21

Browsing through the stratigraphic record

N. S. Ryan*

21.1 Introduction

The production and use of a stratigraphic diagram are well established as important elements in current excavation and post-excavation method. Although there is some variation, particularly in the recording and representation of destruction interfaces and feature cuts, most excavators follow the approach developed by Harris at Winchester during the 1970's (Harris 1979). The diagram is usually built up on a day-by-day basis during the course of excavation and, as new contexts are added, the logical structure of the recorded information may be checked to ensure that there are no loops or disconnected sequences. Any errors in recording can hopefully be corrected before the evidence has been destroyed or, at least, before the details have faded from memory.

Both during and after excavation the formal abstract representation of the site provided by the stratigraphic diagram acts as a valuable key to understanding the physical structure and the relative chronological relationships between excavated materials. Indeed it can function as a form of graphical index to excavated contexts, and can be seen as a significant element in a manual information retrieval system. When considering the material from any single context, the diagram provides an immediate indication of stratigraphically related contexts that might be examined for similar material. The analysis may also proceed in the opposite direction. The extent to which a particular artefact class is concentrated or scattered in the sequence can be examined by superimposing the provenance of all examples on the diagram. Other more formal uses of the sequence information have been described, including the isolation of the longest and shortest paths from surface to sub-soil (Cooper 1987). All of these tasks are facilitated by the use of a suitable diagram but each requires extensive checking which may be both time-consuming and prone to error.

21.2 Automated sequence checking

Several computer-based methods for checking and examining the stratigraphic sequences from excavations have been described but, although a number of working systems have been demonstrated, they have not been widely accepted in excavation or post excavation work. Various reasons for this failure have been suggested ranging

* Computing Laboratory,
University of Kent at Canterbury
Canterbury
Kent CT2 7NF

from an inability to handle the large volumes of data generated on many sites, to a poor match between the functionality of the programs and the needs of the archaeologists.

Surveying the literature on computer-based stratigraphic analysis it is apparent that little has changed since the publication of Wilcock's STRATA program (Wilcock 1975). Certainly there have been discussions of the algorithms employed and improvements in their efficiency (Haigh 1986, Cooper 1987, Ryan 1986, Ryan 1988), but there have been few attempts to examine the needs of the archaeologists and thus to define the range of functions that should be provided by such software (Haigh 1986, Cooper 1987). Some attempts have been made at the automatic production of diagrams using relatively unsophisticated equipment (Rains 1984), but the principal concerns have been the checking of the logical consistency and integrity of the recorded data, and the extraction of sequence information.

The majority of published accounts describe programs that have been conceived as batch processes in which all of the required data is gathered together and fed into a program at one time. The output is normally presented as lists of errors resulting from illogical sequences, lists of links that are necessary to the sequence and of those that can be excluded according to what Harris termed the 'Law of Stratigraphic Succession' (Harris 1979, p. 125). Often these lists have been accompanied by a printed tabular layout showing distinct vertical chains of sequential layers. In order to produce a diagram it has normally been necessary to draw links between layers by hand. Only rarely has the computer been used to produce all elements of a diagram, and even then the sequences represented have invariably been so simple that it is not clear that the problems of diagram production have been adequately addressed.

The use of an automated system to check the consistency of a large and complete set of relationships after the completion of an excavation is clearly beneficial. Even if by that time it is too late to correct all of the errors that might be detected, at least they can be prevented from distorting any subsequent analysis. In contrast, manual checking of the complete set of stratigraphic relationships from even a quite small site is rarely likely to prove satisfactory. There is, of course, no reason why a batch oriented system should not be employed on a regular basis during the process of excavation in order to check those relationships recorded so far. However, this 'black box' approach was roundly condemned by Harris who argued that excavators should not be encouraged to rely on such a system and that the incremental manual construction and checking of the sequence diagram was essential to gain a clear understanding of the progress of the excavation (Harris 1975, p. 33).

