

PIT SEDIMENTS AT DANEBURY AND THE COMPUTER.

M.L. Shackley.

Department of Archaeology
University of Southampton.

J.D. Wilcock.

Department of Computing
North Staffordshire Polytechnic.Abstract

The paper describes the application of computerised information retrieval to the analysis of pit sediments from the Iron Age hillfort at Danebury near Stockbridge, Hampshire. This multivallate structure is nearly 13 acres in area, and present excavations indicate that the total excavation of the site will yield at least 5,000 pits for study. It would be difficult to contemplate the analysis of such a large number of pits, each of which may have several distinct layers, without the help of a computer. Data was recorded in the field direct onto computer coding sheets, then punched on cards and transcribed as card images onto magnetic tape in the Computer Centre. The database management features of the PLUTARCH System are described in detail, with examples of pit sediments data.

1. The site at Danebury

Danebury is a medium-sized Iron Age hillfort near Stockbridge, Hampshire, which was densely occupied from the 5th century BC until the Roman Conquest. The site has been excavated over short seasons from 1969 to 1973 (Cunliffe, 1971). A recent proposal for total excavation is scheduled to start in 1974, lasting for at least five years, as part of a major research project operating in the area (Bowen and Cunliffe, 1973).

Inside the defensive earthworks and massive gates the settlement consisted of rows of timber-built houses, arranged in a regular plan, varying in constructional details with time. These are now evidenced by postholes and gully complexes dug into the chalk, accompanied by large numbers of pits, very variable in size and shape. There is no surface stratigraphy inside the fort and the vast majority of the rich archaeological and faunal remains come from the pits. The storage and retrieval of information from these pit sediments forms the subject of this paper.

2. Types of pits encountered

At present over 500 pits have been excavated, although the postulated total excavation of the site is estimated to yield at least 5,000 more. The majority of the pits were dug initially for grain storage (Bowen and Wood, 1968; Reynolds, P.J., 1967), although some had more specialised functions such as burial or clay-puddling. After their abandonment the pits were either filled with rubbish and occupation material, or were left to silt naturally with chalk weathering products. Occasionally these two processes are combined, a naturally-silted pit being "topped-off" with rubbish, and the surface then levelled. Each pit contains up to 15 distinct depositional horizons, identified by differences in the sedimentation pattern, with large quantities of included archaeological material.

A computer-based system is obviously required for the adequate storage of all this information, and a pilot scheme on this premise was put into operation during the 1973 excavation season. The sedimentological and field recording aspects of the project were controlled by the writer, faunal evidence submitted by Mrs. A. Grant, who was also present during the excavation, and additional archaeological data added by the excavation staff.

3. Data recording in the field

Wherever possible, data was recorded in the field directly onto computer coding sheets, with the object of minimising data loss in transference. In practice this worked well for environmental information, but the need to develop simple coding schemes for features such as pit shape meant that in the initial stages of the project some data needed to be added later. Since these codings are now established it is hoped that feature recording will be much faster, and that the computer scheme will form a much more integral part of the archaeological recording methods than the "optional extra" approach that was taken for the experiment.

At least 20 cards were used to record general features about each pit, including its dimensions, base angle, wall slope and contents, in terms of type of fill and artefacts. An additional 16 cards were then used for each layer within the pit, to record much more detailed information about the sediments, bones and artefacts. The handling system needed to be very flexible, since additional data such as the full pottery dating sequence will be added later, and the recovery criteria may also vary. The computer graphics of the PLUTARCH System will be used in producing histograms of various parameters, as well as in plotting maps of the development of the site at different periods, since the location of each pit is recorded as a six-figure site grid reference tied into the National Grid. Individual pits and layers have unique identification numbers which are punched on the last eight columns of each card.

4. Methodology

A. Introduction:

Computerised information retrieval has now been used for a number of years by archaeologists and museum workers. The museum aspect of this work will not be covered in detail in this paper. A number of workers have used remote terminals for the two-way transmission of archaeological data between a site and a computer (see for example Buckland and Wilcock 1973). This paper will confine itself to the problems of creating archaeological data files on punched cards and magnetic tape, and the construction of Boolean expressions for the searching of these files.

B. Summary of previous work:

Gardin (1962) was the first worker to suggest standard formats for the recording of archaeological data on punched cards. It soon became evident that the computer was necessary for processing archaeological data files and that magnetic tapes or discs were more efficient storage media.

