Welcome to the sixth and final issue of Off the Record!

We have a brilliant issue for you, with a selection of exciting and inspiring articles, which should hopefully provide the sparks needed to counter any ‘writers block’ for those of you still trying to decide on a topic for your dissertations. I hope you enjoy reading the issue.

I for one, will be sad to see the magazine go. But we are moving onwards and upwards and the new blog will provide us with a more continuous presence which will also enable you – our readers – to say something back.

Emily Gresham, Editor
Update from the Chair

This last issue of *Off the Record* as an e-magazine, arrives just after the ARC New Professionals Special Issue gave us a lot of food for thought. The Opening Lines by Dr Melinda Haunton points out the parallels between the ARA’s Framework of Competencies and TNA’s Archive Service Accreditation, reminding us all that both institutions have been working closely together to make the sector stronger. Reading the issue from cover to cover, I felt that it took a provocative approach to starting out in the current technological landscape. It also coaxes established professionals to revaluate their mind-set and invites us all to embrace our continuing professional development for personal and sector advancement.

I will quote Cathy Williams to summarise the spirit of the issue: ‘So if we think about ourselves as new professionals, we can all help shape a new profession.’ My thanks go to the contributors of our first edition of ARC. The team has received positive feedback and it has been widely welcomed by the sector.

With regards to the ARA membership benefits, Sean Rippington and I compiled a list in preparation for a talk about the ARA to students at UCL in March. I don’t need to emphasize that the main benefit, especially for new professionals, is to be a member of a recognized professional body. If you are a member and unsure about the current benefits see a list below. If you aren’t a member yet, see what you are missing!

**ARC Magazine**

**ARC Recruitment**

**ARA Today**

**Registration Scheme**

**Free membership of regional groups and sections**

**Free best practice guidelines**

**Peer-reviewed journal – Archives and Records**

**Access to CPD courses**

**Priority booking for events**

**Discounted fees for trainings and workshops**

**Discount for publications**

**Bursaries to attend national training events and ARA Annual Conference**

**Bursaries to attend international conferences**

If you are an ARA member and interested in attending a conference overseas that is relevant for your professional development, but your current work is unable to support your attendance, I would encourage you to apply for the ARA’s International Conference Bursary. For further information, please check the link: http://www.archives.org.uk/si-international-engagement/section-for-international-engagement.html. I am very grateful to have been granted some funds to attend the 5th Conference on Database and Archives promoted by the Brazilian Association of Archivists in Rio de Janeiro this June. There will be international key speakers, including our fellow Britons Dr Andrew Flinn and Dr Elizabeth Shepherd, in addition to a number of Brazilian practitioners sharing their knowledge.

I will be reporting on the experience of attending my first international conference and key points learned on the *Off the Record* blog. Our blog editor, Emily Gresham, has created a new regular feature *Event Watch*, which the wider community is invited to contribute to. It is a good opportunity for those attending an event (training, workshop or conference) who are willing to share it with their peers. We are particularly interested in hearing from those attending events that are not in the main stream of the profession. *Event Watch* and other features on the blog are good platforms for those wanting to practice their writing skills, so take advantage of it!

Finally, I want to thank those who participated on the peer-mentoring pilot phase, which started last year and finished in early 2013, for taking part in our survey and giving us helpful feedback. The team is taking the scheme further, revamping and re-launching it to better support our community. So, watch this space!

So, without further ado, my sympathies for those writing up their dissertations this summer, good luck for the job hunters and best wishes for those preparing to start their postgraduate studies. For those attending the Summer Seminar on 21st June, see you there.

_Fabi Barticioti, SfNP Chair_
SfNP News:

Summer Seminar 2013

Spaces are filling up fast at our annual Summer Seminar, taking place this year on the 21st June in Edinburgh. If you’d like to find out more about the day, or book a place, visit http://www.archives.org.uk/images/documents/SNfP/ssreg13.pdf

Are we representing you?

The Section for New Professionals exists to support and represent the interests of new information management professionals. We need to know what really matters to you in order to achieve this, so contact us at newprofessionals@archives.org.uk, write on the wall of the Section for New Professionals community on the ARA web site (http://www.archives.org.uk), or post a comment on our new blog and let us know what you think.

