Permanent Archiving of Electronic Publications: Research & Practice[1]

Johan F. Steenbakkers
Director Information Technology and Facility Management, Koninklijke Bibliotheek, National Library of the Netherlands, The Netherlands

ABSTRACT

Results of scientific research are nowadays as a rule published digitally. This is certainly the case in the fields of Science, Technology and Medicine, producing over 80 % of all scientific publications. Digital publishing is causing a real paradigm shift for research institutions and publishers as well as for libraries. As a consequence, these institutions have to develop new policies, new business models and new infrastructure and techniques.

In this paper I will focus on the aspects of infrastructure and of techniques for permanent archiving of electronic publications. The problem is that, at the same rate at which our world is becoming digital, our information is threatened. New types of hardware, computer applications and file formats supersede each other, making our recorded digital information inaccessible. In the past, libraries and archives have undertaken several actions and carried out related studies and research on digital preservation issues. For instance the Koninklijke Bibliotheek (KB) - national library of the Netherlands - has jointly with IBM-Netherlands developed the first standard deposit system (Digital Information Archiving System or 'DIAS'). Using DIAS the KB realised in 2002 an electronic deposit ('e-Depot') and signed archiving agreements with major science publishers for permanent keeping of an important part of the digital 'Record of Science'.

In this paper I will describe the strategy followed and the steps taken to realise the 'e-Depot'. In addition, the new initiative 'PATCH' to integrate and promote further research and standards-development for permanent access will be discussed.

1 THE IMPACT OF ELECTRONIC PUBLISHING

1.2 Libraries - from Collections to Licences

Results of scientific research are nowadays as a rule published digitally. This is certainly the case for Science, Technology and Medicine, the research fields that produce over 80 % of all scientific publications. For the time being printed derivatives of the digital publications are also produced often, not least for archival purposes. The shift from print to digital has induced

[1] Part of the information presented in this document was published earlier by the author or was delivered in a project proposal to the EC by the PATCH Consortium in April 2003.

a shifting of the responsibilities of academic libraries from keepers of collections to licensees of information.

Traditionally, academic libraries have contributed to the keeping of the output of science through time. For deposit libraries - mostly the national libraries - the maintenance of the Record of Science is even a key task. The maintenance of collections by libraries is under pressure as more publications turn digital. One reason is the change of policy of publishers who prefer no longer to sell but rather to license their publications, offering them through sophisticated search & retrieval services. Another reason is that handling and maintaining digital publications requires new skills and a quite different infrastructure than for printed publications. And complicating matters even further is the fact that great research and development efforts are required, as no best practices on the issues are as yet available.

1.2 Publishing - from Physical to Digital

In the preface of the proceedings[2] of the 'ICSTI/CODATA/ICSU Seminar on Preserving the Record of Science', organised on the UNESCO premises at Paris in February 2002, the editors Barry Mahon and Elliot Siegel state the change as follows. "The transfer to digital production is almost complete, it is inconceivable that any scientist would not present the results of his or her work in anything but digital form, even if it is in parallel to a physical copy delivered to a publisher." And they continue. "The changes have given rise to much debate in scientific community – but this debate is not the subject of this compilation, another consequence of the changes is. Preservation or archiving or, to use a more modern term, 'permanent availability' of the record of science as represented by scientific and technical information is one of the processes which is dramatically affected by the change to an all digital world. While once the printed page was the basic medium and was archived by libraries, today the record exists in a number of locations and in many cases is not printed."

1.3 Publishers - the Archive is an Asset

In her presentation at the seminar mentioned above, Karen Hunter from Elsevier Science represented the viewpoints and developments with respect to digital archiving among STM members. She started by explaining that "The International Association of Scientific, Technical and Medical Publishers ("STM") has a wide range of member publishers. It is a mix of commercial and society publishers, small and large organizations. What is clear, however, is that STM *journal* publishers take archiving very seriously. It is sometimes said that publishers only care about the exploitable value of the archive. This ignores the very real concern STM publishers have for their authors. We know we have a responsibility to authors to ensure that their articles will be available to future researchers. As we move toward an e-only publishing future, that means we must provide for digital archiving of our journals." She continues by saying "Publishers are certainly also interested in the archive as an asset. There is value in the collection, although current debates within the scientific community

[2] Published in 'Information Services & Use', 22 (2,3) 49-144, 2002, Special issue, IOS Press, Amsterdam. The Seminar was organised by ICST (the International Council for Scientific and Technical Information) in association with ICSU (the International Council for Science) and CODATA (the ICSU Committee on Data for Science and Technology).

demonstrate the range of opinion about how long that asset should be controlled exclusively by the publisher."

