



národní  
úložiště  
šedé  
literatury

## **New digitization workflow of the National Technical Library in theory and practice**

Řihák, Jakub; Kamrádková, Kateřina  
2012

Dostupný z <http://www.nusl.cz/ntk/nusl-112973>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 30.09.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní [nusl.cz](http://www.nusl.cz) .

# **New digitization workflow of the National Technical Library in theory and practice**

**Jakub Řihák** ([jakub.rihak@techlib.cz](mailto:jakub.rihak@techlib.cz))

National Technical Library, Prague, Czech Republic

**Kateřina Kamrádková** ([katerina.kamradkova@techlib.cz](mailto:katerina.kamradkova@techlib.cz))

National Technical Library, Prague, Czech Republic

**Keywords:** digitization, digitization workflow, OCR processing, digital library

## **Abstract**

This paper describes recent activities in the area of digitization processes at the National Technical Library, Prague (Czech Republic), in particular a new schema of digitization workflow, its outputs and its impact on services provided to the National Technical Library's users. Furthermore, it presents System Kramerius and cooperation opportunities in eBooks on Demand project for the Czech libraries.

One of the significant activities of the National Technical Library (NTL) is digitization of documents and their high quality OCR (Optical Character Recognition) processing. There is an effort to provide an access to a variety of digitized documents in the best quality to NTL's library users. Therefore, all processes within the NTL digitization workflow had to be optimized and improved.

NTL focused on digitization of university textbooks from technical universities in previous years. These textbooks are a considerable part of NTL's library collection. NTL's priority was to digitize the most frequently borrowed university textbooks and new published ones. For this purpose, NTL designated a working place, equipped it with a document scanner and various software for image and document processing (i.e. Capture Perfect 3.0, Adobe Acrobat Professional, Abbyy FineReader 10 for OCR processing, etc.).

At the same time, NTL began an active participation in eBooks on Demand - A European Library Network (EOD) project. NTL is one of the four libraries in the Czech Republic that are involved in the project. Even though other Czech libraries interested in EOD partnership mostly have rare books collections, they do not have necessary resources to join the EOD network (for financial reasons, lack of staff or HW/SW). Thus, NTL offers a cooperation opportunity to these libraries to join the EOD project. Hence, such libraries can also offer their rare books as e-books. These e-books can be published in NTL's digital library Kramerius – online and for free.

NTL began to upgrade and automate the whole digitization process in 2011. Until then the digitization process depended on human labour and it was time-consuming. To overcome

these problems, the OCR process automation was needed. NTL bought a license for Abbyy Recognition Server 3.0 software and began to implement it to the digitization workflow. This software allows NTL to automate OCR processing in the ways that were not possible with Abbyy FineReader 10. Recognition Server was set up on a virtual server, on which Management Console is running. There are six other working stations connected to the server at the time (4 CPUs, each operates with 3,40 GHz frequency, 8GB RAM). Each processing station has one CPU designated for the OCR processing of the documents and is also used as a working station for the library employees during the week. Thanks to Recognition Server implementation, it was possible to decrease the time needed for OCR processing to  $\frac{1}{4}$  of the original time. Consecutive OCR outputs (text files) verification can be executed simultaneously with OCR without waiting for the whole process to be finished. Verification can be done on any computer in the same network which is connected to the Manager Console on the server.

The next step was to upgrade the NTL's digital library Kramerius. System Kramerius is used as a main access point to digital documents in the the National Technical Library as well as in various Czech libraries. It is based on Fedora core (which serves as a document repository), SOLR search platform, and Java-based interface. Digitized documents under the copyright published in Kramerius can be accessed after authentication "in-house", while public domain documents can be accessed externally without restrictions. Further activities will focus on promoting the digital library services.

In connection with previously described activities, the need for significant changes in NTL's digitization workflow emerged. New workflow was supposed to be simple, understandable and helpful so it could make all processes faster and in a better quality. All that had to be done with less staff, due to the budget cuts for this year.

