

NEW CONNECTIONS: THE BT DIGITAL ARCHIVES PROJECT

David Hay

BT Group plc Heritage & Archives

Introduction

New Connections was a £1 million project which ran between 2011 and 2013, a partnership between BT Archives, Coventry University and the UK National Archives. Funded largely by Jisc, which provides digital services for UK education and research, the project was one of the largest of its kind, certainly amongst business archives, and a unique collaboration between the private, public and academic sectors.

The aim was to catalogue, digitise and publish online almost half a million images, a substantial proportion of the archive of photographs and historical documents assembled by British Telecommunications plc (BT) over its 168 year history. The project brought together those with expertise in heritage management and serious games technology, and academic experts in design, language, computing, education and learning resources.

Achieving this aim entailed the management of complex workflows including the development of appropriate metadata schema and a suitable online platform with search functionality and a user interface. The documents in the scope of the project required concurrent scanning, faded text recognition, cataloguing, development of new metadata and integration with the existing online BT Archives catalogue.

The final product is an interactive online catalogue and a major new online digital research resource of scanned archive documents, the BT Digital Archives <http://www.bt.com/btdigitalarchives>.

Background

BT is the world's oldest and most established communications company, with origins extending back to the UK's Electric Telegraph Company, incorporated in 1846 as the world's first national telecommunications provider. BT today is the major provider of telecommunications networks and services in the UK, an established global communications company serving customers in more than 170 countries.

BT Archives documents the leading role that the UK and BT and its predecessors have played in communications technology development from its very beginning, the provision of communications services around the country and across the globe, and their profound influence on society.

Physically located in Holborn in central London, the BT collections occupy over three kilometres of shelving, and include records, photographs and films of BT itself, records of the Post Office telecommunications function, and of the private telephone and telegraph companies taken over by the Post Office in the nineteenth and early twentieth centuries, altogether dating back to 1846. They evidence the development of the UK communications infrastructure and services to overseas from its earliest days, and major milestones such as the development of transatlantic communications from the earliest telegraph service to satellites and fibre optic cables.

BT's research archives, 1878–1995, in particular were included by UNESCO in the UK Memory of the World Register in 2011 as a unique record of over a century of British scientific endeavour and innovation (UNESCO, 2011). The archive as a whole documents a social history of Britain from the early nineteenth century, which was recognised in 2012 by Arts Council England awarding Designation status to all the collections as an important part of the UK's cultural heritage (Arts Council England, 2014).

Records created up to the date of BT's privatisation in 1984 are public records under the Public Records Acts, 1958 and 1967 (as amended by the Freedom of Information Act 2000). BT is almost unique among businesses in the UK in having public records responsibilities. Records created after privatisation are not covered by the Acts. As an approved 'official place of deposit' under these acts, BT Archives undertakes the company's statutory responsibilities to preserve and make available BT's public records to external audiences.

BT has always been proud of its rich heritage, which distinguishes the company from over 500 less well established competitors and supports BT's unique selling point as a company with over 167 years' experience of customer service and technological innovation. The company's commitment to safeguard this heritage on behalf of the nation is detailed online in the BT Heritage Policy (BT, 2009). BT is believed to be the only UK company to have published such a Board approved policy, but has always acknowledged a duty of care to safeguard this nationally significant legacy on behalf of the nation.

As early as 1876, BT's predecessor, the Post Office Engineering Department, was donating artefacts to the Science Museum in South Kensington and this relationship continues to the present day with BT being a Principal Sponsor of the Science Museum's new communications gallery 'Information Age', opening on 25 October 2014. BT also invested over £6 million in the ten year Connected Earth partnership project (www.connected-earth.com) from 2002 in the dispersal and display of its own collection of historical artefacts to partner museums around the UK (Hay, 2004, p. 45-60).

Similarly, BT has also always looked to improve the way its archives collections can be enjoyed and studied, and their true significance and potential fully realised. However, despite the scale and breadth of the archive its reach and usability was limited, hampered by restricted physical access in a shared building, which is also a BT operational telephone exchange, with most of the collections in analogue form only.

As with the Connected Earth project, BT sought to develop external relationships to raise the profile of the archives collection and to engage with a wider audience. As a first step, BT followed the Marks and Spencer plc approach and in 2009 invited an academic appraisal of the collection (Carter, 2013, p. 1-18), undertaken by Professor Peter Scott of the Henley Business School at the University of Reading, which concluded

'The Archive represents BT's heritage, but beyond that reflects the leading role that the UK has played in the communications revolution from its earliest days...BT has a world-class corporate archive which has great potential to be of considerable value to the company, academic researchers and (partly through the publications of those researchers), the wider community (Scott, 2009, p. 2).

With this independent appraisal, BT then sought external partners to further academic and public recognition of the significance of the collections. To the same end, work began at this time on applying for UK Memory of the World inscription and Designation status. As a business,

BT has wide experience of partnership working to bring new insights, skills and knowledge bases, as well as to pool resources and capacity. BT similarly focused greater attention on seeking opportunities to work with a range of institutions to access funding normally closed to BT as a company, and to deliver a greater return on its own significant heritage investment. From BT's wider university research programme it was realised that academic partnerships in particular would have the potential to broaden our audiences as well as to deepen our own understanding of the collections. Although challenging in the difficult economic climate of recent years, the results of these efforts have been fruitful and inspiring.

In recent years, BT Archives has partnered with universities including Salford, Leeds, UCL, University of the Arts, London, Manchester University, Royal College of Arts and others, including again the Science Museum, on a range of collaborative Masters, PhD and other postgraduate projects. Some of these have been funded by BT, but frequently also with the assistance of bodies in the UK such as the Engineering and Physical Science Research Council (EPSRC), the Arts and Humanities research Council (AHRC) and others. Some of these relationships continue, but the biggest partnership project we have undertaken so far has been the New Connections / BT Digital Archives collaboration with Coventry University.

