

ARC magazine

www.archives.org.uk



Archives & Records
Association
UK & Ireland

No: 349
ISSN: 1745-2120
September 2018

20

**Victorian
bloggers:
from
printing
press to pdf**



29

**British
Library Qatar
Foundation
Partnership
programme**



37

**Ingests and upgrades:
technological barriers
and adapting our
preservation practices**



Archives and Technology

City of Culture digital archive: forging connections with digital depositors. Page 24





Queens Hotel, Leeds 28-30 August 2019



Archives & Records
Association
UK & Ireland

CONFERENCE
2019



Welcome to ARC Magazine August 2018

Welcome to the Section for Archives and Technology special issue of ARC, which as ever contains a range of fascinating articles reflecting the diverse projects and initiatives currently being undertaken.

What particularly strikes me about the contributors to this issue is the range of experience they have working with archives and technology. It's great to see volunteers, trainees, new professionals and experienced hands from across the profession working with, and sharing their experiences about, the opportunities technology is now offering.

Along with a number of case studies of digitation projects, the issue features advice on digital preservation and the new technologies available, as well as articles on how access is changing as a result. It is also clear from the contributions how technology is changing the roles of many in the profession, as well as the way institutions and repositories work and plan for the future. All of this is being supported by the Section for Archives and Technology.

Thank you to Elisabeth Thurlow for efficiently gathering the fantastic range of content for the issue and thank you to all the contributors.

Enjoy the issue!



Ceri Sugg
ARC Editor

ARC Magazine is the monthly publication that is published by the Archives & Records Association (UK and Ireland)
Prioryfield House
20 Canon Street
Taunton
Somerset
TA1 1SW
Tel: 01823 327030

ARC Magazine advertising enquiries to:
sam@cabbell.co.uk
or phone Sam Rogers on 0203 603 7943.

Send articles/comments to:
arceditors@archives.org.uk

ARC Magazine design by Glyder www.glyder.org

Front cover: 'Into the Light', a dance piece commissioned for Hull City of Culture 2017 - one of the many thousands of digital photographs in the collection. Image courtesy of Laura Giles.

DISCLAIMER

The Archives & Records Association (UK and Ireland) cannot accept responsibility for views expressed by individual contributors to ARC Magazine. It is a medium for informing members of news, information and ideas relevant to the profession, including archive conservation. It is not an official guide to procedures, concepts, materials or products.

Association News

- 3 Editorial
- 4 Opening lines
- 5 Professional Development Update
- 6 Professional Development News
- 8 Collecting matters

Features

- 9 The Lisbon Municipal Archive

Archives and Technology

- 12 Welcome to the Section for Archives and Technology issue of ARC
- 13 Building digital preservation skills
- 16 Relaunching the UK Government Web Archive
- 18 Essex archives online: web access to recordings from the Essex Sound and Video Archive
- 20 Victorian bloggers: from printing press to pdf



- 23 Safeguarding digital collections with PRONOM and DROID
- 24 City of Culture digital archive: forging connections with digital depositors



- 27 Sticking with the status quo: abandoned experiments in oral history transcription
- 29 British Library Qatar Foundation Partnership programme
- 31 Digitising Kresen Kernow (Cornwall Centre)
- 32 Writing to Wesley: conversion testimonies from 18th century Methodists brought to life in Manchester
- 34 Atomic evolution: making the visual history of nuclear applications accessible
- 37 Ingests and upgrades: technological barriers and adapting our preservation practices
- 39 Transcription and technology: law in the Aberdeen Council Registers 1398-1511
- 42 Bringing qualitative data to life through online access and citation
- 44 Kinaesthetic learning for technical skills in web archiving

opening lines



The huge importance of social media and websites in the digital age has posed new challenges for the recordkeeping sector. While it has allowed us to reach more people than ever before, the Internet's ever-changing nature has made preservation far more challenging. There are now 1.7 billion websites on the internet, with the average life span of a web page being just 90 days, with great swathes of information lost every day.

With the increasing importance of web & social archiving, for not only historical but also legal and compliance reasons, it is of little surprise that debate over the future scope of digital archiving to capture these new media has been growing in importance over recent years. Archivists have rightfully begun to broaden their horizons from the more traditional requirements, with the most exciting trend being the move towards the preservation of entire websites and social media platforms.

Service providers in the private sector have had to innovate to keep up. A recent important development has been the new partnership of MirrorWeb and Arkivum. Both are companies focused on open-sourced digital preservation and intuitive end-to-end archiving

technology enhancements. They have joined forces to provide a hybrid digital preservation and archival safeguarding solution designed specifically for the heritage, libraries and higher education markets.

Already, MirrorWeb has archived every UK Government Department website, along with their Facebook, Twitter and YouTube accounts. In just 10 hours, MirrorWeb's cloud-based software was also able to index 1.4 billion documents from The National Archives (UK). In effect, this technology allows the preservation of snapshots of the internet from certain dates and moments of time. As organisations continue to develop the Internet as their main form of communication, such archives will be invaluable in keeping records of communication to ensure continued legal compliance, as well as information of historical, commercial and brand value are never lost.

The University of Westminster is piloting this new collaboration between MirrorWeb and Arkivum by archiving a selection of legacy websites that are about to expire. The key innovation here is the ability of the archiving department and researchers to have a permanent copy of the assets in a fully searchable and ISO 27001 WARC format in a system for non-expert

users, without needing to rely on the respective websites. The University also now has the archives automatically preserved in the long-term data lifecycle preservation solution, Arkivum's PERPETUA, without any human intervention.

At the ARA Conference in Glasgow, 29th-31st August, MirrorWeb and Arkivum launched the new combined technology. To demonstrate the effectiveness and simplicity of their services, ARA members are being offered the chance of a free archiving service of this technology, with an annual value of £650 (equivalent in Euros), gaining their own best-in-class digital archive, simply through sharing their website's address.

This new technology offers not only an exciting new addition to meeting the archival challenges of the digital age, but also an insight into the sector's direction of travel. As well as benefiting from this offer, ARA members will also be helping contribute towards the ongoing development and R&D roadmap of digital archiving by providing valuable insight and user experiences, thereby helping shape future services and features.

The offer is open to 200 ARA members on a first come, first served basis. To register your interest and/or request a demo, email phil.ogden@mirrorweb.com or paula.keogh@arkivum.com

David Clee

CEO, MirrorWeb

2018/19 ARA MEMBERSHIP RENEWALS

As we are nearing the end of the 2017/18 ARA membership year it is time to think about renewing your membership. Your current membership expires on 30th September 2018 with the new membership year starting 1st October 2018.

To renew you can go online at <http://www.archives.org.uk/membership-renewal> or pay directly with the ARA Office on 01823 327030. If you have any queries please contact membership@archives.org.uk.

We are pleased to confirm that once again, there will be no increase to the membership fees this

www.archives.org.uk

Professional Development Update:

Much has happened since the launch of the ARA's CPD programme in August 2017. This article gives a brief overview of developments.

Competency Framework

The ARA's competency framework sets out the key competencies required by anyone working in the archives and records sector. It is a tool to help those working in the sector understand how to develop their own career, and for managers when planning the training and development of their staff.

The framework summary document sets out all 39 competencies. The three supporting competency documents include a brief description the five levels of experience that sit below each competency, from novice to expert. These descriptions help you understand what experience you need to develop your career. They also help you understand what experience you need to demonstrate to qualify as a Foundation or Registered Member or as a Fellow of the ARA. The documents are available from the CPD area of the ARA website.

Foundation/Registered Membership and Fellow of the ARA

Foundation membership and Fellowships are two new levels of membership that sit either side of Registered membership, so that the ARA can offer professional recognition to all those working in archives and records management, at different stages of their careers. The process to qualify for any of the three levels of membership is known as the professional development programme. Members seeking professional recognition by the ARA enrol onto the programme and develop the evidence they need to demonstrate how their level of experience across a number of competencies meet the required criteria.

Further information is available from the programme website, which is supported by Survey Monkey. Apply here: <https://archivesandrecords.smapply.io/>. Members who enrolled onto the Registration Scheme, but who did not submit their portfolio before the scheme closed, should re-enrol on the new programme.

CPD review

The need for professionals to undertake CPD, and keep their skills and knowledge up to date, is well understood. Previous ARC articles and web content have described the need for existing Registered members to undertake a process of revalidation every five years. Following a pilot of the proposed revalidation process, and to make it clear that this process is light-touch and not formal 're-qualification', the decision was taken to describe this process as CPD Review, and focus more clearly on the continuation of professional development.

Members who have held Registered status for over five years will be contacted over the coming months and invited to submit evidence of their continued professional development activity. Their submissions will undergo a peer review and feedback will be provided. We will be contacting Registered Members in small groups to make the process manageable. All eventual CPD Review submissions will need to be made via the programme website.

Further details on all aspects of the programme are available from the website.

New enrolments

The following members have enrolled onto the Registered membership programme: Emily Weeks, Richard Burman, James Neill, Malcolm Mathieson, Ellie Jones, Helen Swainger, Jo-Ann Vietzke, Sandra Blake, Jennifer Hunt, Chris Oliver, Dawn Sinclair, Fiona Johnston, Garth Stewart, Zoe Fullard, Chris Cassells, Lara Nelson and Vicki Caren. Best wishes to all enrolled members with their progress towards ARA professional recognition.

Chris Sheridan

ARA CPD Programme Manager

Professional Development News – Foundation Qualification (FMARA)

Foundation (FMARA), Registered (RMARA) and Fellow (FARA) are professional levels of membership that offer designatory letters, known as professional qualifications, equal to those offered by many comparable professions. They provide public and professional recognition that you have met industry recognised standards of knowledge and competency.

This month we meet three ARA members currently enrolled onto the Foundation programme.



Teresa Davies, Archive Assistant with Flintshire Record Office.

"After working in the field for over seven years, six at the National Library of Wales and 18 months at Flintshire Record Office, the opportunity of gaining professional

recognition by ARA for my work appealed to me. Having the support of a mentor, learning from them and having discussions about my professional development was something that I thought would be invaluable to me in my future career”.

“I have followed the competency framework for every part of my professional development plan. The self-assessment process was really useful in keeping me focused on what I want from the process, and which areas of my work I wish to develop. It has made me appreciate and value what I have achieved in my career so far.”

“Having the time to reflect has made me realise just how much I have done and gives me the confidence to seek future experiences and opportunities. Since starting the Foundation Membership qualification I have been more proactive in gaining extra experiences at work, going on courses and the Archives and Records Council Wales (ARCW) Forum to meet with other people who work within the sector. This has enriched my knowledge and enthusiasm for my work. I have truly appreciated the input and time my mentor, Dr Melinda Haunton RMARA, has contributed. Having that support and constructive feedback along the way is invaluable and inspiring.”



Shona MacLellan, a Heritage Apprentice for Comhairle nan Eilean Siar, the local authority in the Western Isles.

“I had completed 1 year and 4 months of a Skills for the Future Traineeship with Tasglann nan Eilean (Hebridean Archives) and wanted the skills and knowledge I have gained to be recognised by the sector.”

“Although I have no formal archival qualification, all that

I have learnt through the past two years of training can be recognised by Foundation Membership and used to gain employment within the sector. I hope to continue my studies, and my competency portfolio will demonstrate to potential universities and employers my commitment and experience.”

Shona was the first ARA member to enrol onto the Foundation programme in January 2018.



Nicola Lloyd, Information Governance Officer for Alternative Futures Group.

“As a recent graduate I wanted to continue my professional development in a structured way and Foundation membership seemed like a good way to do this within a specialised framework supported by a mentor. This was a big plus for my employer as it demonstrates they employ someone with a suitable level of knowledge, and connections to a wider network of professionals”.

“I am the sole information professional, and therefore take on responsibility for a wide range of things. I develop training and provide advice on all aspects of information governance across the organisation, and now act as the company Data Protection Officer after running a project to integrate GDPR. I am also responsible for managing the small company archive and the much larger semi-current records store, including dealing with enquiries, cataloguing new material and implementing retention policies. The role is necessarily very varied, and that is what I love about it!”

“If you want to do well then self-assessment will find the areas where you can improve, and help assess how well

you deal with challenging circumstances. It helped identify my strengths and what I really enjoy, helping me define my career goals more clearly. I have realised through the projects I have been involved in so far, and through my past work history, that I am more naturally suited to areas involving in outreach and engagement. This is where my current strengths lie, which helps me to apply to the right jobs that include these elements”.

We wish all our candidates the very best as they progress towards their goal of Foundation Membership. For further information on the ARA's professional development programme please visit the programme website or contact the programme manager chris.sheridan@archives.org.uk

Foundation Membership of the ARA (FMARA) provides a professional recognition for apprentices, volunteers, new professionals and existing professionals who do not yet have the level of experience required for Registered membership.

To qualify, members must submit a portfolio containing six competency submission forms that demonstrate how their experience meets the required competency levels. Candidates are advised to have a minimum of one year's work experience before submitting their portfolio for assessment.

ENVIRONMENTAL MONITOR and LOGGER



Humidity
Light
Ultra-violet
Temperature
Dewpoint

The new ELSEC type 765 enables the measurement of all the conditions that damage valuable objects.

Optional data logging allows over 70,000 readings to be automatically taken at regular intervals, stored on an internal flash disk and transferred to computer by standard USB link.

Humidity as %RH
Temperature in °C or °F
Light level in Lux or Foot-candles
UV as mw/M² or µW/Lumen
Tiny, solid metal construction
Large clear display

Littlemore Scientific Engineering
www.elsec.com elsec@elsec.com

Collecting matters

In an ever-increasing digital world, archival science is advancing in new and intriguing ways. However, many archives are still grappling with how best to provide access to analogue collections and widen access to new audiences.

The National Archives (UK) has been working on a new tool to help archives do just that. Launched in 2017, Manage Your Collections allows archives to create and manage catalogue content in Discovery.

It offers archives without access to a collections management system, a viable way of getting their collections online. Moreover, all archives can take advantage of Discovery's reach, with 8.6 million user sessions last year, connecting their collections to new audiences.

Access to Manage Your Collections is via a Discovery account and archives who contributed collection data to the Access to Archives (A2A) project can access this via the new tool, managing these catalogues as well as collections added.

We have also recently published guidance on our website on collections management and digital asset management systems.

We are beginning to see more open-source CMS and DAMS solutions on the market. More choice provides more opportunity and flexibility for archives to choose the solution that most fits their needs. Conversely, it can be a daunting task beginning the procurement process, so our market survey provides product overviews and key links to documentation and information, to increase visibility to the wide range of systems on the market for management and access to collections.

For more information, see <http://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/documenting-collections/cataloguing-and-archives-networks/>

Caroline Catchpole

The National Archives (UK)

asd@nationalarchives.gov.uk



The Lisbon Municipal Archive. Image courtesy of Paulo Batista

The Lisbon Municipal Archive

Paulo Batista provides an overview of the diverse collection that makes up the Municipal Archive Division in Lisbon.

