

CONTENT, COLLABORATION AND AGGREGATED CONNECTIVITY

Paper for IFLA Newspapers Section pre-conference, Mikkeli, Finland, August 7 to 9, 2012

Patrick Fleming, Head of Operations and Services, London

Abstract

The newspaper industry landscape is changing beyond recognition and news has become a real time worldwide commodity. Once upon a time newspapers were the be all and end all for everyone. Along came radio and television but newspapers held their own and things stayed relatively stable until the arrival of the Internet. That was the single most explosive and transformational change in the way consumers obtained news providing a limitless and powerful platform owned by no-one but used by almost everyone.

Today we expect news to be delivered to us at anytime in any place in any format we choose and social media has given every person the opportunity to be news providers.

Newspapers constitute a unique resource for engagement with the British past at an international, national and a very local level. They provide invaluable information, often found nowhere else, for teachers and researchers at all levels, as well as for the general public. While offering rich opportunities, original newspapers are hard to locate and often survive in rare or unique copies held at the British Library, frequently far from the researcher. Once located they are difficult to consult, the required information being hard to find in the bulk of published material and the original paper often fragile to the point of disintegration. Digitisation opens up access to this valuable resource, unique to the British Library and currently unavailable except by visiting the Library's Colindale Newspaper Library reading room, to the widest possible audience.

This paper will explore how the British Library is collecting, creating and collaborating to fulfil our vision of aggregated connectivity for our newspaper collection.

The British Library Newspaper collection

To understand the British Library's approach to content, collaboration and aggregated connectivity it is important to understand our wider vision:

" In 2020, the British Library will be a leading hub in the global information network, advancing knowledge through our collections, expertise and partnerships, for the benefit of the economy and society and the enrichment of cultural life. "

For all content including newspapers we see a world of fully searchable aggregated multimedia and multiplatform resources, a mix of free and pay to access content. The Library's role is to capture, collect and collate information and open up resources and point researchers to where they can find them around the world.

The vision has five key themes setting out strategic priorities

Theme 1 Guarantee access for future generations

Theme 2 Enable access to everyone who wants to do research

Theme 3 Support research communities in key areas for social and economic benefit

Theme 4 Enrich the cultural life of the nation

Theme 5 Lead and collaborate in growing the world's knowledge base

Storing, preserving and providing access to one of the world's greatest newspaper collections, if not the greatest, has provided the British Library with one of the most challenging tasks in its long history.

The collection dates from the 16th century and contains over 53,198 separate hard copy print titles, 370,000 reels of microfilm on nearly 55 km of shelf space. The collection grows every year by 350 metres as the Library collects more than 1400 titles from the UK newspaper industry through Legal Deposit together with more than 200 overseas titles.

In total the collection now amounts to some 800 million pages of which only 35% is on microfilm and less than 1% is digitised. The majority of the content is stored at the Newspaper Library at Colindale in a run down 1935 building where the conditions are far from ideal and which has reached its optimum storage capability. 15% of the collection is 'unusable' and a further 19% is 'at risk'

The British Library Act 1972 gives a legal requirement to store legal deposit content for the nation.

The main driver for change in our newspaper strategy apart from the evolution of the digital landscape was the need to integrate newspapers in the Library's headquarter building at St Pancras to enable researchers to view newspapers in the context of the wider Library collection.

In February 2007 the British Library Board reconfirmed the following vision:

" The Library will offer an integrated newspaper service based on digital surrogates at St Pancras with hard copy stored to help preserve it for future generations. "

To deliver the vision the Library had to persuade the UK Government to invest £33m to fund a new newspaper storage building at Boston Spa, Yorkshire to store all hard copy newspapers. This was achieved and the building is now being created with completion in 2013.

Access will then revert to St Pancras through microfilm and digital surrogates with hard copy newspapers only being accessible in very controlled circumstances-delivered from Yorkshire 200 miles to London to be viewed by a researcher.

The risk was that in saving the collection by moving it to the best archival conditions the Library would reduce access to unacceptable levels. We knew two things. Digital access was the future but our digital journey would take many, many years to reach an acceptable result.

The road to digital

The digital challenge for the Library isn't just technological but financial. As a national library we are funded by Government but we receive no Government funding for digitisation.

