

Permanent archiving of electronic publications

This article is an extended version of the presentation by the author at the 4th e-COLC Conference, held at Thessaloniki, Greece, in 2002

In recent years more and more scientific journals have become digital, yet still the majority of them are also published in print. An important reason for this is that librarians seem unwilling to give up print even if their users prefer the electronic form, because a library keeps back volumes of journals to guarantee the permanent availability of the information, and librarians cannot yet guarantee permanent availability of the electronic format. Recently the Koninklijke Bibliotheek (KB) – the national library of The Netherlands – has implemented an operational electronic deposit for permanent archiving of electronic publications. The deposit system is compliant with the Open Archival Information System (OAIS) archival standard and has been developed together with IBM Global Services. Acknowledging the necessity of permanent digital archiving, Elsevier Science (ES), a major science publisher, has recently signed a unique archiving agreement with the KB. ES will deposit all its electronic publications at the KB for permanent archiving and access.



JOHAN F. STEENBAKKERS

Director IT & FM of the
Koninklijke Bibliotheek,
The Netherlands

Introduction

In recent years more and more scientific journals, especially those in the fields of science, technology and medicine, have become digital. For some publishers the earnings from the electronic format today exceed those from the printed format. The advantages of the electronic format are obvious, both for users and publishers. Publishers can offer services with added value, like easy searching and instant delivery, and at the same time the production costs are lower so users can be offered more value for money.

Still the majority of the electronic journals, however, are published also in a printed form. An important reason for this is that librarians seem to be unwilling to give up print even if their users apparently prefer the electronic form. This is because a library as a rule keeps back volumes of journals to guarantee the availability of the information, either for a certain period, or in the case of a deposit library, permanently. The assurance of publishers that they will guarantee permanent access to their electronic publications, cannot convince librarians to cancel print

subscriptions. This reluctance can be understood, as nobody has yet in practice developed a solution to the problem of permanent archiving of electronic publications.

Recently in The Netherlands an operational electronic deposit, especially designed for permanent archiving of electronic publications, has been developed and implemented. As a result of a series of projects and other preparatory activities, the Koninklijke Bibliotheek (KB) was able to define and realise the conditions needed for permanent archiving of electronic publications^[1]. One requirement is to acquire a dedicated deposit system, and another to develop practices and tools for permanent access to electronic publications. Acknowledging the need for permanent digital archiving, Elsevier Science (ES), a major science publisher, recently signed a unique archiving agreement with the KB^[2]. ES will deposit all of its electronic publications at the KB for permanent archiving. In this article I will describe the approach followed by the KB to achieve the conditions for permanent archiving and will touch

upon the nature of the archiving agreement and its implications.

The deposit task extended

Before describing the development of the e-deposit of The Netherlands, I will give some facts about the library. The Koninklijke Bibliotheek is the national library of The Netherlands and was founded over 200 years ago. The staff comprises about 260 full time equivalents and the annual budget of the library is about 37 million euro. More information about the KB and its activities is available on the library's website at www.kb.nl.

A key task of the KB is to act as a deposit library for the publications produced in The Netherlands. Even though the library dates back many centuries, the deposit task was given formally to the library only some 30 years ago. The aim of a deposit library is to collect published information, preserve it and provide permanent access to the information for use in research, education or for any other purpose in society.

In most countries publications have to be deposited by law, but The Netherlands has a voluntary deposit based upon agreements of the national library with publishers. This approach has been quite successful, resulting in an almost complete coverage of the publications produced by the commercial publishers in The Netherlands. Traditionally the deposit task concerns printed publications, like journals, books, newspapers. As more and more publications are published electronically, the Koninklijke Bibliotheek in 1994 extended the deposit task to include electronic publications.

Realising the e-deposit

To be able to incorporate electronic publications in its deposit library, the KB needs a deposit system. This requirement is obvious if we consider that a library needs specific facilities, namely stacks, to store and maintain the collection of printed publications. In a similar way a specific facility is required for the storage and maintenance of electronic publications, and that is a deposit system. By transferring the electronic publications from the environment in which they are published (cd-rom, internet etc.) to the specific, controlled environment of the deposit system, the library is

able to manage and control the electronic collection, in order to preserve it and achieve permanent availability of the publications.

Once a specific deposit system is available, it has to be implemented at the library. This means for instance to connect the system to the digital library functions already in place at the KB. There are good reasons to design the deposit system as a separate dedicated system, with clearly defined interfaces towards its digital library environment^[3].