## **Introduction**

Digital documents are taking more significant role in our daily life in recent years. Books, journals, newspapers and other traditional printed media are being published (or created) in a digital form. Libraries are also trying to respond to this trend. Digitization has been a part of library world for some time, but at first it was used merely to preserve precious or damaged library collections for next generations of readers. Today libraries are also trying to create new services or enhance the old ones by using and creating digital documents. The the National Technical Library (NTL) in Prague, Czech Republic, created digital library to preserve its special historical collections and provide them to public. It also can serve students of the technical universities, by providing them digitized textbooks. Because of recent activities in the field of digitization NTL had to enhance and restructure the digitization workflow and also upgrade software and hardware used in digitization process.

This paper is divided into following sections:

- 1) Previous state of digitization in the National Technical Library**
- 2) New digitization workflow and policy**
- 3) Conclusion and future work**

## **NTL and digitization – previous state**

The National Technical Library (previously the State Technical Library) began with digitization of documents in the year 1998 and in following years focused on digitizing its historical collections endangered by paper deterioration. The whole digitization process was small scaled. Only documents in imminent danger or documents which were unique in the Czech Republic were converted into digital form. All the work was outsourced as the State Technical Library had no technical means to digitize those documents in the quality needed for successful preservation of these documents. At first, documents were digitized and copied to the CD-ROM or other portable media and then provided to the public on special computers in the house. These documents created a base for future Digital National Technical Library. The digitization process was therefore primarily focused only on preservation of library collections.

In the following years in the new millennium, the demand for digital media grown larger and libraries in the Czech Republic tried to respond to that fact. The National Library of the Czech Republic, in cooperation with the Library of Academy of Science of the Czech Republic participated in a system creation project for publishing the digitized documents in larger scale and via the Internet. From that project a system called Kramerius originated and since then it is used in many libraries in the Czech Republic (the State Technical Library included) to provide the digital documents to the public. To monitor and organize the documents digitized by various libraries, the Register of digitization was created. This tool was a necessity to avoid duplicity in digitalization and is still in use.

The State Technical Library wanted to maintain the digitization of its historical collections (though still in a small scale and primarily in the means of documents preservation). It was decided to focus also on a digitization of university textbooks, as majority of library customers are students of the technical universities. University textbooks are a vital part of NTL's collections and are one of the most borrowed types of documents. Concerning textbooks, students often have to wait a whole month or more to get their desired book. Therefore it library focused on digitization of mostly borrowed textbooks from our collections and to provide them to our customers via the digital library Kramerius. Based on the loans statistics the list of textbooks was created for the potential digitization. More than ten thousand textbooks should have been digitized. The State Technical Library had to create one special working place for this project and equip it with appropriate hardware and software tools.

This new working place was equipped with a document scanner Canon DR-5010. Scanning software Capture Perfect 3.0 was supplied with the scanner. For quality OCR (Optical Character Recognition) processing, software Abby FineReader was bought. It was decided that the image formats used for digitized documents will be TIFF and PDF, both in grayscale and with resolution 300 DPI. This special working place had only one agenda. It

had to collect, digitize and publish university textbooks and cooperate with other institutions in the field of digitization.

With more digitization projects underway, this independence brought following issues:

- Different image formats used for publishing
  - PDF for university textbooks
  - JPEG for EOD e-books and other documents from historical collection
- Different image quality
- Different naming convention
- Different storage place for archived documents

Despite these differences, all scanned documents had to be published in the same digital library Kramerius, version 3. This generated more work, because the files had to be converted to the same format, they had to be renamed often and there were also problems in finding the documents in the file system because of different storage place.

Between years 2008 and 2011, total amount of 2941 textbooks were digitized, but only 862 of them are imported in the digital library. This was caused by time consuming processes in the digitization workflow, like image OCR processing or metadata creation. All processes in the digitization workflow of textbooks also had to be done by only one library employee.

Because of aforesaid reasons, library began with automation of digitization processes (at first with the automation of OCR processing) and also with updating the former digitization workflow or creating a new one.

## **New digitization workflow and policy**

In July 2011, it was clear that a current state of digitization in the National Technical Library is not sustainable. The amount of documents to be digitized was too big and processes were too slow to maintain them successfully. In that time, NTL also updated its digital library to newer version, Kramerius 4. The main goal was to automate and therefore speed up the OCR processing. NTL chose to buy a licence for Abbyy Recognition Server 3.0 and at the same time began with updating the digitization workflow to meet new demands in that field.

