

In this issue:

- [What's On](#) - Forthcoming events from March 2012 onwards
- [What's New](#) - New reports and initiatives since the last issue
- [What's What](#) - In the Beginning Was the Word, William Kilbride, DPC
- [Who's Who](#) - Sixty second interview with Patricia Sleeman and Ed Pinsent, ULCC
- [Featured Project](#) - SPRUCE, Bo Middleton, Leeds University Library
- [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.

### **DCC Salford Data Management Roadshow**

*20-21 March 2012*

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

The 9th DCC regional roadshow will take place on 20th and 21st March at the University of Salford. The event will be opened by Professor Martin Hall, Vice Chancellor of the University of Salford and Chair of the UK Open Access Implementation Group.

### **Software Sustainability Institute Collaborations Workshop**

*21-22 March 2012*

<http://www.software.ac.uk/cw12>

The Collaborations Workshop gets researchers and software developers working together to solve research problems. If you're a researcher who wants to make more of software, or a

developer who wants to work with researchers, the workshop is the perfect opportunity to meet new collaborators.

### **European Grid Infrastructure (EGI) Community Forum 2012**

*26-30 March 2012*

<http://cf2012.egi.eu/>

The first Community Forum will showcase the role that EGI plays in enabling innovation across the European Research Area. The forum will highlight the services, technologies and tools available to scientific communities to better support their research.

### **Archaeological Research Data Management Day at CAA2012**

*28 March 2012*

<http://www.southampton.ac.uk/caa2012/>

The JISC-funded DataPool project will be delivering an archaeological research data management day as part of the Computer Applications and Quantitative Methods in Archaeology (CAA) 2012. The conference is being hosted by the Archaeological Computing Research Group in the Faculty of Humanities at the University of Southampton from 26-30 March 2012.

### **ISKO UK and the BCS Location Information Specialist Group (LISG)**

*29 March 2012*

[http://www.iskouk.org/events/location\\_march2012.htm](http://www.iskouk.org/events/location_march2012.htm)

In this ISKO UK and BCS joint meeting, we will hear from experts about the current geospatial information landscape and its challenges, some of the standards and frameworks that have been put into place to ensure interoperability and the potential for linking data. We will also hear how some users of GIS systems have applied them in their own organizations.

### **RDMF8: Engaging with the Publishers**

*29-30 March 2012*

<http://www.dcc.ac.uk/events/research-data-management-forum-rdmf/rdmf8-engaging-publisher>  
[S](#)

The DCC's Research Data Management Forum (RDMF) events bring together researchers, digital repository managers, staff from library, information and research organisations, data curators, data centre managers, data scientists, research funding organisations and research networks. The events are organised thematically, with a mixture of presentations and breakout and discussion sessions. RDMF8 will be hosted by the University of Southampton. The programme includes speakers from Nature, Wiley, Elsevier, Dryad, the International Union of Crystallography and Faculty of 1000.

### **SPRUCE Digital Preservation 'Mashup'**

*16-18 April*

<http://www.dpconline.org/events/details/40-SpruceGlasgow?xref=41>

SPRUCE (Sustainable PReservation Using Community Engagement) has been funded by JISC to inspire, guide, support and enable HE, FE and cultural institutions to address digital preservation gaps; and to use the knowledge gathered from that activity to articulate a compelling business case for digital preservation. This event brings together a diverse community to discuss, test, code, plan, and share challenges related to the new types of content entrusted to libraries, archives, and museums to preserve and manage. The focus is around community, communication, and learning from one another for we definitely can't go it alone in the new landscape of digital content. The result will be practical digital preservation tools which meet your specific needs and which are likely to be useful more widely.

### **DCC Roadshow Northeast England**

*23-24 April 2012*

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

The 10th DCC regional roadshow will take place at the Life Centre, Newcastle upon Tyne. The roadshow is being co-organised with Professor Julie Mcleod, Northumbria University School of Computing, Engineering, and Information Sciences.