1.4 Digital Archiving - a Challenge

Nowadays we depend heavily on digital information in a wide variety of forms including text, images, sound, video, as well as combinations of these different media types. At the same rate at which our world is becoming digital, our information is threatened. New types of hardware, computer applications and file formats supersede each other, making our recorded digital information inaccessible. Left in an attic, digital information is bound to become inaccessible, in contrast to information in a (printed) book. Even if the hardware or the carrier-media does not deteriorate within the time frame considered, the technology to access the information will inevitably become obsolete. ICT is developing constantly and rapidly offering us new and appealing applications while rapidly making existing hardware and software obsolete.

In recent years, the digital preservation challenge has been recognised by people outside the traditional memory institutions (libraries, museums, etc). The problem of digital preservation is broad and society as a whole is having to deal with it. This has resulted in the issue of digital preservation being widely discussed today. The problem of digital preservation is an international issue and concerns several sectors of society: governments and cultural heritage institutions, research organisations, data-intensive commercial companies and ICT-vendors. The problem of digital preservation has also gained political support and actions have been undertaken to raise broad awareness, e.g. by the EC[3] and UNESCO[4]. The magnitude of the problem and the diversity of the stakeholders require co-operation (international and cross-sectoral) to develop solutions. But while awareness is growing, solutions will not arise if left to the free market alone. Long-term commitment to action is required. This is only partially compatible with the short-term market orientation of commercial companies. For a lasting and trustworthy solution we need to develop a new approach of embedding the concept of persistence of digital information in ICT applications and procedures. Commercial ICT-vendors and developers are not likely to address this problem until there is a substantial demand in the market. The actual demand comes from some memory institutions that have taken up the challenge of addressing the problem of digital obsolescence.

1.5 Strategy - towards Persistent Digital Information

Memory institutions, like national libraries and archives, contend with a growing quantity of digital documents. For example, most scientific literature is already being published digitally. Ignoring the problem of persistence in digital information puts the 'Record of Science' at risk. Libraries and archives were the first institutions confronted with the problem of digital preservation.

[3] http://digicult.salzburgresearch.at/ Digicult Report. Technological Landscapes for Tomorrow's Cultural Economy, EC, 2002. http://europa.eu.int/eur-lex/en/archive/2002/c_16220020706en.html Preserving Tomorrow's Memory. EC, 2002.
[4] http://www.anu.edu.au/caul/preservation/DigHeritageCharter-prelim-draft.doc UNESCO's Preliminary Draft Charter on the Preservation of the Digital Heritage, 2003.

In essence digital information consists of two things: a formatted bit stream and the functionality needed to decode this logical format and render the information to the user. Even for apparently complicated digital items these are the really vital aspects. Therefore the key actions for digital preservation are:

- preserving the (formatted) bit stream, also called the 'digital object', and
- ensuring accessibility over time to the information embedded in the digital object.

In the past, libraries and archives have undertaken several actions and carried out related studies and research on digital preservation issues. From these various activities an overall approach on the way towards persistent digital information can be distilled. This approach boils down to two adjoining research and development fields of digital preservation. These fields are:

- the maintenance through time of large amounts of 'digital objects', containing the information, and
- the guarantee of 'permanent access' to the digital information in the objects, irrespective of the obsolescence of file formats, software, and hardware.

These two fields reflect the essence of the problem of digital preservation.

For the problem of preserving digital objects, the European project NEDLIB[5] has amalgamated a wide range of internationally acquired research results in the 'Guidelines for setting up a Deposit System for Electronic Publications'. A strategic choice by NEDLIB was to transfer the electronic publications from the publishing environment to an archiving environment, nowadays often referred to as a 'Safe Place'. Note that the word 'place' in 'Safe Place' should not be taken literally, but rather as a concept, indicating an institution that is committed to digital preservation and possesses appropriate infrastructure, resources and skills for the task.

Implementing an archiving environment is the first step towards digital preservation. Guaranteeing rendering[6] of that information is the second - even more complicated - part of long-term preservation. In addition to the safe keeping of the digital objects, access to the information in the objects has to be permanently provided. Tools - techniques and procedures - are needed to provide access to the stored objects now as well as in the future. Research on tools and procedures for permanent access has started, but is still in its infancy.

2 REALISING A 'SAFE PLACE'

2.1 Acquiring the Infrastructure and Skills

In 1994 the KB decided to include electronic publications in its deposit collection. To realise this, specific infrastructure, organisation and skills had to be developed, not only for handling electronic publications but also for guaranteeing long-term availability. From 1995 onwards the KB experimented with electronic publications in co-operation with some publishers. A preliminary agreement was reached in 1996 with 'Elsevier Science' to deposit 270 electronic journals with a Dutch imprint. This was later expanded to 351 titles. A similar provisional arrangement was made with 'Kluwer Academic Publishers' and 'SDU Uitgevers'. To handle

[5] Information on NEDLIB see www.kb.nl/coop/nedlib/
[6] Rendering means retrieval and viewing of digital objects.

the electronic publications, the KB, together with AT&T (involving Bell Laboratories), implemented a small content management system. This system was replaced in 1998 by a more extensive pilot deposit system.