The Lisbon Municipal Archive (AML) has five facilities physically dispersed throughout the city, which together make up one single unit, the Municipal Archive Division. This includes:

- The Lisbon Municipal Archive (including the historical archive)
- Arco do Cego Archive
- The photographic archive
- The Lisbon video library
- Alto da Eira deposit

As a whole, the AML is responsible for preserving and disseminating a significant body of material that dates from the 13th century to the present, and which is indispensable for studying the history of Lisbon and Portugal. Particular highlights of the collection include King D. Afonso II's confirmation of the charter granted to the city of Lisboa in 1179; Decrees and Notices; the Livro de Posturas (Book of Postures); the Measurements Books; the Cartulário Pombalino (Pombaline Cartulary); the Livro Carmesim (Carmesim Book); and the Livro dos Pregos (Book of Nails).



Alto da Eira deposit.
Image courtesy of
Paulo Batista





The Lisbon video library. Image courtesy of Paulo Batista

The historical archive

Its oldest recorded information was obtained through a mixture of donation and purchase and is organised according to the functions of the creator or within a deposit framework. This includes, among others, the administration, the taxes, the Municipal Instruction Service, the Charity Service, the Parish Councils and Borough administration.

The archive also includes political and trade union posters, such as that of José Neves Águas. Águas was a journalist and author of a significant set of political propaganda, which is essential for understanding the post-revolutionary period from 25 April 1974, after the dictatorial regime was overthrown. A further strength of the collection includes the papers of some of the most important Portuguese architects of the 20th century, such as Cassiano Branco, José Luís Monteiro, Ruy Jervis d' Athouguia and Francisco Keil do Amaral. This is a growing aspect of the collection, with material from architects Alberto de Souza Oliveira, Joaquim Bento d' Almeida and Victor Palla recently incorporated into the historical archive.

At the AML, one of the most significant information units is the individual buildings processes. This is a group organised according to each individual building, gathered during the course of each building's existence. It constitutes the largest amount of recorded information produced by the department of any municipality and the one that results in the largest number of reproduction requests by researchers.

The Arco do Cego Archive

The Arco do Cego Archive consists of recorded information of an administrative nature produced by the municipal services from 1820. This includes papers relating to urbanistic studies, social facilities construction projects, surveys, fees and licenses and the recorded information produced by the Lisbon municipality's sporting, financial, urban and administrative services.

The photographic archive

The photographic archive contains unique resources for the history of Lisbon. It constitutes an extraordinary testimony of photographic evolution in Portugal, particularly with respect to urban and daily aspects of life.

A particular highlight of the collection is the Fundo Antigo (Old Fund), by José Candido d' Assumpção e Souza e



The photographic archive. Image courtesy of Paulo Batista

Arthur Júlio Machado, which is the first architectural survey (1898-1912) for the area and includes photographs of buildings in certain areas of Lisbon.

The Lisbon video library

The Lisbon video library has an archive of videographic works of national production and a collection of films of significance to the history of world cinema, containing some 5,000 films with nearly 20,000 hours of images about Lisbon.

Alto da Eira deposit

Finally, the Alto da Eira deposit, which has been closed to the public, contains recorded information relating to the municipal bodies, their financial, urban, human and administrative services, tourism, culture, elections, military functions and the control of economic activities.

Including all of its facilities, the AML it is responsible for a wide range of material, typologies and formats, making it one of the most important archives in Portugal. This archive is the largest municipally archive in the country, and the only nationally run archives that is bigger than AML is the Torre do Tombo/ National Archive and the General Archive of the Army. Thanks to the Lisbon municipal leadership and acquisition policy, the AML is a fantastic collection that continues to grow.

Paulo Batista

Lisbon Municipal Archive and Interdisciplinary Centre for History, Culture and Societies - University of Évora (CIDEHUS.UÉ)

Welcome to the Section for Archives and Technology

The Archives and Records Association can be seen to have officially taken notice of the computer 45 years ago, when a resolution from the South-East Region ‘to set up a committee to consider the question of the application of the computer to archive indexing and cataloguing’ was approved at a Society of Archivists’ Council meeting on 18 January 1973. The resulting Computer Applications Committee was active until the mid-1980s, when it passed on the baton to a Computer (later Information Technology) Group. By 1988 this group felt able to report that ‘archivists are much more confident [...] about their relationship with technology, and we can now go along with a much greater understanding, and perhaps a little useful suggestion, and talk to others in the information technology field’.

Much has happened in the 30 years since this pronouncement was made. Yet, on the one hand little has changed. Reading the articles in this special issue it is clear that the profession is still confident about their relationship with technology. Confident and capable enough to use it to further its own aims (of widening access and participation), but also confident enough to decline to use it when it goes against its principles (of responsible custodianship) - see the report of Canterbury Cathedral’s consideration of YouTube’s close captioning functionality. On the other hand however, everything has changed. Technology has reshaped our entire world and is no longer just a tool to be applied or not, but an all-pervasive

part of our lives. This transformation has left everyone feeling a little less confident about their relationship with technology - witness all the hyperbolic coverage of artificial intelligence in the last year.

Those involved in the profession have a particularly hard role to play in this new landscape. For, not only must they maintain their own confidence in their more individual relationships with technology (perhaps by learning to work at the command line, or the rudiments of coding), but they must also maintain (as they always have) the collective confidence of society itself. Whatever the technology underpinning or associated with information - be that parchment or Python - it is our job to make sure that everyone can have confidence and trust in their relationship with it, relying on it as a basis for both understanding and action. The Section for Archives and Technology seeks to support ARA members as they face this double challenge, and if you would like to learn more, or get more involved, please email us at sat@archives.org.uk

Jenny Bunn

Chair, ARA Section for Archives and Technology

DPC Executive Director,
William Kilbride, with
Leadership Programme
grant recipients at iPRES
2017. © Digital Preservation
Coalition



Building digital preservation skills

Sharon McMeekin discusses the work of the Digital Preservation Coalition in recognising the need to build a well-skilled, robust workforce of digital preservation-ready information professionals.

Digital preservation is an ever-increasing concern and identifying the skills required for this work, and how to gain them, can be one of the greatest barriers to making progress. This is why workforce development is one of our main strategic objectives, and our key outputs include training courses, learning resources and scholarships for members.

To facilitate successful work in this area it is important to monitor the ongoing training needs of our members. Between April and June 2018, we carried out a survey of member training needs and this article will examine the main themes that emerged. It will also identify some

existing resources that are helpful, and the work the Digital Preservation Coalition will do to meet training needs.

One of the most important resources relating to skills for digital preservation is a competency framework developed by a project called DigCurV (<https://www.digcurv.gla.ac.uk/>). If you are looking to identify the skills needed for digital preservation this is a good place to start. DigCurV defines skills at three levels: Practitioner, Manager and Executive. We asked survey respondents to identify themselves according to these levels, with the addition of a Trainee option.



DPC members engage in discussion about Digital Preservation Training and Skills. © Alasdair Watson Photography

The biggest group of respondents, 57%, identified themselves as Practitioners, and a further 36% at Manager level. As expected, this points strongly to a need to build practical skills over theoretical knowledge, as an overwhelming majority of respondents are in roles where they need to make digital preservation happen. As a result we will be trying to incorporate more practical elements into our training courses.

We also delved into quite a lot of detail about the digital preservation training topics that interested the respondents. 52% said they had moved beyond the need for introductory training, but 48% noted that this was still a priority. This means the Digital Preservation Coalition will continue to offer our in-person training course

‘Getting Started...’ and ‘Making Progress with Digital Preservation’, which run on a yearly basis in spring and autumn respectively. We try to rotate these around the UK and Ireland and are always open to location suggestions if you can provide us with a room for a day! As the AARA is a member of the Coalition, there are three free places available at each of these events on a ‘first come, first served’ basis for ARA members. If you cannot wait for a course, we also offer introductory resources through the Digital Preservation Handbook (<https://dpconline.org/handbook>). The handbook offers a broad introduction to digital preservation and includes a ‘Getting Started’ section with short video presentations based on the first of our courses.



Other topics that were highlighted as training priorities included: using digital preservation tools, understanding file formats, metadata for digital preservation, advocacy, and how to preserve particular object types such as audio-visual material. The Digital Preservation Coalition has always offered topic specific events through our Briefing Days, but in response to the survey results we will be expanding and enriching these to cover the topics shown to be a priority. Events on advocacy and audio-visual preservation have already been added to our schedule for the coming year. There is also one free space available at each of these events for an ARA member, with access to these spaces coordinated via the ARA Section for Archives and Technology.

“Training priorities include using digital preservation tools, understanding file formats, metadata for digital preservation, advocacy, and how to preserve particular object types”

More information on all of our events can be found on our website. We also offer a range of ‘Technology Watch’ reports on our website covering preservation of different content types, including social media data, emails, e-books, and personal archives. For those with specific training needs, we also offer member scholarships for training and development opportunities through our Leadership Programme. The Section for Archives and Technology can advise on how ARA members can apply for these grants.

Finally, we asked respondents about how they liked training to be delivered, including options covering in-person training, online and a combination of both. In-person training was clearly the preferred choice, but respondents mentioned barriers to attendance that included a lack of funding, no support from managers, and finding it difficult to get time away. With this in mind, the Digital Preservation Coalition will continue to offer our in-person events, but will also be investigating options for delivering content online. We want to make sure any platform we choose supports an interactive experience that allows of delivery of practical content and high levels of interaction between students and tutors.

Digital preservation remains a developing discipline and the Digital Preservation Coalition aims to meet the changing training needs of the information management community. For more information on the resources we offer and how to join us, please visit our website: <https://dpconline.org/>.

Sharon McMeekin

Digital Preservation Coalition

Relaunching the UK Government Web Archive

The UK Government Web Archive (UKGWA) holds archived versions of UK Central Government funded websites dating from 1996 to present. We continue to capture websites regularly throughout their lifetimes, currently at a rate of around 100 sites each month. The UKGWA is, by some distance, the largest discrete digital collection held by The National Archives (UK) and is free to access online by anyone, anywhere in the world.

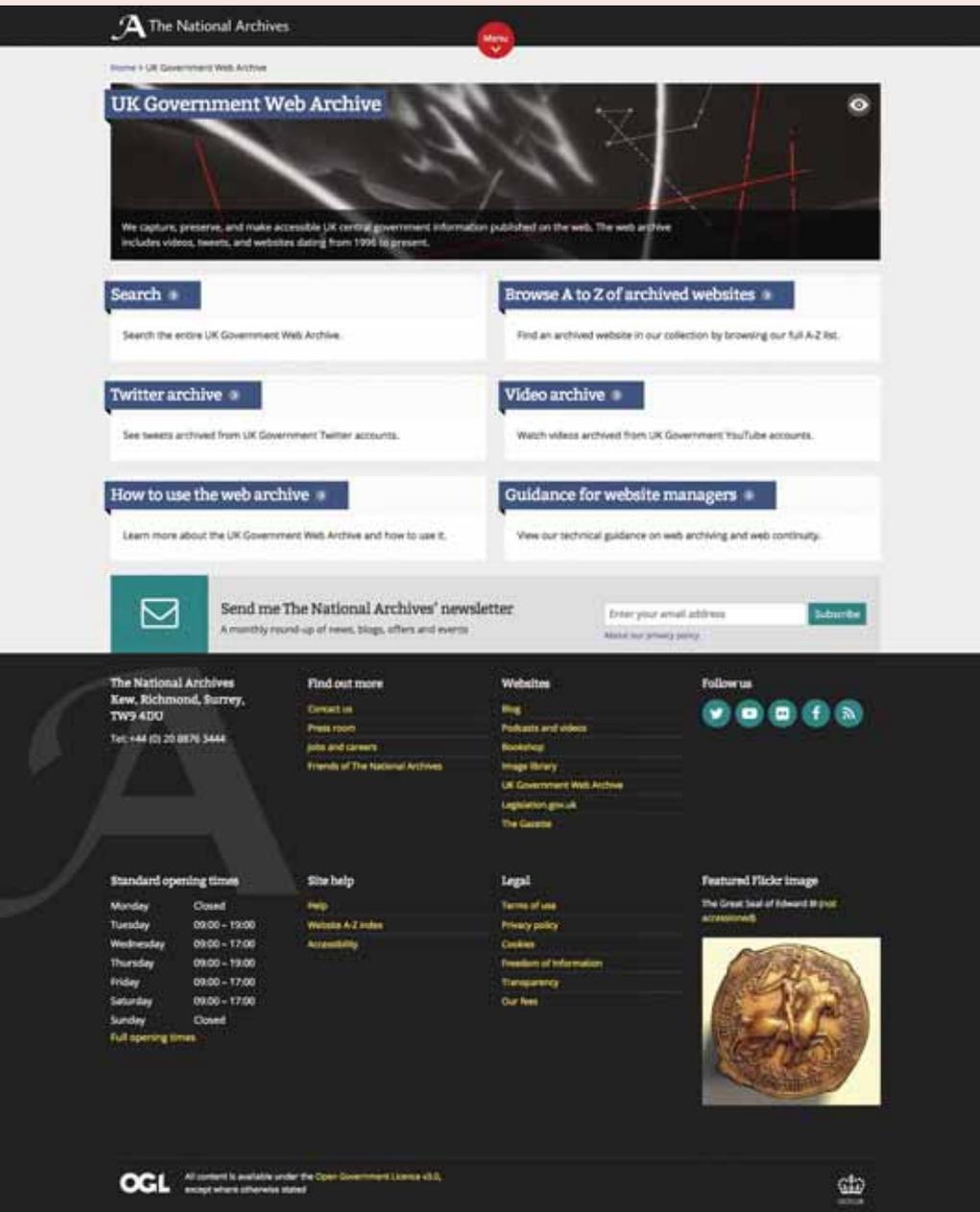
In late 2016, following a competitive tender exercise, we appointed new suppliers, MirrorWeb, to deliver the technical aspects of the UKGWA. Following a lot of work behind the scenes, we were very pleased to re-launch the web archive (<http://nationalarchives.gov.uk/webarchive>) on schedule in July 2017. Here I will reflect on the transition one year on and discuss the progress we have made so far.

