

Handling electronic publications: Practices and Projects

Presentation at the 6th National BAD conference, Aveiro, 7 May

Marco de Niet

**Koninklijke Bibliotheek, National Library of the
Netherlands**

1. Introduction

Firstly, I would like to thank the Portuguese Library Association for inviting me to inform you about some of the practices and projects of the Koninklijke Bibliotheek, the National Library of the Netherlands. This paper was planned to be presented by my director Johan Steenbakkers, but he sincerely apologises for not being able to be here with you today. Johan Steenbakkers is the director Information Technology and Facility Management, and as such he is involved in all the projects and activities related to ICT developments at our library. I am only involved in a few of our projects, but I hope the outlines I will present, will nonetheless be of some interest to you.

Let me give you an impression of the Koninklijke Bibliotheek first. The KB was founded in 1798, which means that we are celebrating our bicentennial this year. During the 19th century and for the greater part of the 20th century the KB was mainly a humanities research library. Only in 1974 the KB founded the national deposit. As you may know, there is no deposit law in the Netherlands. All publications we have in our deposit are sent in by the Dutch publishers on voluntary basis. To achieve this, the KB has made agreements with various organisations in which publishers participate. Although there is no deposit law, we have one of the highest rates of coverage among the national libraries. Of all the commercial publications printed in the Netherlands, we have 99 percent in our deposit. Of all grey publications we have a coverage of about 75%.

Twelve years after the deposit was set up, the Koninklijke Bibliotheek became the national library of the Netherlands. As such, our major activities are:

- Collection development, documentation and management of the national cultural heritage
- Deposit library for Dutch publications and the national bibliography
- Research library for the humanities and social sciences
- Centre of expertise for preservation & restoration
- Netherlands Union Catalogue & ILL-coordination
- Research and development
- International focal point for libraries
- Support for co-operation among libraries.

We have a staff of about 300 employees. There are about 10.000 registered users, and we have about 3.000.000 items in our holdings. Finally, we participate in about 30 national & international projects. Most of these projects deal with information and communication technology and/or preservation issues. I would like to tell you a bit more about the ICT projects, and the practices that resulted from them.

The KB focuses on three activities regarding the setting up of our digital library: the Deposit of Dutch electronic publications, digitisation and the development of an advanced information workstation.

2. Deposit of Dutch electronic publications

The initiative for the development of the Electronic Deposit was taken in 1994. We have distinguished three strands of actions: researching, modelling, and getting hands-on experiences. In the past few years various national and international projects have been set up or carried out to tackle issues pertaining to digital deposit collections.

The first major project was DNEP-IWI. DNEP stands for Deposit of Dutch Electronic Publications, IWI is the national programme for innovation in the provision of scientific information. IWI is funded by the Ministry of Education, Culture and Science, and by participating institutions, such as the major research libraries and SURFnet, the national computer network for education and research in the Netherlands. The goals of the DNEP-IWI project, which ended in February 1998, were to define a workflow model for the KB for electronic publications and to study the changes that will affect the library as

a result of the move towards electronic publishing. All workflow aspects were taken into account: production, selection, acquisition, registration, installation, bibliographic description, de-installation, technical description, migration, storage handling, indexing, retrieval and access. Unfortunately, time won't permit me to elaborate on all these aspects now, so I would like to refer you to an article about our electronic deposit, which is published in the journal *Liber Quarterly*. Instead, allow me to focus on two results of this project: the definition of selection criteria, as defined by the Dutch Bibliographic Centre and storage of electronic documents.

Selection criteria

At first sight, the first criterium might seem obvious, but it is nonetheless a vital decision: it states that selection criteria applied to printed publications should also apply to digital documents. This means that we look at the geographical origin of a digital document (it should be published in the Netherlands), at the distribution pattern (it should be disseminated publicly), at the size of the publication (it should consist of 8 pages or more).

A second criterium is that the documents should be preserved "as is": in the form or format in which it is disseminated by the publisher.

