What’s new ...

Digital Preservation Coalition | Digital Curation Centre Issue 23, 1st March 2010

In this Issue:

- What’s on and What’s New
- Editorial: Preservation and Progress through Harmony (Graham Pryor, Associate Director, DCC)
- Who’s who: sixty second interview with Shane Start, Application Support, The British Library
- One World: iPres 2010 - the Seventh International Conference on the Preservation of Digital Objects, Vienna, September 2010

What’s on:

Digital Preservation Training Programme
4-6 October 2010
http://www.dptp.org/

The DPTP is an intensive 3-day course designed for all those working in institutional information management who are grappling with fundamental issues of digital preservation. It provides the skills and knowledge necessary for institutions to combine organisational and technological perspectives and develop an appropriate response to the challenges that digital preservation needs present. DPTP is operated and organised by the University of London Computer Centre with contributions from leading experts in the field. The next DPTP will take place from 4th-6th of October 2010, at the School of Oriental and African Studies (SOAS), London.

Infrastructure for Research and Information Environment Briefing Day
11 October 2010
http://www.jisc.ac.uk/events/2010/10/researchandiebriefing11oct.aspx

The event will present the various elements of the call and allow for questions. There will be opportunities for networking and sharing advice on bidding for and running JISC projects. The event will also feature one on one sessions with JISC programme managers to explore questions and potential ideas in more detail.

NISO Webinar: It’s Only as Good as the Metadata: Improving OpenURL and Knowledgebase Quality
13 October 2010

This webinar will discuss several efforts in place or underway to ensure the quality of OpenURL data and the knowledgebases that OpenURL links depend on. Is anyone keeping watch over the accuracy and dependability of OpenURLs? Are resource providers and system vendors held accountable for their applications of the OpenURL standard (ANSI/NISO Z39.88)? What can librarians do to bring these problems to the attention of their suppliers? This webinar will address these and other issues of OpenURL accuracy in both theory and practice.

The future of research?
19 October 2010
http://www.jisc.ac.uk/events/2010/10/futureofresearch.aspx

This exciting one day conference will look at the strategic role technologies can play in helping institutions overcome the challenges in supporting the research lifecycle today. Delegates are offered a range of ‘here and now’ advice and guidance and will have plenty of opportunity to discuss, and listen to, key issues within the
sector. Delegates will also have the chance to take part in discussions based around the recommendations of the UUK report ‘The Future of Research’.

SDH 2010 - Supporting the Digital Humanities
19-20 October 2010
http://ztwweb.trans.univie.ac.at/sdh2010/
SDH2010 is the first conference that is jointly organized by the CLARIN and DARIAH initiatives, which are building the European research infrastructure for the humanities and related disciplines. SDH2010 aims to bring together infrastructure providers and users from the communities involved with the two infrastructure initiatives. The conference will consist of a number of topical sessions where providers and users will present and discuss results, obstacles and opportunities for digitally-supported humanities research. Participants will be encouraged to engage with honest assessments of the intellectual problems and practical barriers in an open and constructive atmosphere.

Developing Data Management Expertise in Research Workshop
21 October 2010
http://pekin.cerch.kcl.ac.uk/?page_id=47
This half-day workshop brings together speakers who are actively engaged in supporting the data management to share their experience of working with data creators and managers to embed data management expertise. The workshop will provide an opportunity for attendees to learn from best practice approaches that are beginning to emerge and discuss gaps in current data management provision where further investigation is needed. Speakers will reflect on the challenges faced by data creators in the institution, the strategies that have been adopted to address their needs and the resources that are available for wider use.

RDMF5: Economics of Applying and Sustaining Digital Curation
27 October 2010
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 “Economics of Applying and Sustaining Digital Curation”.