The Cloud

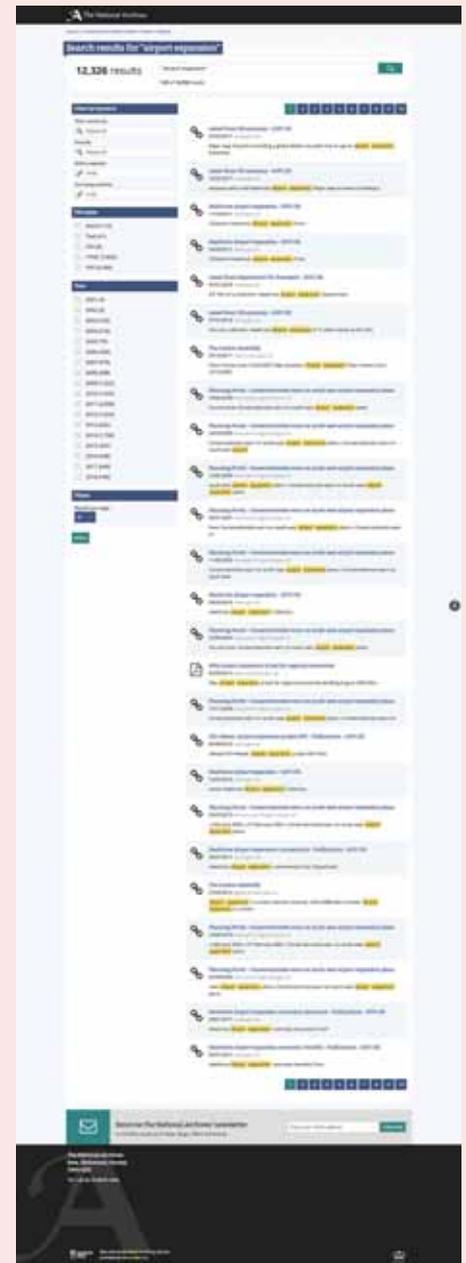
Our digital strategy states our ambition to ‘move our public facing digital services to the cloud’. Migrating over 100TB data was a challenge, but we were very happy to be one of the first services to make the move, as the benefits are huge. The web archive is now faster and more reliable. MirrorWeb make use of their expertise to use the flexibility of Cloud computing to run processes quickly and efficiently. It will also make it easier for us to move web archive data in future, including into our internal digital preservation system, and to make it available to researchers.

Search

Providing a much improved full text search service was a key requirement of our contract with MirrorWeb. The old service was unreliable, used out of date technology and did not provide users with enough functionality. The new service is a big improvement. It runs on up to date Elasticsearch technology and provides users with the ability to filter results by date of capture, type of document, keyword and original website. Above all, it is much more reliable. We recently



The UK Government Web Archive home page. © Crown Copyright



A search results page showing the filters on the left hand side. © Crown Copyright

“We are now able to capture some interactive content such as quizzes and animations that would previously have been impossible.”

updated the search index to cover websites captured since the re-launch and hope eventually to update it on a monthly basis. In the coming months we will make improvements to the search indexing process to ensure all pdf documents are processed by our optical character recognition (OCR) technology. This will make a large number of additional documents searchable.

New technology

We are now capturing websites and replaying the archived versions using up to date versions of the

Heritrix web crawler and the Python Wayback (PYWB) playback tool. The archiving process is more efficient as website captures are now higher quality and need fewer retrospective fixes.

In a very exciting development, MirrorWeb are also able to use the cutting edge Webrecorder tool to capture content which Heritrix cannot process. We are now able to capture some interactive content such as quizzes and animations that would previously have been impossible.

“Users now have the ability to filter results by date of capture, type of document, keyword and original website”

Social media

We now capture the Twitter and YouTube channels of ministerial and non-ministerial departments every day via an application programming interface (API) capture process. Previously, the channels were captured once every six months or annually. As social media platforms are fast moving, the increased frequency of capture minimises the risk of content being missed. We are currently working with MirrorWeb to run quality assurance checks and publish social media content captured since the beginning of the re-launch. In future we aspire to publish newly captured social media content each month.

The future

In line with The National Archives (UK) and wider UK Government commitments, we are very clear that the UKGWA service must meet the needs of its users. The requirements for our contract with MirrorWeb were driven by a piece of user research undertaken on our behalf in 2015. We worked with a user researcher to undertake further research on the re-launched service earlier this year. It revealed several areas for us to work on. Within the next few months, we will use the outcomes to make improvements to the links in the red banner which appears on all the archived pages we hold, the full text search landing page, search results pages and our error page.

We are also working with colleagues across the organisation to increase awareness of the UKGWA as a research resource, and to find out what we can do to help researchers access the collection.

The UKGWA is a huge collection providing a rich source of information about UK Central Government use of the web over the past twenty years. Managing and providing access to such a large amount of data is a huge challenge, as is capturing web content to a high standard in a rapidly changing technical landscape. However, we are happy with the improvements made over the past year and are excited by what the future holds.

Claire Newing

The National Archives (UK)

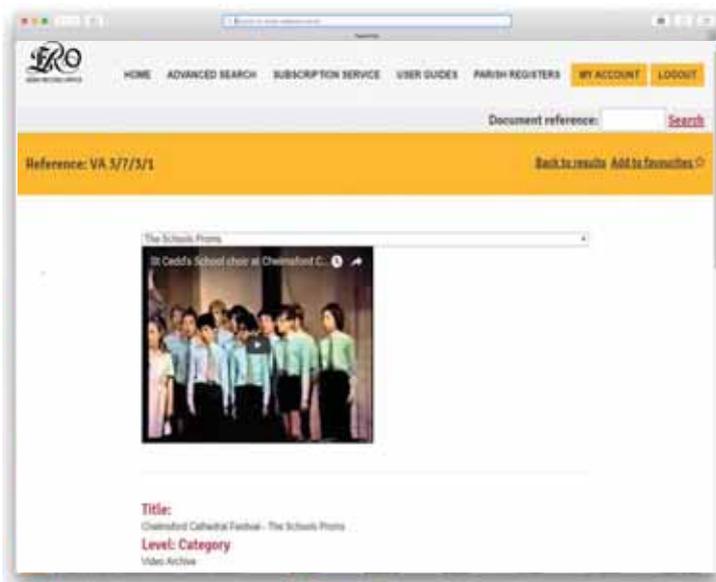
Essex archives online: web access to recordings from the Essex Sound and Video Archive

Sarah-Joy Maddeaux reports on the preservation, user and engagement benefits of digitisation projects

Over the last three years, the Essex Sound and Video Archive at the Essex Record Office (ERO) has been gradually making some of its recordings freely available online, a leap forward that allows us to provide global access to the collections at any time of day or night.

This leap has been made possible through funding received from the Heritage Lottery Fund (HLF) to run our project, You Are Hear: sound and a sense of place. While I have previously written ARC articles about the engagement aspects from the project (audio maps, listening benches and more), behind these attention-grabbing outputs a crucial part of the project has been the digitisation of sound and video recordings for preservation. Once digitised, we can put recordings online - thus improving both preservation and access in one harmonious strategy.

Experiments with uploading recordings predated the You Are Hear project. The ERO's sound archivist occasionally



Video recording embedded in online catalogue entry. Copyright Essex County Council



Digitising a reel to reel tape. Copyright Essex County Council

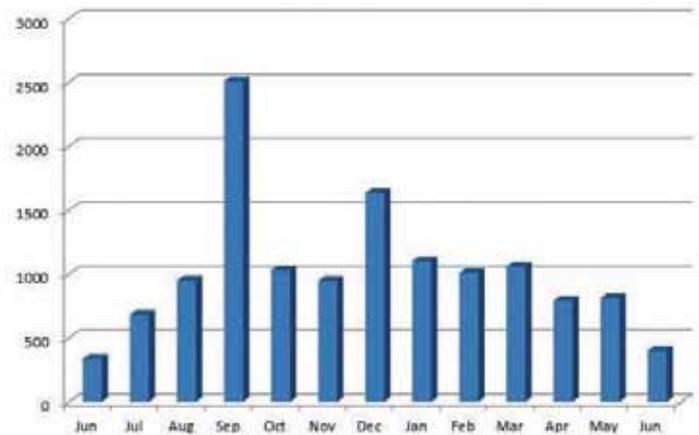
added recordings to our SoundCloud channel, which we could then link to the catalogue entry. But the You Are Hear project allowed us to take that significant step of making the actual recording playable directly from the catalogue entry. This allows users to simply click and listen to gauge if the recording is of interest.

To achieve this, we continue to upload sound recordings to SoundCloud (or video recordings to our YouTube channel). We then embed the recordings in our catalogue. While users have to sign in to play the recordings, they do not have to pay any money. Since any recordings digitised through the HLF grant have to be made available under a Creative Commons (Attribution-Non Commercial) licence, users can also download the recordings and reuse them (though they have to click through to SoundCloud to do this).

The ERO uses a proprietary catalogue developed by Essex County Council's IT team. Adding the function to play sound and video recordings was therefore relatively straightforward. We told them what we wanted and they came up with the solution. The HLF grant funded the software development team's time to implement the solution.

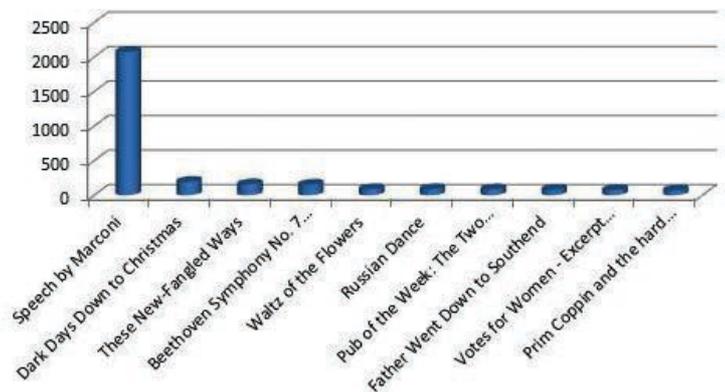
The benefits to providing online access are obvious. Researchers can play recordings at any time of day, from anywhere in the world, no longer needing to visit the ERO in Chelmsford in person - no ID, CARN registration, or ordering required. Adding our recordings to global file-sharing services also means people will stumble across them more frequently than if they were solely available through our catalogue - people who would never think to look at Essex Archives Online, people who may not know we exist. A snapshot of our SoundCloud playback statistics

Top Number of plays on Soundcloud June 2017 - June 2018



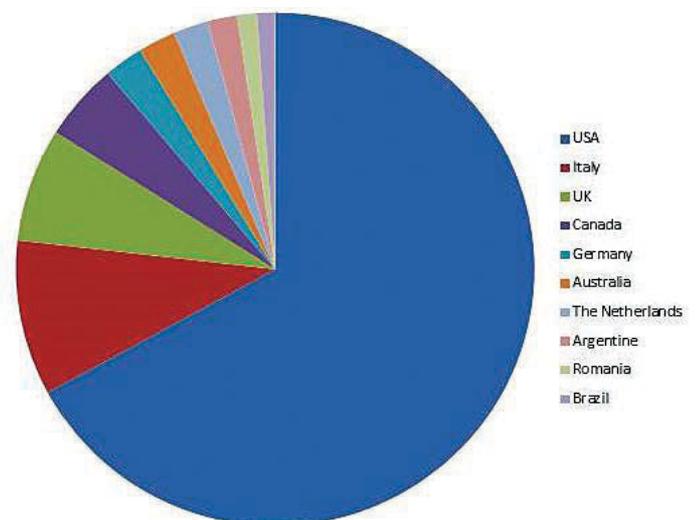
Statistics showing the total number of plays of recordings on the ERO's SoundCloud channel, June 2017 to June 2018. Copyright Essex County Council

Top 10 Tracks Played, June 2017 - June 2018



Top 10 tracks played on the ERO SoundCloud channel, June 2017 – June 2018. Some of these were featured on our 2017 advent calendar. Copyright Essex County Council

Source of Plays by Country (Top Ten) June 2017 - June 2018



Top countries playing recordings on ERO's SoundCloud channel, June 2017 to June 2018. The high number from Italy is due to the recording of Marconi. Copyright Essex County Council

“What might our recordings inspire?”

demonstrates the vast number of plays in the last year and the far-flung locations of the listeners.

As the material is available under a Creative Commons licence, others can share the recordings to reach audiences unknown to us. One of our most frequently played recordings is a speech by Guglielmo Marconi, largely because the recording keeps being shared on radio enthusiast websites. As well as increasing access, this offers exciting opportunities for artistic collaboration: what might our recordings inspire?

Finally, having a bank of digital recordings to hand is helpful to our colleagues running education and engagement activities. We like to encourage use of the sound archive in-house as well as externally, so making it easy for them is mutually beneficial.

Of course, this also means we are not in control over the life of the recording outside the archive. Once downloaded, we are reliant on people's honesty to adhere to the Creative Commons licence. This is a step of faith which not all oral history interviewees or content creators are willing to take. That is understandable and acceptable. However, overall, we believe the advantages outweigh the risks.

Now that the ERO has this functionality, we will keep adding to our online collection of sound and video recordings. We will also continue to look out for ways in which we can make recordings more accessible, particularly for people with disabilities or impairments. We would be happy to exchange ideas, suggestions, and challenges with colleagues, so get in touch if you would like more information - or contact me for notes from my presentation at the ARA Conference exploring the deeper impact of sharing recordings.

ERO's SoundCloud channel: <http://soundcloud.com/essex-record-office/>

YouTube channel: <http://youtube.com/user/essexrecordoffice>

Online catalogue: www.essexarchivesonline.co.uk/

Sarah-Joy Maddeaux
Essex Record Office

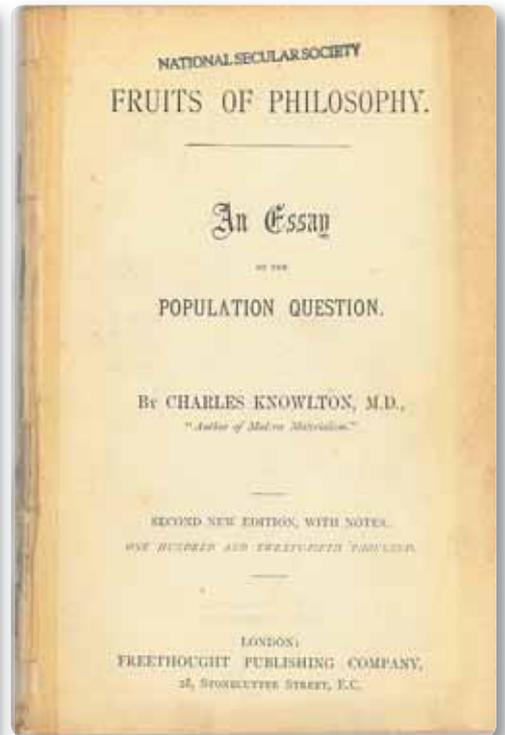
Victorian bloggers: from printing press to pdf

Conway Hall Library and Archives is the only specialist humanist library in the UK and has been a haven for radicals, freethinkers and social and political reformers since the library was established in 1886. We are currently undertaking a Heritage Lottery funded project, 'Victorian bloggers', which will see our collection of over 1300 19th century pamphlets digitised and made freely available online. The collection covers issues such as freedom of the press, secularism, gender equality and political suffrage - many of which are still relevant today. The project draws parallels between 19th century pamphlet printing and 21st century blogging, as two communications technologies used to boost awareness of social and political issues.