Harris' criticism was valid because fully automated systems can often deny the user the opportunity to exercise important professional skills and so to gain understanding from contact with the data. However, the benefits which may be realised with manual systems are invariably accompanied by the need to perform many tedious and repetitive tasks. For example, a major problem in constructing any complex diagram where the content, and thus appropriate layout, of the remainder of the diagram remains unknown, is the need for frequent redrawing in order to maintain clarity and consistency. Fortunately, since Harris voiced this criticism in 1975, there have been considerably developments in methods and styles of computing and the differences between automated and manual systems are no longer so clear cut. With modern interactive techniques it has become possible to develop systems that attempt to combine the benefits and to overcome the disadvantages of both manual and earlier batch systems.

21.3 Interactive diagram production

In an earlier paper an outline design for an interactive graphical tool for the production and manipulation of stratigraphic diagrams was described (Ryan 1986). This was based on *gtree*, a program developed for a similar, but equally specialised application, the production of genealogical diagrams (Ryan 1985b, Ryan 1985a). In both cases the underlying problem can be conceived as one of the construction and representation of a directed graph. A graph is a mathematical structure widely used in many areas of computing science. It consists simply of nodes (vertices) connected by lines (edges). A directed graph is one in which each edge has an assigned direction.

At this level the content and meaning of the nodes and edges is undefined. In the case of stratigraphy, the nodes would represent the individual excavated contexts which are connected by edges representing stratigraphic relationships. The direction of edges would be arbitrarily chosen as either up or down. Unlike conventional abstract representations of graphs, the layout of both stratigraphic and genealogical diagrams is constrained by the normal practice within the application area. For example, both encode a form of relative chronological component in the vertical ordering of nodes, and a range of special symbols may be used to indicate different types of node.

Experience with *gtree* suggested that the major limitations of the program were largely a product of its specialised design. It had been developed to deal with a particular class of data gathered from a single source and was insufficiently flexible in its capabilities. In order to deal effectively with a wide range of genealogical data from sources including both contemporary fieldwork and historical documents a more generalised tool was required. Given sufficient generalisation it should be possible to deal equally effectively with a wide variety of diagram production and manipulation applications including database conceptual models, program call-graphs, genealogical diagrams and archaeological sequence diagrams. It was in part to examine how effective a single tool might be for dealing with such a range of problems that *gnet* was developed.

The basis of *gnet* is a general purpose graph browser, an interactive program that can lay out a graph from data specifying nodes and edges. Diagrams can be created interactively or from data held in files. Nodes may be positioned manually or by the application of an automatic layout algorithm. The algorithm employed is based on one developed for *gtree* and is essentially similar to that described in (Robins 1987). As there can be no definitively correct layout for these diagrams it is limited to the production of a logically correct and, hopefully, intelligible layout. Thereafter the user may edit the diagram to provide a representation that adequately conveys the required information. Nodes and links may be moved, added or deleted as required simply by selecting the required menu function and diagram elements with the mouse. When nodes are moved or deleted the program ensures that the diagram remains consistent by moving or deleting associated links as necessary. Once a satisfactory configuration has been achieved the state of the diagram may be saved either in a form that can be re-used in a later session or as, for example, PostScript instructions for a laser printer.

As such, the program is similar to several other graph browsers that have been described in the computing science literature in recent years (for example, (Rowe *et al.* 1987, Robins 1987). However, rather than displaying abstract graphs, *gnet* is designed

