There need not be a digital dark age—how to save our data for the future

February 24 2015, by Matthew Woollard

Floppies: storage that's about as reliable as a CD used as a frisbee. Credit: orangejack, CC BY-NC-SA

"The internet is forever." So goes a saying regarding the impossibility of removing material – such as stolen photographs – permanently from the web. Yet paradoxically the vast and growing digital sphere faces enormous losses. Google has been criticised for failing to ensure access

to its archive of Usenet newsgroup postings that stretch back to the early 1980s. And now internet pioneer Vint Cerf has warned of a "digital dark age" that would result if decades of data – emails, photographs, website postings – becoming lost or un-readable.

Millions of paper records more than 500 years old exist today. But your entire family photo collection could be lost forever with just a single hard drive failure. Stone tablets, parchment, paper, printed photographs have all lasted through the centuries. But some of our data may not. What do we do about preserving the digital deluge?

Cost v value

Technical solutions already exist, but they're not well known and relatively expensive. How much are we prepared to pay to ensure that digital stuff today is usable in the future? Because if there's cost involved, inevitably we have to think about what has value that makes it worth keeping.

How can we calculate that value? As an example, the holdings of the UK Data Archive include machine-readable versions of all of the General Household Surveys (GHS) carried out between 1971 and 2011. This was a continuous national survey of people living in private households conducted on an annual basis. The cost of the GHS in 2001 was reported as £1.43m, making the value of the survey and its data at least that. As it was the thirtieth year of this survey the value could be said to be higher as it was part of a series, so we could say they survey was worth more than it cost.

The Office for National Statistics transferred the 2001 data to the UK Data Archive in 2002, where we prepared them for preservation and access and published them. Up until today this survey data has been downloaded by 426 people working in government departments, 759

staff working in education, 1,331 students and 109 others for various uses. So benefits accrue from making the data available even after its creators have exhausted their primary value – re-use is a significant benefit from preserving data and adds value.

But there are also cultural and intellectual and not just economic arguments for preserving data. Survey data like these and their supplementary materials provide a window to the concerns of survey designers and, by extension, society at the time. True, cultural arguments for preservation can be expressed more forcefully for artefacts such as images, films, or written works than survey data. But these data stand a good chance of being included within Britain's cultural and intellectual heritage precisely because they have been carefully managed and preserved.

Making digital as long-lasting as paper

How can we improve the chances of something being preserved? Professor Michael Clanchy, writing in his seminal From Memory to Written Record, discusses how the concept of records developed. Owing to the media available to scribes in the Middle Ages they made conscious choices between creating an ephemeral document (on a wax tablet) or a permanent record (on parchment). Today digital media proliferates mainly because it provides the easiest means to transmit a work, and so that distinction has to a point disappeared.

Documents and records are now both digital, but the question remains as to what should be kept for posterity and why. These are hard questions which lead to hard choices, because by their nature the cost of preserving digital materials can be much more expensive than their analogue counterparts. You can't just put them in a box and walk away – the effort and tools required to read a 100-year-old letter is considerably less than the effort required to read a 30-year-old LocoScript popular on

Amstrad computers in the 1980s-90s.

Most born-digital material is, with the right resources, recoverable. However, the chances of born-digital material being usable in, say, 100 years is considerably improved by actively taking steps to ensure that it will – just as medieval scribes made similar decisions in centuries past. Effective digital preservation relies, to some extent, on the activities of the creator as well as the archivist. Today those decisions include providing context, using standard and open file formats, organising material sensibly, and making provision for rights issues to avoid the problem of orphan works.

The future starts now

Organisations can do a better job than individuals, but require a business model and a mandate to do so. Asking someone to pay for something a long time before its value can be realised (if at all) is not an attractive business proposition. What we can do, at a minimum, is try and convince people that it is possible.

Of course neither creator nor archivist can fully understand how future users may approach digital information preserved over time. Social and cultural historians have, by necessity, used records for purposes for which they were not created and often in inventive and interesting ways. Historians are often helped by context, and the digital material we're creating today needs the same contextual information to ensure its usefulness.

This story is published courtesy of The Conversation *(under Creative Commons-Attribution/No derivatives).*

Source: The Conversation