

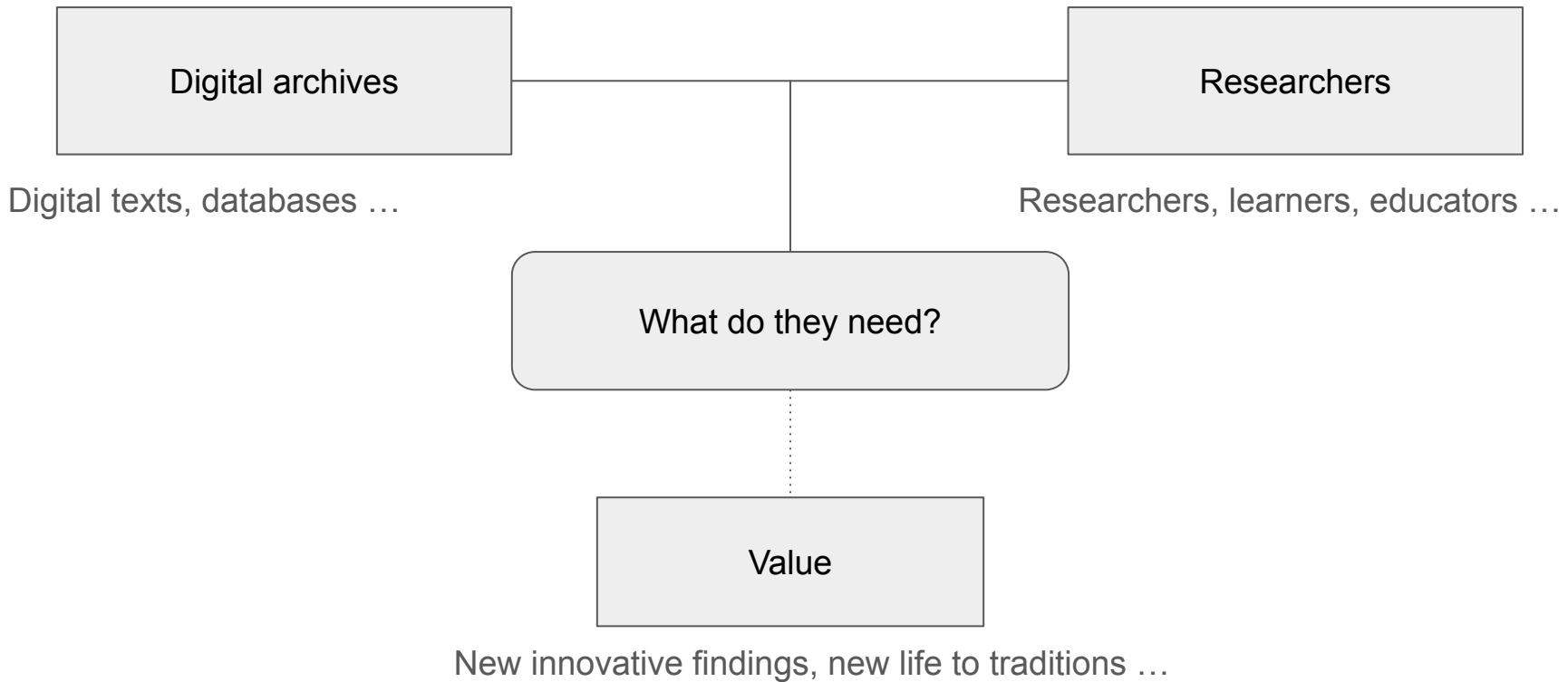
# Adapting newspaper and bibliographic data for historical research: tools and case studies in Estonia

Peeter Tints  
08.02.2024  
Helsinki Digital Humanities Research Seminar

# Outline

0. Thinking about researchers and archives
1. Digitized newspapers
  - a. What is in
  - b. Data explorers
  - c. Example cases
2. Bibliographic data
  - a. What is in it
  - b. Longer case study: emergence of Estonian written language community 1800-1940
3. Discussion

# The use of digital archives



# Thinking about the use

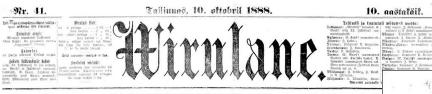
Perspective from data holders --- Perspective from researchers



UNIVERSITY OF TARTU  
Centre for Digital Humanities  
and Information Society



# 1. Digitized newspapers



JUVENILE READER. 37

Lying in bed, suffering great pain, day after day, and week after week, till he was worn quite thin, he began to recover so as to be able to hobble upon crutches.

23 After this long suffering, James could neither run nor play. With others, he could neither run nor play. With others, children, but he had to sit all day long, wishing that he had been more attentive to his mamma's request.

24 It was a great treat to Mrs. Cemetery to see her little son in this unhappy condition; and she often regretted that she had permitted him to grow up so obstinate and obstinate.

ABSURDITY OF PRIDE.

1. EVERY man, let his state and condition in life be what they may, depends on those around him for assistance and support.

Varastati jaalgratas 8. skp. oöl on Wöhmas warastatud Johan | Saewareste kuulutvast rauaõõstusest, Edu | jaalgratas. Waras on tulnud akna kaudu sisse ja tähkunud ukcest, seda seestpoolt awades.

Warastatud jaalgratas oli toodud töökolla | paigutusti. Paigutusti lätepael | seftelittonitabite lätepael Sutfin Lõõtimäe | maja küljel. Lõõtimäe maja |

# Brief historical overview



- 1993 microfilms, 2003 digitization.
  - First major newspapers Perno Postimees (1857-1885), Postimees (1886-1920), Pääwaleht (1905-1912)
- 2004 online indexed reading portal
- 2014 text search portal (85 publications, 100 000 pages)
- 2024 text search portal (total 2787 publications, 5.8M pages, 15.6M articles)
  - Full text access to open data
  - New portal in 2026 includes data mining access in the system

Digitization done by a number of institutions over 20 years following current priorities with technology of the time. Workflow standards have varied over time and across institutions.

Biggest collection is in the National Library, but data is in a few different digital archives

# Example case

You want to use newspapers to understand attitudes to nature in the 1920s.

## Full text access:

0 m., nahk- wõi ülem kalingorköites 1000 m. Eesti **loodus**. 8 iseseisvat artiklit Eesti ranniku, mere, geolo i wärwikiillasemat, meelearurikkamat ja sügawamat **loodus**ilu näinud. Ka kaasreisijate näod kõnelewad sedasa See üsna magus lause on aga wabandataw: on nähtud **loodust** Antoni kaudu, ja tema psühholoogiale lause wastab a oma office'ist wõi Büroost wälja soidab wabasse **loodusesse** uut jõudu koguma. Kes kaugemale sõita ei jõua ole suutnud köigutada isegei ilmasöda. Siin wabas **looduses** andub inglane täieliku innuga oma armsale lõbus rk. Füüsilise maateaduse ülesannete kogu, I wihk «**Looduse maatlussd**». Hind 85 markm Maateaduse ülesannete

Data balance  
(e.g. corpus size)

Data quality  
(e.g. OCR)

Representativeness  
(do we have enough?)

Study  
design

Materials to answer the question

# A user needs to know

A researcher would need to know

- What data is available (many archives)
- The quality of the data (varies a lot)
- Representativeness of the data (archive vs published)

Study  
design

Good to know this in the stage of planning & project-writing

# Learning by doing

Case studies to understand the researcher perspective

E.g. 6 case studies made by 12 students 2 instructors in U. Tartu

Feedback into building the interfaces & updating data

E.g. full text access toolkit, ngram reader

# What is in the dataset?

## General overview

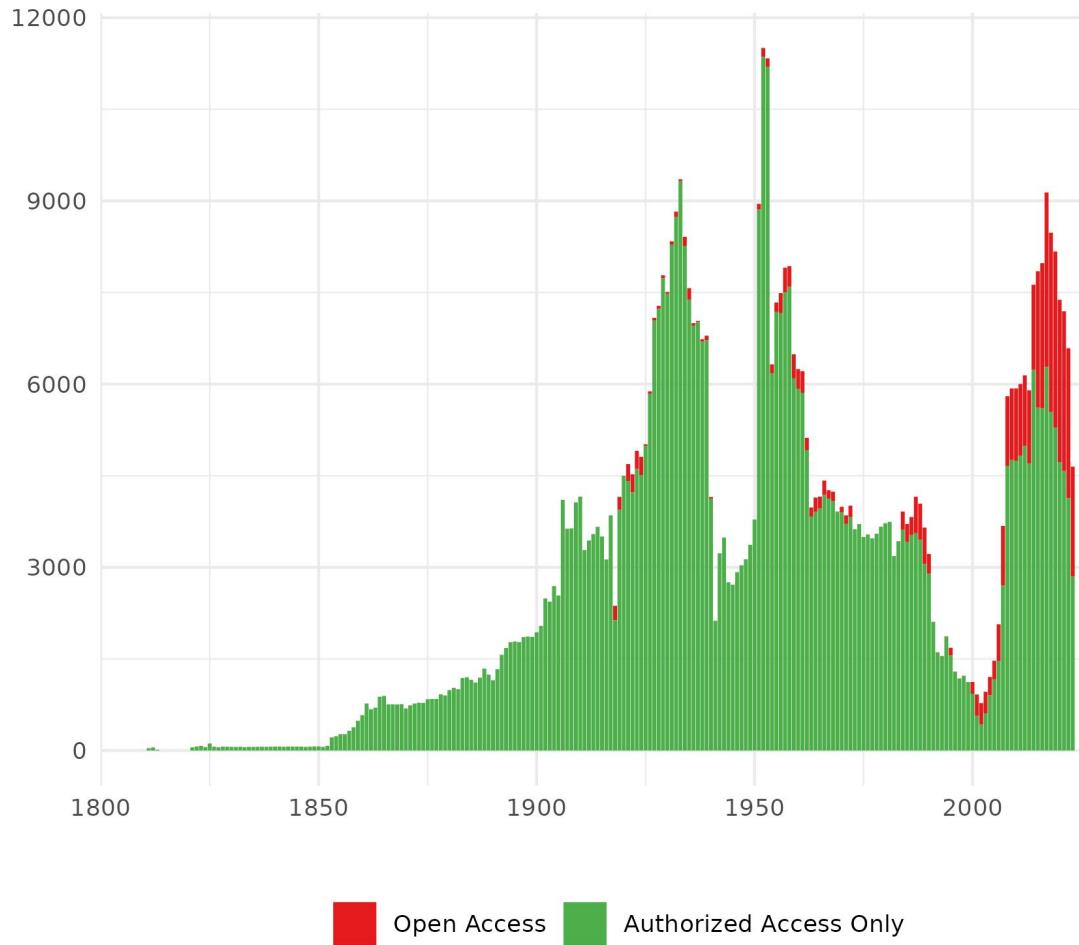
Distribution of issues over time  
(n = 622,061)

