



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://nusl.cz) .

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

210 mm

# New digitization workflow of the National Technical Library

## In theory and practice

### Jakub Řihák

The Development of Electronic Services Department  
National Technical Library, Prague, Czech Republic

SEEDI 2012 Conference, Ljubljana, Slovenia, 17. 5. 2012

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## The National Technical Library

210 mm



- In the centre of technical universities campus
- 150 employees
- Mainly technical collections

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## Outline

210 mm

- Previous state of digitization in the NTL
- Main issues of our digitization
- Changes made
- Conclusion and future work

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## Things that were

210 mm

- Digitization began in the year 1998
  - Small scale
    - Primarily focused on document preservation
- Documents mainly on CD-ROM
  - Provided only on special computers
- Served as a base for future digital library

- An idea to provide something more
  - Our library acquired system for publishing of the digitized documents
    - Kramerius 3
- Change in digitization policy
  - University textbooks will also be digitized
    - Main criterium
      - Number of loans per year
- New working place was created for this purpose
- In 2009 NTL joined „eBooks on Demand: A European Library Network (EOD)“ project

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## Main issues

210 mm

- Different image formats used for publishing
  - PDF for university textbooks
  - JPEG for EOD books
  
- Different image quality
  
- Different naming convention
  
- Different storage place for archived documents

- But ALL digitized documents had to be published in the same digital library
  - Files sometimes could not be found (!!!)
  - They had to be renamed often
  
- Different image formats generated inconsistency in the digital library
  - Great issue during upgrading to the newer version



- Between years 2008 and 2011
  - 2941 documents were digitized
  - Only 862 of them were published
  
- Caused by
  - Time consuming processes (OCR, metadata creation)
  - Only one employee for everything
  
- Solution: **automation and new workflow**

- At the end of 2011
  - Decision made to change the digitization policy of the NTL
  
- Main focus on:
  - Digitization for the EOD project
  - Digitization of thematic historical collection
  - Digitization of newly acquired university textbooks
  
- Update to the new version of the digital library
  - Kramerius 4 (K4)

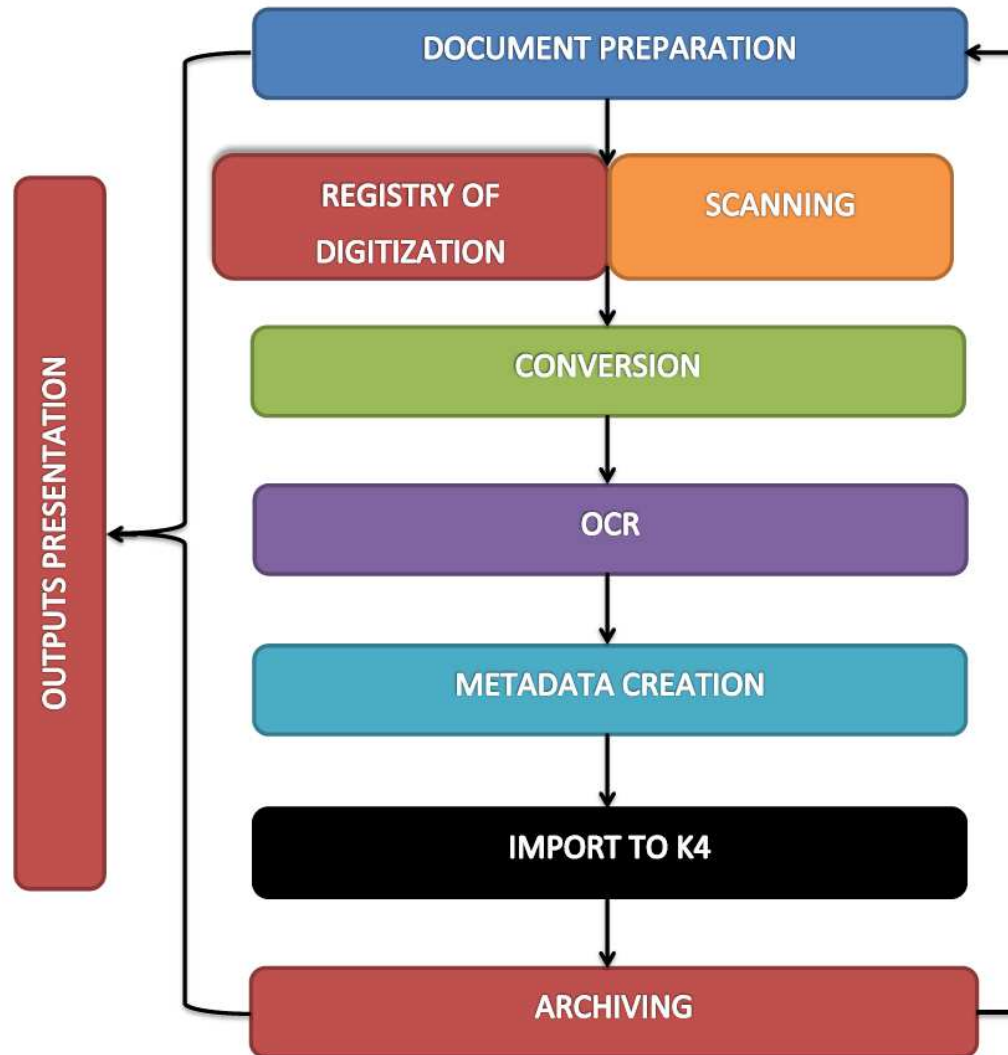
- Documents from Historical Collection (HC) could be published without restrictions given by the Copyright Law via the Internet
- Students didn't seem to care much about older textbooks which appeared in the digital library
  - This was caused by used selection criteria
    - Previously made lists of textbooks to be digitized were obsolete
- K4 is more user friendly