The digitised pamphlets will be a valuable resource, shedding light on some under-researched topics and demonstrating the contributions of humanists towards key social and political reforms. The collection includes annotated copies of the instructive birth control pamphlet, *Fruits of Philosophy 1832*, which were used as evidence by the pamphlet's publishers Charles Bradlaugh and Annie Besant in an 1877 obscenity trial against them. This trial marked a turning point in the dissemination of practical information about contraception. Our copy of Richard Carlile's *Jail Journal* compiles essays written by Carlile during almost ten



The prison writing desk of Richard Carlile. Photograph by Andrew Shaylor



Early birth control pamphlet, *Fruits of Philosophy* by Charles Knowlton. Image courtesy of Conway Hall

The arts and crafts style fireplace in Conway Hall Library. Photograph by Andrew Shaylor



“Students will learn about the history of pamphleteering and related activism. They will discover how similar activist writing has translated to new platforms such as blogs”

years he spent imprisoned for blasphemy as a fierce campaigner for freedom of the press. These essays were most likely written at his prison writing desk, which now resides in our library in the company of the pamphlets and articles created upon it. We also hold a number of early socialist pamphlets, discussing topics such as the campaign for an eight-hour working day and the need for improved housing conditions.

Pamphlets first emerged in the 1500s, following the invention of the printing press, as cheap, small-format publications, discussing topical social, political or religious issues. By the Victorian era, with higher literacy levels and improved printing technologies, pamphlets were produced at a rapid rate. Pamphleteers were the bloggers of their day, adopting the widely accessible format to communicate their ideas to a wide audience. Today, blogs are used in a similar way to share short pieces of writing, often about topical social or political issues, to large audiences online. We are furthering the pamphleteers' dedication to making information affordable and widely accessible, making our pamphlet collection freely available online and searchable through optical character recognition (OCR). The OCR means we are not merely broadening the reach of our pamphlet collection, but presenting it in an improved way, with users able to search within the texts. This is especially useful given the Victorian propensity to waffle!

The digitised pamphlets will be accessible through our digital collections platform, Omeka, as multi-page pdf documents. The open-source platform supports online access to digital collections along with their associated Dublin Core metadata. Digital objects can be arranged into collections, which allows for a basic reflection of the collection's arrangement, and users can browse or

“Pamphleteers were the bloggers of their day”

search using either free text or keyword tags. The process for uploading content is fairly straightforward - we have exported descriptive metadata for the original pamphlets from our catalogue as a CSV file and this will be imported to Omeka and mapped to Dublin Core metadata fields. With the help of volunteers, we will then upload the pdf files to their corresponding set of metadata and supplement the metadata as necessary. We tested the platform during our pilot digitisation project, 'Architecture and Place', which made accessible plans, photographs and other documents relating to the buildings our Ethical Society have called home. We have uploaded our working documents from this pilot to Omeka, to share our experience with other small libraries and archives looking to digitise their collections. Whilst we have found Omeka to be quite simplistic in places, it is generally very user-friendly and an affordable option for smaller libraries and archives.

The learning activities associated with this project centre on how evolving communications technologies have supported social and political activism through the years. We are developing a schools project aimed at 16-17 year olds, looking at citizen journalism in collaboration with the Centre for Investigative Journalism. Students will learn about the history of pamphleteering and related activism. They will discover how similar activist writing has translated to new platforms such as blogs and take inspiration from our 'Victorian bloggers' to produce a special issue of Conway Hall's quarterly journal, the Ethical Record, which is available to read on our website. To improve awareness of humanist and freethought heritage, we are hosting a series of Wikipedia edit-a-thons on themes relating to our pamphlets and their authors. We are also taking the collection as inspiration for two creative workshops, focusing on zine-making and creative writing, the outcomes of which will be presented as online exhibitions.

Through the online media focus of these learning activities and the digitisation itself, this project is harnessing contemporary technologies to further disseminate the ideas within our pamphlet collection and make humanist and freethought heritage freely accessible online. We like to think this free exchange of knowledge reflects the concepts of equality and self-education that many of our 'Victorian bloggers' upheld.

Conway Hall Library and Archives digital collections can be found at: <http://conwayhallcollections.omeka.net>

Alicia Chilcott
Conway Hall

Safeguarding digital collections with PRONOM and DROID

The National Archives (UK) made the digital preservation planning tools, DROID and PRONOM publicly available in 2005. Thanks to the contributions of the worldwide digital preservation community, these tools have grown to become the cornerstone of digital preservation and digital continuity for heritage institutions and records management professionals around the world.

DROID is a file format identification tool that scans collections of digital files and identifies the file formats found within. This is important as it helps you to understand what types of files you have, so that you can plan accordingly to maintain continued access to your files. For example you may identify a need to retain a particular piece of software to maintain access to certain files, or you may decide to substitute, or migrate from one type of format to another, to ensure the content is accessible should the original format become more difficult to use over time.

PRONOM is the database that underpins DROID. It provides information on over 1500 file formats, including descriptions, associated extensions, MIME/Media type, links to file format specifications, and perhaps most importantly, the file format ‘signatures’ used to drive the identification of digital files. PRONOM’s data is found within many digital preservation repository systems, including Archivemata, Libnova, Preservica, RODA, and Rosetta, which make use of PRONOM’s identification mechanisms and unique identifier scheme to drive digital preservation actions, such as file format migration and associating particular formats with applicable rendering software to maintain access over time. PRONOM has also found its way into other types of tools, such as digital forensics software and document management systems.

Both tools are completely free to use. DROID can be downloaded from The National Archives (UK) website at <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid>, and PRONOM is free to browse and search at <https://www.nationalarchives.gov.uk/PRONOM>. Each



Digital files created by older software may not be compatible with the software of today.
David Clipsham

“PRONOM is the database that underpins DROID”

tool has its own Google Group discussion forum, where users can post issues and questions and receive support from The National Archives (UK) and the wider user community. These are at <https://groups.google.com/d/forum/droid-list> and <https://groups.google.com/d/forum/pronom> respectively.

As an open source software application, published on the GitHub software code repository, DROID has benefited from community contributions. External development is encouraged and welcomed via GitHub pull requests via <https://github.com/digital-preservation/droid>.

PRONOM especially has benefited from collaboration across the international digital preservation community. To date over 60 institutions have contributed data to

“DROID is a file format identification tool that scans collections of digital files”

PRONOM, including requests to research missing formats, file format descriptions, provision of sample data, through to fully tested file format identification signatures. The National Archives (UK) maintains a list of historic contributions on the PRONOM release notes web page at <https://www.nationalarchives.gov.uk/aboutapps/pronom/release-notes.xml>.

Recent PRONOM releases have included a number of submissions relating to research data file formats, including specialised scientific measurement formats and statistical analysis formats. In part, these submissions have been driven by a push within some university archives, with support from Jisc, to ‘fill in the gaps’ for format types previously underrepresented in PRONOM.

This community-driven approach to file format research is beneficial to all users of PRONOM. With only a small team of digital archivists, The National Archives (UK) is not able to research every file format out there. As the archive for Central UK Government, there are many types of formats simply not encountered within our deposited collections, but that are undoubtedly important to the wider information management sector. For these reasons, we are keen to hear from those working with collections containing uncommon or specialised digital format types. If, while using DROID, you find any formats that cannot yet be identified, please get in touch with PRONOM team at pronom@nationalarchives.gov.uk with a research request. There is also an online form at <https://www.nationalarchives.gov.uk/contact-us/submit-information-for-pronom>.

Digital preservation is a huge challenge for many institutions, so by working collaboratively, sharing knowledge, and contributing to tool development, we can share the burden and collectively reap the benefits to ensure our digital collections are safeguarded for the future.

David Clipsham

The National Archives (UK)

City of Culture digital archive: forging connections with digital depositors

As the official academic partner of the Hull 2017 City of Culture, the University of Hull is currently engaged in a project with CoSector, University of London, to create a digital archive capturing the history of Hull’s tenure as the UK City of Culture.

The bulk of the records have come from the Culture Company that was set up to deliver the programme of events, but we have also taken in records from key contributors and partners. Though we have amassed a small amount of paper records and artefacts, the majority of the archive is made up of hundreds of thousands of digital records. Through this work we have identified a new sense of immediacy offered by digital archives and have directly encountered a contradictory mixture of nervous caution and eagerness that this can generate in depositors. In response, we have had to relate to our depositors in new ways, involving them in the archive process more than we might have in the past. In this article I will briefly explain navigating this new relationship with depositors and how we have attempted to make this as fruitful as possible.

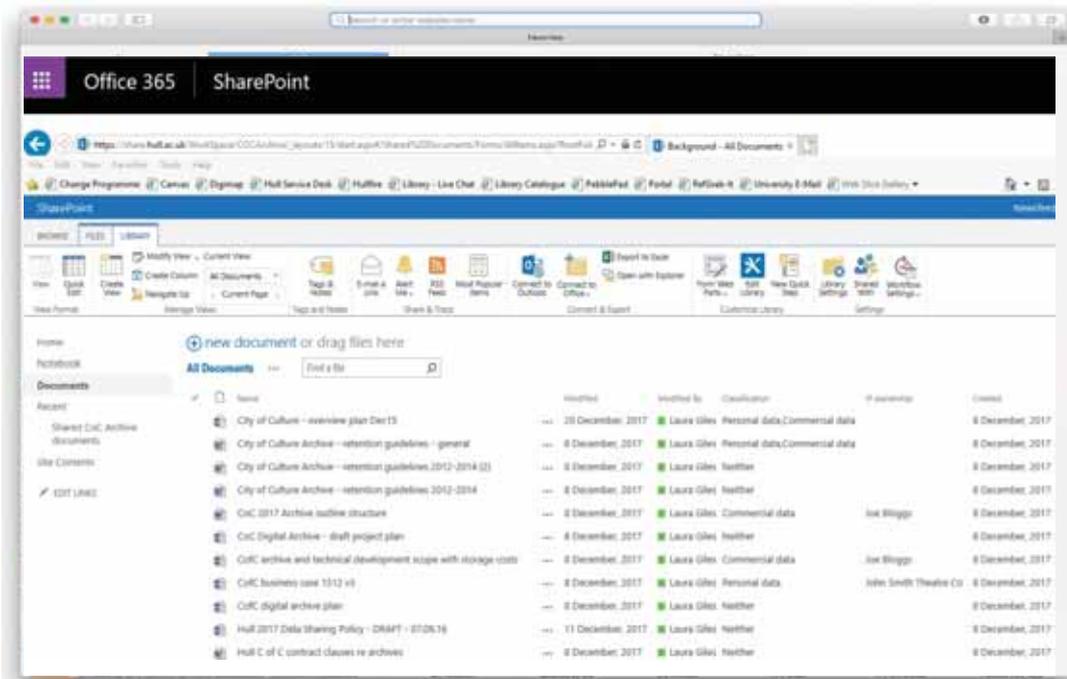
In our experience, the previously mentioned ‘nervous caution’ is caused by a fear that the moment digital records are handed over by depositors, they will be placed online for the world to see. It is increasingly easy to make huge amounts of digital information available online instantly, without even having to have any deep comprehension of what is being uploaded (think WikiLeaks!). Those without an understanding of archival processes can fear that this might be the case with digital

A selection of the eclectic City of Culture collection, including digital records. Image courtesy of Laura Giles



'Into the Light', a dance piece commissioned for Hull City of Culture 2017 - one of the many thousands of digital photographs in the collection. Image courtesy of Laura Giles





“
 For us, the transition to digital records has heralded a new way to relate to depositors
 ”

We used a test SharePoint site to perfect the best way to use columns and tagging. Image courtesy of Laura Giles

archives. The more urgent need to harvest digital records when they are ‘fresh’ (as opposed to paper records which could feasibly sit in storage prior to archival transfer for 20+ years without particular ill-effect) adds to this depositor anxiety.

Whilst we would assume that our depositors will not be deliberately giving us records that contain evidence of wrongdoing, there is still sensitive information - personal and commercial - that they have justifiable instincts to protect. Arguably, paper archival records have traditionally brought with them a perception of being more secure from instances of accidental disclosure for depositors. This is due to their being kept in boxes in secure areas, the requirement for them to be retrieved by a member of staff and their supervised consultation. The reality of course is that if a paper record has been marked as open then these hurdles merely slow down access, not stop it.

Our approach to countering this caution was firstly to emphasise our values as professionals and the continuity between how we approach the security of paper and digital records. This was achieved through discussions at our monthly project meetings with the Culture Company and in our responses to ad hoc queries. We were constantly surprised by how interested our depositors actually were in thinking about digital preservation. Secondly, and more importantly, we worked hard to ensure that the Culture Company were closely involved with flagging up any particular records of concern for us at the creation stage, most notably by working with their corporate operations

manager to enable users to tag records of a personally or commercially sensitive nature - or with external copyright - within SharePoint. This metadata can all be exported and coupled with the records it relates to in order to aid our archival processing.

Any caution we encountered was thankfully outmatched by eagerness - we have found that people working in the creative industries are excited by the idea of contributing to a forward-thinking and evolving field. While they are hyper-aware of issues surrounding ownership, they are also enthused by the thought of other people taking inspiration from what they might find in an archive to create new works. As a further bonus for us, the time-limited nature of the Culture Company (which has already ceased to be) meant that people working there probably had a heightened sense of the ephemeral nature of their work, and understood better than most the need to act straight away to preserve things rather than assuming people that came after would do it.

For us, the transition to digital records has heralded a new way to relate to depositors. Undoubtedly there is a time commitment involved with building that relationship, but it is one that pays dividends when it comes to the richness of the information we now have about our records.

Laura Giles
 University of Hull



Volunteers practice their interview skills. Image copyright of the Chapter of Canterbury Cathedral

Sticking with the status quo: abandoned experiments in oral history transcription

Ashleigh Hawkins provides an insight into Canterbury Cathedral's decision to stick with manual oral history transcriptions.

In 2017, Canterbury Cathedral embarked upon an oral history project, 'The Memories Project', to capture the memories of its intersecting communities. The initial plan was to follow the standardised format of a volunteer-led oral history project with the labour intensive manual transcription, carried out by a team of volunteers based physically in the archives and library reading room. However, an unexpected response was received to the call for volunteer transcribers from colleagues at the University of Kent working as part of Project OPERA.

Project OPERA (Opportunity, Productivity, Engagement, Reducing barriers, Achievement) 'seeks to implement a range of accessibility initiatives to raise awareness of the potential for inclusive design and assistive technologies to improve access to learning for all' (<https://www.kent.ac.uk/studentsupport/accessibility/opera.html>). One aspect of this is using YouTube's close captioning functionality to add subtitles to recorded lectures to support students

with inclusive learning plans. The team at Project OPERA very kindly spoke with us about what they were doing, and demonstrated the potential of using YouTube for transcription. However, they were very clear that they were not recommending it to us as a free transcription tool and, more importantly, stressed that they could not provide any legal guidance on its suitability.

We could identify many potential advantages to using a freely available and widely used tool such as YouTube in a volunteer-led project. Many of the volunteers are familiar with using it and those without accounts would be able to register with the site easily. Being a web-based tool there is no need to install any specific software, enabling volunteers to participate remotely regardless of where they were based around the world. It is relatively simple to both upload and export content and, although in many ways it is quicker to export the subtitles and edit them using a word processor, there is the option of editing within YouTube itself. If



Closed captioning in YouTube

edited within YouTube, the video, with corrected subtitles embedded, can be exported as an mp4 file, thus making it versatile for future use and interpretation.

