



IFLA
2005
OSLO

World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:

<http://www.ifla.org/IV/ifla71/Programme.htm>

2nd Version: June 21, 2005

Code Number:

034-E

Meeting:

150 SI - ICABS (IFLA/CDNL Alliance for Bibliographic Standards)

Nestor and kopal - two national initiatives to ensure long-term accessibility of digital documents in Germany

Reinhard Altenhöner

Head of IT Department
Die Deutsche Bibliothek
Germany

Abstract

The upheaval of the publication market towards the electronic publication via networks significantly changes the task profile of libraries – especially of those libraries that are responsible for the tradition of the cultural heritage of a nation. This initial situation for Germany is specifically serious because a large number of different libraries have been entrusted with the task of collecting, indexing and archiving of the publications there. It is a task that has to be newly shaped by organisational and technical aspects for a changed universe of publications. There is, however, a cross-community understanding that all those activities have to be undertaken in a co-operative effort.

Thus, in the present contribution some prevailing efforts from Germany are being presented: Whereas, on the one hand, it is being worked towards a clarification of the legal framework for the new spectrum of tasks, on the other hand, two initiatives were created: Once, the build-up of a cross-institutional competence network on the aspects of long-term preservation (nestor); second, the co-operative development of a depot system for digital resources (kopal). Both lines of activity in which Die Deutsche Bibliothek is leading involved, are being presented.

1. General perspective

The publication process becomes more and more electronic. Yet, we still do not know how digital publications, works of art, image and sound documents, files, primary data and music can be archived in a way that they will remain permanently readable and thus accessible to all in the future. What we can say is: Whatever strategy will be followed in the future to provide access to the digital content, it will depend on the existence of a bit stream, the integrity and authenticity of which has been kept over the years. On the basis of the preserved bit stream, document rendering will have to be enabled for future access to digital objects. Several strategies are under discussion, which can be summarized in two action lines:

- to migrate the electronic objects in a controlled environment
- or to emulate the historic system environment from the origin time of the object – including the emulation of hardware and system software.

Along with the availability of the bit stream we need additional metadata information concerning especially the technical information about the stored objects, their original technical environment and their storage history. The task to establish a system and an infrastructure for long-term preservation is complex and requires a good deal of money and man power and we need both a network and the exchange of knowledge – national as well as international.

2. Perspective of Die Deutsche Bibliothek.

Die Deutsche Bibliothek¹ is the national library and national bibliographic information centre for the Federal Republic of Germany. The library is responsible for the collection, processing and bibliographic indexing of all German and German-language publications issued since 1913. This task is based on a statutory mandate for the collection, bibliographic processing and long-term preservation of all publications released in Germany or published in the German language abroad. The law also covers digital publications distributed on physical carriers, but has as yet not made any provisions for online publications. At present we are waiting for a new legal basis, which will extend our area of responsibility to all types of net or electronic publications. As a consequence of the new law we expect that many regional libraries with legal deposit for special regions will get a new legal foundation for their collections of digital publications.

We will have to be prepared for this situation. Therefore, in the last years Die Deutsche Bibliothek has started several initiatives to promote the long-term preservation of digital publications in Germany. A number of basic principles applicable to the collection of online publications were defined in preliminary hearings with publishers, library experts, information specialists and government officials and formulated in a policy document passed by the Publishers' Committee of the Börsenverein des Deutschen Buchhandels in June 1997:

- All online publications are to be submitted upon request via data networks or on physical data media
- Online publications available in different forms are to be submitted in the format requested by the library
- Publications with identical contents distributed both on physical media and as online publications are to be submitted in both forms.
- Online publications with identical contents distributed simultaneously by multiple providers need only to be submitted once.

