

What's new ...

Digital Preservation Coalition | Digital Curation Centre Issue 27, 1st July 2010

In this issue:

- **What's on and What's New**
- **Editorial:** Fragmentation and Success (Kevin Ashley, DCC Director)
- **Who's who:** sixty second interview with Tracey Powell, National Archives in Kew
- **One World:** Bjarne Andersen, State and University Library, Denmark

What's on:

Managing Data in Difficult Times: policies, strategies, technologies and infrastructure to manage research and teaching data in a fast changing technological and economic environment

1-2 July 2010

<http://www.ukoln.ac.uk/events/jisc-cni-2010/>

JISC and the Coalition for Networked Information (CNI) are proud to announce the 8th International Meeting to bring together experts from the United States, Europe and the United Kingdom. Parallel sessions will explore and contrast major developments that are happening on both sides of the Atlantic.

TEI @ Oxford Summer School 2010

12-14 July 2010

<http://tei.oucs.ox.ac.uk/Oxford/2010-07-oxford/>

The TEI @ Oxford Summer School is a three day course introducing the recommendations of the Text Encoding Initiative (TEI) for encoding of digital text. It combines in-depth coverage of the latest version of the TEI Recommendations for the encoding of digital text with practical workshops on related technologies. It includes an introduction to mark-up, explanations of the TEI Guidelines, and approaches to publishing TEI texts. Practical exercises expose you hands-on experience of a wide range of TEI customisation, editing, and publication.

Designed to Last: Preserving Computer Aided Design

16 July 2010

<http://www.dpconline.org/details/16-computer-aided-design.html?xref=16>

DPC Members' Briefing Day. Computer aided design and modelling (CAD) have revolutionised design and drawing. CAD now underpins work in fields as diverse as animation, architecture, manufacturing, engineering and heritage management and it intersects with numerous applications such as GIS, simulation, gaming and 3d modelling. The underlying geometry of a CAD model or drawing represents the intellectual property of a significant and sophisticated design process, while the ongoing co-ordination of the underlying geometry with annotation and description brings added value.

1st Karlsruhe Summer School on Service Research

18-22 July 2010

<http://www.service-summer.org>

Through lectures, tutorials and social events, the Summer School will provide a forum for participants to discuss and learn about Service Research. Participants can benefit from the expertise of the lecturers and share experience with fellow attendees. Furthermore, the Summer School will foster interdisciplinary research and collaboration opportunities among international students and researchers interested in the disparate fields within Service Research.

Sudamih Workshop: Data Management Training in the Humanities

22 July 2010

http://sudamih.oucs.ox.ac.uk/training_workshop.xml

This half-day workshop will consider how institutions might meet growing requirements for training in the management of research data within the humanities. The aim is to learn more about research data management training already taking place at UK universities, plans for such training, relevant scoping studies, and related experiences.

Desktop Search Workshop: Understanding, Supporting, and Evaluating Personal Data Search

23 July 2010

<http://www.cdvp.dcu.ie/DS2010/>

This SIGIR 2010 workshop will bring together academics and industrial practitioners interested in desktop search with the goal of fostering collaborations and addressing the challenges faced in this area. The workshop will be structured to encourage group discussion and active collaboration among attendees. Participation is encouraged from people in the fields of information retrieval, personal information management, natural language processing, human-computer interaction, and related areas.

Decoding the Digital: A Common Language for Preservation

27 July 2010

<http://www.dpconline.org/details/18-decoding-the-digital.html?xref=18>

This one day conference, organised jointly by the DPC and the British Library Preservation Advisory Centre seeks to build on the communication between librarians, archivists, curators, conservators and digital preservation managers and establish a shared understanding of the issues surrounding the preservation of digital collections.

Repository Fringe 2010

2-3 September 2010

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

Enjoying two successful years, Repository Fringe is back for the third time... and this year we have moved to the end of the Edinburgh Festival, and to the National eScience Centre. Always innovative, and always looking forward, this year we are looking at the OPEN: Open Data; Open Access; Open Learning; Open Knowledge; Open Content; etc...The programme for the event is, as ever, created by the participants, however there will be a focus on group discussion rather than long talks.