### **ALPSP Seminar: Data publishing**

*24 April 2012*

<http://www.alpsp.org>

The digital world has vastly increased options for what we do with data. Data has emerged from dog-eared notebooks and spawned research institutes, journals, businesses, and controversy. This one day seminar will look at opportunities and challenges in data publishing from the points of view of policy makers, publishers, data generators, and consumers. It will examine everything from access and deposition policies to curation and deposition.

### **10th International Bielefeld Conference 2012**

*24 - 26 April 2012*

<http://conference.ub.uni-bielefeld.de>

The Bielefeld Conference 2012 will provide ideas to renew structures of documents, data, services and organisations. The conference is the 10th in a very successful series of conferences organized by Bielefeld University Library at the Bielefeld Convention Center since 1992. The conferences provide an essential forum for internationally renowned and trendsetting speakers and have gained high reputation among library directors and other senior library and information managers, who wish to discuss future strategies for academic libraries.

## **2nd LIBER international workshop: Partnerships in curating European digital resources**

*7-8 May 2012*

<http://www.rinascimento-digitale.it/liber2012-internationalworkshop.phtml>

This workshop will provide an overview of the best-known collaborative initiatives: the stakeholders involved, the basic set-ups, the legal foundations, the business models – and help you analyse which alternatives are best suited for your organisation, your type of collection and your national culture. We will deal with organisational issues, legal issues, financial issues and technical issues that will influence your choices. Critical questions will be asked by experts in the field, and there will be plenty of time to ask your own questions.

## **DPC Planning Day 2012**

*18 May 2012*

This meeting will focus on the 'Assurance and Practice' elements of the new DPC strategic plan. Full members invited on the evening of the 17th May, Associates and Personal members on the 18th

## **Digital Preservation and Business Continuity Management**

*21 May 2012*

This DPC Briefing Day will examine the intersection of practices associated with business continuity management and digital preservation to explore how these two communities might be more closely aligned, London

## **Digital Preservation Training Programme (DPTP)**

*28-30 May 2012*

[http://bit.ly/DPTP\\_MAY2012](http://bit.ly/DPTP_MAY2012)

The DPTP is a modular training programme, built around themed sessions that have been developed to assist you in designing and implementing an approach to preservation that will work for your institution. Through a wide range of modules, the DPTP examines the need for policies, planning, strategies, standards and procedures in digital preservation, and teaches some of the most up-to-date methods, tools and concepts in the area.

## **DPC Director's Group Meeting**

*28 June 2012*

The Directors' Group provides an extended and informal networking opportunity at which staff, partners, contractors or allies of full members of the Coalition are invited to describe and discuss current, forthcoming and future digital preservation projects. It allows staff, colleagues

and supporters - who might not normally attend Board meetings - to contribute to the Coalition's work plan for the coming year. It encourages the development of bilateral and multi-lateral relationships among members; helps disseminate good practice; and ensures that the work of the coalition remains tied to the changing needs of the workforce.

Full members are invited to nominate up to three delegates. Delegates can be drawn from any department, project, partnership or constituent of the Board Member's institution so long as they are able to contribute to and benefit from an open discussion on digital preservation and cognate issues. Delegates will be expected to present a brief and discursive summary of current and future work.

### **Robust Linked Data**

*29 June 2012*

DPC briefing day on the challenges of ensuring linked data is robust, Cambridge. Details to follow.

---

## **What's New**

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

### **New DCC Guide: 5 Steps to Research Data Readiness**

[http://www.dcc.ac.uk/webfm\\_send/777](http://www.dcc.ac.uk/webfm_send/777)

This introductory guide to the tools and services available through the DCC will allow IT managers to begin identifying the support they need, not only from the DCC but also from other stakeholders at their institution. By working through the 5 steps IT managers will be able to establish the strengths and weaknesses of the services they currently offer and work towards offering the Research Data Management service increasingly demanded by institutions and researchers.

