

Human Traces: How digitising newspapers is transforming family history.

Patrick Fleming, Head of Operations, London, British Library

Patrick.fleming@bl.uk

Abstract:

This paper outlines the British Library's experience of how its digitisation of newspapers has transformed family history. It charts the journey from the Library's early days of newspaper digitisation through to key partnerships with public and private sector organisations evidencing the key connectivity between fully searchable newspaper datasets and the quest of family historians for rich primary sources.

Throughout the world, local and national organisations, including governments, are realising that much of their national heritage, culture and history has been chronicled in newspapers.

Most newspapers have been printed on cheap paper and deterioration of the paper is threatening that newspapers disappear before they are well preserved.

As a result newspaper digitisation projects are popping up all over the world and millions of pages of national and local newspapers are being preserved with access online.

Genealogists benefit from this huge influx of searchable online images through the information that it contained in the newspapers.

Some of the types of information that may be found include articles about births, marriages, death notices, obituaries, social events, sporting events, church activities and affiliations, occupations, legal notices, inquests, court cases, land sales, taxation, military service, businesses and advertisements.

In the past doing research in newspapers was exceedingly tedious. The researcher essentially had to read through page after page of newspapers hoping to find some relevant details. Today most of the online newspaper archives are completely searchable by any word in the newspaper.

This paper explores how the British Library has transformed its approach to digital newspaper access through its partnerships with Gale Cengage, brightsolid and others to bring millions of pages of fully searchable content to life. It will also outline how these initiatives have stimulated more news partnerships and innovations and awakened the Library's desire to consider creating a specialist Family History service in Central London and how the development of news content streams will broaden the appetite for research.

As most of you will be aware the British Library newspaper collection is possibly the greatest collection in the world. The collection contains most of the runs of newspapers published in the UK since 1800: 660,000 bound volumes and

614,283 reels of microfilm containing tens of millions of newspapers together with more than 250 international titles.

As the years have gone by the Library has continued to collect physical newspapers, mainly through legal deposit but also through purchases from other countries and through electronic resources. The physical collection alone continues to grow by 300 metres per year. The newspaper library at Colindale which will close for the last time on November 8 2013, hosts 30,000 visitors per year, many of them genealogists who travel from all parts of the world to London to study newspapers.

For many years the Library has known that collecting an ever growing mountain of newsprint and microfilm could not continue and had to change. With the explosion of the Internet it was easy to see that digitisation was the future. Publishers, libraries and academics worldwide had begun to digitize content to satisfy the growing demands of researchers.

Running parallel to this was the extraordinary explosion of interest in genealogy which sparked a massive interest in resources which grows year on year.

The value of the worldwide genealogy market is estimated to grow from a 2012 figure of \$2billion to \$2.7 billion in 2018.

There are currently estimated to be more than 92 million individual genealogists worldwide. In the United States alone a genealogy enthusiasts typically spends anywhere between \$1,000 and \$18,000 a year to trace ancestors.

There are no definitive statistics to illustrate the likely number of genealogists in the UK but all the indicators are that this is large and a high growth market.

The spend of UK family historians is increasing and the main beneficiaries of this are Ancestry and the Library's partner Find My Past which both offer newspaper content as a primary resource within their portfolios of family history datasets.

Most of the top genealogy websites are focused on the US market. 65 of the top 100 genealogy websites are from the US. This roughly corresponds to the size of the US genealogy market.

Ancestry.com continues to remain the most popular genealogy website in the world and is regarded as a powerhouse in the field of genealogy. They own four of the top ten genealogy websites and 14 of the top 100 genealogy websites. They have 1.7 million paid subscribers. The company expects to grown 2013 revenue by 15% to 20% driven by continuing growth in the genealogy market.

The second most popular genealogy website is My Heritage. This allows you to share family trees and family photos. Although the website caters primarily to the US genealogy market, it is a global website.

The third most popular genealogy website is Family search.org which contains a vast collection of ancestral records from around the world. In addition to the website, Family Search is backed by more than 4,500 family history centres in 70 different countries. The Family Search website has grown rapidly and it now numbers some 2.58 billion names and 466 million digital images.

There are 14 UK websites in the Top 100. The three most popular are Ancestry.co.uk (ranked 8), Find My Past (ranked 13) and Genes Reunited, also

owned by brightsolid (ranked 16). Find My Past also has websites and businesses catering for Ireland, The US, Australia and New Zealand.

What the genealogy, publisher and library worlds have in common in this context is that the overall flow of value has been in one direction: from digitising agency to target audience. Such a situation is not surprising given that until mass digitisation of resources was both feasible and had a critical mass of connected users, 'one directional' flow of value had predominated for as long as collections had existed.

