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

Inge Angevaare

Coordinator, Netherlands Coalition for Digital Preservation (NCDD)
and Editor of LIBER Quarterly, PO Box 90409, 2509 LK
The Hague, Netherlands,
inge.angevaare@kb.nl.

Abstract

On 17 April 2009 LIBER, the Koninklijke Bibliotheek (KB) and the Netherlands Coalition for Digital Preservation (NCDD) co-organised LIBER's first workshop on digital curation/preservation, providing an introduction to the theme of digital curation and different perspectives on the roles research libraries can and/or must play in keeping the digital records of science safe for future generations. Speakers included Eileen Fenton (Portico), Dale Peters (DRIVER), Maria Heijne (TU Delft Library), Jeffrey van der Hoeven (KB, PARSE.insight) and ninety workshop attendees. The paper includes the conference report and some results of the PARSE.insight online survey into digital preservation practices of LIBER libraries.

Key words: conference report; digital curation; digital preservation; research libraries

Introduction

On 17 April 2009 over ninety participants from all over Europe gathered at the Koninklijke Bibliotheek (KB) in The Hague to attend LIBER's first ever conference dealing with issues of digital preservation/digital curation (Figure 1). The event was co-organised by the KB and the Netherlands Coalition for Digital Preservation (NCDD)[1]. The workshop, entitled Curating research: e-merging

new roles and responsibilities in the European landscape (the website includes powerpoint presentations), was a direct result of the Memorandum of Understanding between LIBER and the Koninklijke Bibliotheek which was signed at the Warsaw Annual Conference in July 2007. In the memorandum, LIBER and the KB agreed the following:

> 'LIBER and the KB share the vision of a European research community which is supported by a provision offering perpetual access to digital publications, to the benefit of research libraries and academic researchers.
>
> The KB developed the international e-Depot to guarantee preservation of and perpetual access to the records of science for its designated community, research libraries and their patrons. LIBER recognises the KB as a trusted organisation for preserving digital information.
>
> The two bodies will explore possibilities to work together in the field of digital asset management and curation. They will nominate representatives to explore issues around the long-term digital curation and preservation of materials which are deposited in institutional and subject-based repositories in LIBER member institutions. This work will identify services, technical solutions, work flows, costs and funding opportunities to deliver the partners' vision. …'[2]

Even before the agreement was formally signed, it was decided that furthering knowledge and expertise on digital curation among research libraries would have to be a prime goal of the cooperative effort, and that such knowledge should start at an organisational and managerial level: what is digital curation and why should research libraries get involved – or, alternatively, consciously decide *not* to get involved? Ideas for a joint KB/LIBER workshop were soon developed.

To lay the groundwork for the workshop and introduce the issues involved, this author published an article in LIBER Quarterly 19/1 (Angevaare, 2009), quoting JISC's definition of *digital curation*: 'The term "digital curation" is increasingly being used for the actions needed to maintain and utilise digital data and research results over their entire life-cycle for current and future generations of users'. In other words, digital curation is a broader term than digital preservation, it comprises the cradle-to-grave care digital objects need.

Fig. 1: Carmen Morlon of the LIBER Secretariat (left) registering participants.

The conference programme included four plenary presentations followed by four simultaneous workshops, which were repeated after lunch to enable each participant to attend two workshops, and a wrap-up.

Digital curation is a means to the end of *permanent access*

After gracious words of welcome by conference host Hans Jansen, Director e-Strategy of the KB, the plenary session was kicked off by Eileen Fenton, Director of Portico, a US not-for-profit archive of electronic books, journals and other scholarly content (Fenton, 2009). Fenton started her talk by stressing that digital curation is never an end in itself; digital curation is but a means to the all important goal of *permanent access*. Then she presented her audience with a very clear and concise summing up of the digital curation landscape:

–   Digital information is exploding
–   Digital information is prone to loss
–   We need to manage digital information to safekeep it for future generations.