In 1998 the KB implemented a pilot deposit system based on IBM Digital Library and Tivoli Storage Management software. This pilot system caters to born-digital publications, amongst which are the deposit collection, and scanned images produced by the library's digitisation programme. The pilot system has been maintained and expanded until today (2 TB of storage) but will be replaced during 2003 by the new operational deposit system, acquired recently by the KB.

2.2 Deposit Arrangement with Publishers

In 1999 the Dutch Publisher's Association signed an arrangement to deposit all electronic publications at the KB. The arrangement covers offline and online publications and prescribes restricted access conditions. Readers may only view the publications on-site. The intention is in the near future to develop new business models between national libraries and publishers for access to the digital information.

2.3 NEDLIB - Defining the Requirements

To initiate practical work on long-term preservation, a group of 8 European national libraries, in co-operation with 3 publishers (Elsevier Science, Kluwer Academic Publishers and Springer), started the project 'Networked European Deposit Library' or 'NEDLIB' in 1998. The project was co-funded by the EC and the KB led the project. The aim of NEDLIB was to define the basic functional and technological conditions for a networked European deposit library. The objectives of NEDLIB concur with the mission of national deposit libraries to ensure that electronic publications of the present can be used now and in the future. The NEDLIB project was successfully concluded at the end of 2000.

2.4 DNEP - Developing the Deposit System

In 1999 the KB started the project 'Deposit for Netherlands Electronic Publications' or 'DNEP' to acquire a full-scale deposit system. To prepare for the selection procedure, the library specified the system requirements, which were based on the NEDLIB results and its application of the ISO standard for digital archives, OAIS (Open Archival Information System). The deposit system had to offer large-scale, high quality storage and digital preservation functionality. Also its design should be future-proof and provide for scalability, extensibility and flexibility. In 2000, as a result of a European tender procedure, the KB contracted the development of the deposit system to IBM-Netherlands - Global Services. The development of the system started in October 2000 and was planned to take two years.

2.5 The Deposit System Delivered

IBM developed the deposit system on site on the KB premises. In October 2002 the deposit system was delivered to the KB. Meanwhile the library had created a workflow for electronic publications and had designed interfaces to the catalogue and other digital library functions. The KB deposit system initially has a storage capacity of 12 TeraBytes and is scalable over 500 TeraBytes. The system was constructed using as much as possible off-the-shelf components, like DB2, TSM, WebSphere and Content Manager. As the system is a generic archival system, IBM has branded[7] it with the name 'Digital Information Archiving System' or 'DIAS' and will maintain it as a product.

Parallel to the development of the system KB and IBM have jointly studied and tested Long-Term Preservation issues. The research results were taken into account when designing the deposit system. They were also published as a series of six reports[8]. Also after the delivery of the system, research on digital preservation and long term access will be needed. The deposit system will also be developed further, technically and functionally, including realising specific preservation technologies.

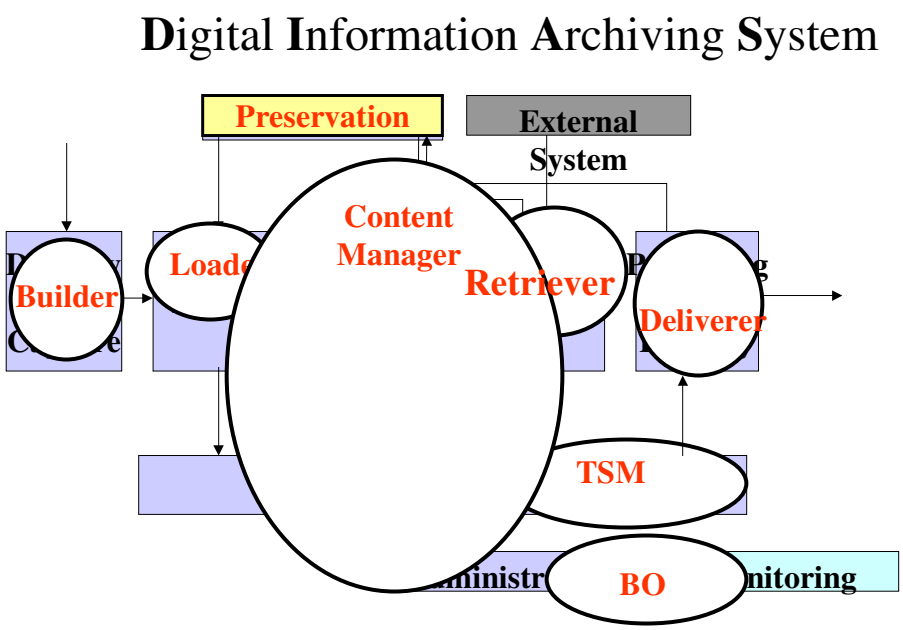

Figure 1: Digital Information Archiving System (DIAS)

