

# Andmete avamine RaRa digilaboris

Peeter Tinitis  
14.10.2024

Digipärandi seminar: masin loeb



# Rahvusraamatukogu digilabor

RARA

Andmestikud

Tööriistad

Blogi

Uudised

Sündmused

Meist

EST ▼



## ANDMESTIKUD



### Andmestike ülevaade

Loe lähemalt digilabori andmestike võimalustest

01.03.2023 • [Bibliograafiad](#), [Hei!](#), [Isikud](#), [Kollektiivid](#), [Multimeedia](#), [Perioodika](#), [Pildimaterjal](#), [Raamatud](#)



### Raamatud

DIGAR-is ja ERB-is leiduvate raamatute metaandmed

01.03.2023 • [Raamatud](#)



### Perioodika

DIGAR-is, DEA-s ja ERB-is leiduva perioodika metaandmed

01.03.2023 • [Perioodika](#)



### Pildimaterjal

DIGAR-is ja ERB-is leiduva pildimaterjali metaandmed

01.03.2023 • [Pildimaterjal](#)

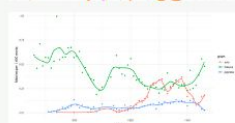
## TÖÖRIISTAD



### Eesti tõlkekirjanduse võrgustik

Eesti tõlkekirjanduse võrgustik 19. sajandi algusest tänapäevani.

13.03.2024 • [Bibliograafiad](#), [Raamatud](#)



### Sõnamitmike loendaja ajalehtedes

Töõriist ajalehtedes esinevate sõnamitmike sageduse visualiseerimiseks

26.03.2023 • [Ajalehed](#)

Vaata kõiki

## Blogi



Digitaalsed rännakud ehk Kultuuripärandi kujutamise kaardidel

Kultuuripärandi kujutamine interaktiivsetel veebikaardidel.

10.06.2024 • [Juhumiuuringud](#)



Eesti tõlkeajaloo digigalaktika toob ilmsiks kultuuri ja ühiskonna aluskoe

Eesti tõlkekirjanduse võrgustiku analüüs.

08.04.2024 • [Juhumiuuringud](#)



Kes oli rahvusliku ärkamisaja keskmes?

Ärkamisaja kirjastamisvõrgustiku silmapaistvad isikud ja keelevalikud.

15.01.2024 • [Juhumiuuringud](#)

# Uuri koos meiega!

[digilab.rara.ee](https://digilab.rara.ee)



RARA

Co-funded by the  
Creative Europe Programme  
of the European Union



# Andmete avamine

Tekstid



Metaandmed



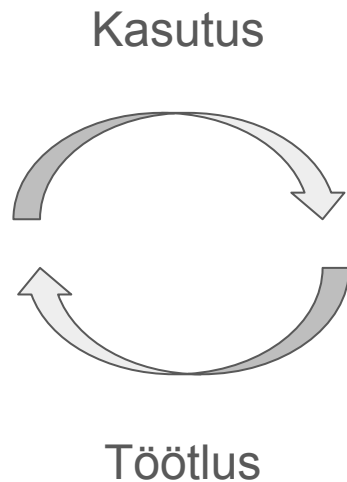
Tööriistad



**F**indable   
**A**ccessible   
**I**nteroperable   
**R**eusable 

# Teekonnad

- Tutvustamine ja populariseerimine
- Andmete töötlemine ja vormistamine
- Andmete kasutamine ja tööriistad



# Tööriistad

digilabor > Tööriistad > Eesti tõlkekirjanduse võrgustik

## Eesti tõlkekirjanduse võrgustik

Rakendus on eraldiseisvana leitav veebilehel: <https://data.digar.ee/tkirjandus/index.html>



### Eesti tõlkekirjanduse võrgustik

[näita kirjeldust](#)

