nestor: Network of Expertise in Long-term Storage of Digital Resources

A Digital Preservation Initiative for Germany

by Susanne Dobratz and Heike Neuroth

Susanne Dobratz (dobratz@cms.hu-berlin.de) is Head, Electronic Publishing Group, a joint effort of the Library and the Computer and Media Services, at Humboldt University, Berlin. Heike Neuroth (neuroth@sub.uni-goettingen.de) holds a Ph.D in Geology and is Head, Research and Development, Göttingen State and University Library (SUB), Germany. This article originally appeared in D-Lib Magazine (April 2004) and is reprinted with permission of the authors.

1. Digital Preservation as a cooperative task

Sponsored by the German Ministry of Education and Research¹ with funding of 800.000 EURO, the German Network of Expertise in long-term storage of digital resources (*nestor*)² began in June 2003 as a cooperative effort of 6 partners representing different players within the field of long-term preservation.³ The partners include:

- The German National Library (Die Deutsche Bibliothek⁴) as the lead institution for the project
- The State and University Library of Lower Saxony Göttingen (Staats- und Universitätsbibliothek Göttingen⁵)
- The Computer and Media Service and the University Library of Humboldt-University

Berlin (Humboldt-Universität zu Berlin⁶)

- The Bavarian State Library in Munich (Bayerische Staatsbibliothek⁷)
- The Institute for Museum Information in Berlin (Institut für Museumskunde⁸)
- General Directorate of the Bavarian State Archives (GDAB⁹)

As in other countries, long-term preservation of digital resources has become an important issue in Germany in recent years. Nevertheless, coming to agreement with institutions throughout the country to cooperate on tasks for a long-term preservation effort has taken a great deal of effort. Although there had been considerable attention paid to the preservation of physical media like CD-ROMS,

technologies available for the long-term preservation of digital publications like e-books, digital dissertations, websites, etc., are still lacking. Considering the importance of the task within the federal structure of Germany, with the responsibility of each federal state for its science and culture activities, it is obvious that the approach to a successful solution of these issues in Germany must be a cooperative approach.

Since 2000, there have been discussions about strategies and techniques for long-term archiving of digital information, particularly within the distributed structure of Germany's library and archival institutions (see [2] or [3]). A key part of all the previous activities was focusing on

using existing standards and analyzing the context in which those standards would be applied. One such activity, the *Digital Library Forum Planning Project*,¹⁰ was done on behalf of the German Ministry of Education and Research in 2002, where the vision of a digital library in 2010 that can meet the changing and increasing needs of users was developed and described in detail, including the infrastructure required and how the digital library would work technically, what it would contain and how it would be organized. The outcome was a strategic plan for certain selected specialist areas (see [1]), where, amongst other topics, a future call for action for long-term preservation was defined, described and explained against the background of practical experience.

As follow up, in 2002 the *nestor* long-term archiving working group provided an initial spark towards planning and organising coordinated activities concerning the long-term preservation and long-term availability of digital documents in Germany. This resulted in a workshop,¹¹ held 29 – 30 October 2002, where major tasks were discussed. Influenced by the demands and progress of the *nestor* network, the participants reached agreement to start work on application-oriented projects and to address the following topics:

- *Overlapping problems*
 - Collection and preservation of digital objects (selection criteria, preservation policy)
 - Definition of criteria for trusted repositories¹²
 - Creation of models of co-operation, etc.

- *Digital objects production process*

- Analysis of potential conflicts between production and long-term preservation
- Documentation of existing document models and recommendations for standards models to be used for long-term preservation
- Identification systems for digital objects, etc.

- *Transfer of digital objects*

- Object data and metadata
- Transfer protocols and interoperability
- Handling of different document types, e.g. dynamic publications, etc.

- *Long-term preservation of digital objects*

- Design and prototype implementation of depot systems for digital objects (OAIS [5] was chosen to be the best functional model.)
- Authenticity
- Functional requirements on user interfaces of an depot system
- Identification systems for digital objects, etc.

At the end of the workshop, participants decided to establish a permanent distributed infrastructure for long-term preservation and long-term accessibility of digital resources in Germany comparable, e.g., to the Digital Preservation Coalition in the UK.¹³ The initial phase, *nestor*, is now being set up by the above-mentioned 3-year funding project.