Coventry University Partnership

Coventry University, established as a university in 1992 with roots back to 1843 and a long tradition as a provider of high quality education and multidisciplinary research, was one of a number of institutions that responded to an invitation in late 2010 through Universities UK and the Russell Group of UK universities to explore partnership working with BT Archives.

Coventry's response stood out partly because of its technological experience and expertise in digital heritage. The university is known as entrepreneurial (later winning the Times Education Award for Entrepreneurial University of the Year in 2011) and is one of the top 25 UK Universities for working with business. It has a record of technological innovation, with particular technical expertise in digital cultural heritage among software designers and engineers at the 'Serious Games Institute', a wholly owned subsidiary of Coventry University and a centre of gaming technology excellence. Successful previous projects have included immersive technology initiatives such as Siobhan Davies Replay (Europe's first digital archive of dance) <http://www.siobhandaviesreplay.com/>⁷. Coventry University had, in fact, already partnered BT on an earlier project in 2005 to develop *Stratford Unplugged*, a Personal Digital Assistant (PDA) based interactive guide for tourists to get information on restaurants and places of interest in the centre of Stratford. The UK's first wireless broadband application for tourists, it used BT Openzone wireless hotspots to deliver relevant information to users as they walked around the town centre (Luft, 2007, p. 10)

What particularly attracted BT Archives to a partnership with Coventry was the University's obvious enthusiasm for the realising the wider potential of the archives collections themselves. In research terms Coventry recognised that the collections could add significantly to the online cultural resources of the HE/FE sector by offering accessible and engaging scientific, cultural and historical material for use by teachers, students, practitioners and researchers within the UK. It was also an opportunity to demonstrate the University's proven digital expertise in new media to develop innovative technologies which to open the collections to a much wider public and academic audience.

Coventry University is particularly enthusiastic about collaborating with BT on this project because it will enable the University to bring its unique expertise in the building of digital archives to an important historic collection for wider public and academic engagement (Coventry University, 2011)

Discussions between BT Archives and representatives of the academic, technology and business planning communities at Coventry University eventually led to a framework collaboration agreement between the two institutions to work together on securing funding for digitisation initiatives, signed in August 2011. However, even before the agreement was formalised, the Jisc Digitisation and Content Programme was identified as a potential source of collaboration funding.

Project bid

Jisc is the UK's technology consortium for colleges and universities. It supports higher and further education by providing strategic guidance, advice and opportunities to use Information and Communications Technology (ICT) to support research, teaching, learning and administration. Under the Jisc Content Programme 2011-13 funding of around £5.4m was available for projects related to the creation and use of digital content (Jisc, 2011).

As the deadline for submissions to the programme was mid-August 2011, the project bid document had to be drafted and agreed within a few weeks, led by a business development manager at Coventry University. The proposed project was ambitious in the scale of its objectives, and it was a remarkable achievement that such detailed proposal requiring the approval of so many stakeholders in both institutions could be drafted and agreed in such a short time.

The proposal was submitted on the date of the deadline, and the successful projects were announced in October 2011. Jisc had received a total of 109 bids from Higher Education establishments across the UK for a variety of digitisation projects. 23 projects eventually gained awards, covering three project strands, with Coventry University and BT submitting one of only seven successful bids chosen from 68 to work on a Strand B: Mass Digitisation funding stream project, which was to focus on embedding content in teaching learning and research. As the lead partner, Coventry University received £744,873 to work in partnership with BT on the project officially named 'New Connections: BT e-Archive project', which had a total value of £1.07million including projected in kind contributions from BT and Coventry.

Approach

The objectives of the New Connections project were;

- the digitisation and publication on a sustainable basis of 445,000 documents and images;
- the embedding of the e-Archive in the University's own teaching, learning and research work;
- the development of resources for teachers, learners and researchers;
- to engage science and technology enthusiasts and the wider community through an online platform which included an interactive gateway and user generated capability.

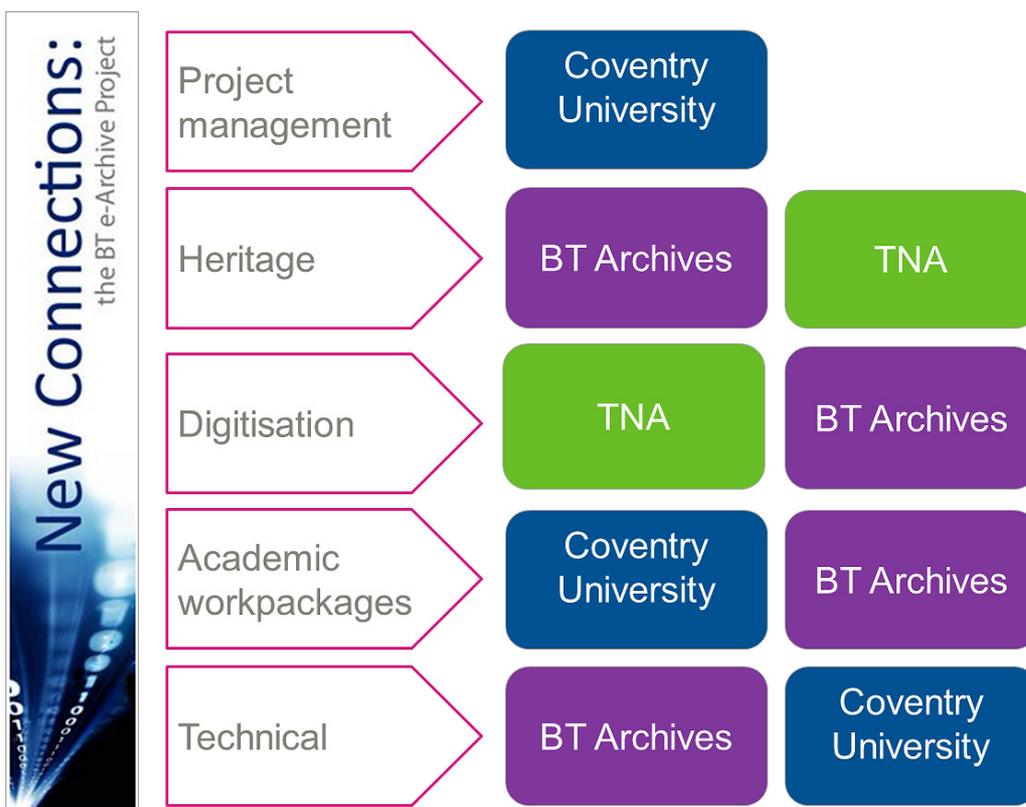
The project had a strictly defined duration of 20 months to conclude in July 2013, and had to begin almost immediately. It was obvious that achieving these objectives would entail the management of complex workflows, and it was quickly decided to invite The National Archives (TNA) to participate as a third partner to manage the actual digitisation element of the project. TNA has wide experience and expert knowledge in scanning archival documents, and their participation as research partner was a critically important contribution to the success of the project. To this point, TNA's digitisation activity was largely around its own collections, so this project was an opportunity to process the digitisation of a large third party collection to increase