Morrison (1968) described a flexible system for storing data on magnetic discs, and retrieving and indexing information. The information was recorded in alphanumeric, with minimum restriction on the format of text and keys, and the system was designed so as not to require the rearrangement of text or index as new material was added. The flexible handling of unrestricted alphanumeric strings is of prime importance in archaeology.

O'Connor, Voigt and Wyuker (1968) described an archaeological information retrieval system with a logical search algorithm employing combination of descriptors (it was not specified in this paper what Boolean operators were available in the system). Conventional archaeological terms were used, allowing cards to be punched direct from field records without intermediate coding, making interpretation easy.

Rusco (1968) described the computer storage and retrieval system set up by the Nevada Archaeological Survey since 1966. Numeric characters were punched on cards, then processed and sorted by each descriptor category for a complete printout, for use as an index. An electromechanical card sorter was used to derive statistics.

A novel language for the setting up of Boolean expressions was described by Reynolds, C.F. (1969, 1970 a-e). CODIL (Context-Dependent Information Language) has variable-length data items of the form NAME=SMITH, YEAR=1945 and DISEASE NOT MEASLES, i.e. each has a set name, an operator and a value. Data items may be nested to several levels, and there is a facility for repeating common leading items (this is important for archaeological data which will have common leading items such as site name, year of excavation, box number, layer number, etc.). During retrieval the CODIL system has a facts statement which is compared with all other statements (called the criteria file). The search is carried out at a speed of about 3,000 lines per second. Logical operators available are NOT (monadic), =, >, >=, < and <=. ANY and NONE are used to denote the complete set and the empty set. The diadic operator OR (inclusive) is implicit in a sequence of items at the same level, AND is implicit in connections between different levels, while OR ABSENT is used to handle situations where the information is undefined. PRINT, TERMINATE and RUN are examples of action statements. In the 1971e reference below an example of the use of CODIL in handling cave fauna records is given.

Data structuring for an archaeological computer file was described by Rogers (1970), using an English-like language TAXIR (TAXonomic Information Retrieval System). A typical query in this language would be phrased as:

```
QUERY, Please list items with genus, Manihot
AND state of collection, Arizona OR New Mexico
AND year of collection, FROM 1949 TO 1968
AND NOT collector, D.J. Rogers.
```

The use of magnetic tapes to record and retrieve prehistoric carvings was described by Walker (1970a & b) and Wilcock (1970a). This system uses an earlier version of the file structure of

the PLUTARCH System described in this paper, where each punched card (recorded on magnetic tape as a card image) has a record identifier, a type designator and a number of fields of information which are free-format alphanumeric, numeric or Boolean variables within predefined field areas. Computer-drawn distribution maps are a by-product of the retrieval runs. The same system used for cave records was described by Wilcock (1970b).

Cuisenier (1970/71) described a hierarchical keyword system for information retrieval for use in museums.

A remote-terminal language for the administration of a data base was described by Hebditch (1972). The syntax is based on the following commands:

find print alter new move why end.

The command find locates a record, print outputs the record, alter changes certain parts of a record, new creates a whole new record, move copies items from one group of records to another, why clarifies error messages and end causes termination. Equivalent facilities to the find, print, alter, new and end commands are available in the system described by Buckland and Wilcock (1973).

Finally, initial remarks about the system described in this paper to retrieve archaeological sediment information were given by Shackley (1973).

C. Methodology of the PLUTARCH retrieval system

The file structure of the PLUTARCH System consists of a serial file, each record of which has a unique record number and consists of a set of card images. Each card image has a type designator (unique within the record) in columns 2 and 3, and the record number in columns 73-80. Column 1 is reserved for a repetition factor, allowing up to ten cards of each type within a single record. Columns 11-70 are divided up into predefined fields, each field being a multiple of ten characters in length and starting in column 11, 21, 31, 41, 51 or 61. This layout facilitates both the design of coding sheets and the checking of records when listed on the line printer or tabulator. The design of a data base for any specific archaeological application commences with the designation of the card type codes and the definition of the starting columns and field lengths for all fields within each type of card. This design function is left entirely to the archaeologist concerned, although he may of course consult a computer scientist in the process. The format is flexible in that new types of cards can be specified at any time and inserted into the system; also each record contains only those types of cards for which data is available (i.e. there are no blank cards) and the various types of card may be in any order within the record.