Fenton offered her audience a number of basic guidelines for dealing with digital information:

- Befriend selection. We cannot possibly preserve everything, nor should we. Selection principles will be guided by the individual missions of all our organisations. To keep the workload to a minimum, we can expect that technical tools will enable us to automate selection.
- Recognise the demands of diversity and scale. What works for a video may not work for GIS data and what works for 1TB may not work for 100TB – yet scale impact moments may be difficult to recognise.
- One size does not fit all: multiple preservation methodologies may be needed. Migration may work for PDF's, emulation may be needed for websites, and for complex databases we may have to just store the bitstream until more advanced tools are developed.
- Understand cost drivers and minimise them. Research indicates that ingest costs may be higher than long-term costs. Also, taking proper measures at the moment of creation will save you a lot of money in the long run – therefore there is a need to work closely together with the producers of digital data.

In conclusion, Fenton identified a number of opportunities:

- Experiment with new approaches, such as the Dioscuri project.
- Share lessons with others, e.g., in the European Planets project or within various digital preservation coalitions (DPC in the UK, nestor in Germany, NCDD in the Netherlands).
- Right-size the solution to the digital resource to be preserved: academic journals may require other measures than newsletters.
- ' Do not go at this game alone' : rely on partners rather than reinventing the wheel at home.

Jeffrey van der Hoeven and Tom Kuipers (KB) presented the PARSE.insight, a European project aimed at gaining insight into the data management and digital preservation practices of researchers, data archives and libraries throughout Europe and developing a roadmap for developing an e-infrastructure in Europe. The PARSE.insight online libraries survey was distributed via the LIBER-ALL mailing list to all four hundred LIBER members.

Significantly, only 59 questionnaires were returned, and this fact in itself may say something about the degree to which LIBER libraries are (not yet) involved in preserving digital content. Of these 59 respondents, 65% report that the organisation has a preservation policy in place, including selection criteria for content to be preserved, rights management, etc. More than 70% preserve published content (books, journals) and only 42% store research data. When asked who should be responsible for digital curation, 74% answered the national library; 59% the researcher's institute; 59% the research library; 46% the government, and 25% the research funder. 56% report that the tools and infrastructures presently available do not suffice to reach the desired goals.

The full details of the survey can be found in the interim report that has been published by PARSE.insight since the conference (First insights, 2009).

Dale Peters (Göttingen, <u>DRIVER</u> project) reviewed the many research projects which are under way to tackle the more technical aspects of digital preservation and put them in the overall framework:

– File format services (GDFR, a global registry of file formats; JHOVE, a tool for format validation; AONS, an automated obsolescence notification system)
– Persistent identifiers (PILIN, an Australian national persistent identifier system)
– Archival concepts/repository models (OAIS, the renowned Open Archival Information System, the basis of any digital repository; SHAMAN, for an open distributed resource management infrastructure framework; CASPAR)
– Metadata (PREMIS, INSPECT)
– Preservation strategies (Planets, Plato, Dioscuri, KEEP)
– Organisational aspects (PARSE.insight, Alliance for Permanent Access, LIFE2/3)
– Scientific data and digital research infrastructures (data resources, e-science verification).

The full list of acronyms and initiatives bedazzled the audience somewhat; fortunately all of these projects have websites to consult when libraries wish to explore them further.

Peters stressed the importance of linking all the information on the web. Also, she mentioned, almost in passing, that of course not every repository must by definition have long-term preservation facilities. She agreed with Fenton that trusted third-party services are not only an acceptable but often an essential part of the digital preservation equation.

Maria Heijne (TU Delft Library) agreed with Hans Jansen in his opening speech that securing long-term access to research data and publications is *core business* for libraries.
Together with the two other technical university libraries in the Netherlands (Twente and Eindhoven), TU Delft Library set up a project for a 3TU Datacentre.

In Heijne's view, libraries have no choice but to engage in data management. She rhetorically asked her audience: who else could do it? It is libraries that have the experience needed, they just need to give their services a digital twist.

This digital twist – as also stressed by Fenton – involves working very closely together with the research communities themselves. They all have very distinct workflows and metadata schemes which are also very different from libraries' traditional schemes, so both sides must do a lot of adapting. Although it is early days yet, the 3TU.Datacentre is hoping to grow into a best practice of research libraries' involvement with data curation. 3TU do important work in developing an entirely new relationship with the research community to create a win-win-situation for researchers *and* research libraries: better quality data during the research process, at the same time enabling data to flow into the digital archive with very little additional effort. In a project which was sponsored by SURFfoundation, the 3TU. Datacentre closely analysed workflows in two sub-disciplines in order to be able to determine the requirements for the 3TU.Datacentre (Waardevolle data en diensten, 2009).

