



## Сеть цифровых гуманитарных наук в Латвии: организации, инициативы и ресурсы

## Digital Humanities Network in Latvia: Institutions, Initiatives, Resources

**Anda Baklāne**

Развитие цифровой науки в Латвии и Украине: современное  
состояние и перспективы, 16.11.2021

[anda.baklane@lnb.lv](mailto:anda.baklane@lnb.lv)

# The structure of this presentation

## **1. Digital resources for humanities and social sciences in Latvia**

*entry points for libraries and archives, most important collections*

## **2. Digital humanities network in Latvia**

*the digitalhumanities.lv website, some of the ongoing activities*

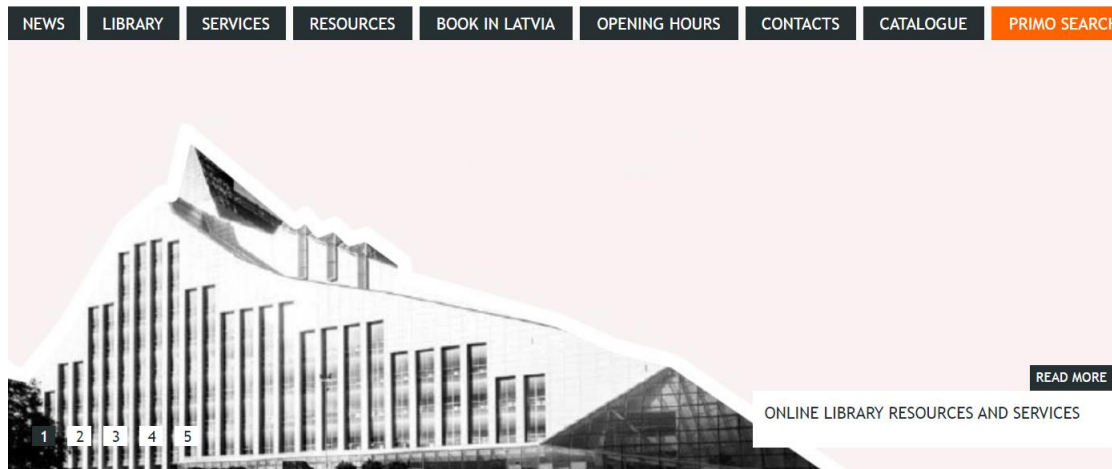




# **Digital resources for humanities and social sciences**

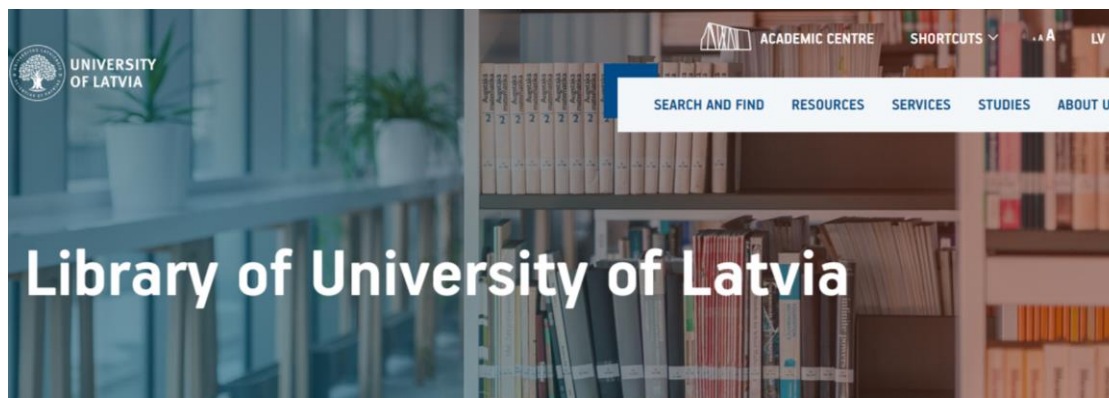


# Libraries



## National Library of Latvia (NLL)

<https://www.lnb.lv/en>



## Library of University of Latvia (LUA)

<https://www.biblioteka.lu.lv/en/>

# National Library of Latvia (NLL): aggregators / single access points

## INTEGRATED SEARCH ENGINE

### National Digital Library of Latvia

<https://www.lndb.lv/>

The largest publicly available repository of **textual** digital heritage in Latvia  
Encompasses resources of NLL and its partners, approx. 120 institutions and 350 individuals; will include all digitized collections of major ALM institutions