UK e-Science All Hands Meeting (AHM 2010)

13-16 September 2010

<http://www.allhands.org.uk/>

The meeting provides a forum in which information on e-Science projects from all disciplines can be communicated and where the capabilities being developed within projects can be demonstrated.

Preservation of Digital Objects (IPRES 2010)

19-24 September 2010

<http://www.ifs.tuwien.ac.at/dp/ipres2010>

The Austrian National Library and the Vienna University of Technology are pleased to host the International Conference on Preservation of Digital Objects (iPRES2010) in Vienna in September 2010. iPRES2010 will be the seventh in the series of annual international conferences that bring together researchers and practitioners from around the world to explore the latest trends, innovations, and practices in preserving our digital heritage.

Spanning the Boundaries of Digital Curation Education, iPRES 2010 Workshop

22-23 September 2010

<http://www.ifs.tuwien.ac.at/dp/ipres2010/program.html>

The primary goal of this workshop is to facilitate the sharing of information and ideas across the boundaries of professional education (national, institutional and educational level). While the number of digital curation educational offerings across the globe has increased significantly in recent years, there are still relatively scarce human resources for developing and implementing educational content. The workshop will explore potential areas of collaboration and will include short summary talks about current educational activities by workshop participants.

PREMIS Events, iPres 2010 Workshop

<http://www.ifs.tuwien.ac.at/dp/ipres2010/program.html>

19-22 September 2010

A PREMIS Tutorial will be held on Sunday September 19, 2010. This full-day educational event will provide an introduction and walk-through of the PREMIS Data Dictionary for Preservation Metadata. A PREMIS Implementation Fair will also take place on September 22, 2010 to promote discussion and share experiences involving implementing PREMIS.

What's New:

Personal Membership of DPC – Application Pack Now Available

<http://www.dpconline.org/members/personal-membership.html>

The DPC has introduced a new class of membership which will allow sole traders, consultants and recently graduated students to participate in the work of the Coalition. Personal membership has been introduced for a trial period of 12 months from 1st August 2010. In the first year, personal membership will be capped at 25 members allocated on a first-come-first-served basis.

New RIN Guide - Government and research policy in the UK: an introduction

<http://www.rin.ac.uk/node/619>

The Research Information Network have published a new guide which describes the nature, roles and responsibilities of the different Government bodies involved in research policy and funding, and the relationships between them.

New Web Archiving Tools Launched

<http://code.google.com/p/liwa-technologies/>

New Web Archiving technology for dealing with streaming video has been developed by LiWA, a European project led by the University of Hannover.

New Repositories Support Project Blog

<http://rspproject.wordpress.com/>

The blog features updates on the RSP team activities, overviews of events that the team have participated in and covers issues of importance to repository managers.

NISO Publishes Special Digital Preservation Issue of Information Standards Quarterly - Selected articles now available for free download

www.niso.org/publications/isq/2010/v22no2/

The National Information Standards Organization (NISO) announces the publication of the Spring 2010 issue of the Information Standards Quarterly (ISQ) magazine with a special issue theme of Digital Preservation. ISQ

Guest Content Editor, Priscilla Caplan, Assistant Director for Digital Library Services, Florida Center for Library Automation, has compiled a stellar set of articles on the topic-authored by experts in the field from the U.S., U.K, Canada, and New Zealand.

Visualization of the Metadata Universe

<http://www.dlib.indiana.edu/~jenlrile/metadatamap>

The sheer number of metadata standards in the cultural heritage sector is overwhelming, and their inter-relationships further complicate the situation. A new resource, Seeing Standards: A Visualization of the Metadata Universe is intended to assist planners with the selection and implementation of metadata standards.

Portico Archive Reaches Milestone in Digital Preservation

<http://www.portico.org/digital-preservation/>

Portico is pleased to announce that 110 publishers, representing more than 2,000 professional and scholarly societies, are now participating in the Portico archive. Furthermore, nearly 15 million articles are now safely preserved in the Portico archive.