A third criterium is that we only deposit "end-products", no semi-manufactured products, such as SGML databases, used as mother databases for a wide range of electronic publishing products. However, we do deposit all versions of these end-products (a publication may for instance be available as an online database, on cd-rom, on CD-I, on tape, or even on paper). It may occur that within a version of a publication, various formats will be used. If there is no difference in functionality, we will only deposit the "richest format", containing for instance information about final lay out or added functionality (e.g. the Word document instead of the version in ASCII). If there is a difference in functionality (for instance HTML versus PDF), we will deposit all these formats.

Finally, we have also decided on types of digital publications which we will not deposit: no films on CD-I with Dutch subtitling, no multimedia publications without self-contained text, and no software without content, such as text processing software. There are other institutions in the Netherlands that will take care of the preservation of these kinds of publications.

Storage of electronic documents

Another aspect that was studied in the DNEP-IWI project was the storage of digital documents. As the deposit contributes to the preservation of the cultural heritage of the

Netherlands, the focus was on long term storage. The only definite thing that can be said about this is that nothing is definite yet. There are still so many uncertainties, that further research is absolutely necessary. Nonetheless, the project helped us to define a few conditions. For instance, storage cannot be separated from retrieval options. In this respect we defined three ways of storage: hot, warm and cold storage. Hot storage means that the information is available for immediate and fast retrieval (e.g. on a hard disk). Warm storage means that the information is also immediately available, but as the information is stored on external media, such as magneto-optical disks, retrieval will be considerably slower. In the case of cold storage, the information is only available on a medium which is not connected to the network, such as a tape. This is used only for documents which will not be requested frequently.

One of the major uncertainties for the design of the electronic deposit is the format in which the information will be stored. If we want to stick to the original format, we will have to set up a library of technical data as well, to simulate hardware and software configurations once they have become obsolete. But we might also decide on storing all documents in one agreed deposit format. This implies migration and conversion, which most likely result in loss of information. And if we do so, how much loss will be acceptable?

These questions cannot be answered easily, and we are convinced that they will not be answered in the near future. We have set up new projects to study developments on the market, for instance regarding data-warehousing.

Modelling

We didn't just draw up some theoretical outlines for the electronic deposit, the project DNEP-IWI also took care of the modelling of the deposit with one hundred digital documents. Fifty off line and fifty on line documents went through the full back office procedures as mentioned above. This way, the cataloguing department could test the procedures that were set up. The bibliographical descriptions were made in the shared cataloguing system, maintained by Pica. In a different project the KB designed the Pica-format for digital documents, which it is now widely used by academic libraries in the Netherlands for cataloguing Internet resources.

Now the DNEP-IWI project has been finished, a few new projects have been initiated to take the model a few step further. I would like to mention two of them.

In the project CERBERUS, we investigate the copyright issues related to the digital deposit. Especially the terms of availability to library users are heavily discussed at the moment. Dutch law offers only limited possibilities to carry out preservation activities

to guarantee the availability of digital documents in the future. We now have lawyers studying the possibilities for the KB in this respect. After all, we do have the national task for preserving the intellectual cultural heritage.

A new major project we set up, is NEDLIB (Networked European Deposit Library), co-funded by the European Union. In this project nine national libraries and state archives participate, together with two software developers and three international publishers. The goal of NEDLIB is the joint development of a basic infrastructure upon which a networked European deposit library may be built.

Practice

To round off this section about the digital deposit: besides modelling the back office, we are also experimenting with available software for storing and retrieving digital documents. Currently, we are focusing on systems that process mainly electronic journals. We have tried three systems so far: Rightpages, designed by AT&T, ICL Commands, and we are currently installing IBM Digital Library. Although the retrieval facilities of IBM Digital Library are not very sophisticated, it is the best product so far in terms of scalability. We currently have several hundreds of journals by Elsevier, Kluwer and SDU stored in IBM Digital Library, which means we have about 120.000 journal articles full text available. The documents are available in (combinations of) Postscript, PDF or TIFF, depending on the way the publishers have deposited them. It may be noteworthy to mention that we were the first national library in the world to sign an official agreement with a publisher for depositing electronic publications. As we did not have to wait for the government to decide on a deposit law dealing with digital documents, we already reached an agreement with Elsevier Science in August 1996 about us safe keeping their scientific journals.