## LIBRARY CATALOGUE

### PRIMO Discovery Service

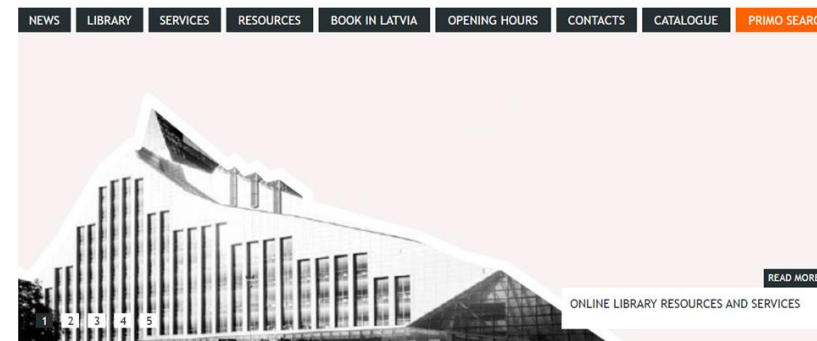
<https://www.lnb.lv> 

The union catalogue for NLL [and LUA], provides access also to digital resources; content can vary when you access the service as a user of NLL vs. user of LUA



<https://www.lnb.lv/en/resources/digital-library>

# NLL: Large digital collections based on the type of the resource



<https://www.lnb.lv/en/resources/digital-library>

VIENOTAIS MEKLĒTĀJS >

GRĀMATAS >

PERIODIKA >

ZUDUSĪ LATVIJA >

AUDIO >

CITI RESURSI >

PERIODIKA

GRĀMATAS

ZUDUSĪ LATVIJA

zīmējumos, atklātnēs un fotogrāfijās

Latvijas vēsturiskie  
skaņu ieraksti

KARTES

*Academia*  
Akadēmiskais repozitorijs

**periodika.lnb.lv (periodicals)**

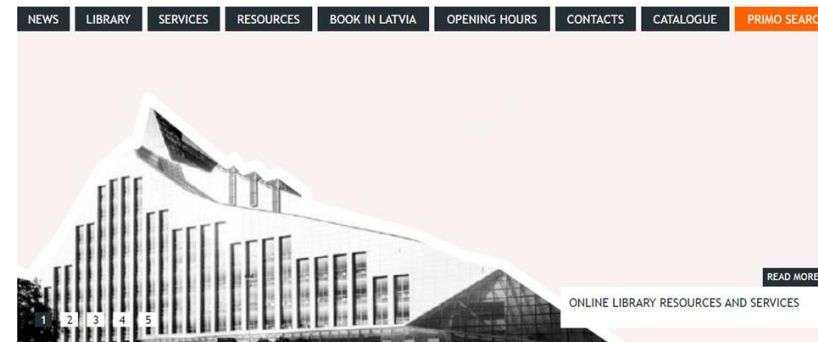
**gramatas.lnb.lv (books)**

**audio.lnb.lv (audio)**

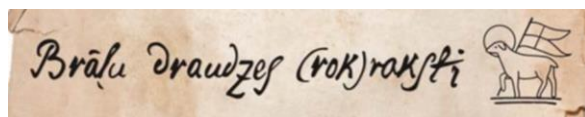
**zudusilatvija.lv (photography)**

**[+ other]**

# NLL: Collections developed around specific topic and/or functionality



<https://www.lnb.lv/en/resources/digital-library>



**DZIESMU SVĒTKU  
KRĀTUVE**

*Gottardus Fridericus Stender.*



Latviešu prozas skaitītājs

**braludraudze.lnb.lv** (Moravian Brethern)  
**stenders.lnb.lv** (G. F. Stender)  
**dziesmusvetki.lnb.lv** (Song & Dance Festival)  
**proza.lnb.lv** (Latvian novels)  
**runa.lnb.lv** (Rainis & Aspazija)  
[+ other]