Green = open access (92.8%)

Red = authorized access (7.2%)

Interactive:

[https://peetertinitis.github.io/reports/nlib/dea\\_info.html](https://peetertinitis.github.io/reports/nlib/dea_info.html)



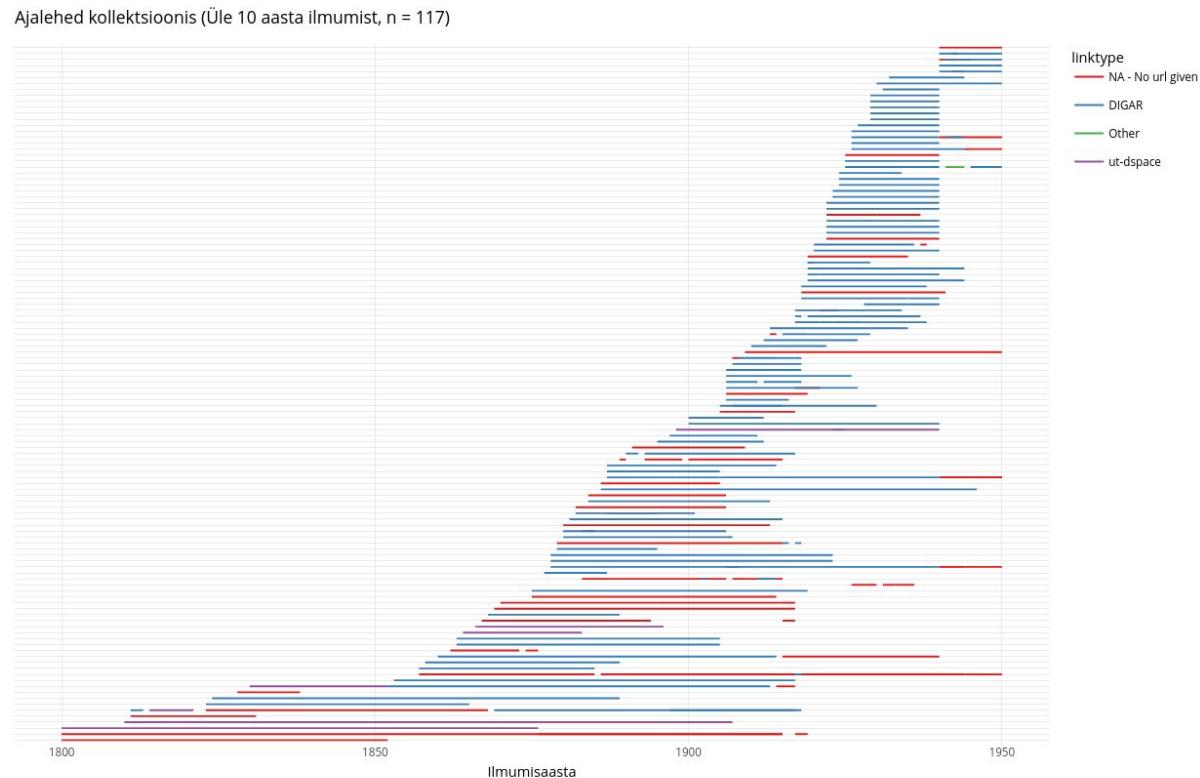
# Which newspapers have been digitized?

## Data in historical context

Data in National Bibliography  
that is linked to digital  
archives (with some links  
added here)

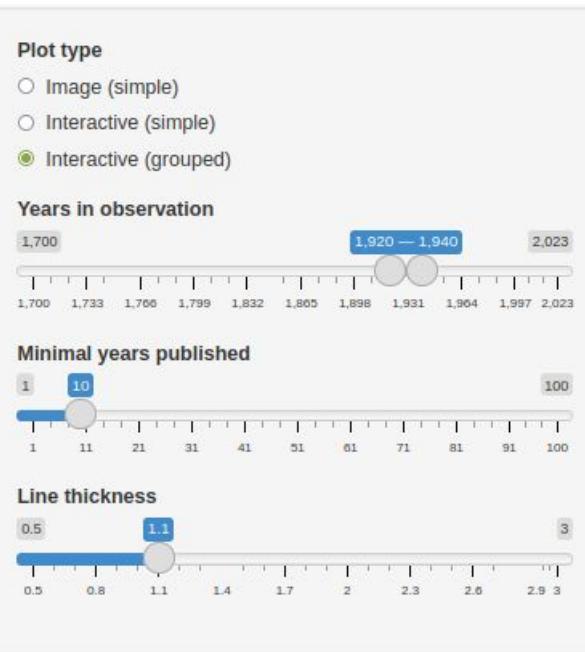
Tool to understand the  
coverage, select sources  
when designing a study

[https://digilab.shinyapps.io/digitized\\_newspapers2/](https://digilab.shinyapps.io/digitized_newspapers2/)

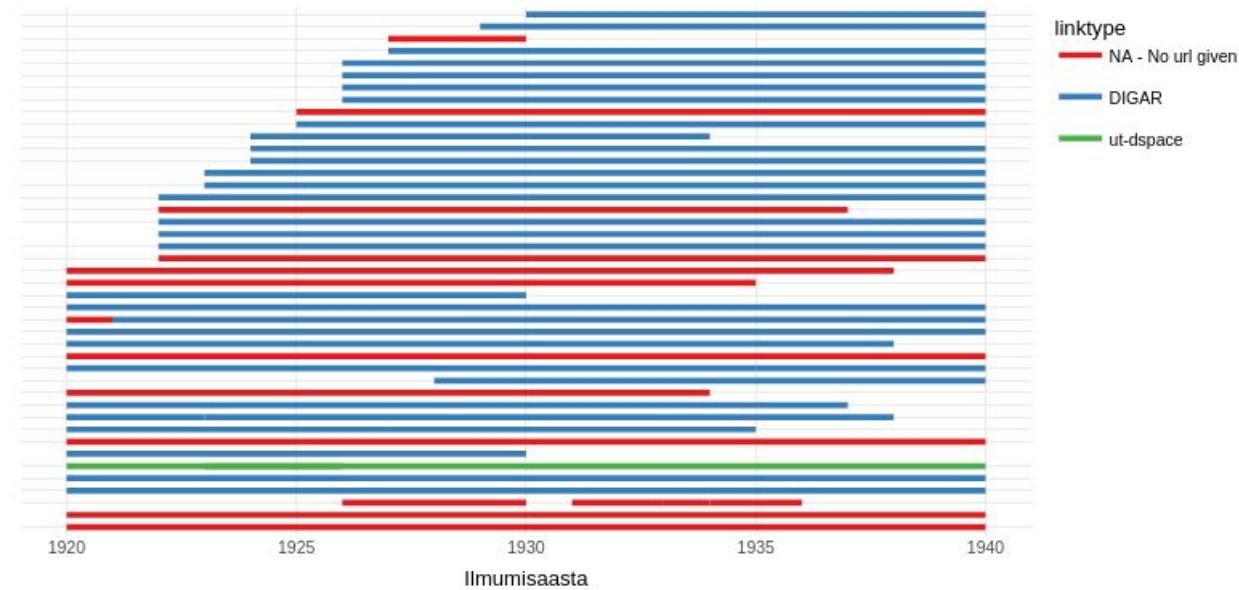


# Which newspapers have been digitized?

An example - newspapers from 1920-1940 published for at least 10 years



Ajalehed kollektsioonis (Üle 10 aasta ilmumist, n = 43)