A new digitization workflow had to be as much comprehensive and simple as possible, because it was planned to cooperate more with other library departments. Another goal of these updates was to standardize all processes, unite the image formats used for archiving and publishing of the digitized documents. Special attention was given to standardizing the naming convention of digitized documents with respect to future workflow automation.

Together with changes in a digitization workflow it was decided to reconsider the digitization policy of the National Technical Library. Current practice of digitization of

university textbooks with certain amount of loans was changed. The National Technical Library started digitizing only newly bought textbooks. This change also reduced the number of documents to be digitized and therefore saved more time for other tasks.

The membership in the [eBooks on Demand – A European Library Network \(EOD\)](#) project must be taken into account in the NTL's digitization policy. A digitization service of public domain books on demand is offered within the EOD project. Four libraries in the Czech Republic participate in the European Library Network, nowadays comprising of more than 30 European libraries. In 2010, one year after the project begun, a short survey has been made among other Czech libraries. The aim was to discover an interest in participation in the EOD project. The survey has showed, that interested libraries have a rare book collection, but mostly do not have necessary resources to join the EOD network (for financial reasons, lack of staff or HW/SW). NTL offered such libraries to join the EOD project via NTL – in order to offer customers a wider range of historical books for digitization and make specialized books available as e-books also to users who do not attend below mentioned libraries.

Currently, four libraries joined in the cooperation with NTL within EOD: the Arts and Theatre Institute Prague, the National Medical Library, the Military History Institute Prague and the Research Library Liberec. An occasional cooperation was made also with other libraries. All EOD e-books are now accessible in NTL's digital library online and for free.

While ordering an EOD digitization from cooperating library, the original of requested book is brought from the library to NTL and here is processed via the EOD service. The customer receives an e-book in PDF with full-text, the library receives the original book back, then PDF with full-text, metadata in XML or TXT format, full-text in RTF and images in TIFF format. The digitization of the book is paid by customer (4 CZK/page, the whole book is scanned in the EOD service, 200 CZK fee). The main advantage for the cooperating library is that it obtains books from its historical collection digitized for almost no costs.

Cooperating libraries has through NTL access to EOD promotional leaflets or posters, could implement the EOD service order button into their online catalogue or make their own dataset of public domain books in their library collection to harvest the records into the EOD search engine or to connect to Europeana.

Therefore, the National Technical Library wants to focus on digitization of historical collections within the scope of the EOD project and thematic digitization of historical collections, thus creating a solid document base for digital library that can be provided to public without restrictions of the copyright.

After two months of work the final version of new digitization workflow was created. The schema is shown in the *Figure 1*. Each section of this workflow is divided into more detailed description. Therefore it is possible to track all procedures within digitization

workflow together with other necessary information. This could be very helpful for employees involved in digitization as well as for supervisors and library management. It could also be used as a guideline for other libraries interested in digitization of printed documents.

