

In this issue:

- [What's On](#) - Forthcoming events from December 2011 onwards
- [What's New](#) - New reports and initiatives since the last issue
- [What's What](#) - The Data Citation Community Gathers Pace - Monica Duke, DCC
- [Who's Who](#) - Sixty second interview with Sally McInnes, National Library of Wales
- [Featured Project](#) - The Irish General Election 2011: Web archiving at the National Library of Ireland
- [Your View?](#) - comments and views from readers

What's new is a joint publication of the DPC and [DCC](#)

---

## What's On

The DCC have a number of events coming up that may be of interest to you. For further details on any of these, please see our DCC events listings at <http://www.dcc.ac.uk/events/>. You can also browse through our DCC events calendar to see a more extensive list of both DCC and external events.

### **DevCSI Life Sciences Hackdays**

*6-7 December 2011*

<http://www.ukoln.ac.uk/events/devcsi/life-sciences-hackdays/index.html>

The event will bring together delegates from the SWAT4LS workshop and tutorials and researchers, developers and anyone else interested in the Life Sciences to work together in teams or individually to use and enhance existing Open Science semantic web applications and tools and possibly develop new ones.

### **7th IEEE International Conference on e-Science**

*5-8 December 2011*

<http://www.escience2011.org/>

Scientific research is increasingly carried out by communities of researchers that span disciplines, laboratories, organizations, and national boundaries. The e-Science 2011 conference is designed to bring together leading international and interdisciplinary research communities, developers, and users of e-Science applications and enabling IT technologies. The conference serves as a forum to present the results of the latest research and product/tool developments and to highlight related activities from around the world.

### **7th International Digital Curation Conference**

*5-7 December 2011*

<http://www.dcc.ac.uk/events/idcc11>

Digital curation manages, maintains, preserves, and adds value to digital information throughout its lifecycle, reducing threats to long-term value, mitigating the risk of digital obsolescence and enhancing usefulness for research and scholarship. IDCC brings together those who create information, those who curate and manage it, those who use it and those who research and teach about curation processes. This year the theme for the conference is "Public? Private? Personal? navigating the open data landscape".

### **JISC RSC seminar: Shared Services: Foundations for Successful Sharing**

*8 December 2011*

<http://www.rsc-southeast.ac.uk/events/2011/01-12-2011-shared-services-session-3.html>

This seminar will provide example tools and techniques so your shared service partnership can lay the foundations of successful shared working.

### **DCC Roadshow: Cardiff**

*14-16 December 2011*

<http://www.dcc.ac.uk/events/data-management-roadshows/dcc-roadshow-cardiff>

The seventh DCC Roadshow is being organised in conjunction with Cardiff University Libraries and Information Services. DCC Roadshows describe emerging trends and challenges associated with research data management and their potential impact on higher education institutions; examine case studies at both disciplinary and institutional levels, and explain the role of the DCC in supporting research data management.

### **JISC RSC seminar: Shared Services: Case Studies and Key Learning**

*14 December 2011*

<http://www.rsc-southeast.ac.uk/events/2011/01-12-2011-shared-services-session-3.html>

The final webinar will provide case studies of successful and unsuccessful sharing in FE, HE and beyond plus the key learning you can gain from 'those who have gone before'.

### **DigCCurr 2012 Public Symposium: CurateGear**

*6 January 2012*

<http://tinyurl.com/3m8ajrm>

CurateGear is an interactive day-long event focused on digital curation tools and methods. See demonstrations, hear about the latest developments, and discuss application in professional contexts.

### **Austin PASIG 2012**

*11-13 January 2012*

<http://sites.tdl.org/austinpasig/>

The PASIG is open to institutions and commercial organizations interested in learning about and sharing practical experiences in the following:

- Comparison of high-level OAIS architectures, services-oriented architecture work, and use cases
- Sharing of best practices and software code
- Cooperation on standard, open, 'in-a-box' solutions around repository technologies
- Review of storage architectures and trends and their relation to preservation and archiving architectures and eResearch data set management
- Discussion of the uses of commercial third party and community-developed solutions

The organization is focused on sharing open computing solutions and best practices. But while sharing information about the state-of-the-art developments in standards and open source is important, this is not a standards-setting organization. It is a place to share practical experiences, successes, pain points, and potential topics for more collaboration

### **Digital Preservation: What I Wish I Knew Before I Started**

*24 January 2012*

<http://www.dpconline.org/events/details/38-studentconference?xref=38>

The DPC and the Archives and Records Association are pleased to invite students and researchers in archives, records management and librarianship to a half day conference on practical workplace skills in digital preservation. Hosted by University College London, and organised in partnership with the University of Aberystwyth and the University of Dundee, this mini-conference will bring a select group of leading practitioners together with the next generation of archivists, records managers and librarians to discuss the challenges of digital collections management and digital preservation. In a lively set of presentations and discussions, each of the speakers will be invited to reflect on 'the things I wish I knew before I

started' - giving students an advantage in their own career development.