Therefore every digital programme we embark on has to be through innovative thinking, collaboration, partnerships and philanthropy. That comes in a variety of ways.

The Library first broke the digital ground in the digital space when it digitised The Burney Collection in 2006.

The largest single online collection of English news media from the 17th and 18th centuries, the Burney Collection, is now available free of charge to Higher and Further Education institutions and Research Councils across the UK as well as within the reading rooms of the British Library. As a result an unlimited number of students and researchers at academic institutions across the UK can now access these essential materials simultaneously bringing significant benefits for research and education.

The Burney Collection offers unique insights into two centuries of history through access to over 1,270 newsbooks, newspapers, pamphlets and a variety of other news materials published in England, Ireland and Scotland, plus papers from British colonies in Asia and the Americas.

Digitised through a partnership between the National Science Foundation and the Library the content was developed and hosted online by one of our partners Cengage Learning.

With the financial support of JISC, the digital version of the Burney Collection has been purchased in perpetuity on behalf of the UK academic and research community at a national level. It is also sold by Cengage Learning to institutions worldwide.

JISC offers leadership and support to UK educational organisations at a local, national and international level. It provides resources, knowledge and expertise that colleges and universities would struggle to source individually due to cost and resource. By staying abreast of developments across information and digital technology it assists the UK education community in making investment decisions that ensure they deliver the learner experience their students demand.

Prior to this access to the Burney Collection had been restricted, both in print and online. Its corpus of printed materials – originally collected by the Reverend Charles Burney (1757–1817) and greatly augmented since its acquisition – has been housed at the British Library. Due to its value and fragility, the Collection was unavailable for general use and the cost of acquiring online access put it beyond the means of the majority of institutions.

The complete Collection (including illustrations) is fully cross searchable, and has granular metadata plus an extensive range of search and browse options which opens up unrivalled and exciting research possibilities.

Among the highlights of this unparalleled collection are:

- Day by day coverage of the financial scandal of the 1720s, the South Sea bubble, with reports in the *Weekly Journal* or *Saturday's Post* of how Parliament decided that if they left the country, the directors of the South Sea company 'shall suffer death as a felon without benefit of clergy and forfeit to the King all his Lands, Goods and Chattels whatsoever.'
- The first advertisement for *The Memoirs of Fanny Hill* in the *Whitehall Evening Post*, 6 March 1750, and then, in the issue of 17 March, a report of how the publisher was taken into custody and all copies were seized
- News of the election of George Washington as president of the United States in the *St James's Chronicle* or the *British Evening Post* on 21 April 1789
- Insight into English attitudes to contemporary events, such as when the *English Chronicle* or *Universal Evening Post* used the unusual device of a headline – FRENCH REVOLUTION!! – for a whole page article on 18 of July 1789. It reported sympathetically on the fall of the Bastille four days earlier, including how the officers were decapitated in 'a sad but necessary spectacle...a solemnity worthy of the highest admiration.'

This transformed our thinking and opened up a world of possibilities. With a grant of £3m from JISC and online support again from Cengage Learning we digitised three million pages of complete runs of 19th century local, regional and national British newspapers. Titles were selected to reflect both the full chronological span of the era and the geographical spread of the UK and include the *Examiner*, the *Pall Mall Gazette*, the *Chartist* and many others. These were made available to further and higher education in a form that allows an unprecedented degree of cross-referencing across almost every imaginable discipline.

This iconic collection contains major historical events from the Battle of Waterloo, the Great Exhibition of 1851 to the opening of the Suez Canal, factual reporting of the Battle of Trafalgar in the *Examiner* and the gory details of the Whitechapel murders in the melodramatic *Illustrated Police News* to name but a few. Many famous authors of the 19th century period wrote for newspapers, and their work is represented, Dickens and Thackeray are included.

The corpus of Burney and the two JISC projects led to a step change within scholarly communities, vastly expanding the range and variety of what can be used in research, at all levels of education.

But at just over three million pages this was a drop in the ocean of more than 800 million pages of hard copy.

We needed a bigger, more dramatic solution.

So I came up with a mad idea. Why don't we go and find a partner who at their own cost and their own risk would invest millions of pounds to digitise many millions of pages and provide a miracle for both the Library and users?

The next hurdle was procurement to satisfy EU rules.