Off the Record - last issue and new blog

This is the final issue of Off the Record in its current form. We are replacing it with a blog which is already up and running, take a look by visiting aranewprofessionals.wordpress.com

We hope that you find the change in format beneficial, as it will enable us to keep you more regularly updated with our activities. It should also provide a better platform for our readers to get involved and participate in discussions.

If you’d like to write something for the blog do get in touch, at newprofessionals@archives.org.uk. Posts can be about anything just so long as they are relevant to any or all of the information management professions. Whether you want to discuss the future of archives in the digital age, career prospects, or a collection you’ve recently catalogued, feel free to submit your post to us for inclusion in the blog.
Growing up I never considered a career in anything related to history or archives. My childhood dream was to become a vet, though in my late teens I realised that my understanding of chemistry and biology would never reach the level needed to do so. By pure chance I secured, at the age of 14, an after school job as a library assistant at the local library, which was located in the small village in southern Denmark where I grew up, for a couple of hours each week. This was a dream job for a teenager who would read 2-3 books per week during the summer holidays. Therefore, when the plans of becoming a vet were abandoned, it was easy to settle on a career as a librarian.

Having finished the Danish equivalent to A-levels I decided to have a gap year as an au pair in the UK. After 18 months as an au pair I decided that I wanted to study here, though I was not sure what I wanted to study. I had researched how to become a librarian in the UK and was advised to get a BA in any subject and then do a post-graduate degree in library science. I narrowed it down to either Scandinavian Studies or History. In the end I settled on studying history at the University of Reading. It was during my studies there that I began thinking of becoming an archivist rather than a librarian.

“I soon realised that paid positions, which would provide the necessary experience, were few and the competition tough”

In retrospect, I am not surprised I ended up deciding to become an archivist. Having grown up in the southern part of Denmark I have always been surrounded by history and one of my early memories of a school trip to a museum dedicated to the Second Schleswig War (1864-1865) is of a bloodstained letter, written by a Danish soldier to his family, found in his pocket after he had been shot on the battle field and sent to his family. Even an early age I was fascinated by the survival of this document. Later, while at secondary school we used copies of original documents and photographs when studying the Cold War and at university I had the pleasure of studying the Paston letters in transcribed and published form. The fact that these very different documents could be used to teach us about the past and give us an insight into how historical events occurred and were seen by people experiencing them intrigued me.

Having decided on a career as an archivist the task of gaining enough experience to be accepted onto the University College London (UCL) course began. I soon realised that paid positions, which would provide the necessary experience, were few and the competition tough, so I sent my CV and covering letters to a selection of archives asking if they could use a keen volunteer. Soon I found myself volunteering at the British Postal Museum and Archive (BPMA), where I was given the task of cataloguing and digitising a collection of photographs, both prints and glass negatives. I had a most excellent volunteer supervisor who introduced me to ISAD(G), Photoshop and CALM with great patience. After a couple of months volunteering I applied to be accepted onto the MA Archives and Records Management course at UCL. Unfortunately I was not successful. I continued to volunteer while applying for paid posts and after a year of volunteering I was offered a job to work on a stamp scanning project.

“there were not a lot of jobs advertised and those suited for newly qualified archivists were in high demand”
at the BPMA. The job involved researching special issue stamps and creating a database of information and images of the stamps. I once again applied to UCL and was this time accepted. When the stamp scanning project finished, three months before the UCL course started, I was offered a contract to work as an archives assistant in the BPMA search room until starting the course. Therefore, when I started the UCL course I had a nice overall idea about how working as an archivist could be.

I chose the UCL course full time mainly due to location. London has a lot of archives and the course made good use of its connections to archives, archivists, records managers and other information professionals though visits to archives and visiting speakers. This gave us students a great opportunity to see how different archives worked and allowed us to ask many questions of visiting speakers. If I had to highlight parts of the course it would have to be the modules on palaeography and preservation, both had excellent and very engaging lecturers and the skills I learned have been used regularly since graduating. On the other hand I was less keen on the records management module, and swore that I would never work anywhere where I had to be involved in records management. It is somewhat ironic that the subject I was least keen on during the UCL course is the subject which I find very interesting now.

During late spring and summer, while writing my MA dissertation, I began applying for archivist positions. It was disheartening to receive one rejection after another. At that time there were not a lot of jobs advertised and those suited for newly qualified archivists were in high demand. After months of applying for jobs without getting to the interview stage I suddenly found myself invited for three interviews in one week in early October 2010. The first interview was at News International and, after a rather nervous interview, I was offered the job on an initial nine month contract which has been regularly renewed since.