The advantages of using YouTube were clear, but further investigation was required to understand fully the disadvantages. An initial online search to identify other organisations using YouTube to transcribe oral history was fairly fruitless, with the American based Oral History in the Liberal Arts (OHLA) providing an article on the process of using YouTube, but little discussion anywhere of the potential rights issues with uploading oral history content onto YouTube. Correspondence with Dr Robert B. Perks, lead curator of oral history and director of national life stories at the British Library, confirmed fears that no one else seemed to be using this method of transcription and echoed our concerns around the rights YouTube asserts over videos posted to its site¹. Furthermore, he questioned the extent that we could exercise control over confidential data posted to YouTube even where the privacy settings were set to private, in theory limiting access to those with express permission.

Nevertheless, we were curious to see how accurate a transcription YouTube could produce. The process requires the video file to be posted to YouTube for a matter of minutes only, as it can be deleted and removed once the transcription is complete and the file exported. Furthermore, interviewees sign an agreement, explicitly stating that their interview can be made publicly available. Regardless, we selected an interview between a volunteer interviewer and a member of the cathedral’s staff as an experiment, to avoid the risk of sharing the data of a member of the public with YouTube.

‘...a solitary bed in one of the youth hostels in town and I had to go under one of the underpasses to get to it and I

thought ooh this isn’t quite what I was expecting of Canterbury. But then when I woke up the second day I had a few hours before my interview, and I went for a wander and came to the cathedral and the weather was just beautiful. It was a clear blue sky, and the background of the cathedral, I mean the cathedral against that background was just gorgeous and I remember wandering along the city walls to Dane John Gardens and it was all very beautiful. And then after the interview I...’

For the time being, we have decided to continue with manual transcriptions of oral history interviews for this project. The risks of not having complete control over content posted to YouTube and it being accessed inadvertently by someone without permission were too great to be overcome even by a highly accurate transcription. The rate of accuracy of the transcription we produced was not very high and significant manual time had to be spent tidying it up, adding grammar, correcting place names and errors made due to the accents of the speakers. However, it was still a quicker process than manual transcriptions from scratch. The potential opportunities presented by YouTube’s close captioning functionality for increasing accessibility and further interpretation work are still intriguing, as is the ease with which we could open up the project to remote volunteers, and we would be interested in hearing from others who have also experimented with using YouTube (thememoriesproject@canterbury-cathedral.org).

Ashleigh Hawkins
Canterbury Cathedral Archives and Library

¹ According to YouTube's Terms of Service section 8.1: "When you upload or post Content to YouTube, you grant... to YouTube, a worldwide, non-exclusive, royalty-free, transferable licence (with right to sub-licence) to use, reproduce, distribute, prepare derivative works of, display, and perform that Content in connection with the provision of the Service ..."

British Library Qatar Foundation Partnership programme

In October 2014, the British Library Qatar Foundation Partnership launched the Qatar Digital Library (QDL), an online bilingual portal that provides free access to material from the British Library's collections

The portal displays content relating to the history and culture of the Gulf and its surroundings, as well as the Library's Arabic Scientific Manuscripts. Among the collections that we are working on are: the India Office Records on Gulf History (Agencies and Residencies), personal papers, maps, photographs, and manuscripts. The portal is fully bilingual, supporting study in both Arabic and English. At the moment, there are almost one and a half million images of British Library material on the portal, comprising over 14,000 records and over 136 manuscripts, with more content being uploaded every week. In addition, the QDL hosts articles from our experts, developed by the

British Library team to help contextualise the collections. There are currently over 140 published articles, with more to come.

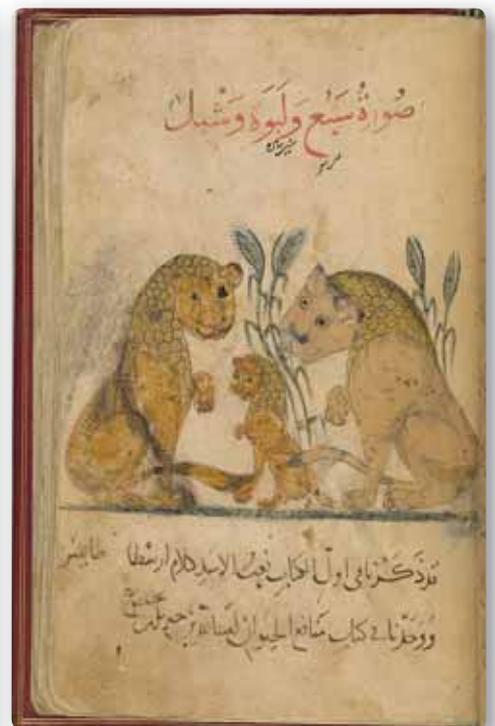
Digitising and publishing the documents on the QDL requires the work of a wide range of specialists. We are an interdisciplinary team, made up of more than forty professionals, including computer scientists, photographers, conservators, curators, archivists, administrators, translators, and specialist historians. Together we are working to give users of the portal a comparable experience to seeing the original documents in person.



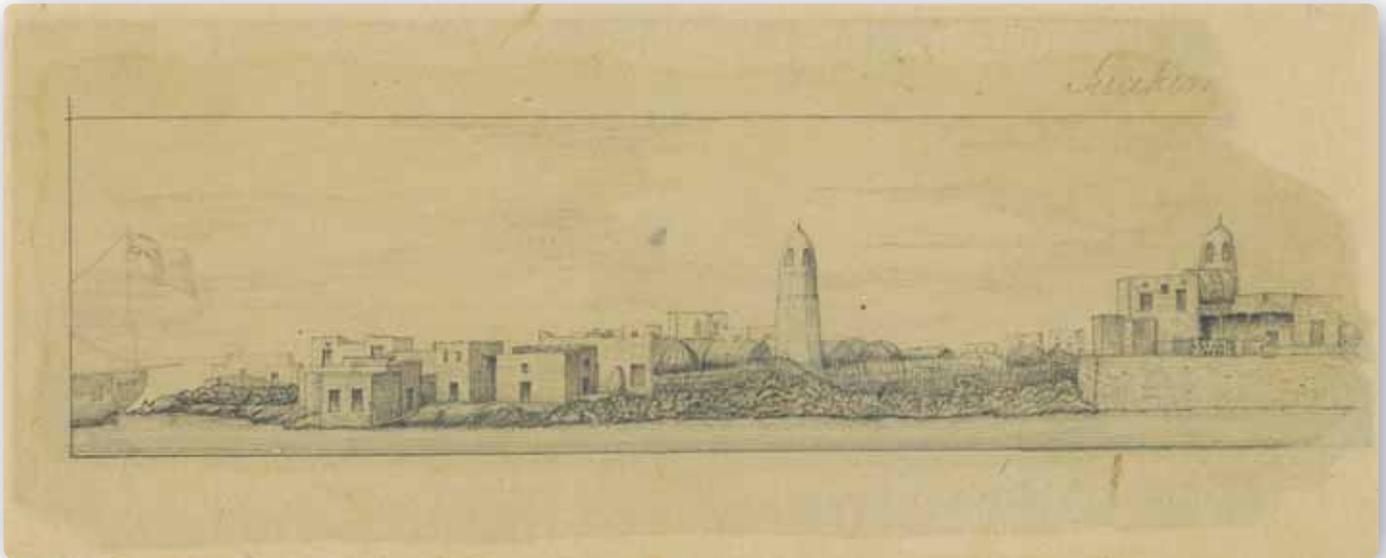
Kitāb fi ma'rifat al-hiyal al-handasiyah. Treatise on practical mechanics by Badī' al-Zamān Abū al-'Izz Isma'īl ibn al-Razzāz, 18th century. Or 116. British Library: Oriental Manuscripts



Kitāb Disqūridīs fi mawādd al-'ilāh. Arabic version of De materia medica, by Pedanius Dioscorides, 735 AH/1334 AD. Or 3366. British Library: Oriental Manuscripts



Kitāb na't al-hayawān, 13th century. Or 2784. British Library: Oriental Manuscripts



View of Suakin and of a peak on the Red Sea coast by Thomas Elwon, 1829-34. WD3529. British Library: India Office Records and Private Papers

The most obvious and important benefit of digitisation is the increased visibility and access to the collections. Users no longer have to be physically present in the Library’s reading rooms in London, but can now view these records from any corner of the globe, on a number of different devices. Since the portal has been active, users have been accessing the site from all around the world, with the top five countries being the United States, Saudi Arabia, Qatar, Oman, and the United Kingdom.

Alongside the digital images, each file is published with a short descriptive catalogue record, created by our team of experts. Cataloguing of this kind allows the library to better understand and document the nature of the collections themselves, improving its own records and highlighting the importance of the material.

When providing free open access to information online, issues surrounding copyright and data protection must be considered. On the programme we have a dedicated rights clearance team, and the programme works with the Library’s information compliance officer to ensure that we are compliant with current legislation and British Library policy. By firstly determining whether the catalogued material is still within copyright or not, our Rights Clearance team then conduct copyright ownership research into the collection items selected for digitisation, tracing and contacting rights holders where possible, such as individuals, companies, publishers, estates and other relevant bodies, working to ensure the correct usage terms are displayed on the portal.



Entrance to Bushire Residency, c.1870. Photo 355/1/34. British Library: India Office Records and Private Papers



Arabia: General Map Showing Communications, Landing Grounds etc. January 1937, IOR/L/PS/12/2160B, f 44. British Library: India Office Records and Private Papers

Moreover, there are further challenges on a digitisation project such as this. There can be challenges in scoping the material: its condition, size, the style of handwriting, and the languages in which it is written may all make a given file difficult to read. These issues can in turn have knock-on effects on the time needed for conservation, cataloguing, and digitisation. Assessing the time needed for an item to make its way from the BL's secure storage onto the portal is no easy task, and requires clear coordination across all teams. To facilitate this, a workflow with three separate streams has been developed, and is now managed through the use of Microsoft SharePoint. Each team also maintains thorough documentation and guidelines to help ensure the consistency of its work.

We are highly aware of the importance of communicating our work to make sure it reaches new audiences. Among our outreach activities, we promote the portal online through social media and in person, through talks and tours of the programme. Many of our specialists also offer presentations at academic and archival conferences, participate in seminars, and write articles and blogs for wider publication. The response of users of the portal is overwhelmingly positive: many researchers and students are using this resource, not only in the UK, but also in the United States and across the Gulf region, and the increased access to this material is allowing for studies of a broader and more comprehensive nature than was previously possible.

Thanks to this project, important historical material from the BL's collections, some of which had not previously been fully catalogued or studied in depth, is now being disseminated and made available to the general public. The partnership has just agreed a further three years for this project, until the end of 2021, during which time we plan to make even more material available. We hope our efforts will prove useful to all who access the portal.

<http://www.bl.uk/qatar/>
<https://www.qdl.qa/en>

Noemi Ortega-Raventos
 British Library

Digitising Kresen Kernow (Cornwall Centre)

Eleanor Wilkins discusses her role and her contribution to Cornwall's Kresen Kernow project

Cornwall Record Office is changing as a result of the Kresen Kernow project. The new location embodies a digital identity embraced by all areas of arts, culture and heritage, but more clearly by archival projects. Planning of this project commenced in 2015, with the investment of an £11.7m Heritage Lottery Fund (HLF) grant, with the primary goal of preserving the archives of Cornwall.

Kresen Kernow is being built on the former Redruth Brewery site and will be an exemplary modern archive and library space, bringing together for the first time the world's largest collection of manuscripts, books and documents relating to Cornwall, as well as Cornwall Record Office, the Cornish Studies Library and the Cornwall and Scilly Historic Environment Record.

Currently, I am a volunteer at Cornwall Record Office, and my role focuses on cataloguing the glass plate collection into a digital form on the record office's database. The glass plate negative collection is an under-explored area, both in the sense of Cornish mining and in general research pursuits.

Cornwall is a hub of creativity and innovation, and throughout history the area has been dedicated to the mining industry. The mining industry is more to the Cornish population than simply a lucrative business. It unites families and friends, as well as communities. Therefore archives surrounding the mining industry allow many residents to reconnect with their ancestors and relive fond memories of building a lifestyle in Cornwall.



Our mining culture shaped our world. Image courtesy of Cornwall Record Office

In my role so far I have been able to successfully catalogue between 40-50 glass plate boxes. In doing so I have been able to observe copious different images epitomising the wealthy Cornish mining industry. The images transcribed from these plates not only act as a beautiful memory in their own right, but also when transcribed they can inform and answer many questions that surround the industry. The origins of these glass plates derive from several locations that embody mining, such as Poldark Mine, Geevor Tin Mine, Levant Mine and Beam Engine, King Edward Mine and Carnglaze Caverns. As such, these precious archives from Cornwall Record Office should be made to have a digital and physical form. I wish to continue investing my time in this new development project, as it helps preserve and safeguard the records of Cornish mining.

The Kresen Kernow Project brings about a new home for the stories of Cornwall so that they can be preserved in their original form, as well as in a new digital medium. The new build will come with new challenges and developments, and it will store and display treasures that go beyond the mining industry and beyond the limits of Cornwall, including a kinship with the National Library of Wales, for example. The Kresen Kernow is an exciting and culturally stimulating project that will encourage an influx of new researchers and focus on community involvement. Therefore, all members of the Cornish community will be able to relish in digital and physical memories through this innovative project.

Eleanor Wilkins

Cornwall Record Office

Writing to Wesley: conversion testimonies from 18th century Methodists brought to life in Manchester

The Methodist Church, compared with Anglicanism and Roman Catholicism, is a comparatively modern denomination. But it is well documented. Among the thousands of documents held in the Methodist Archives and Research Centre at the John Rylands Library, University of Manchester, is a unique collection of 153 letters written largely by ordinary people, who had joined the young Methodist movement covering the years 1730 to the 1780s'. These letters were sent to Charles Wesley, one of the founders of Methodism, who had requested first-hand accounts of conversion testimonies, which could be shared to encourage others to embrace this faith. The voices that shine through are vibrant and expressive, captured in their own hand, and are a wonderful exemplar of rarely heard voices from a significant period in British social and religious history.

The letters are steeped in the language of the period, conveying a knowledge of both the Bible and preaching styles of the day. They also reveal real people with recognisable doubts and problems, but nevertheless illustrate lives transformed by faith and their experience of God, and a real insight into some of the first Methodists.

DDWF 5/14 (3)

Dear Jack,

You own abundantly enough to clear M^{rs} Hutton from any misrepresentation as to You, & me from any misunderstanding her: I was but too right in my Judgement.

