Digital Archiving in the Twenty-First Century: Practice at the National Library of the Netherlands

JOHAN F. STEENBAKKERS

ABSTRACT

Research journals are increasingly being published digitally. The advantage of digital publishing is obvious: immediate accessibility anywhere. Gradually a disadvantage is also becoming clear: digital publishing endangers the continuity of research information. As a consequence of the obsolescence of formats, hardware, software, and carriers, digital information will be lost unless we act. Digital publishing is also causing a shift in the roles and responsibilities of publishers and libraries concerned with archiving digital publications for future use. Archiving digital publications requires a major turnaround in the policy and practice of national libraries. Although some actions have been taken, digital preservation research and implementation are still in their infancy. National libraries will need substantial funding for venture research activities and development of archival infrastructures. They will also have to work together more closely to successfully organize digital archiving in the twenty-first century.

In 1994 the Koninklijke Bibliotheek (KB), the national library of the Netherlands, responded to the risk of losing digital information and started to include digital publications in its depository library. For this purpose the KB designed the "Safe Place Strategy," a strategy ensuring the transfer of digital publications from their publishing environment to a dedicated archiving environment. Through a series of experiments and projects the KB developed skills, procedures, organization, and infrastructure for digital archiving. The digital archive of the KB, the e-Depot, has been fully operational since early 2003. The technical core of the e-Depot is the Digital Information Archiving System (DIAS), a system jointly developed by the KB and IBM. DIAS is the first archiving system available on the

market that is compliant with the Open Archival Information System (OAIS) reference model standard. Elsevier and several other international scientific publishers have entered into a formal archiving relationship with the KB to guarantee future availability of their digital publications. The KB is researching and testing digital preservation methods and solutions that will guarantee permanent access to the archived material.

Archiving in the Digital Era

The Impact of Digital Publishing

In the past decade journal publishing has increasingly turned digital, a change led by publishers in the field of Science, Technology, and Medicine (STM). Generally, all STM publishers now publish their journals digitally with printed derivatives only produced as a secondary format. Without doubt the pioneer of digital publishing is Elsevier (Elsevier, n.d.). Early work on digital publishing began in 1979 with project ADONIS on CD-ROM and, in 1991, with project TULIP for desktop delivery of e-journals. In 1995 the TULIP project was followed by ScienceDirect OnSite and in 1997 by the Web-based service ScienceDirect. New publishing models are under development, such as SPARC and BioMed Central.[1] Even though the final outcome of these experiments is still uncertain (Elsevier, 2004), it is obvious that digital publishing is a success and here to stay. This is evidenced by the increasing number of scientists who use the electronic services offering digital publications.

For publishers, digital publishing has shifted the emphasis from the production and selling of publications toward creating and providing enhanced services for searching, linking, and retrieving digital information. Despite the advantages, however, digital publishing poses a serious threat to the continuity of the "Record of Science." In the past, libraries had assumed the responsibility for archiving the printed journals they bought and collected. Now, with digital resources, libraries do not own the publications but license them from the publishers. Therefore, libraries no longer compile a collection, and the digital publications remain with the publisher. As a consequence the archiving of printed publications implicitly offered by library collections no longer exists for e-journals.

Changing Roles Subsequently, archiving needs to be organized explicitly for e-journals. To accomplish this, some institutions have taken the initiative to investigate and develop the archiving of digital publications. There are several significant examples of such initiatives,[2] and recently several national programs[3] have been set up to encourage the development of digital archiving.

Digital publishing not only affects the archiving role of libraries but also the role of publishers. To cover their share of responsibility to maintain the Record of Science, it is no longer sufficient for publishers to only produce,

market, and sell the publications. An active role is required to plan and organize the archiving of the published information. Scientific publishers have become aware of this change and are rising to the challenge by implementing archiving policies.[4]

Archiving Digital Information Libraries need to encourage the publishers to deposit their digital publications for archiving. However, this is only the first part of the digital archiving story. To understand why archiving of digital information is more complicated than archiving printed information, it is helpful to compare printed publications with digital publications.

A printed publication is a physical object. A digital publication, on the contrary, is not a physical but a logical object stored on a physical medium. Another difference is that a printed publication presents the information so that it is immediately accessible to the human eye and can be read directly. To read or view the information in a digital publication, specific functionality enabled by software and hardware is required. Journalist Richard Poynder described this difference as follows:

> unlike paper or microfilm where the meaning is transparently inscribed on the surface of the medium—digital documents are opaque bit streams only understandable to humans when interpreted by a machine. The hardware and software to do this interpretation, however, is constantly superseded. There have, for instance, been more than 200 digital storage formats alone deployed since the 1960s, with none lasting more than 10 years. (Poynder, 2003)

Another difference is that publications on paper or vellum can survive "by accident" and remain readable for a very long time without any specific actions. Visualize, for instance, a manuscript such as the *Book of Hours* from the fifteenth century, once owned by the princes of Trivulzio living in Milan. The KB received this beautiful illuminated manuscript[5] as a valuable gift, worth at least six million Euro, from an anonymous donor. It was assumed to be lost until the family of the donor acquired the manuscript through an auction some time ago. Would a digital publication survive left sitting on a shelf for hundreds of years? Considering the threats to digital publications, the answer is definitely negative.

Threats to Digital Information One threat to digital information is that the physical carrier of a digital publication will deteriorate much faster than paper or vellum. The format of the digital object can be damaged or lost and may no longer be intact or retrievable. But even before that happens, the technology used to store the publication is likely to become obsolete. Another threat is the loss of the functionality needed to interpret, display, and use the information contained in the digital object. Without this functionality, provided by specific hardware and software, the information will not be available even if the bitstream of the digital object has been preserved. When we are able to address all the threats to a digital publication, we can successfully keep it for future use. If we cannot, sooner

or later we will lose the digital publication itself or the access to the information it contains.

Permanent Access Solutions To guarantee permanent access, a variety of solutions will be required (migration, normalization, emulation, or others). The choice of which techniques to use will depend on the nature of the digital publication (for example, what is the format, is it static or executable, etc.) and also on the requirements of the user (will they want to simply view or also process the information).

Migration (sometimes referred to as conversion) is a commonly applied technique in computer science. However, migration in essence makes changes to the original and implies a risk of damaging the information, a risk that increases if a sequence of migrations is applied. This could potentially mean the loss of the information. Normalization is migration of the information to a specifically chosen format before it is accepted into the archive. This approach explicitly accepts the possibility of loss of specific characteristics and of part of the digital information. However, in some cases it might be useful to hold only a limited number of formats in an archive. Emulation is a totally different approach aimed at rendering the digital information as authentically as possible. The risk of technology obsolescence is addressed by replacing the hardware with a computer written in software (an emulator). The use of emulation as a permanent access solution was suggested a decade ago (Rothenberg, 1995).