Looking back over the last two and a half years at News International it is safe to say that my first professional post as an archivist has been rather unusual.

“I have no doubt that being an archivist is the perfect job for me.”

After the move there was very little time to settle into our new “home” as the News of the World closed as a result of all the phone hacking stories. Subsequently we have been required to provide some of the holdings we manage to the various investigations which have followed. We have also reappraised the records management system, focussing especially on improving how modern records are managed and accessed once they have entered the system.

I now work on two sites; two days per week at the archive in north London where I frequently deal with researchers on a one to one basis, and three days at head quarters in Wapping. This means the company archivist and I only usually see each other once a week. It has taken some getting used to working in this way, and having to deal with first a move and then the issues surrounding the closure of the News of the World means that I have only in the last few months managed to get to know which part of the collection to consult when answering certain types of enquiries.

That said, I have also had the pleasure of being involved in more usual archive work such as cataloguing, preparing exhibitions and my favourite part is writing short articles about all the gems in the archive for our staff newspaper every month.

I have no doubt that being an archivist is the perfect job for me. I think that my love of good stories (remember all those books I read during the summer holidays?) is what attracts me most to being an archivist. There are so many stories just waiting to be discovered and told in the archives throughout the world, and I believe that finding, sharing and preserving these stories is a privilege and definitely the best job in the world.

Anne Jensen, Archivist Assistant at News International
Q: What tips do you have for archivists wishing to work internationally, as you do? What are the biggest challenges with working across different record-keeping cultures?

A: For international work its best to join one of the International Council on Archives (ICA) sections depending on your interests. The individual subscription is quite cheap nowadays 20-100 Euros depending. If there is someone already involved in international work that you know ask them to advise since you may be interested in some form of archival exchange abroad or have some skill you would like to offer to an archival programme. Have a look, for example, at the International Records Management Trust (IRMT) website and also contact Kelvin Smith of the Association of Commonwealth Archives (ACARM).

The challenge remains to understand the archival context of the countries involved and their social and political set ups which, even in North America or Western Europe, are often very different, although the same terms are often used they do not quite mean the same! The real challenge is to accept that other archivists have other priorities to us and that we, if we wish to help them, need to respect that and not impose, however benignly meant, our own solutions to problems which are not the ones they are concerned about.

Q: What was your proudest achievement as Chief Executive of the National Archives?

A: Well I suppose it would be helping to bring about the existence of the National Archives in 2003. Others will have their own views.

Q: The work of the Hillsborough Independent Panel was clearly groundbreaking in its full disclosure of official records. Do you think this will be
the future role of archivists and records managers, i.e. to provide full open disclosure of information which may have otherwise been closed or under retention?

A: The Hillsborough Independent Panel disclosed more than official records and brought together records not normally seen at all e.g. police records which are not public records, except for those of the Metropolitan Police.

It remains to be seen exactly what effect such a discovery process and then digital disclosure actually has on the profession. The Hillsborough records needed to be negotiated into the public domain which may not be what colleagues would like to do, nor have the authority to do.

Q: What are the issues to consider when disclosing records that contain sensitive information to the wider public? Are these issues different today than previous times?

A: This is a big question and, in so far as it relates to digital disclosure on the web, it is indeed different from making records available at a record office or other repository. Disclosure on the web requires the application of data protection and other criteria since the material is published not just archived and made available to individuals. This much is obvious, I’m sure, but drawing up the disclosure requirement according to legal and other provisions and then negotiating it with a number of different authorities with different assumptions can be a very long drawn out process.

Q: The findings of the Hillsborough Independent Panel received a lot of national and international press coverage. To what extent do you think that the public were clear about the role of archivists and records managers in the process?

A: Well I think the public recognised that lots of records had been made available on the website and probably realised that some professionals had been employed. If pressed they would probably have noticed that there were archivist / information specialists on the Panel. They would realise that records had been catalogued but whether they would have appreciated the full enormity of the job I doubt. Certainly the Families and those directly involved were themselves pleased that archivist / information specialists were on the Panel, since they thought correctly as it turned out, this would mean relevant records would be discovered and then made available to them via the website.

