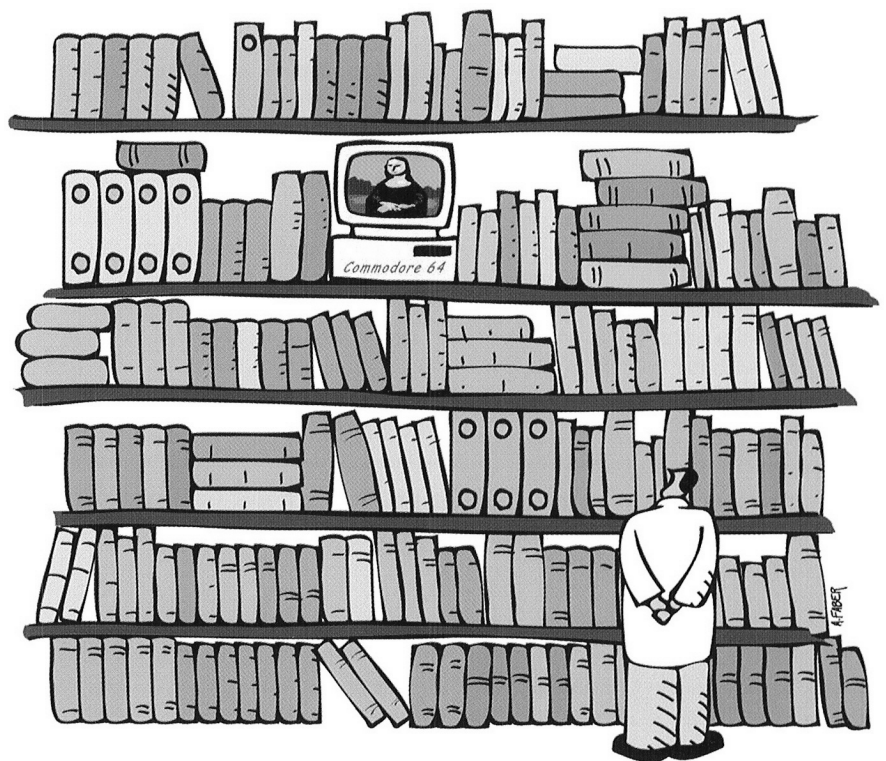


# The great digital information disappearing act

Deborah Woodyard, Digital Preservation Coordinator, The British Library

Could the information age spell the death of information? This is a genuine risk that proper action to store websites and other electronic information can avoid. Now the British have enacted legislation so that electronic publications would be saved for future generations.



How magnificent is a notebook from Leonardo da Vinci? It is full of exquisite illustrations of thoughts and ideas. Nobody could deny it is a unique and valuable record of a great mind.

Who is the Leonardo of 2003, or the modern day Marie Curie or John Lennon? How will we know? Will anyone keep their palm pilots or store their personal web pages? After all, surely today's great thinkers, inventors and creative minds are largely using the latest technology to "pen" their thoughts, instead of ink and paper. You can do so much more in digits and pixels!

Perhaps they were already creating digital documents 20 years ago, but will their old Commodore 64 work when we dig it out of the attic to rediscover their genius in 15 or 50 years time? Leonardo's notebook opens, but would the old computer work? Even if we could turn it on, the old software applications and commands could make it impossible to read the information. Genius would be lost.

Sure enough, there will always be somebody out there who will know how to use it, or even have a Commodore 64 at home. However, we cannot rely on a curious minority of enthusiasts for the

continued maintenance of our documentary heritage and written history, especially in view of the vast amount of digital material being created in the world today.

Look at another seemingly innocent problem: the dead link. You click a hyperlink you have saved, but the document you wished to re-read has disappeared. Whether it has been stored or deleted is one question. The next day, you discover the link is back, but the document seems to have been slightly amended. Has it? Revisionists will be smiling, but for researchers and keepers

of record, this represents a growing challenge. And there are others.

The worldwide web is estimated to contain over 250 terabytes of information and is still growing rapidly. That is equivalent to 17 times the volume of the print collections in the United States Library of Congress. This alone sounds mind-boggling and yet, according to the same research source, e-mail creates roughly 400,000 terabytes of information a year!