DCC Roadshow: Institutional Challenges in the Data Decade
2-4 November, 2010
http://www.dcc.ac.uk/events/data-curation-roadshows/dcc-roadshow-2010-2011
The DCC is organising a series of inter-linked UK workshops aimed at supporting institutional data management, planning and training. The event will run over 3 days and will provide Institutions with advice and guidance tailored to a range of different roles and responsibilities. The first DCC Roadshow be held 2-4 November 2010 at the Bath Ventures Innovation Centre Carpenter and will be open to participants from HEIs in the south-west region of England. The Road show will be opened by Professor Kevin Edge, Deputy Vice-Chancellor of the University of Bath.

What’s New:

W3C Library Linked Data Incubator Group - Call for Use Cases: Library Linked Data
http://www.w3.org/2005/Incubator/lld/
Are you currently using linked data technology for library-related data, or considering doing it in the near future? If so, please consider filling in the W3C Library Linked Data Incubator Group questionnaire (preferably before October 15th, 2010). The information you provide will be influential in guiding the activities the group will undertake to help increase global interoperability of library data on the Web.
Keeping Research Data Safe (KRDS) Factsheet
This four-page factsheet on the costs and benefits of digital preservation is intended to be suitable for senior managers and others interested in a concise summary of our key findings. It will be relevant to all repositories and institutions holding digital material but of particular interest to anyone responsible for or involved in the long-term management of research data.

WePreserve and Metafor: Team Digital Preservation and the Metafor Common Information Model
http://www.youtube.com/watch?v=76MCRXK4Itc
DPE are delighted to announce the sixth Team Digital Preservation adventure, Team Digital Preservation and the Metafor Common Information Model, is now ready for viewing. Blizzard and his team of chaos causing henchmen have escaped yet again and this time they have stolen all of the Climate Model Data being used to plan and build the Really Big Dam. The beautiful and flexible CIM comes to the rescue and makes a big impression on Digiman!

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

How to set up and run a data service – DPC scholarships available
http://www.dpconline.org/training/leadership-programme
The Digital Preservation Coalition is pleased to support three fully funded scholarships to attend ‘How to set up and run a data service’ at the UK Data Archive in November 2010. Applications are welcomed from DPC members and associates. The scholarship covers tuition fees, course materials, access to online resources, lunch and refreshments. Successful applicants will be asked to help promote the course and the work of the coalition.

DCC Unlocks Open Science
The Research Information Network (RIN) and the National Endowment for Science, Technology and the Arts (NESTA) have published the results of research undertaken by the DCC into the benefits and barriers to using ‘open science’ methods. The investigation aimed to identify what motivates researchers to work (or want to work) in an open manner with regard to their data, results and protocols, and whether advantages are delivered by working in this way.

ESRC releases a new research data policy
http://www.esrc.ac.uk/ESRCInfoCentre/Images/ESRC_Research_Data_Policy_2010_tcm6-37350.pdf
The ESRC has released a new research data policy which draws on the OECD principles to promote open access to publicly-funded research data. The policy is accompanied by detailed implementation guidance and outlines the specific responsibilities assigned to grant applicants, grantholders, the ESRC, and the data service providers it funds.

D-lib Magazine
http://www.dlib.org/
The September/October issue of D-Lib Magazine contains five articles and a conference report. Also in this issue you can find 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 "IN Harmony Sheet Music from Indiana" collection courtesy of the Lilly Library, Indiana University, Bloomington.
CETIS Briefing Paper on the Semantic Web, Linked and Open Data
http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf
This briefing paper provides a high level overview of key concepts relating to the Semantic Web, semantic technologies, linked and open data; along with references to relevant examples and standards. The briefing is intended to provide a starting point for those within the teaching and learning community who may have come across the concept of semantic technologies and the Semantic Web but who do not regard themselves as experts and wish to learn more.

OpenAIRE Guidelines 1.0
These guidelines provide orientation for repository managers to define and implement their local data management policies in compliance with the Open Access demands of the European Commission. Furthermore, they will comply with the technical requirements of the OpenAIRE infrastructure that is being established to support and monitor the implementation of the FP7 OA pilot. OpenAIRE would like to receive comments on the Guidelines, and experiences with the implementation of them.

