Digital legacy: Respecting the digital dead

Updated 10:47 06 May 2011 by **Sumit Paul-Choudhury**
For similar stories, visit the **Death** Topic Guide

Read more: "Digital legacy: The fate of your online soul"

It's what's inside that counts (Image: Rex Features)

At first glance, there's nothing remarkable about the computer desktop on the screen in front of me. Its slightly psychedelic wallpaper is half covered by standard, if slightly old-fashioned icons. Many have names that relate to genetic analysis, hinting at their owner's profession, while others are more obscure, perhaps relating to personal affairs. Feeling a bit like an intruder, I watch as folders open, documents pop up, applications start.

The desktop is not just anybody's. It was owned by the late evolutionary biologist Bill Hamilton. When he died, his family were at a loss as to what to do with his digital archives, so they passed them to his friend and former student Jeremy Leighton John, curator of e-manuscripts ("eMSS") at the British Library in London.

Hamilton's is one of about a dozen digital legacies curated by John. He now heads the library's effort to understand the technical and ethical challenges associated with preserving such materials for posterity. That's becoming an increasingly pressing issue: last month, the library hit the headlines when it paid £32,000 to acquire 40,000 emails from the poet Wendy Cope, along with her digital archive

Copied computer

What John is showing me now in the library's eMSS lab is actually a simulation of Hamilton's computer. The actual machine is safe and secure in a vast storage area eight storeys below – and may in due course be returned to Hamilton's family or passed on to another institution.

Researchers who want to understand Hamilton's work will be able to use the simulation to explore his records and even re-enact his analyses without going near the machine. This elaborate arrangement is necessary, says John, because otherwise it's all too easy to contaminate the original data. "As soon as you turn a computer on, you start changing dates," he says. That could destroy clues that future researchers will need to piece together, such as the origins of a notable piece of work.

Not much software has yet been developed for the analysis of digital archives, so John has borrowed many techniques from digital forensics, which takes a similar approach: "Don't change the information; analyse the information without changing it; and show that you haven't changed it." Forensic tools, unlike standard-issue computer maintenance or disc analysis software, are built with these principles in mind – and observe them well

enough for their findings to stand up in court.

Untouchable

But preserving all this material raises its own problems: what should be kept off-limits to researchers? Attitudes vary, says John. Just as some people don't mind having all of their papers open to scrutiny, there are some who are happy to let researchers look through all their digital data. But there are others who would rather have their privacy closely protected.

The problem is that few of us make the distinction between private and public as we store and label files – and given the vastness of our digital estates, it's unlikely that our inheritors will have the time to tidy them up for us. They may not even be able to: John recalls a family who donated a stack of 5¼-inch floppy discs to the library without any real idea of what they contained.

"Archivists tend to see themselves primarily as mediators between the creative individual – the originator of the archive – and the researcher," says John. "So the archivists help protect the privacy of the individual while helping to ensure the authenticity of the information on behalf of the researcher." That might include removing credit card numbers or telephone numbers before releasing the files; forensic software can be used to identify where such sensitive information exists.

In some cases, the creator or donor may prefer that some information be removed entirely, but that has its own complications. John has suggested that organisations like the British Library may need to provide donors with a copy of any digital object before they delete the donated copy, for whatever reason. That way, the library can delete material that seems irrelevant or personal without destroying the existing copy. "We wouldn't want to be blamed for deleting the last copy of one of Einstein's emails, as it were, whatever the justification might be," says John.

Tools for the job

John suggests that part of the solution is technological: we need better tools to help us curate our own data as we go along. "It's amazing that software companies haven't really got to grips with personal information management," he says. The files-and-folder system used by most of us has profound shortcomings, even when supplemented with powerful search capabilities: a third of participants in a recent survey conducted by the British Library and University College London reported a serious loss of data at home. Of these, more than two-thirds had simply misplaced files; hard drive failure, though much feared, accounted for less than a tenth of such losses. Tools that allow us to organise our data more intuitively would not only help to prevent such painful losses, but also make the work of archivists easier.

John, like many other archivists, hopes that standards will emerge that will make it comparatively easy to navigate, preserve and curate a digital bequest. In the past couple of years, this has gone from a theoretical to a practical concern, with John's eMSS lab sharing a floor with the library's digital preservation team and the Open Planets Foundation, a not-for-profit organisation dedicated to developing workaday digital archiving tools.

What might those look like? A promising two-pronged approach, says John, is to develop file formats and software that cater to the particular needs of archiving and preservation and can be run on "universal virtual machines". That would allow digital disc images, like that of Hamilton's Macintosh, to be booted up without being tied to any particular platform.

The virtual machine can be repeatedly "ported" to newer machines, ensuring the files will remain accessible even as the original hardware becomes obsolete or unusable.

If all of that works, a researcher at the British Library may one day experience the spooky sensation of poking around my desktop long after I am gone. And perhaps yours too.

Sumit Paul-Choudhury *is the editor of newscientist.com*

› Back to article