Sometimes a combination of emulation and controlled migration can be used, an attempt to combine the best of two worlds. An example of such an approach recently developed in practice is the Universal Virtual Computer (UVC) approach from Raymond Lorie, a researcher at IBM (Lorie, 2002, 2004). The UVC is a program for a virtual computer that performs all the essential functions of a regular computer. The UVC program is available for testing and evaluation at Alphaworks.[6]

Application of the UVC The KB and IBM have applied the UVC to offer permanent access for selected formats. The approach requires that an easily readable technical description of the UVC is preserved. Using this description, an UVC emulator can be created at any time for any given platform. The UVC emulator will run the programs that have been written for the UVC and that have also been preserved. In figure 1 the components of such a permanent access solution are shown. In the archiving phase, the original image, the UVC Specification, the Decoder, and the Logical Data Schema are stored. In the future, in the delivery phase the stored elements will be used to build an UVC emulator to run the Decoder that generates the Logical Data View. Finally, a Viewer is built that will restore the original image. The KB and IBM have recently developed and tested the first operational permanent access UVC tool for JPEG (Wijngaarden & Oltmans, 2004).
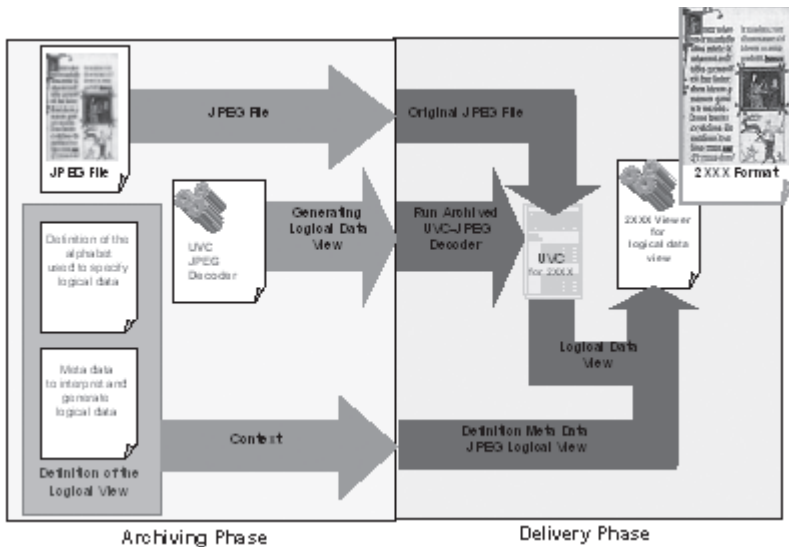
Figure 1.   The UVC Approach

Long-Term Preservation Research

   Achieving Permanent Access As was explained before, digital preserva-tion requires both the permanent maintenance of the bitstream of the digital publication as well as the maintenance of functionality to render the information from the bitstream now and in the future. Working on permanent access solutions is complicated and costly. Hence, as informa-tion technology (IT) continues to develop, ongoing research is needed to address the obsolescence of formats, software, and hardware. It is essential that institutions researching permanent access solutions cooperate closely and exchange experience and results. To support this, a joint terminology and joint requirements for interoperability would be very useful, as well as a joint framework for the research.

   A Framework for Permanent Access Research The idea to create a joint framework for permanent access research was introduced by a consortium[7] in a project proposal to the European Commission, under the title Permanent Access Toolbox for Digital Cultural Heritage, or PATCH. The Permanent Access Toolbox, or PATbox, is a concept representing the functional and technical requirements to which permanent access solutions should comply (see figure 2). The aim of the PATbox is to promote interoperability and increase the usefulness of the results of permanent access research, regardless of where or by whom it was performed. In addition to manifest-ing a framework of requirements, the PATbox also represents the collection

PATbox

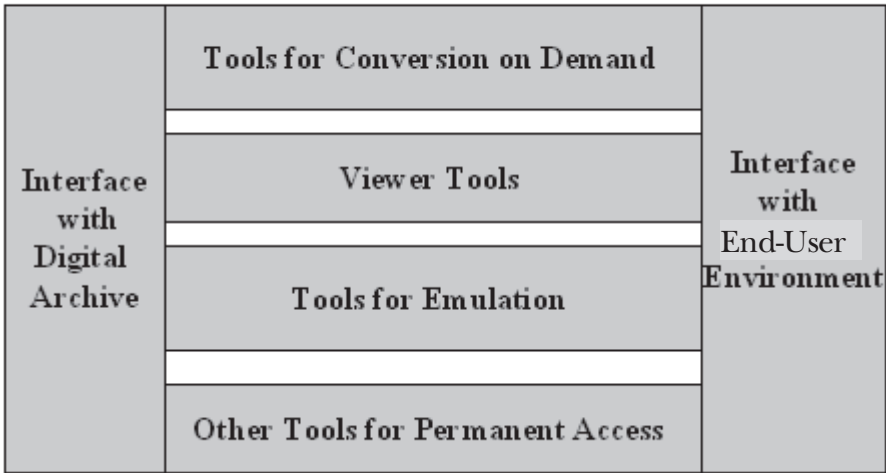| Interface with Digital Archive | Tools for Conversion on Demand | Interface with End-User Environment |
	Viewer Tools	
	Tools for Emulation	
	Other Tools for Permanent Access	

Figure 2.   Preliminary Schema Model of the PATbox Model

of generally available permanent access tools. Access tools may be based on different techniques or approaches, such as conversion, emulation, or other technical solutions.

ORGANIZING ARCHIVING IN THE TWENTY-FIRST CENTURY

Digital preservation of and permanent access to digital information are complex problems that can only be addressed successfully by the co-operation of several parties. National libraries have come forth as a major stakeholder. Their core business is to keep published information available in the long term. Another integral party is comprised of the governments and international political bodies who are needed to support national libraries with appropriate policies and legislation and who should also assign the resources needed to implement digital archiving in practice. Some important examples are discussed later. The third party involved is the publishers. In contrast to the world of printed publications, archiving digital publications requires the active involvement of publishers (Steenbakkers, 2004). Permanent archiving will also require close cooperation with key players in the IT industry to develop new technological solutions and new standards. Their support will be essential to organize the registration of file formats[8] and software versions and to realize global software repositories. Hopefully, academic research institutions will also contribute to the development of durable digital information as this problem is not restricted to publications but also concerns digital research data in general. Not only will the parties directly involved benefit from these efforts, but also society

as a whole. Digital information, either for business or for private use, needs to become durable.

Academic and national libraries have always collected and preserved publications, but archiving of digital publications requires a new, proactive approach. Considering the apparent progress in archiving technology, it is feasible that more and more academic libraries will be able to keep digital information in institutional repositories for the (relatively) short term. Commitment to long-term digital preservation, however, can only be expected from institutions like national libraries, which are dedicated to permanent safeguarding of information.

The national libraries at least will have to take responsibility for archiving the digital publications from their own country, often written in the national language. Just a few years ago the preservation of digital information was still considered to be so complicated and costly that making prints of digital texts and images on acid-free paper was suggested to be an appropriate solution (Gorman, 2001). As off-the-shelf archiving systems are becoming available today (Hodge & Frangakis, 2004, pp. 41–44), it is feasible that national libraries can start implementing an archive of their national digital publications and contribute to the universal control of digital content in the twenty-first century. Under the umbrella of the IFLA-CDNL Alliance for Bibliographic Standards (ICABS), a group of national libraries is looking into the matter of bibliographic and resource control in the digital era (ICABS, 2004).

For international digital publications, such as those that are part of the international Record of Science, preservation will have to be organized beyond national boundaries at a global level. One solution being developed is for publishers to establish a formal archiving relationship to work with a limited number of institutions distributed across the world that qualify as international archiving centers (Hunter, 2002). A substantial number of major international science publishers are already following this model and archiving their e-journals at the KB.

