

Lost in cyberspace

Technology and society: Amid the explosive growth of digital content on the internet, little thought has been given to preserving things for posterity. Will historians of the future wish that web pages had been preserved more carefully?



Tracking down early web pages can be a problem. The Economist's first website, for instance, was built by the paper's California correspondent and went live in March 1994. Eighteen months later, it was reconfigured and brought in-house. All records of the original website were subsequently lost. So much for the idea that the internet never forgets. It does.

There are not even any screen shots of the world's first web page—the one that actually launched the World Wide Web in August 1991. Type in its address and you will see a modern site that provides details of Tim Berners-Lee's seminal achievements at CERN, the European Organisation for Nuclear Research where he devised the first web browser and server.

Amid the explosive growth of internet services such as e-mail, music downloads and video streaming, along with the growth of the web itself, little thought has been given to recording information for posterity. The rapid turnover of content on the web has made total loss the norm. "Civilisation is developing severe amnesia as a result," says Stewart Brand of The Long Now Foundation. Danny Hillis, a pioneer of parallel computing and machine intelligence, fears the world has become stuck in a digital dark age, with few cultural artefacts from its digital past to point the way.

Lacking cultural artefacts, society has no mechanism to learn from previous mistakes. It is not hard to see where that can lead. The Library of Alexandria—built during the 3rd century BC to house the accumulated knowledge of centuries—reputedly had a copy (often the only copy) of every book in the world at the time. It burned to the ground sometime between Julius Caesar's conquest of Egypt in 48BC and the Muslim invasion in 640AD.

It remains unclear how, when or why the fire started. But it destroyed many of the works of Aristotle, Aeschylus, Euripides, Sophocles and countless other ancient astronomers, mathematicians, poets, playwrights and philosophers. All that remains today stems from a small fraction of the Alexandrian archives that had been backed up in a daughter temple called the Serapeum. Some historians believe the loss of the Alexandrian library, along with the dissolution of its huge community of scribes and scholars, created the conditions for the Dark

Ages that descended across Europe as the Roman empire crumbled from within. A millennium of misery ensued, with ignorance and poverty the rule until the Renaissance dawned.

No one is saying that today's digital dark age portends any such disaster. Nevertheless, there could be serious ramifications for education, scholarship, government and even national security. All are legitimate concerns for the future.

It is not really surprising that attempts to dig up The Economist's early web pages have come to naught. The original site ceased serving pages several years before today's search engines came into being. Those around at the time (eg, Aliweb, JumpStation and WebCrawler) have long been pensioned off, or subsumed into other services. Even the Wayback Machine—a search engine that allows users to call up old web pages which have vanished from the scene—came up empty handed.

The Wayback Machine's inventor, Brewster Kahle, is an internet entrepreneur, philanthropist and computer whizz who helped design Mr Hillis's ground-breaking Connection Machine in the 1980s. In 1996 he founded a non-profit organisation, the Internet Archive, to create a free internet library capable of storing a copy of every web page of every website ever to go online. The Wayback Machine allows users to view the library's archived web pages as they appeared when published. Today the Internet Archive also includes texts, audio, moving images and software. At the last count, its collection contained more than 150 billion items.

An interesting spin-off from the Internet Archive is the Open Library, which aims to provide a web page for every book in existence. The Open Library is not to be confused with Project Gutenberg, founded by the late Michael Hart, the inventor of the electronic book back in 1971. Project Gutenberg offers some 40,000 e-books that can be downloaded free in any of the popular e-reader formats.

Open sesame

Open Library, by contrast, is essentially an editable catalogue. The organisation works with various libraries around the world to catalogue their book collections and scan in various texts. So far it has amassed details of over 20m titles and scanned in the contents of some 1.7m books that are in the public domain, and therefore free to download. If a book is still in copyright, it can be checked out as a digital loan for a couple of weeks—rather like a book from a bricks-and-mortar library.

But why is Mr Kahle doing all this when Google, Amazon, Apple and others are putting civilisation's creative outpourings online as fast as their editing, scanning and recording machines can cope? The obvious answer is because these commercial entities charge for access to some information, whereas non-profit archives are generally free. Money aside, there are other reasons for encouraging open-source archives. For one, commercial outfits can be picky about granting search engines, other than their own, access to content they have archived. And even with material old enough to be in the public domain, users of proprietary archives can still be denied the right to copy or distribute it.

As the Internet Archive notes, without libraries, people would find it hard to exercise their "right to remember". As more and more public information moves from printed to digital form, it is vital that virtual libraries of all kinds archive as much of it as they can in the interests of future reference and accountability. For its part The Economist has made available digital copies of every issue going back to the day it was launched in 1843. But it is not alone in lacking copies of the web pages it produced just a few years ago.

Fonte: The Economist, London, v. 404, n. 8800, p. 11, 1-7 Set. 2012.