Periood: 1800 - 2024 vali lähtesta

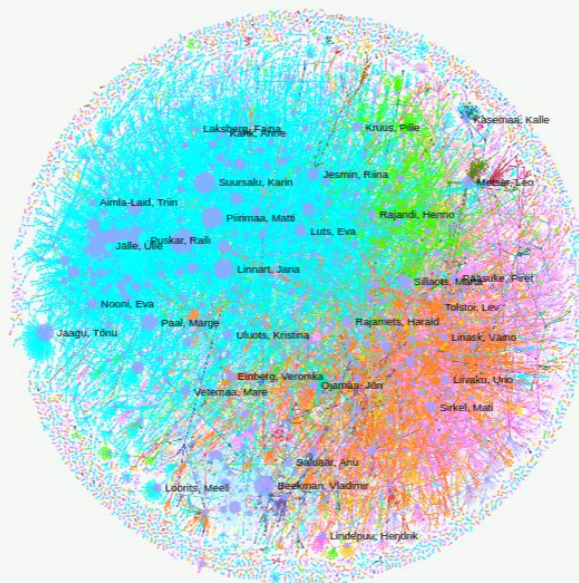
### Otsi isikut

Sisesta nimi...

### Filtreeri keele järgi

vali kõik tühjenda kõik

<input checked="" type="checkbox"/> inglise	45.84 %
<input checked="" type="checkbox"/> vene	13.39 %
<input checked="" type="checkbox"/> saksa	11.88 %
<input checked="" type="checkbox"/> prantsuse	5.81 %
<input checked="" type="checkbox"/> soome	3.85 %
<input checked="" type="checkbox"/> rootsi	3.52 %
<input checked="" type="checkbox"/> hispaania	1.37 %
<input checked="" type="checkbox"/> läti	1.18 %
<input checked="" type="checkbox"/> itaalia	1.17 %
<input checked="" type="checkbox"/> poola	1.17 %
<input checked="" type="checkbox"/> norra	1.18 %
<input checked="" type="checkbox"/> ungari	1.04 %
<input checked="" type="checkbox"/> taani	0.97 %
<input checked="" type="checkbox"/> tšehhi	0.74 %
<input checked="" type="checkbox"/> leedu	0.63 %
<input checked="" type="checkbox"/> hollandi	0.57 %
<input checked="" type="checkbox"/> rumeenia	0.32 %



# Tööriistad

digilabor > Tööriistad > Digiteeritud ajalehed Eestis

## Digiteeritud ajalehed Eestis

[Demo](#) [Juhend](#) [Dokumentatsioon](#) [Andmed ja kood](#)

Eesti kultuuripärandist on digiteeritud hulk ajalehti, aga kaugelgi mitte kõik. Seda informatsiooni on osaliselt kajastatud ERB perioodika [andmekogus](#). Siin on loodud visuaalne vahend sellest ülevaate saamiseks.

Rakendus on eraldiseisvana leitav veebilehel: [https://digileb.shinyapps.io/digitized\\_newspapers/](https://digileb.shinyapps.io/digitized_newspapers/)

### Plot type

- ☐ Image (simple)
- ☐ Interactive (simple)
- ☒ Interactive (grouped)

### Years in observation



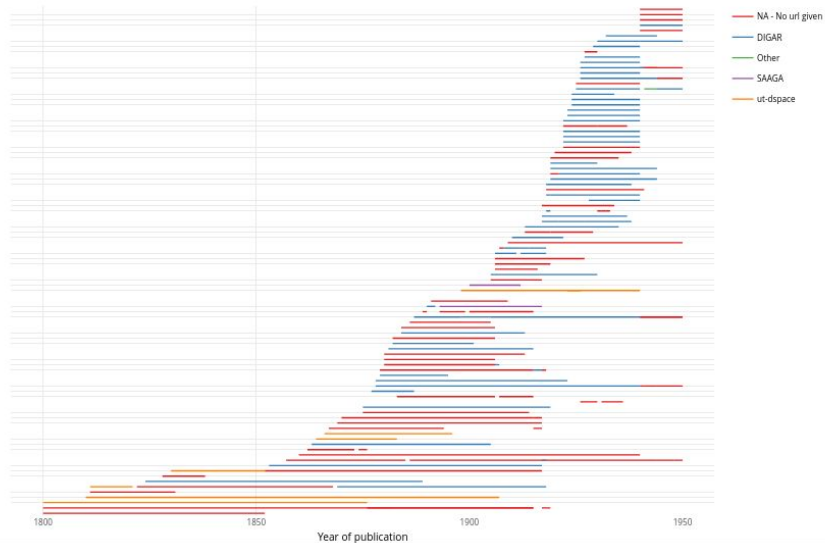
### Minimal years published



### Line thickness



Newspapers in collection (More than 10 years of publication, n = 96)