their own experience of working with different and varied collections and formats, and to attract wider interest in future external collaboration with other partners.

With a project of such breadth and complexity the project plan laid out nine distinct work streams, but broadly there were five concurrent intersecting elements:

1. Project Management – keeping the project on track and reporting to Jisc
2. Heritage –selection and transfer of archives to TNA
3. Digitisation – conservation, scanning, and transcription/OCR
4. Academic - exploration of the use of the new digital content by Coventry University for research, teaching and learning
5. Technical – design and build of the online platform, and integration of the digital content and catalogue metadata.

To facilitate management of the project there were four groups reporting into the overall Project Team: Heritage (BT); Digitisation (TNA); Technical (Coventry/SGL); Academic (Coventry); and Project Management (Coventry). As the illustration below demonstrates, each partner took the lead for at least one broad work stream, although in practice each of the partners was involved in all of them to some extent.



New Connections work streams © BT Archives

Project Management

As the lead partner and the official channel with Jisc for communications and the actual funding, Coventry University managed the official administration and reporting of the project, the cost of which was included in the Jisc grant. An experienced project manager with administrative support was allocated by the University, under the overall supervision of the Project Director, whose full time role was Research Director for the University. All three partners were represented on the Project Team, including representatives from the four subgroups. A fully collaborative approach was taken, with progress, issues risks and next steps discussed and agreed between the partners at monthly meetings. In addition, an Advisory Group met three times during the course of the Project, a diverse mix of representation from academics from other universities, industry experts, the heritage sector and the media.

Heritage

Priorities for digitisation, agreed by BT, and Coventry University academic and technical staff with reference to the Advisory Group, were based on the value of various elements of the collection for academic and wider public use. Given the scale of the collection it was decided to concentrate on three categories of content:

Category A - Photographic images, 1900 - 1975

BT's photographic collection is particularly rich and consists of up to 500,000 individual prints, negatives (glass and safety film) and glass lantern slides, in various subject (Engineering, Publicity, BT Tower, Cable Ship, Satellite, Training School, Lecture etc) or date series . The greater part were produced in-house by the Post Office and BT and, because the main Post Office reprographics function was in the Engineering Department in the telecommunications business, it accompanied the crossover into BT on separation in 1981 and subsequent privatisation in 1984. The core of this collection is in fact the original Post Office photograph negative library, including all the original photographic registers. Many of the old series continued under BT following the split from the Post Office in 1981, ceasing around 1997 when photographic work was outsourced.

For the New Connections project it was decided to concentrate on as producing as many sequential images as possible from the two main series in the photographic archive; the Engineering Series, 1900 – 1969, the original series produced by the Post Office, and the Publicity Series, 1934 – 1975, set up when the Post Office began its publicity machine in earnest in the early 1930s (Hay, 2011). Both series assimilated numerous earlier telephone and telegraph company images, so in fact some of the selected images date well into the 19th century, the earliest being around 1865. Subjects covered include various aspects of communications technology over time, but also shows the technology in use by consumers and workers, street scenes, working practices and so forth and so are a valuable source for social and local history. For example, the numerous views of war damage to buildings and streets during the WWII Blitz. In the event, around 50,000 negatives or prints were digitised under the project, and BT Archives additionally contributed around 3,000 existing digital images from other photographic series, extending the date range into the 1990s.

Category B - Research reports, 1878 - 1981

Both BT and Coventry were anxious to feature the Post Office and BT research report archive in any collaborative project. The reports are specifically acknowledged in the UNESCO UK Memory of the World Register as being particularly significant. The earliest reports dating from 1878 are adhoc manuscript reports which reflect early Post Office investigations into telegraphy,

following the takeover of UK domestic private telegraph companies in 1870 which gave the Post Office a monopoly over telegraph communications. They also reflect the Post Office's early experiments in telephony shortly after its introduction in the UK in 1877. Other early manuscript reports cover Post Office experiments with wireless telegraphy before the wireless radio pioneer Guglielmo Marconi travelled to the UK in 1896, and then with Marconi's system soon after his arrival (Marconi's first public transmission of wireless signals was from the Central Telegraph Office in London, now the site of BT's headquarters building, on 27 July 1896). Subsequent reports cover the work of the GPO Dollis Hill Research Station from the 1920s, and from 1975 at Martlesham Heath in Suffolk at what are now BT's research laboratories following the move from Dollis Hill.

The breadth of the overall subject matter is extensive, covering not only telegraph and telephone technology, but related topics including radio and satellite communications, lasers and masers, data processing and early computer systems, time measurement systems, television broadcasting, early optical character recognition, videoconferencing and video phones, advances in semi-conductors including thermionic valves and transistors, and world first developments in optical fibre cables. Produced by an organisation that held a monopoly in providing telecommunications services for much of the twentieth century, they are a unique record of British scientific research and contribution to developing communications and related technologies.