## **Trust in Post-Cancellation Access Services for E-Journals**

*31 January 2012*

<http://www.dpconline.org/events/details/39-trust?xref=39>

Tools and services have taken a long time to develop for the wider digital preservation community and most sectors report significant gaps in the infrastructure necessary to deliver lasting impact from highly prized and valuable digital resources. Of all the content types with which digital preservation has dealt in the last decade, the e-journal and e-books sector is perhaps the most advanced. Over and above the growing experience in fixing technical challenges, there is a well developed - if complicated and at times dysfunctional - value chain that connects authors, publishers, sellers, purchasers and consumers. A range of service providers and tools now aim to secure this supply chain with digital preservation tools.

Outsourcing - specifically knowing how to trust services that claim to provide digital preservation - has been one of the key barriers to preservation being adopted more widely so the experience of the E-Journal community is of much wider relevance than just the library and academic community. If the E-Journal market has genuinely solved the 'trust' question then everyone needs to know about it: and it has not then consideration of the issues will at least enable a more nuanced reflection on how we might want to develop trust in the preservation of other types of content.

## **Joint Conference on Digital Libraries 2012**

*10-14 June 2012*

<http://www./jcdl2012.info/>

The ACM/IEEE Joint Conference on Digital Libraries is a major international forum focusing on digital libraries and associated technical, practical, organizational, and social issues. JCDL encompasses the many meanings of the term "digital libraries", including (but not limited to) new forms of information institutions and organizations; operational information systems with all manner of digital content; new means of selecting, collecting, organizing, distributing, and accessing digital content; theoretical models of information media, including document genres and electronic publishing; and theory and practice of use of managed content in science and education. The call for papers is currently open and information on deadlines and suggested topics can be found using the above link.

## What's New

For more information on any of the items below, please visit the DCC website at <http://www.dcc.ac.uk>

### **OAIG Repositories Resource Pack**

<http://bit.ly/nq4Q5i>

The UK Open Access Implementation Group's (OAIG) repositories resource pack aims to help universities take immediate action to support wider access to UK research, by ensuring that as much of their research output as possible is made openly available via their institutional repository. The resource pack brings together all the information and guidance that UK universities might need in taking the policy decisions and practical steps for this to happen.

### **Success of open access presented in over thirty compelling stories**

[www.oastories.org](http://www.oastories.org)

Open access - the free, immediate, online access to the results of scholarly research - can transform scholarship and its impact. More than 30 compelling stories have been collected from across Europe and further afield to show this transformation in action. The stories are from over 11 countries and are told by a wide variety of stakeholders from individual researchers and journal editors to publishers and companies and cover a multitude of disciplines.

### **Data Centres: their use, value and impact**

<http://www.jisc.ac.uk/publications/generalpublications/2011/09/datacentres.aspx>

This report was co-sponsored by RIN and by JISC through the Managing Research Data Programme. The report suggests that researchers value data centres highly, and that they derive considerable benefits from the ready availability of curated datasets. However, it remains difficult to articulate this value, both upwards - to policymakers and funders who can provide funding and mandates to support data centres - and downwards - to the researchers themselves, many of whom remain reluctant to populate centres with their own data.

### **RIN study on Research Supervisors and Information Literacy**

<http://www.rin.ac.uk/our-work/researcher-development-and-skills/information-handling-training-researchers/research-superv>

This study investigates the role of PhD supervisors in the drive to ensure that research students possess the necessary level of information literacy to pursue their careers successfully in academia and beyond.

### **Career profiles: data management skills in various professions**

<http://www.dcc.ac.uk/training/data-management-courses-and-training/skills-frameworks>

<http://www.rin.ac.uk/our-work/researcher-development-and-skills/data-management-and-information-literacy>

As part of its work on the JISC/RIN funded Data Management Skills Support Initiative (DaMSSI), the DCC and RIN have produced a series of career profiles that aim to demonstrate how data management skills contribute to and underpin high-quality performance in a number of professions.