JISC MRD Workshop slides now available

<http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmevents/programmemeetingmay.aspx>

The slides from presentations at the JISC Managing Research Data Programme Workshop, held in Manchester on 17-18 May 2010 are now available. The purpose of the workshop was to consider how best to meet the challenge of managing research data in UK Universities. Participants were informed of progress made under the JISCMRD Programme to examine researchers' data management requirements and improve planning in a number of academic and institutional contexts.

Austrian National Library and Google cooperate in digitisation

<http://www.onb.ac.at/ev/austrianbooksonline/>

As one of the first national libraries worldwide the Austrian National Library will digitise its complete holdings of public domain books from the 16th to the 19th century. This project will make one of the five internationally most important historical book collections available online.

LoC Digital Preservation Newsletter

<http://www.digitalpreservation.gov/news/newsletter/201006.pdf>

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

Ireland's National Portal for Open Access to Research Goes Live

<http://www.rian.ie>

Ireland's national portal for Open Access to Irish published research (RIAN) will act as a single point of access to national research output, and contains content harvested from the institutional repositories of the seven Irish Universities and Dublin Institute of Technology. RIAN will significantly increase the visibility and impact of Irish research and will expand to harvest content from other Irish Open Access providers as the service develops.

New report from The National Archives on digital continuity risk in large local authorities in England

<http://www.nationalarchives.gov.uk/recordsmanagement/publications.htm>.

Digital Continuity requires strategic alignment, senior understanding and commitment and effective working relationships between Senior Information Risk Owners, ICT Managers, information assurance and governance officers and those responsible for business processes as well as records and information management. This report is not part of the central government-funded Digital Continuity project but was commissioned to provide an evidential basis for future dissemination of that project's findings to the wider public sector.

Digital Preservation Award 2010: open for nominations and judges announced

<http://www.dpconline.org/advocacy/2010-digital-preservation-award.html>

The Digital Preservation Award is awarded to the nominee who, in the eyes of the judges, best demonstrates excellence and innovation that will help to ensure our digital memory is available tomorrow. The judging panel for the 2010 awards include Kevin Ashley, Director, Digital Curation Centre, University of Edinburgh; Adrian Brown, Assistant Clerk of the Record, Parliamentary Archives; William Kilbride, Executive Director, Digital Preservation Coalition; Pip Laurenson, Head of Time-based Media Conservation, Tate; Zoe Lock, Lead Technologist for ICT, The Technology Strategy Board; Eefke Smit, Director for Standards and Technology, Scientific, Technical and Medical Publishers' Association; Dave Thompson, Digital Curator, The Wellcome Library; Matthew Woolard, Director Designate, the UK Data Archive; and Richard Wright, Senior Research Engineer, BBC.

SPARC Open Access Newsletter

<http://www.earlham.edu/~peters/fos/newsletter/06-02-10.htm>

This June 2010 issue takes a close look at the unanimous faculty votes for institutional OA policies.

Europeana Publishes First White Paper: Knowledge = Information In Context

<http://version1.europeana.eu/web/europeana-project/whitepapers>

Europeana's first White Paper looks at the key role linked data will play in Europeana's development and in helping Europe's citizens make connections between existing knowledge to achieve new cultural and scientific developments. Without linked data, Europeana could be seen as a simple collection of digital objects. With linked data, the potential is far greater, as the author of the white paper, Prof. Stefan Gradmann, explains.

Editorial:

Editorial: Fragmentation and Success

(Kevin Ashley, DCC Director)

One of the reasons for producing the monthly "What's New" is that it can be challenging keeping up with developments in a field that continues to spread its boundaries. An increasing range of new entrants to the field bring new insights and perspectives, but also sadly sometimes repeat the work that has already been done by others. Different professional groups have come to an awareness of the need to think about digital preservation at different times, and have taken differing amounts of time to turn concern into action. In some cases the action preceded considered thought. For instance, part of my first job working for the medical research council included the rescue of old digital research data. I learnt a great deal from older colleagues who had need to do this many times before, but none of us used phrases like 'digital preservation' to talk about what we did. When I first began to consciously use the term in the mid-1990s, there were few professional events one could attend to discuss such matters. The few that existed, such as the [DLM-Forum](#) which first met in 1996, gave one the feeling that the world's entire expertise was contained in the one room; and that probably wasn't too far from the truth. It would be impossible to do that today, and increasing specialisation probably means that many of those involved would not recognise others as engaged in the same field.



