

## Section 5: Archiving CAD Data#

### 5.2 The Selection, Retention and Discard of CAD models#

Many large projects make extensive use of CAD and during their life span literally thousands of CAD models may be created and saved as part of a project's archive. For example, CAD is often used to digitise the hand-drawn plans and sections from large archaeological excavation projects. An excavation may have thousands, or even tens of thousands, of individual contexts, each of which are often digitised and saved as a separate CAD file. During the process of post-excavation these files are often agglomerated into sets of group, sub-group and phase plans, and CAD models may be saved at each stage in the process. All of these later models are essentially composites of the earlier context plans. The appropriate use of a layer-naming system (see Section 4.3) can help to reduce the need for large numbers of separate files. Rather than generating new files each time, new layers can easily be added to existing models to reflect changes in interpretation and, in the process, maintain a close relationship between the underlying data and the interpretations derived from them.

At some point, all project managers need to consider the question of whether it is really necessary to archive and, of course, to document every model. There are obvious cost implications associated with these decisions. During the final stages of a project there should be a process of data selection, where the overall archive is worked through and individual files are either selected for retention in the archive or are discarded. This process is a standard part of the preparation of the non-digital project archive for deposition, and should also be part of how large digital archives are dealt with. For example, there are arguments to be made for the inclusion of every set of group, sub-group and phase plans in the archive, despite the fact that they are often simply agglomerations of the individual context plans. Essentially this will lead to a lot of duplication of data in the archive. Nevertheless such composite plans represent the cumulative results of interpretative decisions made by the archaeologists and as such are important building blocks towards the overall understanding of the site. Consequently it is important that these files are archived as they have a high re-use potential. It is also important that the individual context plans are archived alongside the group, sub-group and phase plans as they can be used to question the original archaeologists' phasing of the site and as such can be re-used to attempt radical re-interpretations of the archaeology.

Nevertheless, there will be CAD files that are appropriate to discard and omit from the final archival deposit. Such models include test and unfinished versions of later plans or earlier versions of phasing, which have been superseded by later interpretations and would consequently lead to false impressions of the archaeology of the site.

It is important to document the selection, retention and discard policy for a given project.

#### 5.2.1 Data Currency

The issues of which files to include in a final digital archive and which to leave out neatly brings us to the issues surrounding data currency. Often during the process of interpretation things change and there is not always time to go back to previous versions of files to bring them up to date. For example, during post-excavation interpretations evolve and those writing the site up tend to work from the latest version of the plans, so that previous versions become orphaned. Under such a scenario the individual context plans, once incorporated into sub-group or phase plans, may not be revisited. Consequently as interpretations change, older versions of files tend to become fossilised with the interpretations at the time of their last save. In an extreme case a sub-group may be initially thought to belong to a particular phase during post-excavation but additional information, such as refined dating evidence, may in fact mean that it belongs to a completely different phase. Such things happen; in the ideal world of course there would be a constant process of revision of earlier files to

ensure that the entire archive is of the same currency. In the majority of cases, however, this would not be practical or financially viable. Consequently during the process of data selection, mentioned above, there should also be an assessment process during which the currency of the information contained in each file is considered. After this a policy of data revision may be put in place if time and cost constraints permit.

The presence of large volumes of orphaned CAD files within an archive could give a false picture of the archaeology of the site for the unwary user of the data. For some projects a programme of data revision may be out of the question, and consequently serious questions should be asked and the issue of whether such demonstrably flawed datasets should be retained or discarded from the digital archive must be confronted. Where the datasets still contain valuable information about the site, or where the archive would be seriously deficient without the inclusion of such files, it may be possible to retain orphaned CAD models as long as they are deposited with a 'health warning'. Such a warning should document which files are orphaned, why the information contained within them is not of the same currency of the rest of the archive and how they may be brought up to this level. The process of the creation of this documentation may be long and complicated and it may well be as time consuming as a programme of data revision. Nevertheless it is up to individual projects to reach a decision as to which policy is best for them.

### **5.2.2 Copyright**

The data deposited with a repository may be the sole copyright of the depositor or copyright may be jointly held. Arrangements for fair use of the digital data will generally be specified at the time of deposit.

There are real difficulties in enforcing copyright of digital data. It is virtually impossible to detect some kinds of copyright violation and the legal framework for dealing with digital materials has not matured. Nonetheless copyright exists and should be asserted. In the case of commercially valuable data, legal advice should be sought at the time of deposit.

Some data sources used in preparing CAD models (e.g. maps, drawings and photographs) are likely to be copyrighted by others. Project directors must not only have permission to use these sources but they must be certain that all necessary permissions to use the new version of the data have been granted.