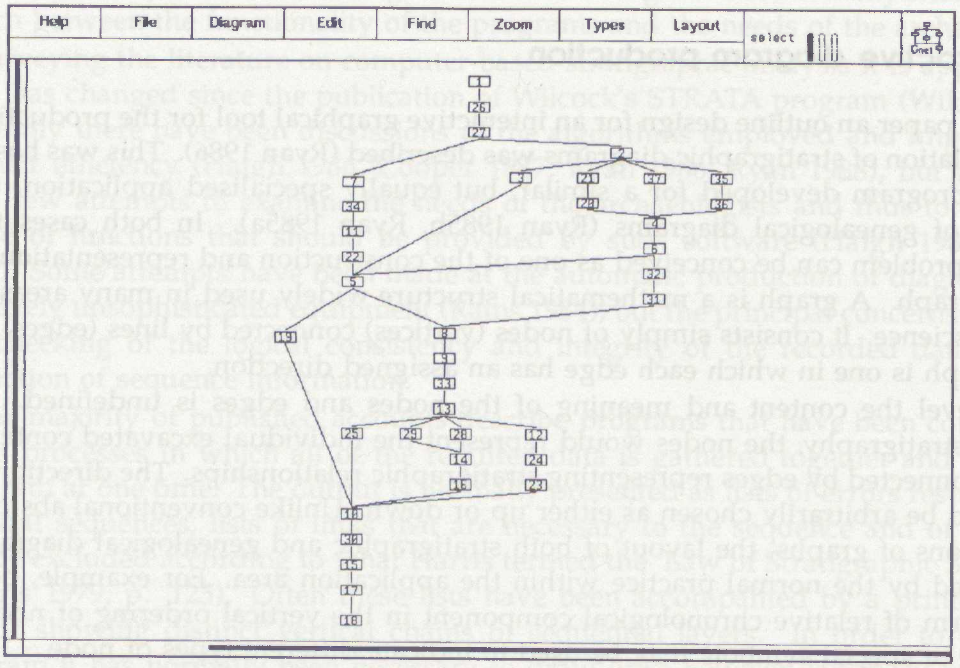

Figure 21.1: *gnet* display of the stratigraphic sequence from a hypothetical site (after Harris 1979, fig. 32 and 33)

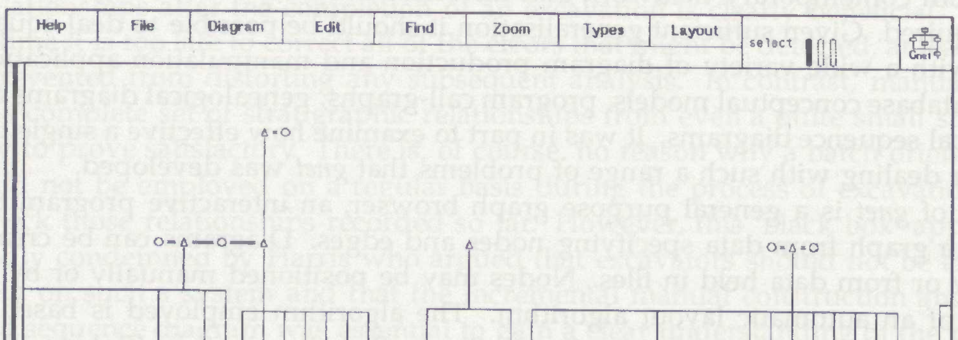

Figure 21.2: Typical *gnet* display showing part of an extensive genealogy. Note the use of different node symbols and link styles to those in the stratigraphic example

to be configured to produce diagrams in a form appropriate to the application. Thus the archaeologist can be presented with a plausible representation of a sequence diagram with each context represented by a labelled box (Fig. 21.1), whereas an anthropologist could see a conventional genealogical diagram with the nodes displayed as triangles or circles according to the sex of the represented individual (Fig. 21.2). The two types of diagram also differ in the layout of the links representing relationships between nodes. The application-specific configuration is achieved by the use of a set up file that contains a number of rules describing the shape, size and colour of displayed nodes and the meaning of the different types of relationships contained in the data. These latter cover both the appearance of the displayed links and information on the precedence and inverse forms of links. For example in the stratigraphic application the file would include the information that if layer *A* was recorded as being *ABOVE* layer *B*, then *A* should be displayed higher up the screen, and that it can be inferred that the inverse relationship *B BELOW A* must also exist.

The program consists of three main modules: a user-interface, a set of general-purpose functions, and a file input/output module. A range of functions that perform graph construction and manipulation operations are organised as a library, thus any or all of these may be incorporated into other programs that need to perform similar operations. The separate user-interface module contains the main event-driven control loop of the program together with all machine dependent aspects of the display of information and user interaction. By maintaining this as a separate module the problems of implementation across a range of machines are kept to a minimum. Similarly, a separate file handling module contains all functions that deal with reading and writing data. These can be adapted to make the best use of available file access methods, or alternatively to link the main program to a database management system.