### **Preserving Email: A new technology watch report from the DPC**

<http://dx.doi.org/10.7207/twr11-01>

Email is a defining feature of our age and a critical element in all manner of transactions. Industry and commerce depend upon email; families and friendships are sustained by it; government and economies rely upon it; communities are created and strengthened by it. Voluminous, pervasive and proliferating, email fills our days like no other technology. Complex, intangible and essential, email manifests important personal and professional exchanges. The jewels are sometimes hidden in massive volumes of ephemera, and even greater volumes of trash. But it is hard to remember how we functioned before the widespread adoption of email in public and private life.

Institutions, organizations and individuals have a considerable investment in - and legal requirements to safeguard - large collections of email. IT managers and archivists have long recognised that email requires careful management if it is to be available in the long term but practical advice about how to do this is surprisingly sparse. So a new 'Technology Watch Report' from the Digital Preservation Coalition (DPC) will be of wide interest.

### **Preserving Motion Pictures and Sound: Member preview of a new technology watch report from the DPC**

[http://www.dpconline.org/component/docman/doc\\_download/742-preservingdigitalsoundandvisionpreviewtwrmarch2012](http://www.dpconline.org/component/docman/doc_download/742-preservingdigitalsoundandvisionpreviewtwrmarch2012)

(Members only, login required. To register for a login, see <http://www.dpconline.org/login/registers>)

The report discusses issues of moving digital content from carriers (such as CD and DVD, digital videotape, DAT and minidisc) into files. This digital to digital 'ripping' of content is an area of digital preservation unique to the audio-visual world, and has unsolved problems of control of errors in the ripping and transfer process. It goes on to consider digital preservation of the content within the files that result from digitization or ripping, and the files that are born digital. While much of this preservation has problems and solutions in common with other content, there is a specific problem of preserving the quality of the digitized signal that is again unique to audio-visual content. Managing quality through cycles of 'lossy' encoding, decoding and reformatting is one major digital preservation challenge for audio-visual as are issues of managing embedded metadata. The report will be made more widely available in due course.

### **Draft of semantic interoperability standard is now online for public comment**

<http://drafts.bsigroup.com/>

Part 2 of the international standard ISO 25964 *Thesauri and interoperability with other vocabularies* builds on the foundation

established in Part 1  
*information retrieval*

*Thesauri for*

, which was published in 2011. It is also well aligned with the mapping properties of the W3C standard SKOS. The draft is now available for public consultation until the end of April 2012.

### **APARSEN Exchanges**

<http://aparsenexchanges.net/>

A new micro-site has been launched from the APARSEN project providing details of the APARSEN Staff and Experience Exchange. It provides a register of exchanges and the research currently in progress in this EU-Funded Network of Excellence.

### **New international standards to aid data sharing**

<http://www.nature.com/ng/journal/v44/n2/full/ng.1054.html>

Led by researchers at University of Oxford (UK) and the Harvard Stem Cell Institute (HSCI) at Harvard University, (USA), more than 50 collaborators at over 30 scientific organizations around the globe have agreed on a common standard for integrating biological data sets. This will make it possible to consistently describe the enormous and radically different databases that are compiled in the biosciences in fields ranging from genetics to stem cell science, to environmental studies.

### **Demystifying the data interview**

<http://www.emeraldinsight.com/journals.htm?articleid=17015015&show=abstract>

As libraries become more involved in curating research data, reference librarians will need to be trained in conducting data interviews with researchers to better understand their data and associated needs. This article seeks to identify and provide definitions for the basic terms and concepts of data curation for librarians to properly frame and carry out a data interview using the Data Curation Profiles (DCP) Toolkit.

### **The University of York join DPC**

The Information Directorate of the University of York are the newest member of the Coalition. 'Accessing and preserving digital information is one of the great challenges of the 21st century,' explained Chris Webb of the University of York. 'We recognise the importance and scale of the challenge, and we're pleased to join the DPC, which is a key partnership that enables these difficult areas to be tackled for the benefit of all.'