# Univerisity of Latvia Library

## Repository of the University of Latvia

<https://dspace.lu.lv/dspace/>

Scientific papers, doctoral thesis, other UL documents

Materials of the Academic Library  
(incl. Misiņš Library) at UL Dspace

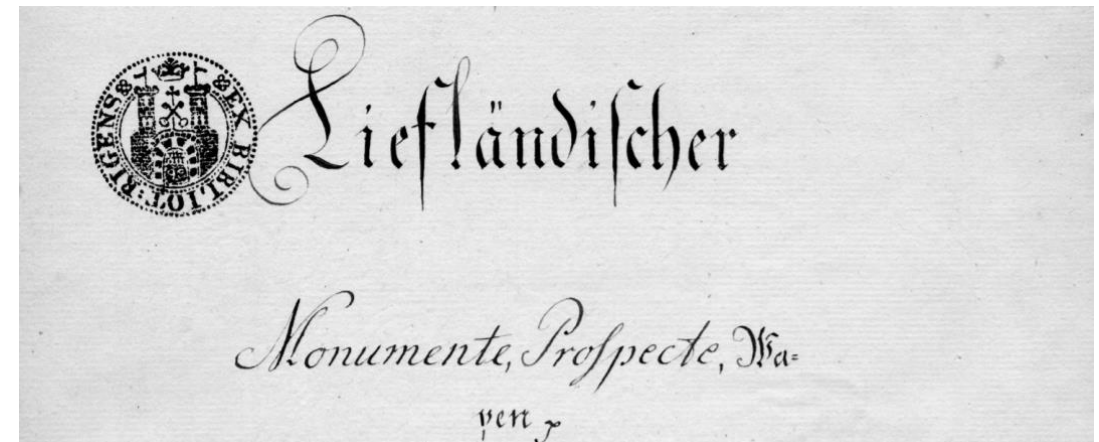
<https://www.acadlib.lu.lv/>

Contains historical **monographs**,  
**manuscripts**; periodicals are part of  
periodika.lndb.lv



## E-resource repository of the University of Latvia

E-resource repository of the University of Latvia provides the possibility to collate, preserve and access freely the scientific achievements of the University of Latvia. It contains publications of the staff members of the University of Latvia - published articles, doctoral thesis and abstracts, conference proceedings, administrative, scientific and funded project reports and other electronic documents.





# Archives

## The National Archives of Latvia

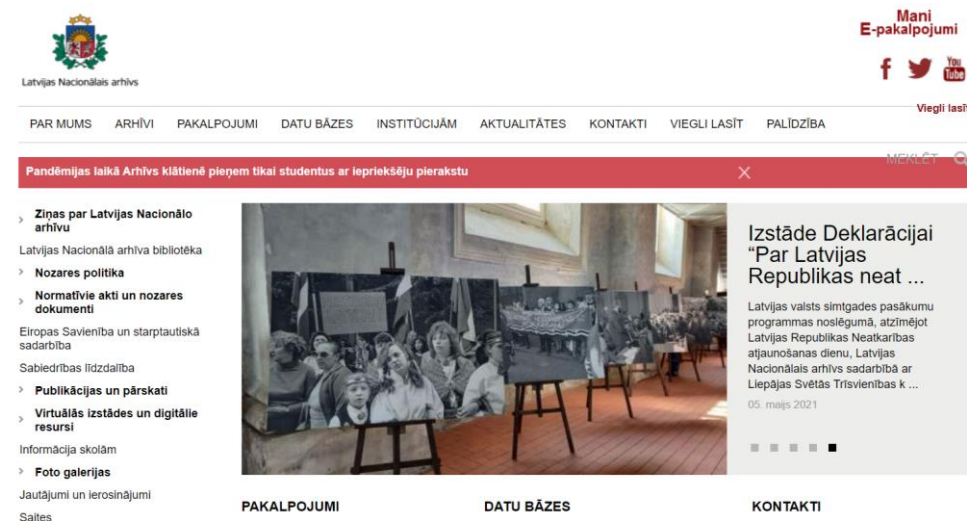
<https://www.arhivi.gov.lv/>

SINGLE ACCESS POINT TO AUDIOVISUAL DOCUMENTS

**See, Hear Latvia!**

<http://www.redzidzirdilatviju.lv/en/>

Digital archive of **photo**, **audio** and **video** documents of the Latvian State Archives of Audiovisual Documents; information on 9115 films, 58911 photographs, 507 audio recordings



## Two other popular databases of National Archives

## Raduraksti ("Lineages")

<https://www.raduraksti.arhivi.lv/>

Church records, census data,  
graduates of the University of Latvia,  
documents of Riga Polytechnical  
Institute, lists of deported citizens

# KGB document archive of LSSR

<https://kgb.arhivi.lv/>

Card indexes, the dossiers of recruited KGB agents, the KGB employees' phonebooks, manuals for agents and other materials on the Central Committee of the Communist Party of Soviet Latvia.

