

DIGITAL DOOMSDAY: THE END OF KNOWLEDGE

by Tom Simonite and Michael Le Page

New Scientist February 2, 2010

"IN MONTH XI, 15th day, Venus in the west disappeared, 3 days in the sky it stayed away. In month XI, 18th day, Venus in the east became visible."

What's remarkable about these observations of Venus is that they were made about 3500 years ago, by Babylonian astrologers. We know about them because a clay tablet bearing a record of these ancient observations, called the Venus Tablet of Ammisaduqa, was made 1000 years later and has survived largely intact. Today, it can be viewed at the British Museum in London.

We, of course, have knowledge undreamt of by the Babylonians. We don't just peek at Venus from afar, we have sent spacecraft there. Our astronomers now observe planets round alien suns and peer across vast chasms of space and time, back to the beginning of the universe itself. Our industrialists are transforming sand and oil into ever smaller and more intricate machines, a form of alchemy more wondrous than anything any alchemist ever dreamed of. Our biologists are tinkering with the very recipes for life itself, gaining powers once attributed to gods.

Yet even as we are acquiring ever more extraordinary knowledge, we are storing it in ever more fragile and ephemeral forms. If our civilisation runs into trouble, like all others before it, how much would survive?

Of course, in the event of a disaster big enough to wipe out all humans, such as a colossal asteroid strike, it would not really matter. Even if another intelligent species evolved on Earth, almost all traces of humanity would have vanished long before.

Let's suppose, however, that something less cataclysmic occurs, that many buildings remain intact and enough people survive to rebuild civilisation after a few decades or centuries. Suppose, for instance, that the global financial system collapses, or a new virus kills most of the world's population, or a solar storm destroys the power grid in North America. Or

suppose there is a slow decline as soaring energy costs and worsening environmental disasters take their toll. The increasing complexity and interdependency of society is making civilisation ever more vulnerable to such events.

Whatever the cause, if the power was cut off to the banks of computers that now store much of humanity's knowledge, and people stopped looking after them and the buildings housing them, and factories ceased to churn out new chips and drives, how long would all our knowledge survive? How much would the survivors of such a disaster be able to retrieve decades or centuries hence?

Fogbank fiasco

Even in the absence of any catastrophe, the loss of knowledge is already a problem. We are generating more information than ever before, and storing it in ever more transient media. Much of what it is being lost is hardly essential -- future generations will probably manage fine without all the family photos and videos you lost when your hard drive died -- but some is. In 2008, for instance, it emerged that the US had "forgotten" how to make a secret ingredient of some nuclear warheads, dubbed Fogbank. Adequate records had not been kept and all the key personnel had retired or left the agency responsible. The fiasco ended up adding \$69 million to the cost of a warhead refurbishment programme.

In the event of the power going off for an extended period, humanity's legacy will depend largely on the hard drive, the technology that functions as our society's working memory. Everything from the latest genome scans to government and bank records to our personal information reside on hard drives, most of them found inside rooms full of servers known as data centres.

Hard drives were never intended for long-term storage, so they have not been subjected to the kind of tests used to estimate the lifetimes of formats like CDs. No one can be sure how long they will last. Kevin Murrell, a trustee of the UK's national museum of computing, recently switched on a 456 megabyte hard drive that had been powered down since the early 1980s. "We had no problems getting the data off at all," he says.

Modern drives might not fare so well, though. The storage density on hard drives is now over 200 gigabits per square inch and still climbing fast. While today's drives have sophisticated systems for compensating for the failure of small sectors, in general the more bits of data you cram into a material, the more you lose if part of it becomes degraded or damaged. What's more, a decay process that would leave a large-scale bit of data readable could destroy some smaller-scale bits. "The jury is still out on modern discs. We won't know for another 20 years," says Murrell.

Most important data is backed up on formats such as magnetic tape or optical discs. Unfortunately, many of those formats cannot be trusted to last even five years, says Joe Iraci, who studies the reliability of digital media at the Canadian Conservation Institute in Ottawa, Ontario.

Iraci's "accelerated ageing" tests, which typically involve exposing media to high heat and humidity, show that the most stable optical discs are recordable CDs with a reflective layer of

gold and a phthalocyanine dye layer. "If you go with that disc and record it well, I think it could very well last for 100 years," he says. "If you go with something else you could be looking at a 5 to 10 year window."

Gone in a flash

The flash-memory drives that are increasingly commonplace are even less resilient than hard drives. How long they will preserve data is not clear, as no independent tests have been performed, but one maker warns users not to trust them for more than 10 years. And while some new memory technologies might be inherently more stable than flash, the focus is on boosting speed and capacity rather than stability.

Of course, the conditions in which media are stored can be far more important than their inherent stability: drives that stay dry and cool will last much longer than those exposed to heat and damp. Few data centres are designed to maintain such conditions for long if the power goes off, though. A lot are located in ordinary buildings, some in areas vulnerable to earthquakes or flooding. And if civilisation did collapse, who knows what uses the resource-starved survivors might find for old hard

drives ?

The physical survival of stored data, however, is just the start of the problem of retrieving it, as space enthusiasts Dennis Wingo and Keith Cowing have discovered. They have been leading a project, based at NASA's Ames Research Center in Moffett Field, California, to retrieve high-resolution images from old magnetic tapes. The tapes contain raw data sent back from the five Lunar Orbiter missions in the 1960s. At the time, only low-resolution images could be retrieved. The tapes were wrapped in plastic, placed in magnetically impervious metal canisters and remain in pristine condition. "It is a miracle from my experience with similar commercial tapes of a similar age," says Wingo.

