

THE 1965-1978 MUCKING EXCAVATION COMPUTER DATABASE

J.P.J. Catton, M.U. Jones & J.C. Moffett

Mucking Post-Excavation, Thurrock Museum, Orsett Road, Grays RM17 5DX

Introduction and background (MUJ)

The Mucking sites were dug at a period of expansion in British excavation. The curator of aerial photography at Cambridge (David Wilson) has described the results as the largest cropmark excavation in the world (Fig. 1). Although there are other extensive and continuous cropmark landscape palimpsests, none seems likely to be excavated on a comparable scale in the foreseeable future. Therefore the Mucking database needs to be considered with present and future applications in mind.

The immediate need concerns the production of the definitive report. Following the recommendations of the Frere Report (Anc. Mon. Board 1975), of increases in printing costs and of changing technology in printing, it is intended that a computer bank - available as printput on paper or on microfiche - will provide the Level III archive. A computer bank should also meet the need for archive security and ease of retrieval.

Simple analysis, making full use of graphical techniques, will provide the foundation for initial synthesis at Frere Level IV. The basic need is to plot all artefacts (whose context is recorded by grid co-ordinates) as dot distributions, to be compared one with another and both with the site plan. To a large extent this treatment will take the place of the vertical stratigraphy which is fundamental to deep urban excavations, but so often lacking in landscape sites.

Future needs can only be suggested. The bank will contain several categories of finds and features which are of key significance. To take two examples: Mucking has the largest assemblage of late Bronze Age settlement ceramics in Britain, including the earliest well authenticated evidence for salt production; it has also the largest sample (more than 200) of Saxon sunken huts in Europe.

The 15ha site plan (Fig. 1) is already in use by John Haigh of Bradford University in connection with research into computer plotting of cropmarks; while Derrick Riley of Sheffield University is applying the Mucking excavation results to his studies of cropmark causation, and their archaeological interpretation. The Mucking site plan could also be the basis for sampling experiments. Different areas could be blocked out to enable comparisons of interpretation to be made between say a 2ha plan of selective and discontinuous excavation; an 8ha plan, which was 50% excavated; and the reality of the 15ha plan which contains areas of which the excavation ratio ranges from nil to over 90%. At this conference John Haigh discussed the need to make archaeological field surveys compatible with excavation results (Haigh, this volume). The Mucking data will provide key material for such a theme.

The Mucking excavation took place in the (archaeologically speaking) pre-computer era. However, the recording system used (grid co-ordinates) links all three main categories of data: field records; plans, sections and notebooks; features and finds; and is entirely suitable for computerisation.

MUCKING, ESSEX 1965-1977 (INTERIM)

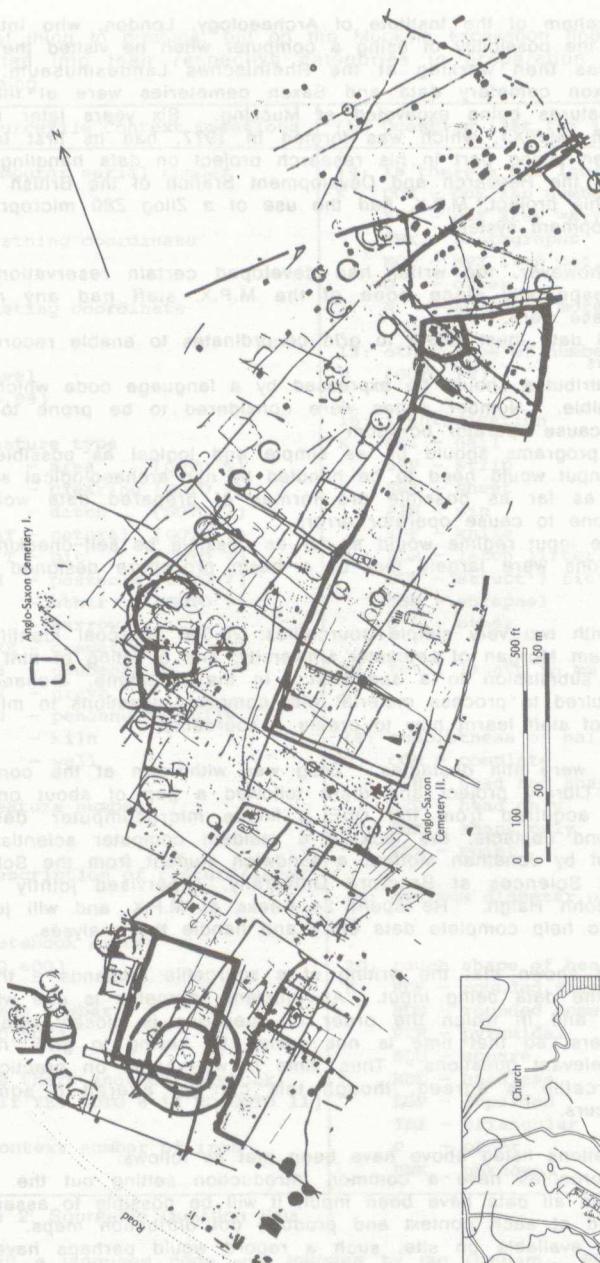
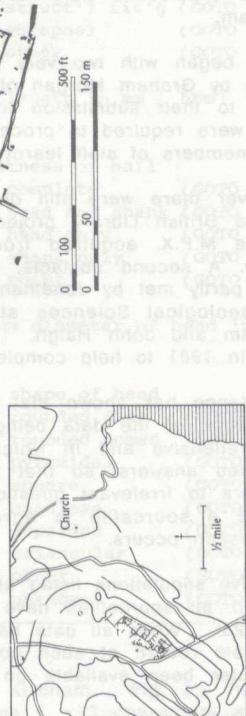


