

Cometh a digital dark age?

Tony Davies

External Professor, University of Glamorgan, UK. c/o Creon Lab Control AG, Europaallee 27-29, 50226 Frechen, Germany

I was fortunate enough to represent IUPAC at the recent ICSTI seminar on the Digital Preservation of the Records of Science hosted by UNESCO in Paris over 14 and 15 February. The topics covered were an eye-opener for an analytical spectroscopist. I had thought that over the years we had managed to supply our field with a range of widely implemented international data standards capable of guaranteed long-term digital archiving. I suppose I was rather proud of what we had achieved as a community of users, manufacturers and industry. I now realise we are the lucky ones. The rest of the scientific world are currently running scared of what now appears to be the advent of a so-called "Digital Dark Age". In this issue I will highlight the reasons for the meeting at UNESCO and what urgently needs to be achieved on a global scale.

There is a general worry in the international scientific community that the moves towards electronic production and presentation of scientific data will lead to serious deficits in the archiving of the records of science. The first meeting on this topic was organised in January 2000 by ICSTI and a progress review in 2001 established the urgent need for a second meeting, which was hosted by UNESCO this February.

The Objectives of the February meeting were outlined as follows:

- To ensure all the interests in digital preservation in Science are aware of all current activities in the field
- To evaluate the needs for co-ordination of the efforts
- To create any necessary structures and work programmes to ensure co-ordination of the activities

And the meeting should also deal with the following issues:

- What are the varieties and future uses of scientific and technological information that must and will be archived?
- What is the minimum amount of information (data fields) needed to locate and identify information and who is creating what kinds of stan-

Abbreviations



ICSTI The International Council for Scientific and Technical Information (<http://www.icsti.org>).



United Nations Educational, Scientific and Cultural Organization.

UNESCO United Nations Educational, Scientific and Cultural Organization (<http://www.unesco.org/>)



CODATA Committee on Data for Science and Technology (<http://www.codata.org/>)



ICSU International Council for Science (<http://www.icsu.org>)



IUPAC International Union for Pure and Applied Chemistry (<http://www.iupac.org>)

dards related to location and basic identification?

- What business and information models are appropriate and how should access to the digital archives be arranged?
- Where are the common issues with the preservation of more general cultural archives and how can these be accommodated?

The seminar

The seminar started with the usual welcoming speeches and an explana-

tion of the interests of the sponsoring organisations. There then followed two days of specialist presentations from interested scientific organisations, international representative bodies and renowned speakers from the scientific publishing industry.

For me, one of the most worrying revelations during the two-day meeting was the current acute fear amongst science historians reported by William Anderson of CODATA. He used a phrase, which at the time was completely new to me in revealing that there is the imminent danger of the

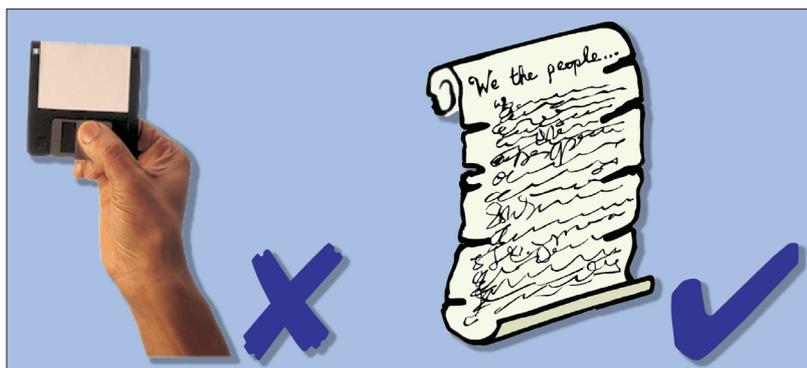


Figure 1. Unfortunately we are currently in the situation that science archivists have well-established practises for handling paper legacies but currently have terrible problems when presented with digital content.

arrival of a new “Dark Age” with our scientific cultural heritage being permanently lost through the exclusive use of electronic media. This will become more severe when electronic laboratory notebooks finally become integrated into the normal working environment. This was highlighted by an example of the problems archivists

are now struggling with (see Figure 1).

A worrying example was used to highlight this problem. Following the death of an eminent British scientist his widow presented his archive material to the British Library for posterity. The problem is that there is effectively no infrastructure available in what is this

national archive for handling two old personal computers and boxes of old format disks!

The debate on what to archive?

A large amount of time was devoted to discussion on exactly what should be archived with no general agreement being reached. The data community (probably heavily influenced by the FDA 21 CFR Part 11 rules currently revolutionising pharmaceutical IT) thought that all information needed to be stored whereas the traditional archivists looked at the logistics and demanded that only selected content land in the electronic archives.