2.6 The 'e-Depot' Realised

The deposit system 'DIAS' is the technical heart of the electronic deposit or 'e-Depot' of the KB. The contribution of several members of the KB staff, defining the requirements and testing the system during the development phase, has been a critical success factor. But after the system was delivered the real job started. The processing department had to implement the

[7] For information about DIAS see www.ibm.com/nl/dias/
[8] IBM/KB reports see www.kb.nl/kb/ict/dea/ltp/reports/reports.html

new workflow and acquire the skills needed for processing the electronic publications. The ICT department had to run a system larger and more complicated than they had ever done. The challenge offered by the electronic publications has a major impact on the KB organisation and on the tasks performed by its staff. But next to being a challenge, this development also offers an exciting opportunity to position the KB in the digital area.

2.7 Archival Agreements with Publishers

In the past months both Elsevier Science and Kluwer Academic Publishers have signed a unique agreement with the KB on long-term digital archiving of all their electronic publications[9]. At this moment the digital publications of Elsevier and Kluwer are being loaded in the 'e-Depot', involving more than 2,200 journals, containing over 5 million articles. In a press release of August 20th 2002 Elsevier states that by signing this agreement it wants to provide "… assurance for Elsevier Science titles, which constitute an essential part of the core of scientific literature." And that "The KB was the natural partner, as it is clear leader world-wide in the experimentation with and the investment in digital preservation."
In a press release of May 19th 2003 Kluwer states: "… the agreement between the KB and Kluwer is yet another step forward in keeping digital archives available in perpetuity." Other publishers have meanwhile contacted the KB about similar archiving arrangements.

3 RESEARCH AND STANDARDS FOR PERMANENT ACCESS

3.1 Digital Preservation: a Sustainable Approach is Needed

The early ICT users were large companies and government. These organisations are still 'keeping alive' old computer systems, sometimes for several decades at great effort and expense, in order not to lose crucial business information. However, given the widespread use of digital information in society and the increased rate in hardware and software changes, it is clear that this is not a sustainable approach for digital continuity in general. Many ICT users have already experienced the loss of valuable digital information because formats of documents have become obsolete or because the required hardware was no longer available.

Backward compatibility in some cases offers a temporary delay of (a part of) the loss. For a lasting and trustworthy solution however, we need a new approach that embeds the concept of persistence of digital information into ICT applications and procedures. However, commercial ICT-vendors and developers will not easily start addressing this problem because of the long return-on-investment-time involved. Hence, to achieve a breakthrough it is essential that national and international governmental bodies take responsibility and actively support research and standardisation of digital preservation. Substantial funding is needed to achieve further progress.

[9] For more information see www.kb.nl

3.2 Guaranteeing Permanent Access

ICT companies have only recently become aware of the problem of relatively short-term accessibility of digital objects. The standardised archival system developed by KB and IBM, is designed to preserve and control digital information for the long term. Still to be resolved however is how to guarantee permanent access to the stored information. How can we render digital information for users in the future? Up till now, the problem of permanent access has been addressed by several, scattered and small-scale, initiatives[10]. To accelerate this development a group of national libraries, archives, universities, research institutions and ICT companies have joined forces to create tools for permanent access. These organisations already have some practical experience in archiving digital information and have also started small-scale research in several digital preservation areas, including development of rendering solutions. To avoid duplicate efforts, to benefit from the experiences, but also to realise a generic approach for permanent access, these organisations propose the integrated research and development project 'Permanent Access Toolbox for the digital Cultural Heritage' or 'PATCH'[11].

3.3 An Integrated Approach

The aim of PATCH is to create a technological framework (the 'PATbox') that will promote continuous development of solutions for permanent access. In parallel with this development, a range of tools for permanent access will be built and tested in a variety of environments (archives, libraries, and scientific institutions). During the project the tools will be compared, general guidelines will be drawn up and interoperability will be reviewed. PATCH will also try to induce a change in the way in which the ICT sector designs systems and software. This change requires that persistence of digital information is a design criterion for systems, rather than an inevitable problem for which at a late stage, makeshift measures have to be provided.