The fields may be alphanumeric, numerical or Boolean. The alphanumeric fields allow the archaeologist to write full plain-language words or phrases if he wishes, or to insert codes of his own design for shorthand use. It is an interesting point that archaeologists often begin by asking for the plain-language facility to help interpretation of the records, but quickly devise shorthand codes when faced with the task of writing out

all their records on data sheets. If supplied with a list of codes and their plain-language equivalents the computer can provide expanded printouts. The words and phrases are completely free-format within the predefined field areas.

The numeric fields allow the insertion of integers or reals (number containing a decimal point) for the specification of quantitative variables. Only one number is allowed per field, but it may start anywhere within the field.

The Boolean fields are used for presence/absence or true/false indications. If the field is completely blank or contains only a decimal 0 the variable is taken to be false or absent; otherwise it is taken to be true or present.

During data validation and information retrieval it is necessary for the values of specified fields to be checked to see if they are within the range of allowed values (in the case of data validation) or if the required criteria occur (in the case of retrieval). The alphanumeric check consists of a search of the whole field for the occurrence in any position of a specified string of characters of any length, from a single character up to the maximum field length. If the sequence is detected the search variable is set true, otherwise it is set false. The field on the card image may contain many more characters than the specified string. Thus if the field contains the word POTSHERDS the search will respond true for the string SHERD, and it will also respond true for the string POT, but it will respond false for the string POTTERY.

The numerical check consists of a scanning of the field for a single number and the conversion of this number from decimal digits (and decimal point where applicable) to a variable in the usual computer form. The variable is then compared with a specified number using a specified operator from the set

> ≥ < ≤ = ≠

If the comparison succeeds (e.g. if the variable is greater than the specified number when operator > is specified) the search variable is set true, otherwise it is set false.

The Boolean check consists of a scanning of the field for the first character. If the field is completely blank the search variable is set false. If a character is found, and if it is decimal 0 then the search variable is set false, otherwise it is set true.

The specification of the search criteria is performed via the teletype at run time. Each criterion consists of a type designator, a field start column, a length of field count, an operator (A for alphanumeric, B for Boolean, or one of >, ≤, <, ≥, = or ≠ for numerical comparisons), and a value (e.g. BONE for alphanumeric, 123.4 for numeric, true for Boolean). As it is specified each criterion is allocated a serial number, starting at 1, and this serial number appears as a light button on the line-drawing display unit, together with the parameters of the criterion.

When all criteria have been specified light buttons also appear for the Boolean diadic operators AND (.) and OR (v), the monadic operator NOT (-), and the multiple operator "n things out of m" (n/m).

The operator then assembles the Boolean expression by which the criteria are to be joined using the light buttons for the criteria and for the operators. As he does so an instruction list is generated for use during retrieval of records or data validation, and the interim form of the expression is displayed. A RESET light button is also available in case the operator makes a mistake, and a RUN light button initiates the retrieval run or data input with validation, when the operator is satisfied.

During the retrieval run the generated instruction list is used to operate a Boolean array working as a nesting store. A nesting store is a stack which operates on a "first in, last out" basis, rather analogous to the insertion and ejection of cartridges in a rifle magazine. The variables are evaluated by searching each record before the nesting store operation begins. Then the instruction list causes the required variables to be brought into the top cell of the nesting store, other cells nesting in, or causes a Boolean operation to be carried out on the top cell or cells of the nest; the diadic operators AND and OR operate on the top two cells, nest out the remaining cells by one cell and place the result in the top cell; the monadic operator NOT just reverses the sense of the top cell; and the operator "n things out of m" sets a counter to zero, then nests out m-1 times, counting the number of true cells, and finally replaces the value in the top cell by true if the counter is greater than or equal to n, or by false otherwise. When the instruction list is exhausted, the top cell contains the result of the whole Boolean expression; a true value is the signal for the acceptance of the record (in the case of data validation) or for the printout, storage, etc., of the record (in the case of retrieval), since it satisfies the criteria. A counter is allocated to each instruction in the list, which is increased by one if the top cell holds the value true at the completion of the instruction. At termination of the run, the counters are output, giving an indication of the number of times each part of the expression has been satisfied by the scanned records. This is particularly useful if few records are retrieved, for it indicates how the criteria may be relaxed to give a higher recall. However, there is a disadvantage in this counting procedure, necessitating as it does the complete execution of the Boolean expression every time; in some instances because of the values of the variables it is possible to determine at an early stage of the execution that the final Boolean value must be false, without the necessity for completion of the execution. In these cases premature exit would be possible, but this is precluded if the counting procedure is to be carried out in all cases. Perhaps in future the choice could be offered the archaeologist whether to carry out the full evaluation with its informative statistics at a relatively slow speed, or to allow premature exit at increased efficiency with no statistics.

