Group and conquer – a method for displaying large stratigraphic data sets

Irmela Herzog
Rheinisches Amt für Bodendenkmalpflege, Bonn, Germany

When dealing with a large stratigraphic data set, it is difficult to get a general idea of the excavation's main features and their chronological sequence because the Harris diagram will become very big. It is hard to understand a diagram consisting of hundreds of equally sized boxes, if no hints are given how to structure them. Large Harris diagrams are seldom published owing to the high costs, and they cannot be displayed properly on a web page. In addition, multilinear or floating sequences (Harris 1984, p. 128) pose a problem, i.e. the network of stratigraphic relationships may be displayed in a large number of diagrams reflecting different chronological sequences.

These problems can be overcome in part by grouping stratigraphic units (Roskams 2001, p. 258). During post-excavation analysis, archaeologists combine individual stratigraphic units to form groups: for example, all the postholes belonging to a building are allocated to a group. Often lower level groups are aggregated to create higher-level groups, resulting in an overview display of the main features of an excavation. British archaeologists call these overview displays *group sequence diagrams* (Hammer 2000, pp. 167-168). According to Hammer, a group summary including dating is vital for understanding a site. Clark (2000, p. 157) notes that the approach which "may be described as the 'Grouping Hierarchy' system … is now familiar and widely used, largely by default".

Though stratigraphic diagrams showing groups as rectangular boxes have been used for decades (an example from 1980 can be found in Hammond 1993, pp. 146-147), no computer program is yet available that handles both groups and stratigraphic data. The only exceptions are graph editors, which allow the user to drag the stratigraphic unit boxes manually so that all the boxes forming a group are within a rectangular area. But these programs do not support any consistency checks, nor do they allow the user to contract or expand a group.

This paper presents features of the new computer program *Stratify* (see also Herzog 2002), which not only provides various methods for defining and checking stratigraphic units (also called contexts in this paper) and their relationships, but is also capable of working with a hierarchy of groups and with phases. *Stratify* allows the user to define "earlier than", "later than", "contemporary with", "equal to", and "part of" relationships. The hierarchy of groups is created using "part of" relationships: each stratigraphic unit may be part of one and only one group, which may be part of another group, and so on. Andresen and Madsen (1992, p. 50) argue that data structures for recording archaeological excavation data should be able to deal with multiple functions of a single stratigraphic unit: for example, a wall may have been part of two buildings. This means that a context may be part of several groups. In practice, the proportion of contexts belonging to two or even more groups is fairly small. It is recommended to split such a context into several virtual contexts each belonging to a different group and to connect the contexts by "contemporary with" relationships.

Stratify can show the group hierarchy in much the same way as a directory structure is displayed in the Windows Explorer, allowing the user to expand and contract branches of the hierarchy (Fig. 1). This group hierarchy plays an important role for group layout and the checks that are necessary as a prerequisite. It is checked that at least one group is present, that groups are not part of themselves, that each group contains at least one context or another group, and that equal contexts belong to one group or to no group at all.
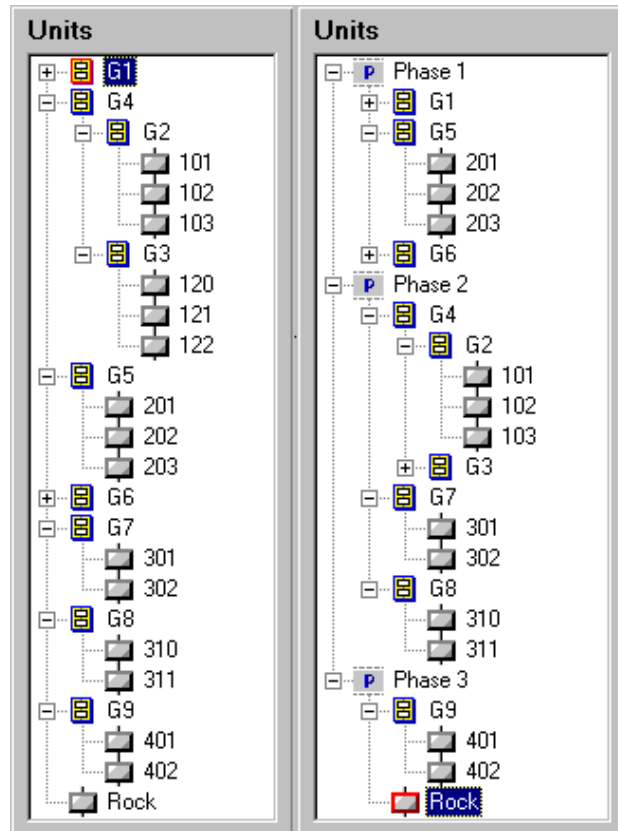
Fig. 1: This figure shows two group hierarchy diagrams created by *Stratify*. On the right, phases are included in the hierarchy. Some of the groups are contracted, indicated by a "+"-sign to the left of the group name, i.e. the members are not shown. Different symbols are used for phases, groups, and contexts.