The total costs for PATCH are estimated at 12,700 K€ over a period of 3 to 4 years. This sum exceeds the budgets of the institutions involved. The Patch Consortium is actually looking for co-funding by the EC of at least 7,500 K€. (This is a modest investment if compared to the co-funding granted by the US Congress for the integrated plan for digital preservation in the USA[12].) The status at the moment of writing this report is that, although PATCH has been qualified in the evaluation process as a strategic and very good proposal, it is still uncertain if the European Commission will co-fund this initiative.

[10] For instance, parallel to the development of the deposit system, KB and IBM jointly performed some research on long-term preservation issues. The results look promising and offer a good starting point for further research on and development of permanent access solutions. See www.kb.nl/kb/ict/dea/ltp/reports/reports.html

[11] The project PATCH 'Creating the Permanent Access Toolbox for digital Cultural Heritage' was proposed as an 'Integrated Project' by the 'PATCH Consortium' in April 2003 to the EC for co-funding within the 6th Framework Programme. For more information see www.kb.nl

[12] The project NDIIP 'Preserving our Digital Heritage. Plan for the National Digital Information Infrastructure and Preservation Program. A Collaborative Initiative of the Library of Congress.' (October 2002). The costs of this plan will amount to a total of up to 175,000 US K$ and are co-funded with 100,000 K$ by the Congress of the USA. Information on the status of NDIIP can be found on the web site of the Library of Congress www.loc.gov.

3.4 A Toolbox as an Enabling Framework

Obviously just one tool or strategy for permanent access to digital information will not be sufficient. A variety of tools and techniques will have to be designed and developed as an ongoing process to keep up with the progress of ICT, and guarantee future access to different data-sets, formats and applications. PATCH proposes an approach to develop an iterative process that creates both several concrete tools for permanent access as well as the 'PATbox' as a general framework for permanent access technology. This approach will guarantee that designing, building, refining, testing and delivering specific preservation tools will be performed in a co-ordinated manner. Additionally, by connecting the toolbox with the OAIS-standard (ISO 14721:2002), the solutions for securing permanent access to digital objects will be fit for use with digital archiving systems that are built according to generally accepted standards.

By focussing on the toolbox as well as on individual tools, exploitation and further development will be facilitated. As a consequence, the toolbox will not be the responsibility of a single institution but will be in the public domain, so that future developers can contribute to the toolbox. As such, the PATbox will support a general and ongoing development of digital preservation technology.

During the PATCH project, tools will be created for four main digital object types, fixed-format digital items, scientific data sets, web resources, and applications. Within these main types, certain formats will be chosen for the creation of the first tools. General usability of these tools will be guaranteed by the requirements imposed by the toolbox. For instance, the PATbox will define the generic characteristics of the access-component needed to interact with the infrastructure for digital archiving and preservation.

The PATCH project will deliver a set of preservation tools, compliant with the requirements of the PATbox. Initially the draft of the PATbox will provide the requirements for the tools, after which the development process can start. This will most certainly result in feedback on those requirements. So both the PATbox and the tools will be developed in an iterative process. The set of tools will allow managers of digital archives to take action when a certain digital format threatens to become obsolete, for whatever reason (e.g. medium obsolescence, technology obsolescence or format obsolescence). The PATCH-consortium is convinced that this is the best way to organise a project with a strong innovative character, aiming at persistence of information in ICT applications.