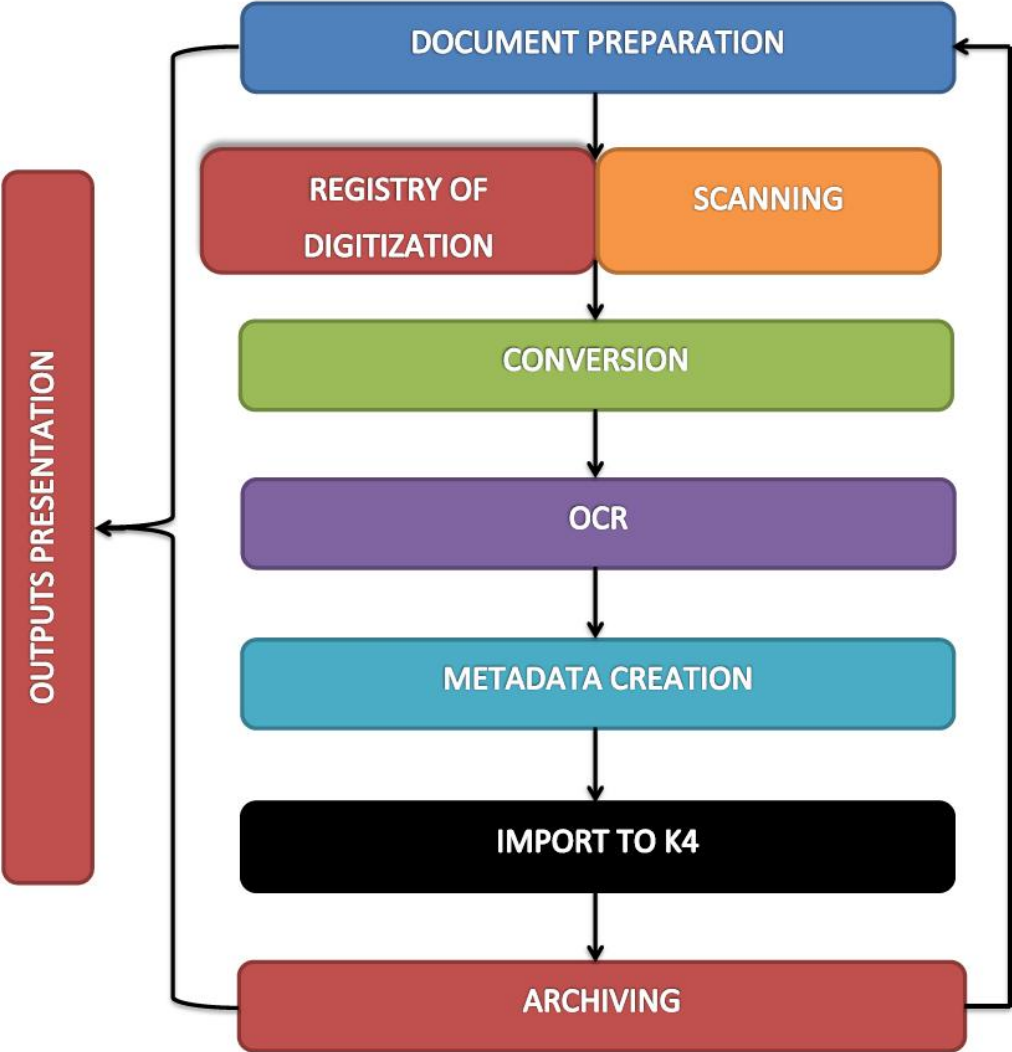


Figure 1: Digitization workflow general scheme



Main processes within digitization workflow are following:

- 1) Document selection and preparation
- 2) Registry of Digitalization update
- 3) Scanning
- 4) Conversion to other image formats
- 5) OCR processing
- 6) Metadata creation
- 7) Import to digital library Kramerius
- 8) Archiving
- 9) Outputs presentation

In the following part of this paper, these processes are described in more detail.

### **Document selection and preparation**

Documents selected for digitization are following:

- Newly bought university textbooks
- historical collection documents endangered by paper deterioration
- historical collection documents ordered for digitization through the EOD project
- historical collection documents covering an interesting topic, e.g. collection of historical maps of Austrian-Hungarian Empire

After the selection of documents to be digitized it is necessary to make sure, that these documents have their bibliographic records in a library catalogue. If not, it is compulsory to create it before advancing further in the digitization process. This was not common in the previous workflow.

The second part of this process is also to add identifiers based on which we can generate OAI-PMH sets of bibliographic records for every kind of digitized documents, i.e. historical collections, textbooks and books from the EOD project. These sets can be then used for updating the Registrar of Digitization. After necessary updates to bibliographic records, documents can go for scanning.

In the case of new university textbooks this process is slightly extended. For textbook scanning the traditional desktop document scanner is used, therefore all documents has to be cut (their binding has to be removed) before continuing with digitization. It was bought an automatic paper guillotine for that purpose, but still in some cases this work has to be outsourced.

## Update of Register of Digitization

With the OAI-PMH set generated it is possible to update relatively easily the Czech Register of Digitization. This Register is primarily used for monitoring the digitization projects in Czech Republic and therefore avoiding the duplicity in digitization. Currently, using the OAI-PMH for updating the register is probably the fastest way. With defined OAI-PMH sets it is also possible to easily and quickly generate datasets for metadata editor.

In the case of already scanned material, which still has no record in the Register of Digitization, this process has been moved to the end of the whole workflow, because a mass record creation or update is not possible, due to the difficulties in document identification. By doing so, it is possible to keep track of the changes and of the current state of all NTL's records in the Register.