Figure 1: All period plan of the crop mark site at Mucking, Essex.

It was Ian Graham of the Institute of Archaeology, London, who introduced the author to the possibility of using a computer when he visited the site in 1972. He was then working at the Rheinisches Landesmuseum, Bonn, handling Saxon cemetery data and Saxon cemeteries were at that time among the features being excavated at Mucking. Six years later Mucking Post-Excavation (M.P.X.), which was formed in 1977, had its first taste of computers when taking part in his research project on data handling in the humanities, for the Research and Development Branch of the British Library.

As part of this project, M.P.X. had the use of a Zilog Z80 microprocessor software development system.

Since 1972, however, the writer had developed certain reservations and stipulations, especially since none of the M.P.X. staff had any relevant expertise. These were:

(i) that all data must relate to grid co-ordinates to enable recording by dot distributions

(ii) that attributes should be expressed by a language code which would be self-intelligible. Number codes were considered to be prone to errors and liable to cause operator boredom

(iii) that programs should be as simple and logical as possible since much of the input would need to be handled by non-archaeological staff.

(iv) that as far as possible pro formas of prepared data would be avoided as prone to cause operator errors.

(v) that the input regime would as far as possible be self-checking. These stipulations were largely met by a basic procedure designed by Ian Graham.

Input began with two very simple sourcefiles: 20,000 charcoal identifications made by Graham Morgan of Leicester University; and a listing of flint pieces prior to their submission to a specialist. In the meantime, archaeological staff were required to process material with computer questions in mind and two members of staff learnt how to create a sourcefile.

However there were still obstacles. Zilog was withdrawn at the completion of the British Library project and there followed a gap of about one year before M.P.X. acquired from the DoE a Midas microcomputer described below. A second obstacle, the lack of a resident computer scientist, was then partly met by Jonathan Moffett, a sandwich student from the School of Archaeological Sciences at Bradford University, supervised jointly by Ian Graham and John Haigh. He spent 24 weeks at M.P.X. and will join the staff in 1981 to help complete data entry and handle the analyses.

Experience has shown that the writing of a sourcefile demands a thorough knowledge of the data being input. An efficient sourcefile is one which is comprehensive and in which the order of questions is closely related to expected answers, so that time is not wasted by having to give negative answers to irrelevant questions. Thus, time is well spent on practice runs before a sourcefile is agreed, though this can be altered if something unforeseen occurs.

The five stipulations listed above have been met as follows:

(i) all sourcefiles have a common introduction setting out the context (Fig. 2a). When all data have been input, it will be possible to assemble a complete record of each context and produce dot distribution maps. Had a computer been available on site, such a record would perhaps have been

the first thing to produce, but on the Mucking excavation finds were initially separated into their respective categories in preparation for traditional processing.