2. Focal points regarding content

The major aims of *nestor* regarding content can be summarized as follows:

1. *Increasing awareness.* Awareness of the problems related to, and the exigency of, digital long-term preservation within museums, archives and libraries will be increased and this information will be disseminated to policymakers to convince them of the need for action in this field.

2. *Creating a network of information.* Information concerning technical, organizational, legal and other aspects of digital long-term preservation will be acquired, and the existing information about current research, projects and “best practice”-results will be bundled and disseminated.

3. *Cooperating with national and international institutions.* National (German) and international strategic alliances will be established between archiving institutions (libraries, archives, museums) on one side and industry and research facilities on the other in order to solve together mutual challenges with respect to the long-term preservation of digital resources.

4. *Coordinating strategies.* Subject to the international context, an interdisciplinary forum will be set up for the development and coordination of strategies for long-term preservation of digital sources in the German cultural sector. Services, technologies and standards for long-term preservation of digital resources will be developed.

5. *Presenting a long-lasting organizational model.* At the end of the project, *nestor* partners will present a concept for a long-lasting organizational model for the network of excellence. In addition, network partners will vote on which tasks are going to be shared and which will be completed solely by individual organizations, in view of the delimitation between libraries, archives and museums. In order to record the current state of long-term digital preservation efforts and current demands being made on the network of excellence, 7 expert reports were commissioned at the end of 2003. The first ex-

pert reports are expected to be completed by mid-2004.

The stated aim of the expert reports is to analyse in detail the requirements in Germany as to specific topics or questions and to make recommendations, if necessary. These recommendations will then be discussed with the wider public at special workshops. Following the workshops, necessary work in these areas (project applications, groups of experts, etc.) will begin. The expert reports address the following topics:

1. *Electronic Journals.* Electronic journals (e-journals) are an important publication type and are of high interest for communication within science and research communities. It is therefore one of the major aims of scientific libraries to make this type of digital resource (e-journal article) available for the long term. Questions concerning the quantity, the technical standardization of the objects, the transfer of data, etc., will be the focus of the expert report on electronic journals.

2. *Perspectives of long-term preservation of multimedia objects.* The term "multimedia" covers all objects that are not based on texts – from pictures, audio, video, visualizations and animations to complex applications. The problems that face those tasked with the archiving of multimedia include a particularly high cycle of innovation, a multitude of technical formats, and insufficient international standardization. Specific questions like sustainability, media types, technical formats, metadata, organization, etc., are dealt with in this expert report.

3. *Development of a descriptive profile for a national long-term preservation strategy (Preservation Policy).* There are many different guidelines with varying scopes and validities for the numerous aspects of long-term preservation. Which criteria/questions are described in the

already existing policies? Can the main focus of the existing guidelines be identified? What are the differences between the published policies on which common ground can be found? What does the business model of long-term preservation look like? Which expense factors need to be considered, and how can they be financed?

4. *Digital long-term preservation and law.* The basic legal conditions for long-term preservation of digital resources in Germany have to be followed. The expert report on this topic will reveal the legal requirements for digital preservation and will assess the legal deficiencies that may exist in Germany. Concerning copyright, the following points are relevant: the right to make archival copies, the right to carry out modifications (migration of data), the removal of restrictions by the digital rights management (DRM) system (e.g., copy protection, individualization of users) and the conditions imposed for the access of archived resources.

5. *Study of the state of existing research data and raw data from scientific activities concerning their requirements and suitability for being archived or put at disposal in Germany.* In this expert report the following questions will be examined in cooperation with Germany's large research institutions: How is it possible to provide scientifically relevant research data in Germany, and how can such data be archived? Which data are relevant for archiving? What is the expected volume of data? How are the research data going to be stored (e.g., in databases, on tapes)? In what different formats are these data? Will third parties have access to the data? Are there already archiving approaches in existence for these data? For which purposes have the raw data been reactivated out of organizational archives up to now, and what might these data be used for in future? Which form of availability, what reaction time and which access

mechanisms will be necessary in this field? Do the raw data already contain a means for self-documentation? Does this make sense in the view of former experiences with aged data, and if so, to what extent?

6. *A comparison of existing archiving systems.* There are already many archiving systems in operation, e.g., D-SPACE, Fedora, IBM-Dias, etc. Therefore, having a descriptive overview and a comparison of existing commercial systems as well as of non-commercial systems is of great interest. This expert report will stress chosen software solutions.

