

What's new ...

Digital Preservation Coalition | Digital Curation Centre Issue 31, 1st November 2010

In this issue:

- **What's on and What's New**
- **Editorial:** Here Comes the Tide (William Kilbride, DPC Executive Director)
- **Who's who:** sixty second interview with Laura Molloy, HATII, University of Glasgow
- **One World:** Dr. Dinesh Katre, Centre for Development of Advanced Computing (C-DAC)

What's on:

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.

Managing your Digital Image Workflow

9th November 2010

<http://bit.ly/9CadR>

Is your collection of images disorganised and giving you a headache? Are you drowning in digital images and unable to find the images you need, when you need them? Do you find images you want to use but have no idea who owns the copyright? If so, this course will help you get your collection back on track.

JISC Advance - Digital Media and Copyright

12 November 2010

<http://www.jiscdigitalmedia.ac.uk/blog/entry/new-copyright-and-digital-media-seminar-12th-nov-london/>

This seminar aims to offer practical approaches and considerations for using digital media in teaching and learning, digitising digital media and digital rights management. It is aimed at those working in developing institutional policy, disseminating copyright information to staff, and those working with digital media on a day-to-day basis.

Open Planets Foundation's Practitioners' Workshop and Developers' Hackathon

15-17 November 2010

<http://www.openplanetsfoundation.org/november-workshop>

On Day 1, digital preservation users, practitioners and developers are invited to join us to explore the technical challenges organisations face when preserving digital content and participate in discussions to define new and existing requirements for OPF's products and services. Days 2 and 3 provide an opportunity for developers to meet face-to-face with the creators of the Planets tools and services to address practitioners' requirements and influence the direction of OPF's product roadmap.

Managing and sharing social science research data

16 November 2010

<http://www.data-archive.ac.uk/news-events/events.aspx?id=2602>

As part of its current Researcher Development Initiative (RDI) grant, the UK Data Archive is organising a workshop at the University of Leeds on managing and sharing social science research data.

Preservation Assessment Survey Workshop

16 November 2010

<http://scottisharchives.org.uk/?p=137>

The Preservation Advisory Centre will be holding a free workshop in Edinburgh to provide a practical demonstration of the survey process and will provide the opportunity to see how a sample is assessed.

eSciDoc Days 2010

16-17 November 2010

www.escidoc.org

The Max Planck Digital Library and FIZ Karlsruhe kindly invite you to register for the eSciDoc Days 2010 which will take place on November 16-17, hosted by The Royal Library in Copenhagen, Denmark. The eSciDoc Days are targeted at both existing eSciDoc users and newcomers in the field of eResearch environments, publication infrastructure, research data management and scholarly collaboration.

Alliance for Permanent Access (APA) Conference

22 November 2010

<http://www.csc.fi/english/pages/apa2010/welcome>

The theme of the 2010 annual conference is scalable infrastructures for digital preservation. Important research projects inspired by the APA strategic research plans have been undertaken over the past several years. The APA is recognised as having an important role in bringing communities together, understanding requirements and developing strategic plans. Collectively the APA represents a significant part of the research data and expenditure in research and implementation in Europe. This conference will be an opportunity to take stock of where we are and where we will be going in this next phase.

National Grid Service Innovation Forum '10

23-24 November 2010

<http://www.jisc.ac.uk/events/2010/11/ngsif10.aspx>

This 2-day event will showcase the impact that the National Grid Service (NGS) has had on research in the UK, allow delegates to find out more about using the NGS in applied research, enable IT staff to find out how their institution can benefit from the NGS, and how you can contribute to and influence the future development of the NGS.

RSC CICAG/RSC Historical Group/CSA Trust Event: Celebrating the History of Chemical Information

29 November 2010

<http://www.rsc.org/ConferencesAndEvents/conference/alldetails.cfm?evid=106668>

The RSC Chemical Information and Computer Applications Group, the RSC Historical Group, and the CSA Trust are organising a joint one-day meeting celebrating the history of chemical information. A superb panel of speakers is being assembled for this very special meeting which will provide a unique opportunity to hear from some of those who have contributed to the very significant developments which have occurred in the last few decades.

