Grimley, Brendan J. &
Haigh, John G.B. 1982

'A General Purpose Data Management System for Archaeologists'. University of Bradford.

The importance of adequate documentation of archaeological data has been recognised for many years, as has the use of pre-printed recording sheets, (for instance, Harris 1979,pp25,27,111; DoE 1975,p3; Gardin 1980,p32). Often selected artifacts are illustrated, and plans and photographs produced to complement text; but it is the written record which conveys the relevant information to the archaeologist, e.g. stratigraphical relationships and pottery typologies. The information itself is of value when access can be gained to large collections of data, which is where computers fulfil a particular need: a more efficient method of providing information to users.

Past research has shown that computers are quite capable of handling archaeological data, (Newell and Vroomans 1972; Wilcock 1973). What is still a subject of some debate, however, is whether or not to use large machines, (so-called mainframe computers), minicomputers or microcomputers as archaeological data handlers. This paper introduces the work currently being undertaken at Bradford University on the potential of microcomputers as data management systems.

One of the most important specifications for the design of the system was flexibility: that the data manipulation programs could be used for several archaeological applications without any alteration. This entails keeping the details of the data completely independent of any operations performed on that data.

Consider two typical basic requests for information from a data set:

(i) locate all sites containing fabric type T1 pottery,
(ii) list all features cut by feature W28.

In each statement, some relation, (e.g. "containing" and "cut by"), links items of data, (e.g. "sites" and "fabric type"). At the simplest level then, the request takes the form "obtain 'elements' bearing relation to 'elements'". In order to comply with the request, the same fundamental operations can be employed upon data items from different applications, though the actual operations are independent of the details of the data. Thus, if the system application programs contain only data manipulation routines, then provided details of individual data sets are made available to the system, the the operations can be performed. From this, a two stage process is apparent; (a) data definition stage, (b) data manipulation stage; implying that for each archaeological application, if the archaeologist defines the data set and makes the definition available to the system, then the data manipulation can be satisfactorily handled.

In order to facilitate the data definition, time has been spent examining the structure of data on pre-printed recording sheets, which are designed to provide the user with a recognisable form of the data. The data structures can be best understood by looking at the logical view of the data.

Consider a very simple type of record card (Figure 1) based on a more complex version used for recording lithic implements. It can be seen that the blank card, (i.e. without entries), logically comprises groups of related sections, e.g. in Figure 1 the X and Y co-ordinates and level all define the location of the object, i.e. the data items are related. Each record card for separate applications can contain one or more such groups, and for identification purposes, each card and each group can be named, (e.g. Class and Use of Figure 1 represent two named groups). Each group can contain any number of related sections, called fields, which are identified by field-names. Data are entered into the named fields, and as such each data item can be uniquely identified, (e.g. in Figure 1, X co-ordinate (field) of location (group) within lithic (card)).

In order to represent the logical view, some database terminology is introduced, which defines the structure of the data definition, not, strictly speaking, the structure of the completed record at this stage. The term 'schema' represents the entire structure of a data set, which is effectively a record card; and a 'subschema' is a subset of the data, i.e. a 'group' in the above example. The logical view is such that any schema can contain a number of subschemas, and each subschema can contain any number of fields, e.g. Figure 2. This structure refers solely to the logical form and does not govern the actual physical representation in any way. (Figure 3 is an example of the card of Figure 1.)

Application programs have been designed to guide the user through establishing the data definition, which is then stored as a dictionary. The physical structure, or internal representation, is quite different from the logical structure and is handled transparently. This allows the archaeologist to concentrate on the details of the data set.

Each dictionary, (or schema in the logical view), contains information for the computer on all sections requiring entries, and the types and permitted subsets of entries. Each subschema is identified by a name and can be one of two types, (a) 'U'-type, unique, or (b) 'R'-type, repeating, allowing inclusion of repeating observations in the record, (e.g. the same details of wear analysis from several locations on the same artifact).