## Scanning

The scanning process follows after the document selection and preparation, in parallel with updating of the Register of Digitization and creation of the datasets for metadata editor. The scanning process divides into three main branches, depending on the projects NTL is conducting or participating on. The branches are following:

- 1) Scanning of documents within the scope of the EOD project
- 2) Scanning of thematic historical collections (HC)
- 3) Scanning of university textbooks

The whole scanning process takes place in the Interlibrary Services Department. There is placed one scanning station for digitization of university textbooks and four stations equipped with book scanners for digitization of historical collections.

The scanning output image format is TIFF without compression. It is possible to convert image in TIFF format to any other image format, this is being done in the case of documents from historical collections. The only exceptions are university textbooks where it is possible to produce uncompressed TIFF file and JPEG used for publishing at once.

The resolution and colour depth of the images varies depending on the document type. Particular resolution and colour depth for given document type shows following table:

<i>document type</i>	<b>EOD document</b>	<b>other HC document</b>	<b>textbook</b>
<i>resolution</i>	300 DPI	400-600 DPI	300 DPI
<i>colour depth</i>	24 bit	24 bit	256 Level Grayscale

**Table 1: Resolution and colour depth used for output images in the scanning process**

Together with the format and image properties it was also needed to determine where and how the outputs will be stored. Since the National Technical Library has no Long-term preservation (LTP) system, all files are stored in the file system. Previously it was common to store output files on different places in the file system, depending on the project and department that digitized the original document. It was one of the main goals of new workflow to standardize and unite the storage place and naming convention of the output files to keep track of what has been done in scanning process or other following processes within the digitization workflow. Therefore only one “storage” folder was created with subfolders representing the current digitization projects.

Naming convention for output files has been standardized for every type of document. Folder names for digitized documents as well as file names consist of nine-place system number which is used for identifying bibliographic record of the given document in a library catalogue. Therefore bibliographic information about the document can be quickly retrieved by simply copying and pasting it into the catalogue search box. It is also possible to automate some processes which are using folder names and file names, because there are no problems with alphabet and special characters in the names. For university textbooks, the signature mark is also used in the folder and file names, divided from the system number by underscore character. This signature mark consists of one alphabetical letter and 3-6 digits. Because of these characteristics it is also a suitable identifier for automated processing. The last part of the file name is usually a file number within the subfolder. It consists of four digits and is also divided from the rest of the file name by underscore character.

## **Conversion**

The conversion process is only applicable on documents from library’s historical collections. It was decided to provide the documents to the public on Image server in JPEG2000 format. This format can be successfully used for viewing large images and streaming them over the internet. For this conversion, the Kakadu software is used. The specifications of converted output file were adopted from the specifications of the National Digital Library. (Hutař, 2012)

For partial automation of this process a Perl script has been made. This script can use command lines for creating the image with a given specification. These command lines were also adopted from the National Library of the Czech Republic and are available online at (Hutař, 2012) and (Vychodil, 2012). As parameters of this script, the folder name of digitized document is used, together with abbreviation of the desired quality of the output. The „mc“ abbreviation is used for „*master copy*“ (or „*Archival copy*“ in (Hutař, 2012)) and „pmc“ abbreviation is used for „*production master copy*“ ((Hutař, 2012) and (Vychodil, 2012)) Documents in this quality are used for publishing. This production master copy has a compression rate 1:8, but still has a quality comparable to original TIFF file (it is visually lossless). Since the NTL decided to use a TIFF format for archival images, master copy JPEG2000 is not used.

Perl script gets the folder name and „mc“ or „pmc“ as the first and second parameter, then scans the folder for any TIFF file and converts it to JPEG2000 (JP2) with the same name. This batch conversion minimizes the time needed to convert TIFF images to JP2. All JPEG2000 files are stored within subdirectory of the original image folder and uploaded to the Image server in the later process.