(a) Sourcefile Context Questions	(b) Question list for iron objects
1: computer serial number S	14: is there a DoE lab/Xray/no. K DOE - DoE lab number (GOTO 13) XRA - DoE Xray =A:? (GOTO 13) PHX - photographs (GOTO 13) MCT - MPX iron cat. (GOTO 13) O - other (GOTO 15) N - no or no more (GOTO 16)
2: northing coordinate N	
3: easting coordinate N	
4: level R (0.99)	15: other type of number T (GOTO 14)
5: feature type K A - area (GOTO 5) CL - clearing (GOTO 5) D - ditch (GOTO 7) NAT - natural (GOTO 7) PIT - pit (GOTO 7) PH - posthole (GOTO 7) O - other (GOTO 7) B - barrow CRE - cremation GH - grubenhaus GR - grave PG - penannular gully K - kiln W - well	16: identification K NAI - nail (GOTO 17) SRP - strip (GOTO 33) SHE - sheet (GOTO 40) PIN - pin (GOTO 43) KNF - knife (GOTO 43) LLR - latch lifter (GOTO 43) STF - struct'l fit'g (GOTO 43) SHN - shrapnel (GOTO 26) O - other (GOTO 43)
6: feature number N	17: total length in mm (5mm) N
7: description of feature/layer T	
8: notebook number R (0,600)	18: completeness of nail K COM - complete (GOTO 21) HSH - head and shank (GOTO 21) HED - head only (GOTO 19) SHK - shank only (GOTO 23) PNT - point (GOTO 23)
9: page number R (0,500)	
10: is there another notebook ZN (if YES GOTO 8 if NO GOTO 11)	19: maximum diameter of head in mm N
11: context number of iron N	20: rough shape of head K RDF - rounded flat (GOTO 26) RDM - rounded domed (GOTO 26) PYR - pyramidal (GOTO 26) SQU - square (GOTO 26) HDL - headless (GOTO 26) LOP - lopsided (GOTO 26) TRI - triangular (GOTO 26) O - other (GOTO 26) UNK - unknown (GOTO 26)

Figure 2: Sourcefile Question lists

(II) a language code was initiated by Ian Graham. Fig. 2b illustrates some of the keywords used. Each keyword uses 1 - 3 characters and has a unique meaning throughout all the sourcefiles. This does not mean that each assistant must learn a complete new language; only the keywords

related to each sourcefile on which they work. If any keyword is forgotten, the relevant list of keywords can be called up immediately on the VDU.

(iii) the decision to use keywords fulfills this stipulation.

(iv) the need to prepare data for input varies with the subject. Two outside specialists have produced data sheets or cards for charcoal and animal bone identifications. In some cases data are written on finds bags beforehand, for instance the measurements of an iron nail. In others, finds are partly prepared by the appropriate archaeological assistant, who calls out the data. Where the data are comparatively complex, as in bags of pottery, it is intended to place a copy of the relevant printout within each finds bag. The contents of the field note books, about 300, and the field plans, about 5000, are simply read out. When these are complete and indexed by computer they will simplify much of the preparation of the textual side of the report.

(v) It is hoped that the problem of errors becoming fossilised within the data base has been met by using two people on each sourcefile, whose work is signed. One is the caller who, with some types of material, must be an archaeological assistant. The other operates the keyboard. Most keyboard operators are part-time workers with typing experience. Both glance at the VDU to check accuracy. A logbook is kept of all input, and to avoid any possibility of loss due to malfunction, a new file is opened for each period of input. In practice these extend for several hours.

The creation of sourcefiles and related problems (JPJC & JCM)

Following discussions with staff processing various classes of data, a list of questions with acceptable answer ranges was produced. The aim was to be comprehensive yet succinct, always bearing future analyses in mind. Thus, for example, one would want to know of an object of fired clay: its category; dimensions; completeness; fabric; colour; and perhaps weight.

The question list is scrutinised to find the most direct route for each possible set of questions, bypassing irrelevant questions. Directions can then be programmed into the sourcefile, and the question and answer list is written onto the floppy disk.

