FISH

Forum on Information Standards in Heritage

FISH
Fact Sheet

ads

A Six Step Guide to Digital Preservation:
An introduction to digital preservation by the Archaeology Data Service on behalf of FISH

This document provides a short description of the strategies that need to be in place to ensure the long term durability of archaeological data. It assumes a **data migration strategy** and places an emphasis on the influence a data creator may have on the prospects for digital preservation.

1. Planning
The best thing you can do to ensure the long-term preservation of data is to plan for its re-use in the future. Studies show that re-use of data is the single surest way of maintaining the integrity, and subsequent longevity of data. Planning for re-use may have cost implications for any archaeological project and, although the costs can be minimised, it is worth ensuring that timetabling and budgeting take account of the process of quality assurance.

Once you have decided that your data will be of long term value, there are five routine practices that should be planned, budgeted for and carried out during your project: **Backup, Security, Documentation, Refreshment, Migration**. Provision for digital preservation should be built into all the planned stages of a project, not left until the end when time, funds and expertise may not be available.

2. Backup
Backup ensures that there is a 'snap-shot' copy of your data held somewhere else, in case of disaster. These copies are important in the life span of the project, but are not strictly the same as long term archiving. Although a necessary part of any exercise in the creation or enhancement of digital data, backing up files will not solve their long term archival problems.

The most widely used back up strategy is the so called **Grandfather-Father-Son** rotation strategy. The system works by employing a rotation of full and partial backups taken daily, weekly and monthly. The most recent full backup, the 'father', contains a snap shot of the whole network or data set at the start of a week. 'Sons' are more frequent, normally daily, backups containing only the changes to the data set executed on that day. These copies don't have to be kept in perpetuity, but can be recycled every time a new father is created. Once a month or so, a permanent complete snapshot is taken, which should be stored in perpetuity and would not normally be recycled. This monthly backup is the "grandfather", and can be brought out in moments of real crisis. It is best practice that the weekly and monthly backups be stored away from the office, preferably in a secure, fireproof, anti-magnetic environment. For a small data set, or one that changes infrequently, such regular copying is excessive. The system can be tailored to individual requirements and the time periods expanded or contracted as necessary.

Magnetic tape is the most common back up medium. Other low cost alternatives such as zip drives, high capacity floppy disks, CDs and DVDs are just as effective. Networks provide a further alternative, with disk arrays and storage areas managed remotely.

3. Security
Archaeological data can be fragile information, especially if the digital version is a master copy. It is important to protect the system on which the data is held from physical damage or theft and 'virtual' damage, such as viruses. Don't be careless with passwords and if you run a large network, monitor the network for unusual activity at unusual times.

Steps to prevent viruses:

- Install anti-virus software on your computer and keep it up to date.
- Be suspicious of any unsolicited programs or files, particularly from email.
- Don't download software from the internet unless strictly necessary.
- Scan all files received via email, shared discs or USB pen drives.
- Ensure software suppliers will underwrite (reasonable) damages incurred should the software contain a virus, or that the product comes with an anti-virus guarantee.
- Have a backup strategy in place.

4. Documentation

Documentation (metadata) is essential for large and complicated data sets. Abbreviations may not make sense to someone re-using the data. Documentation would include an expansion of codes or abbreviations used, and a (brief) description of each file, explaining how they fit together. Many programs feature the option to include documentation at the creation of the files, usually via the 'Properties' option in the 'File' menu. The type of documentation created will depend on the type of data used; that for a text file may be quite straightforward, but for a database quite intensive. The documentation should describe the rights management framework of the data (copyright and IPR) and consistent file naming conventions should be used to ensure users can navigate large numbers of files.

5. Refreshment

Computer hardware changes rapidly. All hardware formats are subject to decay. Magnetic formats such as floppy discs, tape or hard discs are notoriously fragile in the long term, while optical media like CDs and DVDs, though often suggested as a more robust technology, are also prone to long term decay. It is necessary to think about the physical medium and the hardware that supports it. It is advisable to employ a medium term plan to move the data onto new discs from time to time, before the reliable life of the discs has expired. In practical terms it is unlikely that a data creator would suffer from problems of data refreshment, so long as their backup strategy is robust. Refreshment becomes the responsibility of the data archive after the project is complete.

6. Migration

The architecture of hardware changes rapidly, but not as rapidly as software. Data created or gathered in a proprietary software format is hostage to the long-term viability of that brand. Certain types of file have been earmarked as industry standard formats, while in other cases there may be open formats available.

The following principles should be kept in mind:

- Preservation formats should be non-proprietary (i.e. Open Source) - that is to say that they do not depend on a specific software company.
- Formats should avoid compression, especially where this results in data being discarded, or use a 'lossless' compression format.
- Where such open source or uncompressed or lossless standards are not available, market leaders provide a safer bet than obscure or bespoke formats.
- Supply metadata with the files. Some data formats are in essence self documenting, others, like GIS files require metadata to be entered before being used, while others, like images, almost always need external metadata. The file format will have implications on the sorts of metadata required.

Useful Links and Related Resources

ADS / AHDS Guides to Good Practice - provide guidelines on the creation, preservation and reuse of a range of data types relevant to archaeologists.
http://ads.ahds.ac.uk/project/goodguides/g2gp.html

ADS Guidelines for Depositors - advice on archival file formats and metadata.
http://ads.ahds.ac.uk/project/userinfo/deposit.html