### **LoC Digital Preservation newsletter**

<http://www.digitalpreservation.gov/news/newsletter/201111.pdf>

The November 2011 issue of the Library of Congress Digital Preservation Newsletter is now available.

### **Manchester and Elsevier team up on text-mining tool**

<http://www.manchester.ac.uk/aboutus/news/display/?id=7627>

The University of Manchester has joined forces with Elsevier, a leading provider of scientific, technical and medical information products and services, to develop new applications for text mining, a crucial research tool. The primary goal of text mining is to extract new information such as named entities, relations hidden in text and to enable scientists to systematically and efficiently discover, collect, interpret and curate knowledge required for research.

### **New Head of JISC appointed**

<http://www.jisc.ac.uk/news/stories/2011/11/newheadofjiscappointed.aspx>

Martyn Harrow, Director of Information Services at Cardiff University, has been appointed as Head of JISC for a fixed term of 9-18 months from 1 February 2012. Martyn will succeed Dr Malcolm Read who retires as Head of JISC in January 2012 after 18 years in post. Martyn will see the organisation through its transition into a 'new look' JISC, following the recommendations of the Wilson Review (February 2011).

### **Knowledge Exchange Report: A Surfboard for Riding the Wave**

<http://www.knowledge-exchange.info/Default.aspx?ID=469>

Following the report "Riding the Wave: How Europe can gain from the rising tide of scientific data" that was released 2010 by the high level expert group on research data, the Knowledge Exchange (KE) partners have embraced this vision and commissioned a report that translates Riding the Wave into actions for the four partner countries and beyond. The Report "builds on

the 2010 report and presents an overview of the present situation with regard to research data in Denmark, Germany, the Netherlands and the United Kingdom and offers broad outlines for a possible action programme for the four countries in realising the envisaged collaborative data infrastructure.

### **SageCite publisher interviews**

<http://blogs.ukoln.ac.uk/sagecite/publisher-interviews/>

The SageCite project, funded from August 2010 by JISC under the Managing Research Data CLIP Strand, is releasing interviews with the editors of two leading journals in the Biosciences. The two interviews explore a large range of issues concerning data, scholarly communications and publishers, the links between data and publications and interoperability between data repositories and publishers.

### **D-Lib Magazine**

<http://www.dlib.org/>

This issue contains four articles, three conference reports, several short pieces in the 'In Brief' column, excerpts from recent press releases, and news of upcoming conferences and other items of interest in 'Clips and Pointers'. This month, D-Lib features the Digital Memory of Catalonia collection.

### **UKeiG Announces winner of its second award - the UKeiG Jason Farradane Award**

<http://www.uk eig.org.uk/awards/jason-farradane#2011Winner>

The 2011 UKeiG Jason Farradane Award has been awarded to the United Kingdom Council of Research Repositories (UKCoRR). Founded in 2007, UKCoRR is a professional membership-driven organisation managed for and by those staff working throughout the UK as Open Access repository administrators and managers. UKCoRR facilitates communication between the membership and fellow information providers, LIS professionals, the research community and scholarly publishing stakeholders by providing a collective voice that can speak on members' behalf to publishers, funding councils, institutions, and other relevant community stakeholders.

### **JISC Inform**

<http://www.jisc.ac.uk/inform/inform32/contents.html>

This issue of JISC Inform features an interview with the DCC's Director, Kevin Ashley, who talks about the importance of research data management. The issue also includes an interactive quiz to help you assess your institution's research data management activity and infrastructure.



## **JISC and UK Research Councils to build a robust repository infrastructure for the future**

[http://www.jisc.ac.uk/whatwedo/programmes/di\\_researchmanagement/repositories/rioextension.aspx](http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/repositories/rioextension.aspx)

Tracking the UK's research outputs will become easier in the future thanks to JISC and Research Councils UK (RCUK) working together to utilise their expertise. Over the coming months a piece of work called the RIO Extension project will take place to scope the issues and requirements from universities, funders and researchers in managing the information about research outputs. The aim of the work is to provide the UK education and research sector with clear, practical guidance on recording and sharing information about its research outputs, so that it can be reused for a variety of purposes, including by the systems used by the Research Councils.