Q: What areas would you be focussing on now if you were starting your career again? Do you have any regrets in your career?

A: It is difficult to envisage starting again since the situation now both professionally and economically is so different from the 60s when I began as a volunteer in
the Essex Record Office and then went to the British Museum, the British Library and finally the Public Record Office (via a short project with the Cabinet Office). I cannot imagine that it would be possible to do what I did since the professionalism and commitment required from young archivists is immense.

With the softening effect of time, I don’t really have any regrets but could wish I had got things to work better from time to time and we had got some newer legislation but the times were not propitious. I am enjoying my present involvement in archives and records management and long may it continue.

Q: Do you think entry routes in to the profession now compare favourably with the routes in your early career? How do you feel about the Masters / Diploma as the gateway to the profession?

A: Yes is the short answer. But of course the employment situation is pretty awful and the opportunities for advancement remain limited.

Q: You’ve had high profile roles in libraries, galleries and museums. What have been the main differences between these working environments, if you think there are any?

A: I could write an essay on this. The main differences are to do with our different objectives and who we work for. The museums and libraries collect things or books / archives etc. they want and display them or make them available for users.

For archivists and record managers our main job, is in my opinion, to get the relevant records for long term preservation via selection policies etc. from the organisations, authorities, et al. we work for. We do this for the benefit of public and for future publics for reasons of providing reliable, authentic evidence of transactions which we can verify by our activities (custody etc) - (the who, the why, and the when) and to provide the raw material for historical related research / education / entertainment etc.

In our public access role we are much the same but not in our archival role of selection disposal etc.

Q: Your career seems to have really transformed when you began taking on more management responsibility. How would you recommend that new professionals get experience in this area? Do you think the profession may suffer from a lack of management expertise?

A: Well it did, not that I was that keen on it I have to say. I was volunteered and couldn’t refuse to help out. You can only get experience of management doing it and this means maybe doing something else. It’s possible the earlier shift to managerialism of the 80s may shift back a bit, given how it has turned out in some areas, but without some managerial experience I do not see any great opportunity for progression except perhaps as consultants on specific issues requiring our skills.

For more information on Sarah’s past and current work visit http://www.sarah tyacke.co.uk/index.html
Linking Lives: The Story So Far

Jane Stevenson, Archives Hub Manager at The University of Manchester

Linked Data is an innovative and potentially revolutionary technology that enables diverse datasets to be seamlessly connected in a standard and Web-based way, so that a range of sources can be brought together to enable researchers to make new connections and discoveries. It is an emerging technology with huge potential, and currently many different sectors are investing resources into exploring what Linked Data can do for them and their audiences. For archives it has exciting possibilities, as essentially it is about enabling researchers to explore across diverse datasets, pulling together context and content, so it fits in very well with the idea of creating narratives through history.

The basic principles of Linked Data applied to archival catalogues results in the identification of people, organisations, places, subjects, events and other concepts as individual entities. This identification is done in such a way that machines can work with the data. It is done through giving all of these entities identifiers known as HTTP URIs (i.e. Web identifiers) and then creating semantic links between them. So, an individual can be linked to another individual, who is linked to a place, and the place is linked to an event, and the event links to another individual, and so on. The directions a researcher could take with the data become endless.

The Archives Hub has been at the forefront of creating Linked Data for archives, and this process has been documented in detail on the ‘Locah’ project blog (http://archiveshub.ac.uk/locah/). Our ‘Linking Lives’ project (http://archiveshub.ac.uk/linkinglives/) aims to create a front-end onto the Linked Data, based on bringing together information on individual people to create a biographical resource. The project has helped us to gain a real understanding of the challenges and the practical benefits of this technology.

Interface Design
We are working on a fairly simple interface where one page represents one person. We wanted something sustainable and extensible, where we could pull in a variety of external data types – text, images and links. Our interface uses the content boxes that are a familiar feature on many websites, and using these enables us to present different data sources as discrete parts of the interface, which is important if we want to be able to clearly identify the source of the data. (See Figure 1.)

The name displays at the top of the main display, and below this a box contains key information that comes from the archive descriptions: life dates, occupation or status, family name, and title. We decided to add place of birth and death as additional core information, provided by DBPedia, which is a Linked Data version of Wikipedia. We have a large box to contain the all-

Figure 1: Wireframe for the Linking Lives interface
important biographical notes for each individual that are typically created by archivists when they catalogue the material. We placed an image in the centre, as we felt this would make the interface more visually engaging. We intend to have a tab to list alternative names, which are provided by various sources, including VIAF (the Virtual International Authority File).