Starting with this basic collection of online journals, we hope to build a system that will host a large variety of digital documents. It may be that a small portion of these documents will be digitised by the KB. This leads me to the policy regarding digitising our own collections.

2. Digitisation

To be honest, so far we have not given digitisation much priority. We have done some small scale experiments, for instance by digitising one of our coffee table books, *A Hundred Highlights from the Koninklijke Bibliotheek*. The book shows a hundred

master pieces from our collections, accompanied by a short informative text. The book was published in 1994, the online version was launched in April 1995. To our own astonishment, the amount of hits is now, after three years, still rising. We have about 15.000 hits on this virtual exhibition each month. As a sequel, we set up a second virtual exhibition to provide information about various special collections we have. As small exercises, they gave us the opportunity to experiment with scan hardware and software, resolutions, etc.

There are currently two larger digitisation projects at the KB. One is a joint project with the British Library, in which we digitise two similar atlases from the 17th century, with geographic and historic maps. The atlases carry the names of the compilers of the atlases (Van der Hagen and Beudeker). In total about 450 maps will be digitised from both atlases.

The biggest digitisation project deals with illuminated medieval manuscripts. Since we have an important collection of these manuscripts we have decided to create a tool to provide extensive information on the 6.500 illustrations in these manuscripts. For the retrieval of these illustrations, the Iconclass system will be used.

Finally, we are also digitising radiographic scans of watermarks in books printed before 1540, the so-called incunabula and post-incunabula. These watermarks help define the date of printing of these books. For the time being, these scans will only be available on site. All the other images will be available across the Internet in the near future.

3. Advanced Workstations

Besides the digital deposit, we have invested quite heavily in the development of an advanced workstation. We started the project Advanced Information Workstation in 1993 with external funding, and the first phase ended last year. In the second half of 1998 we will start phase 2, to enhance the functionality of the workstation. The workstation offers integrated access to our electronic services, and is primarily designed for end users. As "integrated" is more or less the magic word in libraries nowadays, I would like to show you how we define integration. I would also like to tell you briefly how we take the wishes of our users into account.

Integration

Integrating library services for end users implies in our view three sorts of integration: technical, functional and bibliographical integration. With technical integration we refer mainly to the connecting of local and wide area networks and their protocols. These are

of no relevance at all to the end users, but still they often determine what information can be accessed or not. When we started with the workstation project in 1993, the KB was struggling with five network protocols, namely X.25 for our Pica library system, X.400 for email, TCP/IP for the Internet (still Gopher then!), Ultraset for our CD-ROMs and Novell for the various back office software such as data editors. The advent of the World Wide Web helped us tremendously, and we are standardising as much as possible on the HTTP-protocol nowadays. We have incorporated our workstation completely in our World Wide Web service. The PC's in the library on which all the local facilities can be used, are based on Windows NT, but we use Netscape and our web-site as interface to the workstation. Although we have not solved all the problems concerning technical integration, such as implementing a single digital payment device for all our information systems, we are confident that the developments on the Web will help us in this respect.

As important as technical integration, but often underestimated by libraries is functional integration. Many libraries I have visited tend to focus on creating technical gadgets with their workstations, instead of creating a tool for their visitors to make optimal use of the information they find while using the library's services. A properly designed workstation supports the users during every phase of their research. We have distinguished five of these phases:

- searching - the workstation serves as a quality gateway
- acquiring - the workstation offers various retrieval facilities
- processing - the workstation serves as a personal desktop
- communication - the workstation serves as a virtual address
- publishing - the workstation serves as a media provider for authors

Note that only the first two phases contain the traditional library functions, such as providing bibliographical information through printed or online catalogues. The concept of a workstation shows that libraries are transforming from information providers to service providers more and more. The workstation can thus be called the natural successors to the OPAC's.