### **LoC Digital Preservation Newsletter**

<http://www.digitalpreservation.gov/news/newsletter/201202.pdf>

The February 2012 Library of Congress Digital Preservation Newsletter is now available.

### **FIDO Version 1.0.0 Released**

<https://github.com/openplanets/fido/zipball/master>

Version 1.0.0 of FIDO (Format Identification for Digital Objects), has been released. FIDO is a Python command line tool to identify the file formats of digital objects.

Please do try it out and let us know what you think. We welcome feedback, requests for new features and also bug reports. For more information about the new features and also the future of FIDO, see the OPF blog post: <http://www.openplanetsfoundation.org/blogs/2012-02-27-fido-version-100-released>

### **REDm-MED RDM requirements specification released**

<http://opus.bath.ac.uk/28040>

The Research Data Management for Mechanical Engineering Departments Project (REDm-MED) is pleased to announce the release of its first deliverable. It is entitled 'Research Data Management Plan Requirements Specification for the Department of Mechanical Engineering, University of Bath'; as well as the requirements themselves it also describes how the project arrived at them and how they have been validated. More information is available on the project blog:

<http://blogs.bath.ac.uk/redm-med/2012/03/05/requirements-for-dmp/>





## **What's What - Editorial - In the Beginning Was the Word**

***William Kilbride, Executive Director, DPC***

I've been thinking about data storage quite a lot recently and the difference between storage and understanding.

It's not one of my usual topics but I've learned quite a lot since the start of the year. It started at the PASIG conference in January and it's been focussed in particular by having to develop a presentation for later in March on the broad topic of 'preservation and the cloud'.

PASIG - which incidentally seems to have avoided turning into the corporate rally that some had feared - is an unusual forum which unites digital preservation and the data storage industry. It's a proper intellectual 'mash-up' of policy wonks, collections managers and engineers. I am sure that the SNIA types would have been bored by the reprise of the paramagnetic effect and the difficulty – apparent impossibility - of the perfect 1:1 atom to bit ratio. But it's new to me and I loved it. I heard about storage technologies that could soak up 10,000 separate tapes and hundreds of petabytes (I had to look up what a petabyte is), I heard about the economies of

tape versus flash, and I heard about all the massivest highest performance computing of all time. It's reassuring to know how big everything is in the Great State of Texas.

A highlight for me was a brief presentation by Gary Francis of Oracle who placed these most recent discussions of tape versus disk storage on a longer term trajectory of computing facilities. There's been a lot of change since the 1951 invention of the Remington Univac tape, and the 1956 IBM RAMAC disk – but perhaps the biggest surprise is that the same debate over tape versus disk continues 60 years later.

I'm sure that some of my library and archive colleagues would have wanted me to stick my hand up and point out that magnetic media, optical media, flash media and atom-level storage are part of a trajectory that started with the invention of the printing press.

They'd be wrong on at least one count – the question of storage is older than paper and it's a lot older than printing. I'm not just making some recondite point about the historical precedence of monastic manuscripts, the carving of stone inscriptions or the moulding of clay tablets. The origins of data storage, access and retrieval are probably congruent with the origins of language and culture itself.

Understanding the prehistoric origins of language – how's that for a grand challenge? It's hard work. Four brief and connected points from different perspectives which help us along the way:

Archaeologists are unable to put a precise date or place on the origins of language or on the origins of what might be termed 'culture'. But it's a very long time ago. There is clearly a moment in the archaeological record where brain size and connectedness enabled a step change in thought and there is also a moment when complex thought gives rise to a range of new phenomena which might be loosely categorised as the appearance of culture, religion and art, even if these categories are very problematical and lead to some surprising conclusions. But information storage understood as art and symbolic deposition is in there from early days.