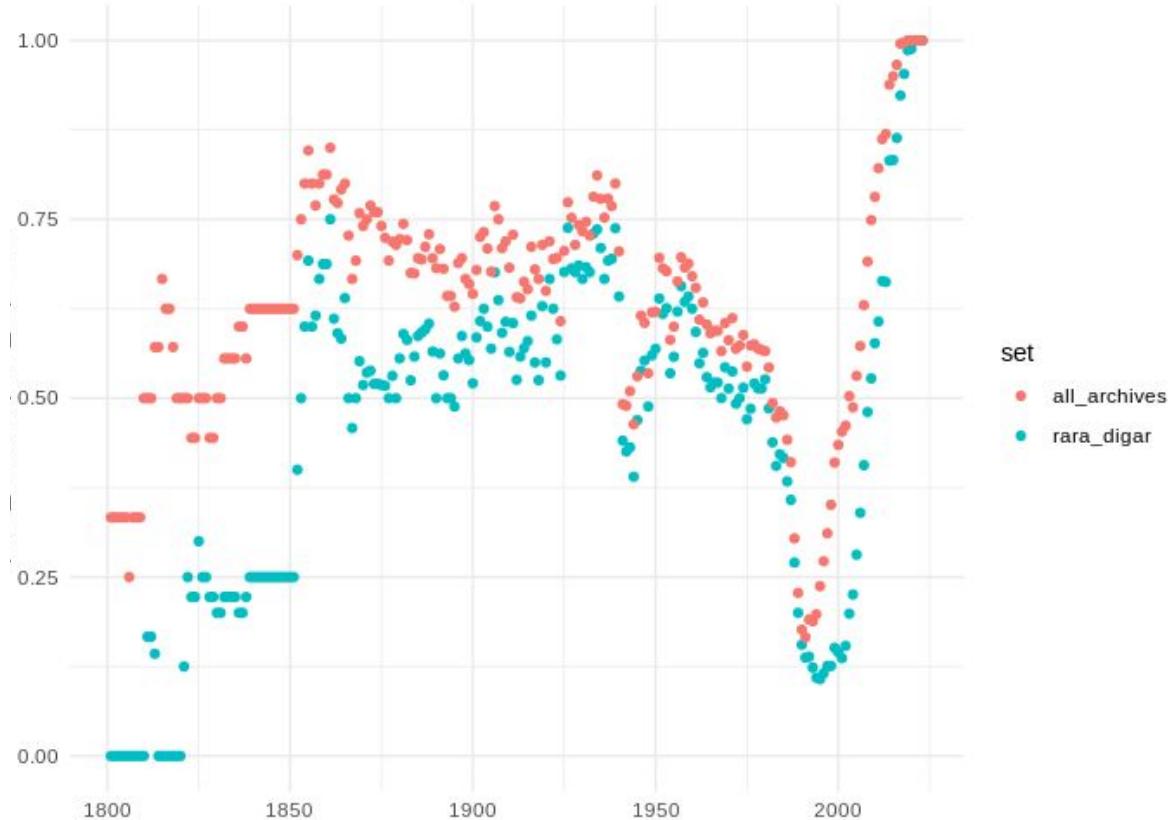
# How much has been digitized?

## Coverage of digitization

For each year, the share of newspapers in digital collections.

E.g. in 1900, 43 newspapers are known, 27 of which are digitized = 62.8%

1800-2020, 58.7% newspapers digitally available in some form



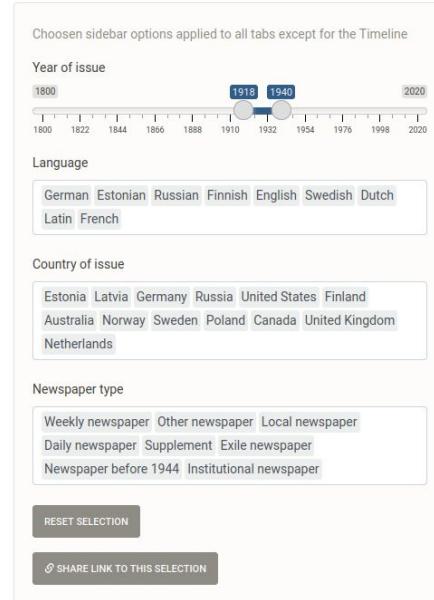
# What is in the dataset?

## Metadata explorer

Explore dataset with your own parameters in mind.

Interactive tool

[https://digilab.rara.ee/  
en/tools/newspapers-  
metadata-browser/](https://digilab.rara.ee/en/tools/newspapers-metadata-browser/)



## Explore DIGAR collection

HOME METADATA NEWSPAPER SEARCH TIMELINE LANGUAGES COUNTRIES TYPES

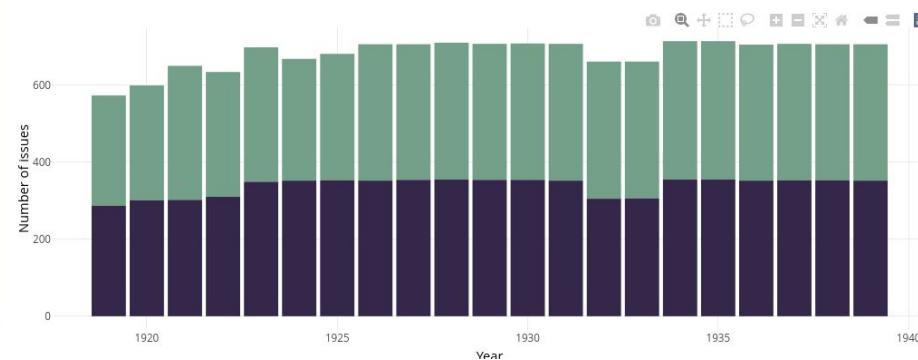
### Search by title

Here one or multiple newspapers can be searched by title and filtered using sidebar. The output displays the issues distribution on a timeline and newspaper's metadata in the table

All titles are included in the search, but if there is no output in the plot and the table, the sidebar options such as year range and language should be loosened. The table below shows summarised information for each newspaper.

Choose newspaper(s):

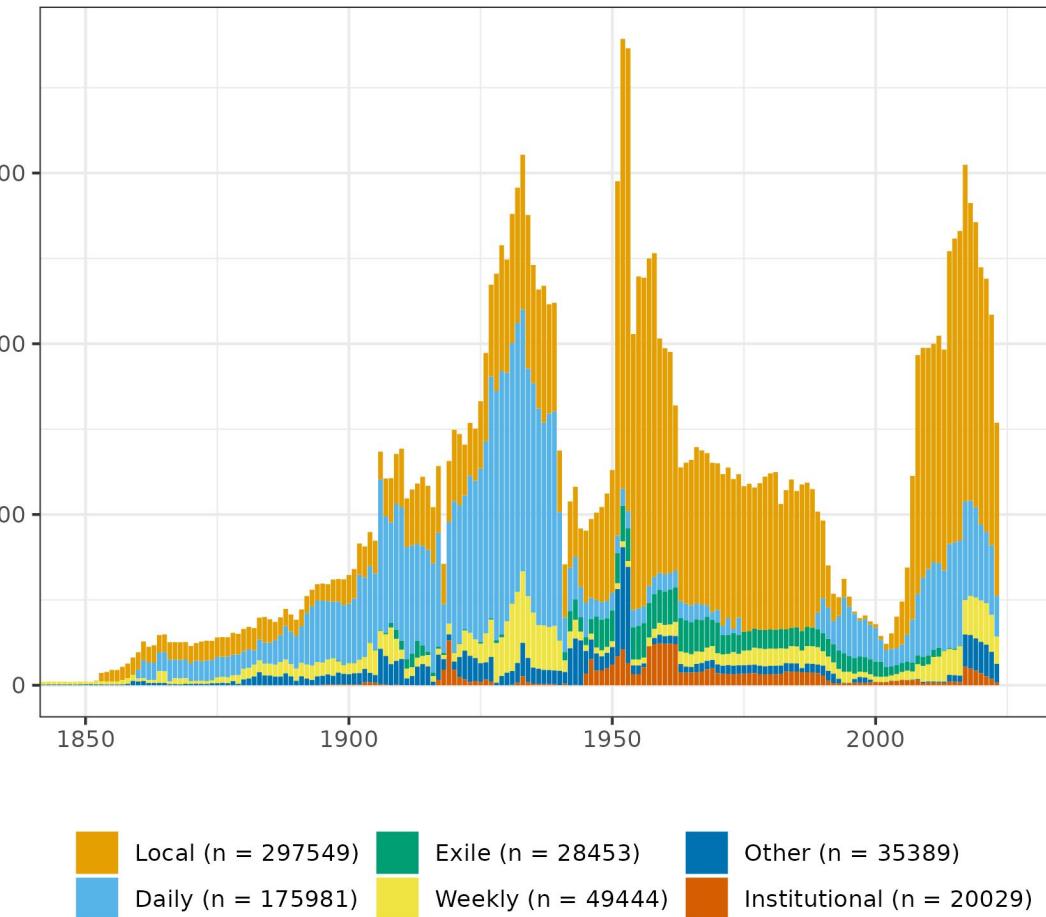
Päevalteht Postimees



# What's in the dataset?

## Overview by publication type

- Daily newspapers
- Weekly newspapers
- Local newspapers
- Exile newspapers
- Institutional newspapers



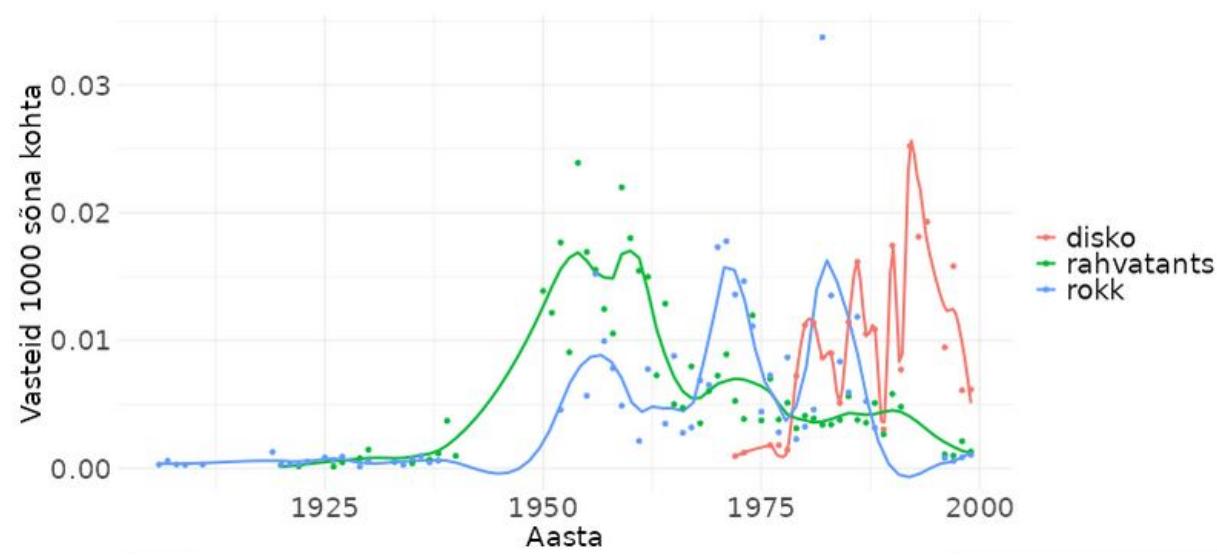
# What's in the dataset

## Ngram search

Quick peek at the text

Terms per 1000 words

Selected newspapers  
over the period to  
balance the dataset.