---



## **What's What - Editorial: The Data Citation Community Gathers Pace**



**Monica Duke, DCC**

I first became involved in researching and thinking about data citation in early 2010, when Liz Lyon, director of UKOLN and Associate Director of the Digital Curation Centre (DCC) asked me to join a working group to consider data citation in the context of Sage Bionetworks [1]. Sage Bionetworks are behind an effort to promote and support the sharing of data and analysis required to help advance the understanding of disease and treatments [2]. They recognised early on that data citation would be one of the building blocks that would contribute to shaping both the culture and infrastructure around data sharing. Hence their working group on data citation which researched data citation and discussed it during the first Sage Bionetworks Congress [3]. I did not know then (but should have guessed) that Liz's invitation would be the start of nearly two exciting years of working in this field, leading up to a time when a community seems to have formed around the topic, and understanding of data citation and its needs appears to be reaching a significant point. Shortly after the Congress I became part of the SageCite team [4] funded through the JISC MRD programme strand on Citing, Linking, Integrating and Publishing (CLIP) data. [5] Through SageCite I was fortunate to attend or follow some great events.

In August the National Academy of Sciences, through the Board on Research Data and Information and the CODATA-ICSTI Task Group on Data Citation Standards and Practices, organised an international symposium and workshop on Developing Data Attribution and Citation Practices and Standards. [6] For anyone new to the topic, I would most like to recommend the presentation by Christine Borgman, who was tasked with parsing Data Citation. Her thoughts comprehensively dissected the current status and questions on data citation. Borgman first of all characterised three aspects of the data citation landscape. Firstly, the data deluge, with its appraisal and selection issues - Borgman rightly recognised how easily discussions on citation and attribution can stray into more general discussions on data management; secondly, the different types of data under discussion (experimental, observational, image, surveys, computational...), acknowledging that deciding what data should be shared, reused or are worth citing or making citable (or even deciding what IS data) is a whole discussion in itself; thirdly, the Infrastructure with its many dimensions along different axes (quoting Star and Ruhleder), ranging from technical to social, global to local, and combinations in between.

With these reflections as a basis, the remainder of Borgman's talk presented data citation questions within nine areas: Social Practice, Usability, Identity, Persistence, Discoverability, Provenance, Relationships, Intellectual Property and Policy. It was particularly satisfying to see discoverability and usability, two oft-overlooked characteristics, making the top list. Discoverability is a topic I have some experience in through previous projects with a focus on

discovery services and metadata. Usability was a core research area at the University of York where I first formally began studies in Information Processing. This opening talk, with accompanying notes, is available from the event agenda web page and a detailed report from the meeting including all other presentations and discussions should also be available in the Spring.

This symposium followed close on the heels of the Future of Research Communication meeting in Dagstuhl [7]. This event was geared towards the wider question of scholarly communication, but as the white paper [8] illustrates, linking data (and related research objects like software) to publication workflows, and deriving new methods and metrics for evaluating quality and impact - both topics that are relevant to data citation - are preoccupations very much at the forefront of this wider discussion. A report on the talks to accompany the white paper will also follow - useful dissemination for those (like me) who were not attendees at the event.

It is no co-incidence that several of the key people active in these events were also present at Beyond The PDF in San Diego in January 2011 [9]. With a focus on moving publications on to the next generation by making the underlying data more accessible and interactive, this event addressed the theme of the future of scholarly research communication, once again touching on issues of how to link and cite data, how to motivate data sharing and attribution and citation. Philip E. Bourne, was one of the key people behind the Beyond The PDF meeting, and is also a co-editor of [7]. Phil will be giving one of the keynotes at IDCC in December 2011 [10]. The SageCite project has very recently published an interview with him [11]. Whilst on the topic of the SageCite interviews, the second of the two interviews was with Myles Axton [12], who as editor of Nature Genetics earlier this year edited the issue which published the landmark paper that featured Microattribution as a method to assign credit in a scientific study [13,14].

Three technical infrastructure themes can be identified in all the above discussions:

1. attribution, credit and incentives: what is the best format for data (and other research object) citations so that they can be used to (a) appropriately attribute contributions (b) support tools that are needed so that attribution can be computed into credit and new measures of impact?

2. making data re-usable and interactive: how can citations for research objects (other than publications) support the discovery and re-use of data? What are the models for integrating them with the narrative of science to take that narrative to the next level, and what infrastructures should data citations interoperate with?

