ELECTRONIC ARCHIVING AND MEDIEVAL GENEALOGY: AN UNLIKELY BUT USEFUL PARTNERSHIP

by Steven Edwards[1]

ABSTRACT

Electronics and computer technology develops with sometimes bewildering speed, so any form of e-archiving needs to be adopted with great care. A regular programme of review, refreshment, data transfer and migration to new platforms when necessary is vital to ensure that all digitised forms of data storage remain accessible and free from corruption over long periods of time. The paper suggests how electronic archiving might be applied to materials stored within the FMG.

Foundations (2003) ***1*** *(2): 106-110* *© Copyright FMG*

Despite its title, The Foundation for Medieval Genealogy (FMG) is a 21st Century organisation intent on making best possible use of the latest digital technologies to realise its objectives. Specifically, this includes a major focus on development of the FMG website[2], which has both open access and password protected sections, together with publications in electronic format that are available either via the website or on CD-ROM. In addition, the FMG is building up a significant, specialised library of current and historical works, many of them obscure and difficult to locate. Although we fully intend to preserve the originals in excellent physical condition, we are also looking for ways to make the texts more accessible to members anywhere in the world.

The historic assets themselves and the results of new research published or coordinated by the FMG should lend themselves to long term storage in some type of electronic format. This paper summarises the results of a short project that I undertook using mainly web-based research to find out about state-of-the-art e-archiving. Starting with current technology, I have set out the steps necessary to ensure long term stability of electronically stored material.

With possible rare exceptions, the FMG will not be the repository of primary and irreplaceable source materials such as medieval manuscripts. The electronic copying and e-archiving of such materials is highly specialised and should be left to expert curators. The materials that the FMG may need to deal with fall into three main categories:

- **Administrative records**
  These include the constitution, minutes of meetings, correspondence, financial and membership records etc. for the Foundation. Although typically generated electronically, and preserved on disc with backups, the primary archive of these will, for the moment, remain the hardcopies held at the registered office of the FMG. Electronic versions may be retained for convenience, but will not be subject to the more rigorous archiving maintenance procedures described below. Eventually an e-

---

1 As well as editing ***Foundations*** the author is the Honorary Secretary of the Foundation for Medieval Genealogy and one of its founding trustees.

Contact details: see cover pages of this journal.

2 http://fmg.ac/

archive of the more significant papers may be created as a record of the FMG's establishment and early activities.

- **Digitised copies of paper genealogical source materials**
  These will normally be produced by scanning, subject to the rules of copyright. The electronic copies may be considered as secondary archives. The purpose of digitising documents is to improve their accessibility for FMG's members and users for study purposes. Where the primary paper source is also held by the FMG, then it will be subject to the Foundation's policy on library holdings, and will be treated as an historical asset that should be retained indefinitely.

- **Electronic datasets generated in the course of the FMG's projects**
  The FMG will be producing genealogical outputs from its sponsored projects, either as narrative articles or in some type of database format. The Medieval Genealogical Registry is a good example of the latter. These are cases where the data is held in primarily electronic form. The Foundation should establish a clear archiving policy and data storage plan for these materials.

Approaches to e-Archiving

A clear distinction needs to be drawn between routine maintenance of current electronic files and the establishment of electronic archives. Current file maintenance should include a procedure for regular data backup with storage of backups at a separate physical location to guard against fire, flood, theft, etc. E-archiving is quite different in that it involves long term storage in secure conditions, with retrievals limited to a minimum, as the act of retrieval itself may be a threat to data integrity in some circumstances. The issues have been well summarised by Suzanne Keene[3].

