

What is Digital Archiving?#

Archaeology is in a special position with respect to archiving because archaeological fieldwork, which creates archaeological data, also destroys the primary in situ archaeological evidence itself. Increasingly, the digital record may be the only source of information about archaeological research materials. It is essential, therefore, that the digital records that describe archaeological resources be made accessible and that their preservation be ensured. Providing for the accessibility of archaeological data and its long-term preservation are the goals of digital archiving.

Digital archiving is different from traditional archiving. Traditional archiving practice seeks to preserve physical objects (e.g., artefacts, samples, paper, photographs, microfilm) that carry information. Digital archiving seeks to preserve the information regardless of the media on which that information is stored. Computer disks and other magnetic and optical media degrade, and the information on them is lost unless it has been moved to other media. Software and hardware change rapidly: the physical media on which digital data are stored are impermanent. Other methods are necessary to ensure wide access to and long-term preservation of digital data.

The Goals of Digital Archiving#

The overall goals of digital archiving are simple:

- Permit easy and wide access to digital archaeological data for cultural, educational, and scientific purposes.
- Ensure the long-term preservation of digital data so that it remains accessible for appropriate uses in the future.

The Principles of Archiving Digital Data#

The points below present an outline of the key issues to be considered when creating a digital archive.

- Ensure that existing digital data are safeguarded and deposited in an appropriate digital archive.
- When creating a new digital archive, ensure that it conforms to existing standards and guidelines on how data should be structured, preserved and accessed.
- All digital archives should ideally be deposited in a digital archiving facility or collections repository where they can be properly accessed, curated, and maintained for the future.
- The key to successful digital archiving is thorough documentation of the data, how they were collected, what standards were used to describe them and how they have been managed since collection.
- If there are concerns that some data (e.g., specific site location information) needs to be kept confidential (as required by the Archaeological Resource Protection Act (ARPA) in the US), a means of easily separating these data from non-confidential data must be developed for reports, analytical datasets, and for displaying site locations on maps. It is also essential that this process is documented and deposited as part of the archive.
- There is generally no need to preserve interim versions of final digital files. Exceptions to this include interim datasets where either data or text is subsequently discarded or decimated to final publication. These issues are discussed in the later section on [Preservation Intervention Points](#).
- Data already held safely in paper archives do not need to be digitised, except to provide a digital security copy or online access to the data. When digitising or scanning from paper records, do not automatically discard the paper originals when complete. Offer them to relevant documentary archives.
- Although the digital, paper and archaeological resource archives may be dispersed, the integrity of the complete archive must be ensured by cross-referencing between physical collections and digital records.

In accordance with the principles presented above, digital archives should at least provide an index to archaeological sites, finds and paper archives and at best provide access to digital records of data, material, documentation, interpretation and analyses. It is recommended that the collection or creation of digital datasets be planned at the outset of a project and incorporated into project scopes of work and specifications. It is recognised that funding agencies must acknowledge such requirements if widespread implementation is ever to be achieved.

Two examples, one from the UK and another from the US, illustrate potential problems when planning for a digital archive is not incorporated as part of project planning and execution.

The Newham Archive: A Case Study of the Loss of Digital Data#

This problem has been effectively demonstrated through work to rescue the contents of the Newham Museum Archaeological Service digital archive . The Archaeological Service was closed down in 1998 and, although its physical collections are still curated by the London boroughs of Newham, Redbridge and Waltham Forest, the digital archive was passed to the ADS. The digital archive represented all the work that was digitised during Newham Archaeological Service fieldwork and post-excavation analysis, along with project designs over a period of about ten years. This archive was delivered to the ADS on 230 floppy disks containing over 6000 files and totalling over 130Mb of data. Much of the data was held in archaic formats or in proprietary software and significant time and effort were required to rescue these files. Unfortunately around 10-15% of the files are still inaccessible and the data that they contain are effectively lost. An additional problem was that the archive was inadequately documented and it was often difficult to reconstruct which files belonged to each project. As a result, there are a number of 'orphaned' datasets, including a large cemetery database, which have been rescued but have little reuse potential.

The Newham Museum Archaeological Service digital archive had two main problems:

1. Data held in non-preservation file formats, i.e. proprietary file formats that have gone out of use;
2. Non-existent data or project documentation.

The Newham digital archive is probably typical of the digital information resources of archaeological units. There are many archaeology units with archives of files in redundant formats, without explicit information relating them to sites, containing unexplained coding and in unknown states of completion (cf. Condron et al. 1999). These files may also be stored on unsuitable media in poor physical storage conditions. In short, there may be large amounts of 'archived' archaeological information that can never be accessed again.

The Newham Museum Service digital archive is a depressing and salutary tale. It developed as a working tool to help the Service write up and manage its archaeological projects, and in this respect, the archive was fit for its original purpose. The concept of digital project archiving was still in its infancy when the Newham archive developed. As there were no published strategies or methodologies to ensure the effective preservation of digital data at the time, the poor condition of the Newham archive is understandable.