One of the resources that I found invaluable in those (for me) early days was [Michael Day's bibliography on the preservation of electronic information](#), something which Michael stopped updating about 10 years ago because the proliferation of publications now meant it was impossible to keep up to date. The bibliography

(and many of the works it identifies) still makes useful and interesting reading. The early work of those such as Bell, Dollar and Fishbein show that the problems that machine-readable information might present were apparent to some in the 1970s. But for almost three decades thereafter, it was feasible for one person to have read and digested all of the literature that had been published in the field, whether it emerged from the world of libraries, archives, computer science or industry (to name just a few sources.) By the end of the century that was becoming an impossible task; the fragmentation we now see between electronic records management, digital libraries, research data curation and the preservation of computer games (again, to give just a few examples) was already under way.

The fragmentation brings risks with it, but it is also necessary to some extent. We have moved from a time when most discussion was simply about acknowledging that there was a problem and attempting to describe it, to one where a great deal of real activity is taking place on a large and small scale, in many different types of organisations. We need larger numbers of people who are trained to do specific jobs and do them well, rather than theorising about what those jobs might be. In such an environment, it's important that (say) virtual world preservers can talk to other virtual world preservers about the issues that arise in their day-to-day tasks and about how to tackle them, without worrying about whether this does or does not have any bearing on the preservation and curation of meteorological satellite data. But it's also important that we have some people who are taking a broader view and ensuring that work is not replicated between domains and that information is effectively shared.

'Different professional groups have come to an awareness of the need to think about digital preservation at different times, and have taken differing amounts of time to turn concern into action.'

"What's New" is one tool to help deal with this task of information exchange. As time goes by, I find it more and more valuable. But the Digital Preservation Coalition itself is another key factor in preventing the fragmentation of knowledge. Many of its member organisations have knowledge dissemination and exchange as part of their remit. At the [DCC](#), for instance, we try to ensure that those active in research data management are kept aware of each other's activities and also of developments elsewhere that may be relevant to them. We can also bring lessons learnt in that field to a wider audience, whether it be in local or national government, national libraries, information science or industry, where the lessons are meaningful outside that particular field. The DPC has a key role to play in helping each of its partners maintain that exchange of knowledge and practical advice. It's why we all continue with our membership. As we enter difficult times, when conference attendance or journal subscriptions may be even harder to maintain, the DPC's role is even more important to us.

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 ...

Tracey Powell

The National Archives

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

I work at The National Archives in Kew, as the Digital Preservation and Resource Discovery Business Manager.

Tell us a bit about your organisation

The National Archives is a government department and an executive agency of the Ministry of Justice (MoJ). As the official archives of the UK government, we care for, make available and 'bring alive' a vast collection of over 1000 years of historical records, including the treasured Domesday Book. Not only safeguarding historical



information, we also manage current digital information and devise new technological solutions for keeping government records readable now and in the future. We provide world class research facilities and expert advice, publish all UK legislation and official publications, and are a leading advocate for the archive sector. The role of the Digital Preservation team is to lead the permanent preservation of the born-digital records that come to The National Archives. We have a number of tools and services that help us to do this, such as the Digital Repository Transfer System, which deals with ingest and preservation of digital records, and we produce presentation copies of the transferred records. We also provide the PRONOM Technical Registry, which offers definitive information about file formats, software products and other technical components required to support the active preservation of digital information.

How did you end up in digital preservation?

I joined the organisation almost three years ago, working in the IT Strategy Department, and took up an opportunity as the Business Manager in the Digital Preservation Department in early 2009. I joined initially because I had experience in, and enjoy, the variety of responsibility that a role like that provides, for example budget management and business planning. However, I quickly found that issues around digital preservation are really interesting, and it's been great to have the chance to develop expertise in this area.

What projects are you working on at the moment?

