



THE LAST WORD

The Sunday Times Online

THE SUNDAY TIMES recently joined the growing band of newspaper titles that are putting their back-catalogues and collections online. At a time when parent company News International has hit the headlines for the wrong reasons, it is interesting that they have led the way when realising the commercial value of old news: *The Times* Digital Archive has been available online now for many years, covering the period from 1785 to 2006 in 1.4 million pages. Of course, Find My Past's parent company, brightsolid, have taken on the task of digitising local newspaper collections for the British Library from the 18th century up to the 20th century, bringing an even greater range of articles, stories and personal notices to the attention of the family history community (see www.britishnewspaperarchive.co.uk).

I spoke to Seth Cayley from Cengage Learning, the company that took on the daunting task of digitising *The Sunday Times*, to find out more about the project – including the plans for the collection, and some of the technical challenges faced during the process. However, the first question was, why this particular collection? Although I read a daily paper, there simply isn't enough time to wade through the Sunday newspapers, as most of them seem to be pre-written features or weekly summaries. However, it was exactly this in-depth level of reporting that people had been requesting ever since *The Times* first appeared online in 2003.

The material in the supplements tends to be more analytical, providing context to the stories featured in the daily editions as 'news', and includes some incredible regular features – such as a freemasonry column that ran from the 1850s to the 1920s every week, full of names, places and meetings, which is a great way to find out whether your ancestor was involved. On the other end of the spectrum was the 'Insight' team that investigated and broke some of the big stories of the day, such as the exposure of Kim Philby as a spy.

For anyone who's ever wondered about the amount of work that goes into creating a digital archive from old documents, it was fascinating to talk about the process behind the project. The date range covered by the newly released material starts in 1822, and has been digitised up to 2006. Although there had been microfilm access to the archives, no indexes survived before 1973, so as well as digitising the 600,000-plus pages, there was a need to create both an index and a means of presenting the

information once a search produced some possible results. The process took about a year to complete, and one of the trickiest elements was ensuring the metadata was sufficiently sensitive – Optical Character Recognition being essential to ensure the quality of search results – and that there were various means of displaying information, both clipped articles plus the context of the remainder of the page. This approach makes it much easier to create a citation for researchers, or send to others. Whilst I cannot reveal the exact price per page, this was low in relation to the amount of work involved, being a six-figure sum.

Cengage have made a name for themselves with their work on newspaper digitisation. They have 13 titles in their collection, including treasures such as the *Illustrated London News*, as well as other historic datasets like the State Papers held at The National Archives. However, there is a drawback – their market is primarily libraries, higher education institutions and public bodies, mainly due to the subscription fees they charge. This is not an organisation geared towards the private researcher market in the way Find My Past and Ancestry are. Yet an increasing number of public institutions hold

subscriptions to not just one title, but entire bundles of newspaper collections that are available for cross-searching, which means you should be able to pick up references to ancestors in any title in which they appear. Cengage is also interested in providing content to schools, so that students can contextualise their work by

finding easy access to the stories behind the historic headlines. Again, is this any use to family historians? Well actually, yes, if we are keen to encourage the next generation of family historians to take an interest in the past and link their family stories with major moments in history.

As family historians, we rely too heavily on the tried-and-tested name-rich lists and government records for our information. There are thousands of datasets out there, records that have been digitised and indexed for different, often academic, audiences. We need to find ways to locate and unlock them for a wider audience.

It was exactly this in-depth level of reporting that people had been requesting

WE ARE INTERESTED to hear about any datasets you've discovered during your work that are not specifically geared towards a genealogical audience, but perhaps would be of much wider use. Contact us using the form at www.your-familyhistory.com/yourstories.