The Challenges of the Source Data
Working with aggregated archive descriptions that come from so many sources, created over a long period of time, and often migrated between different systems, is a real challenge. The data is inevitably inconsistent and there are errors that interfere with the data processing. Creating Linked Data really highlights any problems with the source data, because you need to identify all of the entities within the data in a way machines can process, which requires a high level of rigour. There are, broadly speaking, two alternative approaches to working with problematic data: 1) you can find ways round inconsistencies through the transformation process itself, or 2) you can address the problems at source by improving your data before you create the linked data. We have written about some of the issues with the data on our blog (http://archiveshub.ac.uk/linkinglives/).

Working with External Datasets
When working with data that comes from external sources you are unlikely to have any control over the data. You may have problems if it is inconsistent or if it changes. This is one of the major issues with linked data. The persistence of URIs has often been cited as an issue. (How often do you go and visit a Web page you have bookmarked, only to find it has disappeared?) Although this is certainly not a problem unique to the linked data approach, it does become particularly problematic when the aim is to present an interface that relies on data from external URIs.

We have had variable success with linking to different datasets and pulling in data. To do this you need relevant content and you need the right “hooks” to pull data into the interface. We found that a number of data sources providing linked data do not provide as much of it in this form as one might hope. Simply looking at the linked data website can be misleading; you have to dig into the data and see what you can link to (essentially, which entities have URIs you can use).

Problems of Identity
One of the biggest challenges around our linked data work has been identifying individuals; a particular focus for us because Linking Lives is based upon biographical content. The URIs used to identify persons in the Linked Archives Hub dataset have their origins in the names of persons occurring in the Archives Hub documents. Different forms of the name can legitimately be used to refer to the same person, and so you can end up with multiple URIs for one individual.

In addition to this, use of the name within the URI does not avoid any issues of ambiguity. You may have something like Mary Jones, b 1901 and M Jones, 1901-1980 in two different archive descriptions, both adhering to the same rules for name construction and referring to the same person. You may also have John Smith, b 1945, engineer in two different archive descriptions, which would create the same URI, but it may not be the same person. A human reader can distinguish that the...
name is the same but the person is not the same, because of the context provided, but if you end up with the same URI for two people, a machine will treat them as the same person.

A further problem is that names may change when death dates are added if the life dates are part of the URI. This means the subsequent re-transformation of the data will generate a different URI from that generated by the previous process using the initial form of the name.

One approach to these data processing problems would be to see the transformation stage as only the first part of a larger process, to keep track of the URIs generated over time, and build in a stage of processing to reconcile the URI generated from Scott, James, 1950–2012, Sir, biologist this week from that URI generated from Scott, James, 1950–, scientist in the previous version of the document six months ago. A further possible step we are considering is to match and replace the reconciled URIs in our data with a more authoritative version, such as from VIAF. This still leaves us with the problem of what to do with people in our data who do not exist in an authoritative version, of which there are many. There are other approaches that can be taken, and many examples of linked data initiatives working on this exact problem. Certainly the issues surrounding identification of persons are many and complex. Our Linking Lives project has helped us to understand the practical implications of using our linked data, but we are not yet in a position to say that we have found a sustainable and reliable way to identify individuals. This is not ideal when you are trying to make something work in a practical, cost–effective way.

Conclusions
Part of the motivation behind Linking Lives is to assess whether linked data really does provide an alternative way forward. We believe that we have created a valuable linked data source, and we have connected to external datasets using Linked Data principles. Linking Lives enables us to give archives a different context, putting them into a broader knowledge domain. Our hope is that it provides a useful case study for others who are undertaking similar projects.

We have continued to find linked data work challenging, partly due to the fact that it is a new and developing area, with few templates or tools to utilize, partly due to the challenges of working with various external data sources and partly because of issues within our own data. I would say that the biggest single factor in terms of additional work has been cleaning up our own data. The inconsistencies within data created by so many institutions over such a long period are compounded by the complex nature of hierarchical finding aids.

