
Also in this section
68 Hazards of youth
68 Big video brother

Digital archiving

# History flushed

**The digital age promised vast libraries, but they remain incomplete**

IN 1086 William the Conqueror completed a comprehensive survey of England and Wales. "The Domesday Book", as it came to be called, contained details of 13,418 places and 112 boroughs—and is still available for public inspection at the National Archives in London. Not so the original version of a new survey that was commissioned for the 900th anniversary of "The Domesday Book". It was recorded on special 12-inch laser discs. Their format is now obsolete.

The digital era brought with it the promise of indefinite memory. Increased computing power and disk space combined with decreasing costs were supposed to make anything born digital possible to store for ever. But digital data often has a surprisingly short life. "If we're not careful, we will know more about the beginning of the 20th century than the beginning of the 21st century," says Adam Farquhar, who is in charge the British Library's digital-preservation efforts.

The most obvious problems for digital archivists have to do with hardware, but they are also the easiest to fix. Many archives replace their data-storage systems every three to five years to guard against obsolescence and decay. This is not as expensive as it sounds: hard drives are cheap and reliable. The threat of hardware failure is overcome by keeping copies in different places. The British Library has storage sites in London, Yorkshire, Wales and Scotland.

Collecting digital material is trickier, particularly online. Archivists can only harvest those parts of the web that are freely accessible. Anything requiring user inputs—passwords, searches, forms—is off-limits. Streaming media, such as online videos, are hard to capture.

Changes in software and file formats create more hurdles. "Many of the digital objects we create can only be rendered by the software that created them," says Vint Cerf, a pioneer of the internet who now works for Google. If the original program has gone, an archive of mint-condition files can be useless. By the time software is more than a decade old, running it usually requires hardware emulation—essentially fooling programs into thinking that they are running on old hardware.

Although technical problems can usually be solved, regulatory obstacles are harder to overcome. Laws force copyright libraries, such as the Library of Congress, to seek permission before archiving a website. Regulation can be even more damaging when it comes to preserving such things as computer programs, games, music and books. These often come with digital-rights management (DRM) software to protect them against piracy. Archivists who want to circumvent such programs can find themselves on the wrong side of the law. America's Digital Millennium Copyright Act (DMCA) makes such circumvention a criminal offence.

Copyright and DRM will loom even larger as the nature of information systems evolves. The original internet was by default an open environment, making copying easy. The mobile world, with its widely popular smartphone apps, is much less so. As companies more fiercely protect their wares, contemporary digital artefacts run the risk of never being archived. Libraries have no mandate to collect apps, such as Angry Birds or Instagram, which form part of popular culture.

Despite all these difficulties, the world's libraries have tried for over a decade to conserve some aspects of their national digital heritage. America's Library of Congress started its digital-preservation programme in 2000 with $100m from the government. Its web archive currently stands at around 10,000 sites, many of them owned by the American government, and therefore exempt from copyright. Privately run sites are more difficult to include. For some archiving projects, only a fifth of webmasters reply to e-mails seeking permission for a copy.

### Digital pack rats

Following the Library of Congress, most national libraries in rich countries now have some sort of digital-archiving programme. In Britain, for instance, the National Archives keeps copies of all government websites. The British Library is archiving all British online material.

Yet the best-known digital preservation effort is the Internet Archive, a private non-profit effort. Its servers are home to the Wayback Machine, a popular web service that lets users see how a website looked on specified dates in the past. Founded by Brewster Kahle in 1996, Internet Archive collects, stores and provides access to billions of web pages as well as other digital media such as books, video and software. ▸▸

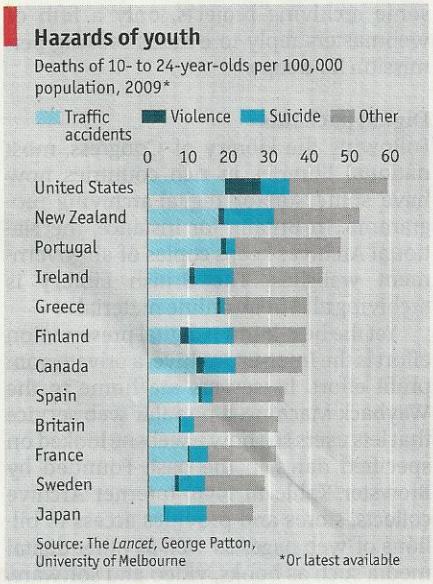
The collection stands at roughly 160 billion web pages. It operates on the principle that it is better to seek forgiveness than to ask for permission.

More recently, geeks have rushed in where official agencies fear to tread. They have always been pack rats. Today they gather on websites such as TOSEC (short for "The Old School Emulation Centre") to collect old software. But these collections have their own limitations. They focus heavily on games and operating systems; people tend not to have the same nostalgia for early versions of spreadsheet applications as they do for Super Mario Bros. More important, the material is very much under copyright.