Due to the increasing body of digital collections, the rapid adoption of the Web as a source of knowledge and information and the changing relationships between content and the individual that Web 2.0 tools have heralded there are now digitisation projects where audiences have been involved in the service creation process and where such services have been enriched through the on-going involvement of new audience groups. Both of these approaches show how the traditional value chain can be reformulated to increase public benefit through engagement and content repurposing. This is precisely how the genealogy market is progressing.

Public value for the British Library lies in linking our content and collections to user groups.

With newspapers the key is to understand the potential of the value chain.

The first challenge, as most libraries working with newspaper collections know, is to get to get a mass of content digitised and available to users.

Funding for such projects is a critical issue. Digitisation is not cheap.

The British Library does not get direct Government funding for digitization so innovation and support from both the public and private sector has played a crucial role in our development.

The start of the process for the British Library started at the turn of the millennium when we received £3 million worth of funding to digitise three million pages of 19th century newspapers from Jisc- a registered charity and champion in the use of digital technologies in UK education and research. This created a fully searchable collection of 70 national and regional newspapers hosted and marketed by Gale Cengage.

The final selection of newspapers selected by the British Library in consultation with the range of academic experts was designed to provide a cross-section which represents the development of the newspaper as an important contributor to social and political life in 19th-century Britain. The selection of newspapers from the 19th century, which are no longer in production, avoided both issues of copyright as well as potential areas of duplication with contemporary newspapers which are making their own historical holdings more accessible to online searching.

This resource was celebrated through academic circles worldwide. It taught is two things. One there was a powerful new and expanding audience of genealogists using the resource and that at three million pages with a collection of some 800 million pages, three million was a drop in the ocean.

Following an innovative open procurement the British Library began a partnership in May 2010 with brightsolid to create the www.britishnewspaperarchive.co.uk. This ten year partnership has the ambition to digitise 40 million pages. Currently there are seven million fully searchable pages with a further eight million to arrive by the end of 2015 covering the early eighteenth century through to 1950.

Whilst Gale Cengage reached the institutional market, brightsolid sought to target the business to consumer market.

The easy logic for the Library was to see how we could reach out to both audiences and maximise our connectivity with audiences including genealogists.

Firstly we were able to embed the Jisc collection in the brightsolid collection. Tranches of the brightsolid collection are now available through Gale Cengage with more to follow.

Importantly from a genealogy perspective the content of the British Newspaper Archive is available through brightsolid's Find My Past website directly targeted at family historians.

Findmypast.co.uk has digitised many family history records and now offers access to over a billion records dating as far back as 875AD. This allows family historians and novice genealogists to search for their ancestors among comprehensive collections of military records, census, migration, occupation directories, and current electoral roll data, as well as the original comprehensive birth, marriage and death records and of course, with the British newspaper archive content.

Expanding the content world to enable our newspaper collection to be accessed by as many people as possible requires us to find other way of establishing collaboration and creating a virtuous circle with our partners.

Both brightsolid through the British Newspaper Archive and Find My Past and Gale Cengage have ongoing digitisation projects with the Library that will grow the content base with and beyond newspapers. From the beginning of the partnership with brightsolid has had its own scanning and production studio near to the collection at Colindale and will move to Boston Spa when Colindale closes.

This enables both organisations to act swiftly to meet customer demand and of course to collaborate with other organisations.

This is illustrated perfectly by an announcement made in August 2013 by the British Library, brightsolid and The Center for Research Libraries (CRL). The three organisations have formed a pathbreaking partnership to digitise and make accessible historical news content from Africa, South Asia, and Latin America of interest to scholars and researchers at UK and North American universities. This initiative will dramatically expand and enrich the digitized collections in the World Newspaper Archive, an initiative by North American research libraries and Readex (a division of NewsBank) to preserve and provide persistent access to historical newspapers from around the globe.

Collected and preserved over three centuries of activity, the British Library's newspaper collection represents an unparalleled archive of British and overseas newspapers, encompassing more than 750 million pages of local, regional and national newspapers. CRL's own collection of more than 15,000 U.S. and foreign newspapers represents the largest circulating collection of news resources worldwide. The combined resources of these two institutions constitute a body of

historical and cultural evidence that is not, and could not be, replicated elsewhere.

The initial phase of the partnership will focus on newspapers published throughout Africa for the period 1798–1900. The British Library has extensive holdings of African newspapers, including many from the 19th century. The collections are strongest for South Africa and former British colonies, but there is also material from Francophone and Lusophone Africa. Holdings are predominantly in English, but also include material in Afrikaans, Zulu, and other African languages.

Titles provided by the British Library will be supplemented by holdings from CRL and its member libraries, including the New York Public Library, Yale University, University of California, Berkeley, and others. When completed, the World Newspaper Archive module African Newspapers, Series 2 (1798-1922) will encompass more than 1.2 million pages from sub-Saharan Africa as well as from North African countries.

Newspapers from other world areas will be digitized over the next several years, including content from South Asia, Latin America, and potentially 19th and early 20th century newspapers published in London, Hong Kong, and China relating to commercial and political relations between East and West. The scope of the BL-CRL effort will eventually encompass up to 7 million pages of international news content, subject to the level of investment of the library community in support of this effort.