Electronic data differs fundamentally from physically archived material, in that it is not enough to simply preserve the storage media. The method of data retrieval must also be borne in mind. Whilst experts may puzzle over the interpretation of Linear A and B scripts or Mayan stele, the eye can still read them hundreds or thousands of years on. In contrast, within 5 years even the simplest digitally-encoded compressed image may never again be accessible because the machines to read them have become obsolete. Betamax video tape systems, for example, were popular and widely available only a few years ago, but the tapes are now largely inaccessible even if physically intact. Will VHS follow as DVD takes over? Similarly, 5¼" floppy discs for computers are now obsolescent and we may soon lack the technology to read them. Who remembers 8" or 3" discs? The 3¼" disc is still popular but rapidly becoming supplanted by CD technologies, while the plug-in solid-state memory chips that are just starting to become available may be commonplace in 10 years time.

The internet, and its associated service providers, is a dynamic system not suitable for archiving purposes. Material placed by the FMG on internet sites must always be backed up by locally held files and, where appropriate, FMG archive copies. Although there are now "web archive" sites (eg http://web.archive.org, cited by Marks and Kahle, 2002) that maintain a snapshot view of the whole worldwide web at a specific date and time, this cannot be relied upon as an indefinite resource. Importantly, such facilities do not preserve password protected sites.

---

[3] http://www.suzannekeene.info/conserve/digipres/index.htm

The required elements of any e-archiving policy are complex, and have been analysed in detail by those who are involved in serious archiving on a national scale (see for example Fresko, 1996[4]). At present it appears there are no agreed standards for digital archiving, although pressure for their development is growing.

An overview of principles of e-archiving recommended for use in the FMG, which could equally be applied to any other small genealogical or local history society, are:

(a) keep it simple;

(b) keep it consistent (all work of a similar nature should use the same software systems);

(c) use mainstream systems by major software suppliers that are likely to provide future continuity (Bennett, 1997[5])

(d) establish and operate a regular programme of review, refreshment, transfer and migration to new platforms when necessary.

A proposed e-archiving plan for the FMG

Currently, there are three main features that should be incorporated into an FMG archive plan:

The choice of physical storage medium

A medium must be chosen that is durable and resistant to some fluctuations in environmental conditions. Magnetic storage on tape or disc is highly vulnerable to stray magnetic fields and the physical media can fail without warning. Optical discs are probably the most secure form of digital archiving with present technology. CD-ROMs have life expectancies of over 70 years, they store large amounts of data in limited physical space, and they are not susceptible to magnetic fields. CD-R discs may be less stable, but still have a long life expectancy. They have the advantage over CD-RW discs that the data cannot be overwritten and so seem to represent the most accessible form of archive for the near future for small scale use such as the FMG e-archive. Archive discs should be kept in dry conditions, protected from light, avoiding extremes of temperature and, in particular, widely fluctuating temperatures. Special archive quality wallets and corrosion inhibiting film are important for physical conservation. Sticky labels should not be attached as the adhesive can react chemically with the plastic. Discs should not be subject to any bending pressure nor should hard tipped writing implements be used for labelling. It is strongly advised to create duplicate archive discs and to store them in physically separate locations.

The choice of software system to encode the data

The data needs to be encoded to include sufficient information, sometimes referred to as “metadata”, to enable it to be read by future technologies. It may sometimes also be useful to provide a “viewer” program linked to the data, although that presupposes the

---

4 http://www.ukoln.ac.uk/services/papers/bl/rdr6238/paper.doc

5 http://www.ukoln.ac.uk/services/papers/bl/jisc-npo50/bennet.html

viewer will itself operate in future systems. A printed version of the metadata should also be stored with the archived item.

A maintenance and review plan

Archivists must ensure that each archived item is physically and functionally intact. This should include a process of technological upgrade to maintain accessibility of the material as hardware and software systems develop, but without compromising the data integrity. A maintenance and review plan for the FMG should include three elements, review, refreshment and transfer and migration:

Review

Each item within the FMG archive should be formally reviewed every year. Both the review and its outcome should be recorded. This does not require recovery of the data from the archive, just a consideration of the following questions for each archived item:

- Q Is the software and operating system used to store the data still in current use by the FMG?
  A: yes – no further action required
  A: no – mark the archive for migration

- Q Is the media showing signs of physical deterioration or obsolescence?
  A: no – no further action required
  A: yes – mark the archive for refreshment or transfer

Refreshment and Transfer

Refreshment or transfer should be carried out at 5-year intervals, regardless of the state of the media. Refreshment involves making an identical copy of the archived data onto a new item of the same medium. Transfer involves copying the archived data, in identical electronic format, to a more stable carrier (for example transfer of archives from floppy disc to CD-R discs).