3. improving the validation and the record of scholarly communication: how do citations of research objects fit in with current peer review and publishing processes, and could they enable or accelerate alternative models of review and validation?

All of these technical issues can be unpacked further, and must be considered alongside the social questions such as those addressed by Borgman, with consideration given to cultural change issues.

Three other developments to mention before closing are (1) the completion of the first round of the afore-mentioned JISC MRD CLIP strand, which delivered among others outputs from DRYAD UK, ACRID and FISH.link and (2) the Data Citation Principles meeting in Harvard in May [15] which should also be reporting soon. Last but not least, the DCC has published a briefing paper and guide [16, 17]. The drafts of these publications were opened to community review, and the level of commentary and interest was outstanding (thank you again to any reviewers out there!), making the final result all the better for it, but also confirming the timeliness of the guides and the engagement of the community with the topic.

My next project after SageCite is in collaboration with Microsoft [18] and is based on Jim Gray's idea of the Fourth Paradigm [19]. Interestingly, back in 2007, Gray devoted the second part of his last talk, as reported in the book, to scholarly communication. He outlined a vision in which data and literature 'interoperate with each other', and highlighted the need to foster digital data libraries. With all the data and literature for science available online, the 'fourth paradigm' of data-intensive science would be accelerated. All of the above-mentioned activities are good steps taken on the path to achieving this vision.

After these two years of activity it seems to me that the key questions about data citation have now been articulated [as exemplified by Borgman's talk, 20 and 7], some challenges have been identified [8, 21], and the community is well on its way to having a defined identity, sustained by successful meetings and reaching critical mass in its influence. Does this mean job done? Of course not. As Whyte reflects [22] when reporting from Beyond Impact [23] (incidentally, another in this years' related initiatives worth following), it is good to come out of these meetings with identified actionable items. Things are happening, but where will we go next? This seems like a good time to stop and ask, what should be the next priorities for the data citation community? How can we best channel the palpable energy, enthusiasm and commitment into concrete actions that will make a difference? And how can funders and services best support the process?

[1] Sage Bionetworks <http://sagebase.org/>

For an introduction to Sage Bionetworks see also: <http://blogs.ukoln.ac.uk/sagecite/the-application-domain/>

- [2] <http://www.sagebase.org/commons/repository.php>
- [3] Citation report at the First Sage Bionetworks Congress, 2010 [http://sagecongress.org/Presentations/E\\_Citation.pdf](http://sagecongress.org/Presentations/E_Citation.pdf)
- [4] SageCite project <http://blogs.ukoln.ac.uk/sagecite/>
- [5] The JISC MRD programme CLIP Strand <http://www.jisc.ac.uk/whatwedo/programmes/%7E/ink.aspx?id=150B7D18009A4336B74E2BA69D421CAE&z=z>
- [6] The National Academy of Sciences International symposium and workshop on Developing Data Attribution and Citation Practices and Standards. Berkeley, August 2011 [http://sites.nationalacademies.org/PGA/brdi/PGA\\_064019](http://sites.nationalacademies.org/PGA/brdi/PGA_064019)
- [7] Future of Research Communications meeting, Dagstuhl, August 2011 <http://www.dagstuhl.de/en/program/calendar/semhp/?seminr=11331>
- [8] Bourne, P. et al. Force11 White Paper: The Future of Research Communications and e-Scholarship (October 2011) <http://force11.org/sites/default/files/attachments/Force11Manifesto20111028.pdf>
- [9] Beyond The PDF meeting, San Diego, January 2011 <http://sites.google.com/site/beyondthepdf/>
- [10] International Digital Curation Conference, Bristol, UK 5-7 December 2011 <http://www.dcc.ac.uk/events/idcc11>
- [11] SageCite interview with Philip E. Bourne <http://blogs.ukoln.ac.uk/sagecite/publisher-interviews/interview-with-philip-e-bourne-editor-plos-computational-biology/>
- [12] SageCite interview with Myles Axton <http://blogs.ukoln.ac.uk/sagecite/publisher-interviews/interview-with-myles-axton-editor-nature-genetics/>
- [13] Giardine, B., Borg, J., Higgs, D.R., Peterson, K.R., Philipson, S., Maglott, D., ... Patrinos, G.P. (2011). 'Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach'. Nature Genetics, 43, 295-301. doi:10.1038/ng.785 <http://www.nature.com/ng/journal/v43/n4/full/ng.785.html>
- [14] Axton, M. (2011) 'Crowdsourcing human mutation'. Nature Genetics, 43, 279 <http://www.nature.com/ng/journal/v43/n4/full/ng0411-279.html>  
editorial on microattribution
- [15] Data Citation Principles meeting, Harvard, May 2011 [http://projects.iq.harvard.edu/datacitation\\_workshop/](http://projects.iq.harvard.edu/datacitation_workshop/)
- [16] Ball, A., Duke, M. (2011). 'Data Citation and Linking'. DCC Briefing Papers. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking>
- [17] Ball, A. & Duke, M. (2011). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>
- [18] <http://communitymodel.sharepoint.com/>
- [19] The Fourth Paradigm. (2009) Eds. Tony Hey, Stewart Tansley and Kristin Tolle. Publ Microsoft Research. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>
- [20] [http://projects.iq.harvard.edu/datacitation\\_workshop/pages/discussion-questions-and-recommended-readings](http://projects.iq.harvard.edu/datacitation_workshop/pages/discussion-questions-and-recommended-readings)
- [21] [https://docs.google.com/document/d/1sH3JOW5Luki4i37Ve1mOnI2wNZJbaUOx1T42S\\_7txQ0/edit?hl=en\\_GB](https://docs.google.com/document/d/1sH3JOW5Luki4i37Ve1mOnI2wNZJbaUOx1T42S_7txQ0/edit?hl=en_GB)
- [22] Whyte, A. 'Beyond Impact and the Catch-22 in Data Reuse' DCC Blog, May 2011 <http://www.dcc.ac.uk/news/beyond-impact-and-catch-22-data-reuse>