1. You was not a Christian before May 24. but are so now. in a Sense of the Word You call obvious; which was so far from it, that it astonish'd all who heard You then; & which I deny to be so much as true.
2. You hold the Witness of the Spirit as containing a clear Information of Adoption whereof Pardon is a part to be absolutely necessary to Your Salvation & that of others, unless excus'd by invincible Ignorance. Enough! Enough! Yet
3. You apply Joel amazingly, tho You give up such Visions as I speak of: yet not allowing me to call such Bad Branches of a Bad Root. That I may not be guilty any more of putting them more or less into every Letter, I'll discuss this matter fully by it self once for all, desiring You in the meantime ^{to say} what other Scripture Dreams or Visions You would insist on, whether all between Genesis & Revelations? I am afraid Ahabs lying Spirits may be but too pertinent.

That You were Not-a-Christian before May ^{in Your sense} any one may allow; but have You ever since continued Sinless? or has not the Dominion? Do You never then fall? or do You mean no more than that You are free from presumptuous Sins? If the former I deny it; if the latter who disputes?

Your Misapplication of the Witness of the Spirit is so thro' roughly clear'd by B^p Bull that I shall not hold a Candle to the Sun. What portion of Love Joy &c. God may please

In 2017 a joint project between the Methodist Church in Britain and the John Rylands Library witnessed the digitisation of the 153 letters, making them available online. The original letters in the form of digital pdf files of mid-18th century handwriting were very difficult to read as they were very often a stream of religious consciousness making significant use of abbreviations, using lots of biblical allusions and idiosyncratic use of capitals. To improve accessibility, 60 volunteers, mainly from the north

west of England, were recruited to transcribe the letters. At least three transcripts of each letter are required in order to ensure consistency and validation of the original source and context. By comparing the letters with other sources, the letters were placed in their 18th century context, providing a better sense of the lives lived by their writers, and adding significantly to our understanding of the place of Methodists in wider society.

What then have we learnt from this project? Thanks to the excellent work of the digital imaging team at the University of Manchester, the digitisation process was the most straightforward part of the project. The transcription process was more complex and less reliant on technology in that it took the transcribers time to get to grips with the challenge of transcribing difficult to read handwriting, as well as understanding the religious revivalist fervour of the 18th century that brought about Methodism. The use of enthusiastic volunteers working with the Methodist Archive and Research Centre, and colleagues in the John Rylands Library, is one of the project highlights and has led to other Methodist-related archive and document projects that can only be achieved by the use of volunteers. Possibly the greatest achievement of the conversion testimonies project is that it has given a resonance to a group of men and women whose voices are no longer hidden, but can now sing and illumine a formative period of British social and religious history.

For further information, please access the project website at <http://www.library.manchester.ac.uk/search-resources/manchester-digital-collections/digitisation-services/projects/rapture-and-reason/>

For further information please contact Owen Roberts, Methodist Heritage Officer, Methodist Church at robertso@methodistchurch.org.uk or Dr Gareth Lloyd, Archivist at the John Rylands Library, University of Manchester at Gareth.lloyd@manchester.ac.uk

Jon Purcell
Methodist Archives

Atomic evolution: making the visual history of nuclear applications accessible

The visual history of nuclear science is dominated by images of mushroom clouds and signs warning of radioactivity. Yet from the 1950s onward, the idea of ‘Atoms for Peace’, promoting peaceful uses of atomic energy and nuclear applications to better the world, offered an alternative path forward for nuclear science. The International Atomic Energy Agency (IAEA), founded in 1957 with its dual mission to promote the peaceful uses of nuclear science while safeguarding against the potential abuses, collected and shared images of the many different applications of nuclear technology from its earliest days. This exchange of information is enshrined in Article VIII of the IAEA Statute, declaring, ‘The Agency shall assemble and make available in an accessible form the information made available to it...’. Technological changes have made digital information the preferred ‘accessible form’ for users across the globe. By digitising the historical photographic collection, the IAEA Archives will make the images publicly accessible as intended, fulfilling the call of the Statute.

Beginning in March 1958, the IAEA wrote letters to representatives of member states requesting the contribution of photographs to ‘form a basis for the coverage of activities in member countries in the atomic field’. The photograph collection grew as staff documented their work and organisations around the world contributed images and requested others in a scientific



On Exchange Experiment during the Mobile Isotope Laboratory stop in Vietnam, 1962. IAEA Archives/ Haeupl, Josef.



Scientists of the West African Research Institute, Tafo, Ghana, demonstrating to members of an IAEA preliminary assistance mission the effect of mealy bugs on cocoa plants. The Institute is using radioisotopes in studying the migration of these vectors of the virus disease ‘swollen shoot’, IAEA/July 1961/ No. 798B. IAEA Archives



Jointly with the Food and Agriculture Organisation of the United Nations (FAO), IAEA is promoting rice tests carried out by Hungarian scientists at Szarvas (Hungary). Checking the uptake of a radioactive fertilizer by rice plants. In the background research workers, wearing protective clothes, are distributing radioactive fertilizer on an experimental plot. IAEA/July 1963/No. 1194. IAEA Archives



Uranium prospecting team in Tunisia, undated. IAEA Archives.



Transport plane, Air Zanzibar, used for an aerial release of sterilized tsetse flies, 1997. IAEA Archives



Sorghum crops, part of a mutation breeding project in Venezuela, 1980s. IAEA Archives

exchange. The images were often used in the IAEA Bulletin, a regular publication documenting the activities of the agency. By the 1980s, the IAEA received fewer outside images, relying instead on staff members and paid photographers to capture meetings, laboratory work and equipment, and the numerous technical cooperation projects. At the turn of the century, digital photography replaced analogue, and in time the analogue photo collection was accessioned to the archives. The IAEA Office of Public Information and Communication (OPIC) continues to produce images while also maintaining the digital images taken since 2000.

For the past several years images have been shared on the IAEA's Flickr and Instagram accounts as well as in the quarterly Bulletins, but there was no place for staff or the public to centrally search for and view images. The historical photographic collection consists of about 30,000 images, only about 20% of which had been catalogued in a legacy database, and only a small percentage had been digitised on an ad hoc basis. The

Mobile Isotope Laboratory at University Santo Tomas in the Philippines, 1961. IAEA Archives/Haeupl, Josef.





This picture, taken at Seibersdorf, shows Dr. Kiyoshi Tensho (Japan) measuring the amounts of oxygen taken in by plants with their leaf shoots removed in comparison with the intake by those with normal shoots, IAEA/July 1965/No. 1499. IAEA Archives/Skrein



IAEA at Seibersdorf, near to Vienna, Austria. Analysis for trace elements in plant material by a distillation method (Helga Axmann, IAEA staff member, Austrian), IAEA/April 1964/No.1299. IAEA Archives

interest in and requests for images has rapidly increased in the past few years, making the historical photograph collection an obvious candidate for digitisation. After the decision to digitise, it became clear that a platform to display these images would be necessary. The archives and OPIC agreed to acquire a digital asset management system to use jointly to describe and store some 60,000 historical and current images. The initial launch is planned for the end of 2018.

The ultimate purpose of the project is to unify and improve finding aids and search capabilities by standardising metadata, thereby providing unmediated access to images and their metadata. Using a common digital asset management system presents several advantages. Not

“The ultimate purpose of the project is to unify and improve finding aids and search capabilities by standardising metadata, thereby providing unmediated access to images”

only will metadata of historical images be harmonised, templates will prescribe standardised descriptions for future entries.

The metadata will be further enhanced through keywords based on an evolving taxonomy from OPIC. The taxonomy is a condensed version of more technical nuclear science ontologies used within the agency, and it reflects the structure of the IAEA website. Using the taxonomy will enable images to be displayed on the appropriate pages of the website based on keyword queries. While some keywords have been applied manually, the Archives, in a pilot project within the archives and records management section of the division of general services in the IAEA, plans to harness semantic tagging to auto-tag images based on their descriptions. Initial testing has shown promise, yet it has also demonstrated that the taxonomy would benefit from greater use of alternative terms to produce more accurate results. Other ontologies already in use provide models for better structuring and linking our data.

The fascinating collection of images spanning over 60 years of nuclear science have the power to change the way we see nuclear applications, their development over time and their effect on everyday life. For instance, despite gender disparities, the images show the participation of women in all fields of nuclear science. The transfer of knowledge in the fields of agriculture, health and energy, among others, has always been part of the IAEA mandate. Now we can rediscover the historical images documenting this mandate, learn about the history of science, global politics and diplomacy, and witness how nuclear science continues to evolve. Through digitisation and the application of descriptive metadata, including linked data through semantic technologies, the pool of images available to the public which depict nuclear applications in all their forms will expand, broadening the perspective of the current and historical uses of nuclear technologies and recommitting to the transfer of knowledge.

Elizabeth Kata

The International Atomic Energy Agency

Ingests and upgrades: technological barriers and adapting our preservation practices

Zoe Fullard provides a case study of the Transport for London's digital preservation strategy

Business case accepted. Funding received. Digital repository system sequenced and configured. This was our position at Transport for London (TfL) corporate archives 18 months ago. Since then we have ingested over 58,000 digital files into a cloud-based system accumulating to 692 gigabytes, which can now be actively managed and preserved long-term. Technology has been crucial in achieving this, but what are some of the technological challenges we have faced and how will this inform our digital preservation strategy in the future?

Ingesting our digitised staff registers and internal staff magazines was the logical place to start (after an initial testing period). The staff magazines in particular are an extremely popular primary source. Name-rich, they contain a wealth of information documenting staff activities in both a professional and social capacity and the pdf versions are embedded with optical character recognition, providing an incredibly useful access point for users. Tiffs, jpegs, and jpeg 2000s had also been created for presentation, access and preservation purposes, respectively. However, this content had been sitting on external hard drives for three to five years - uncomfortably close to the life span of the notoriously fragile external hard drive.



Front cover of London Transport Magazine souvenir edition celebrating the Victoria line opening, April, 1969. Transport for London Group Archives LT000030/049, image reference LT000030_076_00801

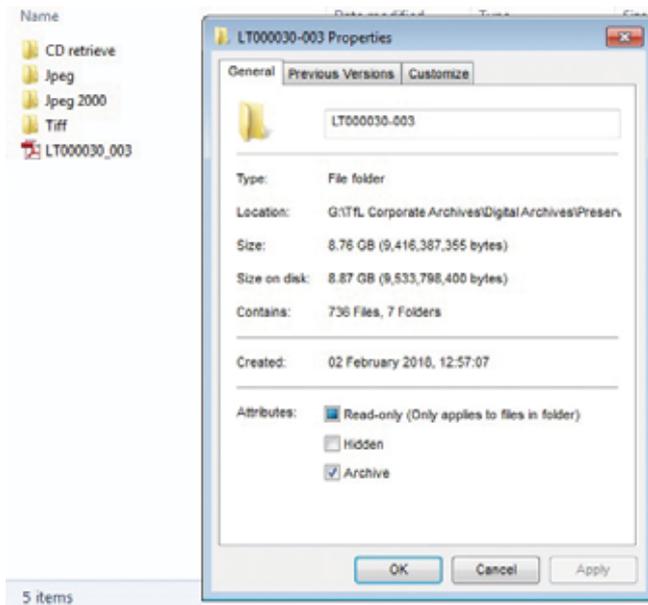


Article from London Transport Magazine discussing Christmas preparations undertaken by London Transport's Food Production Centre. The Centre was responsible for the manufacture and distribution of 30,000 Christmas puddings sold at London Transport canteens. Article includes the recipe for these popular puddings. December, 1968. Transport for London Group Archives LT000030/048, image reference LT000030_076_00742

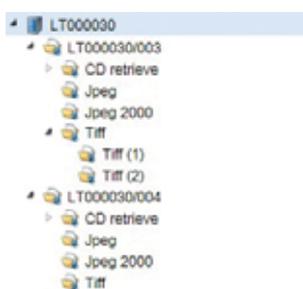


“Sheer scalability of digital record creation means that we will need to process and manage collections using more powerful, automated technological systems”

Selection of external hard drives containing internal staff magazines and staff registers digitised as part of a project beginning in 2012. Zoe Fullard, Transport for London Group Archives.



Screenshot showing typical structure of staff magazines as received from our digitisation supplier. The size of the parent folder exceeds the Submission Information Package (SIP) cloud-based ingest capacity and must be split up into 'ingestible chunks'. Transport for London Group Archives



Screenshot displaying replicated structure post ingest into the digital repository. Note that the Tiff folder is split in to 'ingestible chunks' with content arranged within a new parent folder. Transport for London Group Archives

Our ingest approach is straightforward: to replicate folder structure arranged by file format within the repository, delete un-archival computer support files to strengthen quality of the collections, and ensure material can be rendered and accessed post ingest. Although technology has been imperative to this process, it has also produced challenges.

Despite our choice to opt for a cloud-based repository, client-based software is also required to facilitate ingests. This on-site software must be regularly upgraded in line with the cloud software to ensure compatibility. Within a corporate organisation this activity falls within the remit of the information technology department, where certain processes must be complied with - and rightly so. Their concern is with minimising risks to the organisation's digital infrastructure and they implement the appropriate measures needed to protect this space responsibly. For us, sustaining the link between our cloud-based digital repository and our on-site 'feeder' software relies upon sequencing and upgrading digital systems embedded in a complex, intangible environment. Current business priorities, organisational change, staff turnover, and installation issues cause delays - not good in our time crucial digital environment.

As a response to this, TfL corporate archives has been using an alternative cloud-based ingest method. Despite the automated environment technology provides, this practice currently requires a high-level of manual intervention. Although useful as it does not rely on in-house upgrades, our cloud-based ingest method limits the size of Submission

Information Packages (SIPs) we can create. These limits have meant that the archives team have spent a total of 49.16 hours simply preparing content for ingest, including:

- Creating working copies of material to reduce accidental deletion and retain the ability to replicate original structures (26.05 hours)
- Splitting and zipping folder structures and their related content into ‘ingestible chunks’ (20.93 hours)
- Re-creating original structures post ingest (2.18 hours)

This labour-intensive approach will almost certainly not be sustainable going forward. We have accumulated eight terabytes worth of material over a six-year period. In addition to this, the crossrail archive consisting of billions of documents, is being transferred to our custody and will eventually be ingested and preserved long-term in our digital repository. Sheer scalability of digital record creation means that we will need to process and manage collections using more powerful, automated technological systems.

Nevertheless, working on such a manual and granular level throughout several stages of the ingest process has been beneficial for us. It has provided the opportunity to comprehensively document our actions, identifying where resources are being spent and where improvements can be made. This has informed our decision-making in regards to perceived technological barriers, enabling us to create a more effective digital preservation strategy, including:

- Raising upgrade requests as soon as the most recent is completed.
- Building relationships with IT colleagues and advocating our position.
- Encouraging IT colleagues to log actions undertaken to create upgrade process documents and safeguard this knowledge against staff turnover.
- Ingesting content in bulk when full systems are available.
- Maximising time between upgrades to develop policies and procedures, train volunteers, assign ingest priorities to material, and assess metadata schemas.
- Liaising with developers and providing suggestions about our future requirements.