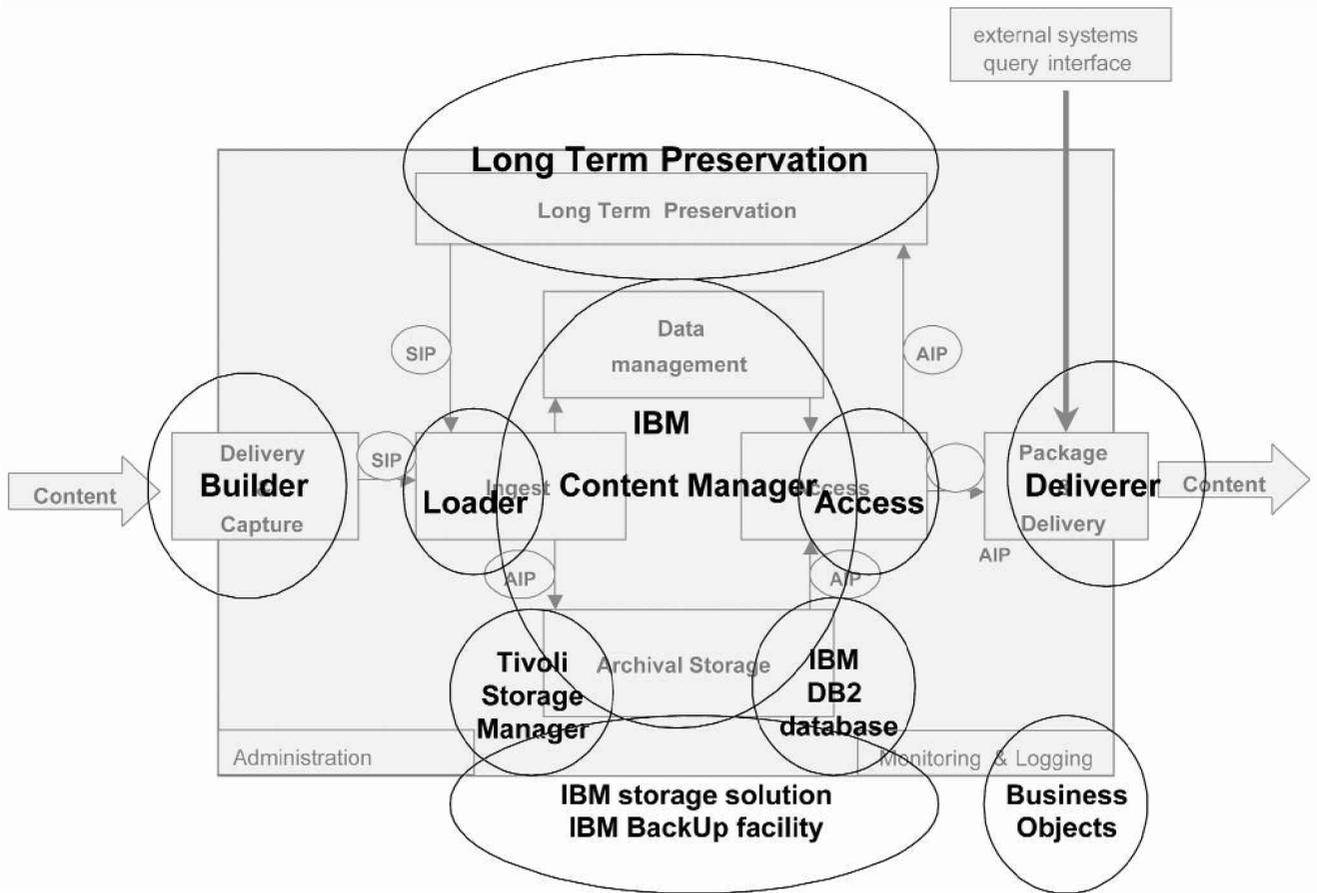
An obvious step in the realisation of the e-deposit is to acquire the electronic publications. So, depositing agreements with the producers of the electronic publications, e.g. commercial publishers, universities and governmental agencies, have to be developed^[4].

Acquiring the deposit system

From market research performed at the beginning of 1999, the KB learned that a deposit system could not be bought off-the-shelf. At the same time it became clear that the ICT companies, involved in the market research, recognised digital archiving as an important topic and that they seemed interested to develop solutions. Therefore the KB decided to proceed with the acquisition of the deposit system.

The first step taken was to describe in detail the functional requirements of the deposit system. This was done, taking into account the results of various studies and research projects on digital archiving and long-term preservation. The next step was for the KB to look for an appropriate technology partner to design and develop the deposit system. At the end of 1999 a European call for tender was published which resulted in 4 proposals being received which qualified. More information on the tender process is available on the KB website. From the offers received the KB eventually selected IBM – Global Services – as the best potential technology partner and started contract negotiations.

After a period of negotiation a contract was signed and the development of the deposit system started in October 2000. IBM developed the deposit system on site at the KB premises. Parallel to the system development, IBM and KB jointly studied a number of practical issues on long-term preservation. This study was an integral part of the KB-IBM contract, intended to support the



design and development of the system [5]. The system was delivered to the KB on schedule in October 2002.

