

## Section 5: Archiving CAD Data#

### 5.1 Planning the Curation and Preservation of CAD Data#

As described in the previous sections of this guide, CAD models are created from data which are collected, manipulated and developed in a digital environment. This puts them in a special position with regard to archiving as the ever-increasing pace of change in computer hardware and software means that in a few years' time these precious research materials may be lost forever. The best strategy for long-term preservation of data in digital formats is for them to be systematically collected, maintained and made accessible to users operating in very different computing environments. For all practical purposes, data from any project will only continue to be available if the data have been archived. It is important to put archival storage plans in effect from the moment data gathering begins.

#### 5.1.1 The Archival Need

It is not obvious to all that digital data must receive special care. However, the problems that can arise were demonstrated in the United Kingdom through ADS work rescuing the contents of the Newham Museum Archaeological Service digital archive (see the general section [What is Digital Archiving?](#)). The Newham example is particularly relevant to CAD as it included a series of site matrices that were produced in an early version of the TurboCAD software. These files could neither be converted into DXF nor read by newer versions of TurboCAD and thus the data that they contain are effectively lost.

Digital archives require special care for the following reasons:

- Magnetic and optical storage systems for digital data have finite lives and data files must be copied onto new storage media on a regular basis to prevent loss
- Digital data formats change rapidly, some becoming obsolete in a few years. A decade is a long period in the digital world. Data that are held in non-preservation file formats, i.e. proprietary file formats, can become irretrievable as versions of software packages go out of use
- Non-existent or inadequate documentation makes it difficult to reconstruct which data goes with which project and limits the potential for re-use.

The absence of a standard file format for CAD (see [Section 3.2](#)) is a particular problem. Although DXF is the most widely used format for CAD it is a proprietary standard developed by AutoDesk and has changed slightly with virtually every new release of AutoCAD.

The Newham Museum Service digital archive makes a salutary tale, but it is important to remember that it was compiled at a time when digital archiving was in its infancy. The purpose of these *Guides to Good Practice* is to put strategies and methodologies in place to ensure effective digital archiving of project data.

#### 5.1.2 Planning for the Creation of Digital Data

From the moment a project begins, careful thought must go into the preparation of the digital archive that will be delivered at the project's conclusion. Planning should include:

- **Preparing a project design** that documents the tasks necessary for the successful completion of the project at its outset, and includes a summary of the types of digital data that will be created. It is important to update this documentation throughout the life of the project.
- **Defining and documenting areas of responsibility** for creating and managing digital files at all stages of their life.

- **Planning the file formats that will be used** for both the secure archiving and the dissemination of data. The formats used for these two activities may be different.
- **Checking** with the digital archive facility destined to receive the files **to see if there are any guidelines or standards that should be followed**. If local guidelines do not exist, it is recommended that the guidelines in this document are followed and that the ADS or Digital Antiquity are consulted for up-to-date information.

Data, accompanied by adequate documentation, should be deposited in a digital archive as quickly as possible after the conclusion of the project. There are two reasons for this:

- Some kinds of digital degradation can occur quickly and prompt archiving is desirable
- Prompt archiving helps project personnel and archive staff to make sure that adequate documentation has been provided for long-term archival care of the files. If too much time passes before deposit, it may be difficult for project personnel to reconstruct the information required by the archive.

### 5.1.3 Storing Digital Datasets

During the working life of a project, digital data may be created on the hard disks of standalone PCs, on laptop computers or on network drives. Data may be acquired or stored on various electronic media. Whatever the initial storage media, ideally digital files that are in use should be routinely backed up and this may involve transferring them onto a network drive. A detailed discussion of general considerations can be found in [Planning for the Creation of Digital Data](#).

### File-naming Conventions

Digital files should be given meaningful titles that reflect their content. Plan to use standard file-naming conventions and directory structures from the beginning of a project. If possible, use consistent conventions across all projects. File-naming is discussed in [Section 4.3.2](#).

### Version Control

It is extremely important to maintain strict version control when working with files, especially with CAD models which may be processed using a series of different treatments.

There are three common strategies for providing version control: file-naming conventions, standard headers listing creation dates and version numbers, or file logs. It is important to record, where practical, every change to a file no matter how small the change. Versions that are no longer needed should be weeded out, after making sure that adequate back-up files have been created.

Another aid to version control is to use separate directories for raw, working and archive data. All primary field data (including the first loading in a CAD drawing) should be 'archived' as it comes in from the field and a copy taken as the 'working file' for editing. Then each identifiable product, for example, aggregations of single archaeological context drawings into group or phase drawings, can be archived separately. The key is to have separate folders for files in each directory to ensure that the data cannot be accidentally updated and overwritten. An index should be created for each directory.