6th International Digital Curation Conference: Participation and Practice: Growing the Curation Community through the Data Decade

6-8 December 2010

<http://www.dcc.ac.uk/events/conferences/6th-international-digital-curation-conference>

Digital curation manages, maintains, preserves, and adds value to digital data throughout the lifecycle, reducing threats to long-term value, mitigating the risk of digital obsolescence and enhancing usefulness for research and scholarship. This year's International Digital Curation Conference (IDCC) will be presented jointly by the Digital Curation Centre, UK and the Graduate School of Library and Information Science, the University of Illinois at Urbana-Champaign, and in partnership with the Coalition for Networked Information (CNI).

What's New:

New DPC Case Note: Practical Preservation at West Yorkshire Archives Service

<http://www.dpconline.org/advice/case-notes>

The DPC, West Yorkshire Archives Service, JISC and MLA are delighted to announce the release of a new Case Note in Digital Preservation. This new case note examines the practical experience of a local government archive when receiving a mixed archive containing digital and traditional materials.

DCC How to guide on Selection and Appraisal

<http://www.dcc.ac.uk/resources>

The Digital Curation Centre has published a new guide on '*How to appraise and select research data for curation*' written by Angus Whyte (DCC) and Andrew Wilson (ANDS). We were pleased to collaborate with the Australian National Data Service on this, the first in a new series of DCC guides intending to provide 'working knowledge' of current approaches, issues and challenges.

Research Data and Freedom of Information (FOI) - JISC Draft FAQs

<http://foiresearchdata.jiscpress.org/>

Research data can be the subject of Freedom of Information requests, as recent high profile cases have shown. As a researcher, how should you respond if faced with such a request? This document sets out to answer this question and some others you may have. Details of particular circumstances can make a major difference, so conclusions reached in an individual case may well differ from those suggested here. This document does not constitute, and should not be construed as, legal advice.

Revision of Encoded Archival Description (EAD) – Call for Comments

<http://www.archivists.org/standards/ead/eadRevisions.asp>

In February this year the Society of American Archivists charged a new subcommittee of the Standards Committee, the Technical Subcommittee for Encoded Archival Description (TS-EAD), to undertake a revision of the standard within a period of 5 years. To ensure the greatest possible input from EAD users around the world, the subcommittee has an extensive international membership and is calling for proposed changes to the current version, EAD 2002. The deadline for change proposals is 28 February 2011.

JISC white on e-Journal Archiving for UK Higher Education Libraries

<http://www.jisc.ac.uk/whatwedo/programmes/preservation/2010ejournalwhitepaper.htm>

The aim of this white paper is to help universities and libraries implement policies and procedures in relation to e-journal archiving which can help support the move towards e-only provision of scholarly journals across the HE sector. The white paper is also contributing to complementary work JISC and other funders are commissioning on moving towards e-only provision of Journals. Although focussing on the UK sector, many of the economic and emerging best practice issues it addresses will also be of interest to university libraries and research institutions in other countries. Comments are invited on this draft up until 12th November 2010.

JISC Digital Media Guidance

<http://bit.ly/bkO3sY>

JISC Digital Media has recently finished releasing ten new advice documents. The eLearning-related series covers such diverse topics as Mobile Learning, Audio Feedback and considerations for the delivery of digital media online, as well as offering how-to guides on topics such as adding multimedia to .pdf files.

LoC Digital Preservation Newsletter

<http://www.digitalpreservation.gov/news/newsletter/201010.pdf>

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

Special Issue of New Review of Academic Librarianship

<http://www.informaworld.com/smpp/title~db=all~content=g928391363>

JISC has teamed up with Taylor & Francis to produce the first Open Access issue of the New Review of Academic Librarianship, edited by Graham Walton from Loughborough University. The special issue on "dissemination models in scholarly communication" is guest edited by Hazel Woodward, university librarian and director of Cranfield University Press.