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## New digitization workflow schema

210 mm



- We needed to write down each process in the workflow
- Image outputs were standardized
  - We only produce TIFF files
    - Archived
    - Converted to other formats when needed for publishing
      - JPEG
      - JPEG2000
        - Automated batch conversion
        - Stored on the image server

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## What changes we made?

### Image server

210 mm

- Used for providing digitized documents in JPEG2000 format
  - Allows streaming of large images via the Internet
- Very useful when providing digitized documents from our historical collection
  - In high resolution
- Short example

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## What changes we made?

## Naming convention

210 mm

- Redefined
  - Because of the automation and metadata creation
  - System numbers from library catalogue
  - Signatures marks added to the name in case of textbooks
    - Metadata editor works with them
- Finally no special characters and diacritic (!!!)

## What changes we made? Storage of the documents

210 mm

- Only one storage place
- Predefined structure in the file system
  - Easier to find the documents and track what is done and what is still underway
- This folder can be used in the OCR process as a „Hot Folder“
  - All documents in this folder are processed



- Previously
  - Abbyy FineReader 10
- Now
  - Abbyy Recognition Server 3.0
  - Network Shared „Hot Folder“
  - Six processing stations
  - One verification station (more are needed)
  - 25 000 processed pages/month
- Output
  - Text files (.txt)
  - We will change it to ALTO XML format

