum
12 Lexden Road, Colchester CO3 3NF

Introduction

Large urban excavations like those seen in Colchester in recent years involve thousands of different contexts and several tonnes of pottery sherds, bone fragments and other objects. After the end of the excavations, it is usually a major task to produce lists of contexts ordered in terms of phase. These lists, in conjunction with summary lists of the contexts, are used to produce phase plans for the report and also serve as the primary documentation for the finds specialists when preparing their reports. The site notes are sorted manually on paper and parts copied out several times over and in several formats to produce requisite lists. The specialists also extract parts of the lists supplied to them to compile their own lists geared to their particular category of find. In total, the effort expended in repeating the same information over and over again is substantial and time-consuming.

The use of a database program on a microcomputer is an obvious way to speed up the whole process of preparing excavation reports. Central to this approach are the site records which, for the method to work, must be converted into a suitable database. Once this is achieved, the data can be sorted in the required orders and the desired lists can be produced automatically. Ideally the site records should be put straight into the computer without an orthodox intervening paper record. This would have the following advantages:

- up-to-date summaries of the site notes including phase lists would be available as the excavation progresses, useful in interpretation

- final phase lists would be available very soon after the end of the excavations

- it would provide a printed document which would serve as a neat archive, unlike ordinary site notes

- if the dating evidence could be combined with the site notes as the evidence became available, then the quality of the site interpretation could be improved

- overall the speedy availability of the phase lists would give the report work a substantial head start

- perhaps most important of all, site notes in the form of a database could be directly accessed by the finds specialists and the relevant parts incorporated into their own databases, thereby avoiding the need for the specialist to re-key site data

Software

The database management program chosen at Colchester was INFOSTAR which comes from the same family as WORDSTAR and CALCSTAR. It is a combination of DATASTAR (the data entry program) and REPORTSTAR (output). The latter

includes FORMSORT, a cut-down version of the fast sorting program, SUPERSORT. INFOSTAR is a flexible database program for microcomputers which, although designed for use by non-programmers, has the reputation for being quite hard to use. My own experience with it is that although most of its features are easy to operate there are some processes in its more advanced usages that are tricky to fathom. I am unable to compare it with other database management programs because it is the only program of this kind I know well. The problems I have encountered are due to faults in the manual which is a combination of DATASTAR and REPORTSTAR but lacks a comprehensive overview explaining the principles of the whole package.

The determined novice would probably take a few weeks to master INFOSTAR to the advanced level. They could then expect to spend anything from a few hours to several days setting up each new application. The current Culver Street system took several days to complete whilst our accounting system (see below) was organised in a few hours. The most time-consuming aspect tends to be the testing which any new application needs. If a program such as INFOSTAR is set up by the user, that is the archaeologist not a programmer, one big advantage is that they are then able to modify it as required.

With INFOSTAR it is possible to merge datafiles, sort by any field or any combination of fields up to a maximum of 24, insert or delete fields, write reports to disk for subsequent reworking, include automatically calculated fields, perform low-level statistics and a variety of many other features. It is one of the fastest sorting programs available for microcomputers. It takes about a minute or so to sort a thousand records on one field. Also it incorporates unusually powerful validation and verification features which can be invoked if required. The fields in its datafiles are each written on a new line and separated by commas. This means that INFOSTAR datafiles can be switched to many other database management packages whilst INFOSTAR itself can be used to rework data from other sources. WORDSTAR is a most useful supplement to INFOSTAR since in non-document mode it can be used to edit raw datafiles. Particularly useful in this respect is the 'find & replace' facility of WORDSTAR and the 'column-block copy/move' feature which under certain circumstances provides an easy way to insert new fields into datafiles or remove existing ones.

Transferring orthodox site notes onto a computer: first method

Initially, to test the practicalities of INFOSTAR, a series of suitable data entry and reporting forms was devised to match the standard site recording forms in use at that time. The site in question, Gilberd School, was of average size for a Colchester excavation, about 5000 contexts. Most of the data were entered onto the computer after areas of the site were finished. This had to be the method of working because the site notebooks could never be spared for long enough during the excavation for data entry to be performed concurrently. This experience indicated ways in which the recording system might be modified to make computerisation easier without compromising on content. It also made apparent the desirability of keeping the computerisation of site records up-to-date with the excavation and, if possible, bypassing the paper record to avoid the complications and extra work caused by illegible handwriting. Finally, a major problem related to memory available in a standard 64K 8-bit micro emerged during operation. It became clear that such machines have insufficient RAM to merge large datafiles under INFOSTAR. For this operation under INFOSTAR, a major site would require the larger memory which the 16-bit machines can offer or access to an additional hard disk.

The original recording system used at Gliberd School had three separate series of numbers: find numbers, feature numbers and layer numbers. As a result it needed three separate sets of files for data entry. This would have made it slow to enter data live on site since the operator would have had to keep leaving and entering different files.

A more serious difficulty arose from the number of files associated with features or layers. This could vary from none to several dozen. A fixed number (12) had to be settled on and built into the INFOSTAR feature and layer forms. For those contexts with more than 12 find numbers, the next form had to be used. This was a clumsy solution to an awkward problem. The structure of the forms also meant that file merging according to find number took much longer than necessary.

