

Digital Library of Latvia and the Experience of Digitisation of Periodicals at the National Library of Latvia

Ilze Pētersonē^[0009-0005-6757-2876]

National Library of Latvia, Mukusalas Street 3, Riga, Latvia
Latvian College of Culture at Latvian Academy of Culture, Bruninieku Street 57, Riga, Latvia
ilze.petersonē@lnb.lv

Abstract. The article provides an overview of the development of digital collections at the National Library of Latvia, highlighting experiences in digitising periodicals and making digitised material accessible. The article is based on the author's experience in digitising periodical collections in the period 2007-2024, as well as on information available online.

Keywords: Digital Preservation, Periodicals, Digitalization, National Library.

1 Introduction

According to research on the history of digital libraries, the first attempts at digitisation in the world took place in the late 1980s and focused on digitising scholarly publications (mostly journal articles) for academic use (Krūmiņa, 2012). Digitisation is the physical transformation and preservation of a text, image, sound, object, environment or process into a digital format. Digitisation can also be described as the practical steps taken to transform a physical cultural heritage object into a digital object and to obtain archive, user and metadata files of a digital copy of the cultural heritage object (Latvijas Nacionālās bibliotēkas Digitālās attīstības departaments, 2022).

Nowadays, digitisation is also a process of information management. Digitisation increases the accessibility of documents, making it possible to use a document not only in person but also online. This makes it possible for a document to be used simultaneously by several readers in different parts of the world. Digitisation also serves as a tool for document preservation (Valsts aģentūra "Kultūras informācijas sistēmas", 2010). Among Latvian libraries, the National Library of Latvia and its Digital Development Department are the most active in digitisation.

2 Exposition of the Investigation

Digitisation at the National Library of Latvia (NLL) began in 1999 with the creation of the digitised newspaper collection “Mantojums-1” (“Heritage-1”) (Fig. 1).

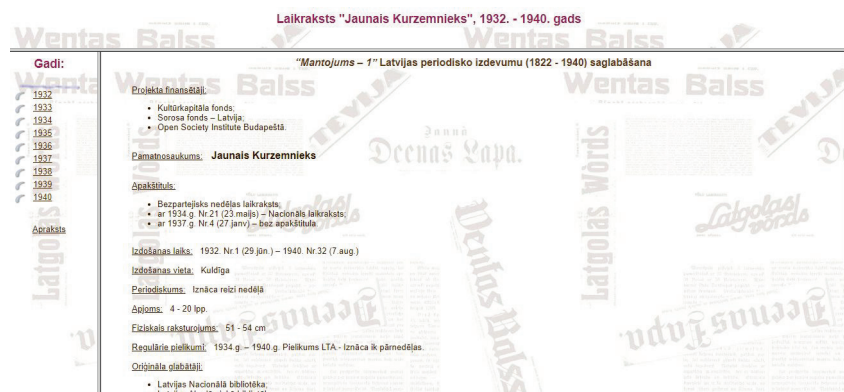


Fig. 1. Collection of digitised newspapers “Mantojums-1” (Available at: http://data.lnb.lv/digitala_biblioteka/laikraksti/JaunaisKurzemnieks/index.htm).

The database contained 100 titles of newspapers published in Riga and the regions in Latvian, German and Russian. The database was relatively simple, the text was not optically recognised, so it was not possible to search for names in the full text. The database is no longer maintained and efforts are being made to process and make available the material contained in the database in the same way as other material digitised at a later stage. We plan to do this through outsourcing within the next two years.

The mass digitisation of newspapers began around 2007. We estimate that more than 700 000 issues of periodicals have been digitised, which is about 80% of the entire periodical collection in Latvia. The digital collection is currently available on the website *periodika.lv* (Fig. 2). The collection of periodicals is expanded with new titles every year. This is possible thanks to a number of successful ERDF-supported projects, as well as smaller cooperative projects. For example, in 2023, a project on the digitisation of the Jewish Historical Press published in Latvia was implemented in cooperation with the National Library of Israel. This project continues this year.