Despite this, until recently the research activity of the Post Office and BT has attracted relatively academic interest, the only real exception being Colossus, the World's first programmable digital electronic computer designed and built by a team of telecommunications research engineers led by Tommy Flowers at Dollis Hill for the Government code breaking centre at Bletchley Park in the Second World War (Copeland, 2006). For this reason, it was an early decision to include as many of the research reports as possible. Ultimately, all the reports from 1878 to 1981 were included in the scope of the project, including a separate series of radio research reports, 1925–1960 which document the Post Office's pioneering work in this field. Altogether, around 190,000 individual digital images were created in this category.

Category C - Policy and correspondence, 1840s – 1980s

This category as a whole broadly covers all policy files, from the first encounter of the Post Office with electrical communications to near-current BT board records and papers. They are particularly strong from the nineteenth to the late twentieth centuries, during which time the Post Office followed the civil service registry system of records management, a tradition that was carried on for some years by BT following privatisation. An incredibly rich and complete archive that charts the development of telecommunications in the UK and overseas has been left from an analogue age of limited reprographics, when the control of information on paper was tightly controlled, and when the organisation was highly centralised and rigidly hierarchical. The files cover every aspect of the business, including the takeover of the telegraph and telephone companies; introduction of telephone to UK; advances in technology from the earliest telegraphs to optical fibres, imperial and global communications from cables to satellites, broadcasting and media relations to list only a few of the subjects covered. Wider social themes are also documented, such as race relations and employment of ethnic minorities, and the role of women in the workplace and gender relations. As the organisation was so centralised for much of this period, the papers in files typically reflect an issue moving up the management chain for a decision, often to the Postmaster General or Managing Director of Telecommunications, and then back down the organisation as it was implemented, progress being recorded at every stage. Later files also cover BT's transition from a nationalised industry to a privatised global company, and its transformation to compete with newly created competitors in the market place.

Unlike categories A and B, which were relatively straightforward, selecting which files were to be included in the scope of the project for Category C was much more complex. Candidates under this category numbers upwards of 7,500 files in total, and could never all be covered within the available budget and timescales. Many of the large files incorporate numerous sub-files extending to several boxes, with multiple drafts and duplicates. Originally, a thematic approach was considered similar to the one adopted in the earlier 2008-09 TNA Cabinet Papers project (TNA, 2008), an obvious point of reference and source of inspiration for the BT project, using selected papers from a wider number of files. However, it was quickly realised that this process was completely unfeasible within the existing financial and time resources available, and the result would be a wholly subjective set of records of limited research value. Following discussion with the external Advisory Group, it was decided whole files should be scanned rather than selecting certain papers, so that higher levels of credibility and lower levels of personal subjectivity would be achieved. This ultimately meant that, although we scanned the same number of individual documents, far fewer Category C files were included than originally hoped. It also meant selecting which files should be included was much more difficult, and was itself inevitably a subjective process.

Consulting with the Advisory Group, a list of research criteria was agreed

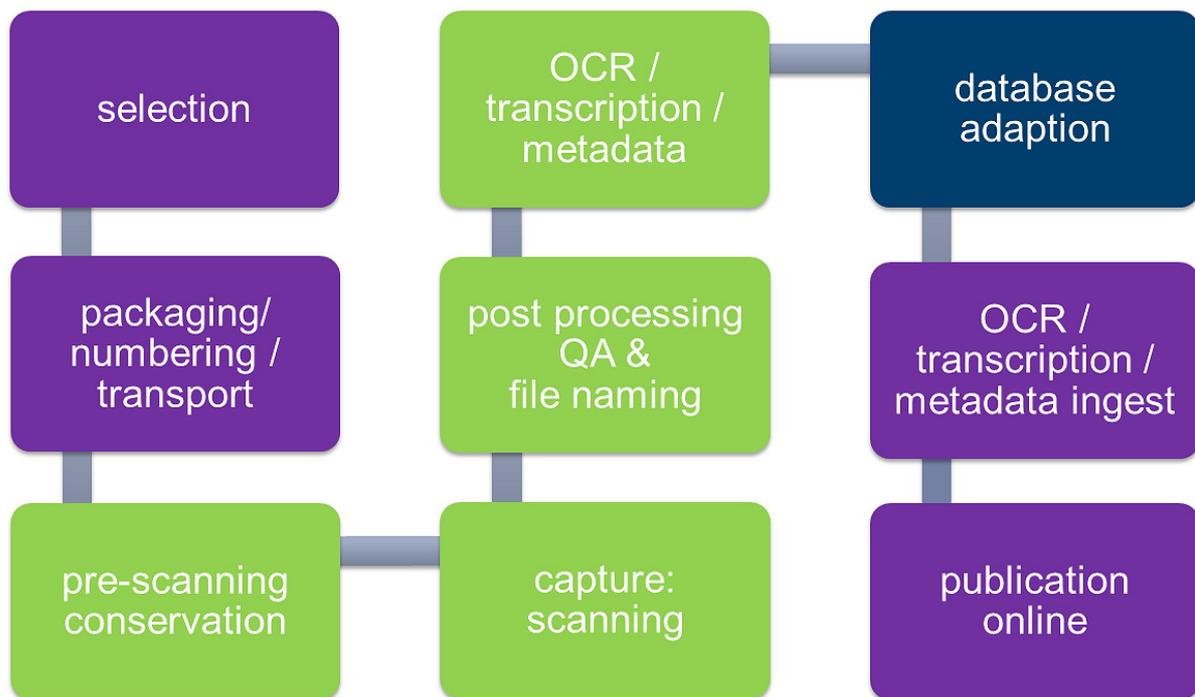
- technology milestones
- national events
- international reach
- Government policy
- industrial relations
- diversity
- iconic products & services

The most appropriate files which met these criteria, and which complemented the major themes of Categories A and B as far as possible, were selected by the BT Archives team based on their knowledge of existing and potential research value in the collection. Ultimately, 440 files were digitised, encompassing around 230,000 individual digital images. The files ranged in size from folders of a few individual pages, to one extreme example that consisted of 16 archive file boxes.