Editorial:
Preservation and Progress through Harmony
(Graham Pryor, Associate Director, DCC)

An editorial for the Digital Preservation Coalition? Well, fine, I am seriously intrigued by the concept of coalitions and the very practical ideal of progress through harmony, a more than challenging thought in the broader context of these testing times. But preservation, now there’s a word to capture the attention of a sexagenarian. Time was, it was such an inconsequential notion. Back when everything in life was fresh, new and hopelessly unplanned, preparing for inevitable decay had no place in any of it. When one was out on the town as a young man, knowing that the heads of young women might turn – or at least one could reasonably kid oneself that that was the case! - why worry about preservation then? But at sixty-one – admit it, by then you’ve become invisible. Like orphaned data your tags are no longer working, your links have broken and what you do or say could so easily be misinterpreted by someone who’s only just getting into appraisal and selection. Especially today, when standards have changed so dramatically.

Standards. That’s what we need. The application of standards so that everyone can read me again, can understand my codes, and I’m able to maintain those essential transitions between the diverse versions of what is my received epistemology. I’ve just got to curate myself more thoughtfully, that’s all, I must heed the words of the style gurus. Like Shane Watson from the Sunday Times. What was her wise advice last month? Older men shouldn’t wear jeans? Well ok, but proceed with caution: it’s fine if you avoid the Tony Blair smart-casual look, and be sure to go for straight not skinny. Long hair is a definite no-no too, the rock star manqué look is far too much of a cliché; but not too short either, that’s just giving in to being a grown-up! What else is to be avoided? Second-skin leather – now who over thirty can pull that off? And brogues and beige chinos. Boring! Is the list endless? Ah - blouson jackets, that’s the thing, so long as you’re trim. OK, no gain without pain.
Good, no-one’s going to find me semantically challenged now. I’ve installed my very own self-assessment toolkit; now to get out and affirm my value proposition.

OK, I have to admit that I’ve enjoyed plenty of access, use and reuse over the last forty years or so - though I could do with a bit more of the latter now, thank you very much. But my persistent identifiers really have worn a bit thin. When I decided what comprised my ontology all those years ago - how I thought about me, what I meant and what my values are - how did I know that even the persistence of memory would prove so unreliable? So here’s the wake-up call; now’s the time to effect a transformation. Again – yes, it’s essential periodically to re-evaluate, refresh and reconceptualise. And look, the Government is going to abolish compulsory retirement. Now there’s a major new advantage coming from that particular coalition. Yippee, they’re saying I’m still to be regarded as an asset. I can keep on struttin’ my stuff.

So, come on, who’s up for a little data sharing…? (As long as your own infrastructure has been kept up to scratch of course.)

**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…*

**Shane Start, Application Support, The British Library**

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

I work for [The British Library](http://www.bl.uk) based in Boston Spa. My job title is Application Support working for the Technical Services Team supporting and delivering both the [Digital Library System](http://www.bl.uk) and other digital asset management 3rd party software and services.

**Tell us a bit about your organisation**

The British Library is the national library of the United Kingdom and one of the world’s greatest research libraries. It provides world class information services to the academic, business, research and scientific communities and offers unparalleled access to the world’s largest and most comprehensive research collection. The Library’s collection has developed over 250 years and exceeds 150 million separate items representing every age of written civilisation. It includes: books, journals, manuscripts, maps, stamps, music, patents, newspapers and sound recordings in all written and spoken languages.

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

I started working in digital preservation when I joined The British Library in 2003, working on the predecessor to the [Digital Library Programme](http://www.bl.uk) (DLP) which started in 2008. The Digital Library System (DLS, which the programme has designed and implemented) is a long-term preservation system for digital assets. The DLS is a distributed solution that has over 107 TBs of data and contains over 3.5 million distinct digital objects.

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

There are lots of [Digital Projects](http://www.bl.uk) that are happening within The British Library. I am currently focused on projects related to the Digital Library Programme. Some of the projects that we are currently working on are:

---

... because good research needs good data
I am also working with a 3rd party product called DigiTool (supplied by Ex Libris) which is used for Voluntary Deposit of non-print publications.