Soil Systems, Inc.: A Study in Data Recovery#

Soil Systems, Inc. (SSI) was a cultural resource management firm that conducted archaeological projects throughout the American Southwest for more than twenty years. Based in Phoenix, AZ, the firm concentrated its work in the state of Arizona and completed a number of very large archaeological compliance projects in the Phoenix metropolitan area. The firm was perhaps most widely known for its extensive and thorough excavations at Pueblo Grande, one of the largest and most influential Hohokam sites in the Phoenix Basin. SSI's work at Pueblo Grande spanned at least five separate testing and/or data recovery projects that took place over the course of more than a decade. The firm's efforts resulted in an impressive set of data that covers most of the known extent

of Pueblo Grande. This dataset alone may rival, in extent and in detail, any other data collection from a single site in the American Southwest.

Unfortunately, SSI closed in 2008, a victim of the worldwide financial crisis that began that year. Most of SSI's physical collections and records (i.e. field notes, paper data records, and artifacts) were curated either at the Arizona State Museum, University of Arizona or at Pueblo Grande Museum, Phoenix, AZ. When digital data were created, some of the finalized data tables were converted to text formats and passed to the Arizona State Museum or Pueblo Grande Museum along with physical records and artifact collections. However, the majority of SSI's digital data and records that document the archaeological projects they completed in Arizona, New Mexico, Colorado, Utah, and Nevada remained in proprietary formats on local hard drives and servers. In addition, nearly all of the digital data associated with several large projects that SSI was completing at its closure were not passed to state or municipal repositories.

SSI's digital data for most of its archaeological projects were stored in Advanced Revelations (AREV) version 3.1, an early relational database platform. Provenience and artifact analysis data for more than 50 separate projects, including several Pueblo Grande projects, were held in AREV file formats. Spatial data were stored in other formats on SSI's server and on individual, local hard drives. Since SSI produced most of their site maps and report figures in AutoCad, vast quantities of spatial data were stored in now-outdated AutoCad file formats. Finally, individual analyses, integrated data tables, draft reports, and final reports were stored in multiple file formats on the company server.

Thus, at SSI's closure, digital data for at least a hundred archaeological projects and the extensive, yet still un-integrated, Pueblo Grande dataset were threatened with increasing chances of loss. The knowledge and software required to interface with AREV formats and to export the data grew more and more scarce with the passing of time. In addition, the hardware on which the data were stored, which was capable of running the original software programs, was growing older and more obsolete.

A grant project sponsored by Digital Antiquity is currently attempting to rescue the digital data associated with several large SSI projects at Pueblo Grande. Project participants have interfaced SSI's server and hard drives with current computing hardware and extracted all of SSI's digitally stored data. In addition, they have booted the AREV database program in a Windows 7 environment and re-learned how to work with AREV to extract data tables from its relational datasets. They are migrating all of the recovered data to more stable formats and creating multiple copies to ensure the long-term preservation of the archaeological data SSI created. Moreover, the project will integrate large portions of the Pueblo Grande dataset and curate the originals with the Digital Archaeological Record (tDAR).

The Soil Systems, Inc. digital data collection faced three primary threats that may have led to data loss:

1. Data stored in non-preservation file formats, i.e. proprietary file formats that are no longer commonly used;
2. Large amounts of digitized and "digital-born" data stored locally, on internal servers and internal hard drives;
3. Lack of resources to migrate large amounts of digital data to a curation facility/repository capable of adequately storing these data.

The problems that threatened SSI's digital data collection highlight critical issues that are common to CRM and other private firm digital data archives. In most instances, private firm archaeological data are entered into, created by, and stored in widely available commercial software. At present, these data are frequently "born" in digital environments, and are growing increasingly complex as firms have access to ever more sophisticated and powerful software packages. Although data are often backed up on servers and stored as multiple copies, they are infrequently converted to preservation-minded formats (i.e. exported from proprietary formats and into more stable, persistent formats). Second, private firm digital data are stored primarily on the hardware purchased and

owned by the firm. A firm may provide copies of finalized digital datasets to an agency or repository along with its final project report for individual projects. However, these datasets likely represent only a subset of the digital data actually created during the completion of an archaeological project. In addition, the submitted digital data are often divorced from the project and site metadata when placed in a repository facility. Finally, many private firms lack available resources to independently undertake large-scale digital curation efforts on a long-term basis. In particular, these firms have almost no resources at their disposal for digital data conversion and migration when they close their businesses. Their data are then under immediate threat of obsolescence and loss.

The Newham Museum Service and Soil Systems, Inc. case studies describe common problems with aspects of current digital archaeological archiving practices. These Guides have been developed to provide better preservation strategies for archaeological project data that will enable individuals and organizations to avoid problems and create useful and easily-preserved digital data. One clear step toward improving archaeological practice in this area is recognition that the road to long-term preservation begins not at the end of a project but at its inception.

Bibliography

Condrón, F., J. Richards, D. Robinson and A. Wise (1999) *Strategies for Digital Data - Findings and Recommendations from Digital Data in Archaeology: a Survey of User Needs*. Archaeology Data Service, York.

'What is Digital Archiving?' Edited by Kieron Niven with contributions by Mason Scott Thompson
Archaeology Data Service / Digital Antiquity (2011) *Guides to Good Practice*