The fields can be one of four types:

(i) 'N'-type, numeric entry
(ii) 'S'-type, string or keyword entry
(iii) 'A'-type, alphanumeric entry
(iv) 'V'-type, variable entry

A domain of permitted entries exists for each of the 'N'-,'S'-and 'A'-type fields.

Application programs exist to enable the user to display the dictionary at the terminal or printer in an intelligible format, (useful if any abbreviations occur in the domains). Figure 4 illustrates two subschemas from an existing dictionary with fields and associated domains.

Data capture routines enable the user to interactively enter data at the console. The routines interrogate the dictionary and prompt entries for each field within each subschema. Each entry is then checked against the domain and the user is informed if the value is not permitted. At any stage during the data capture it is possible to update the dictionary to include unforseen entries, or expand numerical ranges.

Records are compiled internally and stored on the available secondary store. No internal coding takes place, which ensures that completed data sets can be made available to other users and interpreted without a comprehensive knowledge of the dictionary.

References.

DoE 1975. 'Principles of Publication in Rescue Archaeology'. Department of Environment.

Gardin, J.C 1980 'Archaeological Constructs: An Aspect of Theoretical Archaeology'. Cambridge University Press.

Harris, E. 1979 'Principles of Archaeological Stratigraphy'. Academic Press.

Newell, R.R. & Vroomans, A.P.J. 1972 'Automatic Artifact Registration and Systems for Archaeological Analysis with the P1100 Computer: A Mesolithic Test Case'. Anthropological Publications.

Wilcock, J.D. 1973 'The Use of Remote Terminals for Archaeological Site Records'. SCIENCE AND ARCHAEOLOGY 9: p25.

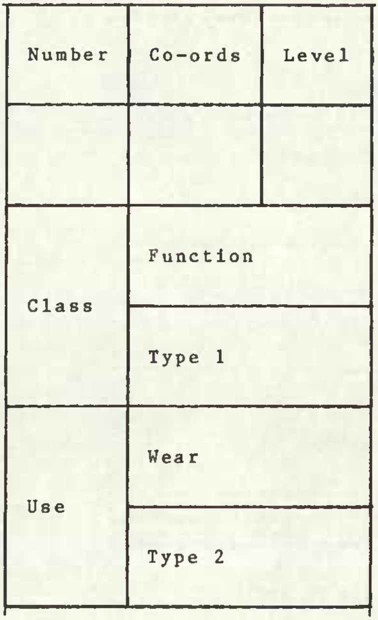

Figure 1. A simple record card
for recording lithic implements.

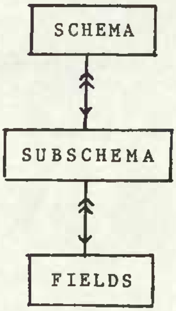

Figure2. Representationof
a record card.
<-->> = a one to many relationship

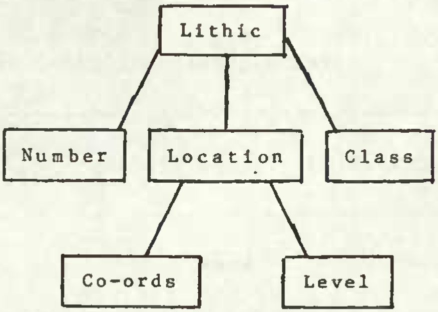

Figure3. Representation
of Figure 1.

```
=======================================

s.s.n.   <LOCATION>   type U

---------------------------------------

<X-COORD> type N
          from 200 to 300

---------------------------------------

<Y-COORD>  type N
           from 600 to 700

---------------------------------------

<LEVEL>    type N
           from 0 to 100

---------------------------------------
=======================================
=======================================

s.s.n. <MATERIAL> type U

---------------------------------------

<MATERIAL>   type S
          FLINT,  OTHER

---------------------------------------

<QUALITY>    type S
      VITREOUS, GRANULAR INCS, GRANULAR

---------------------------------------

<CONDITION>   type V

---------------------------------------

<CORTEX>      type V

---------------------------------------
=======================================
```

Figure 4. Section of existing dictionary.