

Permanent Access to the Records of Science - the International Role of the E-Depot at the Koninklijke Bibliotheek, National Library of the Netherlands

by GERARD VAN TRIER

ABSTRACT

It is expected that library staff are qualified to offer high quality services to users visiting the physical library. Likewise, it is expected that they have substantial knowledge and skills needed for developing and maintaining electronic services and for dissemination of relevant services and facilities requested by the web-user. Serving remote library users calls for additional competencies, such as marketing, branding and communications skills in the electronic environment as well as knowledge of measuring and evaluation of the use of electronic services. It is a challenge to the staff to match particular needs and demands from different user groups, but also to library management staff to ensure that the competencies and skills are available in the organisation to match the needs of the user – wherever s/he might be located. Competencies are, in this context, defined as the combination of knowledge and experience that make the individual able to take the right actions in the daily working environment.

What education and training needs emerge from the changing roles and new tasks? How might we identify the needs for continuing professional development? And how can we maintain and update skills and competencies acquired maybe 25 years ago? These are key questions – not only to be addressed to library managers, but also to be considered carefully by those institutions responsible for continuing education and professional development of library staff.

THE KB DEPOSIT COLLECTION

In the Netherlands the practice of depositing publications at the national library is relatively new. Only in 1974 the Koninklijke Bibliotheek (KB) started to build a deposit collection with the aim of preserving the cultural heritage of the Netherlands. Publishers had opposed against the adoption of a deposit law, which would have made the deposit of publications obligatory. As an alternative to legal deposit the Dutch Publishers Association agreed to promote voluntary deposit among its members. And they kept their promise. Agreements were signed with the KB and the Publishers Association, which still form a solid basis for the deposit collection. The voluntary scheme is successful: recent research shows that 97% of the production of official Dutch publishers is included in the deposit collection. This success was a key factor in the decision of the government to officially grant the KB the title of National Library in 1982, at the occasion of the inauguration of the new building next to The Hague Central Station.

In 1994 the KB decided to include electronic publications in its deposit collection. This was considered to be a logical extension of the existing deposit of printed publications. Small-scale experiments were set up with facilities for handling electronic publications, at first mainly CD-ROMs. We did not have a clear picture of all the problems but it was considered essential to gain hands-on experience. The aim was 'learning by doing'. In the beginning temporary solutions were implemented, first AT&T Right Pages (software now long forgotten) and later IBM Digital Library. These systems did not have the functionality of a full-scale deposit system. Such a system was not available on the market. In order to develop a system with the functionality that was needed, a European tender was organised. IBM made the best offer. In the project that followed the expertise of the KB and the technical knowledge and research forces of IBM were combined. In December 2002 the current e-Depot was delivered.

THE E-DEPOT

The technical 'heart' of the e-Depot is the IBM Digital Information and Archiving System (DIAS), which is compliant to the OAIS reference model (Open Archival Information System). It is integrated with other library modules and has the following functionalities:

- ingest of electronic journals, e-books, CD-roms (installable);
- preserving integrity and authenticity, standard formats (PDF, XML);
- automatic validation (checksum, JHOVE, error handling);
- metadata conversion, batch delivery.

The e-Depot is now fully operational and imbedded in the KB organisation. In the experimental years - that is until 2002 - the e-Depot contained only Dutch electronic publications. This was in line with the mission of any national library: to provide permanent access to the cultural heritage of the country. The KB and the Dutch Publishers Association first agreed on an arrangement for the voluntary deposit of *offline* publications. All the members of the Publishers Association were advised to deposit their electronic publications at the KB. At the same time the KB started discussions with Elsevier Science, because we wanted to have the content of Elsevier *online* journals in the e-Depot.

In 1996 the first bilateral archiving agreement with Elsevier was signed, which allowed the KB to load electronic journals *with Dutch imprint* in our first experimental deposit system. Soon after other Dutch publishers like Kluwer Academic and SDU Uitgevers (the former governmental publishing agency) followed. In the meantime the agreement with the Publishers Association was adapted as to include also online publications. So, at the end of the 20th century there were several agreements on the deposit of Dutch electronic publications. How did these agreements work out in practice?

