

# What's new ...

Digital Preservation Coalition | Digital Curation Centre Issue 23, 1<sup>st</sup> March 2010

## In this issue:

- **What's on and What's New**
- **Editorial:** Something old something new (Chris Rusbridge, DCC)
- **Who's who:** sixty second interview with Polly Parry, Natural History Museum.
- **One World:** Sabine Schrimpf, NESTOR
- **Your View:** commentary, questions and debate from readers

## What's on:

### Research Data Management Forum: Dealing with Sensitive Data: Managing Ethics, Security and Trust 10-11 March 2010

Retrieved 27 February 2010 from: <http://www.dcc.ac.uk/events/research-data-management-forum/research-data-management-forum-dealing-sensitive-data>

This will be the fourth meeting of the Research Data Management Forum (RDMF). Aimed at 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 event will address the topic "Dealing with Sensitive Data: managing ethics, security and trust."

### Digital Preservation Training Programme, London 29-31 March 2010

Retrieved 1 February 2010 from: <http://www.dptp.org/about/dptp-london-soas-march-2010>

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*. The course content is aimed at multiple levels of attendee: people from technical and archival professions come together, to learn the same standards and methods for digital preservation. The overall aim of the course is *critical thinking*, assessing ways of acting and planning at an organisational level.

## What's New:

### DCC phase 3 - a new phase, a new perspective, a new Director

Retrieved 1 February 2010 from: <http://www.dcc.ac.uk>

As the DCC begins its third phase on March 1<sup>st</sup> 2010 we are delighted to announce the appointment of our new Director, Kevin Ashley, who will succeed Chris Rusbridge upon his retirement in April 2010. Kevin Ashley has been Head of Digital Archives at the University of London Computer Centre (ULCC) since 1997, during which time his multi-disciplinary group has provided services related to the preservation and reusability of digital resources on behalf of other organisations, as well as conducting research, development and training. The group has operated the National Digital Archive of Datasets for The National Archives of the UK for over twelve years, delivering customised digital repository services to a range of organisations. As a member of the JISC's Infrastructure and Resources Committee, the Advisory Council for ERPANET, plus several advisory boards for data and archives projects and services, Kevin has contributed widely to the research information community. As a firm and trusted proponent of the DCC we look forward to his energetic leadership in this new phase of our evolution.

### Panton Principles for Open Data in Science

Retrieved 1 February 2010 from: <http://www.pantonprinciples.org/>

Science is based on building on, reusing and openly criticising the published body of scientific knowledge. For science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made open. To this end, the Panton Principles have been developed by Peter Murray-

Rust, Cameron Neylon, Rufus Pollock and John Wilbanks with input from the Working Group on Open Data in Science.

#### **Researchers of Tomorrow: A three year (BL/JISC) study tracking the research behaviour of 'Generation Y' doctoral students**

Retrieved 1 February 2010 from: <http://explorationforchange.net/index.php/current-projects/researchers-of-tomorrow/researchers-of-tomorrow-home.html>

Funded by the British Library and JISC, this three year study on the research behaviour of the 'Generation Y' scholar is being undertaken by Education for Change, in collaboration with The Research Partnership.

#### **Library of Congress Digital Preservation Newsletter**

Retrieved 1 February 2010 from: <http://www.digitalpreservation.gov/news/newsletter/201002.pdf>

The February 2010 issue of the Library of Congress Digital Preservation Newsletter is now available.

#### **New EPSRC Website**

Retrieved 1 February 2010 from: <http://www.epsrc.ac.uk>

The Engineering and Physical Sciences Research Council (EPSRC) will be launching its new-look website on 24 February. A fresh, clean design will now greet visitors to the site. Strong emphasis has been placed on making information easier to find. Under the skin there have been many changes to move towards full compliance with the Government's Central Office of Information's web guidelines which aim to make the site useable for people with visual, auditory, physical, speech, cognitive or neurological disabilities.

