

ASIS&T 2011 PLENARY SESSION

Preservation: The Final Frontier

by Steve Hardin

EDITOR'S SUMMARY

Tom Wilson, professor emeritus at the University of Sheffield, explored the perils and challenges of information preservation in a talk at the 2011 Annual Meeting. Wilson described 2011 technology as quaint from a future perspective and retraced the evolution of known document formats from cave painting and papyrus through clay, wood, paper, microfilm, laser discs and today's variety of digital record formats. Not only is the continuing evolution in formats a challenge in itself, but the content of bygone eras may not be comprehensible to those who follow. Decisions about what to preserve come into play, involving judgments at a point in time that may not be the same from a different temporal viewpoint. These issues, together with migration through versions of hardware and software, suggest keeping lots of copies in safe places, leading to impossibly huge amounts of data to store indefinitely. Wilson mused that reverting to paper, following digital librarian Brewster Kahle's lead, might be our best option.

KEYWORDS

materials preservation
physical media
electronic documents
emulation
migration

Steve Hardin is the interim chair of reference/instruction at Cunningham Memorial Library, Indiana State University, in Terre Haute. He can be reached at Steve.Hardin@indstate.edu.



Whenever a decision is made to preserve information for the long haul, things get complicated quickly. Plenary speaker Tom Wilson addressed a number of issues, successes and failures surrounding preservation at the 2011 ASIS&T Annual Meeting on October 9, 2011, in New Orleans. Wilson, a leader in information behavior research, is professor emeritus at the University of Sheffield in the

United Kingdom (t.d.wilson@shef.ac.uk) and has been a visiting professor at universities in several countries.

He began with a fairy tale from the future, told to children in the year 3511. "Once upon a time, it was normal to keep records on paper [laughter from children]. People worried that paper might decay or be destroyed, so information was transferred to computers [puzzled noises from children]. They weren't like the intelligent robots we have today. Messages were sent along thin strands of metal called "wires." As you children can imagine, much of what was recorded then has been lost. Useful things survived, but we don't know who created them. Newton, Einstein and Dalton – they may be the names of research machines. I think that now, what we know will be preserved forever, because everyone knows everything."

Wilson pointed out that what we know of the past is largely a matter of accident. No previous society has managed to survive a record of its culture. What we have is only a small percentage of what was created. We have works from only four Greek dramatists. Aristotle mentions many more, but we know next to nothing about them.

Tomb paintings survive because they were designed to be hidden. Cave paintings and stone carvings appear to be excellent preservation media. Paintings more than 30,000 years old have been discovered, but we don't know what they mean. They appear to be works of art, but maybe they were works of magic: perhaps the artists drew the animals to have success in hunting. More than 2200 years ago the Rosetta Stone attempted to establish the divinity of Pharaoh Ptolemy V in multiple languages. It provided us with the clue to hieroglyphics.

Papyrus records and clay tablets survive. Wood is also a good preservation medium if you find it in the right place. Wilson showed an image of Hadrian's Wall in Northumberland. Romans wrote letters and lists on wooden tablets about the size of a postcard. About 700 have been found in the midden. They are now regarded as the British Museum's greatest treasure. He showed an image of the "Birthday Letter," from one Roman officer's wife to another, inviting her to attend her birthday celebration. Part of it is in the inviter's hand – the only surviving example of a Roman woman's handwriting.

Paper was invented in China. Some artistic cutouts have survived 1500 years.

To summarize, Wilson said, physical materials can have very long lifetimes, if conditions are right. Preservation is of little value if what is preserved cannot be interpreted. The biggest threats to preservation come from calamities: natural and man-made.

Wilson noted a digital record is also subject to the decay of the media on which it is recorded. DVDs may last 100 years – we don't know yet. Digital linear tape may last 300 years. Ordinary magnetic tape can have a very short life – cassettes from the 1970s can lose their magnetic coating. Microfilm may last 500 years. Photographic slides may last 100 years. Acid-free paper may last 1000 years, even longer. But there are intelligibility issues. There are two aspects to the problem, he said. First, there are obsolete file formats. His word processor offers him 19 different file formats for saving. There are more than 80 image formats. There are in fact thousands of file formats. How many of them are obsolete? And what about video and film formats? Second, how can we understand what is in the files? For instance, how many English speakers understand Anglo-Saxon?