After a data set has passed these checks successfully, *Stratify* deduces the relationships between the groups from the relationships between members of these groups: for example, if a context, which is part of group 1, is later than a context belonging to group 2, then group 1 is later than group 2. When contradictions are found, i.e. group 1 is both later than and earlier than group 2, there are two possibilities: with the strict group chronology option, *Stratify* will show an error message and will not proceed with the layout. Otherwise, the two groups will be set as contemporary. Even if no cycle is detected in the context relationship network, a cycle may be present in the group chronology deduced from these relationships. In this case no group layout can be created. Some of the deduced relationships between the groups may be redundant, and these relationships are deleted (for detailed explanations of cycles and redundant relationships see Herzog 1993, pp. 207-208).

The group layout starts with the highest level groups in the group hierarchy, then the groups within these top level groups are laid out and so on, until the lowest level, the stratigraphic units, is reached. The algorithm is based on an idea published by Sugiyama and Misue in 1991 (a summary of this paper is given in Kaufmann/Wagner 2001, pp. 210-215): Imagine that the history of the stratigraphic units is written down in a book with several chapters that are numbered. The top-level groups correspond to chapter numbers 1, 2, 3 and so on, the groups within these top-level groups are assigned chapter numbers 1.1, 1.2, 1.3, etc., if they are part of a chapter 1 group. Finally, each stratigraphic unit receives a chapter number. For example, in Fig. 2, context 103 is assigned the number 2.1.3: it is the third context in group

G2, this group comes first in group G4, and G4 is in the second "chapter" of the top-level group layout. In contrast to a book, a chapter number may occur several times. This happens whenever two groups or two contexts are positioned on the same depth level in the Harris diagram. With the chapter numbers the vertical sequence of the contexts and the groups is determined. If the nesting levels of the groups differ, a compaction procedure reduces the white space within each group, so that for example, in Fig. 2, contexts 249 and 103 are positioned on the same depth level, though their chapter numbers differ (249 has chapter number 2.4, i.e. 249 is part of group G6 which is in the second "chapter" of the top-level group layout, and within group G6, context 249 is on the fourth depth level).
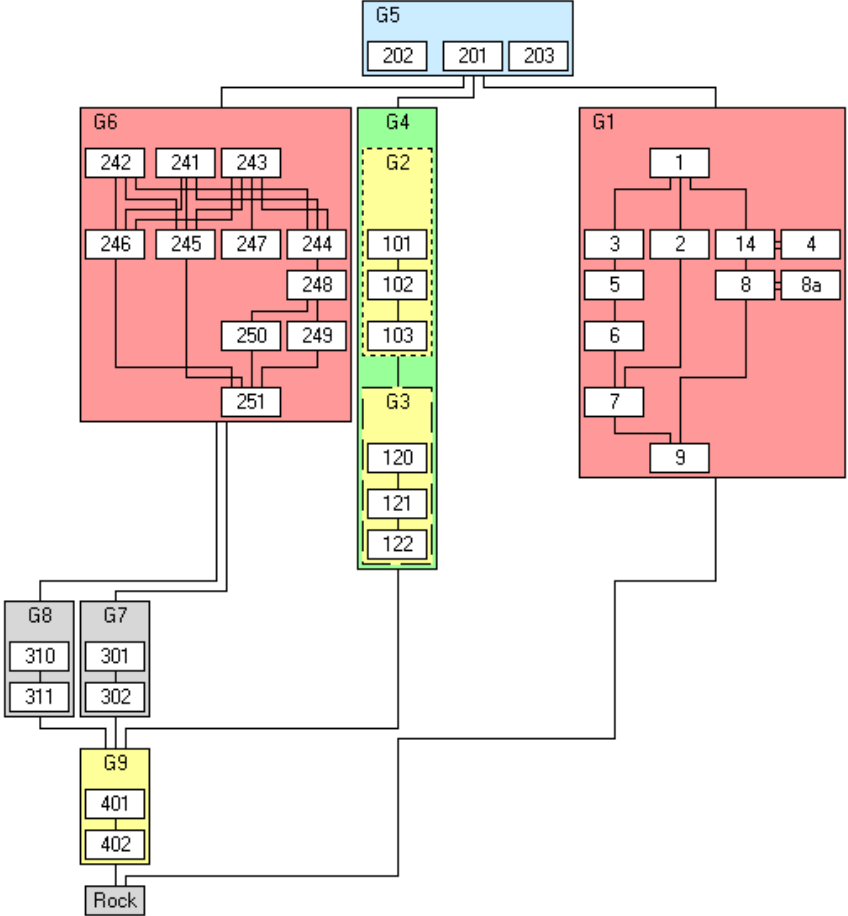