## OCR Processing

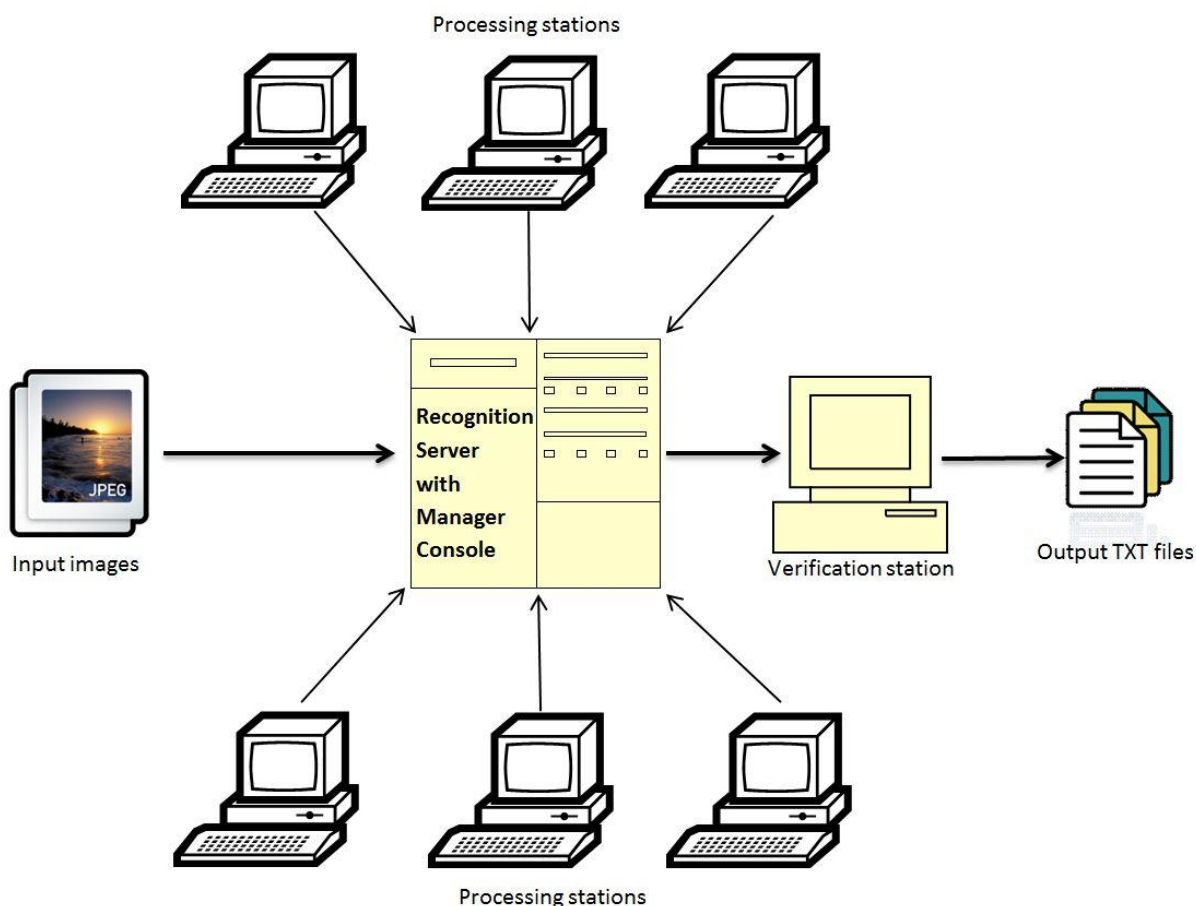
The OCR processing of the digitized documents is vital in the whole digitization process. Therefore the OCR has to be done in a best quality possible. The National Technical Library used Abbyy FineReader software, but the possibilities of automation of this process were very limited. In december 2011 it was decided to buy Abbyy Recognition Server 3.0 (Abbyy RS).

Abbyy RS is installed on a virtual MS Windows 2008 server. Then additional six computers were connected to the RS as a processing stations. Each station has 4 CPU's, each operates with 3,40 GB frequency, 8GB RAM and has MS Windows 7 OS installed. Each processing station provides one CPU (or more if needed) for the OCR processing when the OCR workflow is running and has a batch of documents to process. There might be connected more processing stations in the future. These stations are also used as working stations for the library employees. With the Recognition server implementation it was possible to decrease time needed for OCR processing to  $\frac{1}{4}$  of the original time. Processing time of 1000 pages is now 15 minutes instead of one hour.