#### **Results of Digital Preservation Costs Survey Now Available**

Retrieved 1 February 2010 from: <http://www.beagrie.com/jisc.php>

The findings from the Keeping Research Data Safe 2 ("KRDS2) survey of digital preservation cost information are now available.

#### **Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines**

Retrieved 1 February 2010 from: [http://escholarship.org/uc/cshe\\_fsc](http://escholarship.org/uc/cshe_fsc)

Since 2005, the Center for Studies in Higher Education (CSHE), with funding from the Andrew W. Mellon Foundation, has been conducting research to understand the needs and practices of faculty for in-progress scholarly communication (i.e., forms of communication employed as research is being executed) as well as archival publication. The complete results of this work can be found at the Future of Scholarly Communication's project website. This report brings together the responses of 160 interviewees across 45, mostly elite, research institutions to closely examine scholarly needs and values in seven selected academic fields: archaeology, astrophysics, biology, economics, history, music, and political science.

#### **Knowledge Exchange: Successful Reuse Of Research Data Requires Integrated Approach**

Retrieved 1 Feb 2010 from: <http://www.knowledge-exchange.info/Default.aspx?ID=62&M=News&NewsID=77>

On 23-24 September 2009 an international discussion workshop was held in Berlin, prepared and organised by Knowledge Exchange. The main focus of the workshop was on the benefits, challenges and obstacles of re-using data from a researcher's perspective. The presentations are now available to view online.

### **Editorial:**

#### **Something old, something new (Chris Rusbridge, DCC)**

I am very pleased to have this opportunity to write an editorial for "What's New". The plan is to alternate the editorial between the DPC and the DCC, and this month it's the DCC's turn.

Let me start with something old and something new. I'm writing this on the last full working day of Phase 2 of the Digital Curation Centre, and by the time you read it, Phase 3 will have started. As ever, the end of a project is a frantic rush to get those deliverables finished; I won't tell you how many there were when I checked this morning, but it was an



uncomfortably large number, so I'm hoping the keyboards will be running hot all weekend!

Not long ago the DPC unveiled its new web site, and even more recently we had a "soft launch" of our revised web site. It's a big change from the previous site, and there is still a lot of work to be done making sure that everything is available and works properly. If you spot any problems, please do let us know. I've been unkind to various people about broken links, so I'm sure some people will be checking to make sure all their old bookmarks work (it was a design aim that they do work, but of course in the many changes some may not, either deliberately for a withdrawn item, or accidentally... but do let us know!). Much as William Kilbride described in the last editorial, we arranged with the excellent UK Web Archiving Service to collect the whole web site a few days before the transfer, so if we did drop anything you should be able to find it there. (<http://www.webarchive.org.uk/ukwa/target/118444/source/search> if you are interested).

We have come a long way since the DCC was first established in 2004 with Peter Burnhill as Interim Director. In those early days there was ambiguity on the extent that we were for research data curation and the extent we were for digital preservation (data or not). We're much clearer now that we are about improving the management and curation of research data; we haven't given up on the long term, but given the realities of research projects, we believe that long term benefits will accrue if short to medium term practice improves. If we can convince researchers that good practice will benefit their current research, with lower costs and better results, we'll be going in the right direction. We no longer have a research role, or a development role in core funding, although we do hope to get funds for specific curation-related development tasks. I see Phase 3 as a real maturing of the mission.

On the subject of maturing, you may know that I will be retiring shortly as Director of the DCC, so this is not only my first but also my last editorial here. I started out working on computers with 6-bit characters (yes, the 8-bit byte is or was not ubiquitous, many early computers had 12, 24 or 60-bit words, and I seem to remember the 4-bit nybble) so 2<sup>6</sup> seems a very appropriate age! I've thoroughly enjoyed my time at the DCC, and I'm grateful to so many of you for the support you've given me and the DCC. I'm extremely pleased at the appointment of our new Director (see elsewhere in this issue), who will start in mid April. He will bring years of data curation experience with him: he's been there and done that, definitely has the T-shirt!