Technology impacts our environment and poses challenges that are sometimes beyond our control. At TfL corporate archives, we have learnt that by taking action, no matter how big or small, we can begin to assess where we need to adapt our practices to ensure we are managing our digital records responsibly and within our resources. Being proactive in our digital preservation mission is better than taking no action at all.

Zoe Fullard

Transport for London Corporate Archives

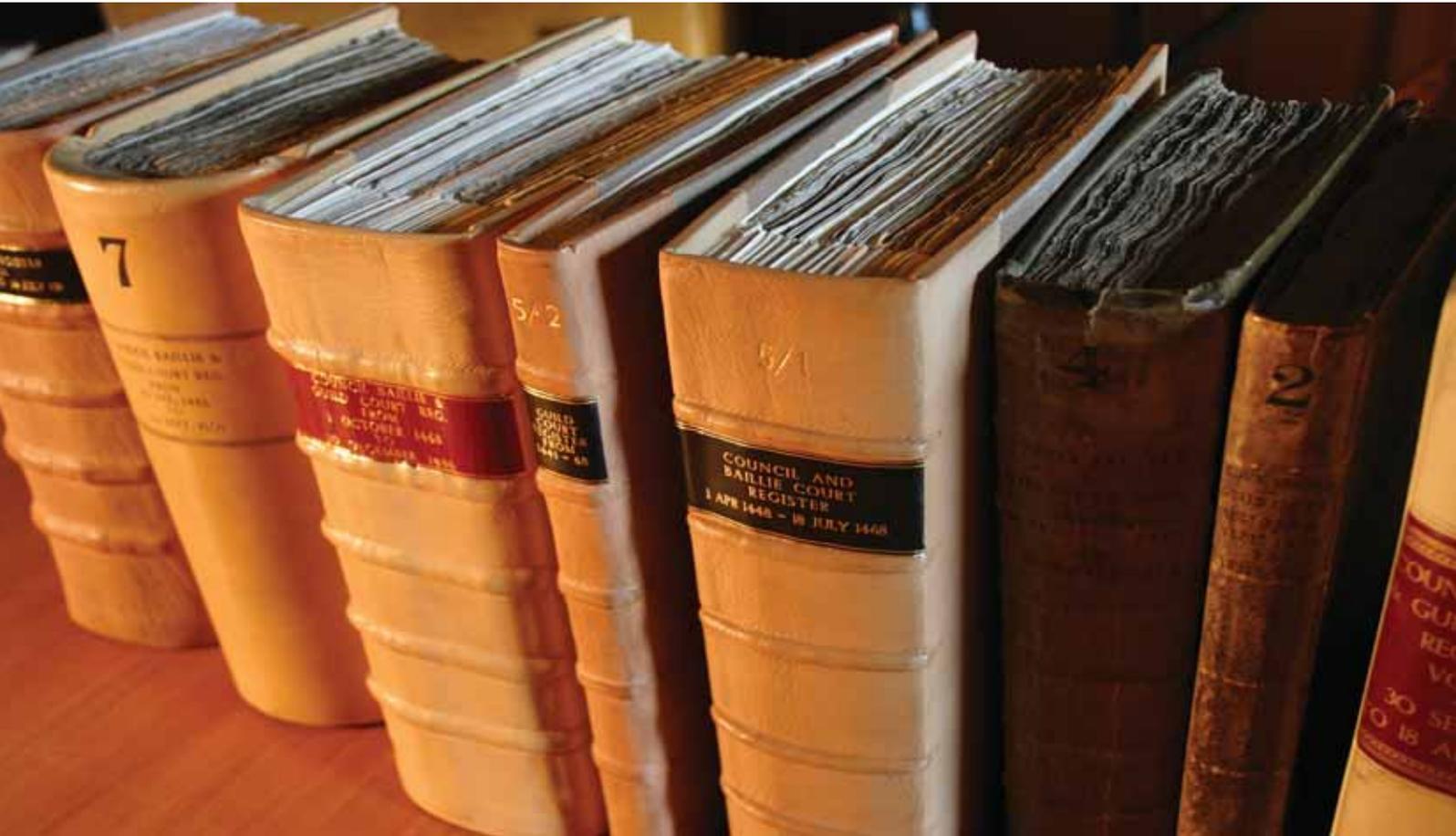
Transcription and technology: law in the Aberdeen Council Registers 1398-1511

Phil Astley explores the digital transcription project of Aberdeen City and Aberdeenshire Archives

The Aberdeen Council Registers, 1398-1511, are the single most important source for the study of Scottish urban history for the late medieval period. While some records do survive for other Scottish burghs between these dates, they are fragmentary and lack the continuity and completeness of the Aberdeen material. Indeed, the unique nature of these eight volumes, which contain the proceedings of the Baillie, Guildry and Head Courts, saw them inscribed on the UNESCO ‘UK Memory of the World Register’ in 2013.

In 2015 Aberdeen City and Aberdeenshire Archives joined forces with the University of Aberdeen in a successful application to the Leverhulme Trust for £310,000 of funding to create a digital transcription of the eight volumes. This is the latest and largest of three projects between the archive and the university all of which have had the early records of Aberdeen as their focus.

To give some sense of the scale of the task of the current three-year project, the volumes comprise 5,238 pages containing some 1.4 million words in Latin, Middle Scots and even some Dutch thrown in for good measure. The project team is drawn



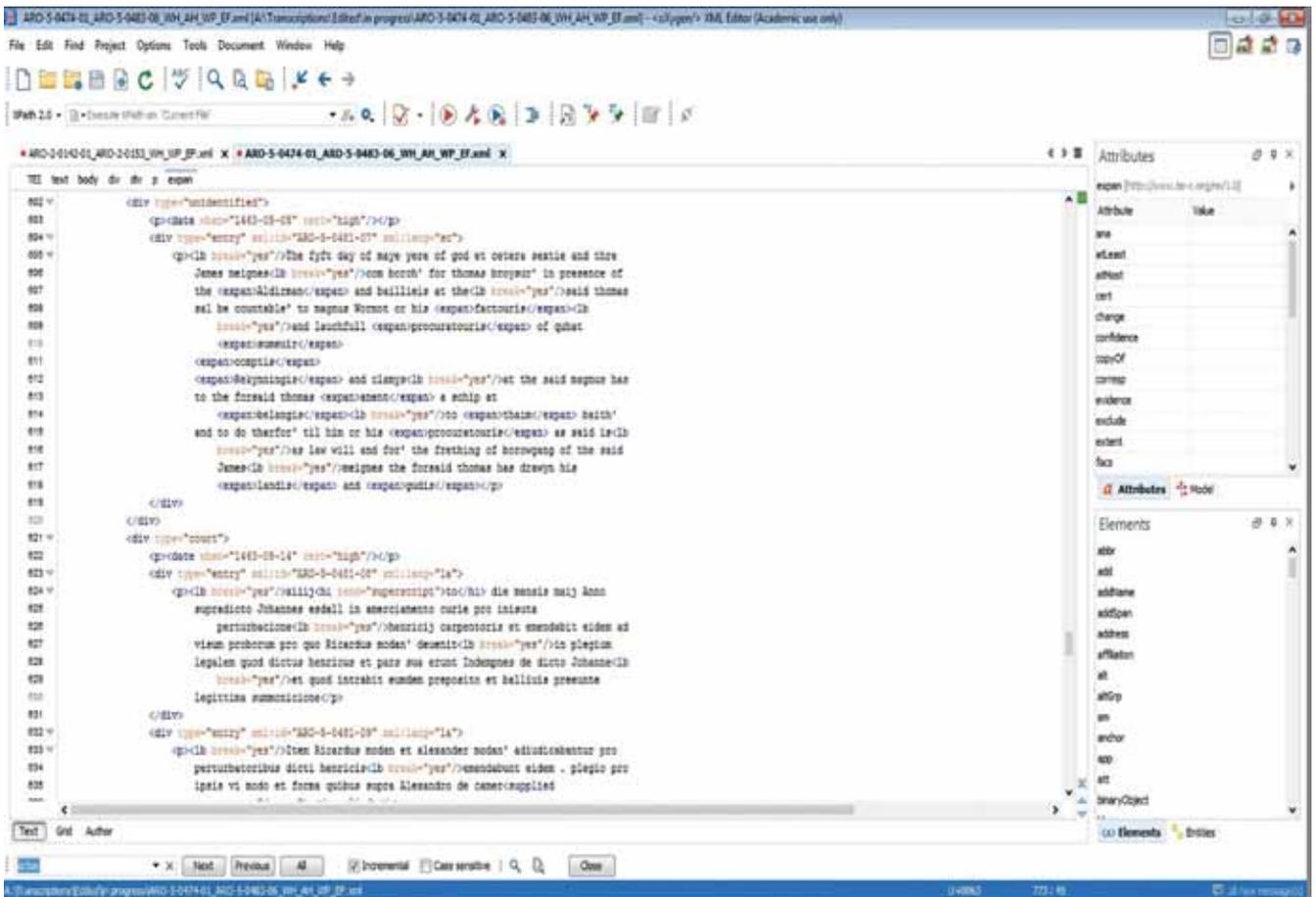
The eight volumes of the Aberdeen Burgh Registers, 1398-1511. Vicky Gray Armstrong



The corresponding image of page 481 of volume 5/1. Aberdeen City & Aberdeenshire Archives

from various disciplines including, archivists, legal historians, linguists and computing scientists. By creating a digital transcription, the aim is to make the records accessible to a much wider audience: only a small percentage of the records have previously been published by, for example, the Spalding Club, which produced a volume of extracts in the 19th century. The project team use the ‘Oxygen’ extensible markup language (XML) editor with the ‘HisTEI’ add-on, which enables transcriptions to be made in a word-processor-like view and which are compliant with the Text Encoding Initiative (TEI) guidelines. In this way, a corpus of text is produced that is both machine-readable and machine-searchable. It is a painstaking and time-consuming task: tests on automatic transcription using the software package ‘Transkribus’ have been tried as part of the project, although the accuracy is such that manual transcription is still quicker due to the amount of subsequent checking required with the automated process.

What are the benefits of a digital transcription? Why do we go through all this effort when the pictures of the Aberdeen Burgh Records are already available online? Apart from the obvious benefit of a digital,



A portion of the transcribed and marked-up text from page 481 of volume 5/1. University of Aberdeen

transcribed version of text being much easier to read than the original handwriting, it allows for information to be added to the text. With the help of so-called ‘tags’, a text can be enriched with all kinds of structural annotations and metadata. Tagging here means adding XML annotations to the text.

For example, the textual passages in the Aberdeen Burgh Registers, which are mainly written in Latin or Middle Scots, can be marked up as such, using the ‘xml:lang’ tag. A researcher who is interested in the use of Middle Scots in these registers could then search for and find all Middle Scots sections in the corpus very easily with the help of a text analysis tool such as ‘AntConc’ or ‘SketchEngine’ without having to plough through the sections written in Latin. More generally, enriching the text with tags means that a researcher does not have to read through all the 5,238 pages of the Aberdeen Council Registers to find what they are looking for. A machine-readable and machine-searchable text not only saves time when researching a particular topic, but is also generally more flexible than a printed version because further tags can be added and unwanted tags can be hidden.

To genuinely open up the records to a much broader

audience, the method of interrogating the text needs to be more user-friendly. As a parallel to the main project, and as part of their degree course during the 2016/17 academic year, a class of third year computing science students from the University of Aberdeen was given the task of designing a search tool which could be used by the public. What they produced is a very powerful and flexible search engine for the Aberdeen Burgh Registers. It will be hosted by Aberdeen City Council and launched later this year. It can contend with the many spelling variations inherent in medieval text and displays relevant transcribed passages in response to searches, as well as images of the corresponding original pages.

Although the current project comes to an end in March 2019, the digital transcription that has been created has laid the foundations for future collaboration and research into this unique and fascinating collection of records.

Phil Astley

Aberdeen City & Aberdeenshire Archives

Bringing qualitative data to life through online access and citation

The UK Data Service (UKDS) makes available collections of primary research data to social scientists. It has around 1000 records relating to qualitative data and mixed methods collections that include transcribed interviews, original survey questionnaires, open-ended questions, essays and photographs. These can be found using the data catalogue, which holds a detailed metadata record for each study/collection and enables download of the whole data collection.

While many datasets can be downloaded, content cannot be interrogated until the individual data has been downloaded. In 2014, the ‘Digital Futures’ project built and launched QualiBank, an enhanced and user-friendly platform for publishing, presenting and searching the content of qualitative data. The QualiBank system offers a streamlined data publishing workflow for data files and metadata marked-up with known schemas.

QualiBank functionality

The UK QualiBank (<https://discover.ukdataservice.ac.uk/QualiBank/>) offers a search and faceted browse interface allowing searching and retrieval of the content of textual files and of the metadata attached to these objects, such as a description of an image or of an audio recording. It enables hyperlinking to related objects and includes a persistent citation feature enabling users to select a whole text or an extract of the text.

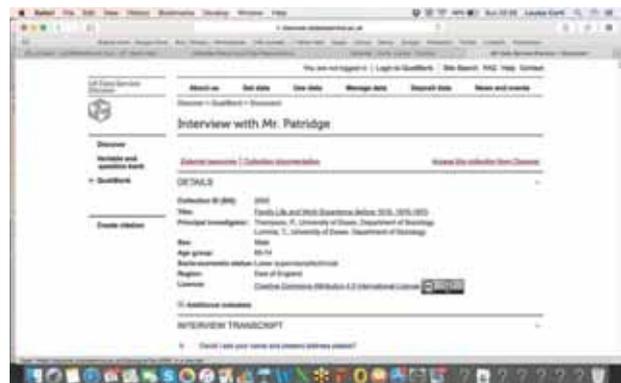
Summary metadata is displayed below the title of the object and where search terms appear in a text object (e.g. an interview transcript) they are highlighted in the text in yellow.



QualiBank search results. UK Data Service, 2018



A UKDS Data Catalogue record for a qualitative data collection. UK Data Service, 2018



Interview metadata from qualitative text in the UK QualiBank. UK Data Service, 2018

Linked resources (e.g. text, audio and images) are identified by a unique icon. Target pages, such as an interview transcript, display core metadata and view additional metadata.

Materials related to a text object such as audio, images, or web resources are also clearly displayed.

INTERVIEW TRANSCRIPT

I: Could I ask your name and present address please?

R: John Perry Partridge, 11 King George Avenue, East Tilbury

I: And your marital status, you have been married?

R: Married twice

I: Could you give me the year of your marriage?

R: 1926. That was the first one. 1951, the second one.

I: And your year of birth?

R: 1902

I: And the place you were born?