Годъ и мѣсяцъ. 1894.		Нумеръ; имя младенца; имя, фамилія и вѣро- исповѣданіе родителей; званіе, чинъ или ремесло отца, или матери, или представившаго младенца къ крещенію; прозваніе Провождника, совер- шавшаго Св. Крещеніе, и гдѣ оное совершено; имена, фамилія и званія, чины или ремесла восприимниковъ.	Законно- рожденные.		Незаконно- рожденные.		Мертворож- денные или умершие до крещенія.	
День и часть рожденія.	День крещенія.		Муже- скаго пола.	Жен- скаго пола.	Муже- скаго пола.	Жен- скаго пола.	Муже- скаго пола.	Жен- скаго пола.
Воспитана Шрида, Января того 11.7.60. Января		№ 5. Азизъ, Насторатъ; (Fustamat) Имла Юсталантъ Род: купцы Виллоль Юсталантъ и его ж. Каринкоу. Имя, о ба св. мот. Воспр: (Ксеріе Берль, Инвица Э. Мало Музенекко, Давача А. Истел, Гавица)		2				

[illegible]

Uzzinu kods: LV LVA F1 US30 GV4-9 0002

Паспорт

ШПЭ

33

№ 302

Ф. И. О. ШУМАНУС РОМАН

Дата рождения 14.01.1931

Место рождения г. Рига

Дата рождения 11.09

Место рождения г. Рига

Адрес г. Рига, ул. Мейеровска 12-4

Место работы и должности

Порт ЕА КИСС

Ученое звание и звание СССР

Образование высшее

Оформление личного дела и дата за уст

Зас. № 1

Секретарь

Аббревиатура подписи, форма 3

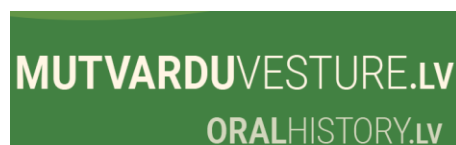
Uzzinu kods: LV LVA F1 US30 GV4-9 0003

# Digital Archives of Latvian Folklore

## Garamantas ("Folklore")

<http://garamantas.lv/>

Incorporates collections of Archives of Latvian Folklore as well as several other institutions



The screenshot displays the Garamantas.lv website interface. The header includes the site name, a search bar, and navigation links for Collections, Persons, Illustrations, Audio, Organisations, Map, and Other. A sidebar on the right offers options to 'TAKE PART!' in various projects like LFK jautā, Buramvārdi, Dziedi ar arhivu, and Mūsdienu kalendārs. The main content area features a 'TRANSCRIBE!' section with a 'SUGGESTIONS FOR TRANSCRIPTIONS' list and a 'RECENT TRANSCRIPTIONS' list. Two featured articles are visible: 'Digital Catalogue of Latvian Charms' (dated 29.01.2021) and 'A Collection of Ethnographic Ensembles' (dated 04.01.2021). The website uses a green and white color scheme with a wavy pattern background.





# **DH community in Latvia: network and activities**



# DH community in Latvia

**Digitalhumanities.lv** network

<http://www.digitalhumanities.lv>

- The website currently displays most active members and a list of digital resources and tools
- The list of resources is not exhaustive but can give an idea of the main actors and resources

**National Library of Latvia**

**Institute of Literature, Folklore, and Art (UL)**

**Artificial Intelligence Laboratory (IMCS UL)**

**Latvian Language Institute (UL)**

**Livonian Institute (UL)**

**Riga Technical University**

**Rezekne Academy of Technologies**

**Liepaja University**

**Tilde**



# Digitalhumanities.lv on social media



Digitālās humanitārās zinātnes Latvijā

39 abonenti

PIELĀGOT KANĀLU

PĀRVALDĪT VIDEOKLIPUS

SĀKUMS

VIDEOKLIPI

ATSKAŅOŠANAS SARAKSTI

KANĀLI

DISKUSIJA

PAR



Augšupielādes

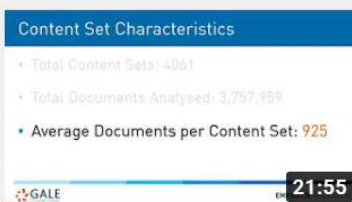
▶ ATSKAŅOT VISU



Periodikas digitalizācija |  
Digitization of Periodicals

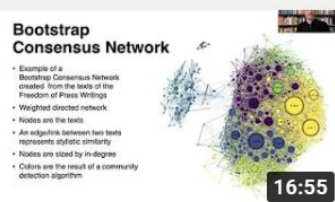
333 skatījumi • pirms 1 mēneša

Subtitri



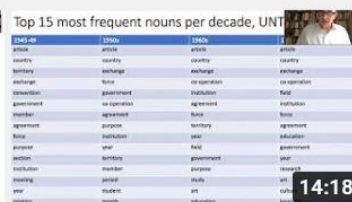
Evaluating a DH Tool: The  
First 18 Months of Gale...