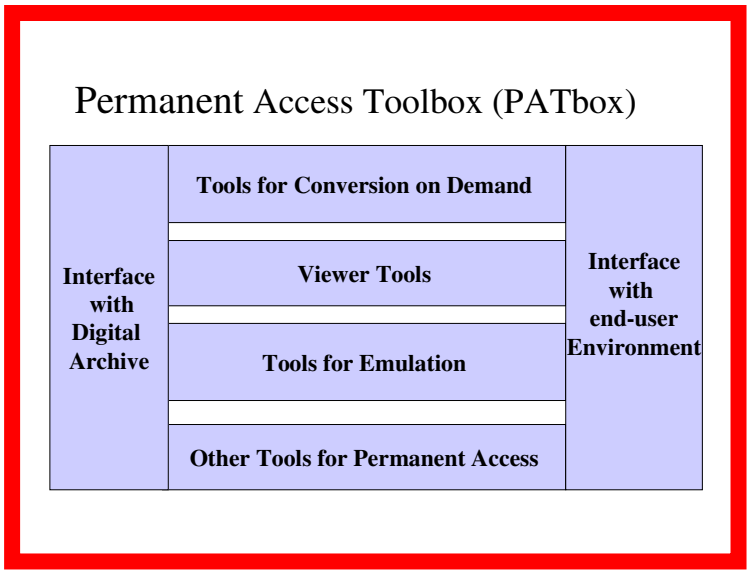

Figure 2: Graphical representation of the PATbox

3.5 Permanent Access Solutions

The experience with the technology for permanent access is still limited. A few promising concepts have been defined that - either separately or in combination - can be used to develop applicable solutions for guaranteeing permanent access to the information in digital objects. Probably the most general concept is to emulate obsolete hardware on future computer platforms and at the same time organise the safeguarding of the rendering software. However, before the emulation concept can be turned into real solutions there is still a lot of research and development to be done. In the one of the NEDLIB reports Jeff Rothenberg has analysed the feasibility of using emulation as a means of preserving digital publications in accessible, authentic, and usable form within a deposit library.[13] In the report a description is given of three subsequent iterations of an emulation experiment. However, because of the substantial amount of effort and resources needed, only little progress has been made since the emulation approach for preservation was first suggested.[14]

Another widely applicable approach is ‘data preservation’ with the 'UVC' (Universal Virtual Computer) proposed by Raymond Lorie (IBM). The KB and IBM have done some experimenting with this approach and the preliminary results look promising.[15] The real proof however, will be to build a first version of such a tool for a concrete type of digital information.

The essence of the data preservation concept is to extract the data from the format in which it has been published in order to render it in a different ICT environment, also in future times. The data preservation approach involves analysis of the content and logical structure of

[13] Jeff Rothenberg, *An Experiment in Using Emulation to preserve Digital Publications*. NEDLIB Report Series 1, Koninklijke Bibliotheek, 2000. (Full text at http://www.kb.nl/kb/resources/frameset_kenniscentrum.html)
[14] Jeff Rothenberg. Ensuring the Longevity of Digital Documents. Scientific American 272, 1995, pp. 24-29.
[15] Raymond Lorie, *The UVC method for preserving digital documents: proof of concept*, IBM/KB Long-Term Preservation Study Report Series 4, 2002. See http://www.kb.nl/kb/resources/frameset_kenniscentrum.html)

original file formats and conversion of the original into a clearly structured, documented, well-defined and self-contained format.

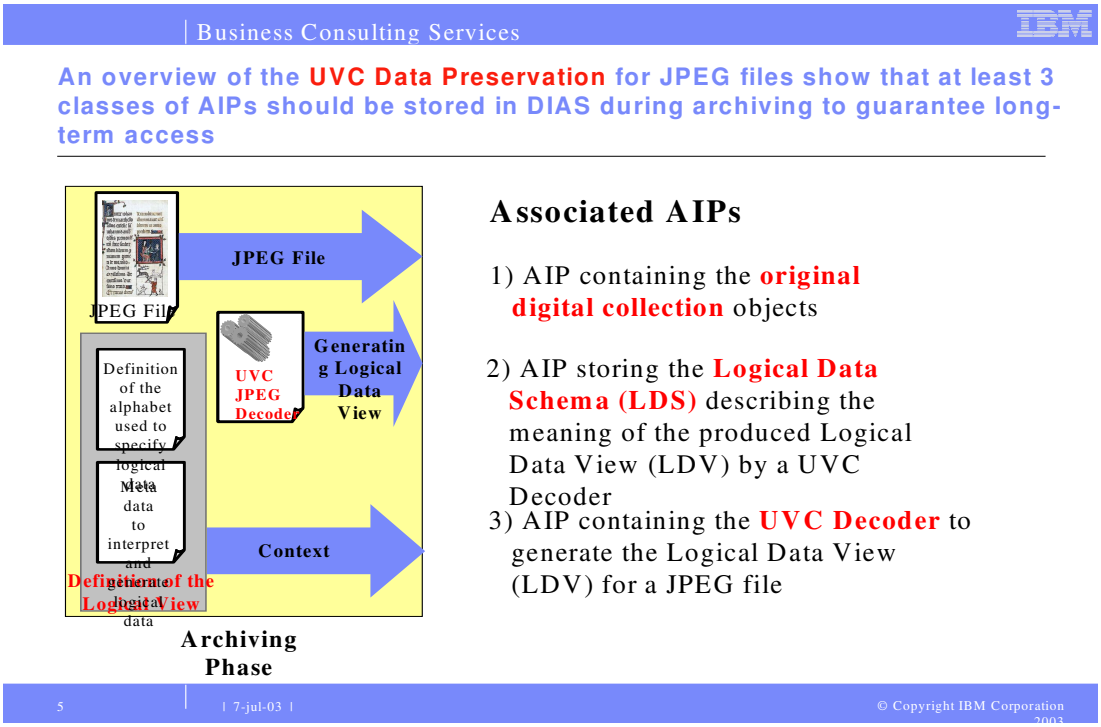

Figure 3: UVC data preservation (figure by courtesy of Raymond van Diessen, IBM-Netherlands)