From a neurological perspective, it is surprising the degree to which the external storage mechanism we call literacy imposes itself on the internal memory system we call the brain. The clues to this are the varying effects of dysphasia upon those with phonographic (e.g. Western, Semitic and Demotic) and those with logographic (e.g. Hieroglyphic or Chinese) alphabets. The

way you were taught to read actually stamps itself on the biology of your thought processes and it influences the pathology of degenerative illnesses in later life.

From an anthropological perspective, much has been made of the impact of literacy on societies – especially those that encountered the written word for the first time in the nineteenth and twentieth centuries. This tends towards a sort of technological determinism which is surprising given the diverse and sophisticated uses of literacy in the ancient world and the truly exotic configurations apparent in colonial encounters.

Sociologists, on the other hand, tend to emphasize the discursive nature of reading and writing as historically and socially constituted practices which don't just convey information but mask inequalities and render privilege ineffable. 'Literacy' created 'illiteracy' where it never existed before.

That's all very fancy but why does this matter to digital preservation? It's an elaborate way of saying that there's more to knowledge than storage. There always has been.

Here's a fun, frequent, but not very compelling half-truth that you hear from time to time in digital preservation sales pitches: if incised hieroglyphs / cuneiform tablets / palace seals / parchment could last thousands of years then what we need is a storage solution that mimics those qualities.

It sounds plausible but it's not the whole truth and in fact there's a kind of madness lurking at the end of it.

We need to be able to read the storage media too. Archaeology provides just as many if not more cases where the media survives excellently but the meaning is still lost. Pictish symbol stones for example – they are robust, numerous, beguiling and virtually impossible to interpret. The Phaistos disk is another example where the symbols survive in some detail but the meaning does not. Over-exposure and over-interpretation these can induce a sort of mania.

Even where we can make a reasonable effort to constitute meaning, the interpretation is not

straightforward. I remember the mix of horror and excitement when a colleague discovered a small carved stone in a secure 6<sup>th</sup> century context in Tintagel Castle in Cornwall which bore the name 'Artognou'. Let me assure you that this is etymologically very far from 'Arthur': but that's not what the journalists reported. Reading and interpreting ancient – and modern – texts is a specialised activity and one that requires a great deal of skill, and expertise. And given the rapid expansion of the digital universe these are going to be in short supply.

We've learned a great deal, sometimes through bitter experience, that some of the digital preservation challenges we thought were round the corner turned out to be less problematic in actual execution: for example that file format obsolescence is more likely to be a question of workflow and capacity than genuine loss. This is not one of those topics. Terms like 'representation information', 'significant properties' or recourse to the 'designated community' may seem like strange - even distracting concepts. Efforts to model descriptive information about context, provenance and authenticity may seem obtuse, and may spoil the fun we can have with data now. But we are right to investigate these topics: archaeology and history give us every reason to act.

You might think that the world has been slow to respond to the challenge of digital preservation: but I think we're still not sure about the implications of language.

---



**Who's Who: Sixty Second Interview with Patricia Sleeman and Ed Pinsent, Senior Digital Archives Specialists, ULCC**

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

We work at the University of London Computer Centre (ULCC), part of the University of London. As to our job title, we're both Senior Digital Archives Specialists, and we manage a range of related projects.

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

The University of London was founded by Royal Charter on 28 November 1836 and is the third oldest university in England. ULCC was founded in 1968, and was the first supercomputer facility established in London for the purpose of scientific and educational research by all of the colleges of the [University of London](http://www.ucl.ac.uk). It currently provides central IT and Web services to the University and its Schools, as well as offering e-learning services and a state-of-the-art data centre to a wide range of customers. ULCC's Digital Archives team was established in 1997 and has contributed to a large number of digital preservation activities in the last 15 years. We currently offer digital preservation training and consultancy, and institutional repository services for many of the University's institutions.

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

**Ed:** For the UK Parliament, I'm just completing a report on the feasibility of digitising a large collection of papers which they would like to put online. It was nice for me personally to get back to appraising physical papers and books on shelves, which I haven't done for a long time.

