

Protect and Enable: Opening up The National Archives

Catt Baum: Conservation Manager Digitisation Services; and Gary Thorpe: Partnerships Manager
National Archives of the UK

Over the last 12 years, we've seen a huge increase in people accessing National Archives content online, with image downloads increasing dramatically (from 1.3 million in 2004/5 to over 130 million last year), and our presentation focusses on how we have successfully managed this through radical responses and strong collaborative relationships.

A short history of digitisation at the national archives

We have worked with third parties to ensure that the main series of records that relate to family history research are available to view and download online, and have leveraged £60million from the private sector to do this. As a result, Family History has become very much an online pursuit. In 1998 we were the first archive in the world to provide a fully online version of its catalogue and since then we have made increasing efforts to provide both improved catalogue data and copies of records over the internet.

We started digitising on a mass scale in 2002 using the vault of microfilm surrogates to front-load the process relatively cheaply, and once that project was complete, we moved on to digitise our primary physical collections. As we have progressed, our programme has needed not only to become more creative and diverse, but opportunistic and demand-driven.

The number of records digitised in partnership with commercial publishers has resulted in greater online access to our collections and raised revenue at a time of greatest need when we have a 25% cut in central government funding. In order to achieve this it has been necessary for us to undergo a seismic cultural shift in our thinking. It has been both a cultural revolution for us as an organisation and a different way of thinking for our commercial partners too. Not least in that we have had to convince those partners of the necessity and value of them picking up the bill for all conservation work prior to digitisation.

Duty of care

It's important to understand that conservation cannot be rushed so we take a very pragmatic approach to conservation in the digitisation environment. Our treatments are minimal but meet both professional ethical standards of conservation while taking into account commercial requirements. An initial assessment determines whether any conservation is needed. If so, a survey of every item is undertaken. This gives our partners and us a clear view of exactly how much and what treatment is required, and which pieces within the selection need any work before imaging. Each item is given a specific allocation of work-time so that we can predict the duration and cost of the conservation work. We have to get this as accurate as possible as you can't take a short cut even if time runs out.

Our processes are consistent but flexible to meet differing demands of individual projects and partners, and constantly refine things as we learn lessons or are faced with new challenges.

The work to create and maintain these partner relationships has been successful; their understanding of the importance and value of our work has noticeably improved, although some challenges remain due to the very specialist nature of our work. As with all collaborations, building these relationships is an on-going process.

Over several years, the number and complexity of digitisation projects at the Archives has grown exponentially. This factor, and the need to clarify for our partners the distinctions of

conservation for digitisation versus other kinds of conservation, forced us to re-evaluate and re-think our approach to the conservation element. The response allowed us to maximise resources and flexibility, and also to enhance our partners' experience of working with us through a much more efficient system, as well as take pressure off our Collection Care Department (CCD).

Development of conservation for digitisation

Before 2009, there was little direct contact between conservation and partner. Oversight for any conservation needed in a digitisation project sat with the Preservation Officer in Collection Care, who also had other duties. By 2009, an increase in the number of projects meant that the Preservation Officer could not keep up with demand, causing a negative impact on the Licensing team's ability to deliver projects and generate income.

In 2009, a permanent Digitisation Support Conservator role was created and I was appointed specifically to look after management of conservation in digitisation projects, although management of the project conservators still sat with the Preservation Officer. At that time, there were 2 large projects with 1 dedicated contract project conservator each, and the new role was responsible for managing those and any small additional projects. Within a fairly short space of time, the demand on this small team had outstripped resource and space.

By 2011, we appointed a second permanent conservator to work on 4 projects and assist in managing the team of 6 contract conservators.

Demand continued to outstrip resource and in late 2011 it was realised that a more radical solution was required to remove the bottleneck and relieve the pressure on CCD. As a result of, and in response to, the growth of digitisation as a tool for access, preservation and revenue, TNA developed a unique solution to the demands mass digitisation makes on a Collection Care department, which is not traditionally set up for commercial production-line working.

Radical change

The response was to completely re-design and divide our Collection Care functions, so that the digitisation conservators would be separated as a stand-alone team, and moved into the Commercial Delivery department to unify all digitisation processes within the same department. A separate studio was built for the team closer to the scanning teams, and all the management of both team and projects were consolidated under one Conservation Manager for Digitisation Support. This created a unified approach, allowing digitisation conservation methodology to be refined and changed freely as appropriate. The team would still always be entirely populated and managed by qualified paper conservators, and would closely liaise with CCD, ensuring the professional standards of conservation were adhered to, but the team would be separate and distinct from CCD. This meant they could continually improve both conservation and recruitment processes to maximise efficiency and keep up with demand, and be closer to the scanning teams to respond to queries and ensure implementation of the handling training all operators undergo before handling documents. The numbers of projects and project conservators continues to fluctuate, and the team has had a minimum of 7 project conservators for the last 3 years. To my knowledge, it is unique to have a conservation team that sits outside the main Collection Care department in order to focus on a specific conservation process.