The first group of 11 questions (see Fig. 2a) is standard to all Mucking sourcefiles and deals with the context: grid co-ordinates; level; layer; feature type; notebook and page references. This is followed by the questions concerned specifically with the kind of data being entered, whether it is of finds, features or records (see Fig. 2b).

An important consideration when constructing a sourcefile is deciding which question type to use. Questions requiring a presence/absence response may be given a Yes/No format, but this could lead to problems in reading the data file as in this example:

Question	Yes/No reply	Keyword reply
Does this flint have a cortex?	Y	CTX
Is this flint burnt?	Y	BNT
Type of flint artefact?	FL	FL
Is there another flint in this context?	N	N
Keywords: FL = flake; CTX = cortex; BNT = burnt		

With Yes/No replies the data file here would read: Y Y FL N. Of the four only one FL is self-intelligible. However, by using a keyword question type the positive attributes are at least readable. Yes/No questions are, therefore, best used for basic divisions when directing questions and not for recording specific attributes.

A third type of question allows a free text reply. One such question is included in each cycle of a sourcefile, since it is impossible to foresee every attribute. Although a free text reply can be printed out in full, it is of no use for mechanical analysis, and so it is employed only as a last resort. If the same answer occurs often enough, a fresh keyword can be created.

Once the sourcefile is approved a manual is produced. This gives explicit instructions for each question. Working drawings are included to ensure consistent answers for forms and types.

Data entry is normally carried out by two people, of whom one is the archaeological assistant handling a particular artefact. This assistant calls out the data to the operator. To what degree it is already prepared depends on the material being input.

As entry proceeds, the caller can check the VDU screen for errors. If any occur, they can in most cases be corrected there and then. As each finds bag is computerised, it is endorsed with its unique computer number. A computer log is kept, recording the names of the sourcefiles, how much entry has been completed, time spent and caller and operator names.

So far 14 sourcefiles have been written comprising: ARTEFACT files for animal bone, charcoal, fired clay, flint, iron, pottery (Prehistoric), pottery (Romano-British), slag and iron; FEATURE files for cemeteries (Anglo-Saxon) and cemeteries (Romano-British); RECORD files for notebooks, plans and pre-computer catalogues.

The M.P.X. microcomputer system (JCM)

This microcomputer system is described in MDA Occasional Paper 4 (Stewart 1981) but it has recently been modified. Table 1 gives a description of the system as it is at present. It is so arranged that each of the microprocessors controls one of the VDUs and one of the disk drives, so that in effect there are two systems. They can communicate with each other, but each disk requires all the files that are needed at any one time, unlike a dual disk system. This does have its disadvantages. In particular during program development, but it does mean, it is hoped, that the rate of data-entry can be increased ultimately to twice that of the single system.

Table 1: M.P.X. microcomputer system description

- 2 Z80 microprocessors each with CP/M 1.4 operating system and 64Kbyte memory store
- 2 8" soft sectored, double density disk drives
- 2 VDUs with 24 80 character screens
- 1 Dot matrix printer
- Supply of floppy disks
- Languages: PASCAL
 - FORTRAN
 - Intel 8080 Assembler language
 - Z80 Assembler language

The Data-Entry And Retrieval (D.E.A.R.) Package

Although data-entry has the highest priority at M.P.X., in generating a useful database data retrieval facilities are also required. Further, unless a D.E.A.R. package has already been written for a microcomputer then time must be set aside to develop it. This was the case at M.P.X.. There was no suite of programs available when the original Zilog was in use. A data-entry program was written by Ian Graham, in Z80, but when the Midas system was bought it was decided to modify this. A revised version was written by the author. Development of this package has continued so that it presently incorporates two principal programs:

INPUT - a data-entry program

RETRIEVE - a data-retrieval program, which was developed at Bradford University using the Department of Archaeological Science's R.M.L. 380Z microcomputer.

The two above programs are written in Z80 Assembler language, but it is anticipated that they will be rewritten in PASCAL/Z.

