HYEF 93/94

Database integrity

The huge size of the databases (the tables for the context database alone run to 14,549 Kbytes, not including indexes, supporting files and linking programming) has meant that not every field of every entry could be checked for accuracy. Context records were checked for basic information only (type of context, stratigraphy, grid reference) and errors amended. Where other errors were noted haphazardly during analysis, any that would actually affect report production have also been amended. Among those fields which were systematically checked, the levels of input error noted ran at around 1%, and in some fields, around 4%, so that for 17000 contexts each with 50 fields, 8500 to 34000 errors may be expected, very few of which have been corrected. For the finds data, the same level of initial input error may be assumed, but fewer of these errors will have remained, as most finds records have been revisited in the course of analysis and identifications tightened up or altered, and other errors corrected at the same time. For the specific pottery databases, where the types of entry tend to be very similar and easily confused, a higher rate of error may be expected but more intensive checking has also occurred here and most errors are believed to have been eliminated.

Thus almost all quantification within this report must be assumed to be no more accurate than $\pm$ 1% at best, perhaps ±5%. It is also worth noting that the process of tailoring individual and idiosyncratic records to standardised database formatting loses some of the character of the originals, including (albeit rarely) actual nuances of meaning that may be significant: where an excavator has, e.g., 'Pit/Ditch', the database will have only 'Pit' and a check in a Query box. Also some excavators, and some inputters in this situation, aware that 'Pit/Ditch' was not 'acceptable' to the database, will have settled for 'Unknown', which is clearly worse.

THE REPORT

This report is presented in ?two volumes. The first volume consists of thematic discussions of the evidence, supported by much but by no means all of the detail required to assess the arguments. The second volume is devoted to detailed presentation of basic data. However, even within this format, a considerable amount of selection has been necessary, and the reader will look in vain for detailed catalogues of every find in every context or descriptions of every context recorded. Selection at every point has been driven by consideration of the needs of current research aims as highlighted in a number of research agendas which were coming to publication during the course of the project, for example, *An Archaeological Research Strategy for the Eastern Region* (ref), *Understanding the British Iron Age: an agenda for action* (James and Millett, forthcoming), etc. In all sections, consideration has been given to what are seen as the **key elements** of the site or of its various material assemblages.

It was recognized at an early stage that an all-encompassing analysis of all assemblages and aspects of the data-collection was not possible and, arguably, not necessary (*c.f.,* the "need for critical selection of data" recognized in English Heritage 1991, 1). Extensive discussion at the assessment stage, on which aspects were most worthy of detailed attention, narrowed the focus of analysis to manageable and deliverable proportions. This report therefore presents the results of the pursuit of only the perceived **key** aspects of the site and those parts of the material assemblages which provide information pertaining to these. Therefore, specific research aims were outlined before the analysis began. These were approached, as far as possible, as a series of distinct but linked topics for which detailed stratigraphic, artefactual and/or ecofactual analyses were carried out on a question-specific

basis, integrated and co-ordinated by the Project Manager. This approach ensured that the large volume of complex analysis work was broken down into manageable and quantifiable tasks, the progress of which could be monitored and reviewed relatively easily both internally and by English Heritage.

Finds are discussed by functional class rather than by the more usual division by raw materials. Reports on individual classes of evidence have been prepared using guidelines drawn up by various special interest bodies (The Study Group for Roman Pottery, *etc.*) but within these guidelines, not all evidence could practically be published. Some classes are published in full (coins, brooches) simply because this has always been the case. Full catalogues and reports for every class of material are available in the project archive.

All excavated features appear on the general area plans. Only those specifically discussed in the text, however, are labelled on the Period Plans. This is purely for practical reasons: over much of the site any attempt to provide a plan with a label for every feature simply produces too much information to fit onto a page legibly. The area reports describe all the important landscape features and all features with important finds assemblages, but this still omits hundreds of post holes, minor slots and gullies, and many layers of deposit which can only be vaguely understood at best. The stratigraphic/structural section of the report describes all the major site features selected as above. Where possible, details that can be gleaned from the published plans have not been described in text. Descriptions have been kept short, but do include discussion of interpretations and problems. In many cases the description may be dependent on the interpretation as much as *vice versa*. It is simply not possible (or honest) to offer a decisive split between 'objective' observation and 'subjective' interpretation for many site features. Groups of similar types of features within the same period have tended to be described *en masse* rather than individually.

However, it remains the case, paradoxical as it may seem to us, that smaller sites with less information to impart receive disproportionately higher levels of reporting than massive projects where sheer size deters commensurate investment. Quite apart from the obvious shortcomings of not having excavated more than a fraction of the threatened archaeology, the publication programme itself is constrained by physical and financial factors. While the authors remain responsible for errors of commission or omission relating to the research that was undertaken, the financial realities of modern archaeological research must shoulder responsibility for the absence of much more that could have been attempted, and for much that was but simply cannot be presented.

F. THE ARCHIVE

Deposited in xxxxxxxxxx, the full site archive consists of tens of thousands of pages of text, xxxx CDs of digital data, including the whole of the context and finds databases, all the digital plans, scanned images of selected other illustrations, etc. Original context record sheets alone occupy 58 lever arch files, while the original plans and sections occupy two tanks, 4 drawers and 24 lever arch files. Sufficient working documents have also been preserved to enable some tracing of an audit trail of thoughts and processes through post-excavation, but again some selection has occurred here and not every version of every draft of every document is preserved. All working notes from the field however have been preserved to allow anyone foolish/brave enough to retrace decisions over problematical stratigraphy or on-site interpretation. All of these have been indexed to a degree which ought to permit easy access. Records relating only to project resource management have not been deposited with the archive but retained by ECC FAU.

Record type	Total
Pre-excavation plans	349
Post-ex plans - multi context	721
- single context	4962
Section sheets	881
Stratigraphic Matrix sheets (most A0 size)	56
Other illustrations	
Must be lots more…..	
Colour film	476
B&W film	458