INTERNATIONAL SCIENTIFIC PUBLICATIONS

I will focus on electronic journals because they have been our highest priority. A selection was made of electronic journals, which were considered to have a Dutch imprint. We looked at the title page of the printed equivalent. If Amsterdam or any other Dutch town was mentioned as place of publication, the journal was selected for deposit. These journals (a few hundred) could have been edited and printed anywhere in the world, but for the deposit they were considered to be Dutch. However, the distinction between Dutch and non-Dutch electronic journals seemed more and more artificial, arbitrary, and also impractical, because through the years the place of publication of a journal could change from Amsterdam to New York or vice versa. It is a fact of life that electronic journals of international scientific publishers have no obvious country of origin. It was a logical step therefore that Elsevier agreed in 2002 to archive all Elsevier journals at the KB e-Depot (so-called Dutch imprint or not). For the deposit of scientific journals the geographical criterion (Dutch imprint) had lost its relevance.

In 2003 Kluwer Academic Publishers signed a similar agreement. The next publisher was BioMed Central. This contract signified again an important step, in two ways. First, it underlined the international role of the national deposit system. BioMed is not a Dutch publisher, nor does it have its roots in the Netherlands. Furthermore BioMed is an open access publisher. This also was new. Other publishers followed: Blackwell, Oxford University Press, and Taylor and Francis in 2004; Sage, Springer and Brill Academic Publishers in 2005. On the basis of these agreements the e-Depot will eventually hold 9 million articles. The annual increase in the number of articles from these publishers will be around 400,000.

CONDITIONS FOR ARCHIVING AND ACCESS

Publishers have to deposit their publications free of charge. On the other hand, the KB has accepted restrictions on access. But there is a bottom line: the KB offers onsite access for any registered user and availability for interlibrary document supply within the Netherlands. Remote access can be provided if the publishers allow this. For example, BioMed Central allows free remote access to over 100 open access journals covering all areas of biology and medicine. Retrieval, access, printing and downloading is for private use only, systematic reproduction is not allowed. These conditions guarantee that the e-Depot service does not interfere with the publisher's commercial interests. At the same time the KB promises to safeguard permanent access and to protect the authenticity and integrity of the content delivered. That is the benefit to the publisher.

BENEFIT TO LIBRARIES AND USERS

It is evident that registered users who are able to visit the KB have access to a huge amount of journal articles in the field of science, technology and medicine. But that is not the primary objective of the e-Depot. The KB is not an STM library and has no ambition to develop end-user services for this kind of publications. The objective of the international e-Depot is preserving the records of science. The main benefit is for future users. For the short term it is primarily a business-to-business service: if a catastrophic disaster would occur and the publisher is inoperable for a long period of time, the KB would be part of the interim service system. The official archive thus serves as a guarantee to all licensees worldwide, by safeguarding the access that licensees have paid for. Finally, should the publisher or a successor cease to make these journals available, the KB could provide access to all on a walk-in or remote basis. In this way, the KB secures permanent access to both libraries and users.

SAFE PLACES MODEL

I have explained how the early and successful implementation of the e-Depot and the commitment of Elsevier and Kluwer put the KB in a natural position to assume an international role. How does this role fit into worldwide developments regarding the preservation of the digital heritage? Gradually, more and more national deposit libraries will build electronic deposits for permanent access to their national cultural heritage and also to their *national* scientific publications. However, as was noted before, international STM literature in electronic form has no longer a country of origin that can be easily identified and therefore has no obvious guardian. When we stick to the traditional model of dividing preservation tasks on the basis of geographical frontiers there is a huge risk of the records of science being lost forever. This is of course unacceptable. For preserving the records of science a new model, a systematic and more concentrated approach, is needed. This is called the Safe Places Model.

This model is directly derived from the requirements for permanent archives. Permanent archives presuppose permanent commitment. A permanent archive should provide a reasonable guarantee for continuity. Substantial investment is required, not only financially, but also in the form of building up specific skills and expertise. Moreover, the preservation function will require an unremitting research and development commitment. From these requirements it follows that permanent archiving should be taken care of by a limited number of institutions: Safe Places. A Safe Place is dedicated to permanent archiving. Permanent archiving is prominent in its mission. For geo-political reasons Safe Places should be wisely spread around the globe.