The present program has been developed as an X client process and is used primarily on Sun and DEC VaxStation workstations. As such it is portable to any display/host combination that supports the X window system. Although the workstations are the preferred development and use environment, it should be noted that similar performance can be achieved on current single-user microcomputers using 68000 or 80286 processors. The major limitation of these machines when compared with the workstations lies in their relatively slow disk access speeds, but adequate performance can be achieved simply because their processors do not have to service multi-user operating systems and window managers.

The program is receiving extensive use in an examination of kinship and other social relations from data contained in sixteenth and seventeenth century English historical documents. Its use in examining archaeological stratigraphy has so far been limited to artificial test data and a sequence from an incompletely excavated site. Further data sets are being sought to enable more extensive testing. A third application, that of the automatic production of program call-graphs has also been investigated, and appears to have potential in the development and maintenance of large softer projects. Indeed, it has been been used to good effect as a tool in the later stages of its own development.

Although such a general purpose tool can be expected to cater for a large proportion of the requirements of any application, from time to time a need for application-specific functions will arise. In practice many of these turn out to be of more general value than is at first anticipated. An example of this is the removal of links that are not necessary to the stratigraphic sequence. As mentioned above, several earlier stratigraphic programs have performed the automatic removal of those relationships between layers that do not add information to the sequence. For example:

A above B
A above C
B above C

Of these three links only the first and third provide *sequence* information. That *A* is above C can be inferred from its position above *B*. Removal of such links clarifies the overall sequence. Nevertheless it would seem unwise to irrevocably dispose of these links as information on contact between, or adjacency of, contexts could be of paramount importance in other aspects of post-excavation work such as finds analysis.

The ability to switch between display of all links and of only those essential to the sequence was built into *gnet* with the intention that it would only be of value in archaeological applications. However, it was subsequently realised that the removal of grandparent–grandchild links, where parent–child links also existed, could be advantageous in the production of genealogical diagrams. Similarly, in the program call-graph application the removal of such links converts a full call-graph into one showing only the basic dependencies between modules.

It would be unwise to claim that all specialised functions might find a use in a sufficiently wide range of applications to justify their presence as standard functions. Too wide a range of functions offered in the menus could make the program appear to be more difficult to use. The problem of overloading menus with unwanted functions can be tackled by providing for application specific menus to be requested in the start-up file. Thus if experience dictates that a particular application requires a number of functions that are not used elsewhere, then these would be added to the function library but only offered in a single menu.

21.4 Beyond diagram production

So far the description of *gnet* has concentrated on the production and viewing of diagrams as if this were an end in itself. This is a useful if only a minimal application for this type of software. The act of selecting a displayed node or link by a point and press mouse operation can be used to invoke a function such as displaying stored information or, as has been done in the call-graph browser application, to invoke an editor with the cursor positioned at the start of the selected program module.

In the introduction to this paper it was suggested that a frequent use of a conventionally produced sequence diagram was in providing a key or graphical index to the excavated contexts. A similar function can be provided by the computer displayed diagram which can form the basis for a simple graphical query language. In the present version this is limited to a facility whereby pointing at a node and pressing the appropriate mouse button causes the information stored in the node data file to be displayed in a second window on the display screen. Given a suitable database management system instead of the simple indexed files in current use the range of possible queries could be be significantly extended.

When used as part of an integrated excavation database the results of queries could be presented in either textual or graphical form as appropriate. For example, all or part of the stored information on a selected context might be displayed in a text window, whereas a request to show all contexts containing a particular combination of artefact types might result in the highlighting of symbols on the diagram. If the single context planning approach is used during excavation, and these plans are stored in the database it would be possible to retrieve and display single or composite plans representing any stage of the excavation. Indeed, it would be possible to step forwards or backwards through the stratigraphic sequence revealing or covering layers as necessary. Many aspects of the excavation could be repeated on the workstation screen. Such a system would have considerable educational as well as practical application. It is anticipated that work in this area will begin in the near future.

Another area for further investigation concerns the potential uses of three-dimensional stratigraphic diagrams. The sequence diagram has proved to be a valuable abstraction of the physical structure of a site, but we might question whether many of the uses of such diagrams might be better served by avoiding the dimensional distortion of the original three-dimensional structure. Given a pair of x,y coordinates representing, say, the centre of a context on the horizontal site grid, the third dimension can be provided by the relative position in the stratigraphic sequence. That is, the vertical position on the conventional two-dimensional diagram.

