THE COMPUTING OF DANEBURY.

Gary Lock,
Research Centre for Computer Archaeology,
North Staffordshire Polytechnic,
Stafford.

1. Introduction.

The excavation of Danebury Iron Age hillfort, Hampshire, is a long term project during which the total excavation of the interior is planned. The area revealed by the first ten years of excavation, about ¼ of the interior, has contained a high density of features and artifactual evidence (Cunliffe.B.W. 1976) including 947 rock cut pits and over 100,000 pieces of animal bone. This makes Danebury important not only within Iron Age studies but also from a methodological perspective. With such large amounts of data even the simplest of analyses would be impossible by hand thus justifying the extra resources necessary for computerisation. This paper is intended as a review of the problems and achievments experienced during the computing of Danebury. What has become obvious, however, is that computerisation must be an integral part of large scale excavation which is given due consideration right from the initial costing and planning stage. The computing specialist must become as regular a member of the future post-excavation team as the environmentalist has over recent years.

The computing of Danebury has been an on-going project supporting one full time worker since 1978, it is currently financed by The Danebury Trust. Work is carried out at the above establishment on an ICL 2966 running under VME 2900.

2. Information collection and storage.

The preparation and entry into a computer of large amounts of data is an arduous task. Danebury has shown just how time consuming, expensive and error prone this operation can be when not incorporated into the overall excavation strategy from the beginning. The three stages of transcriptior ie; site records, formated data sheets and punch cards could, by following current trends, be reduced to just site records and then machine entry via a keyboard with the use of a micro. Any micro used, however, would have to be compatible with the mainframe as the processing of, for example, the 5 megabyte animal bone file on a micro would be extremely tedious. The use of rigidly formated data collection sheets or data collection using a micro on site has been rejected mainly due to pressures of time existing during the annual one monthly excavation and restrictions imposed by existing data collection sheets. More time is needed for the data from the excavation and basic post-excavational analyses, which are collected by human observation and decision making to be organised into the consistent and unambiguous manner that make them machine compatible. The Danebury data sheets are not, therefore, for primary site recording but for alrea processed data. Like most other archaeological data those from Danebury consist of a mixture of quantitative and qualitative units. While the use of a computer forces the archaeologist to organise and record data in a standardised logical way this can give a misleading impression of objectivity and validity which may be shielding fundamental problems in the recording system used. Computers do not solve problems of standardisation, for example between the subjective valuations of different workers on site which when computerised are no more objective than when initially entered in the site recording book.

2.1. File organisation and structure.

The term 'data base' is much abused in archaeology and will not be used here to describe the Danebury files. True data bases are, however, becoming more commonplace in archaeology (Gaines.S.W. and Gaines.W.M. 1980) as in other fields where large amounts of information are stored and retrieved. A data base is essentially automatic. The user need not be aware of the efficient internal organisation of the data but is simply asked questions and presented with answers in natural language usually on a VDU. Data bases are also capable of continual expansion with new data being entered as a series of answers to questions which are then automatically verified and stored in the appropriate place. Flexibility is another essential attribute. Data bases should not be application specific but open to interrogation through any level of the data with retrieved results capable of ranging from the general to the specific.

The Danebury data forms a series of files in a simple two-level hierarchy. The main file contains information on individual pits as well as a series of sub-file indicators which control access to the sub-files. This allows any file to be used in a 'stand alone' capacity for sequential processing or for any of the sub-files to be accessed randomly using the index key (Pit Number) from the main file. At the time of writing six sub-files exist containing information on animal bones, pottery, small finds, briquetage querns and post hole co-ordinates. The total storage space used is 10 megabytes with the animal bone file accounting for half.

The data structure of each file was designed in conjunction with the specialist worker concerned to contain the maximum amount of information with free text where needed

Data within each file is structured in the standard hierarchical way of each file being a collection of logically related records with each record being a collection of elementary data items grouped together to describe one thing. Each elementary data item is the smallest piece of information relevant to the current process whether input, retrieval or analysis. These are, therefore, variable and equivalent to the term 'field'. The records in all Danebury files are stored sequentially by context number. A detailed description of the structure of each file appears in Appendix 1.

3. Information retrieval.

Information can be retrieved from each file at different levels:

- The file in full as it stands.
- Selected complete records.
- Selected fields from all records.
- Selected fields from selected records.