Editorial:

Editorial: Here Comes the Tide (William Kilbride, DPC)

In late September I was part of a small panel session on 'Green Digital Preservation' at the iPres conference in Vienna. It was a lively discussion and a number of colleagues have been kind enough to ask for copies of the brief 10 point manifesto which was my main contribution. I have taken the liberty of polishing it up a little before opening it to wider scrutiny and debate.

Here comes the tide.

Sometimes it's better if friends don't understand your job. How many of us, on revealing an interest in digital preservation, have ended up running impromptu surgeries on digital photographs? It can get out of hand. I remember once being asked to help write a disaster plan to replicate data to multiple sites. My witty interlocutor – always one step ahead – asked what would happen if a nuclear bomb dropped on all three places? Knowing that this gloomy scenario would involve simultaneous strikes on my home and office, I answered that I'd likely be worrying about other things. I'm good, but don't ask me to mitigate risks that are measured in megatons.



A thought-provoking discussion at the [iPres Conference in 2010](#) asked whether climate change was in the category of manageable risks for digital preservation. At first inspection, the tools of the digital preservation community seem comically puny beside the enormity of the problem. But there is something in the claims that our data sets will help study the phenomenon. More importantly, a recent study funded by JISC ([McDonald et al 2010](#)) points to the role of ICT – and in particular the costs of replication and storage – as a contributor to our carbon footprint not to mention a direct cost to our employers. ICT can be power hungry and expensive: one recent report proposed that the UK Higher Education sector spent £116m on ICT electricity bills in 2009, equivalent to 500,000 tonnes of carbon dioxide ([James and Hopkinson 2009](#)). Storage and duplication are prominent themes within the green information management and also within digital preservation, so perhaps there is something here? Is digital preservation is a net contributor to climate change?

The literature on green information management comes in two unequal parts: a very well developed set of rules about procurement that provides technical specifications for low emission and frugal equipment; and a less well developed set of proposals about data management designed to keep data storage in check. Storage is prominent in the procurement literature so the greening of digital preservation and the elimination of costs can start quickly with a sensible set of purchases. But long term storage is a small part of an organisation's total storage costs so it will be hard to isolate the savings associated with a green digital preservation strategy. But nonetheless, if you can make sense of sometimes fragmented and complex procurement processes and of the various tests and measures associated with low emissions storage, the results are quantifiable and comparison between different approaches is viable.

That there should be an interest in green related matters should come as no surprise but collections managers who deal with physical collections probably understand the environmental impact of data storage better than digital repository managers. That's because the rules and regulations about the situation and performance of buildings are very well developed (perhaps too well developed). Anyone commissioning or extending a new repository – a real hard repository which hurts when you kick it – will be well-versed in topics like the environmental impact of their favoured location; the relative performance of materials; the long term costs of robotic versus manual storage; the need to dispose safely of nasty substances like asbestos; the benefits of insulation, stable temperatures and how to achieve them; and some pretty byzantine topics like landfill taxes, water dispersal and traffic management. There are two reasons why they are likely to know chapter and verse on these arcane topics: they want to know that the building will not be a nightmare to maintain over a 30 to 50 year horizon; and even if they didn't care, the planning process forces them to take account of the environment. Like them or loathe them, these regulations tend to be developed on the basis of verifiable

research and have developed through some form of public consultation. But if your repository is a digital one then you can ignore all that intrusive democracy and environmental science.