Following selection the next task was to parcel up and transport the archives to The National Archives in Kew, where the documents and photographs were to be assessed for conservation pre-scanning and then digitised. For the small BT Archives team this was a huge logistical exercise involving inserting protective wrapping and padding for the files, reboxing where necessary, and liaising with TNA over the safe transport and return of the boxes to and from Kew. All of this activity had to be carefully tracked. Altogether, a total of 1006 boxes across all three categories were transported between January and December 2012 in 16 separate batches. Notes accompanied each batch detailing the contents and arrangement of the records in that consignment, metadata requirements, image file-naming conventions and scanning protocols to be followed.

Digitisation

Although managed as a separate work stream led by TNA, restricted timescales meant that it ran concurrently with the Heritage activity, with teams from both BT Archives and TNA liaising closely at all times, particularly given the high degree of flexibility that was required from both as the project progressed. The combined workflow is broadly reflected in the illustration below



Digitisation workflow © BT Archives

TNA has a vast collection of documents – largely archival but also some artworks and surprising objects often found in police evidence files. Almost any of TNA’s collection can be put forward for imaging so the digitisation team are adept at adapting techniques to a wide range of objects. BT’s collection consisted largely of paper documents in different formats, and a huge collection of negatives, both glass plate and plastic media, all of which are familiar to TNA from its own collection. Nevertheless, this was the largest digitisation project of an external archive undertaken by TNA, so the TNA production process had to be reviewed and streamlined, which entailed adapting the standard procedures and buying new equipment to speed up processing time.

Everything digitised at TNA is first condition checked and assessed for conservation work at the start of a project. Pre-scanning conservation input is critical to the success of a project to assess the risks of damage to documents from the physical digitisation process, and to identify existing damage which would affect the legibility and successful OCR scanning of a document. It was important to apply same standards to BT’s archives to ensure that future workflow would not be disrupted, costs would not escalate, and that TNA’s own reputation for quality of work would not be compromised.

An initial visit to BT Archives allowed the lead TNA conservator on the project to spot check a random selection of BT documents before they were transferred to TNA to look at the format, material, condition and any issues may affect scanning, to plan what would be needed in terms of conservation. She also spoke to the BT archivists to get a sense of the collection in terms of format and condition, to obtain as much information as possible so the TNA conservators and camera operators would know what to expect. Following transfer to TNA for scanning, the documents were assessed more fully before digitisation. Records without damage were sent

straight to camera, those requiring work were directed to the custom built conservation studio for digitisation support, and then to camera once repaired.

For the digitisation itself different processes were used to meet the different requirements of the Category A photographic records and the Categories B and C largely paper records.

- The Category A records were a mix of glass plates, acetates and photographic prints. They needed minimal preparation and the biggest workflow challenges were around the careful handling of the documents. ICAM Guardian cameras were used for capture, with around 50,000 scans being produced over seven months. It is important to note that this was largely conservation scanning of original archive negatives, with no subsequent retouching so that the images produced were not to gallery standard.
- Category B and C records presented more of a challenge. The documents are in a variety of formats – tagged files, bound volumes, stapled pamphlets etc – with numerous foldouts. There are varying qualities of paper with both manuscript and typescript text on one or both sides, often in the same file. Considerable preparation was often required pre-scanning ICAM Guardian digital camera workstations were used to produce around 425,000 scans over nine months.

Image capture began in March 2012 and finished in December, a total of 10 months. In all three categories, the output for each record was a 400dpi tif file, a 72dpi jpg and a resized thumbnail image.

Other activity undertaken by TNA as part of the project included;

Transcription

Not all of the Category A photographic records had catalogue metadata, so TNA scanned nine volumes of original handwritten registers to both the Engineering Series and Publicity series of photographs and sent the 6,500 scanned pages for off-shore transcription to be ingested into the final online catalogue to link to the relevant scanned images.

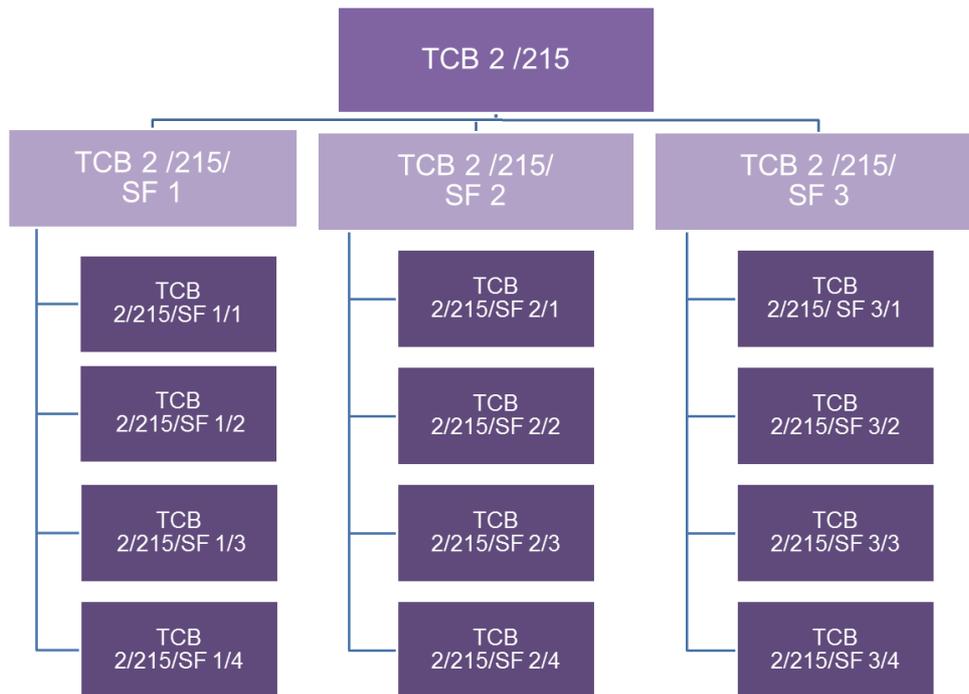
Cataloguing

Similarly, the Category B research reports were similarly largely uncatalogued, so TNA staff recorded the title, author and date of each report as it was scanned, again to ingest into the final catalogue.