Information retrieval is here taken to include basic descriptive statistics such as range, percentages, mean and standard deviation as well as the ability to perform nested sorts on key data items and single or multivariate listings with counts. Each retrieval program is designed in conjunction with the specialist worker concerned who has the advantage of being familiar with the data structure because of their involvment in its design. This makes the Danebury computing system application or, in this case, site specific. While this is not necessarily seen as a good thing in view of current attempts to standardise archaeological recording and retrieval systems (Department of the Environment. 1982.) it does have certain advantages. A higher level of recall and precision is attained by these programs than through an automatic data base. Recall being the proportion of relevant material actually retrieved in answer to a search request and precision being the proportion of retrieved material that is actually relevant.

COBOL is used for all Danebury information retrieval programs because of its strengths in handling strings of text. It also has a high powered sort facility enabling nested sorts on specific key data items by one or two simple lines of programming. Once a COBOL program is written for information retrieval, which is a lengthy process initially, it can be easily altered to perform different retrieval exercises.

3.1. Phasing.

Contexts have been phased by observer imposed phasing according to the pottery typology as described in Cunliffe.B.W. (Forthcoming.). Individual phase ranges for different forms, fabrics and surface finishes were added by program to every record within the pottery file along with an overall context phase range. This is logically established from the highest of the low range for all the forms, fabrics and surface finishes within the context and the lowest of the high range providing the latter is not lower than the former. The lowest of the context range is then taken as being the preferred ceramic phase (CP) for the context which in reality could, obviously, belong to that or any higher phase. Phasing information was then disseminated to workers with the following computed for each context;

- The preferred CP.
- The total number of sherds.
- The number of sherds of the preferred CP as a count and as a percentage of the context total.

From this information all other files have been phased and it is possible for workers to judge the reliability of the preferred CP for any context.

3.2. Processing individual files.

The following summarises all information retrieval exercises carried out so far on the Danebury data;

3.2.1. Pottery.

For all CPs

- Total sherd and weight counts.
- Each context type sorted by phase.
- Pits sorted by sherd count.
- Counts of each fabric type (9) within each form (50) produced as a 50x9 table.
- A listing and count of pits containing pots of specified forms.

For CP1-3 pits

- The percentage of decorated sherds per pit.
- The number and percentage of sherds per pit with specified surface finishes plus the overall highest, lowest and average percentages.

For CP7 pits

- The percentage of decorated sherds per pit.

For CP8 pits

- A listing of pits with sherd counts.

3.2.2. Briquetage.

For CP1-3, CP4-5, CP6, CP7-8 and all phases

- The total weight of each fabric.
- The percentage of each form occuring in each fabric.
- The number of sherds of each form.
- The total weight of each form.

- The percentage of each fabric occuring in each form.

3.2.3. Pits.

For CP1-3, CP4-5, CP6, CP6-8, CP7-8 and all phases
- Pits arranged into shape types.
- The total number of pits.
- The number excavated.
- The number of pits of each shape type and sub-type.
- The total volume of all pits.
- The total volume of all pits of each shape type.
- The total number of pits with tool marks.
- The number of pits of each shape type with tool marks.

For each of the three main pit types, ie. beehive, cylindrical and sub-rectangular
- Histograms of depth, base and volume dimensions for each of the following;
    - All pits with grain layers.
    - All pits with animal bones.
    - All pits with human bones.
- The total number of pits.
- The number of pits with artificial fills.
- The number of pits with natural fills.
- A histogram of the volumes.

3.2.4. Animal bones.

The programs written to analyse the animal bones are being modified to form a package that will be available for use on bones from other sites. Output is in the form of tables and listings with distribution maps of the excavated area where appropriate. The programs are based on the manual methods of analysis used by the specialist worker concerned, see Grant.A. (1975 and Forthcoming.) for further details.

Programs may be used on all bones, bones from any particular feature and/or phase and are summarised below
- The sorting of bones into bone types within each species.
- A summary of the bone content of each feature for quick reference.
- The species represented. Four different methods are used, two versions of the 'epiphyses only' method and two versions of the 'total fragments' method.
- The bone types represented.
- Age details.
- Metrical analyses.
- Butchery.
- Gnawing.
- Burning.
- Erosion.
- Disease.
- Articulation.