In addition to printing out certain fields or the whole of a successful record, the program may be instructed via the teletype to store specified fields for future processing. In this way a series of grid references may be stored for distribution map plotting, quantitative data may be listed for the production of histograms and piecharts, or statistics may be recorded for use in clustering or multidimensional scaling runs.

The PLUTARCH System, of which the database management features described above are only a part, is described elsewhere in this publication. It can be seen that in addition to providing a

database management system, the PLUTARCH system is designed to process and display the retrieved data using statistics and graphics, and its major end products are publishable diagrams summarising the retrieved data.

5. Anticipated Results

The success of the coding and recording scheme for the 1973 season is most encouraging, since from now on excavation and computer-based recording will proceed hand-in-hand. The development of the coding system and the checking of the pilot listing were the most tedious parts of the process (for an example of sediment data see Figure 1). Where possible automatic data validation (specified by the archaeologist) is to be used in the future. Apart from providing a means of data storage the PLUTARCH System will be used to analyse data and provide information on the detailed evolution of the site. The sedimentological implications have been fully discussed elsewhere (Shackley, 1975 forthcoming). The system will be used to test hypotheses concerning pit shape and size differences with time, together with correlations between sedimentation patterns, faunal remains and the archaeological evidence. This can be done either between pits or between sediment horizons within individual pits. It is hoped to comment on such matters as the number of pits open at any one time, the rate of silt accumulation, evidence for seasonal abandonment of pits, the relationship between pit size, shape and function and the differential preservation of material in sediments of different types. The project will yield information vital to the proper interpretation of such a large and complex site, and will provide a model for future integrated recording of environmental and archaeological evidence.

References:

- | | |
|---|--|
| Bowen, H.C. &
Cunliffe, B.W.
1973 | The evolution of the landscape.
Antiq. Jnl. 53(1), 9-13. |
| Bowen, H.C. &
Wood, P.
1968 | Experimental storage of corn underground and
its implications for Iron Age settlement.
Bull. Univ. London Inst. Arch. 7, 1-14. |
| Buckland, P. &
Wilcock, J.D.
1973 | An experiment in the use of a computer for
on-site recording of finds.
Computer Applications in Archaeology 1.
Science and Archaeology 9, 22-24. |
| Cuisenier, J.
1970/71. | Feasibility of using a data processing system
in the Musee des Arts et Traditions
Populaires, Paris.
(UNESCO) Museum 23(1), 33-36
(Museums and Computers). |
| Cunliffe, B.W.
1971 | Danebury, Hampshire: First interim report
on the excavations 1969-70.
Antiq. Jnl. 51, 240-253. |
| Gardin, J.C.
1962 | Cartes perforées et ordinateurs au service
de l'archéologie.
La Nature 3331, 335-357. |