So it is relatively easy to set up a digital repository, it can be quite another thing to operate it. Energy consumption is a major consideration for storage suppliers and anything that reduces this is likely to be good news for immediate budgetary reasons and longer term climatic ones. Put simply, the larger the repository, the greater the energy requirements, the larger the electricity bill and the greater the carbon footprint. It's not a simple calculation: improved data capacities means that, for the long term, there is not a direct correlation between the quantity of data stored and the size of the electricity bill. And there is a considerable difference in energy performance between online, near line and offline storage. [Deleting a gigabyte is not likely to save a polar bear](#): but modest savings are possible and reasonable. In the current financial climate, anything that reduces our energy bills is going to be welcome.

There probably is something that digital preservation can offer here, though we really ought to clarify a few things. [McDonald et al \(35-36\)](#) quote one commentator who links duplication and preservation, implying that the extent of redundancy equates to the effectiveness of a preservation plan. That's nonsense. Uncontrolled proliferation is not a reliable route to robust data. Multiple copies are a sensible part of a preservation plan, but so is keeping careful track of them and checking them routinely. It follows therefore that uncontrolled proliferation will rapidly become an obstacle to the delivery of a preservation plan. In any case most of our time is spent on other issues like characterisation, access, ingest, metadata management and validation—topics which do not necessarily lead to proliferation.

What do we have to offer? I'm not sure if it's a secret we should keep to ourselves or a boast we should trumpet from the roof tops, but let me put it bluntly: *digital preservation is about confident deletion*. It's only when a data centre has decided what it really needs to keep that it can be confident about clearing out the duplicates and worthless junk that is of no value. Reducing the volume of data we hold may or may not reduce the costs of data storage, but trying to reduce the storage without a preservation plan telling you what to dispose of is reckless and sometimes illegal. If your digital preservation plan is what empowers deletion then it is likely to be a key element in any green information management strategy. I'd go so far as to say that you can't have one without the other.

"Sometimes it's better if friends don't understand your job. How many of us, on revealing an interest in digital preservation, have ended up running impromptu surgeries on digital photographs?"

A digital preservation strategy might help stem the rising tide of data but it is not the only tool. It is certainly the case the any archive will always end up looking like its owner. So if an organisation relies on bloated and inefficient file formats then its archive will come to be like that too. Open standard, non proprietary and well documented file formats do not necessarily have to be large to work, but it's not something we can influence after the fact. It's another reason to work with data creators to help them turn their short term aspirations into long term digital outcomes without costing the Earth. Digital preservation can be tough on data and tough on the causes of data.

So what shall we do when the tide comes in? Does digital preservation have anything to offer the climate change debate? Ultimately this is shared problem that requires a response from everyone in the personal and professional lives. Digital preservation whether directly through manual reduction of data volumes or indirectly through procurement of more efficient systems can help save money in the short term and help save more than that in the long term.

James, P and Hopkins L 2009, *Sustainable ICT in Further and Higher Education, A Report for JISC*, University of Bradford, online at <http://www.jisc.ac.uk/media/documents/publications/rptgreenictv1.pdf> last visited 28/10/2010

McDonald, D McCulloch E, and MacDonald A 2010 *Greening Information Management: Final Report*, University of Strathclyde, online at: <http://www.greeningim.org.uk/Portals/82/GIMFinalReport.pdf> last visited 28/10/2010

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 ...

Laura Molloy, *Incremental* Project, Humanities Advanced Technology and Information Institute (HATII), University of Glasgow



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

I'm a researcher at [HATII](#) at the [University of Glasgow](#). I'm a member of the JISC-funded [Incremental](#) team, working with the [University of Cambridge](#). I function on the team as the non-specialist user, to try to represent the point of view of a researcher who doesn't have a background specifically in digital preservation or archiving. This way, I can help our resources to use appropriate language, and point out the things we can't reasonably expect the general researcher population to know. Hopefully, this will help our resources to be as usable and accessible as possible!

