

Digital archiving: storing for the future

Mr Tony McSean, the Director of Library Relations of publishing giant Elsevier, shared his thoughts on the future of digital archiving at an Executive Management Programme in February, themed "Technology and the Profession". This is an edited extract from his presentation.

In the old world of paper, the preservation of materials was simple, broadly understood and widely accepted.

Libraries would buy and keep the print materials, which existed in fairly stable medium and long-lived formats. Back files could be stored and preserved in great libraries, and scholars could visit the libraries or archives to consult and gain access to the materials. It was a process and system that was easily understood.

In this paper world, publishers were never involved in the process of archiving, and while the "technology" has become obsolescent, it was, and still is, resilient.

The advent of new technologies, however, has changed the face of archiving. The aims remain the same, but the options and demands pose a new set of challenges that paper archiving never had to deal with.

Today's archives and libraries have to address the needs involved with the availability, storage and preservation of intellectual content in digital documents. These are produced worldwide, and libraries must still attempt to ensure that the access to this content is permanent, resilient, guaranteed and assured.

The new problems that digital archives face are manifold.

Digital media still represent a relatively new, unknown and volatile environment, with media that are unstable and changing formats. With digital archiving, it is data that does the travelling, not the user.

Because of the new way in which digital documents are produced, shared and published, a host of questions have emerged that have to be answered, such as: Where can the back files be kept? And, who should take the responsibility for keeping the files?

Another significant difference between archiving in the paper world and the digital world is the involvement of the publisher. Libraries, archives and publishers have found that digital archival has to be a library-publisher partnership. E-publishing has changed everything, and publishers are increasingly finding that they have to take the responsibility of archiving – something that has not been part of the publishing culture in the past.

Publishers like Elsevier have found that as they enter the world of archiving, they have had to reassure customers who were concerned about retaining access to the archived material, assure authors that their work would stay accessible for the future, and also reassure libraries that they would get it right.

Since 1999, Elsevier has adopted and developed a formal archiving policy. This includes maintaining a permanent archive of the journals it owns, placing the archive in one or more independent, librarian-approved depositories, and trying to preserve subscribers' access to journals which have transferred to other publishers.

Digital archives can take several forms, some of which include:

- Documents are stored in the depository of the publisher, who has to manage the risks and scale of archival. Users or subscribers get access to the files.
- All its digital documents are kept additionally in Elsevier's production system – its Electronic Warehouse.
- Publishers' customers store the documents locally, but with an access commitment only to their own communities.
- Self-designated "national" archives such as libraries or other institutions keep an archival copy locally, as a national or regional security measure.
- Publishers and trusted archival institutions form formal, contractual relationships that ensure permanent retention and access to the digital files for future generations.
- Trusted third-party providers run permanent archives that provide access to publishers and for library members on a subscription basis.

Whatever form the digital archives take, publishers, libraries and other providers need to keep to agreed standards to ensure that the archives serve their function. For example, they must ensure that the digital archive lasts more than 100 years, and that sufficient resources are committed to archiving the content itself.

Archives, which should drive technology migration, should also not compete with publishers, yet they should not be totally "dark" – that is, they ought to be used.

There is an urgency for digital archives to be set up and kept alive and functional. When publishers go out of business, there may be no-one to take up their assets, and the digital documents could be lost forever. Natural or man-made disasters, or unforeseeable catastrophes, too could destroy the intellectual content.

As publishers, libraries and providers constantly review the developments, standards and changes in requirements for digital archiving, they need to keep

working on the appropriate protocols, procedures and agreements – and listen to the to the library community – to ensure that digital documents are preserved adequately and over the very long term.