

In this issue:

- [What's On](#) - Forthcoming events from June 2012 onwards
- [What's New](#) - New reports and initiatives since the last issue
- [Who's Hiring](#) - Job Vacancies from DPC Members
- [What's What](#) - Big Data, Big Deal? Marieke Guy, DCC
- [Who's Who](#) - Sixty second interview with Neil Grindley, JISC
- [Feature](#) The 2012 Digital Preservation Awards, William Kilbride, DPC
- [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.

### LOV symposium

*18 June 2012*

<http://klington.uc3m.es/hive-es/wiki/index.php/Symposium-en>

A day long-symposium to address challenges, solutions, and research specific to linking and opening vocabularies on the global web. Participation is free and open to all attendees, presenters, and general participants/viewers. Registration is required. Attendance can be in person or virtual.

### SKOS-2-HIVE workshop

*19 June 2012*

<http://klington.uc3m.es/hive-es/wiki/index.php/Workshop-en>

The SKOS-2-HIVE workshop targets the use of semantic web technologies for representing and describing collections using multiple controlled vocabularies. The workshop focuses on basic understanding and usage of W3C's Simple Knowledge Organization Systems (SKOS), linked data, and the HIVE library of open source applications.

### **Policies and Practices in Access to Digital Archives: Towards a New Research and Policy Agenda**

*2-6 July 2012*

<http://summer.ceu.hu/archives-2012>

This course is intended to serve as a bridge between archivists, curators, researchers, legal experts and policymakers whose work deals with digital records, cultural heritage collections and/or open data. Launching an itinerary to reform the political and statutory landscape by uniting the efforts of key stakeholders is one of the broad purposes of the course.

### **The 7th International Conference on Open Repositories (OR2012)**

*9 - 13 July 2012*

<http://or2012.ed.ac.uk/>

EDINA, the University of Edinburgh's Information Services and the Digital Curation Centre are delighted to announce that the University of Edinburgh will be hosting the Seventh International Conference on Open Repositories (OR2012) from 9-13 July, 2012. The theme and title of the 2012 conference at Edinburgh – Open Services for Open Content: Local In for Global Out – reflects the current move towards open content, 'augmented content', distributed systems and data delivery infrastructures.

### **Links That Last**

*19 July 2012*

<http://www.dpconline.org/events/details/47-linksthatlast?xref=49> This briefing day will introduce the topics of persistent identifiers and linked data, discussing the practical implications of both approaches to digital preservation. It will consider the viability of services that offer persistent identifiers and what these offer in the context of preservation; it will review recent developments in linked data, considering how such data sets might be preserved; and by introducing these two parallel topics it will go on to consider whether both approaches can feasibly be linked to create a new class of robust linked data. Based on commentary and case studies from leaders in the field, participants will be encouraged to consider practical implications for their own work and new directions for research and development in the field.

### **DPC Director's Group Meeting (Invitational - Full Members Only)**

*20 July 2012*

<http://www.dpconline.org/events/details/48-directors-group-2012?xref=50>

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.

### **Advanced Techniques on Data Analytics and Data Visualization**

*August 22-24 2012*

<http://www.icdim.org/workshop.php>

Advanced tools such as data mining, data modeling, data visualization, and information analysis etc. have been extensively researched recently. The objective of this workshop is to provide an interactive platform for academic researchers and industrial practitioners to exchange ideas, disseminate the latest development in data analytics and data visualization. Conceptual models, tools, and current issues are encouraged to be discussed.

### **2nd SPRUCE 'Mashup', London 18-20 September**

*18-20 September 2012*

<http://sprucelondon.eventbrite.co.uk/%20http://wiki.opf-labs.org/pages/viewpage.action?pageId=13041673>

JISC funded SPRUCE Project cordially invites you to the second SPRUCE Digital Preservation Mashup. SPRUCE is organising a series of free events around the UK that will provide support and technical expertise to address the real digital preservation challenges that institutions face. The best work from event attendees will be awarded funding to develop the activity and embed it within business as usual processes. £60k is available for these awards.

## What's New

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

### **DMP Online v3.0 launches**

<https://dmponline.dcc.ac.uk/>

The Digital Curation Centre is pleased to announce the launch of DMP Online v3.0 This new release marks a major progression in the software's functionality. For the first time users can create data management plans incorporating multiple templates, so if your institution, your funder and your publisher all require data management plans, you can now create a single plan to satisfy them all.