8 skatījumi • pirms 1 mēneša



Exploring the Potential of  
Bootstrap Consensus...

5 skatījumi • pirms 1 mēneša



Keeping it Simple: Word  
Trend Analysis for the...

4 skatījumi • pirms 1 mēneša



Digital History of Virtual  
Museums: The Transition...

60 skatījumu • pirms 1 mēneša





# Artificial Intelligence Laboratory

Founded in 1992, Artificial Intelligence Laboratory (AiLab) at [IMCS, University of Latvia](#) conducts research on [natural language processing](#) (NLP) and [machine learning](#) (ML) in several directions:

- **Speech-to-text** recognition and **text-to-speech** synthesis for Latvian.
- **Syntactic** and **semantic parsing** for cross-lingual information extraction, question answering, human-computer interaction.
- Multilingual **text generation** from abstract meaning representations and from data.
- Controlled natural language for **knowledge representation**.
- Creation of annotated **language resources** for NLP and ML: machine-readable **lexicons**, text and speech **corpora**.
- Application of **deep learning** and **reinforcement learning** in **robotics**, **computer vision** and **grounded language learning**.

AiLab has recently implemented multiple research and market-oriented projects on large-scale information extraction and speech recognition. In cooperation with National Information Agency LETA, we are handling **big data** problems arising in the **media monitoring** business. This successful collaboration allowed LETA to join the EU Horizon 2020 project [SUMMA](#), together with leading European universities, BBC, Deutsche Welle and others. Researchers from AiLab were also involved.

We have developed a text processing system for National Library of Latvia, dealing with **part-of-speech tagging** and **named-entity recognition** in a billion word corpus of OCR-scanned texts. We also collaborate with Riga Stradins University, developing platforms for monitoring user-generated content in news portal communities and for analysing transcripts of parliament debates. Likewise, we collaborate with an international defence organisation to develop tools for tracking and countering online disinformation. In collaboration with Riga East University Hospital, we develop Latvian speech recognition and synthesis technology for **medical applications**.

<http://www.ailab.lv/en/>

# National Library of Latvia

<https://lnb.lv/>



- **Main focus currently: digital text analysis**
- **Data sets on demand (incl. full text)**
- Text analysis in the **Jupyter Notebook environment**: [lab.lnb.lv](https://lab.lnb.lv)  
*Collaboration with researchers within projects*

- **Text analysis demonstration tool**

Latvian Prose Counter <https://proza.lnb.lv/skaititajs/>

- **Corpus analysis platform**: <https://nosketch.lnb.lv>  
*Concordances, word frequency counts for historical and literary corpora (just started!)*

```
44 </revisionDesc>
45 </teiHeader>
46 <text><body>
47 <div type="chapter"><head>1.</head>
48 <p>Ārīgas bedrainajā sānu ielā, kuru apdzīvo žīdi un strādnieki, šodien savāds gaisa
apkārtņi tik jauka, kairinoša cepeša smarža! Tur pie zemā valējā loga divi žīdu kurpr
kairinošo gaisu, šai mājelei turpat blakus atrodas bārdzīņa darbnīca, izkārtne pie
laudis steidzas, likt noskūt bārdas vai apcirt matus. Bet te viens otrs deguns, saju
parādību, paliek uz ielas stāvam. Baznīcas pulksteņi iekšpilsētā un tepat no tuvējām
zvani zvana ikkatru svētdienu tik svinīgi, tomēr no tāda jauka, smaržojoša gaisa nav
<p>Ikviena sajūt ko ārkārtīgu, ko neparastu, ko sevišķu. It labi zināms, ka te tuvie
te ļaudis apkaimē viens otru it labi pazīst un pārzina, ko katrs iespēj un ko katrs s
```

**CONCORDANCE**

simple **sieviete** 1,382 (407.32 per million)