ANNEX

Digital Preservation: State-of-the-Art

Development of archival systems
Memory institutions, as the custodians of our 'collective memory', were amongst the first to concern themselves with creating architecture and processes in the field of digital preservation.[16] An example is the European project started in 1998 by national libraries: NEDLIB (Networked European Deposit Libraries). The aim of this project was to define the basic infrastructure-components upon which a networked European deposit library could be built. The objectives of NEDLIB coincided with the mission of national deposit libraries to ensure that current electronic documents could be used now and in the future. The NEDLIB project created a consensus in the way of handling and storing digital documents. The digital documents should be separated from their original carrier or environment, which is intended for publishing and not for archiving, and stored in a controlled archiving environment.[17] Such a controlled environment is currently defined as 'a safe place'.[18]

Extension of OAIS with a preservation function
The NEDLIB study resulted in a number of reports, published in the year 2000, with instructions on how to set up a deposit system for electronic publications. The NEDLIB project has contributed to the development of the draft OAIS, originally designed by NASA (CCSDS 2001), by adding a Preservation Management Function. A process model for a deposit system has also been designed, based on the OAIS. The basic principle of the NEDLIB's process model[19] is to separate archiving from other functions i.e. searching, authentication and authorisation. The archiving system has to be integrated in the larger ICT-infrastructure of the institution that provides the other functions. This design, using separate components, ensures durability: every single part can be built and replaced as time goes by without being dependent on one provider and without it being too complex.

Implementations of OAIS
Several aspects of digital preservation have been researched in the CEDARS project. This was a co-operative project of the universities of Leeds, Cambridge and Oxford (April 1998 - March 2002). The project focussed on raising awareness, the creation of a Distributed Archive Prototype, based on the OAIS system and employing a Metadata Specification.
In 2000 the KB started the DNEP-project together with IBM to build a deposit system based on the OAIS-standard. In 2002, the system resulting from this project – the Digital Information Archiving System (DIAS) – was implemented for the 'e-Depot' of the KB. Other examples of digital repositories in development are the e-archiving project of the Delft Technical University, Dspace of Michigan University (US) and Fedora of Cornell University (US). Another example is the Digital Archive of the National Archives in the UK, which has meanwhile reached the phase of implementation.

[16] A more comprehensive listing of the various initiatives on digital preservation can be found on the PADI website: http://www.nla.gov.au/padi/.
[17] Johan Steenbakkers, *The NEDLIB Guidelines - Setting up a Deposit System for Electronic Publications*. NEDLIB Report Series, 5, Koninklijke Bibliotheek, 2000.
[18] UNESCO's draft 'Guidelines for Digital Archiving of Cultural Heritage' compiled by Colin Webb, in preparation.
[19] Titia van der Werf, *The Deposit System for Electronic Publications. A Process Model*. NEDLIB Report Series 6, Koninklijke Bibliotheek, 2000.

Current status of development
Summing up there are two requirements for long term preservation of digital information. These requirements are: a digital archive ('Safe Place') for storage and maintenance of the digital objects, and functionality and procedures ('Tools') for permanent access to the information in the digital objects.

Safe Place strategy
The word 'place' in the 'Safe Place strategy' (SPS) should not be taken literally. A 'Safe Place' is a concept, indicating institutions that are committed to digital preservation and possess appropriate infrastructure, resources and skills for the task. There will of course be more than one safe place to achieve redundancy of safekeeping. The functions and procedures of a 'Safe Place' have meanwhile been described as the standard Open Archival Information System (OAIS), the ISO standard 14721:2002.[20] An example of an operational safe place is the 'e-Depot' developed by the KB, the national library of the Netherlands, with support of IBM-Netherlands.[21] Major scientific publishers[22] have formally acknowledged the KB's e-Depot as the best approach to guarantee permanent access to their digital publications. Other implementations of SPS are 'JSTOR' [23] at the Universities of Michigan and Princeton, the 'Digital Archive' at the National Archives UK and 'PANDORA'[24] at the National Library of Australia, even if they have not been designed to comply fully with OAIS.

Even if it is not a preservation strategy as such, the project LOCKSS ('Lot of Copies Keep Stuff Safe') should also be mentioned here.[25] This approach is still very experimental and there are no real-life implementations yet. LOCKSS does not address the issues of maintenance and permanent access, but focuses on a continuous process of harvesting copies of networked digital data, as a method of replicating multiple copies of data across institutional or geographical boundaries.

Research on permanent access
While the development of 'safe places' is well on its way, the development of 'tools' for permanent access is still in its infancy. With permanent access we mean guaranteeing the rendering and viewing of stored digital objects.