The project started fully in May 2008 when the Library project team defined how the formal procurement would run. The Library decided that a public procurement under a new process called Competitive Dialogue would be the best solution.

This had to be a process which was:

- transparent
- unbiased and giving best value
- enable market and technical knowledge to be gained from the suppliers – calling on as wide a pool of potential providers as possible
- Commercial issues were undefined and the Library was seeking a supplier to take on the project at their own financial risk

Within the Library's overall Newspaper Programme the digitisation project was required to meet the need for bulk digitisation and:

- Grow the use of surrogates (to meet a KPI for the Library of 80% of reader requests from the newspaper collection to be met from a surrogate by the end of 2017)
- To "improve" the user experience following the completion of the Programme and the move of the physical newspapers to a new Newspaper Storage Building at Boston Spa
- Test the market for solutions where the supplier could be "self-funding" as experience with JISC had shown digitisation of newspapers to be expensive and time-consuming and requiring specialist skills and equipment
- Provide the Library with a clear exit strategy if it didn't work or no supplier could be found to undertake the work

A full business case for the project was created to PRINCE2 levels.

The project then derived a firm definition of what the Library wished to achieve and this was enshrined in the OJEU Notice which was the call for suppliers to respond as "interested" in the project and to take part in a Pre-Qualification process driven by a Library. This was advertised in September 2008.

We evaluated the PQQ responses (of which there were 18) to identify a shortlist of potential bidders who were then provided with a full definition of the project and invited to a Supplier's Day where the Library provided an introduction to the project, the requirements, and the Competitive Dialogue process.

We then moved into the full Competitive Dialogue process rejecting fourteen bids with four serious contenders for the contract:

After much work the Dialogue process reached the point where two suppliers were able to submit formal (and binding) Best and Final Offers.

On March 30th 2010 nearly two years later a contract was eventually signed with brightsolid opening up an extraordinary private sector public sector partnership to digitise 40 million pages over a ten year period with options to extend the contract in future five year tranches.

As a result of the contract:

brightsolid is digitising content from the Newspaper Archive to populate a commercial offering: brightsolid is responsible for the selection of material to be digitised based on their business plan.

brightsolid is operating entirely at their own commercial risk with no direct costs to the Library

The Library receives:

- a. Copies of all scanned materials for ingest (with accompanying OCR text and metadata)
- b. Free access from the Library premises to the contractor's online service
- c. For the duration of the contract the out of copyright wall was set at 1900
- d. A royalty on all paid-for accesses to out-of-copyright materials: this revenue is designed to recoup the additional costs to the Library (for example for the provision of materials to the scanning operation)

brightsolid is responsible for all negotiations with copyright (or other IPR) owners and for all due diligence on the copyright status of materials: the Library is indemnified for claims for breach of IPR or other damages

brightsolid may acquire materials direct from the publisher or rights holder and is expected to negotiate the right of ingest on behalf of the Library for these Acquired Images.

The contract is non-exclusive: that is, the rest of the Library's Newspaper Archive is available to be digitised. However, we have built in certain protections to help ensure commercial viability. This means that:

Any request or proposal for digitisation from the Newspaper Archive will be referred to brightsolid to enable them to take on that work if they wish and it agrees with their business plan but where they do not wish to be involved then digitisation may go ahead separately provided the content (and type of exploitation) does not directly compete with brightsolid's business model.

The full scanning operation came on stream in mid-December 2010 and www.thebritishnewspaperarchive was born. Access is free in the Library's reading rooms and through micropayments and subscriptions.

Since then the site has progressed to 5.5 million pages today with content ranging from the mid 1700's through to 1950. It is on target to reach some eight million pages by the end of 2012.

The inclusion of content from 1900 has been achieved through brightsolid's contractual agreements with rights owners. A unique aspect of this project is that the site has new content every day advancing knowledge and extending the potential for further research in real time.