# Andmed

digilabor > Andmestikud > Raamatud

## Raamatud

ERB DIGAR

### Eesti rahvusbibliograafia

#### Eestikeelsed raamatud

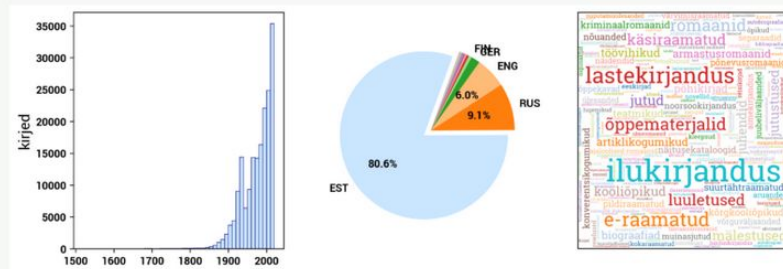
Eesti rahvusbibliograafia alamhulk, mis sisaldab eestikeelsete ja osaliselt eestikeelsete raamatute metaandmeid. ERB-is leiduvate eestikeelsete raamatute metaandmestik on kättesaadav:

- Zenodo repositooriumis TSV-failina
- ZIP-failina MARC21XML formaadis

TSV

MARC21XML (ZIP)

- OAI-PMH protokolli kaudu MARC21XML formaadis: <https://data.digar.ee/repos/OAIHandler?verb=ListRecords&set=raamat&metadataPrefix=marc21xml>



Andmestiku ajaline, keeleline ning temaatiline jaotus

Kuidas andmestikule viidata:

Eesti Rahvusraamatukogu. (2023). *Eesti rahvusbibliograafia – eestikeelsed raamatud* [Andmestik]. <https://doi.org/10.5281/zenodo.8228805>

# Andmed

zenodo

Search records...

Communities

My dashboard

Published November 11, 2024 | Version 1

Dataset

Open

Curated Estonian National Bibliography - books

Kruusmaa, Krister<sup>1,2</sup> ; Tinitis, Peeter<sup>3</sup> ; Nemvalts, Laura<sup>2</sup>

Show affiliations

Contributors

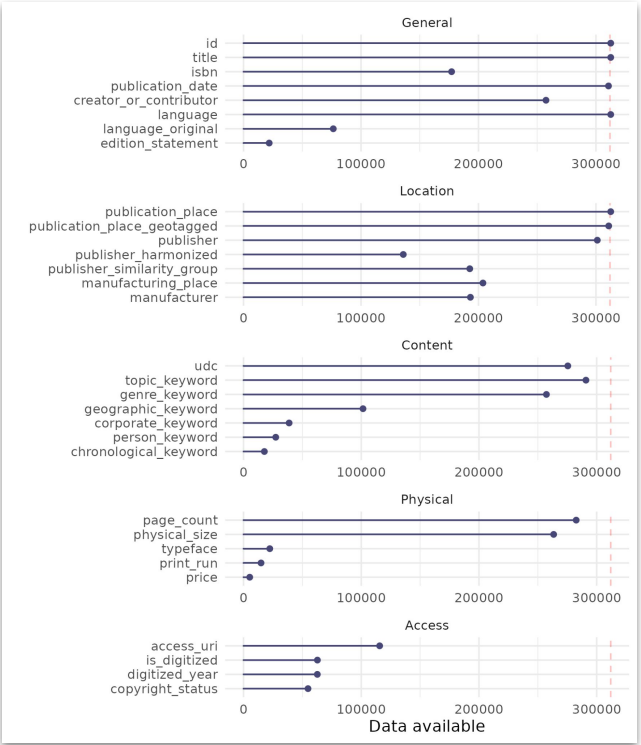
Data collector: National Library of Estonia

This curated dataset is derived from the books subset of the Estonian National Bibliography (ENB), a comprehensive catalog of publications written in Estonian, published in Estonia, or focusing on Estonian culture and people. Designed for computational analysis, this dataset adapts the original catalog for research and cultural exploration.