Revised recording system: second method

To overcome the problems encountered during the testing with the Gliberd School site notes, a revised recording system was instituted in which there was only one sequence of numbers rather than the previous three. The system was essentially the same as described above except that the features and layers did not make two unbroken sequences of numbers starting with 1. When a number is needed as a label for a new feature or layer, the next one in the unified series is used. A special report written to disk rather than the printer is used to correlate the find numbers with the appropriate features and layers. From the common input form by selective suppression of fields according to context type three different context formats are produced: features & layers; finds from features/layers; other finds/contexts.

For each area of the Culver Street site about 30 different files have been set up to cover the various input and output forms required. The principal reports take the form of four summaries. These serve not only to help the site finds specialist provide up-to-date lists relating to pottery, small finds, coins, building materials and clay pipes, but are also used principally in the fourth summary which brings together all the dating evidence in terms of phase.

Direct input via a radio link

There are practical problems which make the input of site records onto a computer direct from the site rather difficult. When our system was being set up, there were no low cost portable computers powerful enough, robust enough and cheap enough to risk using on site. A new generation of portable machines has now appeared which can withstand the rigours of on-site recording. The recording system used in Colchester is predominantly text-based. Conversion into code would compress the volume of data to the extent that a portable computer might be possible but the gains would be far outweighed by the loss of intelligibility in the resultant site notes. Also software would probably have to be specially written and practical problems overcome in transferring the data from a portable to a larger machine for analysis and listing. Moreover, since few if any of the site assistants likely to use an on-site micro have any typing skills, the value of an on-site machine would be substantially reduced. On balance, it was felt best and simplest to contrive to put the data straight onto a fully-fledged micro with disk-drives and an off-the-shelf database management package. This would avoid any programming problems and at the same time a disk-based system would permit convenient interrogation of the database.

As a rather bold, possibly even half-baked, experiment a radio-link was tried as a means of inputting data directly into a mains-powered micro located in a nearby office. The computer operator was a trained audio-typist so that the equipment was used to its full potential rather than in a one-fingered fashion by site assistants. Two types of radio-link have been tried. A British Telecom approved cordless phone permitted convenient, legitimate, two-way conversation as on a normal line but without the expense of connection. The base unit was very useful because it enabled the typist to carry on a two-way conversation with site personnel without having to turn switches or hold a receiver. However, although the cordless phone is claimed to be effective up to 30m, the quality of reception proved very poor except at very short range. CB radio was also tried, with special switches at both ends for ease of use. This was cheaper to set up than the cordless telephone (about £ 75) and provided greatly improved reception. However, it only allows one-way conversation and is subject to breakers, annoying clicks and crackles.

Immediately after each dictation session, the site assistants were given printouts of the new contexts. These printouts became the paper site record and were used to record subsequent changes and additions. Periodically the amendments were entered into the database and, if necessary, new printouts were produced. Temporary labels were used to describe the phases, these were later automatically replaced using the find & replace feature of WORDSTAR when the excavation had finished and the full structure of the site phasing had become apparent.

Despite the test-run with the Gilberd School site, the Culver Street system had the disadvantage that, being based on a new recording system, it was still under development. This meant that minor but solvable problems and ways of improving the system only emerged as the excavation progressed.

The radio systems were in operation for about 3 months and over this period the site database was built up in a fairly satisfactory manner. However, there was always a reluctance to use the system live and a tendency to use rough notes as a script for the 'broadcast'. Unfortunately, the radio-links always proved rather unreliable mainly because the cheapest option was taken each time. They thus tended to push the site assistants towards the paper record which the system was intended to supplant. In previous excavations, it was normal practice to label the finds and write the preliminary site notes in two or more main sessions during the day. At Culver Street, the preference was for find numbers to be allocated sporadically throughout the day with the effect that live broadcasts did not fit very well the work pattern which had already evolved. There is no doubt that the use of the radio-link was dependent upon the confidence and ability of the site assistants to dictate their site notes live.

Input via a preliminary site record form: third method

Rather than pursuing the radio-link by making it more reliable, the preliminary notes made by the site assistants prior to a broadcast were formalised as a special printed form and the radio-link was dispensed with. Each day completed pages of the new form were given to the computer operator. They were then entered into the database and replaced by printouts given to site assistants for recording changes and alterations in the normal way. The result was that, in return for a working regime in which the site assistants felt more comfortable, the problems associated with paper records reappeared. A backlog built up, mistakes and gaps in the printout arose from illegible or poor handwriting and the pressure to keep notes up-to-date was removed.

Future developments

Despite substantial problems, the computerisation of site data at as early a stage as possible in Colchester is clearly here to stay. It will be a regular feature of all future Colchester excavations. When the practice becomes routine and more reliable, finds specialists will be expected to alter their work patterns to take greater advantage of these ready-made databases by designing their own data in such a way that details about contexts and phases can be input directly from the site archive. We shall not try a radio-link again, at least not in the near future, but the dual level of recording which developed from these experiments will be maintained.