Potentially such a representation is semantically richer than the conventional diagram as it maintains valuable information on the spatial positions of contexts. However, the effective display of three-dimensional diagrams requires considerably more processing than the two-dimensional form. A balance must be struck between speed of display and rotation of diagrams and the degree of visual realism that can be obtained on a particular system. This development is as yet only at a very early stage and that balance has not yet been reached. Indeed, I have yet to be convinced that it offers much more than novelty; only extended experience with varied data sets can determine whether it is a worthwhile approach.

My lack of confidence in this particular style of presentation as a *replacement* for existing methods comes from a firm belief that the best graphical tools are those that permit the experienced practitioners to continue doing what they presently do, but provide an environment for such work in which the tedium and drudgery of the task is significantly reduced. In other words, good graphical tools should simply make the task easier with the minimum disturbance to what is currently considered good working practice. New working methods may be suggested by the availability of particular hardware or as a result of problem analysis prior to the implementation on a computer, but the implementor's role is not to impose new solutions. Rather they should be made available as alternatives that must then be rigorously evaluated in the field.

21.5 Conclusions

Computer methods for dealing with archaeological stratigraphy have typically been seen as error-checking systems. In the preceding sections I have sought to show that, although useful, this is only a minimal application. Indeed, it is little more than a necessary side-effect of the construction in the machine's memory of a model of the physical structure of an archaeological site. Representations of a site's structure are central to many parts of the excavation decision making and post excavation processes. A computer model of that structure can play a central role in integrating the currently disparate uses of graphics and database systems in archaeological excavation.

As a first step the model may be used in the interactive generation and manipulation of stratigraphic diagrams. These diagrams can then be used as a means of access to conventional information retrieval systems. A suitably configured graph browser can be used as the basis for an interactive front-end to a database in which textual and graphical data are combined to provide an integrated model of an excavated site. Such a system goes well beyond the expectations of earlier stratigraphic software and should have considerable practical and educational potential

References

COOPER, M. A. 1987. "Using archaeological stratigraphy", *Archaeological Computing Newsletter*, 12: 3–13.

HAIGH, J. G. B. 1986. "The Harris Matrix as a partially ordered set" *in Computer Applications in Archaeology 1986*, pp. 81–90. University of Birmingham, Birmingham.

HARRIS, E. C. 1975. "Stratigraphic analysis and the computer", *CAA: Proceedings of the 1985 Conference*, 3: 33–40.

HARRIS, E. C. 1979. *Principles of Archaeological Stratigraphy.* Academic Press, London.

RAINS, M. J. 1984. "Home computers in archaeology", *CAA: Proceedings of the 1984 Conference*, 12: 15–26.

ROBINS, G. 1987. "The ISI Grapher: A Portable Tool for Displaying Graphs Pictorially", *Proceedings of Symboliikka '87, Helsinki, Finland.*

ROWE, L. A., M. DAVIS, E. MESSINGER, C. MEYER, C. SPIRAKIS, & A. TUAN 1987. "A Browser for Directed Graphs", *Software–Practice and Experience*, 17 (1): 61–76.

RYAN, N. S. 1985b. "Interactive Graphical Tools in the Social Sciences". *in* Johnson, P. & Cook, S., (eds.), *People and Computers: Designing the Interface.* Cambridge University Press, Cambridge.

RYAN, N. S. 1986. "Some thoughts on an archaeologist's toolkit", *CAA: Proceedings of the 1985 Conference*, 13: 126–32.

RYAN, N. S. 1988. "Data structures for stratigraphic analysis", *Archaeological Computing Newsletter*, 14: 1–11.

RYAN, N. S. July 1985a. "Gtree: A system for the interactive display and editing of kinship information", *Bulletin of Information on Computers in Anthropology*, 3: 6–15.

WILCOCK, J. D. 1975. "Archaeological context sorting by computer". *in Computer Applications in Archaeology 1985*, pp. 93–7. Institute of Archaeology, University of London, London.