The resource contains:

- News Articles - national and international events, as well as issues of local and regional importance. News articles into daily life in historical Britain
- Family Notices - search for your family's birth, marriage and death notices plus related announcements including engagements, anniversaries, birthdays and congratulations
- Letters - read letters to the editor written by the newspaper's readers, including illuminating contemporary debates, aspirations and anxieties.
- Obituaries view a wealth of contemporary information on the lives of notable individuals and ancestors
- Advertisements - these include classifieds, shipping notices and appointments..Illustrations - see photographs, engravings, graphics, maps and editorial cartoons

It has been an outstanding success: Success looks like this:

- More than 100 million pages have been viewed
- brightsolid has more than 50000 unique subscribers
- The Library is receiving royalties from brightsolid
- brightsolid is receiving significant revenues
- The project has already opened up a treasure trove of content
- It has transformed search for researchers all round the world
- It has proved that where there is a will there is a way

But that is only the beginning of the Library's developing story for newspapers. Our current challenge is to ingest live born digital newspapers.

To achieve this the hurdle we need to jump is UK legislation.

In April 2013 we are hoping that the Government will pass a new piece of legislation known as Non Print Deposit which will permit the Library to collect the entire UK web domain of over 10 million websites, born digital free newspapers and paid for newspapers published as digital surrogates of their print editions including those behind paywalls eg *The Financial Times*, *The Times* and the *Sunday Times*. We will also be permitted to collect ipad and tablet editions of newspapers.

But what about live radio and TV news? In the true spirit of our vision of aggregated connectivity we want to move on from just collecting news to enable researchers to view and research newspapers alongside multimedia platforms.

To that end we have been working with radio and television companies to capture broadcasts which will become available free to view in our reading rooms. To maximise our collection exposure we are also collaborating on a wider scale and have ongoing conversations with the BBC about linking into their archive from our reading rooms. We aspire to have similar linking to the British Film Institute.

Another area of activity is collaborations with other publishers and institutions like ourselves to create cross platform aggregation of digital assets. A fine example of this is Cengage Learning's latest product Nineteenth Century Collections Online (NCCO).

This contains British Library newspaper content and other collection material alongside content drawn from other institutions. It is a multiple year digitisation and publishing project that will culminate in an enormous number of primary source material from the "long, long" 19th century (mainly 1789 to 1914 but with some content stretching beyond that period) arranged in cross-searchable online collections by topics or themes. It includes books and monographs, broadsides, diaries, financial accounts, hand written sheet music, manuscripts, maps, pamphlets, periodicals, personal letters, photographs, poetry, statistics and much more.

Material is in Western and non-Western languages and will be released over a period of years. The collections available are "*British Politics and Society*" (1.7 million pages), *European Literature, 1790-1840*, "*The Corvey Collection*" (5.2 million pages of material in English, French and German. Future releases to follow this year are: "*Asia and the West: Diplomacy and Cultural Exchange*" (2 million pages) and "*British Theatre, Music and Literature: High and Popular Culture*" (1.5 million pages). There will be further releases in 2013 and 2014.

The system includes textual analysis tools and an image viewer to enable the manipulation, viewing and editing in a full screen mode. Alongside these initiatives we are, of course, very aware of the need to exploit our content to maximize the benefits for researchers. Some of the areas we are looking at to facilitate this are:

- The creation of a UK newspaper resource dataset, locally searchable via advanced text mining and semantic research tools optimised to identify new areas of investigation for digital scholars
- To enable experimentation and trialling of new digital research and data mining technologies with developers attempting to identify meaning or patterns in large scale textual datasets
- To generate statistics information packages and data visualisations derived from the full text designed to present the data in a way that inspires greater interest and further research
- Integration of the metadata/dataset content with external resources (e.g. Wikipedia, British Museum & National Archive linked datasets, search engines etc) in order to present search results from the data in a wider context to researchers and create new perspectives
- To link with internal resources such as UK mapping data, Newspaper Library catalogue and UK sound archive recordings to present an integrated UK resource for local historians, genealogical researchers etc
- As a key specialist resource for researchers interested in the history of the UK newspaper industry
- To integrate with other BL resources in the creation of interactive content for exhibitions with a UK theme (e.g. Writing Britain)
- To generate metadata to link with wider British Library catalogue data in providing first reviews of authors' published works, historical background for those researching authors' biographies etc
- To assist with scientific historical research e.g. climate change (via agricultural/weather reports), disease outbreaks/mortality and living conditions, natural phenomena (comet sightings, earthquakes etc)

The British Library vision is to collect, create and collaborate, to possess but also to point researchers to where content exists.

Patrick Fleming
Head of Operations and Services,
The British Library, London