3.3. The site archive.

The primary data files and some of the files generated by basic processing will form part of the eventual Danebury archive. At the time of writing back-up copies of all files are held on magnetic tape at The Computer Centre, North Sta1 Polytechnic.

4. Specific statistical applications.

Specific computer and statistical applications have helped to elucidate three areas of analysis where traditiona methods were either not feasible or seemed to be unsatisfac1 due to the large amounts of data involved.

4.1. The seriation of the pits.

The 947 pits recorded during the first ten years of excavation at Danebury represent approximatly 500 years of occupation. Due to the lack of stratigraphy in the interior of the fort these had to be relatively dated, this was achieved by seriating the pits according to the six types of pottery fabric found within them. The whole process is described in detail elsewhere (Lock.G.R. Forthcoming.) and here summarised.

The first stage of the seriation technique consists of performing average link cluster analysis on the 947 pits (cases) using the percentages of each fabric type within the pit as the six variables (University of California.1977.). At an amalgamation distance of 12·3 there occurred 24 clusters (henceforth called Pit Types) which appeared to be intuitively sensible after experimentation. The second stage of the process is the manual ordering of the Pit Types using the standard seriation graph technique (Ford.J.A. 1962.). This produced the classic battleship curve as shown in Figure 1. This seriated sequence was then used to test the relative chronology of the Ceramic Phases and the three Site Phases (Early = CP1-3, Middle = CP4,5 and Late = CP6,7 and 8) which are based on the pottery typology. This also acts as a check on the seriated sequence itself, a further one being provided by the excavated evidence of intercutting pits.

As the seriated sequence was validated by these checks a similar two-stage procedure was used to seriate all pits of the Late Phase to investigate pottery changes in detail. Five sub-phases were established which in total represent a period of 150 years. This enabled a very detailed analysis of changes in pottery form, surface finish and decoration to be carried out within temporal limits rarely before enjoyed within a prehistoric context.

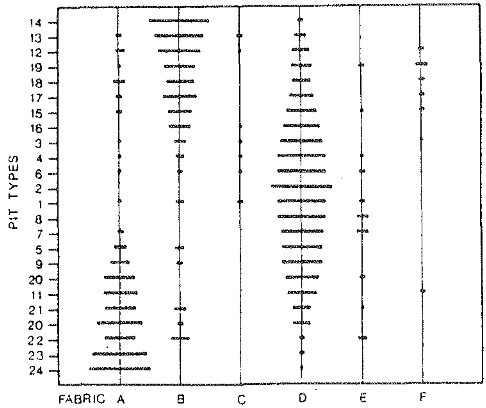

Figure 1. The seriated sequence.

4.2. Pattern perception within the post hole distribution.

The density of post holes within certain areas of the interior is such that searching for 4 and 6 post built structures by eye seems far too subjective. Programs have been developed to identify every occurence of rectangular post built structures within a given tolerance. These are based on programs described elsewhere (Fletcher.M. and Lock. G.R. 1981.) which were developed to identify possible round houses.

The results of the programs have been compared with;

- The results from five randomly generated post hole distributions of similar density and clustering.
- The predictions of statistical theory.
- The analysis 'by eye' which appears in the Danebury report (Poole.C. Forthcoming.).

This establishes the statistical significance of the results by identifying how many would be expected to occur by chance. The analysis then moves on to test the following hypotheses;

- That 4 and 6 post built structures are usually approximately square.
- That the dimensions of 4 and 6 post built structures are usually between 2 and 4 metres.

Both are supported by the results, a fuller account of which is in preparation (Fletcher.M. and Lock.G.R. Forthcoming.).

4.3. Probability sampling of the pits.

Work is in progress to try and establish what percentage of pits need to be excavated to recover an acceptable level of information, this is obviously important as it will affect future excavation. Different probability sampling strategies are being performed on the distribution of 947 pits with those selected then subjected to analyses already performec on the total population. The analyses being used are those concerned with pottery, animal bones and pit morphology and contents. Results are showing that different sizes of sample need to be taken according to the type of data in question.

Appendix 1.

List of data categories.

Main file.

General:
- Year of excavation.
- Feature number.
- Grid reference.
- Relationships.