Based on this policy, Die Deutsche Bibliothek has tested procedures for the submission, collection and long-term preservation of online publications in co-operation with publishers

¹ Die Deutsche Bibliothek: <http://www.ddb.de/>

and producers in a test phase lasting several years. In the process, the 'Electronic Deposit Library' working group explored and established the conditions necessary for Die Deutsche Bibliothek to become a deposit library for online publications as well:

Since 1998 online dissertations and theses (36,000 until now) have been collected, archived, and made available on a document server. Since 2000 electronic periodicals have been collected, and since 2001 Die Deutsche Bibliothek has been operating a submission interface for online publications. During the submission procedure, Die Deutsche Bibliothek has also been asking for technical metadata relevant for preservation purposes. A compromise has to be made between the publishers who are willing to bear workload under the conditions of voluntary submission, and the extensive requirements of future preservation processes in the deposit system. Furthermore, Die Deutsche Bibliothek participated in the European Nedlib-project, in which the OAIS-model was adopted, and workflow suggestions for the integration of library procedures were developed.

Other important experiences:

- Die Deutsche Bibliothek gathered experiences with its Multi-Media Access System (Multimedia-Bereitstellungssystem - MMB). MMB enables storage and access for digital objects on physical carriers. Different object types (workstation image, application installation kit, file collection, presentation object) have been implemented to provide for the rendering of complex digital objects (applications).
- Another activity covers the development of a persistent identifier infrastructure for Germany. The use of persistent identifiers guarantees the stability and unequivocal possibility to address a digital object in the net².

3. Federal structure of Germany

Germany has a federal structure with important elements of self-government in the federal states concerning especially the education-system and the science and research sector. The existence of many regional libraries with the legal deposit for their regions is yet another part of the federal structure. Considering the importance of the task of long-term preservation within the federal structure of Germany it is obvious that a successful solution can only be achieved by a co-operative approach. This applies primarily to the organisational aspect, but there are many practical and technical reasons why we try to distribute the responsibility to collect the electronic objects in the net.

Thence, the partners we need to implement a long-term preservation strategy and infrastructure in Germany have to cover the organisational as well as the operational area.

4. National initiatives in Germany

There are two initiatives that deal with the problem of long-term preservation:

- On the organisational side there is nestor (Network of Expertise in Long-Term Storage of Digital Resources) that was established with the goal to build up a platform of competence to exchange knowledge and experience in the field of long-term preservation and to exchange experts and expertise between different types of cultural heritage institutions.
- The kopal project (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen or Co-operative Development of a Long-Term Digital Information Archive) on the operational side was established to build up a running long-term archival system based on OAIS.

² So embedded in the project "EPICUR - Enhancement of Persistent Identifier Services - Comprehensive Method for unequivocal Resource Identification" persistent identifiers become a part of a metadata framework for electronic publications

5. nestor

Financed by the Federal Ministry of Education and Research, nestor, the alliance for Germany's digital memory, started in 2003 and will be terminated in 2006. Under the leadership of Die Deutsche Bibliothek nestor's partners come from the library area (Bavarian State Library, Lower Saxony State and University Library), from a media centre (Computer and Media Service of Humboldt University, Berlin), from archives (Bavarian State Archive – Head Office) and for the museum area (as a multiplier the Institute for Museum Studies, Berlin). In addition there is the nestor Advisory Board constituted by publishers, representatives of science & technology, museums, archives, libraries and universities and as well as members of culture & politics and research institutions / computing centres.

It is the aim of the nestor project to bring together existing knowledge, people and expertise on long-term storage of digital resources as a starting point for a future alliance for Germany's digital memory.

Thus, within the project we

- create a network for information and communication about present and future long-term preservation activities in Germany
- establish a cross-sectoral community to promote and support long-term preservation activities and to raise awareness in society
- trigger synergies between on-going activities in Germany and co-operate with international partners and projects
- develop strategies for co-ordination of long-term preservation activities in Germany
- propose a long-lasting organisational model to continue the service as a network of excellence after the end of nestor project in 2006

The tasks in detail are:

- Collecting and presenting information
- Consolidating areas of expertise and making them visible and available
- Promoting co-operation and supporting a common solution
- Preparing materials on technical, organisational and legal issues
- Presenting models, putting them forward for discussion and encouraging widespread best practice
- Promoting standards and representing Germany on international standardisation committees
- Developing collection guidelines and selection procedures for the storage of digital sources
- Coordinating responsibilities for long-term tasks
- Raising awareness of the problem in specialist documentation circles and in the public
- Preparing a permanent organisation which co-ordinates and represents the concerns of long-term storage

The working groups and the expertise are an important part of nestor. At the moment the following groups are operating:

- nestor Working Group on Trusted Repository Certification
- nestor Working Group on Multimedia Archiving
- nestor Working Group on Preservation Policies and Selection Criteria (in preparation)

The expert reports address the following topics:

1. Electronic Journals
2. Perspectives of long-term preservation of multimedia objects
3. Development of a descriptive profile for a national long-term preservation strategy (Preservation Policy)
4. Digital long-term preservation and law
5. Study of the state of existing research data and raw data from scientific activities
6. A comparison of existing archiving systems
7. Digitisation and preservation of digitized material in German museums

Some of these reports have already been published and are available on the web.

6. kopal

kopal faces the problem of long-term preservation from a more practical approach in a collaborative way – technically and organisationally: Supported financially by the Federal Ministry of Education and Research, kopal is supposed to develop an innovative technical solution in the form of a reusable long-term archive for digital data. The solution is based on DIAS (Digital Information and Archiving System), jointly developed by IBM and the National Library of The Netherlands in The Hague.

Within the project, which started in July 2004, two of the project partners, i. e. Die Deutsche Bibliothek and the Lower Saxon State and University Library (SUB Göttingen) will file digital material of all kinds via mass processes into the long-term archive. It ranges from digital documents in PDF, TIF or TeX formats to complex objects such as digital videos.

The system will be implemented in accordance with international standards for long-term archiving and metadata within the OAIS framework (Open Archival Information System). A transparent integration into existing library and information systems is guaranteed and is a fundamental objective of the project.

To get an open, upgradable and adaptive solution several types of partners have been integrated in the project. Starting from the different kinds of intention it was decided to separate the aspect of system running from that of development. I. e. our partner for the operating of the system (the computer centre partner Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG)) wants to get experience with a well documented and scaleable system, in order to gain further partners, who will use the benefits of the system for their own needs. On the other hand there is a neutral platform, where the developing partners (Die Deutsche Bibliothek and SUB Göttingen) have to find common solutions for the software and the preservation procedures. Additionally, the system has to become capable to handle multiple users in a way that guarantees the independency of the single partners.

Concerning the software architecture, a separation has been determined between the core functionality of the archiving system and the environmental tools, which handle the

homogenisation and the transfer of digital objects into the system. The separation of a core functionality (DIAS-Core) demands well defined interfaces and an open usability for future partners: The tools to build standardized Submission Information Packages (SIPs) and import them into the system will be developed by the partners with an open source licensing (GPL) method. For the presentation system for objects we follow the same rule: We want to become independent from special dedicated solutions, towards well defined interfaces, open for other partners and systems.

The development partner for the DIAS-Core is the IBM Deutschland GmbH. This will ensure a professional adoption of software components and provide stable long-term support.

In this way, from the start kopal has been integrating several partners at different locations. Thus, in the future this long-term archive for digital information will provide other institutions with the opportunity to keep their data available on a long-term basis. Consequently, kopal ensures an opportunity for academic, business and administrative use that extends beyond libraries.

Important complementary components to the existing DIAS-system are:

- realisation of monitoring functions to prepare a long-term preservation of digital documents
- flexible data import and export functions based on the object description scheme METS (Metadata Encoding and Transmission Standard of the Library of Congress, USA) and LMER (Langzeitarchivierungsmetadaten für elektronische Ressourcen)³
- possibility of reuse for interested institutions

For these reasons much effort is spent on defining a powerful data model and architecture within the system. In the course of the project several information events and workshops are being planned on a national as well as on an international level.

³ <http://www.loc.gov/standards/mets/> and <http://www.ddb.de/standards/pdf/lmer12.pdf>