One of the international legal loopholes currently urgently needing closing is that although there is a legal requirement for publishers to make a deposition to their national archives of all material printed in a particular country, there exists no equivalent law requiring deposition for electronic-

only publications. This will apparently be addressed during the Spanish presidency of the EU and it will be interesting to see how the Council of Ministers deals with this thorny subject.

On an international level it was clear that the classical role of the librarian as archivist is outdated and being continually undermined by the digital presentation of scientific publications. An ever-increasing proportion of the library budget is being spent on digital-only subscriptions to peer-reviewed scientific journals. These electronic journals are maintained off-site and accessed through the Internet often on a pay-per-view basis. The librarians cannot archive this material, as it never physically lands in the individual organisations. It was generally agreed that it is foolish to expect the publishers to take over the role of archivists and so another mechanism needs to be put in place.

A series of presentations dealt with individual limited-term projects that were or had been run in various countries funded by the Mellon Foundation, the EU and by different national governments. What was strikingly clear was that the projects were not coordinated and any benefit would probably end with the funding.

Not just a problem for scientists!

Having only just become aware of the phrase “Digital Dark Age” you can imagine my complete surprise when browsing through one of the bookshops at Newark international airport two weeks later, I discovered a brand new book *Dark Ages II—When Digital Data Die* by Bryon Bergeron, a teacher at Harvard Medical School and MIT (published by Prentice Hall PTR, Upper Saddle River, New Jersey 07458, USA, ISBN 0-13-066107-4, www.phptr.com).

This is an interestingly written book with many anecdotes, but with much which directly addresses the problem of long-term data archiving. It is not a tacky techie tome for IT freaks but has good advice for everyone from home computer users to managers of corporate networks written in clear normal language. He attacks “Bloatware” succinctly and provides many useful links to more detailed information sources such as the US NARA (National Archives and Records Administration) Center for Electronic Records guidelines. Table 1 is extracted from the

Table 1. Extracted and adapted from *Dark Ages II*, Chapter 3, page 82. Expected Media Lifetimes under Ideal and Typical Conditions.

Storage medium	Ideal lifetime (years)	Typical lifetime (years)	Comments
CD-R	5–100	2–30	Dye less stable than pits used in commercial CD-ROMs
CD-ROM	30–200	5–50	Uses pits on a metal surface to encode data—fragile surface
DVD	100	20	Higher-density susceptible to environmental changes
DVD-R	20–30	10	As with CD-R less stable than commercial media
Hard disks	?100	10–20	Lifetime is down to stability of the mechanical parts
Magnetic tape	30–100	5–20	Rewind periodically to release tension
WORM	30–200	5–50	Formats not as standardised as for CD-ROMs and DVDs
Paper buffered	?500	50–500	!
Photographic print	?200	?100	Assuming non-acid paper and stored out of light (not Polaroids!)
Microfilm	500	100–200	Standard for archives

book to give an idea of the level of the advice to be had.

Meeting outcome

One of the messages which came out of the meeting was the clear need for a more active advocacy making scientists aware of the encroaching danger and more especially of the heritage value of the work they should be careful to make available to archivists. As digital preservation will not be a cheap exercise it was seen as important that the need be expressed at many levels in order to convince those who control the different budget sources of the vital nature of this work. ICSTI will lead in this area.

The different needs of the text archivists as opposed to the data archivists were clear to all by the end of the meeting. This was especially the case during discussions on metadata content. From my own recent experiences working with FDA 21 CFR part 11 compliant systems I can see that the issues of exactly what metadata is worthy of storage and how to obtain it is still a critical factor in an industry well advanced in archiving digital content. Amongst those sciences just feeling their way into this field, there are those who cannot currently agree on what constitutes metadata!

I was surprised by the depth of thought taken by many of the contributors to the seminar. This was evident in a number of well-constructed arguments such as those presenting the desire for a “technology watch” on current archival computing systems. This technology-watch will need to be established in order to warn in good time of upcoming migration needs when computer hardware or software on which the archives are reliant are about to become outdated.

The needs of the developing countries were reported as being not only in the area of the preservation of science information but also in the desire for more exposure, which they currently lack.

Conclusions

OK, all I can say is worry! Basically, we should all be rather worried about the current status of born-digital scientific information. Fortunately, the current precarious state of our science legacy has been spotted and there are now international initiatives underway at a political level to secure the significant funding required to establish the necessary infrastructure—we can only hope that they are successful. Maybe by talking about the problems with our colleagues we can raise awareness and support those striving to find appropriate solutions.