We are currently looking to extend the use of DigiTool for digitisation projects, making use of DigiTool to ingest vulnerable digital collections into an interim store prior to ingest to the DLS. A small selection of the content that will be ingested under Fast Track includes:

- Cotton Manuscripts
- Maps

What are the challenges of digital preservation for data services such as yours?
The management and long term preservation of digital objects in “perpetuity” presents many unique challenges and opportunities.

Digital preservation is very much in its infancy. The contrast for digital content versus physical content is enormous. I sometimes wonder what future generations will make of this age whereby we are just really starting on the journey of digital preservation. I think the global community is starting to realise the implications that digital over physical records will have and the implications for future generations with respect to the digital dark age.

I think one of the great challenges is the preservation of historic digital records in an electronic era where change and speed is perhaps valued more highly than conservation and longevity.

There are many elements that we must acknowledge and manage. A very small example is:

- The integrity and authenticity of digital content must be planned and executed to ensure the digital objects are what they claim to be, that they are complete and more importantly that they have not been altered.
- To take into account the usual cost implications including people, accommodation, hardware, software and external 3rd party costs.
- We must also consider the selection and prioritisation of digital content; the sheer volume of digital content can be overwhelming! I think the key is selecting which digital resources to preserve and which should not be preserved.
- The file formats that we accept today will in time need to be migrated to new formats. The digital objects will require frequent refreshing and recopying to new storage. The translation into new formats will have a price associated with it; both financially and in the production of imperfect copies to that of the original.
- Rights management and access control to digital content.

What projects would you like to work on in the future?
We are currently in the early stages of integrating PLANETS with the Digital Library System. This will be an interesting project allowing us to use the services and tools to help ensure long term preservation and access to our digital cultural heritage and digital assets.

What sort of partnerships would you like to develop?
The digital information age presents many opportunities for partnerships to be formed. The British Library has aligned itself with a number of strategic partnerships some of which include Microsoft, Amazon, UK Research
Reserve and more recently Bright Solid. The Library has also partnered with a range of firms (Microsoft, HP, Haworth) and the JISC for its forthcoming exhibition on digital research, Growing Knowledge: The Evolution of Research (12 October 2010 – 16 July 2011).

I think the British Library Business & IP Centre is an excellent resource for both business and entrepreneurs – there have already been suggestions that it could form a model for regional business support centres and partnerships with regional organisations.

If just one tool or standard could be brought into existence that would make your job easier, what would it be?
A single unified toolset to perform all digital asset management functions including the ingest workflows, the metadata extraction for all known file formats, preservation activities (migration and emulation), authenticity checking and resolution and also resource discovery. This toolset would provide technical and business management reporting capabilities. It would be a highly available fault tolerant service that has a focus on digital preservation. Is this too much?

If you could save for perpetuity just one digital file, what would it be?
Ok, there are many important digital documents that exist in this world. However, I am going to choose the song Songbird by Eva Cassidy. Although this is by no means my favourite song, it is very sentimental for me.

Finally, where can we contact you or find out about your work?
The British Library web site is www.bl.uk. Further information on the current digital activities within the British Library can be seen on the Digital Library Programme pages within the web site.

My email address is shane.start@bl.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 ...

William Kilbride, Executive Director, DPC

iPres 2010 - the Seventh International Conference on the Preservation of Digital Objects - is a somewhat misleading title for the recent gathering of the digital preservation community in Vienna in late September. For one thing, discussions betrayed a wider range of interests than the functional maintenance of obsolete bits: it is a forward looking event resolutely interested in the future. And iPres is not so much a conference as a concurrence: a six day festival in which iPres is the locus of a dozen or more workshops, tutorials, meetings and exhibitions. There's a lot to pack in.