Fig. 2: Example of a group layout created by *Stratify*.

This method may not only be used to establish the vertical sequence of the contexts and groups, but may be applied similarly to determine the horizontal order. The algorithm used for laying out the top-level groups and the contents of each group mainly draws on methods suggested by Sander (1996b). The "earlier than" and "later than" relationships are depicted as orthogonal lines that connect the appropriate context or group boxes. Some extra effort is required to ensure that any two horizontal relationship line segments do not share a point and, similarly, that vertical line segments never touch.

The result of the group layout algorithm is a Harris diagram of groups, whereby each group can be considered as a Harris diagram of groups and contexts and so on, until the lowest level in the hierarchy is reached. Contrary to a standard Harris diagram, the grouped Harris diagram consists of boxes whose sizes differ. The layout spaces the depth levels based on the height of the tallest group in the level. In the same way, each column width is dependent on the

maximum group width in this column. This may force groups on subsequent levels to be placed further away from their predecessors than necessary. However, it is the only way to show contemporary relationships between groups.

Furthermore, with this layout strategy it is fairly easy to implement a contraction and expansion feature: if the user decides to hide the details of a group, i.e. to contract the group to a so-called black box (Paulisch 1993, p. 77), then the only change is that the members of this group and their relationships are ignored during layout, but the chapter numbers remain unchanged for the rest of the groups and contexts. The layout method outlined in this paper creates a "grey-box" display of the Harris diagram, i.e. the relationships between contexts belonging to different groups are not shown. A "white-box" display, which does include all relationships, requires a more sophisticated layout strategy; the appropriate technical term in graph theory is "compound digraph" (Sugiyama/Misue 1991; Sander 1996a).
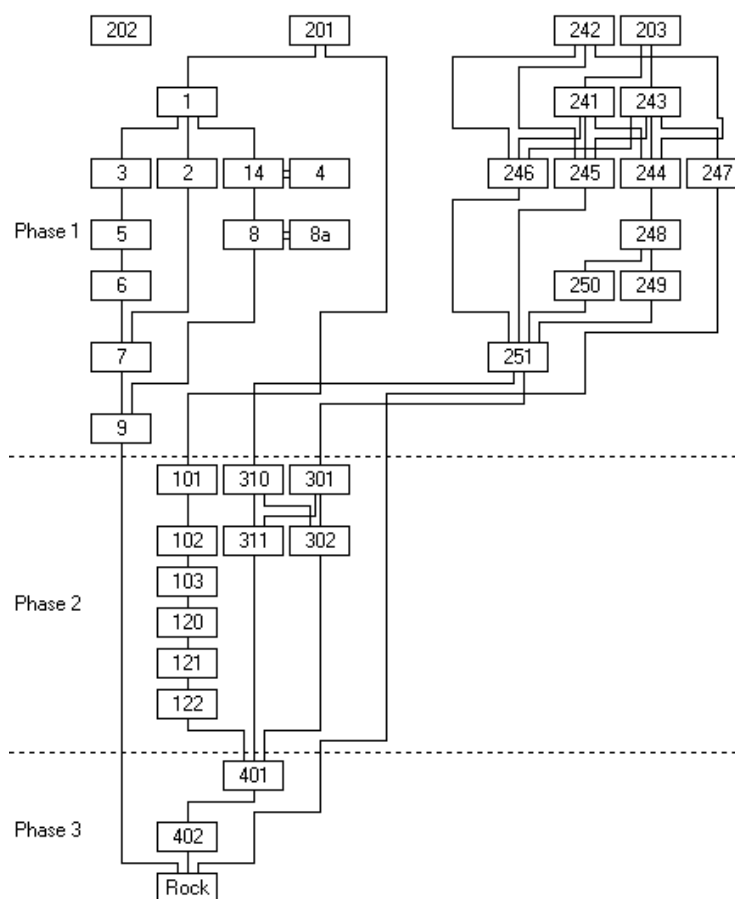
Fig. 3: Example showing the data set of Fig. 2 with phases but without groups.