I've just finished a JISC-funded project called Future-Proofing ( <http://11kitbid.jiscinvolve.org/wp/> ): working with the records manager at UoL, we trialled various open source tools for normalising file formats. It was a very straightforward approach, which is why it appealed to me.

I'm also involved with Web archiving initiatives: I manage the JISC's accessioning to the UK Web Archive, and also advise an EU-funded blog archiving project (which involves most of our team) called BlogForever ( <http://blogforever.eu/> ). It is developing a system for harvesting and preserving blog content. I'm helping to develop the preservation policy; Patricia is developing case studies to test the system.

**Patricia:** I am currently working on the JISC-funded [SHARD](#) project, part of the JISC Digital Preservation programme. We are developing both face-to-face and online training tools for

preservation of research data.

I also manage the Digital Preservation Training Programme (DPTP) ( [www.dptp.org](http://www.dptp.org) ), which Ed and I deliver. Originally a JISC project, we continue to run this popular course, with the DPC's support. We constantly adjust the course to reflect the new developments we encounter in our projects. Next course is in May!

I also work on the [House of Books](#) , an EU project with an NGO called Un Ponte Per...(a bridge towards) and UNESCO. It is looking at capacity building for the Iraq National Library and Archives (INLA) and other Middle Eastern libraries, in relation to digitisation and digital preservation. I contributed to various workshops in Jordan, Iraq and Italy working with groups and have found it a very satisfying and challenging experience.

I have also just begun work for the Enhancing Linnean Online project, working with ULCC's repository specialists to enhance the [Linnean Online collection](#) .

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

**Ed:** I wish I knew! I am not a computer person at all. I trained as an Archivist in 1991 but have been working in paper archives since about 1987. In 2004 I joined ULCC to become a cataloguer for the National Digital Archive of Datasets (NDAD), and from that point on found myself involved in several JISC-funded projects. It's been a steep learning curve for me, but I'm lucky to work with friendly and approachable IT specialists, and we've not only managed to find common ground but even worked together successfully.

**Patricia:** I became interested in digital preservation when studying archaeology in University College, Galway. We did a study on ring forts in Cork and Kerry compiling data about their various aspects trying to see which had the highest, largest etc. Where is the data now? Lost in the mists of time. Obviously Cork won.

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

**Ed:** Well, oddly enough I don't provide digital preservation services directly to the UoL as yet, but we're working on that. We're trying to make inroads into helping the records manager with preservation of digital records that have long-term value; and also to start doing something about research data in digital form. There are many synergies with the work our colleagues are

doing on research repositories for the University.

**Patricia:** Yes, I too think that we should be sharing far more of our expertise within our own institution. In straitened times like these shared services should be prioritised. The management of digital information must be a priority for every organisation especially one like ours.

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

**Ed:** See my previous answer...we'd simply like to offer more digital preservation to the UoL.

**Patricia:** I would like to work more with schools. I did a [workshop](#) last year in my local primary school and found the children to be very aware and insightful.

International partnerships also interest me. There is an awful lot of digital material being created beyond the so-called developed world. A lot of this information is potentially at risk. I would like to see practical multilingual best practise guides for various aspects of preservation of digital information being available in simple accessible language and freely available online on sites such as IFLA, ICA and ICOM.

I would also like to see more partnerships with Museums. They produce a quantity of digital material and yet we see very few museum folk on our training programme. What gives, museum folk?

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

**Ed:** Hmm. If you ask me we need fewer tools, not more of them. What I would like to see is more use being made of the tools and services we already have.

**Patricia:** I would like a translation tool for translating digital preservation jargon for a lay person like me, an archivist from [Clonakilty](#). I think we in the community of digital preservation are often preaching to the converted.

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

**Ed:** It's often much simpler and cheaper than you think.

**Patricia:** Take care of your personal digital archive. Obsolescence is inbuilt into our society and it will affect your personal material unless you do something now.