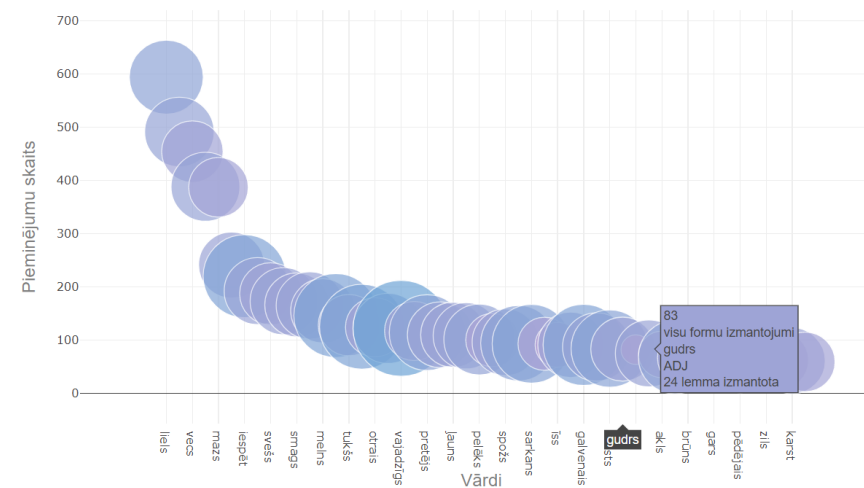
☐ Details ☐ Left context ☐ KWIC ☐ Right context

1	<input type="checkbox"/>	doc#1	ī skaisti trauslās kristalla domas un bālās, nervozās un skaistās	<b>sievietes</b>	, lai tikai varētu atgulties
2	<input type="checkbox"/>	doc#1	šā tomēr vēl dažus, kuņiem tikās to pavadīt. Kā savādi smaidīja	<b>sieviešu</b>	brūnās lielās acis un ner
3	<input type="checkbox"/>	doc#1	auga mātes iespaidā. Māte Dangam bija atmiņā kā sveša, bāla	<b>sieviete</b>	viņu mājās, kuŗa bieži bi
4	<input type="checkbox"/>	doc#1	ņņu un sirds skaidrības. Viņa bija no tām nedaudzajām latviešu	<b>sievietēm</b>	, kuŗām katrs praktiķis pi
5	<input type="checkbox"/>	doc#1	atviešu sievietēm, kuŗām katrs praktiķis pieliek vārdu: nederīga	<b>sieviete</b>	. Grāmatās vēl tagad bij
6	<input type="checkbox"/>	doc#1	nes skaistulēm un tiešām daba bija darījusi visu, lai tā būtu īsta	<b>sieviete</b>	. Viņa nebija liela augun

Izvēlies autoru

☐ Visas vārdšķiras ☐ Lietvārdi ☐ Darbības vārdi ☒ Īpašības vārdi ☐ Apstākļa vārdi

Populārākie vārdi Regīnas Ezers darbos  
Katrā pamatformā jeb lemmā ietvertas visas vārda formas



# Another project of National Library: Machine learning and crowdsourcing for improving optical recognition of Latvian fraktur script

<https://frakturs.lnb.lv/>

Supervised machine learning approach:

- Training data is prepared manually by correcting a sample (at least 10 000 rows)
- Trained model used for recognizing the rest of the corpus

Tesseract software, LSTM neural network.  
Original model based on the German fraktur



to, ka nu teem buhs weena deena swabada, kuru tee warēs	<input checked="" type="radio"/> Pārbaudīts <input type="radio"/> Nesaprotams
īleētot kattris pehz fawas patikfchanas; ka warēs atpuhstees un	<input checked="" type="radio"/> Pārbaudīts <input type="radio"/> Nesaprotams
īspreezatees. Wimu preezīgā tīchalofchana peepildija eelas, un	<input checked="" type="radio"/> Pārbaudīts <input type="radio"/> Nesaprotams
dafchs labs fālīhzis firmgalwis ar labpatikfchanu noluhkojās	<input checked="" type="radio"/> Pārbaudīts <input type="radio"/> Nesaprotams

\s S

]s f

[s l

[a â

]a à

[e ê

]e è

Saglabāt



# What National Library aims to develop in the following years:

## - New functionalities for regular digital collections

*For example:*

Visualizing search results

Topic modeling

Discovery and tracking of similar text strings

*[Which algorithms, techniques should become part of the new **core services**?]*