Hence my own rather late entry this year, arriving after the first batch of papers, not to mention an evening of ballroom dancing in Vienna's stately Rathaus. Tony Hey's opening contribution – which I accessed via the book contained in the conference pack the - focussed attention on the role of digital preservation in scientific research, and the wide and beneficial impact which the sometimes esoteric needs of the scientific community can have on the wider preservation community. Tony Hey, now vice president of Microsoft Research and previously the motive force behind much of the UK's E-Science infrastructure speaks with distinct authority on the topics of scholarship, research data and its management. Marieke Guy quotes her thoughts on the future of librarianship; by quoting General Shineski 'If you don't like change you'll like irrelevance even less'. Preservation should be about the future in more ways than one.
Day two began with Pat Manson of the EC's Directorate General for Information Society and Media. An opposite error triggered a thoughtful analysis of digital preservation as it has developed in the last decade and as it will continue to develop in the next: she had intended to talk about the 'evolving landscape' but found herself talking about the 'involveing landscape' instead. But, she mused, as far as digital preservation is concerned involvement and evolution elide. It is too important to be left to researchers and needed to engage a much wider number of partners. Scalability, efficiency and automation will be required if we are to keep pace with the scale, complexity and demand for digital data, themes which will be thoroughly explored in the Commission's next set of calls. Fundamental to this will be a determined effort to consolidate gains and to extend collaboration across disciplines, sectors and borders.

Epitomising the desire to consolidate gains through collaboration, my highlight of the next session was Jen Mitcham's presentation of the Archaeology Data Service's efforts to update their Guides to Good Practice. This was part of a session where presenters described tools that they had developed to help them along the way – partly telling people about services they could use but also looking for help and advice from peers. Involvement and evolution at the same time.

Three lunchtime meetings followed before I took the stage for the 'Green Digital Preservation' panel. Neil Grindley had asked me to examine the politics of digital preservation from a green perspective. It's hardly surprising to say that politicians don't pivot to green issues when you talk about digital preservation: they pick up on privacy and economic issues if they pick up anything at all. But nor has the development of green information management, represented on the panel vicariously by Diane MacDonald of Strathclyde University, been able to make much of a challenge in any case. One gets the impression that it's not mature enough yet to offer the sort of quantifiable details that will make its case. It is true that the electricity required to keep large quantities of data on spinning disks is a drain on financial and environmental resources. But David Rosenthal made a very clear case that preservation is only a very small part of the costs of data storage and therefore that if we wanted to make progress on reducing the environmental impact we should look more closely at how storage solutions are engineered and procured. Kris Carpenter Negulescu pointed out that increases in capacity means that energy costs at Internet Archive had not grown in line with data storage: historically speaking, data storage and energy consumption have not followed a simple ratio. As I see it green information management has more in common with digital preservation than meets the eye. I'm not aware that anyone in the digital preservation has ever advocated useless proliferation as a serious preservation strategy: on the contrary, well implemented preservation enables confident deletion. More subtly, Malcolm Todd's minimum redundancy paradigm reminds us that the ideal data format contains no unnecessary code, thus clarifying to an infinite degree the significant properties of the data held. But the preservation community is often times left to deal with bloated and inelegant data formats which have been produced with little concern for size. We may be tough on data and we must be tougher on the causes of data.

From here John Kunze sped us through a tumult of lightning talks – a dozen or so three minute presentations with short questions. By peculiar circumstance the lightning session coincided perfectly with ICON's 1400BST embargo on announcing the shortlist for the Conservation Awards. I didn't catch many of the lightning talks as I hurriedly assembled some slides about the shortlist for the Digital Preservation Award: but it was a great way to end the session and the whooping during my presentation was a first.
A pre-arranged meeting over coffee left just enough time to take in one more session. The highlights for me were around preserving web data including Marieke Guy’s work on preserving blogs in the cloud and Martha Anderson’s elegant and compelling presentation on why Twitter matters, and why therefore the Library of Congress has decided to take this on, and why it is a lot more difficult than it first appears. A thoughtful post-paper discussion caused us to consider the ineffectual nature of national boundaries in a networked interdependent world.