Migration

This is critical to the long-term retrievability of the archive. Migration solutions are still in their infancy, and the FMG archivist should maintain a watching brief on developments in the field. In essence, migration comprises the conversion of the data to the standards of data structure in current live use, followed by re-archiving in the new format (Wheatley, 2001[6]). Interestingly the at-the-time much heralded BBC Domesday Project of 1986 is already inaccessible by current computers. Fortunately a wise choice was made for storage media, namely videodiscs, and the CAMiLEON Project[7] has developed software that emulates the original computer and disc player, thus enabling the discs again to be read. Emulation is an alternative to migration, particularly relevant to complex information systems such as the Domesday Project, whereby a modern software system is developed that enables the original stored data to be read without further transformation (with its accompanying risk of corruption).

Industry standard software such as Windows operating systems and Microsoft Office packages usually have upwards compatibility built in, enabling migration of small data

---

6 http://www.ariadne.ac.uk/issue29/camileon/intro.html

7 http://www.si.umich.edu/CAMILEON/domesday/domesday.html

systems to be done easily. Verification is vital, but may be difficult to prove. Word processors and spreadsheets often provide two-way compatibility – the upgraded data can be re-converted to the historic format. Theoretically, it can then be compared with the original although full 100% compatibility in both directions is unlikely. This facility may not always be available, particularly with databases. In practice, some simple checks should be carried out to establish the data integrity in all its aspects in the upgraded format. Caution should be taken with so-called "standard" file formats such as RTF or TIFF, which have a number of variants that are not all readable by all systems.

A fundamental archiving principle is the prior definition, and inclusion within the metadata, of precisely which aspects of the material should be preserved. Thus, for textual documents, the most important facet is the text itself and considerations of layout, font style and so on are secondary. In this case, provided the core ASCII or Unicode coded text is preserved, nothing else matters. Things become more complex when the document contains embedded items such as tables, charts or pictures. Where possible, the embedded objects should also be archived as stand-alone file objects, with the necessary explanatory metadata. For digital images (be they from originals of manuscript, printed text, or pictures) the main requirement is to be able to reconstruct a physical facsimile of the original source. Most complex of all are databases, as both the data themselves and the inherent relationships or data structure must be preserved.

Archiving Timescales

The general principle should hold that, once material is allocated to the archive it is held "in perpetuity". This should not be open to review. Although no absolute guarantee, this reduces the danger of individuals or groups making decisions to discard archive material as of no current interest, simply in response to the whims of fashion or pressures of finance. A historical example might be a hypothetical 13th century librarian who decided that the Domesday Book was no longer relevant and threw it out. Closer to home, many of us have regretted similar acts by members of our family discarding papers or photographs that we would have wished to keep.

It is proposed that items become subject to the archiving policy if they are expected to have an academic value for at least 25 years into the future. The policy should be targeted at data management and preservation over a 50-year period. Beyond 50 years, the potential developments in technology are impossible to predict.

References

Marks, Paul & Kahle, Brewster (2002, Nov 23). Way back when… *New Scientist* (2370) 46-49.

Fresko, Marc (1996, editor). Long term preservation of electronic materials. *Proceedings of the Joint Information Systems Committee (JISC)/British Library workshop,* University of Warwick, November 1995.

Bennett, John C (1997). JISC/National Preservation Office (NPO) studies on the preservation of electronic materials. *British Library Research and Innovation Report* **50**.

Wheatley, Paul (2001). Migration - a CAMiLEON discussion paper. *Ariadne*, Issue 29.