# Rahvusraamatukogu sõnamitmike loendaja

## Otsing

disko rokk rahvatants

## Töötlus

Sõnad (muutmata)

Lemmad (sõnade algvormid)

## Aastad vaatluse all



## Joone kõverus



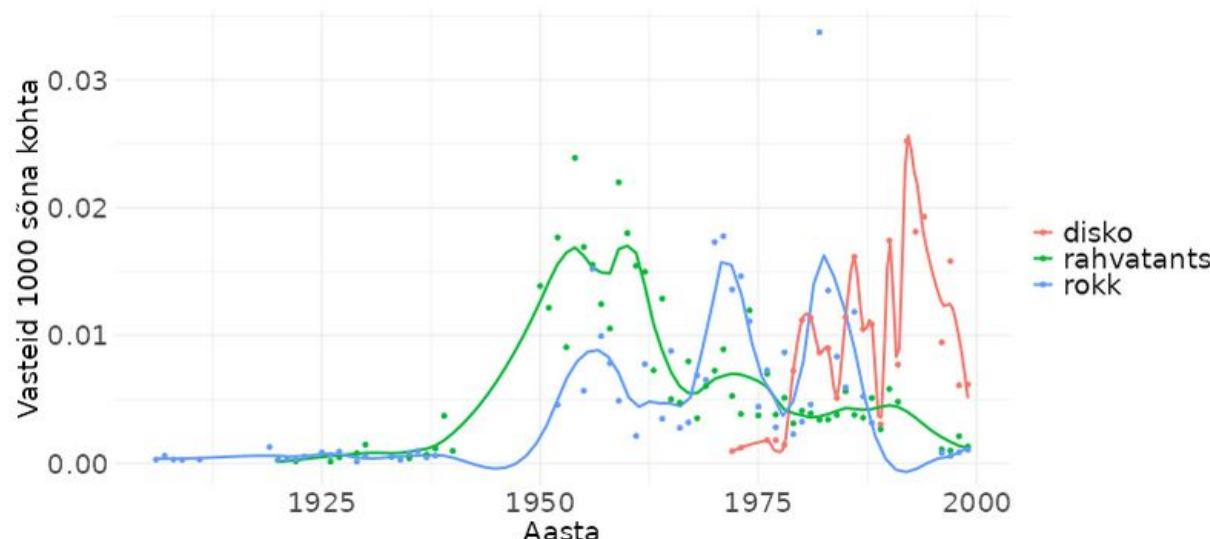
## Lõika joonisel y-telg ära siin



## Märgi joonisel iga mitmes aasta



**RaRa sõnamitmike loendaja** koondab sõnade ja sõnamitmike sagedusi Rahvusraamatukogu ajalehekollektioonides. Sisestades ülal otsingusse mõne sõna, näeb paremal joonisel selle sagedust kollektioonides 1857-2023. Terminite eristamiseks kasuta koma, tab-i või enterit. Otsides sõnamitmikku eralda sõnad selle sees tühikuga. Kui teie otsitud sõna ei kuvata joonisel, ei ületanud ta vajalikku künnist, et



Näita kirjeid 5 kaupa

Otsi (regex lubatud)

Sõna/Sõnamitmik	Tüüp	Mitu korda kokku	Mitmes dokumendis	Mitu sõna
disko	All	All	All	All
55904 disk	lemma	3176	2519	1
85010 disk	word	1991	1676	1
99895 diskor	lemma	1652	1127	1
134944 retrodisko	lemma	1159	593	1
135147 retrodisko	word	1157	608	1

# Full text access

JupyterHub access to freely work with data & download if needed.

Sign in

Username:

Password:

Sign In

## Server Options

Select a job profile:

Default: 1 CPU core, 8 GB memory, 6h timelimit

Start

## Elektriuuringu kood

Me kasutame uuringus Digiteeritud Eesti Artiklite andmekogu Eesti Rahvusraamatukogus. Ligipääsu vahendid on laiemalt lahti kirjeldatud [siin](#).

Kõigepealt loeme sisse vajalikud paketid `tidyverse` ja `tidytext` ning DEA tekstidega töötamiseks loodud `digar.txts`. Neil pakettidel on

```
[1]: suppressPackageStartupMessages(library(tidyverse,lib.loc="/gpfs/space/projects/digar_txt/R/4.3/"))
suppressPackageStartupMessages(library(tidytext,lib.loc="/gpfs/space/projects/digar_txt/R/4.3/"))
suppressPackageStartupMessages(library(digar.txts,lib.loc="/gpfs/space/projects/digar_txt/R/4.3/"))
suppressWarnings(dir.create("plots"))
```

Andmestiku olevatest materjalidest saab ülevaate käsuga `get_digar_overview()` `digar.txts` paketist. Tulemuseks on tabel, kus on kõik

## Andmed

```
[2]: all_issues <- get_digar_overview()
[1] "Issue metadata read"
```

Praeguses uurimuses huvitavad meid Postimehe numbrid vahemikus 1880-1940. Teeme töötamiseks vastava alamhulga

```
[3]: subset <- all_issues %>%
  filter(DocumentType=="NEWSPAPER") %>%
  filter(year>1880&year<1940) %>%
  filter(keyid=="postimeesew")
#fwrite(subset, "data/subset_postimeesew1.tsv", sep="\t")
```

Sellele alamhulgale võtame eraldi välja ka metaandmestiku iga artikli kohta, mis valimi ajalehenumbrites on.

```
[4]: subset_meta <- get_subset_meta(subset)
#fwrite(subset_meta, "data/subset_meta_postimeesew1.tsv", sep="\t")
```

Valimist ülevaate saamiseks vaatame kui palju on igal aastal valimis 1) artikleid ja 2) sõnu.

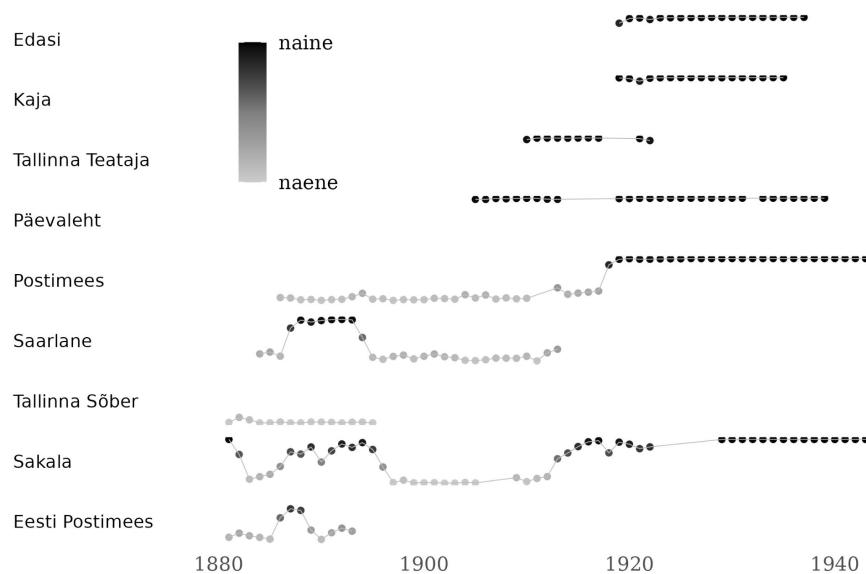
# Full text access

This allows more deliberate studies

E.g. the use of linguistic terms of interest in different newspapers.

Can look for relative frequencies of ngrams within newspapers.

~ Impact of prescription and/or editors on language.



