TomášFoltýn, AnnaVandasová
National Library of the Czech Republic

THE VIRTUAL DEPOSIT LIBRARY – INTRODUCTION OF THE PROJECT AND THE FIRST RESULTS

Abstract: The aim of the project Virtual Deposit Library is to simplify and increase the efficiency of the preservation book collections refilling, which are stored in the Czech libraries. The main goal is to create a special application which enables(based on the harvested data from library catalogues) to evaluate the lists of discarded books, that are offered by the public libraries. To increase the value of Bohemical book collections three certified methodologies for building of preservatives collections will be also developed. The project takes into account not only the presence of the single specimens in the collections but also some attributes of their physical condition, which are continuously examined through the extensive research of book collections. All the activities funded by the Virtual Deposit Library project will enrich preservation collections greatly and support the cultural heritage stored in the Czech libraries.

Key words: deposit library, preservation collections, legal deposit, enrichment of the book collections

1. Introduction of the project

The Virtual Deposit Library is an acronym for the project „*Building the cooperative system for the creation and management of modern preservation book´s collections in the Czech Republic and the development of needed tools*", which is funded by financial mechanism called NAKI[1] organized by the Ministry of Culture of the Czech Republic. The project schedule is divided into two periods. The main goal of the first of them (2012–2013) is to create the beta version of application for preservation collections refilling, which are stored in the Czech libraries; and propose basic methodological tools that will simplify building and management of these collections. The second period (2014–2015) is aimed to optimize developed application and verify proposed methodologies in practice and possibly renovate them at the end of the project corresponding with achieved results. The head of the project is the National Library of the Czech Republic. Because of wide range of this problematic, researchers from the Moravian Library in Brno and the Research Library in Olomouc are participating directly in this project. Technological partner of project is the private company INCAD.

The main issue initiating this project was the incompleteness of the preservation book collections, so far built on legal deposit basis. In these collections book heritage for future generations should be stored. But this premise seems not to be valid. Four preservation books collections are located in the Czech Republic - two copies are stored in the National Library of the Czech Republic, one in the Moravian Library in Brno and the last copy in the Research Library in Olomouc. Only National Archival Collection[2] of the National Library has true

---

[1] Please see http://www.mkcr.cz/cz...na-rok-2012-90646/.
[2] Please see http://conspectus.nkp.cz/nacchar.phtml.

archival character: It is used for reading only in exceptional cases, when no other copy of the requested document is available in all Czech public libraries. The second collection called Universal Collection is used for the public access to the documents, especially in in-house mode. Similar way of using has two collections situated in Brno and Olomouc. That means that preservation books collections are relatively complete for the last twenty years. On the other hand a high number of books coming from 19th century and from first half of 20th century are missing unfortunately.

It exists some ways how to refill these collections. The most common way are lists of books offered by libraries disposing their holdings to other public libraries. Unfortunately the level of usability is not so high, because the information is shared in many different ways e. g. html files, xls sheets, docx documents or xml outputs from catalogues. The second problem is the incredible high amount of records of offered documents. The most of these lists usually contains hundreds of books. In the last three months of 2013 86 lists of eliminated books were shared through electronic conference of acquisition[3] for example. The mentioned lists contain thousands of documents. That means - it is impossible for anyone to check the presence of all the titles of the lists in the content of the own collection. This leaves only two options. Take everything, or nothing. If the second way shall be used, many volumes that are often not represented in the preservation collections will be lost for good.

The key benefit of the project is to prevent the losses, increase the efficiency of refilling preservation collections from lists of eliminated books and try to simplify the process of acquisition of the missing books by the automation of the processes using the software solutions.

2. Application Virtual deposit library

The key milestone of a successful project is to build an application so far called Virtual Deposit Library, which should play an important function within the refilling of conservation book collections and the subsequent creation of physical deposit library in future. Application should permit direct enrichment of preservation collections stored in Czech libraries –in the first phase especially collections of the main project partners. Basic condition of the automated replenishment is to precise mapping the number of copies stored in libraries to the information about their holdings. That is very difficult because of various condition and number of records in library catalogs. The research team had to work on the definition of statuses harvested to the Virtual deposit library, their mapping to a single profile and the subsequent unification of these profiles. Concurrently, detailed algorithms to compare the results from individual harvests and sort them were established. Without the proper usage of the unique identificators so as Czech National Bibliographic Number it will not be possible. Single harvests must occur in regular cycles to ensure the timeliness up-to-dateness of the database content.