All files which are going to be recognized with Abby RS are moved to the designated directory on a network disk. This directory is called „Hot Folder“. After the processing, the text file outputs are moved to the directory labled as output directory and from there they can be moved to the folder containing the original files. The whole OCR workflow can be set to start at a certain day of a week and certain time or it can be active permanently.

The verification of the outputs is an integral part of this process. It can only be done on one computer at the time, because of the licencing. Currently it is necessary to upgrade the licence for at least one more. The verification only takes place in the document pages, which have more than 10% uncertainly recognized characters on one page. This value can of course be adjusted. The verification station is also connected to the Recognition server and when the verification client is connected and active, the uncertainly recognized pages are automatically sent to this station for correction. Implementation schema of the Abbyy Recognition Server is shown in the *Figure 2*.

OCR outputs are plain text files, named according to original images and are stored in the same folder as the original images as well. This is necessary due to the metadata creation process.



**Figure 2: Implementation schema of the Abbyy Recognition Server 3.0 in the NTL**

## Metadata Creation

Metadata creation process is the next step of the digitization workflow. The National Technical Library creates bibliographic and structural metadata in the DTD for Monography and DTD for Periodicals format, which was used in the older version of digital library Kramerius 3. Bibliographic and structural metadata are stored in one XML file, which is created by metadata editor.

Metadata editor used for creating XML metadata files, is a product of the National Technical Library and it uses bibliographic data exported from the Aleph library system. These datasets for metadata editor are generated regularly every week. Therefore it is possible to keep datasets recent. In case that bibliographic metadata are not found it is possible to generate an XML structure containing only structural metadata and manually add the bibliographic metadata later in the process.

The National Technical Library is currently using a newer version of digital library called Krameris 4 which is using the Fedora Commons repository for storing the digital objects. These digital objects are stored in the FOXML 1.1 format<sup>1</sup>. In Kramerius 4 metadata are in various formats as shown in the *Table 2*.

<sup>1</sup> Detailed specification of this format can be found at <https://wiki.duraspace.org/display/FCR30/Introduction+to+FOXML>

<b>metadata types</b>	<b>standard (metadata format)</b>
descriptive metadata	MODS, Dublin Core
administrative metadata	PREMIS, MIX
technical metadata	PREMIS, MIX
structural metadata	METS
OCR files	ALTO XML, TXT

**Table 2: Kramerius 4 metadata formats (Hutař,2012)**

For successful import of the documents to the digital library it is therefore needed to convert metadata from older format to the new one. Kramerius 4 has its own convertor implemented, thus documents can be imported in the Kramerius 3 format and convertor is able to process these files and create digital objects in FOXML. This is only a temporary state, it is planned to implement a new metadata editor created for the Kramerius 4 digital library.

### **Import to Kramerius 4**

Kramerius 4 is developed by Incad company in cooperation with the Library of the Academy of Science of the Czech Republic, the National Library of the Czech Republic and the Moravian Library. Kramerius 4 uses Fedora Commons repository for storing digital objects in FOXML 1.1 format. There are also differences in metadata formats used for describing the documents. Fedora repository is connected with SOLR indexing tool and with PostgreSQL database. System has a user interface that serves as a presentation layer for documents stored in the repository and is available online.

To successfully import the document into the digital library, it is necessary to upload all data (images in JPEG format, TXT files with OCR output, XML file with bibliographic and structural metadata) to the conversion folder on the Kramerius 4 server. Conversion of the files and metadata is necessary because of the differences between older and new system and it is made automatically in the import process. When importing historical collections to the digital library (provided to the public in JP2 format), it is important to upload those JP2 files to the Image server.

Then, the system administrator can login to the administration of the system and begin with conversion and import of the documents. The whole process is automatic and its duration depends on the amount and the size of the imported document. After successful import to the Fedora repository, the document is indexed and made visible in the digital library. If the imported document is a university textbook, system administrator can add a permanent link to the document and special identifier to the bibliographic record. This identifier assigns the document to the designated collection within the library system. This functionality is used in the process of bibliographic records batch enhancement.

If the imported document is an EOD book or other historical document, FOXML files have to be downloaded from Kramerius 4 server. Then, RELS-EXT datastream in the FOXML file has to be edited to link to the images in JP2 format on the Image server. Edited FOXML files are then imported to the Fedora repository. After the successful indexing of the documents, JP2 files can be viewed within the digital library interface and the bibliographic record in the library catalogue can be edited in a way described above.