To achieve permanent archiving of electronic publications, the involvement of political bodies is essential. Support is needed not only from the local or national government but also from international bodies. Several of them are already involved. From the European perspective, two initiatives from 2002 are particularly relevant. One of them was initiated by the European Commission, which produced the experts report "Preserving Tomorrow's Memory" (Smith, 2003), and the other was initiated by the European Council, resulting in an EU resolution (Council of the European Union, 2002). The European Commission has made digital preservation one of the five themes of the last call for proposals of its Sixth Framework Programme (2002–06).[9] An example of political activity at a global level is the UNESCO initiative, which resulted in a charter on digital preservation containing recommendations to governments. In addition, UNESCO has

published the *Guidelines for the Preservation of the Digital Heritage* (UNESCO, 2003).

DIGITAL PRESERVATION IN PRACTICE

Digital Archiving in the Netherlands

The national library of the Netherlands, the Koninklijke Bibliotheek, was founded in 1798. The organization comprises the equivalent of 290 full-time staff, and the annual budget is 38 million Euro. The KB is located in The Hague, the residence of the government of the Netherlands. One of the key tasks of the KB is to act as the depository library for publications. The aim is to collect the publications, preserve them, and provide access to them, now as well as in the future. Depositing of publications at national libraries can either be regulated by legislation or be arranged by voluntary agreements. In the Netherlands deposit is based on voluntary arrangements between the KB and publishers.

Traditionally the depository task of the KB concerned only printed material, like journals, books, and newspapers. In 1994, as more and more publications became digital, the KB decided to start including digital publications in its depository collection. To implement this the KB officially extended the depository task to include digital publications; began cooperation with publishers and IT-partners; and established research and organization of digital archiving in practice.

Breaking New Ground

In the early 1990s when the first digital publications were deposited, the KB treated them as printed books and shelved them in the book stacks. They were "offline" or "handheld" digital publications and were catalogued on the basis of information on the wrapper or box of the publication. In order to learn to handle digital publications appropriately, the KB had to break new ground in a variety of areas. The library staff had to develop completely new procedures and acquire new skills. The KB also had to develop close cooperation with IT partners and had to obtain a dedicated IT infrastructure. Last but not least, the KB had to build the trust of publishers to cooperate in experiments to archive their digital publications (Adams, 2004). A similar tripartite cooperation of libraries, IT companies, and publishers provided the basis for the success of the European project NEDLIB, which is discussed later. To acquire more experience in handling digital publications in practice, the KB started to experiment with "online" digital publications, using samples of e-journals kindly provided by Elsevier and Kluwer Academic. The KB experimented with a number of offline publications, trying to install them in order to check the content (Noordermeer, 1997).

Archiving Dutch e-Journals

The publishing company Elsevier has been involved in the archiving experiments of the KB since 1993. In 1995 the KB and Elsevier discussed the possibility of depositing the e-journals with a Dutch imprint (that is, those originally printed in the Netherlands) with the library. In 1996 a preliminary agreement was signed, and Elsevier started to deposit its e-journals. In 1999, when a similar arrangement was reached with the Dutch Publishers Association, the deposit of digital publications was extended to Dutch publishers in general. Both offline and online digital publications are deposited. Another archiving relationship developed by the KB dates from 2003 when the Dutch universities jointly started the Digital Academic Repositories (DARE) project[10] to create institutional repositories. The KB will archive and ensure the persistency of the digital information published through these institutional repositories.

Archiving International e-Journals

While developing the new digital archiving practices, the KB recognized that imprint (that is, place of publication) was no longer a feasible selection criterion for depositing digital publications by international publishers like Elsevier. Therefore, the KB investigated the adoption of a wider responsibility for archiving all the e-journals published by Elsevier irrespective of their imprint. The issue was discussed with the Ministry of Education, Science and Culture, to which the KB is accountable, and the ministry approved the KB's ambition to archive international digital publications. In 2002, after careful exploration of the mutual interests and capabilities of the partners, Elsevier and the KB signed a unique international agreement (Elsevier, 2002) for the archiving of all of Elsevier's e-journals. The main aims of the archiving agreement are the following:

- A formal archiving relationship to ensure permanent archiving of the publisher's publications
- To guarantee integrity of the digital information and to ensure permanent availability of the publications
- To act as host to provide access for (former) customers of the publisher and also to provide access as an emergency backup for the publisher

Following the archiving agreement with Elsevier, the KB signed agreements with other international publishers such as Kluwer Academic Publishers, BioMed Central, Blackwell Publishing, the Taylor and Francis Group, Oxford University Press, Springer, and Brill Academic Publishers. In these archiving agreements the digital publications archived by the KB may be used as follows:

- Bibliographic metadata about the publications may be included in the KB's online public catalogue and in the National Bibliography.

- Publications may only be used onsite at the KB and only by persons authorized by the KB.
- In the case of Open Access publishers and nonprofit publishers, the onsite restriction does not apply.
- For authorized KB staff, both onsite and remote use is allowed.
- The archived digital publications may be used as a source for print or fax copies for interlibrary loan within the Netherlands.
- Sending or transferring the electronic files outside the library by any means is not allowed.

Safe Place Strategy

   KB's Archiving Strategy Over the years the KB has developed an archiving strategy (Steenbakkers, 2002) that is referred to as the "Safe Place Strategy." This strategy consists of three steps. The first step is to create an archiving environment to which any publication that has to be archived will be transferred. The technical core of the archiving environment is a deposit system that has a similar function to the physical stacks of a library. The archiving environment offers specific and controlled conditions for storage, maintenance, and management of the publications.

   The second step is to organize and execute "perfect" copying of the digital object in order to refresh the storage medium. This has to be done before the old storage medium deteriorates or becomes technically obsolete. Because a deposit collection will continually grow, in the end copying will not be a trivial task due to the large size of the collection (Diessen & Rijnsoever, 2002).

   The third and most complicated step is to ensure that the publications can continue to be used in the future. To achieve this, we will have to register, preserve, or replace the functionality for rendering the digital information. These activities will require ongoing research and development as information technology is rapidly and continuously developing.

   IT Solutions To manage the online digital publications, the KB recognized that a specific computer system was needed and therefore looked for an IT partner who was willing to provide a pilot system. In 1995 the KB succeeded in teaming up with AT&T Solutions and its Bell Laboratories. AT&T provided the KB with a system called Right Pages, which is designed to handle a modest amount of e-journal articles. Besides creating the workflow for the online digital publications, the KB and AT&T jointly investigated the potential to scale-up the Right Pages system in order to manage larger amounts of digital publication.

   The KB and AT&T project made good progress, but in 1997 AT&T closed down the European division that was developing and marketing Right Pages and the KB was forced to look for another IT partner. After a search for alternative products, IBM's Digital Library was selected as a

replacement. The implementation of Digital Library at the KB required major effort from both IBM and the KB. This second pilot system became operational in January 1998, and it was substantially larger and contained basic functionality for handling and managing digital publications (Steenbakkers, 1999). The system, referred to as the Depot voor Nederlandse Electronische Publicaties, or DNEP-system, contained about 1.9 Tb of storage capacity. The DNEP-system was not only used to load and maintain digital publications but also to provide access to the digital content for library visitors and staff.

As described before, in 1994 the KB adapted its deposit policy to include digital publications and experimented with pilot deposit systems to store and handle them. In 1998 a comprehensive pilot deposit system was installed. In January of the same year the international project NEDLIB[11] started; it aimed to define the functional, technical, and organizational requirements for an operational electronic deposit system.

NEDLIB NEDLIB stands for the Networked European Depository Library project and was launched in January 1998 with the aim to define and test the architecture and procedures for capturing, preserving, and accessing digital publications. The NEDLIB project was initiated by CoBRA+, a cooperation of a number of national libraries operating under the umbrella of the Conference of European National Librarians (CENL). The project was cofunded by the Commission of the European Communities within the Telematics for Libraries Programme. The project operated over three years, 1998–2000, and had a total budget of 1,760,195 Euro. The participants consisted of eight national libraries, one national archive, two IT organizations, and three science publishers (Steenbakkers, 2000). The KB was in charge of the project coordination and provided the project management.