Tell us a bit about your organisation

The University of Glasgow is the fourth oldest university in the English-speaking world. It was founded in 1451, and previous staff members include scientist Lord Kelvin, economics legend Adam Smith, pioneer of television John Logie Baird and Scotland's national poet, Edwin Morgan. I'm still looking for the eminent women of the University's history!

HATII is an interdisciplinary research institute. We offer teaching in arts and media informatics, information management and preservation, computer forensics and museum theory and practice. In addition, there is a research team, which I'm part of, which works across a range of UK and EC-funded projects, mostly in digital curation and preservation, digital humanities and related areas. I've previously worked on the Arts and Humanities Data Service (AHDS) – until the cessation of funding, HATII hosted the [AHDS Performing Arts](#) centre – as well as [Digital Preservation Europe](#) (DPE) and the [Planets](#) project. Our current projects include *Incremental*, of course, as well as the [Digital Curation Centre](#), [3D-COFORM](#), [SHAMAN](#) and several others.

What projects are you working on at the moment?

Incremental – it's funded by the JISC under the [Managing Research Data programme](#). I really like working on JISC projects – communication is supported well, and programme management really engage with what we do. *Incremental* is pitching into the currently-hot topic of how researchers manage their data. We're working with researchers across the arts, humanities and sciences at the Universities of Cambridge and Glasgow to see what help they need with research data management – this will help us to build simple but relevant resources to address these needs. The idea behind the name is that we're improving research data management in small, practical increments at both institutions.

In my mythical free time I'm also working on a study of performance practitioners and how they interact with the archive, i.e. whether they document their work, what they do with these records, and how they use archives to find other people's work. I'm really interested in the relationship between an ephemeral event and the archived traces of it.

How did you end up in digital preservation?

I started working at the AHDS Performing Arts centre, drawn in by the possibility of working in an art-related post with a great team and getting to write a lot! And through that, I met all these people across the UK who were passionate about something called 'digital preservation'. The more I thought about it, the more I realised that digital preservation is something that is relevant to almost everybody in the developed world. After the AHDS, I was lucky enough to get involved in Planets, DPE and *Incremental* as well as continuing my own research. I've learned so much about the tools that various projects are working on, as well as the

pedagogical issues – it really isn't an area that the general public thinks about much, but it really does affect almost everybody. I've also met loads of great people with very interesting minds!

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

HATII is unusual in that digital preservation and data management is, for us, not only something deal with as part of our regular work activity like everybody else, but it's also our field of research. So we think about it as a discipline, not just an issue for our organisation: we teach students about it as well as working on projects developing tools and services to improve it out in the rest of the world. Having said that, we do also have working relationships with those who decide the university's policies on digital preservation and data management.

Digital preservation is a bit of an odd research field as it's invisible in some ways – people always say, 'You research *what?*' - but as soon as you stop and think about it, it makes sense that people should work in this area, and develop tools and improve understanding. Almost every professional role, in every sector, now involves email, digital document creation and data storage of one kind or another, but unlike with paper documents, the creation is only the first step in the file's survival. So we need to make the connection in people's minds between the fact that they want to hold onto their information, and the fact that they need to actually do something for that to happen. That practical approach to helping individuals in a specific context is what *Incremental* is doing for researchers at Glasgow and Cambridge.

What sort of partnerships would you like to develop?

We at *Incremental* want to be of assistance to people attempting to improve research data management at their own institutions, so we're keen to share the resources we develop as well as hearing their experiences. If anyone else has developed training materials for research data management, we'd also love to hear from them.

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

The magic funding pot! Particularly if it can genuinely assess interdisciplinary project proposals in a collaborative, interdisciplinary way and provide experimental funding for early career researchers to see what they come up with, even just over the short term. Knowledge for its own sake is valuable as we have no way of knowing how it will be built on in the future, and we're in danger of legislating that kind of approach out of existence.

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

If you want your stuff to be around for at least your own life span, you will need to find out how to choose the best storage for your needs and learn how to look after the stuff you choose to keep on it. There's no way around it!