Another important aspect of the application is its proper functionality, the ability to divide and organize records in the requested order and analyze the needs of individual libraries. At this time the local catalogues from the participated organizations – National Library of the Czech Republic, Moravian Library in Brno and Research Library in Olomouc are harvested. All the harvest procedures are based on OAI-PMH protocol. Of course, to get as much clear data as it is possible is essential to reach all the project goals. But unfortunately, the opposite reality is true one, especially in the case of information about library units. The huge mess of data coming from the three involved libraries, make this step of project

---

[3] Please see https://andrea.vc.cvut.cz/cgi-bin/mailman/listinfo/akvizice.

realization very hard. During some harvest even had happened, that in the harvested data various information about the number of propagation objects (e. g. umbrellas) were involved. For the promised project goals achievement was necessary to unify the data to uniform profile, check the duplicities and deduplicate them and finally adjust the new profiles suitable for the project aim. The research team have made a great work so far, but it seems, that this process is never-ending story. To avoid useless and time consuming human work is challenging. During the project realization one tool, developed from other purposes, was added to project framework. The tool called Prokop is composed from DB/FS working space and parallel special modules compounded of several JAVA scripts, SQL procedures, etc. It provides the requested output, where the results from every single harvest are reached and the eventual inconsistences are highlighted and recommended to repair in the library catalogue.

For these reasons, before the own development has begun, very detailed functional analysis very created and approved. It defines not only the different functional properties of application, but also single business models including their interconnections or detailed schemas describing the sequence of operations. Designed application must contain automated functions to evaluate the number of copies held by every single participated library, with the accent to the libraries with the preservation duties, which will be standing in top of the wish list. When collections are refilling from delivered offers, National Archival Collection of the National Library, as the most important conservation collection in the country, is preferred automatically, followed by another three collections mentioned at the beginning. Preserve collections of other public libraries are standing at the rear of the hierarchy, behind the sources for the future creation of the physical deposit library, which is designed to complement the lost or otherwise destroyed copies. In the physical deposit library currently counts with the insertion of five copies for further use. Books, which will not cover any of above mentioned criterions, are determined to the process of destructive digitization, what guarantees maximum extraction of acquired resources. The offer agenda must also determine, how many copies of requested document institutions should require through the application not to lose the copies in the update system. Finally, a special warning mechanism for appeared documents standing in the absent copies wish list, will be established.

The agenda of demanded Bohemical documents is standing complementary to the mentioned offer agenda. By the above mentioned models the application will be able to simplify the process of documents transferring between the institutions, which have to offer their unwanted collections to other public libraries and on the other hand for the individual conservation´ collections curators to enrich their funds in the electronic form including automatically generated protocols, contact details etc.

All defined agendas can be divided into separate views. Users will be able to save and use them after the authorization to the system. After the initial logon the responsible person will get the list of user views and will be able to manage, edit or share them. An authorized user will be able to evaluate incoming records from other libraries on the basis of the documents uniqueness, possibility of integration to his collections, real state of collections or other predefined attributes. A closer connection to the application for physical state of paper documents measurement (see the next text) is assumed in the following years.That will bring more options how to compile the lists of requested books assigned to the different libraries, based on their level of degradation.
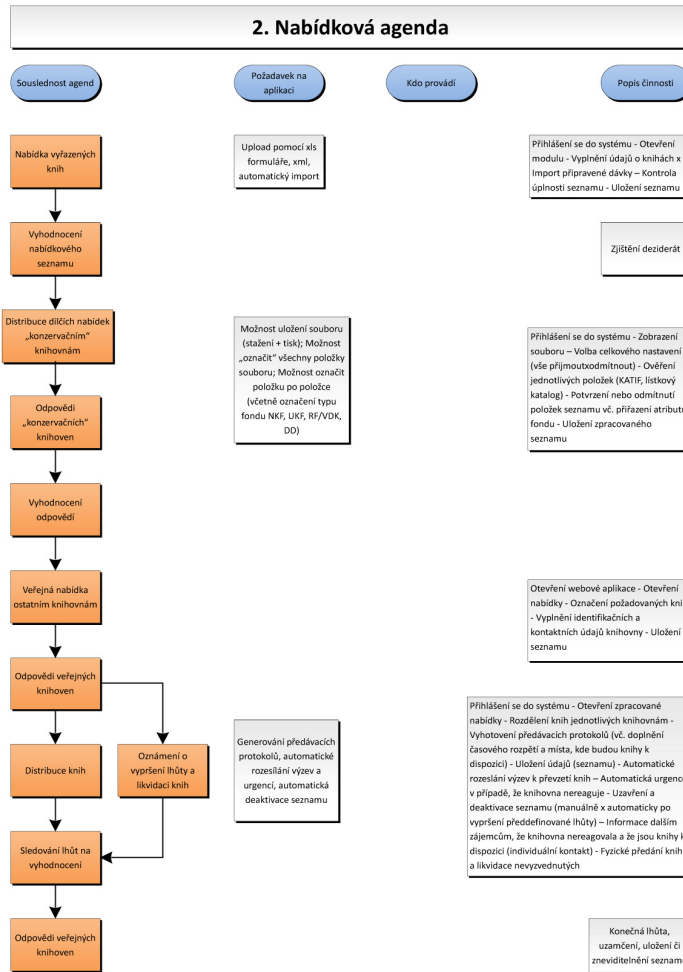
Figure1: Business model of the agenda for books offering