By the way, I've ruminated sometimes about why so little remains to us from some of those early days. The high cost, genuine frailty and often commercially unsuccessful formats of storage form at least part of the reason, but not necessarily in the way we might now think. It isn't just that data and documents stored on those early computers somehow got lost. It's rather that it mostly never occurred to us that storing data or documents on a computer was at all sensible. I remember as a callow youth vaguely sneering at the French term "l'informatique"; what did information have to do with it?

**'I remember as a callow youth vaguely sneering at the French term 'Informatique'. what did information have to do with it? Computers were for computing, for faster maths!'**

Computers were for computing, for faster maths! You built collections of input parameters, ran them through your program suite, and printed out the results. If you needed to change things, you edited the input card deck or paper tape, and ran it again. When done, you kept your printout and threw the card deck or paper tape away. Then off you went to build your bridge or whatever (it's sobering to know that quite a number of motorway bridges in Britain were designed using software that ran in 48 kilobytes of memory with two magnetic tape drives and no hard disks). It wasn't until after widespread use of PCs with reasonable amounts of storage for office and home activities that the French term was finally vindicated, and "computing" relegated to a quaint activity of engineers and researchers!

Anyway, enough reminiscing and back to the point. The DPC and the DCC may sometimes get confused, but we have quite different missions and scope. But both organisations need each other, and it is very pleasing to see cooperation appearing in many different ways, including this publication.

## Who's who:

*In this section we invite a DPC member working on digital preservation to introduce themselves and their work. It will help us build connections between staff working on similar issues, will help promote the work of members, and will help identify priorities for the development of shared tools or services This time we hear from ...*

### **Polly Parry, The Natural History Museum**

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

The Natural History Museum is about a great deal more than simply dinosaurs – although I expect that exerts a significant pull on over 4 million visitors that we get a year. Behind the public galleries we have more than 300 scientists working on a huge variety of research projects, many on a global scale, and curating a world-class collection of over 70 million specimens. We have over 850 staff across Science, Public Engagement and the administration/support departments, based on three sites. From an information management and digital preservation point of view, this gives us a wide range of issues and challenges that we need to address.

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

It is really the natural extension of my work in archives and records management – the next big thing we have to tackle. Almost everything we produce now is digital, so that is the direction in which preservation activities also need to go. Of course, it means balancing it with all the other more traditional information management tasks that we need to keep up with, which adds to the challenge.

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

My level of involvement varies, but the Museum has a number of projects with digital preservation aspects. We have the [Open Repository](#), where our scientists' published material is recorded and made available (subject to intellectual property rights) for further research and private study. We are in the very early stages of implementing a Digital Asset Management System, primarily to help deal with our vast quantities of images. I am most heavily involved in a project to digitise the Museum's accession records to provide a backup copy in case the originals are destroyed, which has raised interesting questions over what file format to use. That's just the digital preservation side of course – beyond that there's much more...

#### **What are the challenges of digital preservation in museums?**

I wouldn't really want to speak for all museums, although I suspect most of our challenges as an individual museum are similar to everyone else in the sector and indeed beyond. In our case, I think a key challenge is simply how big the problem is, and where to start. Some of the challenges are on the technological side i.e. how you can actually make sure material is preserved, that we don't lose important records due to obsolescence of file formats and hardware. Some of them are actually old-fashioned records management issues overlaid with the conditions of the digital environment - how to cope with the quantity of material that is produced, how to decide (and implement) what to keep, how to ensure people take responsibility for their own records and understand their role in management and preservation.

#### **What projects would you like to work on in the future?**

We really need to get to grips with our electronic records management – whether we ultimately acquire a system or develop work-arounds. We are working with staff now to tidy up and restructure shared drives, and trying to encourage people to use them more, as a small step towards the culture change that would be required to implement any kind of EDRMS. On a related note, we would like to establish a digital archive of some sort. Whether we can utilise existing systems or need to acquire something new remains to be seen. What I think will be useful in the planning process is the OAIS reference model, which we explored on the Digital Preservation Training Programme that I attended recently (courtesy of a DPC scholarship, in fact).