### **DPC commissions three new Technology Watch Reports**

The Digital Preservation Coalition and Charles Beagrie Limited are delighted to announce the continuation of their collaboration, producing 3 more Technology Watch Reports. The three new reports will be:

- Web Archiving, Maureen Pennock
- Preserving Computer Aided Design, Alex Ball (jointly with DCC)
- Preservation Metadata, Brian Lavoie and Richard Gartner

### **SHERPA**

<http://www.sherpa.ac.uk/romeo/?la=hu>

SHERPA is pleased to announce that a new Hungarian language version of its RoMEO database is now available. The RoMEO interface has already been translated into Hungarian, and our Hungarian partners HUNOR have started adding RoMEO data directly for Hungarian publishers and journals. Existing RoMEO data for other publishers is in the process of being translated.

### **Request for Comment: IIIF Image API Proposal**

<http://library.stanford.edu/iiif/image-api>

The International Image Interoperability Framework is an initiative driven by several major research and national libraries to enable the rich and robust delivery of digital images through common interfaces, and to spur the development of open source and commercial software solutions in this space. The IIIF Working Group invites comment and feedback on a proposed

API for the the delivery of images via a standard http request. Please send feedback to [iiif-feedback@googlegroups.com](mailto:iiif-feedback@googlegroups.com) by June 22, 2012. Once all feedback is reviewed and incorporated, we will release the specification for adoption.

### **Ringgold becomes the first contracted ISNI Registration Agency for Institutions**

<http://www.ringgold.com>

Ringgold Inc. has contracted with the ISNI International Agency to be the first ISNI (International Standard Name Identifier) Registration Agency for institutional identification. Ringgold will incorporate ISNIs into its Identify database of institutional identifiers and distribute these ISNIs without charge to Ringgold's Identify clients. For Ringgold's clients, this will immediately affect over 300,000 institutions worldwide.

### **Podcast: Cookie law next steps - legal expert and university web manager**

<http://bit.ly/Noc4nu>

Cookie law is now in force in the UK – and we've all got a different way of dealing with it. In this podcast, we speak to John X Kelly, lawyer at JISC Legal, for the definitive guide to the law. We also ask Mike Nolan, head of web services at Edge Hill University, to share top tips from his approach. Read the JISC Legal guidance at <http://www.jisclegal.ac.uk/cookies> .

### **Request for support for a new Question and Answer site for Digital Preservation and Curation**

A proposal has been put forward for new Digital Preservation Q&A site that would provide a neutral, central location where anyone can ask questions about preservation, and receive answers from the experts. The site will use the tried and tested mechanisms provided by Stack Exchange, which supports Q&A for a whole host of topics. These capabilities ensure that questions are focused and on topic, and that the best answers are moderated and voted to the top by other users. The result is a growing knowledge base of information on the topic, that links to more detail on the variety of sources of information about digital preservation that are scattered across the internet. In order for the DP Q&A site to go live, it needs your support and demonstration of commitment to using it. Those interested can help out in two ways:

1. Sign up for the proposal here: <http://area51.stackexchange.com/proposals/39787/digital-preservation?referrer=ikyDT2iSDeEl8-Pye4BdZw2>
2. Build up 200 Stack Exchange reputation points by asking questions and answering questions on another Stack Exchange site. The Libraries and Information Science Stack Exchange site is the perfect place to do this: <http://libraries.stackexchange.com/>

More information about this proposal can be found in a Blog post by Paul Wheatley here: <http://openplanetsfoundation.org/blogs/2012-06-06-question-and-answer-site-digital-preservation>

---

## **Who's Hiring - Job Vacancies from DPC Members**

### **Digital Preservation Technical Architect, the British Library**

**Location:** Boston Spa, Yorkshire

**Position Type:** Permanent

**Specialism:** Information Technology

**Salary:** £34,391 - £39,743 per annum

**Closing Date:** 25/06/2012

For informal enquires please contact Maureen Pennock, Digital Preservation Manager, on 01937 546302

For more information see: <http://bit.ly/LfCIUb>

### **Digital Preservation Technical Lead, the British Library**

**Location:** Boston Spa, Yorkshire  
**Position Type:** Fixed Term until 31 July 2014  
**Specialism:** Information Technology  
**Salary:** £34,391 - £39,743 per annum  
**Closing date:** 25 June 2012

For informal enquires please contact Maureen Pennock, Digital Preservation Manager, on 01937 546302

For more information see: <http://bit.ly/LQNd8a>



**What's What - Editorial - Big Data, Big Deal?**

*Marieke Guy, Research Officer, DCC*

The provocatively titled Eduserv symposium 2012: Big Data, Big Deal? provided a forum for IT professionals, and anyone responsible for managing research data or planning to work with big data in Higher Education, to discuss the meaning of big data and the challenges it presents. Speakers from both the commercial and academic world came together to reflect on big data trends and the implications for research, learning, and operations in HE. Big data is considered to be data sets that have grown so large and complex that they present challenges to work with using traditional database management tools. The key factors are seen to be the "volume, velocity and variability" of the data (Edd Dumbill, O'Reilly Radar).