[23] [http://beyond-impact.org/?page\\_id=64](http://beyond-impact.org/?page_id=64)

---



## **Who's Who: Sixty Second Interview with Sally McInnes, Head of Collections Care, National Library of Wales**

### **Where do you work and what's your job title?**

I am Head of the Collections Care Section, which is situated in the Collections Services Department, at the National Library of Wales, Aberystwyth.

### **Tell us a bit about your organisation**

The NLW, founded in 1907, is one of the oldest and most important of the national institutions of Wales. As a legal deposit library it has an extensive collection of published material, but it also holds manuscripts and archives, maps, pictures, photographs, sound and moving images and an increasing volume of digitised and born digital material. We welcome over eighty thousand visitors through our doors every year and have over two million annual visits to our website.

### **What projects are you working on at the moment?**

As Head of Collections Care, I am responsible for the preservation of all items in the NLW's collections, regardless of format. Preservation is a core function of the NLW and is essential to maintain the usability of the collections. At the moment, I am involved in drafting the preservation policies for the NLW's collections and ensuring that they align with the NLW's

priorities of cost-effective operations, agility and focus upon the user. I am also involved with our 'Space Development Programme' (nothing to do with NASA!) which is an ambitious programme to provide additional storage capacity for our growing digital and analogue collections.

**How did you end up in digital preservation?**

I joined the NLW well before we had computers on every desk, let alone mobile devices! My first job at the NLW was cataloguing medieval Welsh deeds on card catalogues. With the increase in digital content being accessioned by, and created within, the NLW, I became interested in the challenges involved with sustaining access to digital content and I became Digital Preservation Co-ordinator in 2009. When I became Head of Collections Care, I kept the strategic responsibility for digital preservation. I work very closely with our Digital Preservation Team, which comprises three members of staff in the Systems Section of the NLW.

**What are the challenges of digital preservation for an organisation such as yours?**

One of the biggest challenges is embedding digital preservation into Library workflows and business activities. Digital preservation awareness must be built into business areas, including acquisition, data management, ICT and Systems and digitisation. We are engaging with this by developing a workflow for born digital material and by implementing a Sustainable Access Plan which identifies the key areas, tasks and responsibilities and monitoring procedures. Capacity is another challenge, particularly when resources are restricted.

**What sort of partnerships would you like to develop?**

I am a member of the Digital Preservation Group of the Archives and Records Council Wales (ARCW). We have just obtained CyMAL funding to employ a research officer for six months to work in partnership with consortium members to increase capacity for digital preservation across Wales.

**If we could invent one tool or service that would help you, what would it be?**

A one click, easy to use, ingest tool!