With big players like the Library of Congress committing more fully to linked data with their ‘Bibliographic Framework’ project, a certain level of optimism in the promise of linked data is clearly still in evidence, and the community is continuing to expand and evolve. There does also seem to be significant and increasing interest from the LODLAM community (Linked Open Data for Libraries, Archives and Museums, http://lodlam.net/). For archivists the skills required in this type of work may seem very unfamiliar and rather daunting. It requires a different approach, different thinking, and working closely with technical colleagues. Linked data can make us look at our catalogues very differently, and this may be no bad thing. The principle is to work with the advantages that the Web offers, and to move away from our notion of paper–based finding aids that are simply in electronic form towards the amazing promise of the Web for enabling unbounded exploration of content.

Jane Stevenson, Archives Hub Manager at The University of Manchester
This time last year, as I approached the end of my MA in Archives and Records Management at University College London, I was distracted from writing my dissertation – not just by the Olympics coverage, but also by the imminent prospect of finding work in my new profession. The different places we had visited on the course, and the institutions represented by visiting speakers who had come to talk to us, emphasised the vast range of workplaces where our new skills might be required, from small charities to multinational corporations. Where might I end up?

I sometimes wonder about the extent to which competition for vacancies intensifies as new graduates enter the job market each year. Like my classmates, I had started subjecting the JISC NRA archives mailing list to intense scrutiny for possible vacancies. As soon as the dissertation was submitted, completing online application forms (and in one quaint case, an actual letter) became a full time job. Anyone about to embark on a job search would be well advised to start thinking about examples from their work experience that demonstrate proficiency in the skills most commonly in demand: project management, team work, communication and time management.

I knew I wanted to stay in London, my home town, but that it was unlikely I would find a job that exactly matched my interest in twentieth century social history. I decided to keep an open mind and apply for any roles that looked interesting and for which I thought I could do a reasonable application, to gain experience. Besides, after a year of being a full-time student (not to mention all the time spent doing unpaid work as a voluntary Archives Assistant to gain the experience that is a prerequisite for getting a place on the course) the bank account was empty and I needed to start earning again.

I applied for twenty jobs, and was invited for three interviews. For two of them I was required to research and deliver a presentation on a subject on which I felt woefully ignorant, certainly compared to the people who would be on the other side of the desk asking the questions. But I would definitely say that rather than let this (admittedly daunting) prospect put you off, try to view it as an opportunity to talk about what you have learned, from the course and your experience, and what you are able to find out. I was lucky enough to be offered the second job for which I was interviewed, a fixed-term post as a Digital Records Officer at King’s College London.

Job descriptions are often so generic that it is hard to get a feel for what the role actually entails. Even though I asked at the interview about what a typical day would be like, I am not sure I was any the wiser. However I was pleased to be offered the job and determined to positively embrace the opportunity. Now I have been in post five months, I am very glad I did. The post combines two separate strands to make a full-time job.

I spend two days a week on a project funded by the JISC Transformations programme which supports universities in achieving organisational improvement by using information technology. The aim of the Transform project at King’s is specifically to build the business case for making changes to the way records are managed throughout their life-cycle, identifying the changes required to reduce dependence on hard copies and duplicated paper systems and move towards more digital methods of records capture, management and archiving. Most big organisations currently live with hybrid records systems, a situation that is unsustainable long-term because of the difficulties involved in managing information that exists partly on computerised systems and partly in paper form, resulting in duplication and gaps. Consulting people who create and use records, and analysing processes, has been a great way to meet people and gain insight into the way the university functions.

"Consulting people who create and use records, and analysing processes, has been a great way to meet people and gain insight into the way the university functions."
outcome to be a report that no-one will read, I am writing an online case study, making a 5 minute video about what we have learned, and blogging http://transform.jiscinvoke.org/wp/

“This whole work area was recently described as ‘looking after the data for the next generation’ which strikes me as a fundamentally archival undertaking.”

The other three days of my week are taken up with Research Data Management, a new and growing preoccupation for universities. There are many drivers: researchers are generating an ever-greater volume and variety of data, at the same time that policies of government and funders are dictating that publicly funded research data, produced in the public interest, should be made openly available – and all in the wider context of information legislation, and legal, ethical and commercial constraints. What’s this got to do with an archivist? Well, institutions need to capture research outputs - in the form of records that could be anything from interview transcripts to MRI brain scans - with metadata that makes it available and discoverable, and apply appropriate retention periods. Organising and managing the data involves appraisal and selection; curation to ensure results are fit to archive; and preservation to ensure the valuable assets remain fit for reuse – all of which require the skills and expertise of a trained archivist. This whole work area was recently described as “looking after the data for the next generation”, which strikes me as a fundamentally archival undertaking.