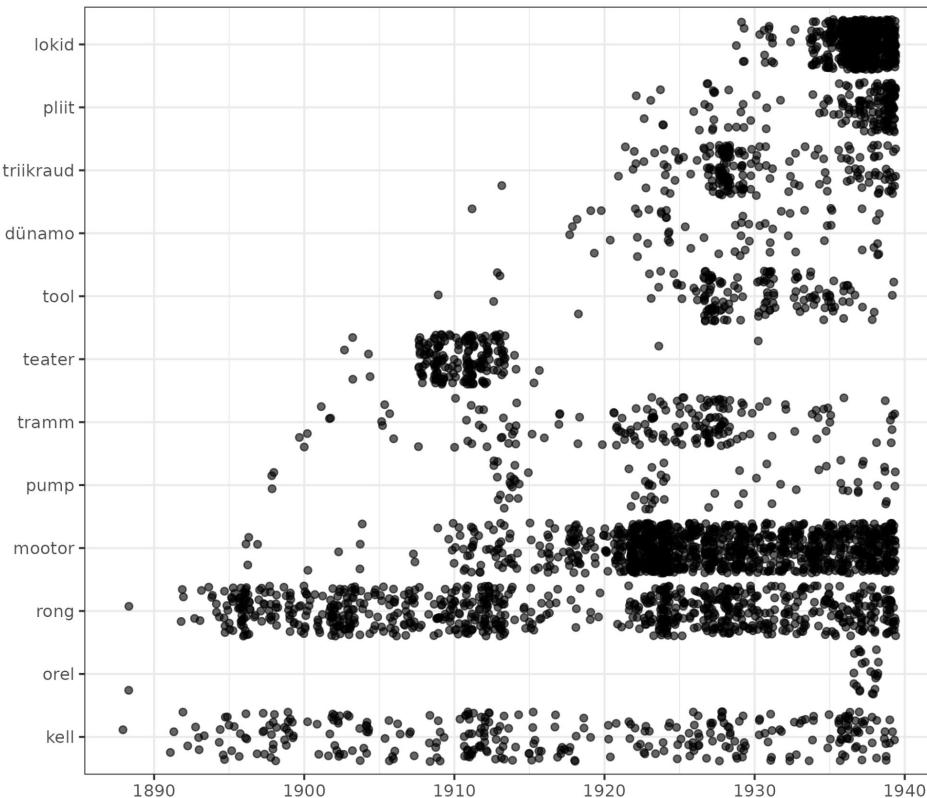
# Full text access

This allows more deliberate studies

E.g. the distribution of electricity related products in early 20th century.

Can look for electricity related terms, classify them into types, and select the meaningful ones.

~ History of technology and advertisement or popular discourse



# Quality of materials

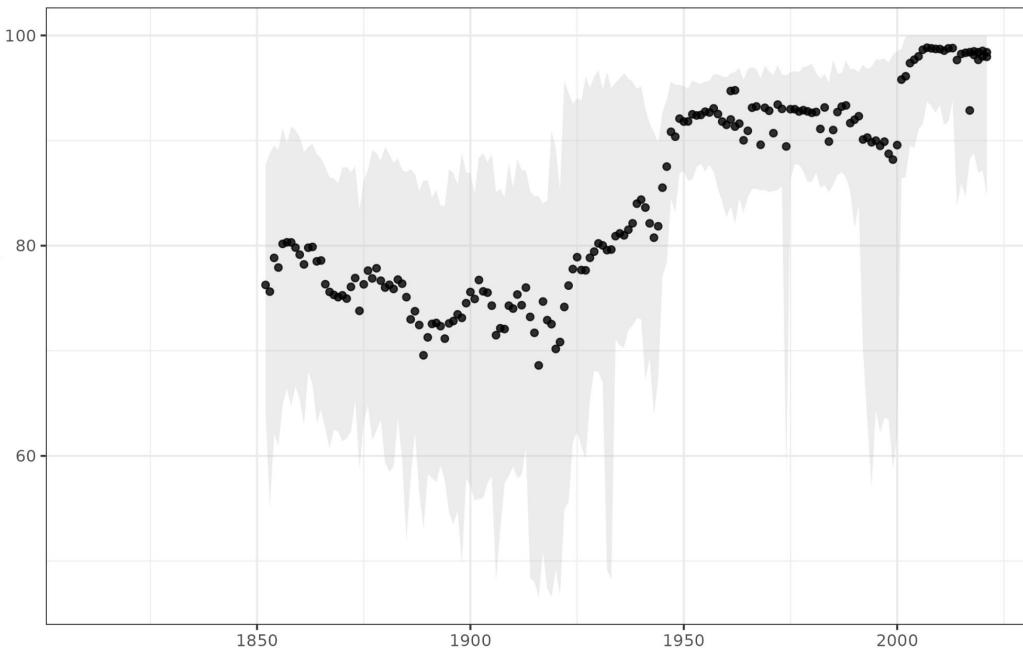
Mean OCR Confidence + spread of 95% of articles

## Data quality differs.

- Earlier digitized works have poorer quality.
- Quality not evaluated during digitization.
- Since ~2014, OCR Confidence stored for articles.

Working to evaluate now.

- Possibilities to auto-correct?
- May need to redigitize either from image or from scratch.
- Problems for some tasks, not others.



# Digitized newspapers

Keeping the researcher in focus:

- Overviews
- Access
- Data explorers
- Data in context

Current topics:

- Data quality

## 2. Bibliographic data

- luggermischeshed ic. v. D. W. Masing. 1821.  
 Oppetus fulta nnd luggermische lehti fassu  
 ga pruufda, v. D. W. Masing. 1821.  
 ABD. Duchi. v. D. W. Masing. 1823.  
 Tätilin. ABD. Ramat. v. D. W.  
 Masing. 1823.  
 Tätilin. Ristiauso oppetus n. v. D. W.  
 Masing. 1825.  
 Palousu. Ramat. Wönni kihelkonna Pä  
 foot last tuluer. v. E. Ph. Kieber. 1827.  
 2) Manuscrite.  
 • Muoro Juuri S. Andres Schreibbuch.  
 1814.  
 • Tamme Andres S. Jakobs Schreibbuch.  
 1814.  
 • Andres u. Jakob's Stylübungen. 3 Bde.  
 1815 u. 16.  
 • Wenne ABD. ja luggermische ramat, von  
 d. h. 1815.  
 • Andres u. Jakob's Geographie, nach Gas-  
 paris erstem Kurius. 1816.  
 • Abram Holters Stylübungen. 2 Bde. 1816  
 und 17.  
 • Nurme. Lahem S. Karel's u. Nissenka

X. N e c h n e n k u n s t.

- 1) D r u c k s c r i f t e n .
- \* A r c o p l o d a m i s t e e s t A r c o m a i s t e . K u n  
P. H. 1806.
- \* A r c o m a i s t e - R a m a t , v. O. W. M a s i n g .
- \* A r o o - e s t A r c h e n d a m i s t e e s t e m p l i d ,  
A b e r . H o l s t e r . 1823.
- 2) M a n u s c r i p t .
- \* A r c o m a i s t e - r a m a t u . v. d. H. 181

X I. P á d a g o g i z .

- 1) D r u c k s c r i f t .
- M a - r a h w a l a k a s t e - k a s w a t a m i s t e u . S  
d i g t e n .
- 2) M a n u s c r i p t e .
- \* Z e i t s c r i f t , d e m e h n i c h n e n S c h u l ,  
h u n g s - u n d B i l d u n g s w e s e n gewidm .  
d. H. E r s t e H e f t . 1817.
- \* P l a n u n g e n S c h u l m e i s t e r , S e m i  
E s h e n , n e c h t e r G e s c h i c h t e d e s s e l  
O. W. M a s i n g . 1827.

<b>title_responsibility_statement</b>	<b>publication_place</b>	<b>publisher</b>	<b>publication_date</b>
on Hansen Tammsaare ; illustrerinud Herald Elma	Tallinn	Eesti Raamat	1974
erl Laar	Stockholm	Välis-Eesti & EMP	1997
nahil Makarov ; [kaas: E. Tali]	Tallinn	Eesti Raamat	1966
rl Buck ; inglise keelest tõlkinud A. Hanko ; [kujundanud...]	Tallinn	Eesti Raamat	1994
ti Luuleosuriiduse Selts Eesti Teaduste Akadeemiasa...;	Tallinn	Välgius	1993
ja Hamsun ; [tõlkinud Henrik Sepamaa ; kujundanud...]	Tallinn	Eesti Raamat	1995
xandre Dumas (junior) ; [õrutanust keelest tõlkinud ...]	Tallinn	Eesti Raamat	1992
Ja Pettit	Tallinn	Välgius	1993
rikkil Böll ; [saksa keelest tõlkinud M. Sirkel ; kujundanud...]	Tallinn	Eesti Raamat	1993
rina Tsetverjova ; [koostanud, eestindanud ja järeltõlg...]	Tallinn	Eesti Raamat	1994
metajad: Raivo Rähni, Sven Kiviöli, Ant Noormets,...	Melbourne	Eesti Meesulaalu Selts Melbourne'is	1994
ckman	Uppsala	Etu Sõna	1993
Ja Pettit	Tallinn	Välgius	1992
es Herriot ; [inglise keelest tõlkis Jaagup Alasto ; ka...]	Tallinn	Välgius	1994
an Jaik ; [illustrerinud H. Mugasto]	Tartu	Noor-Eesti	1934
n, Tammsaare	Tartu	Noor-Eesti	1917
er Brook ; inglise keelest tõlkinud M. Palli ; Eesti Teadu...	Tallinn	Eesti Raamat	1993
erl Laar	Stockholm	Välis-Eesti & EMP	1994
anut by Anne Vabarna ; toimnitsaneer by Paul Hagu-...	Kuopio	Snellman-instituutti	1995
		www.snellman.fi	

```
ind2="" tag="700">>  
" 51 11 11 11 11 11 11
```

">(BLBKB)000564341</mx:subfield>

">1783-1842</mx:subfield>  
">(TSNT)0000000120981326</mx:

```
ind2="1" tag="700">
">><--><--><--></mx:subfield>
">1783-1842</mx:subfield>
">(NLR)RU NLR AUTH 775430</mx:subfield>
```

=LDR 00376n a2200133n 4500  
=001 a11124799  
=003 ErRR  
=008 990405|n|adnnaabn|||||||\n|||  
=040 ||\$aErR\$best\$cErRR  
=043 ||\$csee  
=100 1|\$aAdoiaan. Julius \$d1912-1994.

=LDR 00589nz a2200181n 4500

=001 a11124805  
=003 ErRR  
=008 981223|n|adnnnaabn|||||||a|||||

=043 \\\$ccee  
=100 1\\\$aAdson, Artur.\\\$d1889-1977.  
=400 1\\\$aAdson, Karl Arthur.