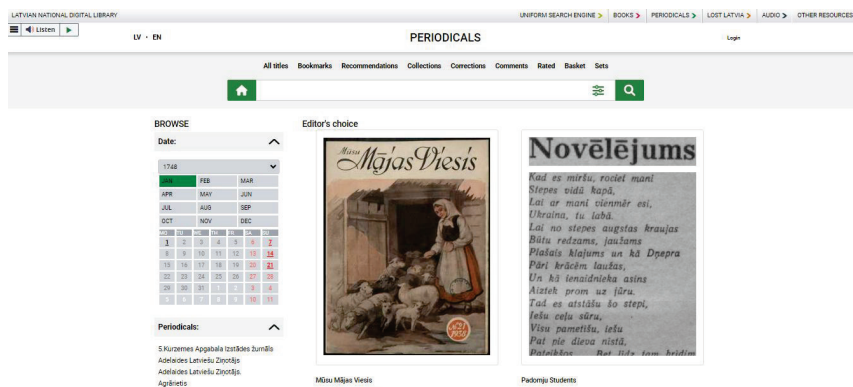


Fig. 2. Collection of digitised newspapers “periodika.lv (Available at: (<http://periodika.lv/#mainPage>)).

The digitisation of text usually involves scanning. In practice, however, a physical newspaper must go through several stages before it can be digitised. These include material selection, physical condition assessment, restoration, scanning, segmentation, optical recognition and other processes.

The digitised newspapers and magazines start their journey to internet users from the NLL Periodicals Department repository. The periodicals are stored in bound format. One year's issues are usually bound in 2 to 4 volumes.

Titles are selected for digitisation based on a number of criteria: the wear and tear of the title, its informational value, its availability (or unavailability) in other libraries, reader demand and other factors. Before submitting a newspaper for digitisation, it is important to go through the volume page by page to assess its completeness and physical condition (Fig. 3).



Fig. 3. The newspaper before it is sent for restoration.

There may be missing pages or even entire issues - we then search for the missing material in other libraries in Latvia or even abroad (e.g. Finland). We assess the need for conservation or restoration. Newspapers are a type of document that was never intended for long-term preservation. It is information that becomes outdated very quickly, the next day. Traditionally, newspapers have been printed on thin, low quality, non-durable paper that disintegrates quickly. Over time, the volumes have developed a wide range of defects - torn pages, cut-out articles, faded text, tight binding covering part of the text column, etc.

The restoration team spends a considerable amount of time on the edition - smoothing paper folds, fixing frayed page edges, repairing tears and, in some cases, even completely disassembling the volume. The Periodicals Department staff also puts notes in the volume with information for scanners if there are any important indications about colours, numbering, appendices, page numbers or other information.

The scanning process itself takes place in the Digitisation Department of the LNN, but for large-scale projects it is outsourced. Professional large-format A1 scanners are used to scan periodicals. Scanners with an adjustable base on which the opening of the volume can be adjusted are particularly useful for working with bound editions. This prevents damage to the volume and the pages instead of the binding. Newspaper pages are large and fragile and cannot be scanned using a scanning robot - each page must be turned by hand.

A resolution of 400 dots per inch is usually chosen for scanning. Higher resolutions are used in special cases, e.g. very small letters. Newspapers are mostly scanned in greyscale. Full colour mode (RGB) is used for the first issues of the year, issues with colour illustrations or colour design elements, and magazines.

Scanning produces images of newspaper pages. They are not yet machine-readable at this stage. Post-processing of the images is necessary to enable full-text searching. The National Library of Latvia uses the *docWizz* software developed by CCS (Fig. 4).

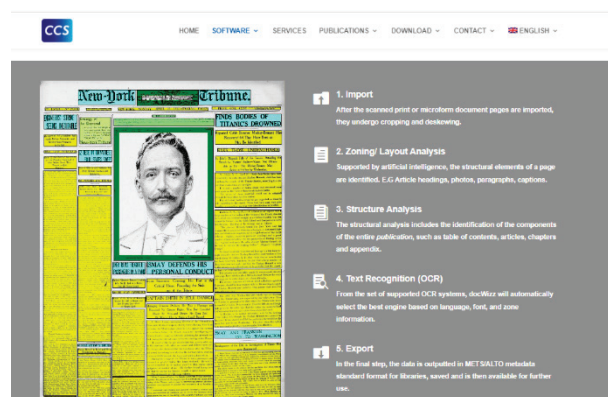


Fig. 4. Newspaper segmentation view and processing steps (Available at: <https://content-conversion.com/software/docwizz/>).