The objectives of NEDLIB were to develop a functional specification and an overall design for a depository for electronic publications and to address the issue of long-term preservation and permanent access. Also within this process a technical description and a prototype of a depository of digital publications had to be delivered (Werf-Davelaar, 1999). By the time NEDLIB had created the first draft of a generic model for a depository for electronic publications, the OAIS Reference Model White Book 1998 was published. After matching the NEDLIB model with the OAIS model, the NEDLIB partners decided to adopt the OAIS model and to contribute to its further development, as well as to apply it for use in national libraries and national archives.

The results of NEDLIB have been published in a series of six reports (NEDLIB, n.d.). These reports in turn provided valuable input for the development and implementation of the KB's operational electronic depository. Through the NEDLIB project the European national libraries were also able to contribute to the development of the OAIS archiving standard.

Acquiring the Deposit System Based on experience with the pilot system, the NEDLIB Guidelines (Steenbakkers, 2000), and the model for a Deposit System for Electronic Publications (Werf, 2000), the KB defined the requirements for its deposit system. In 1999 a scan of the IT market was done through a "Request for Information." The results from the request indicated that a deposit system could not be bought off the shelf. It also showed that the IT market seemed to be interested in designing and developing a deposit system. Encouraged by this, at the end of 1999 the KB decided to move ahead and publish the "Call for Tender *Depot van Nederlandse Publicaties*" (National Library of the Netherlands, 1999).

Requirements for the Deposit System The overall requirement for the deposit system was that it should offer a controlled archiving environment. It should support the maintenance of the digital publications in such a way that no data would be lost or mutilated and that access to the information was guaranteed, now as well as in the future. The detailed requirements were published in the tender document (National Library of the Netherlands, 1999). In summary, the requirements were as follows:

- The deposit system should be designed to handle a constantly increasing variety and amount of digital publication
- The system should be durable, in the sense that it can be technically updated continuously, without affecting the reliability of the archiving process and without endangering the archived content
- The functional design of the deposit system should be in accordance with OAIS (2002) Reference Model ISO14721:2002
- The deposit system should be a separate system interfacing with the digital library infrastructure that will offer traditional functions like Cataloguing, Search & Retrieval, etc.
- The interfaces to its environment should be well defined and easy to maintain
- As much as possible, the system should be constructed with proven technology and off-the-shelf building blocks with a large installed base

Tender Result The result of the European call for tender was a short list of five bidders, from which IBM was selected as the best technology partner and was contracted in October 2000. It took several months of intense but constructive negotiations before the contract could be signed.

The major problem to be solved during the contract negotiation was that the KB wanted an operational, comprehensive, OAIS-compliant deposit system, including full functionality for digital preservation (planning, management, and permanent access). At that time, however, the KB could not define the requirements for the preservation functionality precisely enough to demand its development and delivery. To solve this problem the

contract was divided into the development and delivery of an operational deposit system and an additional study to help define the requirements for long-term preservation and permanent access. This research effort was implemented as the "Long-Term Preservation Study," which ran parallel to the creation of the deposit system. IBM used the results of the study for the design and development of the Digital Information Archiving System (DIAS). The results were also published in December 2002 in a series of six reports.[12] It is interesting to note that in the contract negotiation the KB decided to grant IBM the full intellectual property of the archival system. The KB hoped that this would be an incentive for IBM to brand, market, and update the archival system as an IBM product.

IBM's Digital Information Archiving System (DIAS) Even though IBM developed DIAS in partnership with the KB, it is not a system specific to the KB nor even to libraries but rather is a solution for digital archiving in general.[13] DIAS is, as required, based on the Depository System for Electronic Publications (DSEP) model as published by NEDLIB, making it the first concrete application of the OAIS Reference Model (Werf, 2000). Compliance with the OAIS archival standard (see figure 3) recommends inclusion of the following functions:

- Delivery and Capture: services and functions to receive from the publishers the digital publications and check their quality and to produce SIPs (Submission Information Packages) and present them to Ingest
- Ingest: the services and functions to check and accept the SIPs
- Archival Storage: services and functions for the storage, maintenance, and retrieval of AIPs (Archival Information Packages)
- Data Management: services and functions for keeping, maintaining, and accessing descriptive information for the archived publications and other administrative data
- Administration: services and functions for controlling the daily operations
- Preservation: services and functions for planning, monitoring, and executing preservation strategies and actions
- Access: services and functions to locate and retrieve the archived information and produce the DIPs (Dissemination Information Packages)
- Packaging and Delivery: services and functions for preprocessing the information packaged in a DIP and delivering it to the user
- Monitoring and Logging: services and functions for registering of and reporting on actions

Further information and explanation of the OAIS model can be found in other publications such as Sawyer et al. (2002) and Cornell University Library (2003).
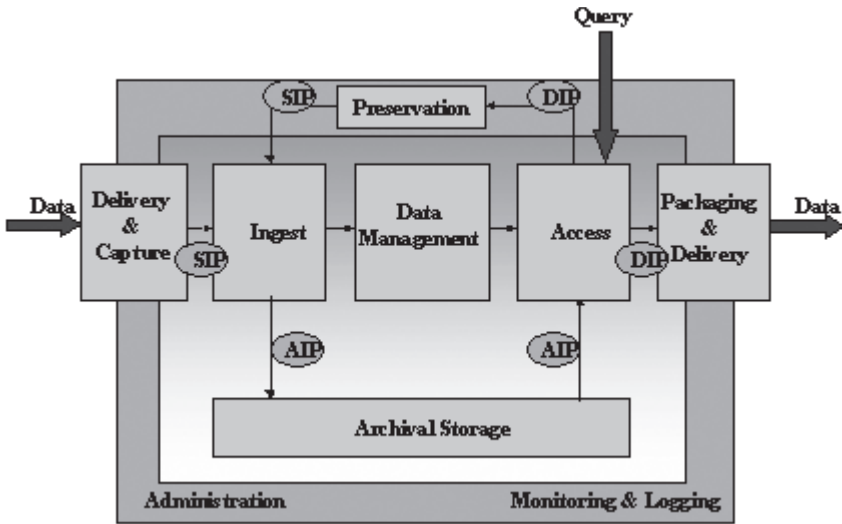
Figure 3.   Model of the Open Archival Information System (OAIS)

A simplified model of DIAS[14] is shown in figure 4. The DIAS Core contains the functions shown in the OAIS model above, except the Delivery and Capture and the Packaging and Delivery, which are shown in more detail on either side.
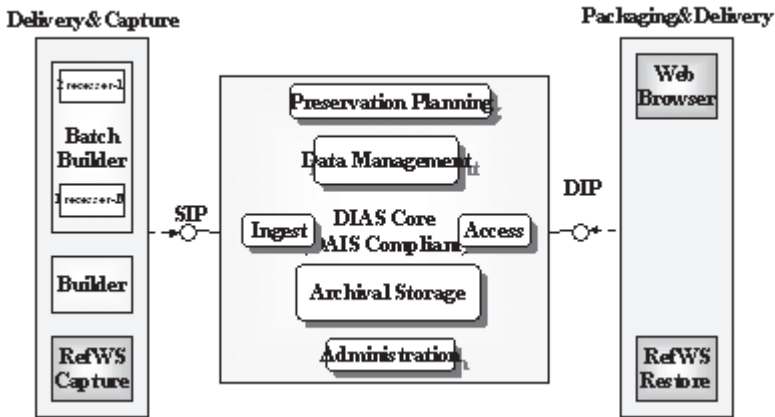