=400 1\\$aAdson, Karl Arthur.  
=400 1\\$aАдсон, Артур. \$d1889-1977.  
-500 1\\$aBellmann, Rector \$opacysud \$d188

=670 \saEesti kirjanike leksikoni, 2000.  
=680 \siEesti luuletaja, näitekirjanik ja kriitik

# Research use

Thinking again in terms of research use <-> data

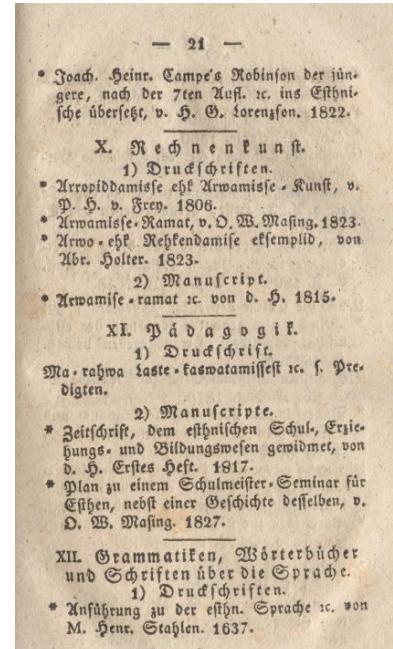
The researcher needs

- Access and usability
- Overviews of the data
- Connections with other datasets
- Representativeness of the data



# National bibliographies

- Register publications connected to a language and/or region
- Collect systematic data (titles, authors, formats, etc.) on printed publications
- Usually maintained by libraries and accessed within the library catalogue search

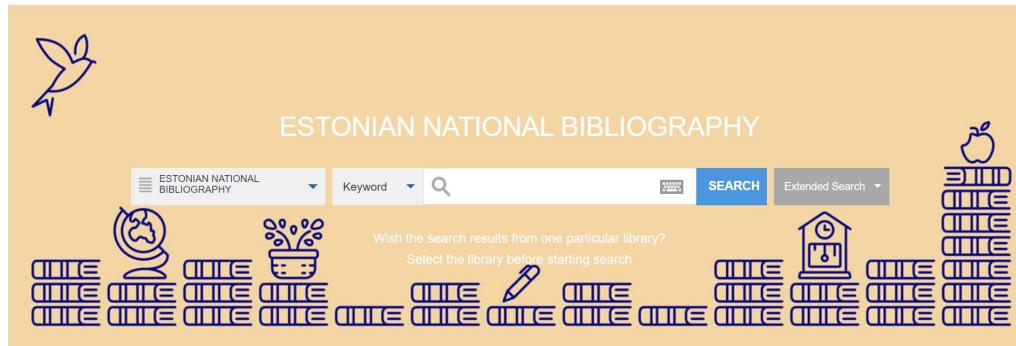


Excerpt from Rosenplänter's article  
“Übersicht der estnischen Literatur” (1832)

# Estonian National Bibliography

Consists of publications that are either:

- a. published in Estonia
- b. in Estonian
- c. by Estonians
- d. about Estonia or Estonians



<https://digilab.rara.ee/en/>

# Estonian National Bibliography

Origins in 1832: H. J. Rosenplänter aimed to collect an overview of all published in Estonian or about Estonia.

Then: 384 books listed in Estonian 1505-1832. Now known: 919 books 1505-1832.

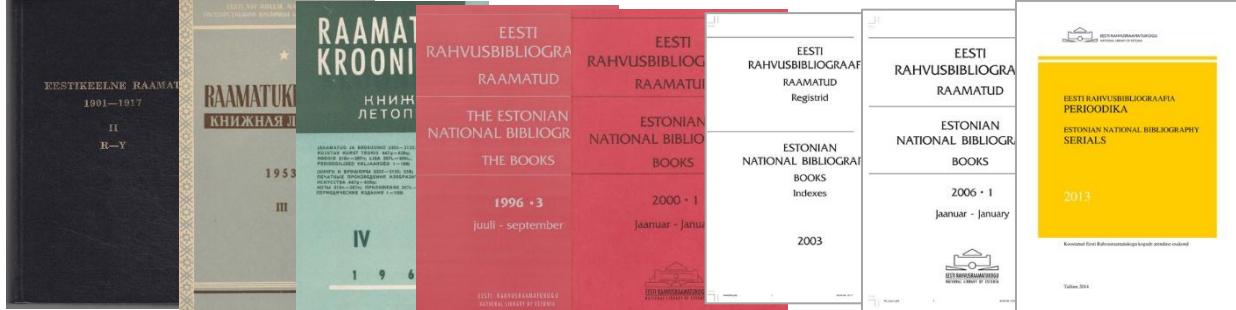
Followed by a number of comprehensive bibliographies throughout the period, since 2004 an electronic database.

- luggermisse leghet ic v. D. W. Masing. 1821.  
 Oppetus fulta ned luggermisse legh fastu  
 sa prufsta, v. D. W. Masing. 1821.  
 ABD. Buch. v. H. B. 1823.  
 Zeteline ABD. Namat ic v. D. W.  
 Masing. 1823.  
 Zeteline Ristissuo oppetur ic v. D. W.  
 Masing. 1825.  
 Paluswe Namat. Wodno kihellema Pa  
 sooli laste tulissu. v. E. Ph. Recke. 1827.  
 1) Manuscript.  
 • Nurme Juri S. Andres Scheibebuch.  
 1814.  
 • Lamme Andres S. Jakobs Scheibebuch.  
 1814.  
 • Andres u. Jakob's Stylsübungen. 3 Bde.  
 1815 u. 16.  
 • Wenn ABD. ja luggermisse. ramat, von  
 d. h. 1815.  
 Andres u. Jakob's Geographie, nach Gas-  
 paris ersten Kurfus. 1816.  
 Abram Holters Stylsübungen. 2 Bde. 1816  
 und 17.  
 • Nurme Abram S. Karel's u. Paluswe

X. Rechnenkunst.  
 1) Druckschriften.  
 • Arrosomise est Arwamise. Ku  
 P. h. v. Frey. 1806.  
 • Arwamise. Namat, v. D. W. Masing.  
 Armo est. Hethendamij esemplif  
 Kor. Holter. 1823.  
 2) Manuscript.  
 • Arwamise. ramat ic. von d. h. 181

XI. Pädagogik.  
 1) Druckschrift.  
 Ma. rohma laste. sotamamatissi ic.  
 digten.  
 2) Manuscript.  
 • Zeitschrift, dem ethnischen Schul-  
 ungs- und Bildungsweisen gewidmet  
 d. h. Erstes Heft. 1817.  
 • Plan zu einem Schulmeister. Sem-  
 Eben, nebst einer Geschichte dessel-  
 D. W. Masing. 1827.

### Excerpt from Rosenplänter's article “Übersicht der estnischen Literatur” (1832)



=LDR 00376n a220013n 4500  
=001 a11124799  
=003 EiRR  
=008 990405[n]dnnaabn|||||||\\|a|||||||  
=040 \\\\$eEiRR\\$b\\$eEiRR  
=043 \\$see  
=100 1\\$aAdojaan, Julius.\\$d1912-1994.  
=400 1\\$aАдояан, Юлиус.\\$d1912-1994.  
=670 1\\$aEesti elulood. Tallinn, 2000.  
=680 1\\$aEesti öigusteadlane Aleksander Adojaan vend.

```
=LDR 00589n a2200181n 4500
=001 a11124805
=003 ErRR
=008 981223|n|adn|naabn||||||||\a\\\|\\
=040 $aErR$rbest$cErRR
=043 $csee
=100 1$aAdson, Artur.$d1889-1977.
=400 1$aAdson, Karl Artur.
=400 1$aAdson, Arthur.
=400 1$aAdson, Karl Arthur.
=400 1$aАдсон, Артур.$d1889-1977.
=500 1$aBollmann, Peeter.$cpeusd.$d1889-1977.
=670 1$aEesti kirjaniku leksikon, 2000.
=680 1$aEesti luuletaja, näitekirjanik ja kriitik. M. Underi abikas
```

# The database format

- Has been maintained in Marc21, following library standards.
- Converted it into more compact tables for research purposes.
- Now as a .tsv in Zenodo
- We are working on a curated, simplified version of it.

-LDR 00376nz a220013n 4500  
-001 a11124799  
-003 ErRR  
-008 990405[admnnaabn]|||||||n|a|||||  
-040 [s]ErR[b]stest[s]ErRR  
-043 [s]See  
-100 1[s]Adojaan, Julius.3d1912-1994.  
-400 1[s]Adojaan, Künyyc.3d1912-1994.  
-670 [s]Eesti elulood. Tallinn, 2000.  
-680 [s]Eesti õigusteadlane, Aleksander Adojaani vend.  
  
-LDR 00589nz a220018n 4500  
-001 a11124805  
-003 ErRR  
-008 981223[n]admnnaabn]|||||||n|a|||||  
-040 [s]ErR[b]stest[s]ErRR  
-043 [s]See  
-100 1[s]Adson, Arthur.3d1889-1977.  
-400 1[s]Adson, Karl Arthur.  
-400 1[s]Adson, Arthur.  
-400 1[s]Adson, Karl Arthur.  
-400 1[s]Adson, Artyr.3d1889-1977.  
-500 1[s]Böllmann, Peter.Scpsied.3d1889-1977.  
-670 [s]Eesti kujanike, näitekirjanik ja kriitik. M. Under abiakasa Kasutanud pseudonüüm Peeter Böllmann.  
  