We've just embarked on a project to enable our PRONOM service to use linked data. This will allow us to expose and share the data in PRONOM in a format that is easier for people to reuse it. At a later stage we will be looking to add functionality so that we can update PRONOM from other sources, to help us to provide a more comprehensive technical registry for the digital preservation community. I think we'll learn a lot along the way about how else to open up the data we hold, and how to link to other data sources, to augment the services we offer around The National Archives' digital records.

The National Archives has also spent the last four years as a consortium partner of the Planets project, which developed practical services and tools to help ensure long-term access to digital collections. We led work on the Preservation Characterisation sub-project, to develop methodologies, tools and services for characterising the significant properties of digital objects, to help in planning and validating preservation actions. We contributed to other parts of the project too, for example in developing the DROID identification tool, to automatically profile the file formats of digital collections. Although the Planets project ended in May 2010, we've spent further time since in preparing for a final review by the EC in July, so it feels to me like the project is still going!

What are the challenges of digital preservation for data services such as yours?

The volume of material we expect to receive – we know we've got lots of data coming our way, and we need to enable our systems to ingest this data on a much bigger scale. This means simplifying our processes to make it easier to manage such large volumes of data. I'd also like us to develop more file format signatures, and to create and evaluate more migration pathways for converting between formats.

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

Well, I'm really looking forward to producing a linked data version of PRONOM, so I'd say that, if it wasn't already on the horizon! As well as that, I'd like us to find a generic solution for preserving emails and websites. On the non-technical side, I think as a community we need to build a better understanding of how the immediate cost of preserving digital information is swept away by the future value of those digital assets.

What sort of partnerships would you like to develop?

We've already got good working relationships with lots of other institutions, particularly through the Planets project – organisations such as NANETH (the Nationaal Archief of the Netherlands) and UzK (University of Cologne). We're also involved in discussions to create a Unified Digital Formats Registry, building on PRONOM

and the Global Digital Format Registry project, so are having productive dialogue with Library and Archives Canada and with Harvard, amongst others. So, we'd like to keep these partnerships going, and build others wherever there's an opportunity to develop simpler and larger-scale methods for preserving digital records.

If just one tool or standard could be brought into existence that would make your job easier, what would it be?

Ah, easy - I mentioned this one before - a robust tool for preserving emails.

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

Wikipedia! A huge 20 Gigabyte dump of the whole site – a rather large file, admittedly, but an awesome database containing heaps of fascinating knowledge. OK, so the provenance of some of the contents may be uncertain, and the database couldn't be updated in the way that it is at the moment, but even that in itself would give a good account of human knowledge as it was at a particular point in 2010. What a great reminder of what's possible when people collaborate.

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

You can contact the team by email digitalpreservation@nationalarchives.gov.uk, and read more about The National Archives at: www.nationalarchives.gov.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 ...

Bjarne Andersen, Head of Digital Resources, State and University Library, Denmark

This report gives a brief overview of digital preservation activities in Denmark at a national level and primarily within cultural heritage institutions. The core set of institutions working together on many aspects of digital preservation includes the Royal Library (<http://www.kb.dk/en/index.html>), the State Archives (<http://www.sa.dk/content/us/>), the Danish Film Institute (<http://www.dfi.dk/English.aspx>) and the State and University Library (<http://en.statsbiblioteket.dk/>) which I represent.



Before jumping into an overview of the current activities I would like to briefly introduce the main parts of the digital cultural heritage collections in Denmark. In 1998 the Legal Deposit Law underwent a revision and from then on static web material had to be saved for the future. This was in practice the first legal deposit of digital material. The two libraries with national obligations had both made pilot digitization projects already but only on a smaller scale. The first legal deposit of web material was collected upon request from publishers who had to register their publications for deposit. Compared to modern day web this meant that only very few publications were archived and from 1998 to 2005 only 500 Gigabytes worth of publications were archived in total.