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

Relationships and partnerships within the museum sector are clearly important, given our similarities and ability to draw on each others' experiences. But I think it's equally as important to build partnerships with other sectors, the academic or private sectors for instance, to enable that much broader cross-fertilisation and leverage.

#### **If just one tool or standard could be brought into existence that would make your job easier...?**



Inspired by having just watched a film which included some time travel (the question doesn't specify realism!), something that would enable us to see into the future to help decide now what material really turns out to be of interest and thus worth the effort of preservation in the long term.

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

I think it would have to be something quite unoriginal and personal like a wedding photo or video, or a photograph of my whole family. I'm sure there are many vital Museum-related files that might qualify, but then how would you pick just one (other than by referring to my previous answer)?

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

[p.parry@nhm.ac.uk](mailto:p.parry@nhm.ac.uk) or the Museum website [www.nhm.ac.uk](http://www.nhm.ac.uk)

## One World:

*In this section we invite a partner or colleague to update us about major work on their home country that will interest readers. In this issue we hear from...*

### Sabine Schrimpf, Nestor, Germany

2009 was a decisive year for nestor, the German network of expertise for digital preservation. After six years as a funded project, the network transformed to a sustainable partner consortium in July 2009. Most of the partners of the project phase and 3 additional organisations signed a cooperation agreement, in which they commit to continue the work of the network. The new partners are ...

- Landesarchiv Baden-Württemberg (Baden-Württemberg State Archive)
- Bibliotheksservice-Zentrum Baden-Württemberg (Baden-Württemberg Library Services Centre)
- Institut für Deutsche Sprache (German Language Institute)



The purpose of nestor is to bring together the existing know-how and the competencies with regard to digital preservation in Germany.

The nestor partners host four working groups, which are open to non-nestor members as well: "Networking and Cooperation" provides a forum for identifying and addressing collective problems in digital preservation, "Preservation of non-textual Media" gathers expertise and best practices from the area of AV and multimedia preservation. Legal experts of the WG "Legal Issues" points to passages of the copyright legislation, which hinder digital preservation, and the just recently established WG "Digital Preservation" discusses how the concept of significant properties can be pragmatically used in practical preservation processes.

A group of now eleven higher education institutions has renewed its [Memorandum of Understanding](#) with nestor to work collaboratively towards a digital preservation curriculum and several e-learning modules. Modules currently under construction comprise an introduction to digital preservation, on formats and data carriers, on metadata generation, and on web archiving. The university partners already deploy some of the developed modules in their academic teaching. The higher education partners also co-organise the nestor summer school; the next one will take place in June 2010.

In 2009, the priority initiative "[Digital Information](#)" of the Alliance of the German Science Organizations gained momentum. Launched in June 2008, it is aimed at the establishment of an integrated digital research environment during the period 2008 - 2012. Archiving of primary research data and a hosting infrastructure for the storage of digital texts are two of the main goals of the initiative. The goals are supported by corresponding calls for proposals for the [Scientific Library Services and Information Systems program of the German Research Foundation](#) (DFG).

Two recently started projects, both funded by DFG, are worth mentioning: The project [LuKII](#) (LOCKSS and kopal Infrastructure and Interoperability) investigates interoperability between the open-source elements of two existing archiving systems (LOCKSS and kopal) in order to combine cost-effective bitstream preservation with an established tool for usability maintenance and format migration.

The project DP4lib (Digital Preservation for Libraries) will advance the organisational and technical further development of the kopal Solution into an integrated service for the temporary long-term archiving of digital information.

### Your View: commentary commentary, questions and debate from readers

Get involved! Comment online at: <http://www.dpconline.org/newsroom/what-s-new/>

### Credits

What's New is a joint publication of DPC and DCC and is compiled by Najla Rettberg. Thanks to contributors. If you would like to feature in What's New or have a contribution to the news or events, then please contact us.

For more information on the DPC see <http://www.dpconline.org/>

For more information on the DCC see <http://www.dcc.ac.uk/>