This does not mean that we should neglect our bibliographical systems. On the contrary. The workstations offer a tremendous opportunity to finally make optimal use of all this bibliographical data that we are creating. After decades of library automation most libraries are still very much in a transition phase. Many important parts of their collections are not yet available through the OPACs, and many libraries have different databases for various types of publications, like books, journal articles, manuscripts,

maps, online resources. Setting up a workstation forces you to think how to connect all these systems, to work towards *bibliographical integration*. One way is to use a common search protocol, such as Z39.50, to allow simultaneous searching in different systems. But in some cases it may be better to set up a metacatalogue, that incorporates all records from various databases. Such a metacatalogue could contain various types of secondary information that link to primary publications. They can be formal descriptions, like in a library catalogue, but they can also be free descriptions, as available on the Internet. If available, summaries could be incorporated and indexed in such a metacatalogue as well. Nowadays there are more ways to give access to primary publications outside a catalogue, for instance through online tables of contents or subject guides. These ways of bringing the users to the publications they are looking for are becoming more and more important in our view, and they should be incorporated in the workstation as well.

Users

So far I have talked about users in the most generic way. Of course, every library has its own target groups, but when developing a workstation, probably four categories are relevant to all libraries:

- Local vs. remote users: should the users come to the library, or do we offer them our services if they are working elsewhere as well?
- Registered vs. anonymous users: do we offer our services only to those who have a library pass? Or do we also help those on the other side of the world? With the Internet, the anonymous use of library systems has grown tremendously.
- Individuals vs. groups: Many users would like to customise the workstation to their own needs, but this often requires a lot of development. In some cases it can be beneficial as well, to use group profiles, based on for instance study disciplines or professions.
- Staff vs. visitors: although we have focused mainly on the end users, the workstation can also be beneficial to staff, who will require special facilities that usually are not available to the visitors.

While designing the pilot version of the workstation at the Koninklijke Bibliotheek, we decided to focus on specific target groups to learn more about the wishes and needs of modern scholars. We invited for instance twenty art historians to evaluate the prototype. In March 1998 we started a new project as an extension to the workstation project. The goal of this project, named Bibliopolis, is to develop a research tool for the history of the book in the Netherlands. All online information systems that can be beneficial for

the study of the book will be incorporated and integrated in this tool, which in its turn, will be incorporated in our workstation.

Conclusions

This leads me to some conclusions and recommendations, based on our experiences:

Firstly, when building a digital library, always work according to the overall policy of the library, in terms of services, automation plans etc. Experimenting is OK, but a system is bound to fail if it will not be supported by the organisation in the end.

Invest in infrastructure: do not focus on gadgets that are superficial. These are bound to be outdated very quickly. Having good local and remote network facilities is essential for the development of a digital library.

Define your virtual collections, as in our case we have defined criteria for our digital deposit collection and our research collection.

Don't forget the user: he is expecting advanced facilities nowadays.

A good way to work is to shift from projects to operational services to projects again. This way, many staff members who are not directly involved in the projects will be kept up to date about new developments, and in return, they can give invaluable feed back.

It is obvious that we should not try to invent the wheel again all on our own. Unlike a traditional library, a digital library with its electronic publications is not restricted to one location. Sharing knowledge, information, technology and means is a good way to progress.

References

More information about the projects of the KB can be found at the website of the Koninklijke Bibliotheek: <http://www.kb.nl>

M. de Niet: A single access point to information resources, The Advanced Information Workstation of the National Library of the Netherlands. In: *Resource sharing & information networks* 13 (1998) 2, p. 29-37.

T. Noordermeer, J. Steenbakkers & T. van der Werf-Davelaar: Electronic Library Developments in the Netherlands. In: *Liber Quarterly* 8 (1998) 1, 57-80.