And materials in that language are only 1000 years old. So what are the chances that 1000 years from now that what we consider to be standard English, French and so forth will be understandable? Language changes all the time. Wilson listed some common occupations of 100-150 years ago, including acreman, backman, baller, badger, beltonist, boniface, boothman, bradener, busker, cafender, owler, pleacher, puggard, souter. How many English speakers know what those words mean today?

The entertainment industry is buzzing about 3D TV these days. Wilson noted that an earlier form of 3D television, called stereoscopic TV, first appeared in August 1928 (see www.bairdtelevision.com). People wouldn't know about this earlier form of 3D TV if they didn't know to search for "stereoscopic television."

So, Wilson said, we have to consider the hardware, the software, the record format(s), standards and physical security of preserved records.

Wilson related a story of preservation gone wrong: the BBC's Domesday Project from 1986. The original Domesday Book was a set of tax records. The 1986 project was to collect information from schoolchildren about the state of Britain. Project designers collected 148,000 pages of text and 23,000 photographs. They were recorded on large videodiscs – the "technology of the future" at the time. It was about the only thing then that could handle the volume of material. The BBC made a bet on the future. But videodiscs didn't succeed on the market, and were subsequently withdrawn. Soon the discs were unreadable by anyone other than the original project designers. The information was lost until 2011, when a process was established for retrieving the information from the discs. Now, the Domesday Reloaded project (links from www.domesday.org.uk) lets you search this information by place and discover what life was like in England in 1986. Wilson said that Andy Finney, who authored that page, maintains that the only way to preserve data is to recopy it onto whatever technology is dominant at the moment. Migration is an old strategy for preserving things for the future – copyists copied decaying papyrus onto parchment, for example – but mistakes creep in. The ability to copy something onto a fresh medium doesn't guarantee that everything will be copied correctly.

Another option: LOCKSS (Lots of Copies, Keep Stuff Safe). If you have

enough copies scattered around the world, losing one copy won't mean the loss of the information. Copies are produced only when you need them. But since only a small percentage of the stuff in an archive is actually used, how long will it be before the unused stuff is lost?

There's also emulation: the idea that computer systems of the future will be able to emulate computer systems of the past. It's being done – one example is the SHAMAN Project funded by the European Union (www.shaman-ip.eu/). Emulation can present files as if they were being presented by the original software. But to do this emulation, you must preserve not only the history, but also the metadata that explains how the older information works. How much effort can we afford to expend on this level of documentation?

And consider the problem of what should be preserved. Wilson said it seems that the unspoken assumption is that everything should be preserved. But is that right? We get along quite happily knowing lots of older information is lost. For instance, we know the leading authors of the past, but not the minor ones. Is that OK?

The 2011 Digital Universe Study (www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm) forecast that by 2011 the digital universe would be 10 times the size it was in 2006, reaching 1800 exobytes (EB) or 1800 billion gigabytes. That's almost the number of stars estimated to be in the galaxy. The study says that in 2007, we reached a

tipping point in that the amount of information being created exceeded the amount of storage to handle it. The gap will only widen. Of course, not everything produced is kept.

Wilson said we don't know what will be of value in the future. Consider Charles Dickens, who got terrible reviews at first, but who's considered a great author today. We value things Victorians didn't, and they valued things we don't.

Preservation is a problem for the immediate future. The first programmable computer – Colossus – was developed in 1944. That was not that long ago, and look what has changed since then in the lifetimes of many people sitting in the audience. In his own life, he's used mainframes, minicomputers, early microcomputers, PDAs (personal digital assistants), laptops, iPads, even analog computers. Do we really expect that things the way they are now will persist into the indefinite future? He doubts that a single mode of transfer will win out. There will be multiple forms of migration and emulation. There's an increasing realization that we can't depend on digital technology for everything. Brewster Kahle converted a lot of digital information to paper because he worried about how digital preservation will work.

Wilson concluded that no matter what we do, we cannot guarantee the persistence of our culture – or our records – over time. If there are things we wish to see endure, we might phone Brewster Kahle and ensure we have enough paper – or perhaps employ some stonemasons. ■