The functional design of the deposit system is based on the international standard for digital archives OAIS (Open Archival Information System reference model) as it has been detailed for national libraries by the European project NEDLIB (Networked European Deposit Library) [6]. According to the requirements of the KB, IBM constructed the system as much as possible using off-the-shelf components. IBM has recently branded the deposit system with the name DIAS (Digital Information Archiving System).

Implementing the deposit system

To achieve a durable solution, the KB has implemented the system, according to the NEDLIB Guidelines[3], as a distinct entity within its digital library environment. The key functions of the DIAS are archival storage and long-term preservation. Other functions needed for handling and use of the electronic publications are provided by the other systems already in place within the library (not shown in the figure). These systems provide the

more traditional library functions like acquisition, cataloguing, search and retrieval, user registration, authentication and authorisation of users.

Archiving agreements with publishers

In August 2002 the KB signed a unique archiving agreement with ES. The aim of the archiving agreement is to ensure permanent archiving of all the electronic journals owned by ES. This means that the KB has become a formal archival agent of ES, providing a number of functions. Amongst these functions are permanent storage, maintenance and preservation of ES journals and ensuring their permanent availability. But the function of an archival agent can also involve acting as host for actual or former customers and guaranteeing continuity of access in case of an emergency, as a consequence of which the publisher can no longer provide the information to the customers.

The work to implement the archiving agreement is in full progress and in the beginning of 2003 the electronic publications of Elsevier Science will be loaded. Most of ES publications are journals (over 1500 different titles) of which both the current and back volumes will be archived.

Meanwhile, other publishers have approached the KB to come to similar archival arrangements.

Status and further plans

To conclude I would like to sum up the actual situation and further plans.

Amongst the results achieved is the realisation of a technically durable infrastructure that offers all basic functions for digital archiving. Also the workflow for processing, maintenance and delivery of electronic publications has been developed and implemented in the library. Last, but not least, publishers have acknowledged the importance of permanent archiving of electronic publications by national libraries, and fully support its implementation.

The results of the practical long-term preservation study by KB and IBM are promising and will be used to develop further the preservation management function of the deposit system^[7]. In addition the KB and IBM will jointly develop the UVC approach into an operational preservation tool^[8]. This tool will support the permanent access of the electronic publications that are stored and preserved in the e-deposit.

The KB will also start experimenting with permanent archiving of information harvested from the internet. Web publications impose additional requirements on an electronic deposit system^[9]. Web-archiving activities executed today in several countries have provided much experience on web-harvesting, however the actual web-archives have not been designed to address long-term preservation. It is this problem that the KB intends to focus on, using the best practices on web-harvesting for collecting the material to be permanent archived.

References

1. Steenbakkers, J., Developing the Depository of Netherlands Electronic Publications, *Alexandria*, Vol.11, no.2, pp93-105, 1999.
2. National Library of the Netherlands and Elsevier Science make digital preservation history. (Press release of Elsevier Science and the Koninklijke Bibliotheek), Glasgow, 2002:
<http://www.kb.nl>
3. Steenbakkers, J., *Setting up a Deposit System for Electronic Publications*. The NEDLIB Guidelines. (NEDLIB Report Series no. 5), 2000, The Hague, Koninklijke Bibliotheek.
4. *Statement on the Development and Establishment of Codes of Practice for the Voluntary Deposit of Electronic Publications*. 1999, Conference of European National Librarians / Federation of European Publishers (CENL/FEP):
<http://www.bl.uk/gabriel>
5. Diessen, R.J.van, and Steenbakkers, J. F., *The Long-Term Preservation Study of the DNEP project. An Overview of the Results*. (IBM/KB Long-Term Preservation Study Report Series no. 1), 2002, Amsterdam, IBM / The Hague, KB.
6. Werf, T. van der, *The Deposit System for Electronic Publications. A process Model*. (NEDLIB Report Series no. 6), 2000, The Hague, Koninklijke Bibliotheek.
7. Diessen, R.J. van, *Preservation Requirements in a Deposit System*. (IBM/KB Long-Term Preservation Report Series no. 3), 2002, Amsterdam, IBM / The Hague, KB.
8. Lorie, R., *The UVC: a Method for Preserving Digital Documents. Proof of Concept*. (Long-Term Preservation Study Report Series no. 4), 2002, Amsterdam, IBM / The Hague, KB.
9. Verhoeven, H., *Archiving Web Publications*. (Long-Term Preservation Report Series no. 6), 2002, Amsterdam, IBM / The Hague, KB.

Johan F. Steenbakkers

**Director IT & FM of the Koninklijke Bibliotheek,
National library of the Netherlands,
P.O. Box 90407 NL-2509 LK The Hague, The
Netherlands
E-mail: johan.steenbakkers@kb.nl**