Post-processing includes page straightening, content zoning or segmentation, and optical text recognition. First, the scanned pages are straightened and the edges of the images are trimmed. If an image is not scanned straight during scanning, it is corrected in this step. The segmentation software then divides the image into zones and defines its different elements: titles, text zones, images, image captions, etc. This is not always accurate, so a human is needed to review and correct the work done by the programme. Optical recognition is used to enable the reader to find words in the newspaper text. For many reasons, the programme can misrecognise text - if the type is poorly printed, blurred or faded, the pages are damaged, different typefaces are used. Due to its huge size, the segmenter cannot correct all errors in the main texts, but usually corrects headings, authors, image captions. Text blocks that have not been recognised by the program at all are also revised.

After scanning and segmentation, the files are exported in different formats. One common newspaper file in *pdf* format, *tif* and *JPEG* files for each individual page, as well as various technical files containing information about the recognised text, the coordinates of each word on the newspaper page, etc. (Fig. 5).

	Slēds ID	DOV ID	Dane	Tips	Nosaukums	Apraksts	Secība	Izmērs	Darības	Danes piekļuves tiesības
Lietojumdatnes										
+ -	1	39402975	2420917	maw1907h99.pdf	Lietojumdatne		0	3.98 MB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)
Arhīvdarbes										
+ -	1	39402933	2420917	p_001_maw1907h99_001.tif	Arhīvdarbe		0	10.07 MB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)
+ -	2	39402934	2420917	p_001_maw1907h99_002.tif	Arhīvdarbe		0	19.78 MB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)
+ -	3	39402936	2420917	p_001_maw1907h99_003.tif	Arhīvdarbe		0	19.7 MB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)
+ -	4	39402938	2420917	p_001_maw1907h99_004.tif	Arhīvdarbe		0	19.53 MB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)
+ -	5	39402937	2420917	p_001_maw1907h99_005.tif	Arhīvdarbe		0	10.28 MB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)
Tehniskās darbes										
+ -	1	39402978	2420917	maw1907h99_001.jpg	Tehniskā darbe		0	1.19 MB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)
+ -	2	39402977	2420917	maw1907h99_001_ats.uvml	Tehniskā darbe		0	49.04 KB	Laijūkātē Apsēt Pielūjves tiesības	Pieejams bibliotēku tīklā (bez iespējas lejupielādēt)

Fig. 5. Technical files in the LNB digital object management system.

Searching for information on website *periodika.lv* allows you to find data for research, family and neighbourhood information, and you can find this information much faster than using a physical printed publication. They provide information on important political and cultural events of the time, everyday life, fashion and other areas of life.

3 Access to and Use of the Digitised Collection

In order to preserve the physical collection, the original digitised titles from the Periodical Collection are no longer issued, but are offered to readers in digital format. This is defined in the LNB Terms of Use: “for the purposes of preserving the Periodicals

Collection, ensuring the preservation of the Periodicals Collection as a particularly valuable set of national and historical materials of national importance as it is defined in the Library Law and other regulatory enactments, originals of the digitized periodicals of the Periodicals Collection and periodicals which are in a very poor physical condition are not issued to readers” (Latvijas Nacionālā bibliotēka, 2024).

Library users have become accustomed to using periodicals in digital format, and statistics show an increase in the use of digital material every year (Table 1).

Table 1. Website Periodika.lv usage statistics.

Year	Unique users	Website visits	Pages viewed
2016	59 475	215 385	967 806
2017	64 840	237 389	1 031 408
2018	87 911	297 752	1 307 098
2019	110 862	369 598	1 644 454
2020	No data available to the author		
2021	138 599	490 204	2 426 232
2022	76 498	309 574	1 573 577
2023	86 593	318 643	1 116 858

It should be noted that the use of *periodika.lv* increased particularly rapidly during the Covid-19 pandemic (2019-2021), when in-person visits to libraries and other memory institutions were severely limited.

Periodika.lv is by no means the only digital collection offered by the National Library of Latvia. Digitised books, periodicals, images, maps, sound recordings, etc., materials that formed separate collections, have been available under a single search engine *digitalabiblioteka.lv* (<https://digitalabiblioteka.lv>) since October 2023 (Fig. 6).

