

## **The KB e-Depot: Building and Managing a Digital Preservation Service**

*Marcel Ras, head e-Depot Department,  
Koninklijke Bibliotheek, National library of the Netherlands,  
PO Box 90407, 2509 LK The Hague,  
The Netherlands; e-mail: marcel.ras@kb.nl*

### **Introduction**

In centuries past publishers and libraries had well-defined roles and responsibilities in the dissemination of information: publishers produced publications, libraries bought them and preserved them in their stacks. In the digital world, that division of labour no longer applies, especially with regard to e-journals. Now, publishers retain ownership of the content and license access rights to libraries.

This leaves research libraries in a vulnerable position: their dependence on electronic information is growing: e-journals have come to dominate academic literature. Although publishers assure perpetual access rights to the content purchased in their licensing agreements, the question still remains who takes responsibility for preserving this electronic intellectual output.

The question is all the more pressing as digital information is fragile and very much dependent on a properly working technical environment – which must function throughout the life-cycle of the digital object. Many threats to such continuity may be identified: hardware and software will eventually become obsolete, publishers go out of business, or their access platforms may fail.

### **Digital preservation**

In contrast to printed materials, digital information will not survive and remain accessible by accident: it requires ongoing active management. The extent to which this active management differs radically from activities in the paper environment is something that is only now beginning to be understood. Digital preservation is the process of active management by which we ensure that a digital object will be accessible in the future. The time span is potentially very short and technology changes rapidly. This will have great impact on future access of our 21st century digital output. Much of the technical infrastructure we use today to create and read our data may be unavailable in the future. Digital media are fragile and dependent on a range of technical processes. A combination of specific software programmes and physical hardware is crucial to read any digital file, and this range of requirements has to be preserved as a whole in order to access the information. Solutions for digital preservation are being developed and effective preservation tools and archives are being built. Teams of research professionals are working hard to solve the challenges of ensuring access to complex digital data such as multimedia.

But digital preservation is not only a set of technical problems for which tools can be created. It relies on organisational objectives; financial and staffing issues, collection management, legal obligations, auditing requirements, strategies and policies.

One group of libraries would seem to be specifically well-placed to assume the role of securing access to the records of science: national libraries. National libraries have a long tradition of preserving national deposit collections and their remit specifically includes the long-term perspective so needed to secure permanent access. All that would be needed is to extend this remit from printed collections to digital publications.

### **The KB e-Depot**

The Koninklijke Bibliotheek, National Library of the Netherlands, was one of the very first cultural heritage institutions to become aware of the emerging importance of digital resources. As early as 1998 it concluded an agreement with the Dutch Publishers Association to extend the Dutch voluntary deposit scheme to off-line digital publications (cd-rom's etc.), and in 1999 a tender was issued for the development of a long-term preservation facility for digital information resources. As no ready-made commercial products were available at the time, the KB embarked on a joint project with IBM to develop the Digital Information Archiving System (DIAS). The 'e-Depot' became operational in January 2003 and it was the first preservation facility specifically designed to store and maintain digital objects not just for a few years but indefinitely, in line with the remit of the Koninklijke Bibliotheek as the national deposit library.

Originally, the e-Depot was designed to preserve the digital publications of the Dutch publishers, in agreement with the Dutch voluntary deposit scheme. In June 2005 an agreement was signed with the Dutch Publishers association to secure the deposit of on-line digital publications as well. Members of the association were found prepared to deposit all digital publications with a Dutch imprint. The KB in turn took on the obligation, within its deposit remit, to preserve these publications in the e-Depot. As was the case with the printed deposit collection, access to digital publications is restricted to on-site authorised users only (open-access publications being freely available online).

Some of the first archiving agreements were signed with major scientific publishers based in the Netherlands, such as Elsevier and Kluwer. As these are international publishers, the question soon arose how digital resources which are simultaneously published all over the world, fit into traditional national deposit schemes. The answer was simple: they do not. The KB decided that a new international framework would have to be developed to preserve digital publications for the long term. As such a framework does not come to be overnight, the KB took a first step by opening up its own e-Depot facilities to *all* digital resources published by Dutch international publishers and to all major international scientific publishers. Quite a few publishers have meanwhile concluded archiving agreements with the KB, i.e., Elsevier, Springer, Blackwell, Oxford University Press, Taylor & Francis and Sage. As of April 2009, the e-Depot has ingested almost 13 million digital objects.

Considering the costs of the infrastructure and the complexity of sustainable storage, the KB has since decided to use the e-Depot for a wider variety of services. The e-Depot now

also preserves masters resulting from major Dutch digitisation programmes, the contents of the Dutch institutional repositories and the Dutch national Web Archive.

### **Organizational Implications**

It soon became clear that the e-Depot was not just another book storage facility that could be added without any organizational implications. New digital workflows had to be designed and as digital preservation was such a new topic, it was immediately clear that a fair amount of research and development had to be organized as well. The KB decided to create two distinct departments in support of the e-Depot: an operational e-Depot department embedded within the Acquisitions and Processing Division, and a research and development unit within the Research & Development Division.

The Operational e-Depot is staffed by ten full-time employees, six of which are collection managers. They are responsible for processing incoming data. Elsewhere in the organization, three employees provide IT support. The Digital Preservation R&D comprises six full-time staff. These numbers do not include temporary staff for international projects and the present programme to set up the next-generation e-Depot.

