

Conference Report: Curating Research: e-Merging New Roles and Responsibilities in the European Landscape

17 April 2009, The Hague, Koninklijke Bibliotheek, The Netherlands

1. Summary of issues relevant for DPC members

- Training is popular but what sort of training will be most effective: what will drive down costs and support our work best?
- Considerations of scale: what is the right size solution to our digital preservation challenges? Do we want lots of small DP facilities or a small number of large ones? How do we collaborate without undermining institutions?
- There would appear to be a lot of policy development which is an important change from a decade ago: but how do we assess the value of these emerging policies and how do we know if they are being applied?
- There is still a policy gap. There are some high level aspirations in the UNESCO Charter and some very detailed guides, but a gap in between. What would be our golden rules for creating digital data?

2. Contributors and their contributions

Eileen Fenton of Portico provided an opening tour of digital preservation. She rehearsed familiar concepts: data continues to proliferate in scale as well as complexity; scale of storage is not the barrier we once thought; and ingest is a significant cost driver. Portico has grown to hold around 11.5 million articles. This scale has been a challenge, but in unexpected ways. Processes that work well at one scale become expensive or unworkable at a larger scale and in unexpected ways: what works for a GIS might not work for video. Whereas publishers tend to be clear about what their needs are, researchers and institutions digitising their own collections find this harder: so even small collections can be a challenge. This leads to two broad observations: we need to find the right size solutions to our digital preservation challenges: there are economies of scale and expertise. Secondly, though less clear, we will need to 'become friends with selection'.

Scale issues were implied in a paper by **Tom Kuipers and Jeffrey van der Hoeven** who reported preliminary results of a survey of European research institutions lead by Parse.Insight Project/Koninklijke Bibliotheek, The Netherlands. Key trends were that many institutions were looking to develop their own long term repositories but lacked the staff or expertise to do so. When asked, many respondents thought that the National Libraries should ultimately be responsible for preservation.

Dale Peters, Scientific Technical Manager DRIVER II, State and University Library Goettingen, Germany brought a different perspective by reviewing the wide range of recent and forthcoming European projects that have contributed to our understanding of the digital preservation landscape. It's clear that there have been movements in the right direction at a policy level: there is a lot more expertise available than hitherto. But there are still serious questions about whether these policies are being effectively applied and even if they are applied, whether their application is evaluated. She argued that data creators are playing a larger role in development than hitherto. Key areas for the future are about co-working

between types of service which are themselves still only emerging, such as linking institutional repositories to long term facilities that will make them more robust.

A case study of practical actions from **Maria Heijne**, Librarian, TU Delft Library/Chair UKB, Dutch Consortium of Research Libraries, The Netherlands shed light on a 3-party collaboration of Dutch technical universities. She argued that Libraries were the natural home of digital preservation for research data and that with enough thought and planning it was possible for institution to tackle the problems of digital preservation, using institutional research budgets, without having to rely on 3rd parties. The key challenges were about the more complicated metadata used by research communities and about the differences between 'dynamic' and 'static' data.

3. Discussion, debate and reflection

There was surprisingly little dispute of **Maria Heijne's** view that libraries were the best institutions to look after research data. I wonder if that would have gone down so easily with a different audience, perhaps outside of a library.

There is also a lesson about scale: Portico needs about 30 staff for a service purely concerned with books and journals. It's hard to imagine every research institution in Europe being able to find or afford that level of expertise to cope with the diverse range of institutional outputs. Given the likely contraction in public sector spending in the next 3 years, **Eileen Fenton's** call for 'right sizing' seemed prudent.

In different ways, Maria Heijne and Eileen Fenton made me think about the **UK Research Reserve**. Commentators have noted that this has tended to provide a library perspective on the problem of preserving digital research data, in line perhaps with Maria Heijne's thoughts about the role of libraries. But it's also clearly based on the assumption that economies of scale are useful and that individual institutions are not going to have the capacity to deal with the complexities associated with research outputs.

Tom Kuipers and Jeffrey van der Hoeven reported that there's a demand for training. It would be good to have a really keenly informed sense of what sorts of training and for whom and a discussion of why. (Is it too heretical to say that training can be expensive and not always as useful as it seems). Do people want training in setting up digital preservation facilities? Is it in fact realistic to set up lots of small DP facilities with large duplicated costs? Would we be better training people to negotiate with DP providers to reduce the ingest cost and avoid the duplication costs too? Discuss...

Keith Jeffrey of STFC and **Peter Wittenburg** of the Max Planck Institute for Psycholinguistics led a vibrant discussion on what we should expect from a national or international infrastructure. This highlighted, inter alia, that decision making frameworks did not have enough clout. A lot of digital preservation advice is stable - at least at a generic level - but that there was a gap in policy level among authorities able to enforce or encourage uptake in this advice. There was a consensus that the DP community could easily draft a generic set of 'golden rules' to append to the UNESCO charter (or indeed stand alone) which could fill this gap.

The conference was, of course, dominated by researchers and research needs and though this is important, there is a risk that the solutions being developed exclude other parties.

Researchers have particular needs and solutions to be cost effective need to be extensible for other users. For example, there was much repeat of the difficulty of moving and replicating large data sets. There is no question that large data sets are awkward: but I tend to think this is a big problem for a small number of people. I'm still to be persuaded that a majority of researchers have this problem. Surely basic cataloguing and effective risk management in modest research is still more of a priority? I'm not sure whether the research community is a special case or not and what the implications might be.

4. Conference organisers

- Koninklijke Bibliotheek (KB): <http://www.kb.nl/>
- Ligues des Bibliothèques Européennes de Recherche (LIBER): <http://www.libereurope.eu/>
- Nationale Coalitie Digitale Duurzaamheid (NCDD): <http://www.ncdd.nl/>

5. References

UNESCO (2003) *Charter on the Preservation of the Digital Heritage*, United Nations Educational Scientific and Cultural Organisation, online at:
http://portal.unesco.org/ci/en/files/13367/10700115911Charter_en.pdf/Charter_en.pdf
(last visited 21/04/2009)

6. See also

Conference pages including programme, overview and presentations:
<http://www.kb.nl/hrd/congressen/curatingresearch2009/index-en.html> (last visited 22/04/2009)

NCDD blog and discussion including links to projects and initiatives discussed:
<http://digitaalduurzaam.blogspot.com/2009/04/curating-research-1-wrap-up.html> (last visited 22/04/2009)

7. About this document

The **Digital Preservation Coalition (DPC)** is a not-for profit membership organisation whose primary objective is to raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. It acts as an enabling and agenda-setting body within the digital preservation world and works to meet this objective through a number of high level goals. Its vision is to make *our digital memory accessible tomorrow*.

This report has been drafted in support of these aims. It is for circulation primarily to the benefit of DPC members and is for wider circulation only according to the document distribution noted at the top of each page.

William Kilbride,
william@dpconline.org
Executive Director of the Digital Preservation Coalition
April 2009