MATERIAL AND STAFF COSTS

One of the requirements for a Safe Place is substantial investments. The bad news is that the fixed costs of a permanent archiving system are relatively high. In order to finance the service the KB had to re-allocate funding within its own budget. Fortunately, the KB receives an additional permanent grant of €1.1 million from the government for system maintenance and operating the service. IBM has been contracted for the maintenance. The total number of staff for handling the system, ingesting the publications, research projects, and management is about 15 full time equivalents.

The good news is that there are considerable economies of scale. As a national library the KB has the obligation anyway to maintain an archiving system for Dutch electronic publications. Now the system has been installed and works smoothly, the storage capacity can easily be expanded and used for preserving a variety of national digital collections of universities, museums etc. Costs per unit are going down.

RESEARCH & DEVELOPMENT

Another important requirement for a Safe Place is a permanent and substantial Research & Development (R&D) effort. The Dutch government provides an annual grant of € 1.4 million, earmarked for research into digital preservation. The funding is expected to increase next year. For maintaining accessibility sustained investment in R&D is necessary because technology keeps changing. New technologies provide new solutions in the future. It's a permanent effort to keep the ever-changing toolbox up-to-date.

The research efforts of the KB are focussed on the full range of available preservation techniques: migration as well as emulation. Migration or conversion changes or updates the format of the object, so that it is available on new software and hardware. New standards and new functionalities can be used. This is important for future users who are only interested in the content and not in the look and feel of the original document. However, migration might affect the content and thereby the authenticity of the object. If authenticity or the look and feel is important, then emulation is necessary. Emulation means that a new rendering tool is instructed to behave like an obsolete one. The authentic form of the publication is preserved, which could be important not only for the end user. It is also highly valued by the publishers, who need a safe place, which preserves the original bit stream of the document. Both models are studied and considered for implementation at the KB. Other issues are the storage of technical metadata about the technical properties of the stored file formats and the extraction of technical information from the delivered publications.

The KB has its own R&D-programme, but this is not executed in isolation. On the contrary, the KB is actively contributing to international collaboration. An important step forward is the European project PLANETS (Preservation and Long Term Access through Networked Services), which is coordinated by the British Library within the 6th Framework Programme of the European Commission. The result of the project will be a distributed preservation framework for the development and application of instruments for preservation planning, preservation tools and content characterisation. It will also include a decision support system, which will help institutions to decide which preservation strategy suits their situation best. The group of partners is quite diverse: European national libraries and archives, leading research institutions, and technology companies. These companies will build tools and the technical infrastructure that will allow us to set up work together in a networked environment. The participation of these commercial partners facilitates the take-up and dissemination of research results.

EUROPEAN TASK FORCE ON PERMANENT ACCESS

The results of Planets will be important, but more research is needed and the cooperation between major stakeholders should be consolidated. The KB and The British Library have formed a high-level European Task Force Permanent Access, with representatives from the science sector, academic publishers and libraries. This Task Force has produced an elaborate R&D programme, which is submitted to the European Commission within the context of the 7th Framework Programme. Furthermore, the Task Force took the first step to the creation of a so called 'Alliance for Permanent Access'. Major stakeholders in Europe across the information chain from producers to archives to users have decided to form this Alliance. The aim is to establish broad consensus on a strategic level as to the main characteristics of a European infrastructure for permanent access to the records of science. Among the initial partners are large scientific organisations like ESA, CERN and the British Council for the Central Laboratory of the Research Councils, but also JISC, the European Science Foundation, the International Association of STM Publishers, the National Archives of Sweden, The British Library and the KB.

The KB also aims for cooperation with other parties that emerge within the Safe Places model. One example of such a party is Portico, an electronic archiving initiative launched in 2005 with support from JSTOR, Ithaka, Library of Congress and the Andrew Mellon Foundation. Portico aims to ingest three million pages by the end of 2006, based on formal archiving agreements with international publishers. The first contacts with Portico have already been made.