**And if you could give people one piece of advice about digital preservation ....?**

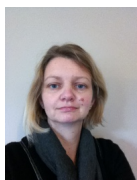
Create a glossary to ensure that all stakeholders agree on the definition of tools and concepts

**If you could save for perpetuity just one digital file, what would it be?**

I would like to keep the digital file that we have just created at the NLW from the digitisation of the Book of Aneurin. Llyfr Aneurin dates from about 1265 and is one of the earliest manuscripts written in the Welsh language. After it was digitised, staff in the Conservation Unit at the NLW created a binding for it which is of such high quality that it is scarcely indistinguishable from the original. This facsimile copy will then be used to extend access to the original through exhibition and outreach. The facsimile is a testament to the marriage between the new skills of digitisation and the traditional skills of the book binder. The original has survived over 800 years and the digital preservation challenge is to ensure that we can access the digital file in the year 3000!

### **Finally, where can we contact you or find out about your work?**

To contact me, please email [slm@llgc.org.uk](mailto:slm@llgc.org.uk) . To find out more about the National Library of Wales, please visit the web page at: <http://www.llgc.org.uk/>



### **Featured Project: The Irish General Election 2011 - Web archiving at the National Library of Ireland**

***Della Murphy, Assistant Keeper and Born-Digital Programme Manager, National Library of Ireland***

The mission of the National Library of Ireland (NLI) is to collect, preserve, promote and make accessible the documentary and intellectual record of the life of Ireland. The Library has a long history of collecting material generated during general election campaigns, including unique holdings of posters, flyers and other ephemeral items which would otherwise not be preserved and available to researchers.



In relation to the Irish General Election of 2011 (GE11), while continuing to collect material on paper such as posters, flyers and leaflets, we also conducted a focused web crawl. This was our first initiative into selective web archiving and it was felt that the importance of this election from an historical and technological point of view needed to be captured. With digital and social media used extensively throughout the campaign, this presented the NLI with a unique and exciting opportunity to record history.

100 sites were identified for inclusion, with pre- and post-election snapshots of these sites taken as part of the project. The web crawling activities of this project were carried out by the Internet Memory Foundation (IM), a non-profit organisation established to actively support the preservation of the Internet as a new media for heritage and cultural purpose. See: <http://internetmemory.org/en/>

## Site Selection

We had general selection criteria for all sites such as, for example, levels of engagement and ease of use from a technical point of view. We then narrowed our focus to more specific categories of website and had different selection criteria depending on the category of site in question. For political candidates sites, for example, apart from a few automatic selections (in this instance the outgoing ministers), we favoured those candidates tipped to succeed as we were keen to include all in both the pre- and post-election snapshots. We also wanted to ensure a good cross-party and geographical spread of candidates.

For the political commentary websites we engaged in a consultation process with academics with an expertise in this area, as well as NLI staff, and the blog awards sites. We also wanted a full range of coverage from the more academic to the less formal, and we wanted to include the view from outside the country. Very important from our point of view was the inclusion of sites that specifically complemented our existing hardcopy collections, whether visual, ephemera or newspapers. This element of continuity and complementarity between our traditional collections and our new digital collections is very important for us as we travel down the road of building digital collections.

For a more complete discussion on our selection criteria please consult <http://www.nli.ie/en/general-election-2011-web-archive.aspx>

## **Access to the Collection**

With ease of use a key objective, provision of full text search to the collection is very important to us. There are a number of ways to access the collection. From the IM's website, users can browse by URL, or carry out keyword searches across the full-text of the collection. Additionally, our Information Systems team has developed a widget based on the OpenSearch protocol which displays results from the web-archive collection directly in our catalogue alongside our other resources. We intend to build on this work in the future to make the collection even more accessible through our own interfaces. Our blog post offers a very useful guide to the collection and how to use it and can be found at <http://www.nli.ie/ge11-webarchive>

## **Potential uses**

The collection is digital in content, and is therefore accessible to anyone with a computer. It will have a broad appeal to anyone with an interest in politics. Users of the collection can compare online content before and after the election and also use the information as a record of campaign information. Students of various disciplines will be able to use the collection to analyse how 'online' the collection was, as well as levels of online interaction with the electorate, and whether or not it made a difference to how people voted. The impact of social media and web 2.0 technologies can also be examined. The NLI hopes to build on this foundation born digital collection to make more of its material available online, and is at present working again with the Internet Memory Foundation on the archiving of websites relating to our recent Presidential Election.

{jcomments on}