Consider the information created in your organisation, the e-mail and work documents that form the basis of your organisation or business. Can you guarantee that you will still have the right computer hardware and software to use these important records and data in the future? It will require a long-term plan.

Electronic Records Management systems definitely improve the immediate management of this material, but how long can they really keep the memory of our organisations? Technology moves at such a

are a good place to look for inspiration in solving this problem. Take the national library of the United Kingdom, the British Library (BL), which has the ethical and legal responsibility to acquire, preserve and make available all printed material published in the UK. This is enforced through legal deposit law. In recent years the national published output has included an increasingly digital component that was not covered by the law. Legislation to extend legal deposit to non-print publications made significant progress through parliament in 2003 culminating in a Royal Assent Act in October, in a bid to ensure electronic publications are saved for future generations.

The BL collection of digital material also extends beyond the deposit of publications. The Library receives digital collection material through several methods: deposit, purchase, capture and creation. It arrives on disk or online, in the form of CD-ROMs, CD-Rs, DVDs, floppy disks, online publications and web sites. It can then also exist in a multitude of legal

pages, or in a digital file, but you cannot see or use it anymore.

New systems are being developed specifically to collect, manage and preserve the digital collections in the Library, such as the recently established Web Archiving Programme. Under the programme, the British Library will take a leading role in negotiating permission with authors and owners to collect UK web sites in partnership with other institutions, nationally and internationally. Given the huge scale and dynamic nature of the web – there are currently more than 4 million UK-based web sites – the Library does not consider it feasible or affordable to aim at truly comprehensive coverage. The approach will be selective and based on voluntary terms negotiated with the web site owners.

The BL is also working towards improving the archiving of digital resources as part of the Digital Preservation Coalition (DPC) in the UK. This is a nationwide consortium of 25 major UK organisations which was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base. Initiatives such as these will ensure a way forward for at least some digital information, but more action will be needed in all countries if we are to avoid a digital “dark age” in years to come. † †

### Ignoring preservation would be like putting a book on the wrong shelf and never being able to find it again, or as bad, storing the book under a water drip.

great pace that keeping up with current and new trends often leaves the older information behind. File formats and software applications are constantly evolving. We need to take these issues into account before we suffer severe digital amnesia.

Simply saving all the bits and bytes is a good start, but it will not necessarily be enough to be able to use the information again in the future. Managing large amounts of data is a considerable challenge also. You need to understand the technical profile of your entire collection, as well as be able to efficiently find a single piece of information within a huge storage system – a needle in a virtual haystack.

#### Librar-e

Libraries have always dealt with large amounts of information, and large ones

formats such as images, text, audio, interactive databases or geographic information systems. This diversity adds several degrees of complication to finding digital preservation solutions.

Preservation needs to be addressed throughout the life cycle of digital material in order to be effective. Appropriate steps must be introduced to acquisition and cataloguing, for example, to ensure the capture and management of technical details and preservation information, and guarantee the digital files are not altered in any way. Ignoring preservation and not performing these tasks at this stage would be like putting a book on the wrong shelf and never being able to find it again, or as bad, storing the book under a water drip. The information may still be between the

#### References

- Digital Preservation Coalition website [www.dpconline.org/](http://www.dpconline.org/)
- Lyman, P. and Varian, H., “How Much Information 2003”, online report by the School of Information Management and Systems (SIMS) at the University of California, Berkeley [www.sims.berkeley.edu/research/projects/how-much-info-2003/](http://www.sims.berkeley.edu/research/projects/how-much-info-2003/)
- Preserving Access to Digital Information (PADI) at [www.nla.gov.au/padi/](http://www.nla.gov.au/padi/)
- Read about the UK's extension of legal deposit to non-print: [www.bl.uk/collections/british/depint.html](http://www.bl.uk/collections/british/depint.html)
- Also see “Historic change in Legal Deposit Law saves electronic publications for future generations”: [www.bl.uk/cgi-bin/press.cgi?story=1382](http://www.bl.uk/cgi-bin/press.cgi?story=1382)
- For more on the British Library Act: [www.bl.uk/about/blact.html](http://www.bl.uk/about/blact.html)