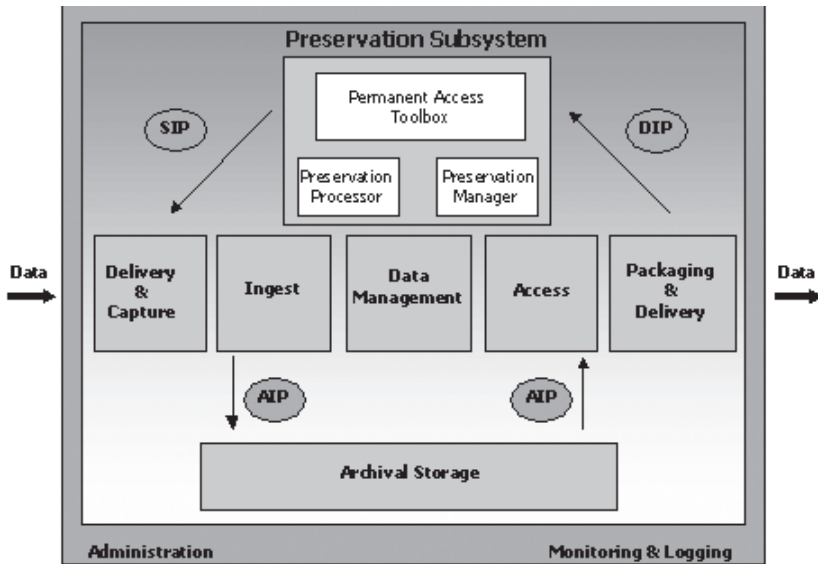*Figure 4.*   Model of IBM's Digital Information Archiving System (DIAS)

Figure 5.   The Preservation Subsystem and DIAS

IBM used off-the-shelf components as well as specifically developed components for constructing DIAS. Off-the-shelf components included Content Manager, Tivoli Storage Manager, DB2 database, and Business Objects. The Delivery and Capture functions (manual or automated capture) and Packaging and Delivery function (delivery through the network or on a specific workstation [such as RefWS][15]) were especially created for DIAS. The Preservation function is under construction.

Preservation Functionality
The Preservation Subsystem The current version of DIAS already contains basic preservation functionality. However, the KB and IBM are developing a Preservation Subsystem for DIAS that will support the registering of technical metadata and will provide other functionality needed for preservation. In figure 5 the model of DIAS is shown with the planned Preservation Subsystem in place. The subsystem consists of three components. In the lower right of the subsystem box is the Preservation Manager, a component that is used to register the technical metadata. At the top of the subsystem is the Permanent Access Toolbox or PATbox, as described earlier. At the bottom left is the Preservation Processor, used to execute the preservation actions.
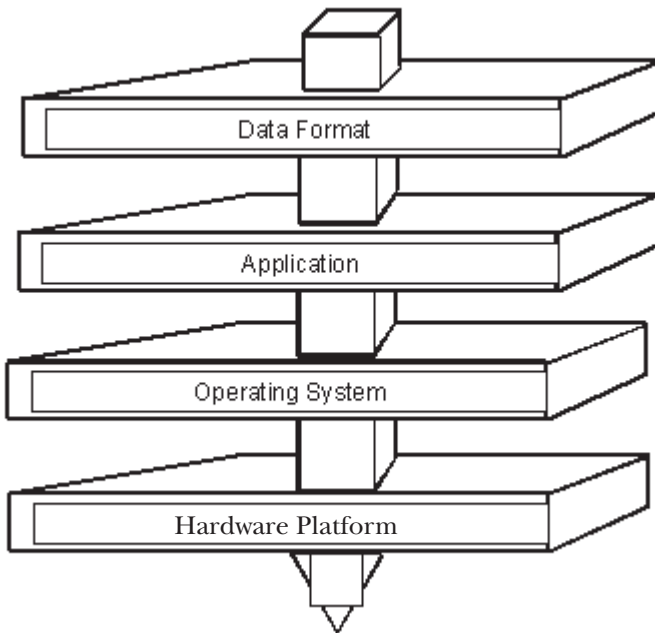
Figure 6.   The Preservation Layer Model

Technical Metadata An important part of the preservation metadata[16] is the metadata for rendering, or in OAIS terminology, "Representation Information." Metadata has to be registered for each archived digital publication. The OAIS concept of Representation Information was further developed by NEDLIB into a specific layered model (Lupovici & Masanès, 2000). In order to apply the layered model in practice, IBM developed it further into the Preservation Layer Model (PLM) (Diessen, 2002), which is shown in figure 6.

The PLM represents the template structures used by the Preservation Subsystem to register, maintain, and manage the technical metadata. A concrete PLM for a specific data format is called a "View Path." For example, a View Path for PDF 1.3 data format in practice may contain the following elements: reference platform = *Intel Pentium,* operating system = *NT,* viewer application = *Acrobat Reader 3.0.* The policy of the KB is to guarantee at least two View Paths for every digital publication archived.

The Preservation Manager The central component of the Long-Term Preservation function is the Preservation Manager, which supports the managing of the technical metadata. It registers the View Paths for every digital publication and monitors their viability. The Preservation Manager

has recently been developed by IBM and has been tested by the KB. It will be integrated into the next version of DIAS. (Diessen, Oltmans, & Wijngaarden, 2004). The purpose of the Preservation Manager in detail is to support registration and management of the metadata for rendering; monitor the availability of the technical prerequisites for accessing the information; and aid the planning of preservation actions, for example, migration or emulation.

The e-Depot: The Operational KB Archive

    The e-Depot's Infrastructure The "e-Depot" is the name of the KB's infrastructure and organization for archiving digital publications, including the workflow and procedures for handling and archiving digital publications. The name "e-Depot" also applies to the archiving service offered by the KB to producers and users of digital information. The e-Depot's infrastructure consists of DIAS with some smaller supporting systems and of other systems offering the usual digital library functions. In figure 7 the technical infrastructure of the e-Depot is schematically presented.

    DIAS is the technical core of the e-Depot, offering dedicated archiving functions. In the schema the functions on the left are for receiving and loading the digital publications: EPO is the Electronic Post Office, BER is the Basic Error Recovery, and NBN is the National Bibliographic Number generator for unique identifiers. The functions on the right are for cataloguing, search, retrieval, and delivery: GGC is the national central cataloguing system of Pica/OCLC, KB-TITEL is the local overall catalogue database of the KB, and IAA is the function for Identification, Authentication, and Authorization of end users.