The package as a whole is designed to be flexible and generates a variable record length database. This has made the development of a data-entry program relatively easy, but in order to be able to retrieve data it is necessary to have a map of the structure of the datafile. This is accomplished by using a Question Sourcefile (Q.S.F.). As described above the Q.S.F. is basically a textfile, comprising a list of questions, each of which is identified by a label and has a codeletter indicating the type of answer that will be accepted. Table 2 provides a list of those codeletters used at M.P.X. When any of the programs is used the Q.S.F. is loaded into the memory of the microcomputer and the program continually refers to it. Since each data type, e.g. flint or pottery, has its own Q.S.F. the program initially asks the name of the Q.S.F. to be used. Since the datafiles and Q.S.F.s. of the different data types are kept on separate disks, there is little chance of confusion.

Table 2: Question types used in data-entry

Codeletter	Question type
N	Numerical question
R	Numerical question within a specified range
S	Serial number, automatically incremented by the program
T	Text question, up to 255 characters
K	Keyword question
Y/Z	Program control questions, used to redirect the program to another question

Fig. 3 is a simple flow diagram of the INPUT program. A new datafile is created for each session, as a safeguard against disk writing errors, and the program then asks the questions from the Q.S.F. in a continuous cycle, until the 'end of session' signal (\$) is entered at the keyboard. The program checks the answer received against the information it has from the Q.S.F. and stores the answer on the disk if it is correct. Then it asks the next question. Or, if the answer is incorrect, or contains an error, the program signals an error and asks the same question again.

It is possible to examine data entered during a particular cycle of questions, but they cannot be altered. However, it is possible to erase the present cycle of data, if several errors are found in it. Otherwise, any necessary corrections to the data in the datafile can be carried out using the text editor.

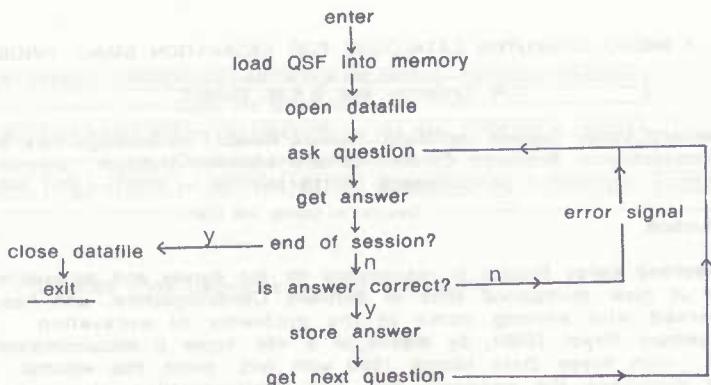


Figure 3: Flow diagram of INPUT.

Conclusions (MUJ)

Any judgement on the computerisation of the Mucking data is clearly limited by the fact that, apart from simple printouts, no actual results are yet available. While, apart from typing a few hundred multi-period bibliographic entries, no word processing has been attempted.

Inevitably one compares computing with traditional processing. There is no doubt that preparing data for computer entry concentrates one's thinking through to the final presentation. Parameters need to be evaluated and selected: what purpose is served, for example, in recording the weight of different kinds of pottery, in addition to, or instead of, the sherd counts?

One suspects that archaeologists will have less need to be articulate in a literary sense when there is a machine which will make the data self-explanatory. Whether this will make archaeological reports more attractive to consult is another matter; although no one with access to a computer will have any excuse for failing to produce a report even if only in printout form.

The aim of a fully computed archive, whether available as printout, as graphics or as microfiche, provides a spur to processing staff, since everyone's work will make a perceptible contribution to any definitive report, and to future research or evaluation. A computer base also, of course, resolves the security angle, since apart from the data bank, copies of printout can be lodged in different places, including the Museums holding the finds. In the case of Mucking, this will be the local, Thurrock and British Museums.

As far as expense is concerned, the capital costs (about £ 6,000) are comparatively slight compared with present day salaries. If aggregate costs should be higher, these must be offset against greater effectiveness of the databank as against manually typed lists and text. Some data handling - which would have been quite impractical with hand labour - becomes quite feasible. By far the most important for the Mucking database will be the production of dot distribution maps for, without a computer, only verbal evaluations would have been practicable. This would have been rather like completing a geophysical survey and attempting to describe the resulting plot of anomalies in words.

ANC. MON. BOARD ENG., COM. RESCUE ARCH. 1975 Principles of publication in Rescue Archaeology. Dept. Envir., London.
 STEWART, J.D. (ed) 1980 Microcomputers in Archaeology. MDA Occ. Paper 4. Duxford, Cambs.