Figure 2: Documents coming from the HorníJiřetín Library offer, which are not available in the National Library of the Czech Republic collections

Virtual Deposit Library application development builds on previous research activities of the National Library of the Czech Republic, especially on the tools and unique identification established in the past. Operations connected with large scale data operations are simplified by the Czech National Bibliography number[4], which prescribes unique identifier of a physical document. The system architecture includes a database system Oracle, application framework Relief IV[5] and search tool MS FAST Search for Internet Sites[6]. These components are long-term used and verified in the environment of the National Library of the Czech Republic. This leads to the reduction of secondary costs connected with technical resources and their management, which saves unexpected amount of money.

Software solution of the Virtual Deposit Library application is composed of two parts – database of the application and search interface. Bibliographic records from the catalogues of libraries are collected in application, as well as the data about users of the system, participating organizations and other structured information about the processes used by the individual agendas. There are special modules for the administration of every single harvest included. On the other hand search interface allows to search and to analyze the status of records across the catalogues of participating organizations.

3.   Other outputs of the project

Together with the fundamental development of the application Virtual Deposit Library other activities necessary for continuous building of archival collections are solved within the project.

3.1. Basic processing of Bohemical Units coming from various library collections. The basic premise of this activity is to process as many Bohemical units from archival (or reserve) collections of participating libraries as it is possible, because often happens, that there is no precise property evidence about them in the library catalogues. Mapping of the Bohemian documents stored in the so called Reserve collections of the National Library is crucial. This not very well explored collection is a rich source of documents for the process of refilling more important collection of the National Library of the Czech Republic, including Archival collection of 19th century documents and the National Archival Collection. During the thoroughgoing processing of the documents more than 600 substitutes for the National Archival Collection has been discovered so far, which were lost in the past. It has been found more than 1,300 titles that were not represented in the National Archival Collection parallel. Till now more than 250 000 units have been processed in the framework of the project. Furthermore, there are also collections of the Research Library in Olomouc, which have not been mapped and do not have such records in an electronic catalog, although they have archival character.

3.2. Mathematical formula and methodology for recalculating of the optimal amount of copies stored in the system of libraries having preservation duties. One of the most important factors for the long-term preservation of the traditional paper documents is a number of copies that must be stored in the system of libraries having preservation duties. Because of the true warranty of the original documents´ long term preservation, which is not

---

[4] Please see http://www.registrdigitalizace.cz/rdcz/uzivatele/texty/rd_jak_ziskat_cCNB.pdf.

[5] Please see https://code.google.com/p/relief/.

[6] Detailed information about MS FAST
http://en.wikipedia.org/wiki/User:Janhoy/FAST_Enterprise_Search_Platform_%28ESP%29.

guaranteed by law in other culture heritage institution, it is not possible to add those copies to any methodically build formula.

Currently, it is possible to use mathematical methods to determine the optimal number of copies that guarantee preservation of one or a few copies within a time period. To calculate the recommended number of copies methodology that guarantees a higher degree of objectivity was developed. Gaining the right number copies recommended for the preservation in predefined time line may affect the general requirements for the storage capacity of the archival collections, which is hardly impossible now because of the current state of knowledge.