7. *Digitization and preservation of digitized material in German museums.* The focus of this seventh expert report will be an inquiry about the digitized objects that currently exist within German museums. Furthermore, it will be interesting to learn whether there are already strategies and guidelines in place for the preservation of these objects.

The expert reports are being coordinated by one of the project partners. In addition, every *nestor* partner must address five work packages within the project period. At the end of the project, work on these five will lead to concrete recommendations, guidelines and best practices for Germany. The five work packages include:

- Developing criteria for trustworthy digital archives
- Recommending procedures for the certification of archives
- Recommending collection development policies and selection criteria for archiving digital sources
- Setting up a work structure for the long-term preservation of digital sources in museums
- Developing a concept for a long-lasting organization of the network of excellence and the information forum beyond the end of the project

3. German Network of excellence for digital preservation

The German network of excellence represents a cooperative approach, bringing together experts, experience, partners, interested institutions, etc., with as their primary goals: development of a common homepage for the network; development of an information platform; and development of a communication platform.

1. Development of a common homepage for the network

The development of a homepage (<http://www.langzeitarchivierung.de>) that combines all information, offers, reports, etc., into a single access platform, is to the fore. This platform serves as a starting point for networks and working groups, and as a place where project ideas can be interchanged and all people interested in digital long-term preservation can be brought together to share challenges, experience and knowledge.

2. Development of an information platform

The information platform consists of the following modules: subject gateway, expert database, reviews, news, catalogue, newsletter, calendar.

- *Subject gateway* (<http://nestor.sub.uni-goettingen.de>) – This module is based on the development of a web-based database (PHP, MySQL) for the investigation of relevant German, Austrian and Swiss information about digital preservation. The subject gateway enables browsing by subject (e.g., preservation metadata, persistent identifier, national strategies) or by docu-

ment type (e.g., article, journal, conference, project, etc.). The provision of the *nestor* gateway is founded on previous experience in subject gateways at SUB Göttingen and is for the most part based on the concepts of PADI¹⁴ (e.g., technology, meta-data).

- *Expert database (under development)* – Topic experts on every subject that is also listed in the Digital Preservation subject gateway will be listed in an expert database. The experts may be nominated by colleagues, or they may nominate themselves. However, the nominations are subject to editorial control to ensure that only those who are valid experts are registered in the database and that the registered experts are available to take enquiries. The expert database is still under development, and during the first stage, the priority is finding experts fluent in German.

- *Reviews* (<http://nestor.sub.uni-goettingen.de/review/>) – The *nestor* project group and selected external experts are able to enter reviews concerning document types and other fixed topics directly into the database. A list of all the available reviews can be generated from the database via an extra button, and in addition, the records in the “Digital Preservation” subject gateway will highlight every subject entry that has a review associated with it.

- *News* (<http://nestor.sub.uni-goettingen.de/news/>) – A news ticker will provide information about national and international events (e.g., the creation of a new Digital Curation Centre in UK). All news items will be archived, and the archives will be searchable in future.

- *Catalogue* (http://nestor.sub.uni-goettingen.de/new_entry/) – Lists of newly catalogued records in the subject gateway. The expert database and the event database are available separately. In the future personalized access to the catalogued records is planned including an alert service.

- *Newsletter* (<http://nestor.sub.uni-goettingen.de/newsletter/>) – A newsletter will be published every two to three months in order to provide information about current developments, references to relevant projects, initiatives, conferences, etc. The newsletter may be obtained automatically via an announcement service.

- *Calendar* – A calendar of events will be automatically generated from the subject gateway. The monthly list will provide information about future conferences, workshops, etc., *nestor*-specific events will be listed in a way that is distinct from other national or international events. (Although not available at this time, the URL for the calendar will be (<http://nestor.sub.uni-goettingen.de/calendar/>).

All of the services described above are being developed in two languages (German and English) and are based on internationally accepted standards.

3. Development of a communication platform

The communication platform provides a place for a two-way exchange between the project partners and the community. The platform enables the establishment of working groups for various topics and supports them by providing an appropriate infrastructure in terms of mailing lists, a cooperative work space system, a web-based questionnaire tool, etc.