-690 [s]Eesti luuletaja, näitekirjanik ja kriitik. M. Under abiakasa Kasutanud pseudonüüm Peeter Böllmann.

title_responsibility_statement	publication_place	publisher	publication_date
Anton Hansen Tammsaare ; illustrerinud Herald Eelma	Tallinn	Eesti Raamat	1974
Mart Laar	Stockholm	Välis-Eesti & EMP	1997
Mihhail Makarov ; [kaas: E. Tali]	Tallinn	Eesti Raamat	1966
Pearl Buck ; Ingliste keelest tõlkinud A. Hanko ; [kujundus..]	Tallinn	Eesti Raamat	1994
Eesti Loodeuseurjate Selts Eesti Teaduste Akadeemia...;	Tallinn	Valgus	1993
Knut Hamsun ; [tõlkinud Henrik Sepamaa ; kujundanud..]	Tallinn	Eesti Raamat	1995
Alexandre Dumas (junior) ; [prantsuse keelest tõlkinud..]	Tallinn	Eesti Raamat	1992
Linda Pettì	Tallinn	Valgus	1993
Heinrich Böll ; [saksa keelest tõlkinud M. Sirkel ; kujund..]	Tallinn	Eesti Raamat	1993
Marina Tsvetajeva ; [koostanud, eestindanud ja järelld..]	Tallinn	Eesti Raamat	1994
tommetajad: Raivo Rähni, Sven Kiviöli, Ants Noormets.....	Melbourne	Eesti Meeslauku Selts Melbourne's	1994
U. Ekman	Uppsala	Elu Sõne	1993
Linda Pettì	Tallinn	Valgus	1992
James Herriot ; [inglise keelest tõlkis Jaagup Alasto ; ka..]	Tallinn	Valgus	1994
Juhan Jaik ; [illustrerinud H. Mugiago]	Tartu	Noor-Eesti	1934
A. H. Tammsaare	Tartu	Noor-Eesti	1917
Peter Brook ; Ingliste keelest tõlkinud M. Palli ; Eesti Tea...	Tallinn	Eesti Raamat	1993
Mart Laar	Stockholm	Välis-Eesti & EMP	1994
laulunut by Anne Valarna ; toimittaneet Paul Hagu..	Kuopio	Snellman-Instituuti	1995
Harald Suulepp ; [kujundanud R. Pangsoop]	Tallinn	Eesti Raamat	1994

```
<mx: datafield ind1="1" ind2="" tag="700">
<mx: subfield code="a">Stendhal,</mx: subfield>
<mx: subfield code="d">1783-1842</mx: subfield>
<mx: subfield code="0">(EBLNB)000564341</mx: subfield>
</mx: datafield>
<mx: datafield ind1="0" ind2="" tag="700">
<mx: subfield code="a">J.-H.-C. Stendhal</mx: subfield>
<mx: subfield code="d">1783-1842</mx: subfield>
<mx: subfield code="0">(ISNI)0000000120981326</mx: subfield>
</mx: datafield>
<mx: datafield ind1="1" ind2="1" tag="700">
<mx: subfield code="a">J.-H.-C. Stendhal</mx: subfield>
<mx: subfield code="d">1783-1842</mx: subfield>
<mx: subfield code="0">(NLR)RU NLR AUTH 775430</mx: subfield>
</mx: datafield>
```

# The data at the moment

- Comprehensively described with good coverage
- Non-harmonised
- A lot of pre-processing and cleaning to be useable

=LDR 00376nz a2200133n 4500  
+001 a11124799

+003 EiRR

+008 990405(n)adnnaabn|||||||\\|a|||||

+040 \\\$aEiRR\$best\$cEiRR

+043 \\\$ccee

+100 1\\\$aAdojaan, Julius.\$d1912-1994.

+400 1\\\$aAdojaan, Юлиус.\$d1912-1994.

+670 \\\$aEesti eluloop. Tallinn, 2000.

+680 \\\$iEesti õigusteadlane, Aleksander Adojaani vend.

=LDR 00589nz a2200181n 4500

+001 a11124805

+003 EiRR

+008 981223(n)adnnaabn|||||||\\|a|||||

+040 \\\$aEiRR\$best\$cEiRR

+043 \\\$ccee

+100 1\\\$aÅdsøn, Artur.\$d1889-1977.

+400 1\\\$aÅdsøn, Karl Arthur.

+400 1\\\$aÅdsøn, Arthur.

+400 1\\\$aÅdsøn, Karl Arthur.

+400 1\\\$aÅdsøn, Artur.\$d1889-1977.

+500 1\\\$aBöllmann, Peeter.\$cpseud.: \$d1889-1977.

+670 \\\$aEesti kirjanike leksikon. 2000.

+680 \\\$iEesti luuletaja, näitekirjanik ja kritik. M. Undri abikaasa. Kasutanud pseudonüümi Peeter Böllmann.

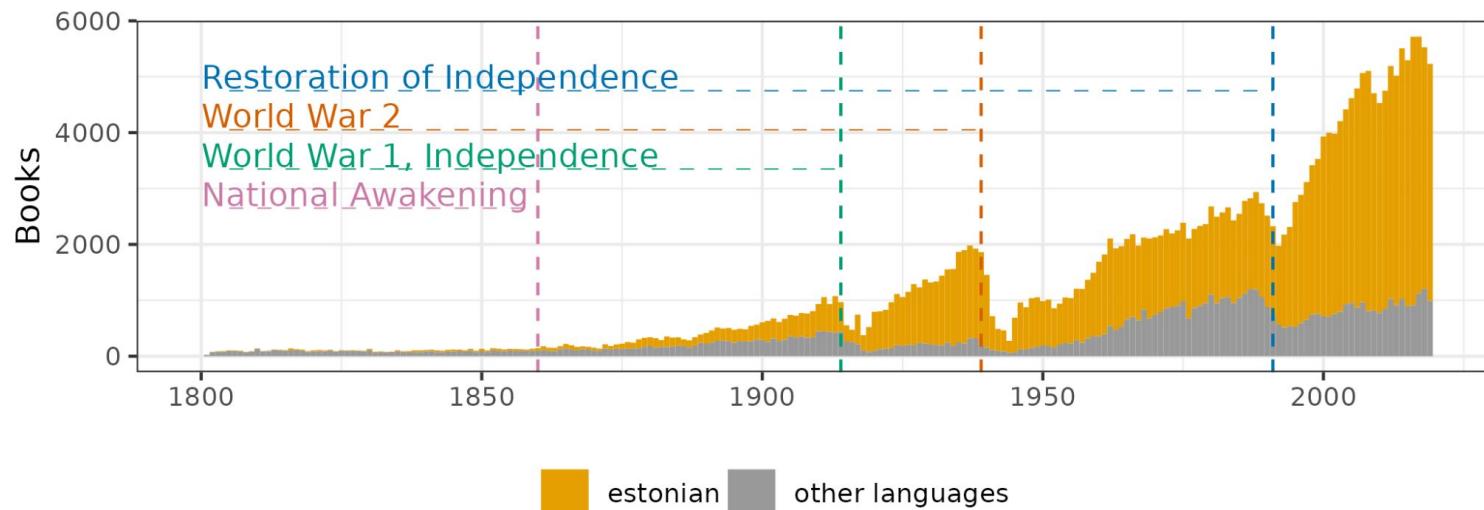
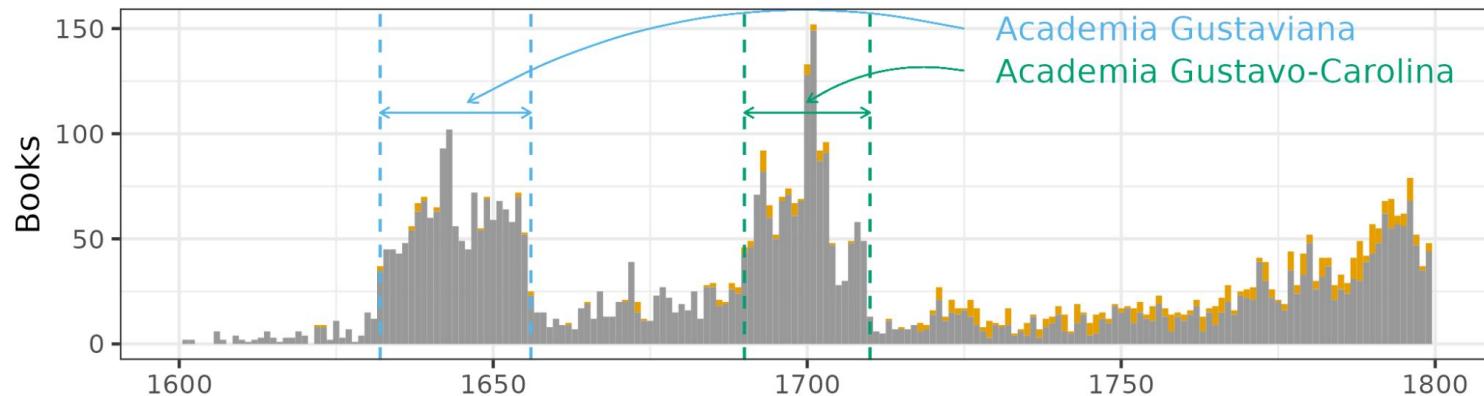
title_responsibility_statement	publication_place	publisher	publication_date
Anton Hansen Tammsaare ; illustreeritud Herald Eelma	Tallinn	Eesti Raamat	1974
Mart Laar	Stockholm	Välis-Eesti & EMP	1997
Mihhail Makarov ; [kaas E. Tali]	Tallinn	Eesti Raamat	1966
Pearl Buck ; inglise keelest tõlkinud A. Hanko ; [kujundatud Eesti Looduseuurjate Selts Eesti Teaduste Akadeemia järgi]	Tallinn	Eesti Raamat	1994
Knut Hamsun ; [tõlkinud Henrik Sepamaa ; kujundanud Linda Pettit]	Tallinn	Valgus	1993
Heinrich Böll ; [saksa keelest tõlkinud M. Sirkel ; kujundanud Marina Tsvetajeva ; [koostanud, eestindanud ja järelõppenud toimetajad: Raivo Rähni, Sven Kiviivali, Ants Noormets...]	Tallinn	Eesti Raamat	1994
U. Ekman	Uppsala	Elu Sõna	1993
Linda Pettit	Tallinn	Valgus	1992
James Herriot ; [inglise keelest tõlkis Jaagup Alasts ; kujundanud Juhan Jalk ; illustreeritud H. Mugasto]	Tartu	Noor-Eesti	1994
A. H. Tammsaare	Tartu	Noor-Eesti	1917
Peter Brook ; inglise keelest tõlkinud M. Palli ; Eesti Teaduste Akadeemia väljapaistvaid kirjandusteadlasi	Tallinn	Eesti Raamat	1993
Mart Laar	Stockholm	Välis-Eesti & EMP	1994
Iiaulannut by Anne Vabarna ; toimittaneet by Paul Hagu ; Harald Suulepp ; [kujundanud R. Fangsep]	Kuopio	Snellman-instituuti	1995
	Tallinn	Eesti Raamat	1994

```
<mx: datafield ind1="1" ind2="" tag="700">
<mx: subfield code="a">Stendhal,</mx: subfield>
<mx: subfield code="d">1783-1842</mx: subfield>
<mx: subfield code="0">(EBLNB)000564341</mx: subfield>
</mx: datafield>
<mx: datafield ind1="0" ind2="" tag="700">
<mx: subfield code="a">J. J. Rousseau</mx: subfield>
<mx: subfield code="d">1783-1842</mx: subfield>
<mx: subfield code="0">(ISNI)0000000120981326</mx: subfield>
</mx: datafield>
<mx: datafield ind1="1" ind2="1" tag="700">
<mx: subfield code="a">J. J. Rousseau</mx: subfield>
<mx: subfield code="d">1783-1842</mx: subfield>
<mx: subfield code="0">(NLR)RU NLR AUTH 775430</mx: subfield>
</mx: datafield>
```