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

**Ed:** I understand where this question is coming from, but it's like the people who used to ask me "what is the oldest document in your archive?" when I was working for the Church of England. That always grated with me. Why single out one document of value? The value of an archive, be it paper or digital, resides in its diversity and its completeness; many documents reflecting the truth (or many truths) about history.

**Patricia:** An audio file of "Scrap Saturday". It was a brilliant satirical Irish radio show from 1990s created by Dermot Morgan.

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

We both contribute regularly to <http://dablog.ulcc.ac.uk/> . Email us at [e.pinsent@ulcc.ac.uk](mailto:e.pinsent@ulcc.ac.uk) or [p.sleeman@ulcc.ac.uk](mailto:p.sleeman@ulcc.ac.uk)

.





## **Featured Project: What the deuce is SPRUCE? [or, *we need more* ]**

***Bo Middleton, Head of e-Strategy and Development, Leeds University Library***

*You know how it is, you're browsing through a JISC funding call and you have an idea, and then 3 months later you've got this mega-project with a weird acronym on your hands.....*

The SPRUCE Project started with [AQuA](#) , a small project focussing on quality assurance for digital content. The AQuA project methodology had its origins in the 'hackathon' where software developers meet up to solve technical challenges over a short period of time. Hackathon events have become increasingly popular in recent years as a way of removing the overhead of traditional project based development, and enabling rapid prototyping and development progress through a combination of collaboration and competition. The digital library community has begun to embrace the hackathon concept, with projects such as

[DEVCSI](#)

, working actively to develop a technical community via supporting activities such as hackathons and programming challenges.

The advent of open data and linked data approaches has encouraged the creation of a similar event model to the hackathon but with a focus on exploiting open interfaces, mashing up data from several sources and providing new and often innovative services. Data Mashup events, like hackathons, typically provide supportive environments for participants to collaborate in small teams and compete to win challenges.

The AQuA Project held two 3-day workshops which built on the hackathon concept – but, rather than being purely technically focused, the workshops brought together digital preservation experts (those that know a bit about it and have experience of appropriate techniques and tools), techies (those who didn't know about digital preservation but were interested in agile development and wanted to understand more about digital preservation), and content 'owners' (again, not particularly au fait with digital preservation – but they had digital content and they knew they should be considering how best to check the content for issues). We called them mashups – because we were 'people mashing'.

It's difficult to get across how successful the AQuA events were – we were aiming on getting a mixture of people together to solve quality assurance challenges and we succeeded in doing that, but the events were highly successful in facilitating collaboration and knowledge sharing. We sowed the seeds for a wider and more diverse digital preservation community whilst also extending expertise into organisations that were previously thinking digital preservation meant having a risk log for their collections! (Leeds being one of those organisations).

And so, that is how SPRUCE came about. When JISC issued a call for a project to “inspire, guide, support and enable HE, FE and cultural institutions to address digital preservation gaps; and to use the knowledge gathered from that activity to articulate a compelling business case for digital preservation”, we thought that the AQuA approach would work – to grow the community, to extend support into organisations that are starting down the road of truly understanding what is meant by 'digital preservation', and to then use evidence from the SPRUCE events to create a digital preservation business case.

The SPRUCE Project is being led by the University of Leeds Library and project partners span the currently active digital preservation community: the British Library, the London School of Economics, the Digital Preservation Coalition, and the Open Planets Foundation.

They'll be a number of mashup events this year - the first one is coming up in Glasgow in April, and we're aiming on going west (Bristol/Bath/similar) in July and then to London in October. Details for the Glasgow and event are at <http://www.dpconline.org/advocacy/spruce> where you'll also find registration details and, later on, information about future events. I encourage you to join in – people mashing benefits all, from newbies to experienced staff, from techies to non-techies.

NB SPRUCE mashups are free to attend and the best work from event attendees will be

awarded funding to develop the activity and embed it within their organisation's processes (£60k is available for these awards).

---

{jcomments on}