During the day some interesting key themes emerged:

We don't need to get hung up on the 'big' word.

While data is increasing exponentially (something a number of scary graphs indicated) this doesn't have to be an issue, we are getting more used to dealing with large scale data. While the large Hadron Collider produces around 15 petabytes of data annually, ecology engineer Simon Metson from the University of Bristol/Cloudant talked about 50 terabyte datasets. In his lightening talk Simon Hodson, Programme Manager at JISC, provided a quick straw poll from two Russell group universities. Both believed they held 2 petabytes of managed and unmanaged data and while one currently provides 800 terabytes of storage the other provides only 300 terabytes. There were concerns from those universities that their storage may be full in the next 12 months. But then storage costs are decreasing, and storage models are changing (often to cloud computing). Guy Coates from the Wellcome Trust Sanger Institute explained that the cost of genome sequencing halves every 12 months and this trend is continuing. It is more than likely that the id="mce\_marker"000 genome will be here in the next 12 months and people will soon be purchasing their own USB stick genome sequencers!

The tools are now available.

During the symposium speakers mentioned tools such as Hadoop, DB Couch, NoSQL which all allow people to work easily with data sets. There was consensus that people no longer need to create systems to deal with big data, but can now spend that time on understanding their data problem better. Graham Pryor from the DCC saw the data problem as being in part about how you get researchers to add planning into the research data management process, these issues



are central to effective data management irrespective of size.

It's all about the analysis of data.

While storage of data can be costly and management of data labour intensive, the analysis of data is often the most complex activity. Keynote speaker Rob Anderson from EMC explained that "If we'd been able to analyse big data we might have been able to avoid the last financial crash". He sees the future as being about making big-data-based decisions and unlocking value by making information transparent and usable at a higher frequency. However while tools have a role to play here analysis still requires human intervention. On his blog Adam Cooper from CETIS advocates human decisions supported by the use of good tools to provide us with data-derived insights rather than "data driven decisions". During the symposium Anthony J Brookes, professor of Genomics and Informatics at the University of Leicester gave an overview of disastrous divide between research and healthcare (i.e. divide in the management of data e.g. the use of different standards), and the need for knowledge engineering (analysis) to bridge the gap. When talking about data as a way of life for public servants Max Wind-Cowie, from the progressive Conservatism Project Demos, explained that many public centre brands have been toxified and big data can help us to better understand that journey. Devin Gaffney from the Oxford Internet Institute provided a number of interesting case studies showing why prescribed analytics often fail to deliver.

We don't yet know what data to get rid of.

Anthony Joseph, professor at the University of California, Berkeley suggested the selection/deletion of data is the most intractable problem of big data. He pointed out that if you "delete the right data no-one says thank you, but if you delete the wrong data you have stand up and testify" giving the US climate trial as an example. We often find it difficult to be selective when curating data because we don't know the question we will need to answer yet.

We need data scientists.

Many of the talks highlighted the need to build capacity in this area and train data scientists. JISC is trying to consider changing the research data science role in its programmes and Antony Joseph asked HEIs to consider offering a big data curriculum. In his summary Andy

Powell Eduserv's Research Programme Director asked us to think carefully about how we use the term data scientist as the label can be confusing. He noted that there is a difference between managing data, an activity we at the DCC are fairly familiar with, and understanding and analysing data.

'Big data' and its associated challenges are likely to dominate the infrastructure landscape in the years to come especially as we start to see a return on investment for organisations who understand their data and are able to unlock its potential value. So far, commercial organisations have been most active in investing in big data infrastructures and, as a result, most successful in exploiting the full value of the data they hold. Better access to guidance and advice on how not-for-profit organisations and HEIs can follow suit will be needed if they are to reap the same rewards.



**Who's Who: Sixty Second Interview with Neil Grindley, Programme Manager, JISC**

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

I work for JISC and I'm the Programme Manager responsible for digital preservation and curation.

**Tell us a bit about your organisation**

JISC is a funding organisation that tries to ensure that UK universities are best placed to take advantage of innovative digital infrastructure, tools and methods. It does this via programmes of funded projects and in partnership with organisations both within and outside of the UK.