As obvious as the implementation of bulk workflows may seem from a strategic point of view, it was to be expected that work floor practices did not adapt as easily. Cataloguers who had for years taken pride in the quality of the descriptions they produce with quite a lot of manual effort, did not readily warm to the idea of automatically generated metadata, and the fragility of the digital media and data themselves worried them as well. These staff concerns were of course quite justified, but with the influx of millions of digital publications it became quite clear very soon that manual cataloguing was not a feasible option. Six years hence it is encouraging to see that workflows for printed and digital materials are converging and that staff are in fact benefiting from each other's expertise. The research team, in turn, greatly profits from the experiences of the day-to-day running of a digital archive.

### **The Next-Generation e-Depot**

On a technical level, the core of the e-Depot is the Digital Information Archiving System (DIAS), which was developed between 2000 and 2002. DIAS is designed according to the OAIS reference model to perform functions such as ingest, archival storage, data management and dissemination. Developing the e-Depot was a truly pioneering effort at the time, as there were no existing models or tools for quality assurance.

During the first six years of the e-Depot's operation, the focus was on processing e-journal articles. Generally these are objects of a similar type and format. The workflows designed to accommodate these e-journals clearly worked, as almost 13 million e-journal articles were ingested since 2003. However, the e-Depot is anticipating a future influx of many more different types of content from a variety of content suppliers, and the present technical environment is clearly not suited for such additional tasks. The architecture of the system requires major redesign, the infrastructure needs to be scaled up to be able to process larger amounts of objects of a wider variety and newly developed tools for quality control need to be implemented. Thus, early 2008 KB embarked upon a large-scale programme directed towards the development of a next generation e-Depot.

Collaboration with the other users of the DIAS system, the German National Library and the State and University Library of Göttingen in this effort was a matter of course. But quite recently the KB has also reached out to other national libraries.

### **National and International Outreach**

The challenges posed by permanent access are quite daunting, both in a technical and in an organizational and financial sense. No single institution or even country is capable of solving the many inherent problems single handedly. Moreover, responsibilities are diffuse. In order to deal with these enormous challenges, national and international cooperation is called for. From the start, the KB actively sought cooperation. The digital preservation research department participates in a number of major European projects designed to facilitate permanent access to digital objects. These projects focus mainly on technical issues: development of tools, services and preservation strategies. In addition, they are mostly of a temporary nature: once the projects are completed and funding ends, there is a risk that the results disappear into cyberspace.

But digital preservation is a long-term game, and definitely not just a set of technological problems. It involves the grander problem of organising ourselves over time and as a society. Digital preservation is also about selecting what materials should be preserved, and in what form (social and cultural issues); what rights are needed to support permanent access (economic and legal issues); who is preserving what (responsibilities). Digital preservation is an ongoing, long-term commitment, shared and met by many stakeholders.

There is a great need for long-term, sustainable partnerships with corresponding funding to take up where temporary (technological) projects leave off. The KB has taken the lead in developing such partnerships at several levels:

An example of cooperation on a practical level within The Netherlands is the network of Digital Academic Repositories, in which the National Library fulfils the role of safe place for academic output stored in institutional repositories. The participants retain responsibility for and control over their own data and provide access to them, while the KB takes responsibility for storage and long-term preservation, thereby enabling the universities to concentrate on their research work. This model is used in a wider perspective within the European DRIVER project.

At the national level the KB participates in the cross-domain Netherlands Coalition for Digital Preservation (NCDD) which rallies major stakeholders from the research community, government, and cultural heritage institutions around a joint strategic agenda for permanent access to digital information in the Netherlands.

On a world-wide level the KB promotes the idea of a Safe Places Network of trusted digital repositories which share the responsibility for keeping digital publications safe on a global scale. The KB is of the opinion that continuous research and development efforts require substantial financial, technical and staffing commitments that exceed the possibilities of individual institutions. As a consequence we expect that only a limited number of institutions will commit themselves to permanent archiving of the records of

science. A network of such 'Safe Places' would ensure a more systematic approach to digital preservation of scientific information.

### **In Conclusion**

The KB's e-Depot has now been in operation for six years, and the invaluable experience gained in these years, both in the day-to-day running of the e-Depot and in conducting research and development, now enable us to evaluate our position and the impact of the e-Depot on our entire organisation. But the rate of change in ICT, in publishing, in libraries and in preservation is rapid and the scale is enormous. Thus this is not the time for us to look back and be satisfied. Rather, new challenges come on our path every day and we are willing to take them on. Ensuring permanent access to digital objects is a complicated task which requires collaboration on a national, a European and a worldwide scale. It is hoped that more organisations and countries will join the existing initiatives to create a sustainable infrastructure to support permanent access to digital information on a global scale.

### **References**

- e-Depot of the National Library of the Netherlands, <http://www.kb.nl/dnp/e-depot/e-depot-en.html>
- *Addressing the Future of Preserving the Past: Towards a Robust Strategy for Digital Archiving and Preservation* (2007), Report by the RAND Corporation, prepared for the Koninklijke Bibliotheek, [http://www.kb.nl/nieuws/2007/rand\\_report\\_e-depot-en.html](http://www.kb.nl/nieuws/2007/rand_report_e-depot-en.html)
- Reference Model for an Open Archival Information System (OAIS) (2002), Consultative Committee for Space Data Systems, <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- DARE, Network of Dutch Digital Academic Repositories, <http://www.narcis.info/>
- DRIVER, Digital Repository Infrastructure Vision for European Research, <http://www.driver-support.eu/en/>
- NCDD, Netherlands Coalition for Digital Preservation, <http://www.ncdd.nl/en/>
- PLANETS, <http://www.planets-project.eu/>