Pit data:
- Layers.
- Shape type.
- Depth.
- Base dimension.
- Volume.
- Fill type.
- Phase.
- Tool marks.

Presence/absence indicators:
- Pottery on pit base.
- Animal bone on pit base.
- Animal joints.
- Complete carcasses.
- Human skull.
- Complete human skeleton.
- Human joint.
- Isolated human bone.
- Pottery.
- Sediments.
- Molluscs.
- Seeds.
- Charcoal.
- Quern.
- Briquetage.
- Daub.
- Slag.
- Pot boiler.
- Whetstone.
- Loom weight.
- Weaving comb.
- Bone object.
- Iron object.

Copper alloy object.
Stone.
Glass.
Amber.
Gold.
Coin.

<u>Sub-files.</u>

Pottery.
Context number.
Form.
Fabric.
Surface decoration.
Number of rim sherds.
Number of body sherds.
Total weight.
Sherd unique numbers.

Briquetage.
Context number.
Form.
Fabric.
Number of rim sherds.
Number of body sherds.
Total weight.
Sherd unique numbers.

Animal bones.
Species.
Number of bones.
Bone type.
Left or right.
Proximal or distal.
Fused or unfused.
State of shaft.
Articulation.
Butchery.
Gnawed.

Measurements:
1. Total length.
2. Proximal width.
3. Proximal width anterior/posterior.
4. Distal width.
5. Distal width anterior/posterior.
6. Shaft width.
7. Miscellaneous.

Burnt.
Eroded.
Diseased.
Miscellaneous.
Feature type.
Feature number.
Pit layer number.
Recognition number.

Small finds.
- Small find number.
- Context number.
- Layer number.
- Material category.
- Object code.
- Complete or fragmentary.
- Decoration.
- Weight.
- Fabric (clay).
- Number of pieces.
- Number of faced pieces.
- Total weight (stone).

Querns.
- Sample number.
- Small find number.
- Context.
- Stone type.
- Weight.
- Weight of small find.
- Type of object.
- Diameter.
- Number of faced fragments.

References.

Cunliffe.B.W. 1976 — Danebury, Hampshire: second interim report on the excavations 1971-5. ANTIQUARIES JOURNAL. 56. 198-216.

Cunliffe.B.W. Forthcoming. — An Iron Age hillfort in Hampshire. COUNCIL FOR BRITISH ARCHAEOLOGY RESEARCH REPORT. VOLUMES 1 AND 2.

Department of the Environment. 1982. — The publication of archaeological excavations. REPORT OF A JOINT WORKING PARTY OF THE COUNCIL FOR BRITISH ARCHAEOLOGY AND THE DEPARTMENT OF THE ENVIRONMENT.

Fletcher.M. and Lock.G.R. 1981. — Computerised pattern perception within post hole distributions. PROCEEDINGS: Xth UNION INTERNACIONAL DE CIENCIAS PREHISTORICAS Y PROTO-HISTORICAS. 63-78. Mexico.

Fletcher.M. and Lock.G.R. Forthcoming. — A statistical analysis of post built structures at Danebury hillfort. OXFORD JOURNAL OF ARCHAEOLOGY.

Ford.J.A. 1962	A quantitative method for deriving cultural chronology. TECHNICAL MANUAL 1, DEPARTMENT OF SOCIAL AFFAIRS, PAN AMERICAN UNION, WASHINGTON.
Gaines.S.W. and Gaines.W.M. 1980.	Future trends in computer applications. AMERICAN ANTIQUITY. 45. 462-470.
Grant.A. 1975.	The animal bones. in Cunliffe.B.W. Excavations at Portchester Castle. Volume 1: Roman. REPORTS OF THE RESEARCH COMMITTEE OF THE SOCIETY OF ANTIQUARIES. NO.XXXII.
Grant.A. Forthcoming.	Animal husbandry in Cunliffe.B.W. Forthcoming.
Lock.G.R. Forthcoming.	The seriation of the pits. The sub-phasing of the Late Phase. both in Cunliffe.B.W. Forthcoming.
Poole.C. Forthcoming.	Rectangular post built structures. in Cunliffe.B.W. Forthcoming.
University of California. 1977.	BMDP.Biomedical computer programs P-Series. USA.