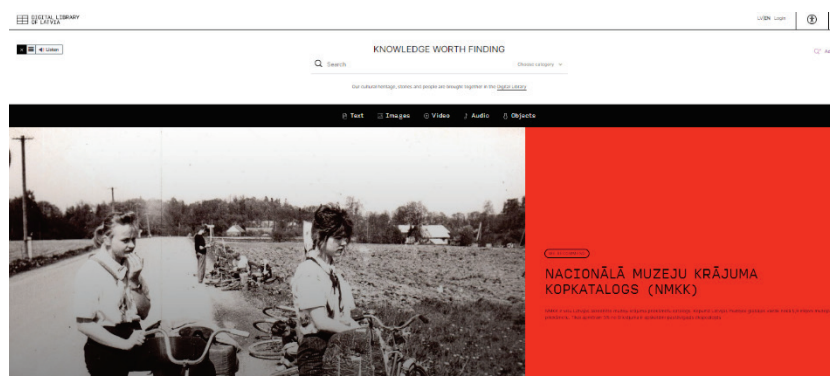


Fig. 6. Digital library (Available at: https://digitalabiblioteka.lv/?set_lang=en).

In addition to the resources of the National Library of Latvia, the Digital Library brings together the collections of more than 500 partners from Latvia and abroad - archives,

museums, libraries, organisations, communities and individuals - images, documents, periodicals, books, maps, museum objects, paintings, sheet music, films, sound and video recordings. The total size of the digital library is around 2.8 million digital items. The NLL brings together this digital content in the Digital Library and manages the Digital Cultural Heritage platform. This ensures the widest possible access to Latvia's cultural heritage.

Digital library content includes:

- newspapers, magazines and other periodicals published in Latvia and worldwide;
- books from the 17th century to the present day and small printings - address books from the first half of the 20th century, newsletters, etc;
- diaries, letters, patents and other documents from archives, museums and libraries;
- photographs, drawings, posters, blueprints, postcards, prints, engravings and other images;
- more than 5 700 maps of the historical territory of Latvia;
- composers' manuscripts, printed, handwritten and digitally created sheet music;
- objects from Latvian museums, including objects in kind and natural objects;
- intangible cultural heritage;
- documentation on the identification, survey and monitoring of cultural monuments, surveys of cultural monuments, including 3D surveys;
- documentary sound and video recordings, music recordings, films and newsreels made in Latvia, video recordings of the latest cultural events;
- cultural and analytical broadcasts, cultural programmes, theatre performances and radio programmes, including a significant part of the archives of Latvian Radio and Latvian Television;
- digital media and digital art archives.
- the Digital Library also includes library catalogue records alongside digitised and digitally created content (Latvijas Nacionālā bibliotēka, 2023).

The digital library also includes a "Learn" section - stories by experts, educational materials, instructions for using digital resources, guidelines for digitising cultural heritage, and other information.

4 Conclusions

The digitisation process has created a large number of digital documents in the form of individual collections. Most of them, including *periodika.lv*, are available online. The digital objects are chronologically broad and rich in content. Most of the digitisation was carried out within the framework of individual projects. When the funding for the projects runs out, the library continues to increase and maintain the digitised volume within its financial possibilities. The inclusion of *periodika.lv* in the general search engine *digitalbiblioteka.lv* allows a much wider search for information. Of course, the Digital Library is still a work in progress and is being expanded on a daily basis.

References

- Krūmiņa, L. (2012). Digitalizācija Latvijā pasaules pieredzes kontekstā [Digitization in Latvia in the context of world experience]. *Bibliotēku Pasaule*, 57, 39-45. <https://dom.lndb.lv/data/obj/file/162387.pdf>
- Latvijas Nacionālā bibliotēka. (2024). *Regulations for the use of the National Library of Latvia*. <https://lnb.lv/en/library/noteikumi/>
- Latvijas Nacionālā bibliotēka. (2023). *Digitālās bibliotēkas saturs* [Digital library content]. <https://digitalabiblioteka.lv/?module=article-Read&id=651eb530b4c75&tab=6>
- Latvijas Nacionālās bibliotēkas Digitālās attīstības departaments. (2022). *Kā uzsākt kultūras mantojuma digitalizāciju? : Rekomendācijas* [How to start digitalization of cultural heritage? : Recommendations]. https://digitalabiblioteka.lv/upload/cms_uploads/20221213/LNB-Digitaliza%CC%84cijas%20rekomenda%CC%84cijas%202022-LV.pdf
- Valsts aģentūra "Kultūras informācijas sistēmas". (2010). *Vadlīnijas dokumentu digitalizācijai valsts arhīvos* [Guidelines for digitization of documents in state archives]. <https://www.arhivi.gov.lv/lv/media/65/download>

Received: March 15, 2024

Reviewed: April 20, 2024

Finally Accepted: May 23, 2024