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

There's no one thing I could choose. My only thought is that we want researchers to have the ability to preserve the files they want to keep for the long term, which is one of the things *Incremental* aims to help with.

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

My email address is laura.molloy@glasgow.ac.uk.

To read more about our project please visit <http://www.lib.cam.ac.uk/preservation/incremental/> or check out our blog at <http://incrementalproject.wordpress.com/> - we genuinely welcome comments and questions.

You can also follow us on Twitter – our name is @JISCIncremental.

To find out more about the JISC Managing Research Data Programme, visit:
<http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>.

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 ...

**Dr. Dinesh Katre, Head & Programme Coordinator
Centre for Development of Advanced Computing (C-DAC)**

National Study Report on Digital Preservation Requirements of India

We are happy to inform the digital preservation community that Department of Information Technology, Ministry of Communications and IT, Government of India has entrusted C-DAC with a sponsored research project to prepare a national study report on digital preservation requirements of India. I am the Chief Investigator for this project at C-DAC, Pune, India. It was a mammoth and complex task as collecting the data pertaining to digital obsolescence and preservation challenges faced by various stakeholder organizations wasn't easy. It was important to capture the technological, legal, organizational and usage imperatives of digital preservation from diverse domains like government, science and education, films, video and audio, health, insurance and banking, cultural heritage, etc.



Therefore, we constituted the national expert group comprising of archivists, technologists and other stakeholders from 30 public sector organizations across the country. Initially, the members of this group were asked to submit position papers stating the challenges of digital obsolescence, current archival practices, preservation priorities in their respective domains and the necessary actions that need to be taken. Structured questionnaires were prepared for different domains and sent across to all expert group members so as to capture relevant information for this report.



A national meet of the expert group was organized at Pune on May 20-21, 2010 during which the members were given an opportunity to present their position papers and offer their recommendations in terms of short term (3 years) and long term (10 years) actions for long-term digital preservation of their data. During the presentations, feedback was given, questions were raised and the members were requested to resubmit the enhanced position papers for inclusion in the report. The details of this meet are published at www.ndpp.in

National Meet of Digital Preservation Stakeholders / Experts at Pune, India

The national expert group unanimously recognized that we are extremely vulnerable to the threat of irrecoverable data loss due to imminent technological obsolescence within next 5 years or so due to absence of reliable digital preservation practices; as Government of India is hugely investing in the computerization and digitalization of its departments at national, state and district level.

Finally the national study report is consolidated and presented in 2 volumes as under:

- Volume-I Recommendations for National Digital Preservation Programme of India (80 Pages)
- Volume-II Position Papers by the National Expert Group Members (165 Pages)
(ISBN for both volumes: 978-81-909383-1-0)

Apart from the domestic study of digital preservation requirements, the report also includes a comprehensive overview of the international digital preservation initiatives like NDIIPP, CASPAR, PLANETS, NESTOR, DPE, etc. We have also referred the useful information published by Digital Preservation Coalition of UK in this report. It also includes some chapters on various metadata standards, preservation practices, OAIS, checklists and auditing approaches to digital repository. It is important to note that the recommendations given by the international panel of experts during the Indo-US workshop on International Trends in Digital Preservation held in March 24-25, 2009 at C-DAC, Pune are also incorporated in the report.

I wish to gratefully mention that we received a lot of support and valuable guidance from Dr. David Giarretta, Project Director, CASPAR, Science and Technology Facilities Council, UK in the making of this report.

The report was presented before a committee at Department of Information Technology, Ministry of Communications and IT, Government of India, New Delhi on 22 August 2010. It has been received very positively and as recommended in the report, we have been asked to make a proposal to establish the Centre of Excellence for Digital Preservation R &D at C-DAC, Pune. The Centre of Excellence will be expected to develop tools, technologies, standards and best practices for digital preservation. We also propose to develop the pilot digital repositories in selected domains.

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/>