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

JISC programmes are often quite extensive in scope but at the moment, the three related topics I'm tackling via commissioned projects are: the business case for digital preservation; the cost/value/benefit of digital preservation; and the sustainability of digital collections.

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

I did a History of Art degree and then in about 1992 I began doing digital cataloguing work in an image library. I was struck early on by the apparent fragility of the work that we were doing. We used anachronistic hardware and some pretty arcane procedures and I kept wondering ... is this going to be OK? Are we doing the right thing here? What happens when this particular system fall over? The hardware is better now and the procedures make more sense but those initial thoughts have stayed with me over the years. Perhaps you need to be a bit of a worrier to be interested in Digital Preservation!

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

We're a funding organisation so our focus is on helping other people to advance and engage with digital preservation. As such, the principal challenge I have is to make sure that the work I commission and the issues that I try and persuade JISC to invest in are the ones that are going to be of maximum current and future value to the broader community. Continuing to do this effectively as JISC works through a process of renewing the foundations of its own governance and funding (see <http://bit.ly/jom3Ht> for more details) will add to the challenge!

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

There are maybe a dozen or so UK universities that regularly engage with JISC digital preservation programme funding. I don't think everybody needs to care about every aspect of the topic but it does puzzle me that the other 150 or so UK HEI's seem so reticent about applying for project funding. It would be great to get more institutions involved. I'd also like to do some international joint funded programmes with US and European funding agencies. Aligning processes internationally can be difficult but the 'Digging Into Data' programme ( <http://bit.ly/4XpAQ9> ) indicates that it is possible.

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

A reporting tool that provided half a page of easily digestible and convincing evidence setting out the cost/benefit case of any digital preservation action undertaken within an institution!

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

Approach digital preservation from your own perspective. It's a great topic because if you just want practical steps that help you to manage your information better – then DP has got some good ideas you can use. Conversely, if you are an information scientist on the lookout for challenging research topics, DP has those too! It's a relatively small community that is tackling a problem that is growing in importance. That's a nice space to work in.

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

It would be amusing for the humanoid-type beings of the year 2012 to stumble across a copy of Oliver Postgate and Peter Firmin's episode of The Clangers entitled 'The Music of the Spheres'. Not a bad representation of late 20<sup>th</sup> century genius.

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

There is extensive material on the JISC website ( <http://www.jisc.ac.uk> ) or you can email me at [n.grindley@jisc.ac.uk](mailto:n.grindley@jisc.ac.uk) . Or better still, [c.uk](http://www.jisc.ac.uk) . Or better still, come along to a few digital preservation events! I'm often there in some capacity or another.



## **Feature: The 2012 Digital Preservation Awards**

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

The Digital Preservation Awards are coming back in 2012.

The DPC was established in 2002 to help agencies meet this new and growing challenge and in 2004 we sponsored a small prize to mark outstanding contributions to the field. It was so popular that we've offered the prize every other year since, and each time the quality and number of nominations has grown. The National Archives won the prize in 2004 and 2007, the PREMIS Working Party won the prize in 2005 and in 2010 it was Los Alamos National Labs. Digital preservation is vital work but it's not particularly high profile: the awards give us an opportunity to celebrate efforts that are not very much celebrated by other groups.

This year, the award takes a new form. In the past a single award was offered as one of the Conservation Awards but this year, as 2012 is the tenth anniversary of the founding of the DPC, we're offering 4 separate prizes. This includes a special 'DPC Decennial Prize' for the most

outstanding contribution to digital preservation in the last decade. There are also prizes for 'Teaching and Communication' and for 'Research and Development' as well as an innovative Digital Preservation Challenge being offered via the Open Planets Foundation.

We're calling on all our friends and colleagues - the whole digital preservation community - to help us get the best possible set of applications.

The criteria are defined broadly, encompassing any initiative that has helped ensure 'our digital memory is available tomorrow', and although the DPC's membership is in the UK and Ireland, this is an international competition. We encourage all manner of proposals – projects, services, ideas, books, methodologies, standards, working groups and campaigns: all are welcome.

Applications are due by the 17th August at which point they will be scrutinised by a judging panel drawn from the DPC membership. A shortlist will be announced in October and DPC members will be invited to vote for their favourite proposals. We expect that the judges will have a difficult job, and we really hope that the community makes it as hard as possible to choose: but it's also an entirely optimistic process.

The winners will be announced at a special ceremony in London on 3rd December.

The application pack is available online at: <http://www.dpconline.org/advocacy/awards>

{jcomments on}