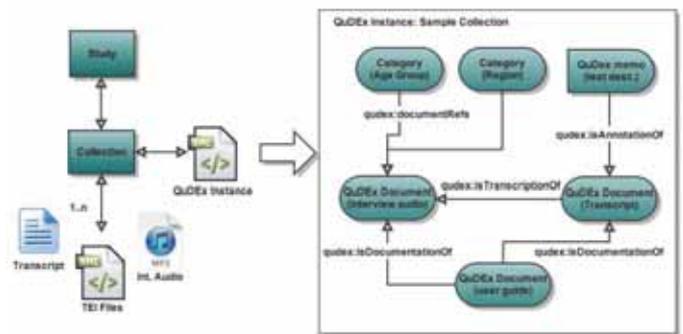
R: Camden Town.

I: Do you know the address?

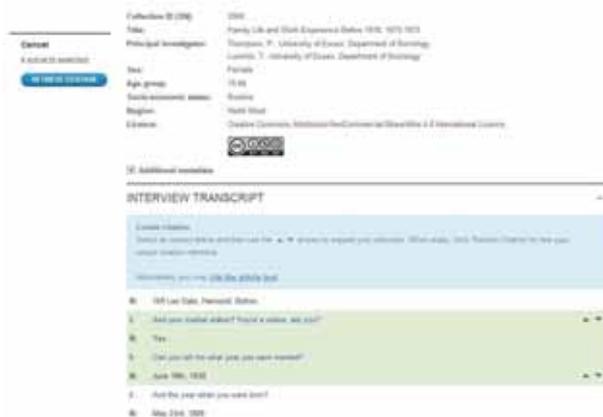
R: Kings Road I believe it was.

I: How many years did you live in Kings Road, Camden Town?

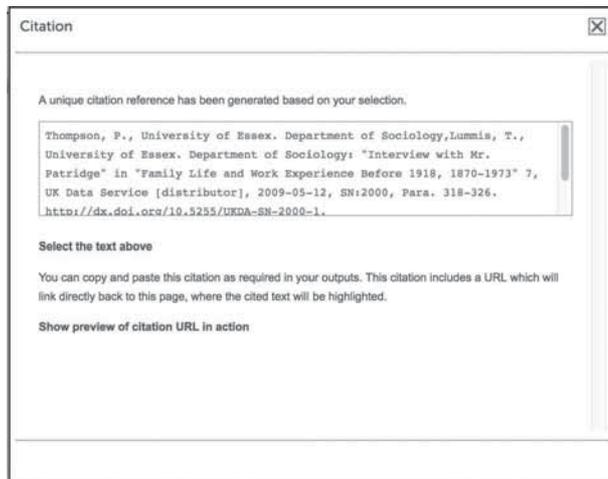
Interview text from qualitative text in the UK QualiBank. UK Data Service, 2018



Use of QuDEx metadata in QualiBank. UK Data Service, 2018



On-the-fly citation created for an extract from qualitative text in UK. UK Data Service, 2018



Citation of an interview extract in QualiBank. UK Data Service, 2018



Simple MS Sharepoint metadata entry tool for QualiBank. UK Data Service, 2018



QualiBank interview marked up in TEI. UK Data Service, 2018

Citation functionality

Users can select an extract of text, based on one or more consecutive paragraphs and dynamically retrieve a citation, which can be pasted into any outputs.

The citation URI created takes the user back into the quoted extract (highlighted in green) in QualiBank, unless the data is safeguarded and needs a login.

Metadata requirements

The ability to provide context for data objects requires sufficient descriptive metadata for each object, and metadata must capture relationships between objects in a collection. To prepare materials for the system, data must be highly structured and consistently marked-up. QualiBank exploits known technologies and utilises three standardised metadata schema:

1. The Data Documentation Initiative (DDI) - an open international standard for describing the data produced by surveys and other observational methods in the social, behavioural, economic, and health sciences, used by all national social science data archives around the world (<https://www.ddialliance.org/>)

2. Qualitative Data Exchange (QuDEx) - an XML schema maintained by the author at the UKDS, which preserves annotations of, and relationships between, data, parts and segments of data and other related objects in a qualitative data collection. It is a baseline data exchange model for the

“In 2014, the ‘Digital Futures’ project built and launched QualiBank, an enhanced and user-friendly platform for publishing, presenting and searching the content of qualitative data”

archiving and interchange of data and metadata. A simple MS SharePoint tool is used to enter metadata in a tabular format. In QualiBank it is used for rich file-level description and intra-collection relationships (<http://www.data-archive.ac.uk/curate/standards-tools/metadata?index=2>).

3. Text Encoding Initiative (TEI) - a standard for the representation of texts in digital form. QualiBank uses a limited bespoke set of TEI XML elements from a massive schema for structural mark-up of text: headers, speaker turn-taking, defining paragraphs and inline corrections and errors (<http://www.tei-c.org/>).

Technologies in use

QualiBank uses a range of commonly used technologies for web interfaces that enables searching and retrieval of data and metadata and deals very well with large quantities of qualitative data. It uses an XML database (BaseX) and Xquery for rapid metadata, and textual data storage and retrieval, Solr indexes used for faceted browsing and TEI text highlighting, and a GUID generator SQL database for QuDEX and TEI elements (that enables the fine-grained citation feature). The system avoids the use of hard-coding, so often found in multi-media archives, making data publishing a much more straightforward process, and accommodating changes far more easily.

Louise Corti

UK Data Archive, University of Essex

Kinaesthetic learning for technical skills in web archiving

Trainee digital archivist Kelly Burchmore on the Bodleian Libraries’ Developing the Next Generation Archivist programme, where she is cultivating digital curation skills through a combination of projects in Special Collections alongside distance learning with Aberystwyth University

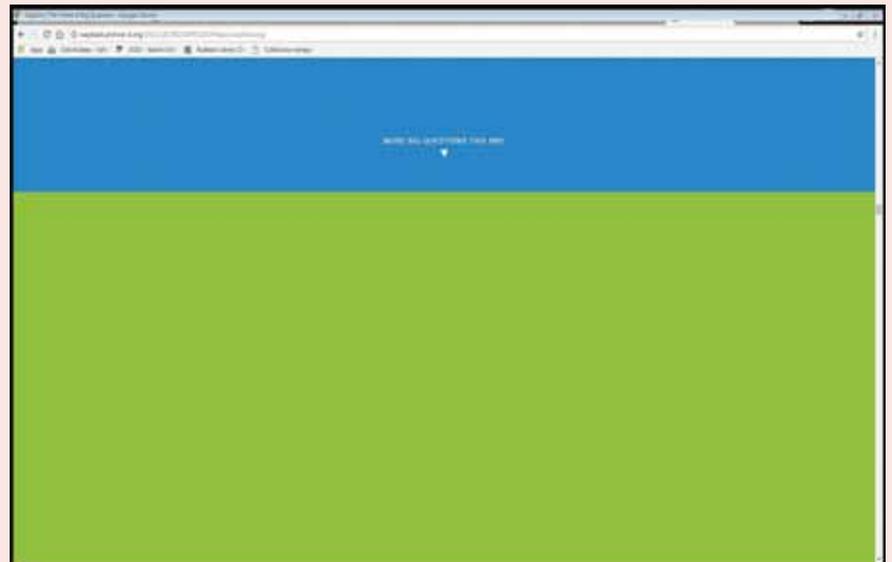
As someone who 18 months ago did not know a gigabyte was larger than a megabyte, (it’s fair to say I didn’t have a technological background!) I have loved learning new skills. I would like to focus on the technical skills I have developed during the traineeship through participation in the Libraries’ web archiving activity, and particularly on how I have consolidated my knowledge and confidence through experimentation.

With Archive-it, our web archive partner, you have the opportunity to run test crawls and follow them up by both reviewing the archived capture and the crawl report. The crawl report breaks down any blocked, queued or out-of-scope hosts (where web content is stored),

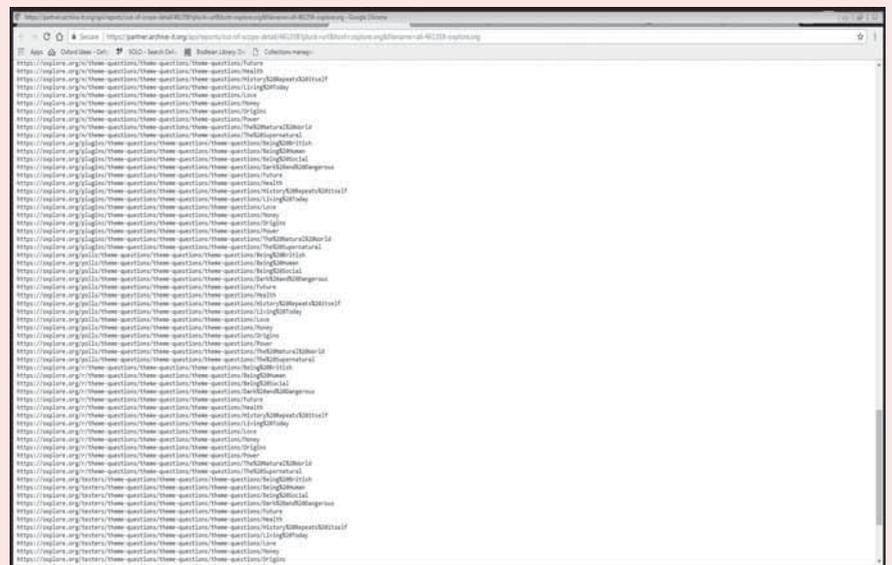
indicates file types archived and shows the total as well as new data archived for the site. This was pertinent to my learning and navigating my way around the workflow. I could try expanding or limiting the scope that would be crawled for each site, ignoring robots.txt crawler exclusions or crawling sub-domains along the main domain in order to fix the archival capture. The experimental nature of this meant I could be bold and take risks instead of being afraid to test the solution for myself and to try something new. If my solution did not work then I could keep a record of what I had tried and delete the test crawl along with any new data.

A frequent challenge the team and I would come up against in working on the Libraries' web archive, was capturing the more dynamic content that would then function and replay in the archival records of the sites. A new crawling technology developed by Archive-it, Brozzler, helped me to try out a different solution for this. Brozzler captures HTTP traffic as it is loaded and uses a real browser to record server and web browser interactions. After hearing about it at the 'Bountiful Harvest: curation, collection and use of web archives' event, hosted by the ARA and Archive-It last year in Manchester, I started looking for potential sites in our collections I could test it on. The crawl technology was enabled for one-time experiment test crawls, so it was easy to take a risk-approach and try it before committing to manually applying it on a regular frequency. I took an ambitious, 'just-try-it' approach.

There was one particular site in our collections, with really exciting and interactive navigation menus and playable media, for which I was keen to capture a quality record - Oxplore.org. This is a site widening participation and access at Oxford University, which encourages and promotes access to bigger thinking for young people. Oxplore's home page is really dynamic and engaging, with a myriad of 'big questions' for users to browse and explore and which is reliant upon Java-script and the human interaction for the user to choose their



A first test crawl of Oxplore.org, with the home page lacking its dynamic content. The University of Oxford, 2017



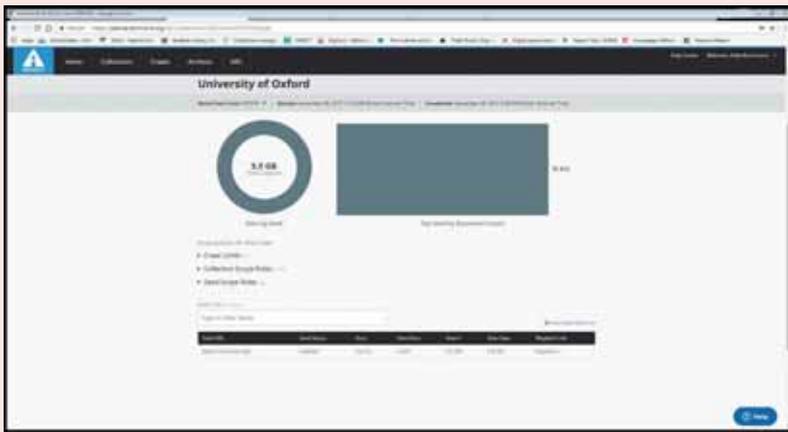
The hosts in the crawl report show many of the plug-ins and interactive polls out of scope of the crawler. Archive-it, 2017



A successful capture after trying Brozzler. The University of Oxford, 2017



Video content captured and replayed! The University of Oxford, 2017



The data jumped from 126.8 MB to 5,5 GB with the crawling technology. Archive-it, 2017

thoughts on each topic. There was also problematic media such as a short video explaining the purpose of the site that was essential for our record. Test crawls with the standard technology and editing scoping rules enabled did not seem to be working, so it was the perfect situation to run a Brozzler crawl.

This resolved both the capture and replay of the site in the Library's web archive. Interestingly, due to the nature of Brozzler, I found that I would have to keep an eye on data budget and make logical adjustments to try to reduce data captured when crawling this site. I still used Brozzler, but instead I tried blocking some hosts that were not in our collection scope. However, due to the effectiveness of the crawler they were captured anyway.

Through experimenting with the technology available, combined with applying logic, persistence and of course careful awareness of data, I think skills for practising web archiving can be learned, developed and retained. I've certainly found out that instead of being afraid of getting it wrong, it is better to embrace and test the technology and tools at hand, to make mistakes, try again (and again!) to solve the problem. I would like to encourage anybody from a non-digital background not to worry about starting or getting to grips with digital archives. The more I experiment, the more I learn, and the more I learn, the more confident I am to experiment.

Kelly Burchmore
Bodleian Libraries

Calling all colleagues!

ARC is always seeking articles reflecting the issues that matter to you most. We would love to publish pieces that reveal the sector's opinion and showcase successful best practice.

If you would like to send something for inclusion in the magazine, please send articles to arceditors@archives.org.uk, or write and let us know what you'd like to read about. Guidelines for articles for ARC can be found on the Association's website: www.archives.org.uk.





Super tough

From capes for superheroes to a christening gown for a princess, our sturdy and durable large textile storage boxes are excellent for keeping fabrics, garments, costumes and other materials safe and secure. The unbuffered blue/grey exterior and white lignin-free bond interior ensures complete protection, while the metal edges provide extra strength and durability.

pel
giving history a future

T **+44 (0)1379 647400**

W **www.pel.eu**

E **info@pel.eu**

@preservation_e 

Preservation Equipment Ltd, Vincennes Road, Diss, Norfolk, IP22 4HQ, UK



CONSERVATION BY DESIGN



TURNING OVER A NEW LEAF



After nearly thirty years, Conservation By Design is now to be known as CXD. We all felt that with continuing global growth in business, it was time for us to consolidate into a single, easily identifiable brand.

Needless to say, as always, we're here to help with our comprehensive range of high quality conservation, preservation, storage and display products.

CARE • CONSERVE • CREATE



BOXES &
STORAGE
PRODUCTS



PAPERS,
MATERIALS &
BOARDS



EQUIPMENT &
ACCESSORIES



STORAGE &
WORKSHOP
FURNITURE



SHOWCASES &
MOUNT MAKING

Timecare Works, 2 Wolsley Road, Kempston, Bedford, MK42 7AD, United Kingdom
T +44 (0) 1234 846300 E info@cxdglobal.com www.cxdglobal.com