Optical Character Recognition (OCR)

The scanned documents on Categories A and B were OCR scanned, again off shore by an external contractor, so that each scan had a text file that would be invisible to users of the ultimate system, but which would make a large proportion of the records searchable by free text searching. Since many of the documents are manuscript, it had to be accepted that this would not be possible for all records, and that manual transcription would not be feasible.

BT Archives and TNA worked closely on file naming conventions to ensure that all of the existing and new metadata would match so that individual digital images could be retrieved. This was a complex process in many cases, as the Category B and C files were not catalogued to the individual page level, nor in the case of Category C files to the sub-folder level, meaning that the additional levels had to link to the existing catalogue records.



Additional metadata levels linking images and catalogue records © BT Archives

Academic

The Academic Team of Coventry University research and teaching staff, supported by the Heritage and Technical Teams, delivered two major higher education specific outputs: research projects which made innovative use of the newly digitised archives resource and a web-based Learning Space with pedagogic tools. The academic work packages were varied and wide ranging.

Problem Based Learning (PBL) scenarios

Five PBL scenarios were developed under the project based on BT archive documents; Disability and inequality in employment (Social Care and Health); Race and racism in healthcare (Social Care and Health); Exercise warm-ups in the 1910s and 1930s (Sports Therapy); Text in advertising (Graphic Design); and Gender in advertising (Marketing).

PBL is a methodology whereby students are given a particular scenario or problem. They then discuss as a group and discover a solution or several solutions. There is no right or wrong answer. Archive resources were found to be an excellent way to prompt students to identify the gaps in their learning needs. Because they are archive-based and therefore generally older, these resources could often contrast with students' disciplinary and cultural understandings. The archives selected prompted discussion and challenged students to think differently about things they had come to take as given. This particular approach to PBL would be extremely difficult without ready access to these kinds of resources.

It is extremely rare for archive resources to be used in disciplines other than history, and it is believed that this is the first example of problem-based learning which demonstrates that archive resources can be used in disciplines other than history. At least three of the five scenarios are

likely to be implemented permanently in Coventry University courses, in health and social care, and graphic design.

Linguistics research project

In the Linguistics work package the academics selected and transcribed 500 letters from the digitised content between 1853 and 1982 to study language change and how the letter genre developed over time in a business context. Correspondence was linked to the Coventry University Word Tree interface allowing users to explore linguistic features such as keywords and phraseology (Coventry University, 2013).

Some of the letters selected for inclusion had obvious historical significance because they were written by, or on behalf of, well-known people such as Guglielmo Marconi or Alexander Graham Bell. Also included was a great deal of 'day to day' correspondence which touched on topics of historical interest such as the employment of disabled workers, the consideration of employees from former colonies, the establishment of the transatlantic telephone service, the design and installation of a new style of telephone kiosk, and the impact on members of the public of the cessation of the telegram service.

Case studies

Research case studies were also produced on design (studying the iconic 1960s Trimphone telephone), employment of women in Post Office Telecommunications in the 1960s, and on computing strategy and procurement in the Post Office and BT.

Technical

The technical team was led by experts from the Coventry University Serious Games Institute (SGI), with support from the other partners. The ultimate purpose of the project was to publish the digitised content online as a research resource for the higher education sector. Jisc were also clear that the final website should have a wider popular appeal and include a degree of interactivity for the wider community and allow user generated content. So the real challenge was to make the archive available, accessible and interesting to all.

Discussions at the Project Team and Advisory Group soon revealed varying though not necessarily conflicting expectations of what the website would deliver. Discussions were also informed by a user survey conducted BT Archives on its existing online catalogue www.bt.com/archivesonline, launched in 2009 as one of the first such catalogues in the business archives community. 194 people responded ranging from academics to general recreational users. Interestingly, the data showed that respondents were more concerned with a good conventional search tool rather than more modern features such as social media. Other important aspects were the ability to view photos and documents, and to download files.

A set of essential requirements was discussed and agreed at a dedicated externally facilitated workshop, which established that the website should be

- visually compelling, to appeal to a non-academic audience
- able to meet search capability expectations of academic users as a research resource
- interactive, and enable crowdsourcing and user generated content
- robust and sustainable
- hosted by BT, complying with BT design standards

- incorporate and enhance the existing BT Archives online catalogue
- able to include a learning space encompassing research projects
- delivered within budget and timescale

In order to meet this specification the Technical team had to consider building a completely new system platform, using open source software such as DSpace, or adopt an existing commercial product. Given the limited timescales it was decided to follow the second route, combined with an interactive search portal that would be developed by SGI to give it the visual appeal that Jisc stipulated for a wider audience.

Since BT Archives was already using the CALM back-end catalogue platform and DServe online catalogue portal, both supplied and supported by Axiell and well known to the archives community, early attention was given to a new Axiell product, the Arena web delivery portal which offered many possibilities towards meeting the project's objectives. Arena was a system originally aimed at the library domain, where it is now in widespread use. For this particular application additional development work would be needed to fulfil the requirements, but Axiell were keen to participate in what would be the first use of Arena in the archives community in the UK, and anywhere on this scale. Following further discussions, it was decided to go with Arena as the front-end catalogue delivery portal, populated from a back-end CALM server with live catalogue data. Axiell were closely involved in developing the solution from that point, becoming more of a partner in the project than a supplier.