Despite the proliferation of archives, digital preservation is patchy at best. Until the law catches up with technology, digital history will have to be written in drips and drabs rather than the great gushes promised by the digital age. ■

### Teenage wastelands

**Adolescence is a difficult time. But in some countries it is more difficult than in others—and more dangerous. Within the rich world mortality rates for teenagers and those in their early 20s vary widely, as do the factors explaining these rates, according to data collected by George Patton of the University of Melbourne for an article in the Lancet, published on April 25th. In Greece suicides are rare, but car accidents common. In Finland it is the other way around. America stands out for having the highest mortality rate, more than twice that of Japan. America has a particularly high rate of traffic deaths, despite laws that bar drinking until the age of 21. Where America is truly exceptional, however, is in its violence. The rate of violent death in America is 18 times that in Britain.**




### Video surveillance

# I spy, with my big eye

**Face recognition is good news for the police, but bad news for privacy lovers**

WELCOME to China, the land of video surveillance. Guangdong province boasts over 1m cameras. In 2010 the city of Chongqing, governed by the now-disgraced Bo Xilai, ordered 500,000. Other provinces have hundreds of thousands, according to Human Rights in China, an NGO. Video surveillance constitutes over half the country's huge security industry, and is expected to reach 500 billion yuan ($79 billion) in 2015. China will soon overtake Britain, with around 3m cameras, as the capital of video surveillance.

Yet China is far from alone. In many democracies surveillance cameras are multiplying, too. And face-recognition technology is proving a wonder tool for both governments and marketers.

A jail in Alabama uses it to check those leaving against prisoner records. Mexican prisons use it to identify visitors. Heathrow airport is installing systems to track passengers through lounges and onto the plane. Brazil has plans to equip police with camera-spectacles that can identify troublemakers at the 2014 World Cup.

As for businesses, Quividi, a French marketer, can measure the age and gender of passers-by who linger at an advert; advertisers vary their offerings based on who is looking. A service called SceneTap gives similar information on the crowd in Chicago bars. The smiles of employees at Keihin Electric Express Railway in Japan are assessed by computer. Facebook, a social network, recognises uploaded photos. The latest smartphones can spot their users.

The technology is improving fast. In 2010, in an assessment by America's National Institute of Standards and Technology (NIST), the best program matched 92% of mugshots to one out of 1.6m pictures. Such results require high-quality still photos, stresses NIST's Jonathon Phillips. But progress continues on fuzzier moving images. The error rate halves every two years—and not just in the West. In the 2010 NIST tests Chinese entrants lagged behind, identifying just 64% of images. But their systems are now "state of the art", says Sharon Hom of Human Rights in China.

System performance depends on the context. Controlled environments, such as jails, are ideal. An experiment in Mainz railway station in Germany got steady shots by mounting cameras over escalators (although the recognition rate only reached 60%). Putting the lens behind an advert is a good way to get subjects who are facing it. Facebook is good at recognising people because they pick names from a limited list of friends.

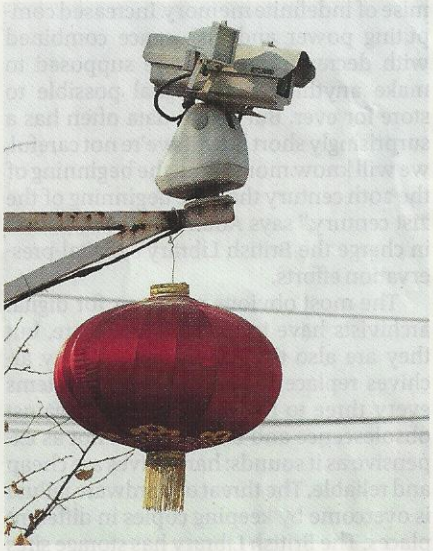
More cameras and better face recognition raise tricky legal and political questions. America places little restriction on the use of face recognition, as legal precedent denies the "reasonable expectation of privacy" in public. But Harley Geiger of the Centre for Democracy & Technology, an advocacy group, says the technology goes beyond normal public scrutiny and could create a world where everyone, in effect, becomes "a public figure".

The industry is aware of reputational risks. Eric Schmidt, Google's chairman, has said that Google will limit its face-recognition services—to avoid "crossing the creepy line". Last year the Digital Signage Federation, a trade group, adopted a strong set of face-recognition privacy standards.

Privacy-loving European countries are less easy-going, and usually require cameras to be matched with signs to tell people they are being watched. Facebook's face recognition has already fallen foul of tough German privacy laws. And America's Supreme Court is uneasy with technology that enables the persistent tracking of individuals in public.

Still, even democratic governments will want to monitor people as technology improves. But losing public anonymity could affect political life. Freedom of speech is reduced when mere physical attendance at protests goes on record. Kelly Gates, the author of a book on surveillance, sees a "chilling effect".

Some countries welcome this. Nicholas Bequelin, a researcher for Human Rights Watch, another NGO, says authorities installed thousands of cameras in Xinjiang province, in China, after riots there in 2009. Its strategy for stability, Mr Bequelin points out, is "to nip protests in the bud". Video surveillance seems the ideal tool. ■



**Keeping an eye on Ai Weiwei**