The mailing list technology used is a Sympa Mailing List Server at Humboldt-University. Initial mailing lists have been set up for the purpose of:

- *Spreading the nestor newsletter to the community*. Currently, there are approximately 200 subscribers to the newsletter from Germany and German-speaking countries like Austria

Figure 1: Screen shot from the *nestor* web site.

and Switzerland. When the newsletter has been translated into English, the project expects the number of subscribers to grow. (To subscribe, see <http://nestor.sub.uni-goettingen.de/newsletter/anmelden.php>). The newsletter archive is available at: <http://nestor.sub.uni-goettingen.de/newsletter/archiv.php?show=1>.

- *Discussing activities and topics within the broader community via email.* (To subscribe to this discussion list, see <http://www2.hu-berlin.de/nestor/mail/index.php>.) The list archive is at <https://appel.rz.hu-berlin.de/sympa/wwwsympa/arc/langzeitarchivierung-nestor>.
- *Contacting the project partners.* To contact the project partners send email to info@langzeitarchivierung.de.
- *Maintaining internal lists for working groups, the advisory board, and experts.*

In addition, in order to support the working groups, a Tiki Wiki system has been set up.

Survey questionnaires for both the work being done on the expert reports and the work of several working groups are hosted on the same server,

to guarantee interoperability of data and tools. The first survey is planned for April/May 2004 within the context of the expert report "Development of a descriptive profile for a national long term preservation strategy (Preservation Policy)."

The communication platform is part of the *nestor* network and can be reached via <http://www2.hu-berlin.de/nestor>.

4. Long-lasting organizational concept

The goal of the *nestor* project is to operate *nestor* like the Digital Preservation Coalition in the UK is operated. Once the project ends and *nestor* leaves the funding cycle, it could serve as a membership model for libraries, archives, museum, and publishers.

Recently, a first step was taken to ensure the organisational stability of *nestor* as a cooperative network into the future: An advisory board has been established with representatives from different communities, in-

cluding libraries, archives, museums, publishers, the scientific community, the open access community, and research funding institutions, and from the Conference of German Federal Ministers of Culture.

5. International integration and cooperation

Several *nestor* services have been planned to make it possible for *nestor* to cooperate with international partners. This includes offering services in two languages as well as developing with an eye to technical interoperability. In addition, a future collaboration between the *nestor* subject gateway and the Australian subject gateway PADI is planned. With that in mind, *nestor* is using metadata compatible with OAI-PMH and Z39.50. This allows for different interfaces to be implemented quite simply within both *nestor* and PADI services, either for metadata exchange or for enabling distributed search via a single interface. All subject relevant metadata (keywords, classification, description, review, document type) are also available in an English language version.

Nestor is already collaborating with other international working groups, such as the working group PREMIS (PREservation Metadata Implementation Strategies, <http://www.oclc.org/research/projects/pmwg/>). The PREMIS working group intends to develop recommendations, proposals and best practices for implementing long-term preservation metadata (PREMIS). The *nestor* project partner SUB Göttingen is also a member of two subgroups (WG preservation

metadata, WG OAIS implementation) and will bring the results of those directly to the German community.

Nestor intends to engage in future cooperative relationships with national initiatives (e.g., the Digital Preservation Coalition, the Digital Curation Center) and international projects (e.g., Erpanet, DELOS). A prerequisite for these cooperative relationships is that the different *nestor* services have reached a stable stage of development, which is expected by April 2004.

6. Contact information

For those who wish more information, contact information is provided below.

Nestor project

The *nestor* project can be contacted via email to <info@langzeitarchivierung.de>.

Individual project partners

German National Library
(Die Deutsche Bibliothek)
Ulrich Tiedau
Die Deutsche Bibliothek,
Deutsche Bibliothek Frankfurt
am Main
Adickesallee 1, D-60322
Frankfurt am Main, Germany
email: tiedau@dbf.ddb.de
phone.: +49-69-1525-1141
fax: +49-69-1525-1010

The State and University Library
of Lower Saxony Göttingen
(Staats- und Universitäts-
bibliothek Göttingen)
Dr. Heike Neuroth
Niedersächsische Staats- und
Universitätsbibliothek
Göttingen
Platz der Göttinger Sieben 1,
D-37073 Göttingen, Germany

email: neuroth@sub.uni-
goettingen.de
phone: +49-551-39-3866
fax: +49-551-39-5222

Humboldt-University
Susanne Dobratz
Humboldt-Universität zu Berlin
Electronic Publishing Group,
Computer and Mediaservice /
University Library,
Unter den Linden 6, D-10099
Berlin, Germany
email: dobraz@cms.hu-
berlin.de
phone: +49-30-2093-7070 / 3274
fax: +49-30-2093-2959