“Information management roles probably suit those who like managing their own time and workload, being proactive about working with other people at all levels, and getting creative about developing better ways of doing things.”

My work in this initial phase at King’s has involved consulting researchers to understand more about what sort of data they are generating, in what format, and how they are storing it. I have found out what other universities are doing and been to conferences and events organised by the research community, the Digital Curation Centre[1] and the Digital Preservation Coalition[2]. I have learned to advise new researchers about how to encrypt personal data, and store it safely, and delivered research data management training. We have formulated a draft policy and developed a business case for providing a new service, which will be considered by senior management in June.

And to think I finished the course thinking I faced a choice between a career in archives OR records management. It transpires that my naïve assumption was a false dichotomy. At King’s, the work of archives and records management professionals is very integrated, in the Archives and Information Management team. I have discovered that there is a whole world of information management out there, a field that brings together archival practice and records management expertise, but it doesn’t seem to be a world we hear much about. When I first looked at the job description I thought it looked suspiciously like something to do with IT, which shows how little I knew. I have since learned that information management professionals are the brokers, operating as objective specialists between the business and technical functions in organisations. I am very glad I wasn’t put off because I feel I am working at a high level and learning a lot from an inspiring manager. Information management roles probably suit those who like managing their own time and workload, being proactive about working with other people at all levels, and getting creative about developing better ways of doing things. These roles are also likely to be better paid than entry level archive roles, something that can be particularly important to the mid-life career changer with some work experience to offer. I will admit to having the occasional pang of longing to be handling more traditional archival material, but it is exciting to be developing skills in managing the data archives of the future. I was always aware that when it came to regular cataloguing roles I was competing with more experienced archivists and that, because so many organisations employ only a single archivist, traditional jobs can be hard to come by. I predict we will see more jobs in the world of managing information and research data being advertised in future, and it is one with great career potential for newly qualified archivists and records managers.

Veronica Howe, Digital Records Officer at King’s College London

References
The World Wide Web plays an undisputed important role in modern society, having transformed the way we communicate. Its impact has been felt in how we publish, learn, teach and research, and many other areas of human activity. Since the mid 1990s efforts have been developed around the world to address the transient nature of the web by archiving and preserving the historical web as a record of our social history. The Internet Archive’s Wayback Machine [1] is the earliest and most comprehensive web archive to date, containing over 240 billion web pages archived from 1996. A survey of web archiving initiatives conducted by the Portuguese Web Archive in 2010 identified 42 web archiving initiatives across 26 countries [2] worldwide. The survey concluded that collectively 6.6PB of web resources has been archived, with the involvement of many national libraries and archives, which traditionally have the duty to preserve a nation’s cultural and scientific heritage. Many are members of the International Internet Preservation Consortium (IIPC) [3]. A registry of the members’ web archives provides an overview of the collections’ geographical and temporal coverage [4].

The UK Web Archive [5], provided by the British Library and partners, currently contains over 14,000 selected UK websites archived since 2004. This is a relatively small archive because prior to the Non-Print Legal Deposit Regulations [6], which came into effect on 6 April 2013, web archiving required the permissions from rights holders, which necessitated a time-consuming and non-scalable administrative process. The National Archives also provide a web archive, which preserves UK government information published on the web [7]. Going forward, the UK Legal Deposit Libraries will be able to archive the web at a much larger scale, crawling the web periodically to preserve snapshots of the UK domain over time.

Social media is the collective name given to Internet-based or mobile applications which allow users to form online networks or communities based on common interest, social or ideological orientations. Such applications take many forms but their main purpose is to support interaction and communication among the members of a community, including the creation and exchange of user-generated content. Twitter and Facebook...
“Social media has become increasingly prevalent in people’s lives and also an important source for scholars to understand our time.”

are key examples of large social networking platforms, which aggregate many forms of media into one place and are used globally for business, research and personal communications. Social media has become increasingly prevalent in people’s lives and also an important source for scholars to understand our time.