The solution to divide conservation functions and move the team into a new department was borne out of strong collaboration and understanding between the various internal teams involved in digitisation at the Archives. Radical solutions such as this one require a high level of trust and cooperation from the top down for successful implementation. The Conservation is one key role in a digitisation project and requires a strong collaborative relationship with Licensing, our Digitisation partners and many other teams, both internally and externally to ensure a constant and efficient workflow for effective delivery.

Cultural shift

A cultural change was also necessary in other National Archive departments, from appealing solely to the seasoned scholar jealously guarding his knowledge, to the modern, online researcher, prepared to impart knowledge to others. Where once all our records experts had to know was the physical records and the information contained within each collection, with the growth of material being made available online records experts are now expected to know where the digitised version can be found online and indeed how each collection is searchable on any given website. We have developed from the role of simply guardian of our collections to that of enabler; the responsibility of guardianship isn't diminished, but has evolved to run in parallel enabling access whilst offering protection to the integrity of the original record. Convincing our stakeholders to buy into that thinking has been a major achievement over the past ten years.

Commercial partnerships

Commercial partnerships are at the heart of TNA's digitisation programme

Two of the major impacts on an archive working with the commercial sector to deliver services and resources is on the skill-set and organisational structure. Licensing is a vital element of leveraging commercial investment in fulfilling both access and preservation goals, so today's archives have to address legal and commercial demands for contract management.

Contract terms are the primary levers we have to manage our partnerships. Whatever shape of programme is suitable for the institution is defined in those contracts.

All of our licences are non-exclusive, and for fixed time periods – we never offer rights in perpetuity and these two conditions are non-negotiable.

Family History forms the bedrock of our partnerships and we are ever increasing the number of records available by partnering with the online genealogical companies – Ancestry, DC Thomson Family History, thegenealogist, My Heritage, to name but a few.

Academic partners

However, we have also been working with Academic partners, for example Cengage Gale, Adam Matthew Digital and Proquest for more than 20 years to provide access to our collections. Where once they took copies of microfilm, today digital images with a mixture of Optical Character Recognition and manual transcription to index key terms create searchable records, mean that academic publishers are now driving forward research as well as technological advances. Cengage's use of so-called 'Term Clusters' are designed to aid researchers with specific search terms – the categories are devised by an algorithm that creates clusters of terms based on the first 100 words of the first 100 search results per content type. In this case the key word is *famine in Ireland* and then you can refine the search further for each category. This is just one of the ways that research is being driven by the academic sector to encourage the innovative use of data. No matter the type of material online users increasingly expect to be able to interrogate data and the use of OCR and keyword search is a vital element of this.

Crowd sourcing

And in a crowd-sourcing project, images of our historic ships' log books are being used alongside books from other collections, to reconstruct historic weather patterns by using an army of volunteers originally on the Zooniverse platform from Oxford University's astronomy department to key in weather readings such as wind speed, wind direction and air

temperature. So we're open to multiple innovative ideas which encourage re-use of our records.

There is a wonderful and particularly appropriate quote from Francis Beaufort, he of the Beaufort Scale, Rear Admiral of the Royal Navy in the early part of the 19th century:

There are at present 1000 King's vessels employed. From each of them there are from 2 to 8 Log books deposited every year in the Navy Office; those log books give the wind and weather every hour.....spread over a great extent of ocean. What better data could a patient meteorological philosopher desire? Is not the subject, not more in a scientific than a nautical point of view, deserving laborious investigation?

More than 150 years later that is exactly what National Archives records, enthusiastic volunteers, meteorological experts and the internet is allowing us to do. In fact they go further. By tracking and comparing changes in wind and weather in the 19th century with today's weather they are able to measure the extent of climate change itself.

Volunteers

The use of volunteers to help sort documents in to date order as the very first step in our digitisation of the First World War Battalion War Diaries has been an immense success resulting in digitising 1.5 million pages to date. The battalion war diaries themselves whilst always extremely popular in our reading rooms have previously been overlooked by the genealogical companies as they were seen as not containing name rich data. With some 10,434 registered users, an army of volunteer 'citizen historians' have applied in excess of 500,000 tags to the pages to date – tags can cover, people, places, dates and activities – and debunked the myth about what the records themselves contain. They are immensely name, place, date rich and shed new light on the importance of these records. Without these 'citizen historians' much of the information within these records would remain untapped.

Conclusion

The success of our digitisation programme is a direct result of innovation and initiatives such as the streamlining of the conservation procedures. The redesign and evolution of our processes combined with the cultural change we have embraced as an organisation means we meet the needs of readers, publishers and online users whilst continuing to ensure the integrity of the public record.