The formula, that is used to calculate above mentioned results, works with several variables that were defined at the beginning of the project and are specified continuously. Elemental variables are - natural rate paper degradation, rate of degradation of other material use e. g. in book binding, number of annual losses, the frequency of borrowing, type of damage, the use of copies (microfilm, digital copy), etc. Variables that can influence the possible reduction of archival copies are taken into account only if their value exceeds a certain limit. However, the researcher often miss the data that are necessary to quantify these variables or are not obviously observed e. g. the relation between the number of loans and the number of annual losses of documents. Hopefully these values will be specified in the near future and it will be possible to optimize the use of the formula for calculating the recommended number of copies. The mathematical formula is designed flexibly, so it can hold not only the variables that influence the preservation of documents in the collection, but also set the target time period for long term preservation, so as the survival probability. After the process of input data verification and subsequent outputs evaluation, it will be possible to determine the number of copies which should be kept inside of the system of deposit libraries. Because the number of obligatory copies should be fixed by law, the determined number of copies will be reflected in the Virtual Deposit Library initially and after that in physical deposit storing.

3.3. Systematic surveys of modern Bohemical collections. Systematic surveys of modern Bohemical collections take place in the National Library of the Czech Republic continuously since 2010, when first launched continuous surveys were funded by institutional support. Individual measurements are recorded to the application known as the "Central Knowledge Base of Examined Documents". This application is used exclusively for monitoring the physical condition of library collections, although is created on the ground of the other tool used in the Czech Republic – Digitization Registry.[7] During the measurements different parameters affecting the physical state of the document are considered. Different types of damages are detected, after that the experts from the Original Documents Conservation Division diagnose their causes and propose possible solutions (reparation, conservation or restoration works, storing in protective packages, deacidification, reformatting etc.). Currently, the database contains more than 160,000 records. Given their importance the researchers are trying to include this application in the software environment of the National Library of the Czech Republic and get some continuous support from Ministry of Culture, which would ensure the full functioning of the application in the future, the clear policy of the data backup and expand its applicability to other Czech libraries.

To obtain key information to the database in the same level of quality, well-defined terminology and uniform structure of issues, new very deep methodology defining the optimal

---

[7] Please see http://www.registrdigitalizace.cz/rdcz/?language=en.

working procedures of these surveys were created and approved by responsible public bodies. The methodology gives to the libraries with conservation duties very detailed information about the best physical conditions in repositories, possible conservation interventions as deacidification or restoration, which can be simply used in the negotiations with their funders about the institutional budgets connected with original documents management. Very important aspect of the methodology is also the actual picture about the ways of storing and state of physical art of the single documents holding in the collections of the biggest Czech libraries (besides National Library of the Czech Republic also from Moravian Library in Brno and Research Library in Olomouc), which provides very interesting feedback to collections´ curators about the different approaches to the curation in past. It enables also broader decision making procedures on the basis of the cooperation of all Czech libraries.

        Besides from these main activities the partner institutions are also engaged in other joint processes that aim to contribute to the long term and adequately preservation collections e. g. a single components analysis, reporting the libraries that belonged to the recipients of legal deposit in the past or mapping its use in specific institutions and disseminating the project itself.

Figure3: Application for survey of collections

4. Conclusion

The activities in the project NAKI "*Building the cooperative system for the creation and management of modern preservation book´s collections in the Czech Republic and the development of needed tools*" are in the full accordance to the Policy of the permanent preservation of the traditional text documents proclaimed by the Ministry of Culture. Both of these activities are designed to minimize damages on the cultural heritage, for effective protection of the printed literary treasures and for better management of the preservation collections storing Bohemian documents. These activities are characteristic to the libraries.

They define their status in a wide range of memory institutions. However encounter different barriers. Unfortunately some of them are hardly influenced, so as the limits associated with costs, which are essential for continuous support. But of course, there are some possible steps realized now. The core activities of the funding mechanism VISK 7 open for all Czech public libraries were extended about the possibility of modern collection deacidification or for the financial support for conservation or restoration treatments.

Some barriers are not associated only with funding unfortunately, but also with some professional activities realized by libraries, which are not solved optimally. In connection with the above mentioned Policy of the permanent preservation of the traditional text documents and realized NAKI project, the inconsistent applications of persistent identifiers (Czech National Bibliography number application or URN:NBN resolver) cause many troubles connected with records merging, which can establish many problems in the next logical step – physical Deposit Library building and the broader network of Czech libraries.