At the same time, SGI were developing the interactive search gateway using largely open source software and ultimately produced the Mosaic, which allows users to browse the digitised archive through genuine mosaic images created by a software algorithm made up a randomized arrangement of thousands of scanned photographic and document images from the archive, allowing serendipitous searching backed by the metadata fed through Arena. Ultimately, the Mosaic interface was integrated with the customised Arena platform and the final product - branded 'The BT Digital Archives' - was launched at an event at the top of the BT Tower in London in July 2013.

The screenshot shows the BT Digital Archives home page. At the top left is the BT logo and the title 'The BT Digital Archives'. To the right are navigation buttons: 'Search BT Archives', 'Academic Study', 'Mosaics', 'My Research List', and 'Help'. Below this is a purple header bar with the text 'Male operators working the switchboard at Buckingham Palace, 1935 (TCB 473/P 0621)'. The main content area is divided into three columns. The left column is 'Sign in', containing a form with fields for 'Username' (containing 'selfgar') and 'Password' (masked with dots), a 'Sign in' button, and links for 'Forgot Password?' and 'Create new User'. The middle column is 'Search the archives', featuring a search bar with the placeholder 'Type here to search', a 'Quick catalogue search' button, and logos for the United Nations Educational, Scientific and Cultural Organization, BT Research Centre collection (1875-1995), and the Memory of the World Programme. The right column is 'About BT Archives', containing text about the project's collaborative nature and funding by Jisc, and a link for 'About BT Digital Archives'. The footer contains logos for Coventry University, The National Archives, Jisc, and BT, along with links for 'Contact BT Archives', 'Site map', 'Privacy policy', and 'Terms of use'.

BT Digital Archives home page © BT Archives

The screen shot of the BT Digital Archives shows a Mosaic image incorporated into the home page. Users can zoom into the image revealing the individual images as thumbnails, any of which can be clicked on to bring up that image itself and its catalogue metadata. A different Mosaic is loaded each time the page is refreshed or visited, and users can choose their own mosaic from a selection of over 150 images. This highly visual interface entices visitors into the archive as each image links back to individual records through the displayed metadata.

For academic users there are both simple and advanced search options for more conventional searches of the Digital Archives. Anyone can quickly search the system and view images of documents and photographs if they have been scanned, or if not view the standard catalogue data. As with its predecessors DServe and CalmView, a graphical tree browser in Arena enables material to be viewed within its hierarchical context.

The site also includes a learning space, incorporating the research studies and case studies produced by the Coventry university academics on the project, and links to almost 100 social media channels from every page so that individual pages, images and documents can be shared easily.

For additional features there are two levels of registration.

General registered users

General Registered Users get immediate authentication on registration. As well as normal searching and viewing they can also;

- view and download 1000 pixel (longest side) images where they are available
- search the OCR digital content of documents that have been scanned
- create and keep their own research lists for future reference and research
- add their own comments, tags and knowledge on individual photographs and records, that other registered users can also search, view and supplement

Academic Registered Users

Users from the academic community, including teachers and trainers attached to an educational or training institution, can get a higher level of functionality. As well as having General Registered User benefits they are able to;

- view and download 2000pixel (longest side) images
- view and download pdf documents of whole reports and files where they are available, for study off-line.

Academic User access has to be validated to make sure those applying are attached to an approved university, college or institution. In the meantime, they enjoy the same immediate access privileges as General Registered Users.

Communications and dissemination

Although not a specific work package, considerable focus was given to engaging with external stakeholders and a communications strategy was drawn up. The start of the project was officially launched in January 2012 with an event for stakeholders at the London campus of Coventry University. During the project two seminars were held in London and Coventry to disseminate the initiative, attended by over 100 people from the heritage sector, the engineering profession and academics community. Members of the Project Team also spoke to a number of conferences and societies in the UK and abroad. Social media was an important part of the strategy, and accounts for Twitter, Pinterest and a blog were set up for the duration of the project, with the partner organisations' main social media accounts also being used to disseminate the project.

In the last few months of the project, a separate team of marketing professionals from the three partner organisations met by phone monthly, and nearer the launch date weekly involving the Jisc press office, to ensure that the marketing message was consistent and that all brand requirements were met. This enabled the project to reach wider, subject specific audiences through the varied types of contacts that each partner had within the academic, technical, museum and archive sectors as well as through mainstream media. This resulted in widespread coverage following the launch of the BT Digital Archives in July 2013 with accompanying co-ordinated press releases by BT, TNA and Coventry, with over 26 press and online articles nationwide including the BBC, The Times Higher Education, The Economist and many specialist channels for the technical, engineering and heritage sectors.

Challenges and lessons learnt

As part of a Jisc funded research programme, one objective of the project was to learn from the experience of managing a large collaborative partnership to inform future initiatives. This was a massive and complex project with varying expectations from many interested stakeholders, so a number of challenges were to be expected from the outset. The biggest overall and most obvious challenge was to complete within the 20 month timescale, and this was made more challenging by a number of unforeseen issues that had significant impact and caused further delays as the project developed. The biggest lesson learnt is to build as much contingency into

a project plan as possible to allow as much flexibility as possible should unexpected issues arise. They always will. Other significant challenges and learning points that could guide future projects include;

Partner cultural differences

There were three official partners to the project. Given that representation from Coventry included academics from the University itself and creative technologists from the SGI, and that the supplier Axiell became closely involved during the Arena development, in practice there were actually five partners. Jisc were also represented at every stage. Behind all these stakeholders were their respective management chains and bureaucracies (and their inevitable re-organisations), and additional representatives such as lawyers and tax experts. From a Project Management perspective working with a team that included such a mix of backgrounds and influences posed some interesting challenges and some initial conflicts, including differences in organisational cultures and different perceptions on pace and depth of tasks. With good will and flexibility on all sides, a mutual respect and understanding developed within the team and any difficulties were overcome to deliver the project successfully.