Through a systematic process of filtering, cleaning, and harmonizing, the ENB dataset is presented in a streamlined tabular format that retains rich metadata while improving accessibility. Fields selected for inclusion are harmonized and, where possible, linked to external sources, offering an optimized and reproducible resource for historical, cultural, and bibliographic research.

Files (188.2 MB)

Name	Size	
enb_books.tsv	188.1 MB	Download
md5:510a1b589be9dabb2846862b9449f007		
README.md	21.6 kB	Preview  Download
md5:52f6f01bb3b73e6b824a64e7345d50		





# Kasutamine + populariseerimine

## Blogi



Digitaalsed rännakud ehk  
Kultuuripärandi kujutamisest  
kaartidel

Kultuuripärandi kujutamine

interaktiiv

10.06.2024

novaaator

UUDISED

MEEDIAAIIP

TEADUSLUULE

ERIPROJEKTID



Eesti tõlkeajaloo digitalaktika  
toob ilmsiks kultuuri ja  
ühiskonna alusko

Eesti tõlkekirjanduse võrgustiku analüüs



Kes oli rahvusliku ärkamisaja  
keskmes?

Ärkamisaja kirjastamisvõrgustiku  
silmapaistvad isikud ja keelevalikud.



Kui palju on eesti keeles ilmunud  
raamatuid?

Ülevaade sajanditepikkusest kirjastamise  
valikutest.

30.11.2023 » Juhtumiuuringud

## Eesti tõlkeajaloo digitalaktika toob ilmsiks kultuuri ja ühiskonna alusko

KULTUUR

novaaator

UUDISED

MEEDIAAIIP

TEADUSLUULE

ERIPROJEKTID

Maris Saagpak, Daniele h  
Kruusmaa (RaRa)  
01.04.2024 13:51

Andmeteadlane vastab: kes oli rahvusliku  
ärkamisaja keskmes?

KULTUUR

Krister Kruusmaa (RaRa)  
08.01.2024 10:41

Digihumanitaar vastab: kui palju on eesti keeles  
raamatuid ilmunud?

KULTUUR

Peeter Tinitis (Rahvusraamatukogu, TÜ)  
23.11.2023 14:15



# Kasutamine + populariseerimine



## Valik lõputööde teemasid:

### OCR kvaliteedi hindamine ja parandamine

Optiline märgituvastus (Inglise keeles *Optical Character Recognition* ehk *OCR*) on trükitud või käsitsi kirjutatud teksti teisendamine digitaalseks tekstiks. RaRa materjalide seas leidub erineva OCR kvaliteediga andmeid. Potentsiaalsed uurimisteemad oleksid järgmised:

+ OCR-I kvaliteedi hindamine (ajalehtede täistekstid)

+ OCR-I kvaliteedi parandamine (ajalehtede täistekstid)

### Kultuuriteemad

RaRa andmekogud pakuvad mitmekülgseid võimalusi kultuuri uurimiseks. Täistekstide kõrval talletatakse ka metaandmeid ehk andmeid andmete kohta, need sisaldavad teose infot (autor, aasta, asukoht jne). Näiteks on võimalik uurida:

+ Tutvumiskoolutuste analüüs (ajalehtede täistekstid)

+ Eesti hstorigograafia ERB-ls (ERB-I metaandmed)

+ Eesti kirjandusklassika (ERB-I metaandmed)

# Aitäh!

Uuri koos meiega!

<http://digilab.rara.ee>

