

## **HOW TO GET VALUABLE CONTENT FOR YOUR DIGITAL ARCHIVE? ESTABLISHING COLLABORATIONS AND PARTNERSHIPS WITH RESEARCHERS AND PUBLISHERS**

KŘÍŽ, Filip, MAIXNEROVÁ, Lenka, HORSÁK, Ondřej, BOUZKOVÁ, Helena

National Medical Library, Prague, Czech Republic (kriz@nlk.cz)

### **INTRODUCTION:**

Acquisition, processing, preservation and accessibility of traditional printed documents in the library environment are well established. Libraries, however, must deal with the threat of loss of documents due to the degradation of acidic paper and to the impairment by frequent use. Effective mean for preserving endangered publications is their digitization and their subsequent disclosure in digital libraries, archives or repositories.

In connection with the development of information and communication technologies libraries must address acquisition, processing, storing and making available new types of documents created already in the electronic form and in particular the problems of long-term preservation and access to digital documents.

The necessity for long-term archiving of the cultural heritage and scientific information in both traditional and electronic documents are formulated in the Commission of European Communities initiative i2010: Digital Libraries Initiative. For libraries in the Czech Republic is intended the *Conception of permanent preservation of traditional library collections and electronic documents in Czech libraries*, which foresees the creation of specialized digital libraries and archives in connection with the *National Digital Library* system [1].

### **METHODS:**

The National Medical Library of the Czech Republic (NML) has prepared a project to build a digital archive for the permanent preservation and access to both traditional and electronic professional medical information. NML initiated cooperation with Czech publishers of professional biomedical journals, which focuses on the possibility of obtaining and disclosure of electronic documents in the digital archive. Equally important is to acquire contents for the archive directly from the authors of scientific information and encourage the use of the archive for the storage and access to research reports, conference materials, presentations, preprints, teaching materials and other materials with professional information.

#### ***Cooperation with publishers of professional medical journals***

The contents of the archive is not restricted to documents resulting from the digitization process only, but the archive can be used to store documents already created in digital form. At present, most of professional published journals have its electronic versions and some are already published only electronically. Permanent and long-term preservation and access to these electronic versions becomes an increasingly important task.

NML has contacted Czech publishers of professional biomedical journals that might be interested in the service of long-term archiving in the digital archive, including the subsequent linking with bibliographic database Bibliographia medica Czechoslovakia (BMC). It is already clear that Czech publishers have strong interest in this service, primarily because they do not have resources to solve the long-term preservation of their electronic publications.

BMC is the national bibliography registering all the professional literature in the fields of biomedicine and health published from 1947 in the Czech Republic and also works by Czech authors issued abroad. BMC is accessible online in the Medvik portal and is also issued on DVD Bibliomedica by private company.

Currently the URL of an electronic version of an article leading to publisher website is stored in the bibliographic record so users can get directly from BMC to full text if freely available. But there are many disadvantages. The publishers may change URL addresses which result in broken links. For some journals it is not possible to refer to a particular article, because the articles do not have their own URL. There might be time lag before the electronic version is available. A possible solution is to archive electronic versions of journals in the digital archive, which would ensure that user would get to the full text, even in the event of publisher disappearance or possible loss of electronic data.

Publishers, who are interested in this service, enter into a contract with NML for the provision of electronic resources. After mutual agreement, the full texts stored in the archive can be accessible to the public or only to NML registered users. There may also be set up a time delay for an issue to be accessible after the actual time of publication according to the publisher.

### ***Cooperation with the authors of professional health information***

Equally important sources of content for the archive are other outputs of scientific and educational activities: the reports of Internal Grant Agency of Czech Ministry of Health (IGA), presentations, conference materials, posters, teaching materials, etc. In these cases, it suffices if the authors conclude with the NML license agreement. NML has contacted several authors of the IGA final grant reports which are of great interest of users whether they would be willing to sign the license agreement with NML. Most of the authors have agreed. At present it is already signed +20 license agreements. In the future, NML will try to negotiate with IGA that authors can sign the agreement at the time of submission of the final report so both printed and electronic versions will be delivered to NML together.