## **Archiving**

Digitized documents are stored in the file system from where they can be mirrored to the server specially designated for this purpose. All files are also archived on tapes. This backup process is made regularly every month. For the nature of digitized documents it is necessary to ensure a proper way of document archival method. Because of the size of the digitized documents and its disk space demands images can not be stored in many copies in different formats. Therefore only TIFF files are stored and archived. By using the mirroring as a backup method it is possible to quickly recover archived documents and convert them in any format on demand.

## **Digitization outputs presentation**

The National Technical Library is trying to present the outputs of the digitization by:

- Enhancing the bibliographic records in OPAC with front pages thumbnails of the digitized documents, thus advertising the digitized content among other documents
- Providing paper flyers to the customers
- Displaying banners pointing out the services of the digital library

The main goal is to show the digital library as an interesting tool for study.

## **Conclusion and future work**

Previous digitization workflow of the National Technical Library was focused merely on preservation of the historical collections, not on providing a special service to the library customers. All processes were difficult to track and very time consuming, due to the lack of automation. With an increasing amount of books and other documents to be digitized, it was necessary to upgrade the old or create a new digitization workflow, to improve all processes embedded in it and to automate them in highest possible rate. NTL also decided to focus on digitization of university textbooks, which represent a significant part of its collections, to provide a solid study base for a big amount of library customers. At first, it was decided to digitize university textbooks in an order given by the amount of loans they had. Together with other digitization projects, as eBooks on Demand and digitization of thematic parts of the historical collections, it became clear, that contemporary digitization processes are not sustainable. It was decided to change the digitization policy and focus more on the digitization of historical collections and also on digitization of new university textbooks, acquired to the library collections.

The change in the digitization workflow began with integration of the digitization workplaces under one department. The main goal of the change was also to create one storage and place for all digitization projects and to standardize and write down all processes embedded in the digitization of printed documents. With this done, it was possible to automate some of the processes, like Optical Character Recognition and image conversion. All processes within the digitization workflow now can be easily monitored and maintained by responsible employees. As a part of the new workflow a new version of digital library was

installed in the National Technical Library, which is able to provide more functionalities for the customers.

The National Technical Library can now focus on enlargement of its digital library and on promoting this service among the library customers. Special effort will be made in automation of even more processes in the workflow. One of the next steps will be to implement a new version of metadata editor developed especially for the digital library Kramerius 4.

The National Technical Library will try to cooperate with technical universities in Prague to make available such university textbooks, which are often used during the lectures and that are not available in university's document repositories. The National Technical Library would also like to use its digital library for providing e-books from various publishers for its customers based on special licence agreements.

NTL's digitization workflow can now serve as a good practice model for other small libraries and memory institutions not only in the Czech Republic. Interested parties are welcomed to contact NTL in order to get a feedback on creation and use of single steps of the workflow.

### **List of interesting links**

- Detailed schemes of the processes in digitization workflow
  - 1) [Document selection and preparation](#)
  - 2) [Register of Digitization update](#)
  - 3) [Scanning](#)
  - 4) [Conversion](#)
  - 5) [OCR processing](#)
  - 6) [Metadata creation](#)
  - 7) [Import to K4](#)
  - 8) [Archiving](#)
  - 9) [Outputs presentation](#)
  - 10) [General scheme](#)
  
- [Digital library of the National Technical Library](#) (Kramerius 4)



## References

1. HUTAŘ, Jan. Nové standardy digitalizace (od roku 2012). NÁRODNÍ KNIHOVNA ČR. *Národní digitální knihovna* [online]. 2012, 10.5.2012 [cit. 2012-05-12]. Dostupné z: <http://www.ndk.cz/digitalizace/nove-standardy-digitalizace-od-roku-2011>
2. VYCHODIL, Bedřich. JPEG 2000: Specifications for The National Library of the Czech Republic. FEDERAL AGENCIES DIGITIZATION GUIDELINES INITIATIVE. *Federal Agencies Digitization Guidelines Initiative* [online]. 2012, 16.3.2012 [cit. 2012-05-12]. Dostupné z: <http://www.digitizationguidelines.gov/still-image/documents/Vychodil.pdf>