The program *Stratify* allows users to create a simple Harris diagram, a layout with phases (Fig. 3), one with groups, or one with groups and phases. The procedures used for laying out a Harris diagram with phases bear some similarity to group layout. The chronological relationships of the phases are deduced from the relationships between contexts belonging to different phases. The aim is to create a consecutive chronological sequence of the phases. If there is no unique chronological sequence of the phases, the user will be asked to sort the phases. When combining groups and phases in a Harris diagram, each group must belong to one phase only. This is checked by creating a group hierarchy including phases, i.e. the phases are the top-level groups in the hierarchy.

The program can be downloaded at www.stratify.privat.t-online.de.

References:

Andresen, J./Madsen, T. 1992. Data Structures for Excavation Recording. A Case of complex Information Management. In: C. U. Larsen (ed.), Sites & Monuments. National Archaeological Records, Kopenhagen:49-67

Clark, P. 2000. Post-Excavation Analysis: moving from the context to the phase. In: St. Roskams (ed.), Interpreting Stratigraphy. Site evaluation, recording procedures and stratigraphic analysis. *BAR Int. Ser. 910, Oxford*:157-159

Hammer, F. 2000. From Field Record to Publication with a Minimum Archive Report. In: S. Roskams (ed.), Interpreting Stratigraphy. Site evaluation, recording procedures and stratigraphic analysis. *BAR Int. Ser. 910, Oxford*:161-172

Hammond, N. 1993. Matrices and Maya archaeology. In: E. Harris et al. (eds.), Practices of archaeological stratigraphy, London/San Diego:104-121

Harris, E. C. 1984. The analysis of multilinear stratigraphic sequences. *Scottish Arch. Review 3 (2)*: 127-133

Herzog, I. 1993. Computer-aided Harris Matrix generation. In: E. Harris et. al. (eds.), Practices of archaeological stratigraphy, London/San Diego:201-217

Herzog, I. 2002. Possibilities for Analysing Stratigraphic Data, CD Workshop Archäologie und Computer Wien

Kaufmann, M./Wagner, D. (eds.) 2001. Drawing Graphs. Methods and Models, Berlin/Heidelberg

Paulisch, F. N. 1993. The Design of an Extendible Graph Editor. *Lecture Notes in Computer Science 704, Berlin/Heidelberg*

Roskams, S. 2001. Excavation. *Cambridge Manuals in Archaeology, Cambridge*

Sander, G. 1996a. Layout of Compound Directed Graphs. Technical Report A/03/96, Universität des Saarlandes, Saarbrücken

Sander, G. 1996b. A Fast Heuristic for Hierarchical Manhattan Layout. In: Brandenburg, F. J. (ed.), Graph Drawing Proceedings (GD '95), Berlin/Heidelberg/New York: 447-458

Sugiyama, K./Misue, K. 1991. Visualization of Structural Information: Automatic Drawing of Compound Digraphs. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21, No. 4:876-892