In 2005 the Legal Deposit Law was revised again. This time the entire Danish web was included and the establishment and funding of netarchive.dk became a reality. Netarchive.dk is a virtual organization driven by the Royal Library and the State and University Library. Furthermore the Legal Deposit Law also covered online capturing and recording of all Danish radio stations and television channels. The national media archive was placed at the State and University Library. During the last five years this has generated rather huge digital collections. The web archiving program has passed 145 Terabytes mainly because of the domain crawling

completed two to three times a year, and the national radio and television archive has already passed 500 Terabytes of digital video and audio recordings.

Together with large digitization projects at all the major cultural heritage institutions in Denmark the legal deposit digital archives have developed a natural need for digital preservation solutions.

The Royal Library and the State and University Library have both been active partners in Planets, a project co-funded by the EC. The Danish libraries have been involved in many activities from user studies within research communities, research in "Context Aware Digital Objects" to practical tool building and wrapping primarily of tools for characterization and migration actions. Furthermore the libraries have developed a Planets connector for the Fedora Commons repository system used in a currently ongoing internal development project at the State and University Library. The Fedora connector allows repositories based on Fedora Commons to hook up directly to the Planets interoperability framework thus bringing preservation functionality to the repository system.

Towards the end of Planets a number of organizations founded Open Planets Foundation (OPF – see the last issue of What's New) and both national libraries in Denmark are founding members of the organization. Furthermore the State Archives has joined OPF, so Denmark is rather well represented, at least compared to the size of our country. We believe in and hope for the sustainability of the good results of Planets and believe that OPF is exactly the right framework for ensuring that.

Aside from the work done and the knowledge gained in Planets State and University Library participated in another EC-funded project too; Digital Preservation Europe (DPE). In this project we investigated many aspects of digital preservation – from legal issues of European digital repositories to the creation of the DRAMBORA audit framework for digital repositories.

During 2009 the two libraries and the State Archives ran a strategy project funded by the Danish Ministry of Culture. This project aimed at defining a strategy and architecture for a nationwide bit preservation facility. The project ended in December 2009 with a one day conference where the results were presented to a broad range of cultural heritage institutions in Denmark as well as invited people from Scandinavia. We had the pleasure of having David Rosenthal from LOCKSS give an introduction to the challenges of preserving bit streams over time.

In late 2009 the Danish Ministry of Culture decided to fund three different projects under the umbrella of digital preservation. All three projects have now started and will run till the end of 2011.

The first project is the actual implementation and establishment of the nationwide bit preservation facility. During the project phase the work will be carried out in collaboration between the Royal Library, the State and University Library and the State Archives but it is possible that other institutions or solution providers might take active part in the setup, once the project switches into a regular service from 2012. The project builds on the results of the earlier strategy project as well as all the knowledge accumulated within the participating institutions by running large storage facilities for several years. The final setup will be offered to other institutions on a commercial and/or non-for-profit basis. The project was disseminated and awarded for a paper at this year's ICDL¹.

The second ongoing project is about cost models for digital preservation. The project builds on the British LIFE2 projects and tries to use the results of LIFE2 in a Danish setting and take the cost model of LIFE2 even further. This project is also collaboration between libraries and archives.

¹ "To be published in proceedings for ICDL 2010 (International Conference on Digital Libraries) 23-26/2 2010 New Delhi"

The third and last ongoing national project aims at building a national knowledge base about digital preservation. A web portal (www.digitalbevaring.dk) will be launched in its first version in November 2010. This knowledge base will collect and disseminate knowledge and information about digital preservation to a Danish audience – primarily within cultural heritage institutions but also to other bodies within Denmark interested in or faced with the challenges of digital preservation and limited in-house expertise. This project will naturally collate international research and results and present these in a condensed form in Danish as well as present the work and experience of the national institutions working with the topic. As with the first two projects both national libraries and the State Archives are involved together with the Danish Film Institute. Since Denmark is a small country we are always interested in sharing knowledge, experience, best practices, tools, and systems so the systems or tools technically developed by us are normally released as open source as we have done with many things in the past – e.g. the web archiving tool set: NetarchiveSuite (<http://netarchive.dk/suite>).

We expect to disseminate results from the national projects on relevant conferences in 2010 and 2011 as well as through other relevant channels – e.g. mailing lists. For questions, comments or request for more information please feel free to contact me at bj@statsbiblioteket.dk

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/>