There is a limited amount of social media content from Twitter, Facebook and YouTube in the UK Web Archive. It often forms part of a “special collection”, which is a group of websites brought together on a particular theme or an event, usually just archived for a fixed period of time. The decision not to systematically archive social media was related to the selective nature of the archive itself, and to resources constraints at the Library. However, both the scale and the complex technology used for systems such as Twitter and Facebook also pose significant technical challenges.

In conventional web archiving, web crawlers are used to capture snapshots of websites. It generally starts from a list of URLs (seeds), which are visited and downloaded. All the hyperlinks within the visited pages are extracted, and these page recursively visited and downloaded. This works fine when a resource is explicitly referenced in the HTML text using an URL. The crawler will be able to download a copy of the file via a simple HTTP request, by going to the right URLs. However the current web archiving technology is yet to catch up with the fast evolving web, which is less and less static and contains an increasing amount of dynamically generated content. Complex technologies are used to publish content and provide functionalities, especially in social media platforms where advanced and cutting-edge user experience is a key driver for user interface design. Barriers are sometimes put in place to prevent robots or crawlers to access all or parts of a website, making it more difficult for archiving crawlers to capture resources completely. Even the limited exemplar content within the UK Web Archive quite often required highly skilled technical staff to develop customised solutions outside our standard workflow.

There is also a whole raft of rights and ethical issues related both to archiving and to providing access to archived social media. For example, Intellectual Property Rights apply to both user-generated content and the social networking platforms that deliver it, which are not always easy to separate clearly. This opaqueness of ownership makes it very difficult for memory institutions to obtain clearance for archiving. In addition, the analysis of aggregated social media datasets may reveal unexpected or unintended patterns and connections, raising concerns about the risk of breaching personal privacy. The lack of transparency in many applications which collect and analyse social media data only deepens such concerns.

“Making use of social networks to help with archiving is surely an avenue worth further exploration.”

There is another strand of our work that relates to social media. The British Library received funding from the IIPC in 2011 and 2012 to develop a prototype application called Twittrvane [8]. This is not a tool for archiving Twitter but is instead an application which is capable of collecting and analysing Twitter feeds relevant to a given theme. Twittrvane looks for URLs in the Tweets and generates a list of the ones most frequently shared, which can then be presented to curators as potential titles for web archiving, saving time and effort required for manual selection. Curators from a number of national libraries evaluated the Twittrvane and found this a useful complementary tool, particularly for event-based collections.

“the current web archiving technology is yet to catch up with the fast evolving web”
While we are gathering wider feedback on Twittervane to further evaluate the approach, the ARCOMEM project, funded by the European Commission, is also developing a tool which makes use of social media to drive web harvesting [9]. Hopefully the set of common issues identified by both projects will find solutions as the projects progress. Making use of social networks to help with archiving is surely an avenue worth further exploration. Its potential has already been evidenced by a number of successful crowd sourcing projects in the MLA sector. A recent inspiring example is the Boston Marathon Archive project initiated by DHCommons, calling for collaborators to build a crowd-sourced archive of pictures, videos, stories and social media related to the Boston Marathon bombing. The strength of the project is that it goes beyond requesting help from the crowd with data entry, but also asks for programming effort in building plugins for an open-source publishing platform (Omeka) to support the incorporation of various types of data [10].

In comparison with some at-risk resources on the Web, my view is that public social media content, again especially on Twitter and Facebook, is currently archived reasonably well by a combination of researchers, companies and individuals using a wide range of commercial or open source tools. At an organisational level, Twitter has been donating its digital archive of public tweets to the Library of Congress since 2010 [11]. For individuals, the new Google Drive Extension allows web archiving and image storage [12] and Twitter has begun rollout of users’ tweet archive option for offline use [13].

Archiving social media is still a significant challenge for the web archiving community. It is clear that our current tools are not up to the challenge and a different approach may be required, taking advantage of the APIs provided by social networks which make data available to developers. Archivists, librarians, technologists, researchers, publishers, representatives of the government and industry have to work together to develop new models and tools appropriate to archiving social media content. As more and more scholars start using social media to study our world today, new models of collecting and donating data for research will emerge, overcoming some of the barriers currently preventing them from using such data.

References

Off the Record

will now continue at
ARANewProfessionals.wordpress.com

Keep up to date by following the blog, all new posts will be sent to your inbox

Get involved and write something for us

Take a look now and see some of the great posts we already have for you to read