    The e-Depot's Organization Within the KB three divisions are jointly responsible for running the service and developing the infrastructure of the e-Depot:

1. The Acquisition and Processing Division is in charge of the day-to-day operations of obtaining, checking, and loading the publications, including their metadata.
2. The Division for Information and Communication Technology is responsible for the technical maintenance of the infrastructure of the e-Depot. This task includes the maintenance of DIAS and expanding its storage capability, guaranteeing backup, and providing media migration. This division also manages integration of the deposit system within the general digital library infrastructure.
3. The Research and Development Division performs studies and experiments to further develop the functionality of the e-Depot. These activities are usually joint projects with the two divisions mentioned above. External technology partners are often involved. The Research and Development Division also organizes or participates in international activities, like development of standards, preservation studies, projects,
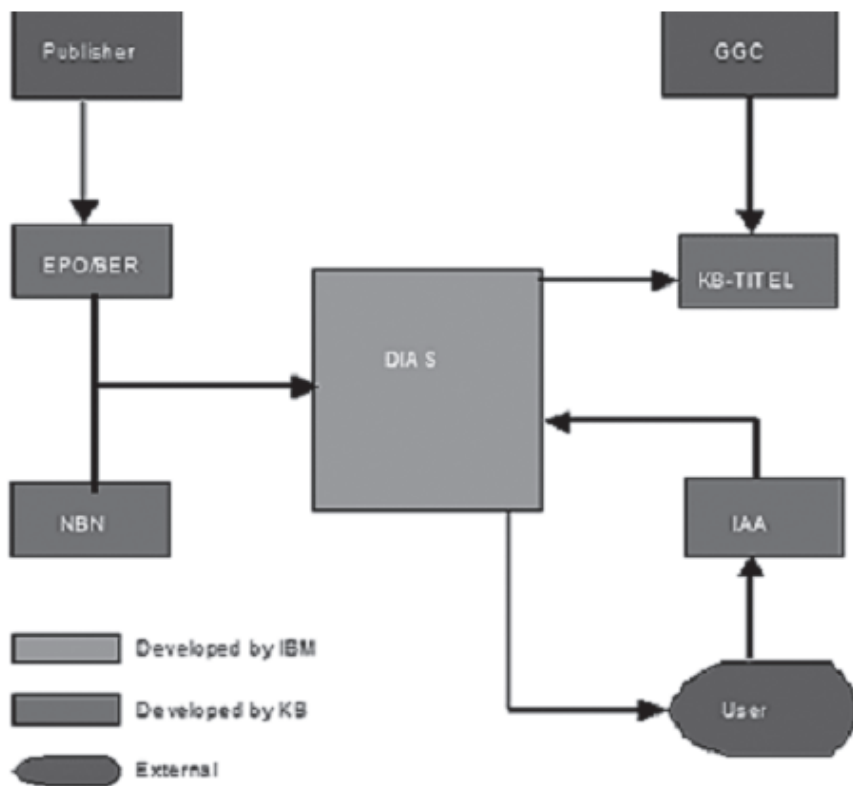
Figure 7.   Schema of the e-Depot Infrastructure

and conferences. For these activities a dedicated Digital Preservation research unit has been created.

The e-Depot is a strategic activity that has a great impact on the KB's policy and organization. To coordinate the activities and policy development concerning the e-Depot, the KB has implemented an e-Depot Steering Board. In addition to the three divisions already mentioned, the User Services Division also participates on the board. This division is in charge of providing access to the digital publications in accordance with conditions specified in the publishers' archiving agreements.

The e-Depot as a Trusted Repository The aim of the KB is to develop the e-Depot to become a "Trusted Digital Repository" (TDR). The concept of the Trusted Digital Repository was introduced in 2002 by the Research Libraries Group/Online Computer Library Center (RLG/OCLC) Working Group on Digital Archive Attributes (Research Libraries Group, 2002).
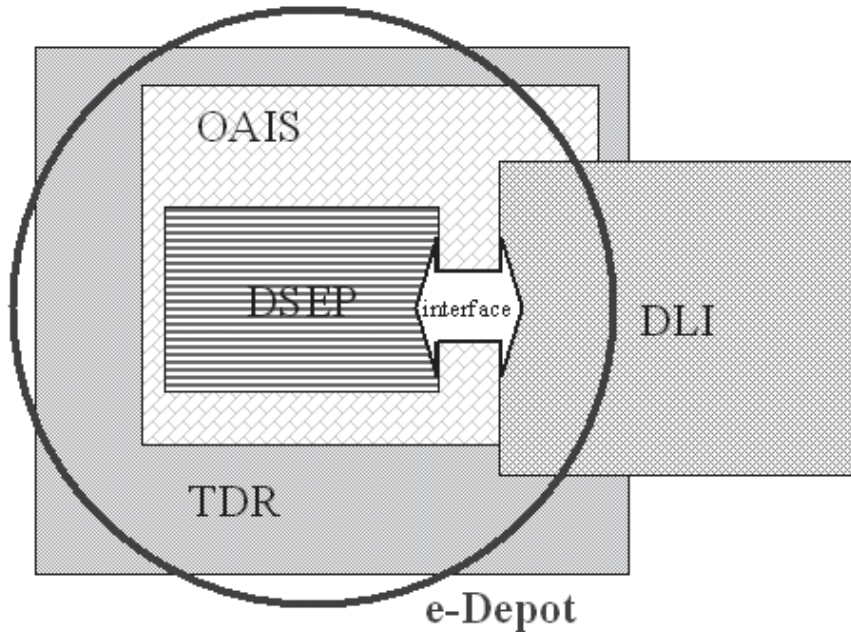
Figure 8.   The e-Depot Related to Archiving Concepts and Models

In its report the working group gives the following definition: "A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future" (Research Libraries Group, 2002, i).

Other archival concepts relevant to the e-Depot are the OAIS reference model and the NEDLIB model for DSEP. Figure 8 depicts the relationship of the KB's e-Depot to the following concepts and models:

- The e-Depot is the operational digital archive of the KB, dedicated to long-term archiving and permanent access. "e-Depot" refers to all aspects involved in operational digital archiving—policy, infrastructure, organization, and services—and encompasses a significant proportion of all the other concepts represented here.
- TDR is a concept describing the various requirements an institution has to meet in order to be considered a trusted repository that is able to archive digital information for the long term. In addition to the functions described in the OAIS model, the TDR concept includes functions like administrative responsibility, organizational viability, financial sustainability, system security, technical and procedural suitability, and procedural accountability (Research Libraries Group, 2002).

- OAIS is an International Organization for Standardization (ISO) standard describing the functional, organizational, and procedural requirements for a digital archive (Sawyer et al., 2002).
- DSEP is the NEDLIB process model that details the functional aspects of the OAIS in a deposit system model.
- DLI is not a specific concept for archiving. It represents the overall IT infrastructure offering the "traditional" digital library functions, like cataloguing processes, access to the catalogue, user registration and authentication, etc.

The Success of e-Depot

In October 2002 DIAS was handed over to the KB. Once DIAS was implemented and the workflow and procedures for handling the digital publications were in place, the loading of digital publications started. The DNEP pilot system was closed down at the end of 2004 as all its functions had been taken over by the e-Depot. The total budget available for the KB to create and obtain an operational deposit system; for professional project management support; for all hardware, software, and twelve Tb (terabytes) storage and seventeen Tb backup, was 5.5 million Euro. IBM developed the system at the KB's premises, for the costs agreed and according to the project planning.

CONCLUSION

In the past few years the stakeholders—either producers, brokers, and users of digital publications or IT producers and governmental bodies—became increasingly aware of the problem of preserving digital publications. More and more stakeholders have started to address this threat to the human digital record. Several publishers have developed an archiving policy and are acting accordingly. National and international governmental bodies are raising awareness of the problem of digital preservation and are creating regulations and legislation to support digital archiving. The scientific research community also has recently noticed the problem of digital preservation. The IT sector is increasingly becoming interested in studying and tackling the problem of digital preservation, as solving the digital preservation issues will have a long-term return on investment.