MORE AMBITIONS

Other ambitions the KB pursues:

- The KB will develop further the technical and organisational infrastructure and even better fulfill the requirements for a permanent archive. An external audit by a group of experts from the US National Archives and Records Administration (NARA) and Research Libraries Group (RLG) in April 2006 gave us confidence that the KB is on the right track.
- The KB will actively try to conclude archiving agreements with more of the major international scientific publishers. The 20 largest publishing companies cover around 80% of the total world production of STM literature. The KB would like to reach that level of coverage in the e-Depot. The KB will also try to obtain the most cited scientific journals, irrespective of the publisher. Next to this active strategy, the KB will accept STM literature of any other publisher, who wishes to deposit material, provided that the publisher is able to deliver the

material in a standard format and with the necessary metadata. And provided the publisher complies with the minimum set of conditions with regard to access.

- Apart from archiving scientific digital publications the KB explores opportunities for the long-term storage of other material, for instance TIFF files, websites and digitised material. The KB also collaborates with the Dutch university libraries in the DARE project. The aim of the project is to provide permanent access to the publications in the Digital Academic Repositories. The object domain for preservation is defined as those objects that are made available on internet through the Institutional Repositories.
- De KB would like to discuss the feasibility of new business models to cover the costs of storage space, processing and management of the material, and R&D efforts.

NEW BUSINESS MODELS FOR PERMANENT ACCESS

In the Safe Places Model, archiving institutions provide a business-to-business service to academic institutions and publishers. In the paper-based model, publishers have never regarded the long-term storage of their material as one of their responsibilities and this will not be different in the digital world. Few publishing houses possess the kind of infrastructure or expertise that would enable them to provide sufficient security and continuity. What's more, subscribers prefer this safeguard to be provided by an independent third party.

At the moment our archiving agreements determine that the parties each bear their own costs. This means that the KB bears the cost of the archive and the publisher the costs of supplying the publications. This means that the e-Depot depends in fact on public funding by the Dutch government. The government should indeed cover most of the costs, because the technical and operational infrastructure is needed anyhow for preserving the digital heritage of the Netherlands. Another argument in favour of the present financial construction would be the long-term character of digital archiving. It is rather difficult to induce the end-user to pay for a service that does not serve any immediate advantage to him. There are simply no direct incentives. Research libraries on the other hand, do have an immediate interest in long-term digital archiving as an insurance against the permanent loss of electronic journals. The Portico archive model builds on this interest by charging participating libraries a one-time and annual financial contribution. So, this is a model that could be explored.

But, above all, the electronic deposit provides a business-to-business service to the publishers. If an international network of safe places is created in which academic publishers deposit their electronic material, it makes sense to assume that the charges for digital archiving will eventually be incorporated into the overall costs of publication. This will effectively mean passing on the expenses to the publisher, and ultimately to the consumers. In practice the sums involved will amount to only a small proportion of publication costs. The KB would like to work on this basis, but this construction could only be put in place after a discussion with all parties involved.

RELATED ARTICLES

Drimmelen, W. van: "Universal Access through Time: Archiving Strategies for Digital Publications". *Libri, International Journal of Libraries and Information Services*. 54(2004)2. <http://www.librijournal.org/pdf/2004-2pp98-103.pdf>

Oltmans, Erik and Hilde van Wijngaarden: "Treasuring the Records of Science. The International e-Depot". *Library Hi Tech*, 2006, no 4 (to be published)

WEB SITES REFERRED TO IN THE TEXT

DARE - Digital Academic Repositories. <http://www.darenet.nl/en/page/language.view/search.page>

European Task Force Permanent Access. <http://tfpa.kb.nl>

NARA - National Archives and Records Administration. <http://www.gpoaccess.gov/nara/index.html>

Planets - Preservation and Long Term Access through Networked Services. <http://www.planets.arts.gla.ac.uk>

Portico. <http://www.portico.org/>

RLG - Reseach Libraries Group. <http://www.rlg.org/>