HMRC and tax implications

Another lesson learned was that rules and regulations regarding VAT and partnership status should be determined at the project writing stage. Despite the fact that this was a research project under a research programme funded by a research body, HMRC insisted on regarding TNA as a commercial supplier to the project and that VAT should be paid. This was a huge surprise since not only were TNA an active research partner in the project, but the HMRC ruling conflicted with precedent and legal advice and appeared to contradict HMRC's own published guidelines at the time. Discussions between HMRC and advisers from the partner organisations held up the scanning workflow of the project for some months, which created further challenges and pressures later on. HMRC proved immutable, and the consequence was that twenty per cent of the scanning budget of c£550, 000 was lost at a stroke. All the partners had to demonstrate considerable flexibility to ensure that the committed number of digital images was delivered. It is ironic that TNA, who were never going to realise any direct commercial benefit, had to work extremely hard in adjusting their workflows to enable cost efficiencies, whilst at the same time not giving any public subsidy which is not permitted under HM Treasury rules. Ultimately, the project delivered and published over 475,000 digital scans, 30,000 more than estimated in the original Jisc application. Given the time, financial and other constraints this was a significant achievement.

Pre-scanning sampling and preparation

Because of the tight timescales only one day was allocated to on-site conservation sampling by TNA before digitisation commenced to plan the workflows necessary. Conservation work needed on the material proved significantly higher than anticipated leading to delays in scanning. Seven per cent of the whole project ultimately required pre-scanning conservation to some degree, rather than the original two per cent estimation, so TNA had to adapt their processes and workflows to meet this challenge.

There was some confusion between BT Archives and the TNA scanning team about both digital file and archive record naming conventions. This caused delays and put further pressure on staff time and resources at both BT and TNA. Some work had to be revisited to ensure that the file names matched the catalogue metadata so that Arena would be able to retrieve them. The lesson learnt is to ensure clarity in naming conventions and to put quality assurance processes in place during the scanning processes as well as at the end.

Digitising uncatalogued archives

A large proportion of the records selected for digitisation in Categories A and B were partly catalogued or not catalogued at all, one of the objectives of the project being to undertake this activity. Time pressure resulting from other challenges impacted this activity so that not all the new catalogue entries created by OCR and manual transcription could be fully proofed to ISAD(G) standard before launch. This is having to be revisited following the formal end of the project.

The number of individual pages in the partly catalogued Category B research reports series was also considerably more than originally estimated. As the digitisation of the reports in full to 1981 was at the heart of the whole project, this meant that the number of records that could be scanned in the other two categories, particularly the Category A photographs had to be reduced, though the total number of scans ultimately delivered exceeded the original estimate.

Conclusion

The BT Digital Archives was a significant achievement, particularly given the amount of work undertaken in the available timescale. It has brought a hitherto largely unseen and under-appreciated collection to a much wider audience. All the partners have learnt a great deal from the project, and have appreciated the opportunity by Jisc to collaborate with new colleagues in other sectors. They are proud that the New Connections has produced in the BT Digital Archives an innovative online resource, freely available under a Creative Commons licence to encourage sharing and the use of the material in education curricula and research. Going forward, BT will continue to develop the site as its new online catalogue, to add further digital content and to improve the user experience.

This JISC Content Programme funded project is an important contribution to initiatives taking place internationally, which are creating digital archives of cultural and scientific significance to enable greater public access and preserve cultural and scientific heritage for future generations. It also supports the recommendation adopted by the European Commission on 28 October 2011 asking EU Member States to involve the private sector in digitalising cultural material; essential to make European cultural heritage more widely available and boost growth in creative industries.

Notes

An earlier version of this paper appeared in *Business Archives*, the journal of the Business Archives Council, in May 2014

Bibliography

Arts Council England (2014). Designation collections list <http://www.artscouncil.org.uk/advice-and-guidance/browse-advice-and-guidance/designation-collections-list> [Viewed: 31 August 2014].

BT (2009). BT Heritage Policy <http://www.bt.com/heritage> [Viewed: 31 August 2014].

CARTER, Katherine (2013). "The M&S Archive: from penny bazaar to research resource", *Business Archives*, 107.

- COPELAND, B. JACK et al (ed.) (2006), *Colossus: the secrets of Bletchley Park's codebreaking computers*, (Oxford University Press).
- Coventry University (2011) "The BT Archive proposal." (unpublished).
- Coventry University (2013). "Word Tree" <http://wordtree.coventry.ac.uk/>. [Viewed: 31 August 2014].
- Coventry University (2014). "Digital Media Projects" <http://www.coventry.ac.uk/research/grand-challenge-initiatives/grand-challenge-initiatives-digital-media/projects/>. [Viewed: 31 August 2014].
- HAY, David (2004). "BT's Connected Earth: a new model for safeguarding a corporate heritage collection", *Business Archives*, 87.
- HAY, David (2011). "The GPO Film Unit and telecommunications culture", in Anthony, Scott and Mansell, John G. (eds.), *The projection of Britain: a history of the GPO Film Unit* (Palgrave Macmillan), p. 206-231.
- Jisc (2011). "Content Programme 2011-13".
<http://www.jisc.ac.uk/whatwedo/programmes/digitisation/content2011_2013.aspx>. [Viewed: 31 August 2014].
- LUFT, Tim (2007). "Stratford Unplugged: a new Wi-Fi experience for tourists in Shakespeare's country", *Innovate: Applied Research at Coventry University*, Issue 1.
<http://www.coventry.ac.uk/Documents/Research%20Documents/Innovate%20magazine/Innovate_Issue1.pdf>. [Viewed: 31 August 2014].
- SCOTT, Peter (2009), *The BT Archive; an academic appraisal*. (unpublished).
- TNA (2008). "The Cabinet Papers, 1915-1984". [Viewed: 31 August 2014].
<https://www.nationalarchives.gov.uk/cabinetpapers/> . [Viewed: 31 August 2014]. This project was an obvious point of reference and source of inspiration for the BT Digital Archives.
- UNESCO (2011). 2011 UK Memory of the World Register
http://www.unesco.org.uk/2011_uk_memory_of_the_world_register [Viewed: 31 August 2014].