The Needs of Researchers

The afternoon split the audience in a number of workshops. Keith Jeffery (Science and Technology Facilities Council, STFC), UK, and chairman of the

Alliance for Permanent Access) and Peter Wittenburg (Max-Planck-Institute of Linguistics, Nijmegen) focussed their attention on research itself; what elements of the research life cycle should in fact be preserved, and who is responsible for preserving them? This is a monumental question, especially as the researchers in this group kept stressing how complicated research data are: only the publication is static, everything else is dynamic and thus difficult to preserve.

Fig. 2: Workshop session on national and international roles.

Some doubts were raised as to whether libraries are in fact best suited for the job of preserving the manifold elements of the research life cycle. Libraries' work flows and metadata schemes, it was suggested, are perhaps too 'library-centric' to serve the research community properly.

So should perhaps the management of live data, including providing *access*, be separated from the archival function? And, more importantly, should communities themselves take care of curation rather than libraries? Krystyna

Marek from the European Commission explained that the e-infrastructure vision of the EU is in fact focussing on the research communities themselves – which reminded this author of Sijbolt Noorda's comments during the 2008 LIBER Annual General Conference that perhaps libraries had missed their window of opportunity in the digital age (Noorda, 2008). At the time this comment elicited a remark from Heijne that Noorda was maybe judging too soon.

Hans Geleijnse of LIBER suggested that we draw up five or ten golden rules of digital curation, to help the community along. UNESCO drew up such guidelines in 1996, but they need modernising and updating. Half the attendees of this workshop volunteered on the spot to help bring this about, but it seems their enthusiasm did not have a follow-up.

Problems, Preconditions and Costs: Opportunities and Pitfalls

Neil Beagrie (Charles Beagrie Ltd., Figure 3) took his cue from David Rosenthal, who recently held a controversial presentation at the Coalition for Networked Information (CNI), saying that our real problems now are not about media and hardware obsolescence, as predicted by Jeff Rothenburg in his famous 1995 article, but rather about *scale* and *cost* and *intellectual property* (Rosenthal, 2009). 'Bytes are vulnerable to money supply glitches,' is a memorable quote from Rosenthal's presentation, especially in these credit crunch times.

So, what does digital preservation cost? Marcel Ras of the Koninklijke Bibliotheek shared his experiences with the KB e-Depot which now archives about 13 million journal articles, thereby providing a sound base for archiving the *published* output of research. Between now and 2012, the size of the e-Depot will grow exponentially, as the e-Depot will incorporate digitised master files and websites. The cost will go up to M€6 a year, which includes 14 full-time staff. In the corridors, some representatives of organisations with emerging digital repositories expressed their wonder at hearing such numbers. They estimated that they would not run into such high costs. Is the KB perhaps still paying the price for its early-mover position?

Fig. 3: Neil Breagrie addressing the workshop on costs.

And what do these numbers say about possible costs for research libraries? Beagrie investigated the costs of preserving research data at higher education institutions in the UK (Keeping research data safe, 2008). Notable findings are that preserving research data is much more expensive than preserving publications. Also, timing is a crucial factor. Good care at creation saves a lot of money in the long run. Beagrie also found that it is very difficult to compare costs between organisations, as they all have their own criteria when it comes to attributing costs to digital preservation.

Another finding: scale matters. Start-up costs are high, but adding content to existing infrastructures is relatively cheap. The Archaeological Data Service estimates that overall costs tail off substantially anyway with time and scale. This is important for our thinking about funding models and up-front (endowment) payment.

Beagrie concluded his presentation with the observation that when it comes to defining a policy for digital preservation, many higher education institu-

tions still have a long way to go; this author would add that the same seems to hold true for research libraries.