The Bavarian State Library in
Munich (Bayerische Staats-
bibliothek)
Dirk Scholz
Ludwigstraße 16, D-80539
München, Germany
email: dscholz@bib-bvb.de
phone: +49 89 28638 2447
fax: +49 89 28638 2672

The Institute for Museum
Information in Berlin (Institut
für Museumskunde)
Monika Hagedorn-Saupe
Staatliche Museen zu Berlin –
Preußischer Kulturbesitz
In der Halde 1, D-14195 Berlin,
Germany
email: m.hagedorn@smb.spk-
berlin.de
phone: +49 30 8301 460
fax: +49 30 8301 504

General Directorate of the
Bavarian State Archives (GDAB)
Dr. Karl-Ernst Lupprian
Generaldirektion der
Staatlichen Archive Bayerns
Schönfeldstr. 5, D-80539
München, Germany
email: iuk@gda.bayern.de
phone. +49 089 28638-2484
fax: +49 089 28638-2615

Nestor will be presented at the European Conference on Digital Libraries (ECDL 2004) at the University of Bath, UK, (<http://www.ecdl2004.org/>) September 12–17, 2004.

References

- [1] "Arbeitskreis Infrastrukturen für Digitale Bibliotheken": Globale Standards, Integration und Kooperation – Konzeption und Gestaltung einer Infrastruktur für elektronische Fachinformationssysteme – Arbeitsbericht des Arbeitskreises Infrastrukturen für Digitale Bibliotheken" eingerichtet und betreut durch dl-konzepte / bmb+f Projekt im Digital Library-Forum, 2001, <http://www.dl-forum.de/engl/foerderung/projekte/digitallibraryplanungsprojekt/ergebnissarbeitskreis.pdf>.
- [2] Dobratz, Susanne; Liegmann, Hans; Tappenbeck, Inka: "Langzeitarchivierung digitaler Dokumente." In: *Zeitschrift für Bibliothekswesen und Bibliographie* 48 (2001) 6, S. 327–332.
- [3] Dobratz, Susanne; Tappenbeck, Inka: "Thesen zur Zukunft der digitalen Langzeitarchivierung" in Deutschland, *Bibliothek. Forschung und Praxis*. 26 (2002) No. 3, p. 257–261. ISSN: 341-4183, http://www.bibliothek-saur.de/2002_3/257-261.pdf.
- [4] *Trusted Digital Repositories: Attributes and Responsibilities*, an RLG-OCLC Report, <http://www.rlg.org/longterm/repositories.pdf>.
- [5] *Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1, *Blue Book*, January 2002, <http://www.ccsds.org/docu/>

dscgi/ds.py/Get/File-143/650x0b1.pdf.

Endnotes

¹ BMBF: Bundesministerium für Bildung und Forschung, <http://www.bmbf.de>.

² *nestor* – Kompetenznetzwerk Langzeitarchivierung, <http://www.langzeitarchivierung.de>.

³ Bau eines virtuellen Archivs kann beginnen, <http://deutschland.dasvonmorgen.de/press/870.php>.

⁴ Die Deutsche Bibliothek, <http://www.ddb.de>.

⁵ Niedersächsische Staats- und Universitätsbibliothek Göttingen, <http://www.sub.uni-goettingen.de>.

⁶ Humboldt-Universität zu Berlin, <http://www.hu-berlin.de/>. Electronic Publishing Group at http://www.edoc.hu-berlin.de/e_projekte_en.

⁷ Benders – Bücher & Medien, <http://www.bsb.de>.

⁸ Institut für Museumskunde, <http://www.smb.spk-berlin.de/ifm/>.

⁹ Die staatlichen Archive in Bayern, <http://www.gda.bayern.de>.

¹⁰ Digital Library Forum, <http://www.dl-forum.de/engl/foerderung/projekte/digitallibraryplanungsprojekt/>.

¹¹ Bericht vom Workshop, 29–30 Oktober und 14 November 2002, <http://www.langzeitarchivierung.de/modules.php?op=modload&name=News&file=article&sid=8>.

¹² In order to express the criteria see [4] for more details.

¹³ Digital Preservation Coalition website, <http://www.dpconline.org>.

¹⁴ PADI (Preserving Access to Digital Information), <http://www.nla.gov.au/padi/index.html>.