## - A platform for niche text analysis projects – more robust than notebooks

Traditional corpus analysis tools

Stylometry

Automatic annotation of named entities

Machine learning based research



# Institute of Literature, Folklore, and Art: Crowdsourcing, transcribing, mapping literature and folklore



LFMI Digital

Crowdsourcing activities to enrich the data of databases <http://garamantas.lv/> and <https://www.literatura.lv>, as well as to engage new audiences

Types of crowdsourcing practiced by ILFA:

- **Technical tasks:** transcription of ancient manuscripts, keywords tagging
- **Knowledge sharing:** enriching metadata, surveying
- **Creative response:** reading aloud, voice recording, interpreting archival sounds



<http://garamantas.lv/en/map/index>

Transcribed  
pages

146275

### How to participate?



#### 1. Register!

Indicate your name, school and class!

#### 2. Choose a file with text!

You can choose files in different languages

#### 3. Transcribe it!

### Choose the text

In Latvian

In Latgalian

In Livonian

In Lithuanian

In Estonian

In Russian

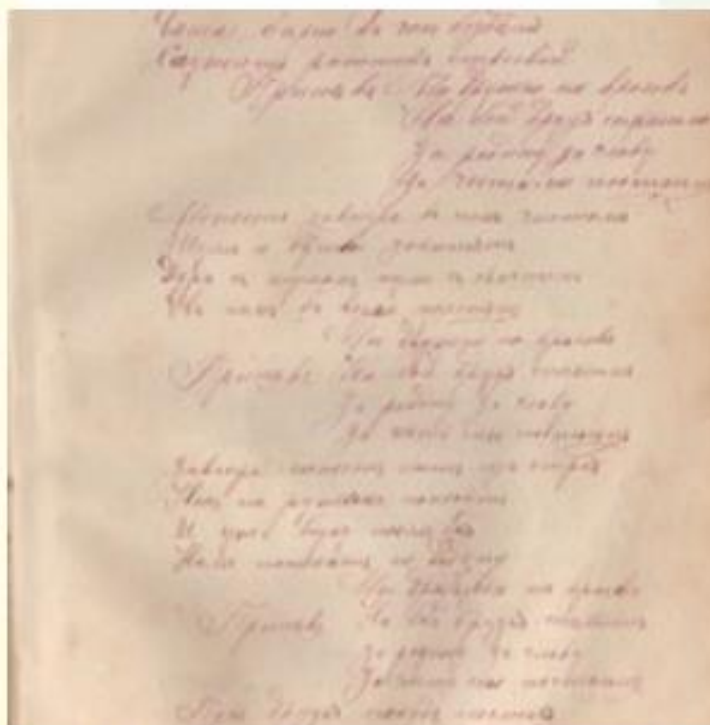
In Belarusian

In Yddish

In Romani

In Polish

In German



Total time spent by  
volunteers

1528 days

05:20:38

### Latest transcriptions



Pirms 7 minūtēm LigaB:0008-  
Lielaucē-pamatskola-02-0170



Pirms 10 minūtēm Aija  
Joņina:0059-Valija-Zakevice-01-  
0041



Pirms 15 minūtēm Aija  
Joņina:1955-Liepājas-ekspedicija-  
01-0055



Pirms 39 minūtēm LigaB:0008-  
Lielaucē-pamatskola-02-0169

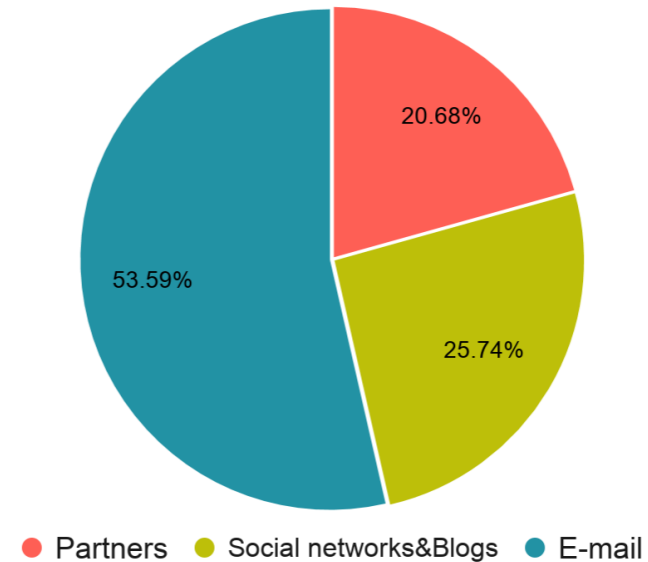
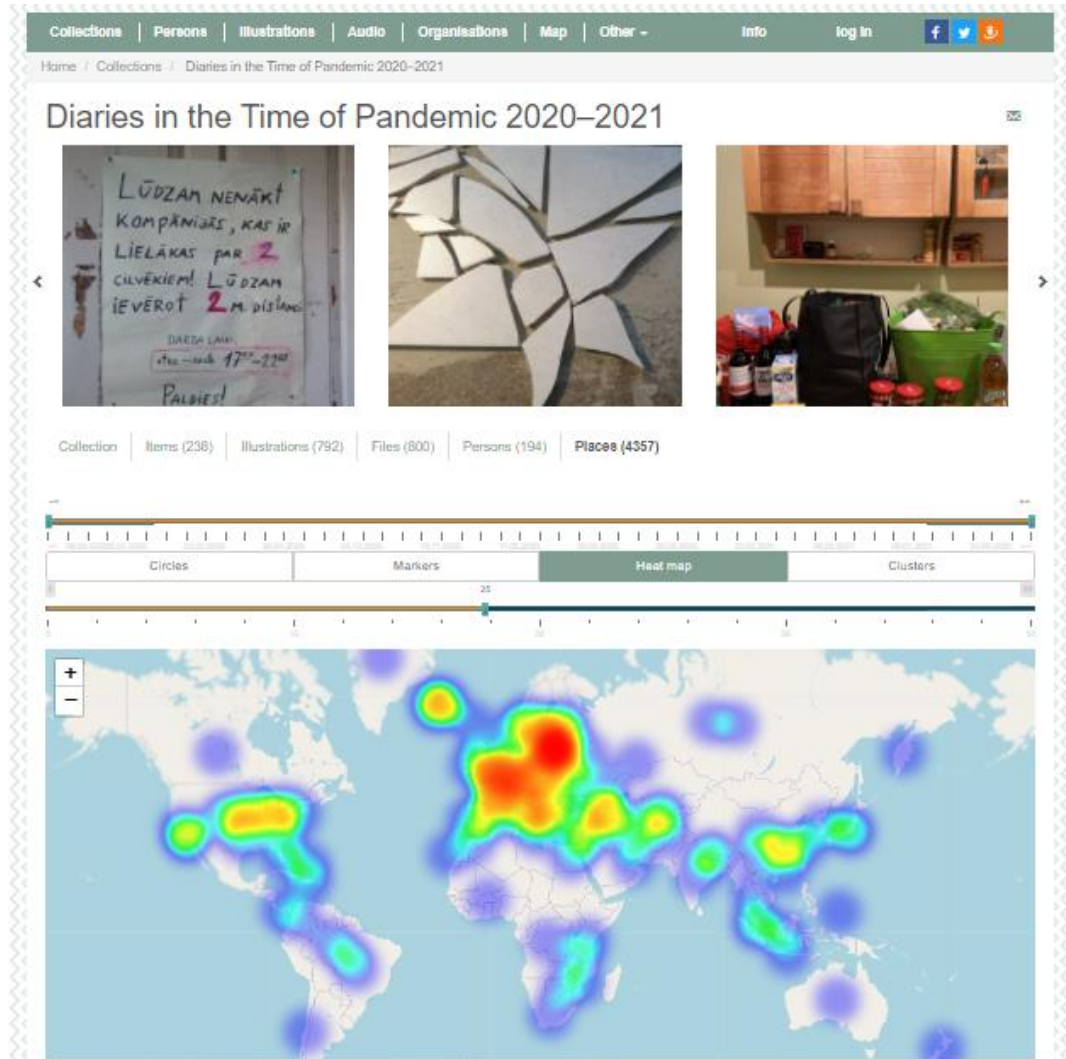


Pirms 1 stundas Aija Joņina:0123-  
Malpils-Kikutu-pamatskola-0005



# Not only historical materials: Pandemic diary project

<http://garamantas.lv/en/collection/1415829/Diaries-in-the-Time-of-Pandemic-2020-2021>



237 contributors (19% unanymous)  
2219 records (by 26.04.2021)  
784 photos and drawings  
4350 places mentioned (annotated)

Дякую!