In near future NML plans to use Czech version of Creative Commons licenses (CC) which have been officially published in April 2009 [3]. CC is a U.S. non-profit organization whose goal is the lawful use and sharing of copyright works. CC understands that there are people who do not want to use all the intellectual property rights, which guarantee them the law. CC offers different licensing schemes copyright holders can choose specifying exactly what of the copyright for the work they want to keep and what in contrary they surrender to the public.

### **RESULTS:**

NML launched the archive in 2008. The archive system has been implemented in Czech open source software Kramerius [2] produced by Qbizm Technologies developed in collaboration with National Library of the Czech Republic, Czech Academy of Sciences Library and the Moravian Library in Brno. Development of Kramerius has been financially supported by Czech Ministry of Culture and Ministry of Education, Youth and Sports.

Kramerius currently supports periodicals and monographs only with plans to endorse other types of documents in the future. It supports DjVu, JPG, PNG, PDF image files and metadata in XML format. The system uses full-text search engine Lucene and allows replication of data between its different instances.

Public interface of the archive is accessible to users at <http://www.medvik.cz/kramerius>.

Archive currently contains complete edition of 5 journals and 17 monographs. The poor physical state of these documents already did not allow their further use.

All 5 titles of periodicals are subject to the protection period of the copyright law, thus it is only possible to view the full texts on devices connected to NML's computer network and only the metadata are freely accessible. The digitization was partially covered by the subsidy program of Czech Ministry of Culture. The information that electronic version of a document is accessible in the archive is displayed in the Medvik portal catalogue with the direct linking to the archive.

The content of the archive can be further expanded to include digital copies of documents from other Kramerius instances under the terms that these documents are or were owned (loss, destruction, etc.) by NML. This way there has been replicated 17 digitized monographs from National Library of the Czech Republic in February 2009. These older monographs are not subject to the protection period of copyright law, the disclosure of the DA is therefore without any restrictions. This year, NML is seeking additional financial support for the digitization of The Medical Newspapers from 1952-2003.

### **DISCUSSION:**

Building a digital archive is not only the selection and implementation of the archiving system, but it is necessary to ensure long-term preservation of digital data, which the system itself does not address. ISO standard for digital storage is a reference model Open Archival Information System (OAIS) [4]. This model offers a complete functional description of a digital archive of information and formulate commitments, which must meet the organization to work as an archive "compatible with the OAIS. In the OAIS model, the digital archive should meet the following minimum requirements: authenticity, integrity, reliability and availability, possibility of re-use. The model does not cater only to the protection of digital data, but also the protection of other related elements such as metadata, identifiers. Kramerius system used in NML does not currently implements the OAIS model but in the next version the system will be based on Fedora repository [5]. The new version will thus allow to archive more different digital objects in conformance with the OAIS model.

### **CONCLUSIONS:**

Digitization, archiving and accessibility of digital documents in NML are fully in line with the National Digital Library project of National Library of the Czech Republic under the National Strategy for Digitization. This project aims to accelerate the digitization and long-term conservation of digital objects and access to them.

Building a digital archive in NML is rather complex and continuous endeavour. It is necessary to ensure not only financial and personal resources for operation of the archive, but also to choose the methods of protection, archiving and accessing data, metadata, licensing models, proper software and hardware, and to develop many operating procedures. The OAIS standard to the NML's digital archive is also one of the conditions to become trusted archive and its putting into service should not be delayed.

However from first responses of our users, we assume that building the archive is good and important decision that will facilitate the communication of scientific information.

Preservation and access to professional health information in the digital archive is an important contribution to the development of science and education in the Czech Republic.

### **REFERENCES:**

1. Project "National Digital Library". Available from: [http://www.ndk.cz/project/view?set\\_language=en](http://www.ndk.cz/project/view?set_language=en)
2. Kramerius. Available from: <http://kramerius.qbizm.cz/>
3. Creative Commons. Available from: <http://creativecommons.org/license/?lang=cs>
4. Reference Model for an Open Archival Information System (OAIS). Available from: <http://public.ccsds.org/publications/archive/650x0b1.pdf>
5. About Fedora Commons. Available from: <http://www.fedora-commons.org/about>