Biggest challenge

But to get the raw data off the tapes, the team first had to restore old tape drives saved by a former NASA employee. That was the biggest challenge, says Cowing. "There was a lizard living inside one of them." Once they began to retrieve the raw data, converting it into a usable form was only possible after a three-month search uncovered a document with the "demodulation" equations.

If today it takes a bunch of enthusiasts with plenty of funding many months to retrieve the data from a few well-preserved magnetic tapes, imagine the difficulties facing those post-catastrophe. Even with a plentiful supply of working computers to read hard drives, recovering data would not be easy. Much data nowadays is encrypted or readable only using specialised software. And in a data centre left untouched for 20 or 30 years, some drives would need disassembling to retrieve their data, says Robert Winter, a senior engineer with Kroll Ontrack Data Recovery in Epsom, Surrey, UK, which in 2003 rescued the data on a hard drive from the space shuttle Columbia.

Indeed, rescuing data if things go wrong can be tricky even in today's fully powered world.

Last year, for instance, after some servers malfunctioned, it took Microsoft many weeks to recover most of the personal data of users of Sidekick cellphones.

Post-catastrophe, the lack of resources -- of people, expertise, equipment -- might be a far bigger obstacle than the physical loss of data. And resources are likely to be scarce. Restarting an industrial civilisation might be a lot harder the second time round, because we have used up most of the easily available resources, from oil to high-grade ores.

Would the loss of most of the data stored on hard drives really matter? After all, much of what we have inherited from past civilisations is of little practical use: the Venus Tablet of Ammisaduqa, for instance, consists largely of astrological mumbo jumbo. Similarly, an awful lot of what fills up the world's servers, from online shops to the latest celeb videos, seems dispensable too.

Even the value of much scientific data is questionable. What use would it be knowing the genome sequence of humans and other organisms, for instance, without the technology and expertise needed to exploit this knowledge? With some scientific experiments now generating petabytes of data, preserving it all is already becoming a major challenge. The vast quantity of material will be a problem for anyone trying to recover whatever they regard as important: while it is relatively easy to find a book you are after in a library, there is usually no way to be sure what's on a hard drive without revving it up.

Top of the pops

What's more, what is likely to survive the longest from today's digital age is not necessarily the most important. The more copies -- backups -- there are of any piece of data, the greater the chances of its survival, discovery and retrieval. Some data is much copied because it is so useful, like operating systems, but mostly it is down to popularity.

That means digital versions of popular music and even some movies might survive many decades: Abba might just top the pop charts again in the 22nd century. However, there are far fewer copies of the textbooks and manuals and blueprints containing the kind of distillation of specialised knowledge that might matter most to those trying to rebuild civilisation, such as how to smelt iron or make antibiotics.

Perhaps the most crucial loss will occur after half a century or so, as any surviving engineers, scientists and doctors start to succumb to old age. Their skills and know-how would make a huge difference when it comes to finding important information and getting key machinery working again. The NASA tape drives, for instance, were restored with the help of a retired engineer who had worked on similar systems. Without expert help like this, retrieving data from the tapes would have taken a lot longer, Cowing says.

A century or so after a major catastrophe, little of the digital age will remain beyond what's written on paper. "Even the worst kind of paper can last more than 100 years," says Season Tse, who works on paper conservation at the Canadian Conservation Institute. The oldest surviving "book" printed on paper dates from AD 868, he says. It was found in a cave in north-west China in 1907.

A century or so after the power goes off, little will remain of the digital age except what's on paper.

Providing books are not used as a handy fuel, or as toilet paper, they will persist for several hundred years, brittle and discoloured but still legible. Again, though, the most popular tomes are the most likely to survive. Imagine risking your life exploring dangerous ruins looking for ancient wisdom only to find a long-hidden stash of Playboy magazines.

It is not just what survives but the choices of those who come after that ultimately decide a civilisation's legacy, however. And those doing the choosing are more likely to pick the useful than the trivial. A culture of rational, empirical enquiry that developed in one tiny pocket of the ancient Greek empire in the 6th century BC has survived ever since, says classicist Paul Cartledge of the University of Cambridge, despite not being at all representative of the period's mainstream culture.

As long as the modern descendant of this culture of enquiry survives, most of our scientific knowledge and technology could be rediscovered and reinvented sooner or later. If it does not survive, the longest-lasting legacy of our age could be all-time best-sellers like Quotations from Chairman Mao, Scouting for Boys and The Lord of the Rings.

Store it for millennia

The current strategy for preserving important data is to store several copies in different places, sometimes in different digital formats. This can protect against localised disasters such as hurricanes or earthquakes, but it will not work in the long run. "There really is no digital standard that could be counted on in the very long term, in the scenario that we drop the ball," says Alexander Rose, head of The Long Now Foundation, a California-based organisation dedicated to long-term thinking.

Part of the trouble is that there is no market in eternity. Proposals to make a paper format that could store digital data for centuries using symbols akin to bar codes have faltered due to a lack of commercial interest and the challenge of packing the data densely enough to be useful.

Perhaps the only data format that comes close to rivalling paper for stability and digital media for data density is the Rosetta Disk. The first disc, made in what its creators call 02008, holds descriptions and texts of 1000 languages.

The nickel discs are etched with text that starts at a normal size and rapidly shrinks to microscopic. At a size readable at 1000 times magnification, each disc can hold 30,000 pages of text or images. The institute is considering creating a digital version using a form of bar code.

If we did have a way to store digital data long-term, the next question would be what to preserve, and how to keep it safe but easily discoverable.