Conclusions

Digital curation and preservation are emerging new challenges for research libraries. LIBER organised this first workshop on digital curation in order to aid libraries in making informed choices about long-term care for their digital collections and possibly for research data. Notable recommendations include:

–   Digital curation is too complex and expensive a task to be taken on lightly. It is recommended that libraries find trusted partners to work with rather than develop a digital repository by themselves.
–   If a library decides to include research data in its long-term collection plan (as some argue is the only way to go for research libraries), it is essential that libraries establish close working relationships with the research communities they serve, as each (sub)discipline has its own requirements.
–   Another reason to get involved in the research process is the fact that measures facilitating long-term access (such as proper metadata) must be taken at the point of *creation* of a digital object; interventions at a later point in time may be impossible or prohibitively expensive.
–   Selection is the key to finding the right balance between available resources and data to be accessed permanently.

In order to continue to further the debate between research libraries on this important new work, LIBER intends to organise a follow-up meeting in two years' time.

CREDITS

Photos courtesy of the Koninklijke Bibliotheek, Jacqueline van der Kort.

References

Angevaare, Inge (2009). 'Taking care of digital collections and data: "curation" and organisational choices for research libraries', LIBER Quarterly 19/1, pp. 1–12; http://liber.library.uu.nl/publish/articles/000278/article.pdf

First insights into digital preservation of research output in Europe: interim insight report (2009), PARSE.Insight, http://www.parse-insight.eu/downloads/PARSE-Insight_D3-5_InterimInsightReport_final.pdf (retrieved 23 October 2009)

Keeping research data safe (2008), by Neil Beagrie, Julia Chruszcz and Brian Lavoie, JISC, http://www.jisc.ac.uk/publications/documents/keepingresearchdatasafe.aspx (retrieved 23 October 2009)

Noorda, Sijbolt (2008), 'The Impact of Digitization from an Academic Point of View', powerpoint presentation at the 2008 LIBER Annual General Conference, Koç University, Istanbul, 1 July, http://www.ku.edu.tr/ku/images/LIBER/istanbul_noorda2.ppt.

Rosenthal, David (2009), 'Spring CNI Plenary: the Remix', DSHR's blog, http://blog.dshr.org/2009/04/spring-cni-plenary-remix.html

Waardevolle data en diensten (2009), Eindrapport, 3TU.Datacentrum, http://3tu.typo3.3xo.eu/fileadmin/documenten/Eindrapportage_WDenD_v10_170709.pdf (retrieved 23 October 2009; only available in Dutch)

Websites Referred to in the Text

3TU.Datacentre, http://3tu.typo3.3xo.eu/en/hoofdmenu/home/

Alliance for Permanent Access, http://www.alliancepermanentaccess.eu/

Curating research, conference website with powerpoint presentations, http://www.kb.nl/hrd/congressen/curatingresearch2009/index-en.html

DRIVER, Digital Repository Infrastructure Vision for European Research, http://www.driver-repository.eu/

KB, Koninklijke Bibliotheek, e-Depot and digital preservation website at http://www.kb.nl/hrd/dd/index-en.html

NCDD, Netherlands Coalition for Digital Preservation, http://www.ncdd.nl/en/index.php.

PARSE.insight, http://www.parse-insight.eu

Portico, http://www.portico.org/

STFC, Science and Technology Facilities Council, http://www.stfc.ac.uk/

TU Delft Library, http://www.library.tudelft.nl/ws/index.htm

Notes

<hr/>

[1] Memorandum of understanding between LIBER and the Koninklijke Bibliotheek, signed 5 July 2007. Unpublished.

[2] The NCDD, Netherlands Coalition for Digital Preservation, is a cross-sectoral, bottom-up initiative of major stakeholders in public digital information intended to promote permanent access to digital information. Members of the Coalition are: 3TU. Data Centre, Netherlands Institute for Sound and Vision, Data Archiving and Networked Services, the Royal Netherlands Academy of Arts and Sciences, the Koninklijke Bibliotheek, the Ministry of the Interior and Kingdom Relations, the National Archives of the Netherlands, the Netherlands Organisation for Scientific Research and SURFfoundation. Associated members include Statistics Netherlands and Cultural Heritage Netherlands. In July 2009, the NCDD published its national survey on digital preservation, a twenty-page English-language summary of which is available from http://www.ncdd.nl/en/publicaties.php.