National libraries such as the KB in the Netherlands were amongst the first institutions not only to raise awareness of the problem of digital preservation but also to address it in practice. In this process the national libraries found the lack of funding the most serious obstacle, which has inhibited the progress in digital preservation research and development. Funding agencies, both in Europe and in the United States, usually do not provide funding for cultural sector venture research projects. They prefer to fund "safe" activities, like performing another study or survey or drafting another report or research agenda, rather than funding the research and

development itself. Nevertheless, national libraries have undeniably taken the lead and emerged as key players in the field of digital preservation. The International Publishers Association (IPA) and the International Federation of Library Associations and Institutions (IFLA) have acknowledged the position of the national libraries (Hodge & Frangakis, 2004, pp. 17–18).

In the past the lack of resources has frustrated the progress of research and development of solutions for a problem that will have a great impact on both our private lives and our business activities. It is about time that the national libraries and the national archives, together with partners in technology and publishing, receive substantial support to move the practice of digital preservation forward.

Acknowledgements

Notes

1. Information about SPARC is available at http://www.arl.org/sparc/; information about BioMed Central is available at http://www.biomedcentral.com/.
2. Examples of initiatives in digital archiving are PANDAS (http://www.nla.gov.au/padi/); JSTOR (http://www.jstor.org); YEA (http://www.library.yale.edu/~okerson/yea/frontmatter.pdf); NEDLIB (http://www.kb.nl/coop/nedlib/); PubMed Central (http://www.pubmedcentral .nih.gov/); DNEP and e-Depot (http://www.kb.nl/kb/resources/frameset_kenniscentrum .html); LOCKSS (http://lockss.stanford.edu/); D-Space (http://dspace.org/index.html); E-Journal Archiving Program (USA) (http://www.diglib.org/preserve/ejp.htm); DOM Programme (British Library) (http://vincent.bl.uk/cgi-bin/htm_hl?DB=website&STEM MER=en&WORDS=digit+object+manag+&COLOUR=Olive&STYLE=s&URL=http://www .bl.uk/about/policies/etmeetjul2003.html#muscat_highlighter_first_match); KOPAL (Die Deutsche Bibliothek) (http://www.langzeitarchivierung.de/); E-Archive (http://www .ithaka.org/e-archive/index.htm); and New Zealand (http://www.natlib.govt.nz/bin/ news/pr?item=1085888952).
3. National programs to promote digital archiving include NDIIP, of the United States (http:// www.digitalpreservation.gov/); DPC, of the United Kingdom (http://www.dpconline .org/graphics/index.html); and NESTOR, of Germany (http://www.langzeitarchivierung .de/).
4. A number of publishers have published their archiving policies online, including Elsevier (http://www.elsevier.com/wps/find/librariansinfo.librarians/libr_policies#sdarchiving); Kluwer Academic Publishers (http://www.wkap.nl/prod/a/newsaboutkluwer#190503); BioMed Central Journal (http://www.biomedcentral.com/info/libraries/archive); Blackwell Publishing (http://www.blackwellpublishing.com/press/pressitem.asp?ref=83&site=1); and Taylor & Francis Group (http://www.taylorandfrancisgroup.com); and Oxford University Press (http://www.oup.co.uk/).
5. For information about the Trivulzio manuscript, see http://www.kb.nl/webexpo/trivulzio-en.html.
6. The UVC can be downloaded via Alphaworks at http://www.alphaworks.ibm.com/tech/uvc.
7. The PATCH Consortium is comprised of Koninklijke Bibliotheek (The Netherlands), consortium leader; British Library (United Kingdom); UK National Archives (United Kingdom); Centre National d'Etudes Spatiales (France); Bibliothèque National de France

(France); University of Leeds (United Kingdom); Digital Preservation Coalition (United Kingdom); IBM-Nederland (The Netherlands); Die Deutsche Bibliothek (Germany); University of Bath, UKOLN (United Kingdom); Bibliotheek Technische Universiteit Delft (The Netherlands); Nationaal Archief (The Netherlands); Det Kongelige Bibliotek (Denmark); Statsbibliotheket (Denmark); Kungliga Biblioteket (Sweden); Schweizerische Landesbibliothek (Switzerland); and Statens Arkiver (Denmark).

8. See the important work of the project PRONOM on File Format Registry at http://www .nationalarchives.gov.uk/PRONOM/default.htm.
9. See http://www.digicult.info/pages/drr_themes.php.
10. Information about DARE can be found at http://www.darenet.nl/en/.
11. Information about NEDLIB can be found at http://www.kb.nl/coop/nedlib/.
12. See http://www.kb.nl/hrd/dd/dd_onderzoek/dnep_ltp_study-en.html for more information.
13. The national library of Germany, Die Deutsche Bibliothek, is actually implementing DIAS for their digital archive; see http://www-306.ibm.com/software/success/cssdb.nsf/CS/ MCAG-65VEUX?OpenDocument&Site=default.
14. See http://www-5.ibm.com/nl/dias/ for more information about DIAS.
15. The Reference Workstation (RefWS) or Reference Platform is a standard computer system configuration designated by the deposit library for installing and running electronic publications. The RefWS is suited for most publications appearing on the consumer market during a given period of time (Werf, 2000, p. 26). The KB considers the use of RefWS a temporary solution for guaranteeing access to archived CD-ROM publications.
16. For a comprehensive framework for Preservation Metadata (or PREMIS), see http://www .oclc.org/research/projects/pmwg/.

References

Adams, G. (2004). Partners go Dutch to preserve the minutes of science. *Research Information, 13*, 18–22. Retrieved February 15, 2005, from http://www.researchinformation.info/ risepoct04archiving.html.

Cornell University Library. (2003). *Digital preservation management: Implementing short-term strategies for long-term problems* [Online tutorial]. Retrieved February 15, 2005, from http://www .library.cornell.edu/iris/tutorial/dpm/index.html.

Council of the European Union. (2002). Council resolution: On preserving tomorrow's memory—Preserving digital content for future generations. *Official Journal of the European Communities, 162,* 1–2. Retrieved August 15, 2005, from http://europa.eu.int/eur-lex/ pri/en/oj/dat/2002/c_162/c_16220020706en00040005.pdf.

Diessen, R. J. van. (2002). *Preservation requirements in a deposit system* (IBM/KB Long-Term Preservation Study Report Series 3). Retrieved February 15, 2005, from http://www.kb.nl/ hrd/dd/dd_onderzoek/reports/3-preservation.pdf.

Diessen, R. J. van, Oltmans, E., & Wijngaarden, H. van. (2004). Preservation functionality in a digital archive. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries, Tucson, Arizona* (pp. 279–86). New York: ACM Press. Retrieved February 15, 2005, from http://portal.acm.org/citation.cfm?id=996350.996416.

Diessen, R. J. van, & Rijnsoever, B. J. van. (2002). *Managing media migration in a deposit system* (IBM/KB Long-Term Preservation Report Series 5). Retrieved February 15, 2005, from http://www.kb.nl/hrd/dd/dd_onderzoek/reports/5-mediamigration.pdf.

Elsevier. (n.d.). *E-Journals at Elsevier: Over two decades of experimentation and development.* Retrieved February 15, 2005, from http://www.elsevier.com/wps/find/authored_newsitem .cws_home/companynews05_00021.

———. (2002). *Elsevier and Koninklijke Bibliotheek finalise major archiving agreement.* Retrieved February 15, 2005, from http://www.elsevier.com/wps/find/authored_newsitem.librarians/companynews05_00020.

———. (2004). *Elsevier's comments on evolutions in scientific, technical and medical publishing and reflections on possible implications of Open Access journals for the UK.* Retrieved February 15, 2005, from http://www.elsevier.com/authored_news/corporate/images/UKST1Elsevier_ position_paper_on_stm_in_UK.pdf.