- Hebditch, D.L.
1972 A simple language for a data-base.
Computer Bulletin 16(1), 36-40.
- Morrison, D.R.
1968 PATRICIA - Practical Algorithm to Retrieve
Information Coded in Alphanumeric.
Journal A.C.M. 15(4), 514-534.
- O'Connor, D.
Voigt, M.M. &
Wyuker, E.
1968 Information retrieval for archaeological
sites.
MASCA Newsletter 4(1), 2-3.
- Reynolds, C.F.
1969 An introduction to CODIL.
ICL., London.
- Reynolds, C.F.
1970 Meeting on CODIL
Advanced Programming Study Group.
Computer Bulletin 14(7), 244-245.
- Reynolds, C.F.
1971 a) CODIL, Part 1. The importance of flexibility.
Computer Journal 14(3), 217-220.
b) CODIL, Part 2. The CODIL language and its
interpreter.
Computer Journal 14(4), 327-332
c) CODIL - a new concept in information languages.
Computer Weekly, May 13, 1971, 6.
d) An interactive CODIL interpreter.
CODIL News 4 (Dec.1971), 1-4.
e) Handling cave fauna records on a computer.
Trans. Cave Res. Grp G.B. 13(3), 160-165.
- Reynolds, P.J.
1967 Experiments in Iron Age agriculture.
Trans. Bristol and Gloucs Arch. Soc. 86, 68-73.
- Rogers, D.J.
1970 Theoretical and practical considerations on
data structuring for a computerised
information retrieval system.
in Gardin, J.C. (ed.), 1970.
Archeologie et calculateurs 145-159.
Editions du CNRS, Paris.
- Rusco, M.
1968 Index for the Nevada archaeological survey
archives.
MASCA Newsletter 4(1), 2.
- Shackley, M.L.
1973 Computers and sediment analysis in archaeology.
Computer Applications in Archaeology 1.
Science and Archaeology 9, 29-30.
- Shackley, M.L.
1975 The Danebury project: an experiment in site
sediment recording.
(forthcoming) Chapter 4 in Davidson, D. and Shackley, M.L. (eds.)
1975.
Sediments in archaeology.
Duckworth, London.
- Walker, M.J.
1970 a) An analysis of British petroglyphs.
Science and Archaeology 2/3, 30-61.
b) An analysis of British petroglyphs (further
remarks)
Science and Archaeology 4, 26-29 (N4/1)
- Wilcock, J.D.
1970 a) Petroglyphs by computer
Science and Archaeology 2/3, 27-29.
b) Information retrieval for cave records.
Trans. Cave Res. Grp G.B. 12(2), 96-98.

FIGURE 1

SAMPLE PRINTOUT OF SEDIMENTS DATA
 MASTER CARDS COMMON TO ALL PITS OR LAYERS, TO BE INSERTED FOLLOWING
 THE CARD WITH TYPE DESIGNATOR GIVEN IN COLUMNS 7-8

01	04	DANFEBURY			STOCKRIDGE		
02	04	HAMPSHIRE	ENGLAND			BRITISH ISLES	
03	04	SU323377			CUNLIFFE	8.73	
21	25	DANFEBURY			STOCKRIDGE		
22	25	HAMPSHIRE	ENGLAND			BRITISH ISLES	
23	25	SU323377			CUNLIFFE	8.73	
34	25	MLS	8.73				
04		385 PIT			8.73		385
05		SMALL ERODED BEHIVE					385
06		IRREG FLATTISH BASE					385
07		UNREMARKABLE					385
09		6R	14	0.602			385
10		5	5	0			385
11		385.1	385.5	8.73	J984790		385
12		80N	85S	1.2			385
13		3.1	01R	0			385
14		1	1	0			385
15		ROOT DISTURRANCE					385

MASTER CARD PECULIAR TO ONE PIT AND ITS ASSOCIATED LAYERS, TO BE
 INSERTED FOLLOWING THE CARD WITH TYPE DESIGNATOR GIVEN IN COLUMNS 7-8

24	25	385			2			
25		1		1	1	7		385.1
27		0	13	5	0	0	0	385.1
28		5v6.2	0	0	0	0	0	385.1
29		1	0	0	0	0	0	385.1
32		1	1.2	1000	0000	0	0	385.1
33		0	0					385.1
35		NO LAB TESTS						385.1
36		LIGHT GRAY SOIL WITH CHALK						385.1
37		H984790						385.1
25		1		2	2	10		385.2
27		1	5	5	0	0	3	385.2
28		0	0	0	0	0	0	385.2
29		1	0	2	1	0	0	385.2
32		1	0	0000	0000	0	0	385.2
33		0	0					385.2
35		DELIBERATE CHALK FILL						385.2
36		H984790						385.2
25		1		3	3	11		385.3
27		3	34	1R	0	0	0	385.3
28		7.5v7.2	0	0	0	0	1	385.3
29		1	0	1	0	0	0	385.3
32		0	1.2	1000	0000	0	0	385.3
33		0	0					385.3
35		GREY SILTY WASH						385.3
36		H984790						385.3
25		1		4	4	3		385.4
27		4	52	3	28E	59W	3	385.4
28		0	0	0	0	0	0	385.4
32		0	0	0000	0000	0	0	385.4
33		0	0					385.4
35		PRIMARY CHALK SHATTER						385.4
36		H984990						385.4
25		1		5	5	1		385.5
27		1	6	1	0	0	0	385.5
28		2.5v7.2	0	0	0	1	1	385.5
29		2	0	0	0	0	0	385.5
32		0	0	0000	0000	0	0	385.5
33		0	0					385.5
35		GREY PRIMARY SILT						385.5
36		H984790						385.5