I had intended a short walk to drop off my bags but I will confess to getting completely lost, missing much of the evening reception at the exquisite Prunksaal in the Hofburg. I arrived shortly after a bombshell delivered by phone: a close friend’s diagnosis with multiple sclerosis. Tony Hey and Pat Manson’s keynote presentations came into sharp focus. Research seems more urgent now. Compassion and courage even more so.

The morning of day three saw me in the case studies session with Eld Zierau describing work at the Royal Library in Denmark and Helen Hockx-Yu describing issues with archiving streaming media embedded within web resources for the British Library. Both brought to mind discussions at the DPC meeting on file formats in December 2009 where one shrewd commentator observed that web harvesting made sense when web pages were precisely that. But now that web sites are complex data streams with multiple sources constantly in motion and constantly personalised so the topics which Helen raised are likely to presage a coming shift in attitudes to preservation of more complex objects.

Lunch on day three - actually a meeting got in the way of lunch - was the official end of the conference. By afternoon the conference had moved across Karlsplatz to the National Library of Austria in the Hofburg complex. Three concurrent workshops, including the Premis Implementation Fair and the International Web Archiving Workshop as well as a meeting of those involved in training. The latter had my attention, led by Cal Lee and colleagues at the University of Carolina at Chapel Hill. A diverse and international group quickly found common ground around complicated issues of helping employers find staff that will be able to close the constantly reported digital preservation skills gap. A formal shared curriculum might be hard to accomplish and quick wins, such as more effective sharing of examples and other resources are within immediate reach. It’s hard to believe that a postgraduate curriculum is likely to make a difference at a time when employers are simply not recruiting new staff. So development of the existing workforce and understanding employers’ needs also emerged as high priorities, as did the need to establish better labour market intelligence. The evening closed with a reception introducing the Open Planets Foundation.

Day four saw more workshops. The Skills Gap and International Web Archiving Workshops continued while another group met to examine collaboration and co-operation within digital preservation. This event, co-sponsored by NESTOR, DPC, NCDD NDIIPP and APA, examined what works - and what doesn’t work - when you try to get organisations to work together. It rapidly became clear that many different activities in many countries had experiences to share. This workshop was in fact the formalisation of a standing dinner date between the co-hosts. We knew we planned to meet at iPres and thought it would be fun to open the meeting to others. We didn’t anticipate that quite so many people would want to join us. It also became clear that while it’s incredibly useful to have a national body configured to tackle legal and political issues so there are international issues that could well be addressed. It’s a group that therefore needs to meet again and most likely with more participants next time. It’s striking just how quickly our community has grown in the last decade. The workshop met under the title ‘Greater than the Sum of our Parts’, which Abbie Potter of the Library of Congress noted almost contains the acronym ‘soup’. The metaphor could get out of hand but there was agreement that a multiplicity of ingredients thoughtfully preparedly mixed and given sufficient time would not only make good soup but was also the recipe for ongoing collaboration.

... because good research needs good data
Day five saw a meeting of International Internet Preservation Coalition to discuss the configuration and development of collecting policies. This has been an active topic for the DPC Web Archiving and Preservation Task Force and the IIPC audience were clearly interested in the energy and openness with which DPC members have engaged in this topic. It rapidly became clear that the local political structures matter a lot in such discussions. The relative clarity of governance in Denmark for example, is surely envied by planners in the UK: but the multiplicity of agencies involved here points to a shared if uneven responsibility and a lively if fragmented desire to preserve which is its own strength. Day five also brought a final bonus - the first lunch of the week without a meeting or planned catch up. At last time to reflect on a successful week and to enjoy the company of colleagues for its own sake.

iPres is about more than preservation and is much more than a conference. Next year it moves to Tsukuba Japan. The International Conference on the Preservation of Digital Objects is more than a conference and about a lot more than digital objects: but its international character is hardly in question.

Credits
What’s New is a joint publication of DPC and DCC and is compiled by Kirsten Riley. 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/

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