Permanent access solutions can be characterised by the following two approaches, separately or in combination[26]:

- Migration or format conversion: migrate digital items to new technology/formats as they become available and as the old technologies/formats cease to be supported by vendors and the user community
- Emulation: emulating (imitating) old and obsolete technologies/formats on current technology platforms

[20] About OAIS: www.ccsds.org/documents/SO2002/SPACEOPS02_P_T5_39.PDF The Open Archival System (OAIS) Reference Model and its Usage, by Donald Sawyer et al. Note that the OAIS standard should not to be confused with OAI, the Open Archives Initiative, a standard for the harvesting of meta-data.
[21] Information about the deposit system build by KB and IBM www.ibm/nl/dias
[22] News on the archiving contract of Elsevier with the Koninklijke Bibliotheek: www.infotoday.com/newsbreaks/nb020903-2.htm Note: the wrongly used term 'Digitization' in the heading of this article, an example of confusing 'Digital Archiving' with 'Digitization'.
[23] See for information www.jstor.org
[24] See for information www.nla.gov.au
[25] See for information www.lockss.stanford.edu.
[26] Raymond J. van Diessen and Johan F. Steenbakkers, *The Long-Term Preservation Study of the DNEP project. An overview of the results.* IBM/KB Long-Term Preservation Study Report Series **1**, 2002.

A variety of initiatives
While projects like NEDLIB focused on a coherent and generic approach of the long-term storage-aspect of digital preservation, the research on permanent access has been scattered over several projects. CAMiLEON (Creative Archiving at Michigan & Leeds: Emulating the Old on the New), a joint project of the universities of Michigan (USA) and Leeds (UK), focused on some aspects of emulation technology and explored options for long-term retention.

DNEP is the name of the KB/IBM project that resulted in two years in a generic, OAIS compliant, archiving system DIAS. During the development of DIAS, KB and IBM also carried out a joint study, in order to define the requirements for long-term preservation.[27] In 2003 KB and IBM continue to work together on the development of an operational preservation module for DIAS.

The Universal Virtual Computer (UVC) is a concept developed by IBM Research. It uses a simple virtual machine as a stable platform to implement software that will decode (convert) the original bit stream of a digital item into a logical (XML-like) version of the content. This logical view of data can be understood without any additional technological assistance. In the future the logical view can be used to transform the content into the then appropriate data formats used by the then generally available programs. The UVC can provide a stable platform on which to run the decoders for the individual bit stream formats and obviates the need to migrate the UVC decoders through technology changes. A proof of concept during the project DNEP has underlined that the UVC concept is a viable approach that should be developed further.

One of the activities of the Dutch Testbed project (ICTU)[28], commissioned by the National Archive of the Netherlands and the Ministry of Interior Affairs (finishing July 1st 2003), has tested (together with IBM) the idea of the Universal Virtual Computer, one of the emulation-approaches. Furthermore, the Testbed project has tested the effect of migrating different documents to different formats.

Digital Preservation is…

The issue of digital information obsolescence is indicated by a variety of terms (long term preservation, digital preservation, digital archiving, permanent access, digital continuity, persistency of information etc.). This is caused by the infancy of this area of research and development and has to be dealt with for the time being. However the issue of persistent digital information is also sometimes confused with quite different topics.

…not digitisation
Digitisation is turning information digital, e.g. by scanning or keying in. Digitisation concerns making a digital copy of originally analogue recorded information (pictures, texts, maps, medieval manuscripts etc.) After being created however, the digital information has still to be

[27] The results of the Long Term Preservation Study has been published in a series of reports. IBM/KB reports see www.kb.nl/kb/ict/dea/ltp/reports/reports.html
[28] See for the Testbed project: http://www.digitaleduurzaamheid.nl/home.cfm

preserved for the long term, similar to born-digital information. And of course any solution for long-term preservation can be equally applied to digitised and born-digital information.

…not record management
By opposing appropriate requirements for the creation and handling of digital items – from the beginning of their life cycle on – their characteristics can be favourably influenced to make preservation easier. Archives in particular have been actively promoting requirements for appropriate record management. Record management is an important issue, but as such it is not digital preservation.

…not rights management
The word 'access' is often used in association with rights management. Then it concerns the rights to access and use the information. It is about ownership and (financial) conditions of usage. This is an important issue but has no direct relation to digital preservation.

…not legislation for depositing
Most countries have legislation (some have voluntary arrangements instead) for the depositing of printed information at the national libraries and archives. New legislation is needed (and in some countries is under development) for digital information. Another issue that needs legislation is the right of the archiving institutions to copy and manipulate digital information for preservation purposes, in spite of rights of the owner or creator. These legislation matters are important, but are different issues than digital preservation.