Content provided by the British Library through this partnership will be accessible free on the premises of British Library within the United Kingdom..

CRL is an international consortium of university, college, and independent research libraries. Founded in 1949, CRL supports advanced research and teaching in the humanities, sciences, and social sciences by preserving and making available to scholars the primary source material critical to those disciplines.

CRL acquires and preserves newspapers, journals, documents, archives, and other traditional and digital resources from a global network of sources. Most materials acquired are from outside the United States, and many are from five “emerging” regions of the world: Africa, the Middle East, Southeast Asia, South Asia, and Latin America.

CRL enables institutions to provide students, faculty, and other researchers liberal access to these rich source materials through interlibrary loan and electronic delivery. CRL loan and electronic delivery are designed to support major research projects, such as the production of scholarly monographs and studies, the writing of dissertations, and seminars at the graduate and advanced undergraduate levels.

Librarians, specialists, and scholars at CRL libraries participate in building this shared CRL collection of research materials through the purchase proposal and demand purchase programs. Participation in these programs and in CRL collection webinars and other Global Resources Forum events enables librarians to benefit from the enormous pool of collection-related expertise and knowledge available in

the CRL community. Collection specialists at major U.S. and Canadian research institutions such as Harvard, Yale, Princeton, the University of Chicago, the University of California, McGill University, and the University of Toronto participate in these CRL programs.

These initiatives provide extraordinary collections that advance knowledge for all researchers.

This, of course is only part of the story for genealogists across the world who seek more and more information and in a variety of formats. Genealogy, of course, has been transformed by magazine, radio, web and television products based around the subject utilising newspaper content as a primary source.

Who Do You Think You Are? is a British genealogy documentary series that has aired on the BBC since 2004. In each episode, a celebrity goes on a journey to trace his or her family tree. The 10th series is currently being broadcast. The show regularly attracts an audience of more than 6 million viewers.

Ten international adaptations of the show have been produced. In September 2007, a Canadian version comprising 13 episodes began. An Australian version aired in January 2008.

An American adaptation of the show premiered in March 2010.

Ireland's national broadcaster RTÉ has aired two series of Irish version of Who Do You Think You Are?

TVP1 has aired a Polish version called Sekrety rodziny,

Sveriges television, the Swedish public service broadcasting network, aired a Swedish version called Vem tror du att du är? in Spring 2009.

Also on 31 May 2009, South African station SABC2 premiered its version, split up into two parts, featuring local personalities.^L

In 2010 the Dutch version of the show started. The show is called Verborgen Verleden (Hidden Past). In January 2011 the Norwegian TV station [NRK](#) started to air the Norwegian version called Hvem tror du at du er?

In September 2010 the Danish TV station Danmarks Radio aired the first episode of the Danish version called Ved du hvem du er?

In 2008, ARD's Das Erste aired the German version called Das Geheimnis meiner Familie (The secret of my family).

Starting on 9 January 2012, Finland's national public-broadcasting company, YLE, aired 15 episodes of the Finnish version titled Kuka oikein olet? in which local public figures searched for their origins.

A furthering of the concept was developed for American public broadcast TV, called Finding Your Roots. This series adds DNA profiling to the format; including Y-chromosome DNA, mitochondrial DNA, and autosomal DNA analyses to infer both ancient and recent genetic relationships.

In January 2013, Czech broadcaster ČT started broadcasting their version of this show called Tajemství rodu (translated into English: The secret of the lineage.)

Starting 15 January 2013, RTP broadcast a Portuguese version called Quem É que Tu Pensas Que És? (Who Do You Think You Are?)

All feature in a variety of ways episodes in which newspapers play a significant role in discovery.

The British Newspaper Archive features in the TV series of Find My Past with two series shown on UKTV. This sees ordinary members of the public go on an extraordinary journey to find out how their ancestors played a role in some of history's key moments. Subjects covered include, The Battle of Britain, Mutiny on the Bounty, Jack the Ripper, Dunkirk, D-Day, The Titanic, The Battle of the Somme, The Dambusters, Gunpowder Plot, Tutankhamun, The Great Fire of London, The Battle of Trafalgar, Scott of the Antarctic and the British Royal Abdication Crisis.

Programmes like this reach many millions of people further stimulating interest in family history.

The explosion of social media, technologies such as geotagging, crowdsourcing and multiple news sources do, of course, take the relationship between newspapers and genealogy to new levels as different content streams throw up new resources.

With Colindale closing and the British Library's news content streams increasing all the time we have no doubt that genealogists will seek to refer to these sources.

The last word comes from genealogists.

Our experience of the relationship between newspapers and genealogy has provoked the discussion as to whether the British Library should have a dedicated Family History Centre in London.

The jury, as they say, is out on that one.