Gorman, M. (2001). Bibliographic control or chaos: An agenda for national bibliographic

services in the 21st century. In *Libraries and librarians: Making a difference in the knowledge age: 67th IFLA Council and General Conference, August 16–25, 2001, Boston, USA.* Retrieved February 15, 2005, from http://www.ifla.org/IV/ifla67/papers/134–133e.pdf.

Hodge, G., & Frangakis, E. (2004). *Digital preservation and permanent access to scientific information: The state of the practice.* A report sponsored by the International Council for Scientific and Technical Information (ICSTI) and U.S. Federal Information Managers Group (CENDI). Retrieved February 15, 2005, from http://cendi.dtic.mil/pubarc.html.

Hunter, K. (2002). STM members and digital archiving. *Information Services & Use, 22*(2/3), 83–88.

ICABS. (2004). *IFLA Core Activity: IFLA-CDNL Alliance for Bibliographic Standards (ICABS).* Retrieved February 15, 2005, from http://www.ifla.org/VI/7/icabs.htm.

Information Society Technologies in the 6th Framework Programme. (2002–06.) *Thematic priority under the specific programme "Integrating and strengthening the European research area."* Retrieved August 15, 2005, from http://ica.cordis.lu/search/index.cfm?fuseaction=prog .simpledocument&PG_RCN=5465040&CFID=3981689&CFTOKEN=15082814.

Lorie, R. (2002). *The UVC: A method for preserving digital documents: Proof of concept* (IBM/KB Long-Term Preservation Study Report Series 4). Retrieved February 15, 2005, from http:// www.kb.nl/hrd/dd/dd_onderzoek/reports/4-uvc.pdf.

———. (2004). Preserving digital documents for the long-term (pp. 88–92). *Proceedings of the Imaging Science & Technology 2004 Archiving Conference, San Antonio, Texas.* Springfield, VA: Society for Imaging Science and Technology.

Lupovici, C., & Masanès, J. (2000). *Metadata for long term preservation of electronic publications* (NEDLIB Report Series 2). Retrieved February 15, 2005, from http://www.kb.nl/coop/ nedlib/results/NEDLIBmetadata.pdf.

National Library of the Netherlands. (1999). *Call for tender Depot van Nederlandse Elektronische Publicaties* [published in accordance with the European procedure under number 99/ S 177-126796/NL]. Retrieved August 15, 2005, from http://www.kb.nl/dnp/e-depot/ dm/callfortender.pdf.

NEDLIB. (n.d.). NEDLIB Web site, Results section. Retrieved February 15, 2005, from http:// www.kb.nl/coop/nedlib/.

Noordermeer, T. (1997). Deposit for Dutch electronic publications: Research practice in the Netherlands. In C. Peters & C. Thanos (Eds.), *Research and advanced technology for digital libraries: 1st European Conference on Digital Libraries, ECDL 1997, Pisa, Italy, September 1–3, 1997* (Lecture Notes in Computer Science, 1324, pp. 361–73). London: Springer-Verlag.

Poynder, R. (2003). Elephants and dung-trucks. *Information Today, 20*(8). Retrieved February 15, 2005, from http://dspace.dial.pipex.com/town/parade/df04/elephants_and_dung .htm.

Reference Model for an Open Archival Information System (OAIS). (1998). Consultative Committee for Space Data Systems, Recommendation concerning space data systems standards, CCSDS 650.0-W-3.0, White Book. Retrieved August 15, 2005, from ftp://nssdcftp .gsfc.nasa.gov/standards/nost/isoas/us12/CCSDS-650.0-W-3.pdf.

———. (2002). Consultative Committee for Space Data Systems, Recommendation for space data system standards, CCSDS 650.0-B-1, Blue Book. Retrieved August 15, 2005, from http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCDS-650.0-B-1.pdf.

Research Libraries Group. (2002). *Trusted digital repositories: Attributes and responsibilities. An RLG-OCLC Report.* Retrieved February 15, 2005, from http://www.rlg.org/longterm/ repositories.pdf.

Rothenberg, J. (1995). Ensuring the longevity of digital documents. *Scientific American, 272,* 24–29.

Sawyer, D., Reich, L., Giaretta, D., Mazal, P., Huc, C., Nonon-Lapatie, M., et al. (2002). *The Open Archival System (OAIS) reference model and its usage.* Retrieved February 15, 2005, from http://www.ccsds.org/documents/SO2002/SPACEOPS02_P_T5_39.PDF.

Smith, B. (2003). Preserving tomorrow's memory: Preserving digital content for future generations. *International Preservation News, 29,* 4–10. Retrieved August 15, 2005, from http://www .ifla.org/VI/4/news/ipnn29.pdf.

Steenbakkers, J. (1999). Developing the depository of Netherlands electronic publications. *Alexandria, 11*(2), 93–99.

———. (2000). *The NEDLIB guidelines: Setting up a deposit system for electronic publications* (NEDLIB

Report Series 5). Retrieved February 15, 2005, from http://www.kb.nl/coop/nedlib/results/NEDLIBguidelines.pdf.

———. (2002). Preserving electronic publications. *Information Services & Use, 22,* 89–96.

———. (2004). Digital archiving: A necessary evil or new opportunity? *Serials Review, 30,* 29–32.

UNESCO. (2003). *Guidelines for the Preservation of the Digital Heritage.* Retrieved February 15, 2005, from http://unesdoc.unesco.org/images/0013/001300/130071e.pdf.

Werf, T. van der (2000). *A process model: The deposit system for electronic publications* (NEDLIB Report Series 6). Retrieved February 15, 2005, from http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf/.

Werf-Davelaar, T. van der. (1999). Long-term preservation of electronic publications: The NEDLIB project. *D-Lib Magazine, 5*(9). Retrieved February 15, 2005, from http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html.

Wijngaarden, H. van, & Oltmans, E. (2004). Digital preservation and permanent access: The UVC for images (pp. 254–58). *Proceedings of the Imaging Science & Technology 2004 Archiving Conference, San Antonio, Texas.* Springfield, VA: Society for Imaging Science and Technology.

Johan F. Steenbakkers, Director of e-Strategy and Property Management, Koninklijke Bibliotheek, National Library of the Netherlands, Prins Willem-Alexanderhof 5, P.O. Box 90407, NL-2509 LK, The Hague, The Netherlands, johan.steenbakkers@kb.nl. Johan F. Steenbakkers is Director of e-Strategy and Property Management at the Koninklijke Bibliotheek (KB), the national library of the Netherlands. In 1969 Johan graduated cum laude in Biology and Biochemistry at the University of Utrecht, the Netherlands. Until 1973 he performed scientific research on bio-membranes. He then started his career in the field of libraries and documentation. Johan joined the KB in 1987 and was involved in restructuring the library's organization. He also renovated the IT infrastructure of the KB and initiated several strategic digital library projects. He was the project coordinator of the European project NEDLIB (1998–2000), which focused on the functional and technical aspects of handling and preservation of electronic publications. Johan initiated the e-Depot of the KB, the infrastructure and organization for preserving digital information. He organized and managed the joint KB/IBM project (2000–2002) for the development of the deposit system DIAS, the first OAIS compliant archival system. He has published several articles about the development of the e-Depot and digital preservation in general. He is a member of the Task Force on Digital Repository Certification, initiated by RLG and the National Archives and Records Administration (NARA) in the U.S.