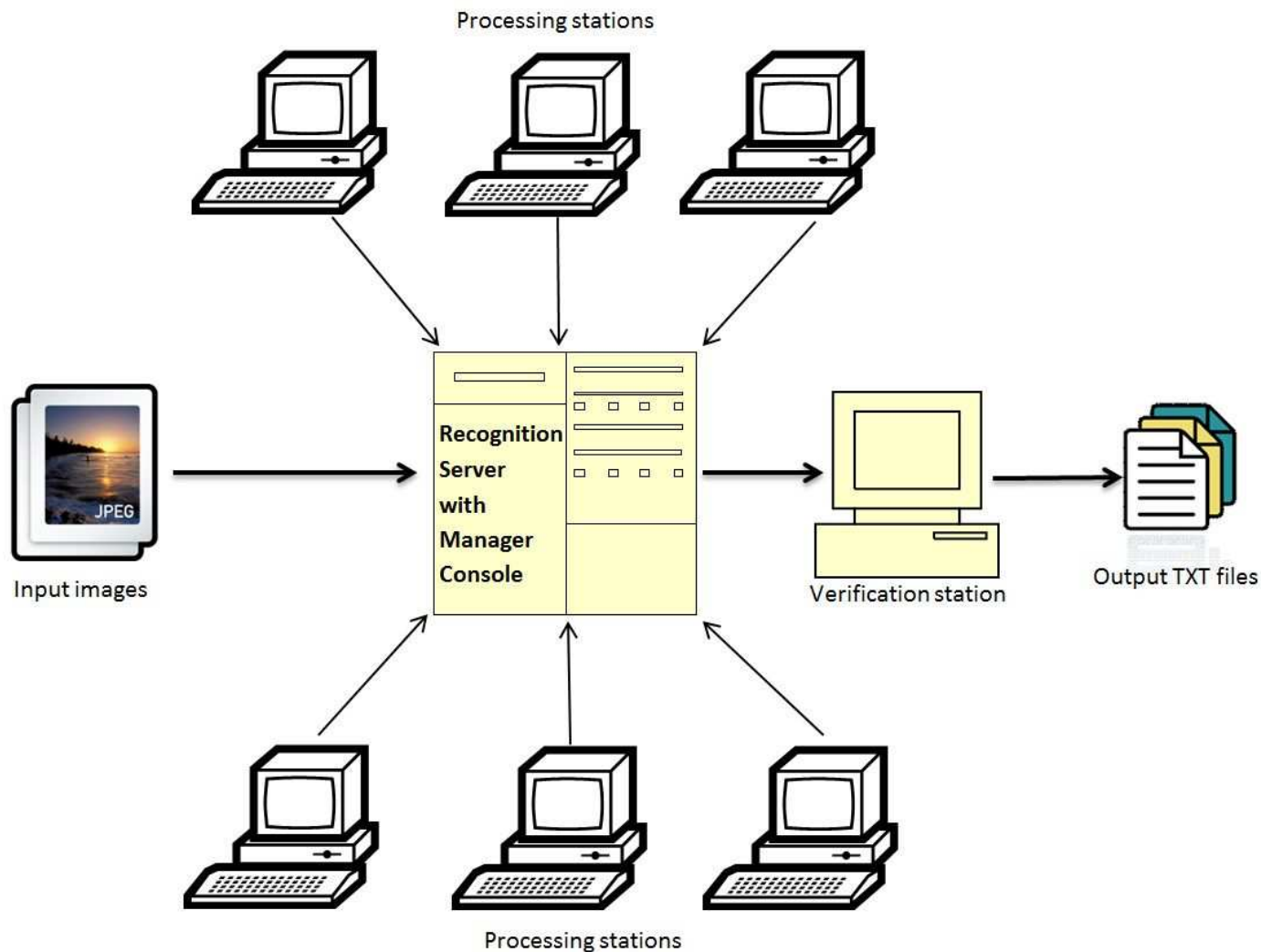
# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## What changes we made?

## Changes in the OCR

210 mm



# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## What changes we made?

### Number of employees

210 mm

- Digitization takes place only in one department
  - 7 employees
    - 1 employee – digitization of textbooks
    - 6 employees – digitization of HC
- 1 intern working on OCR verification
- 1 intern for other administrative work
  - Good experience - we are looking for more

- Digitization focused more on providing an interesting documents to the library customers
- Processes can be easily tracked
- More automation brings more time to solve other issues and develop this service
  
- **Something to do in the future:**
  - Implement new metadata editor for K4
  - More automation
  - Cooperation with technical universities
    - Textbook collections for students
  - More focus on presentation

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Technical Library

## Links to workflow schemes

210 mm

- [Document selection and preparation](#)
- [Register of Digitization update](#)
- [Scanning](#)
- [Conversion](#)
- [OCR processing](#)
- [Metadata creation](#)
- [Import to Kramerius 4](#)
- [Archiving](#)
- [Outputs presentation](#)

# NTK

50°6'14.083"N, 14°23'26.365"E

Národní technická knihovna  
National Technical Library

210 mm

Thank you for your attention! 😊

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

[www.techlib.cz](http://www.techlib.cz)

[www.k4.techlib.cz](http://www.k4.techlib.cz)