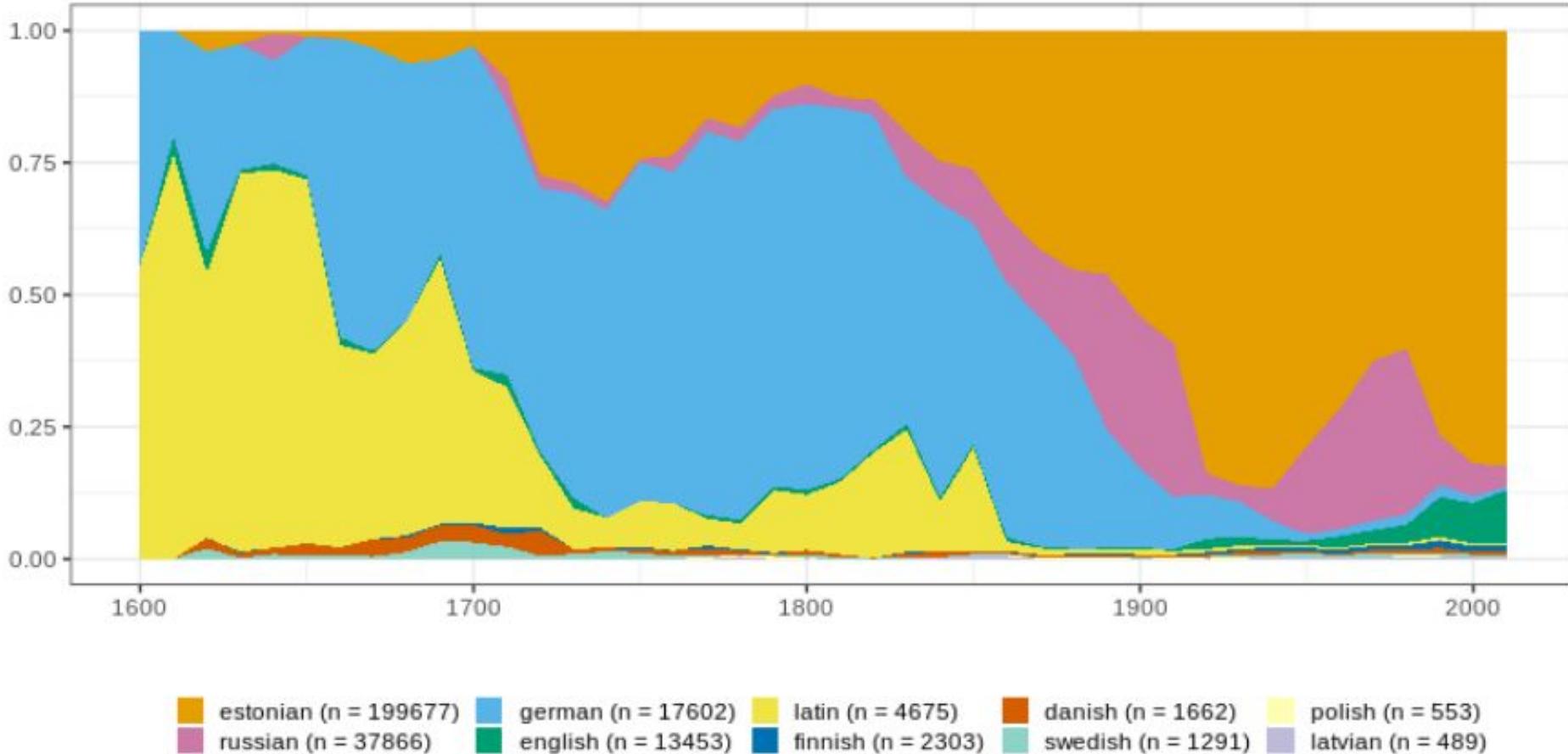
# A look at the data

# Books in ENB (1600-2020)

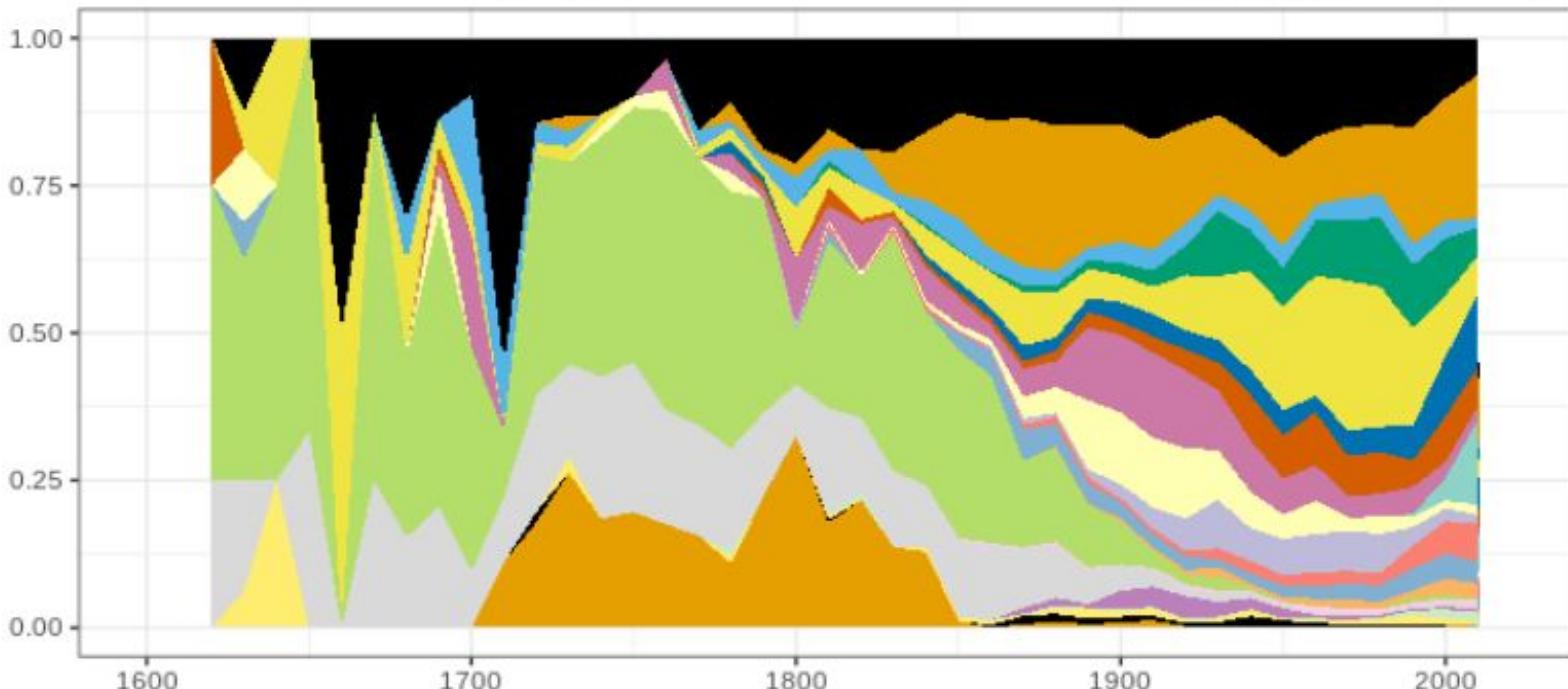
Some historical periods and events added for comparison



## Languages in Estonian National Bibliography (1600-2020)



# Genres of Estonian language texts in Estonian National Bibliography (1600-2020)

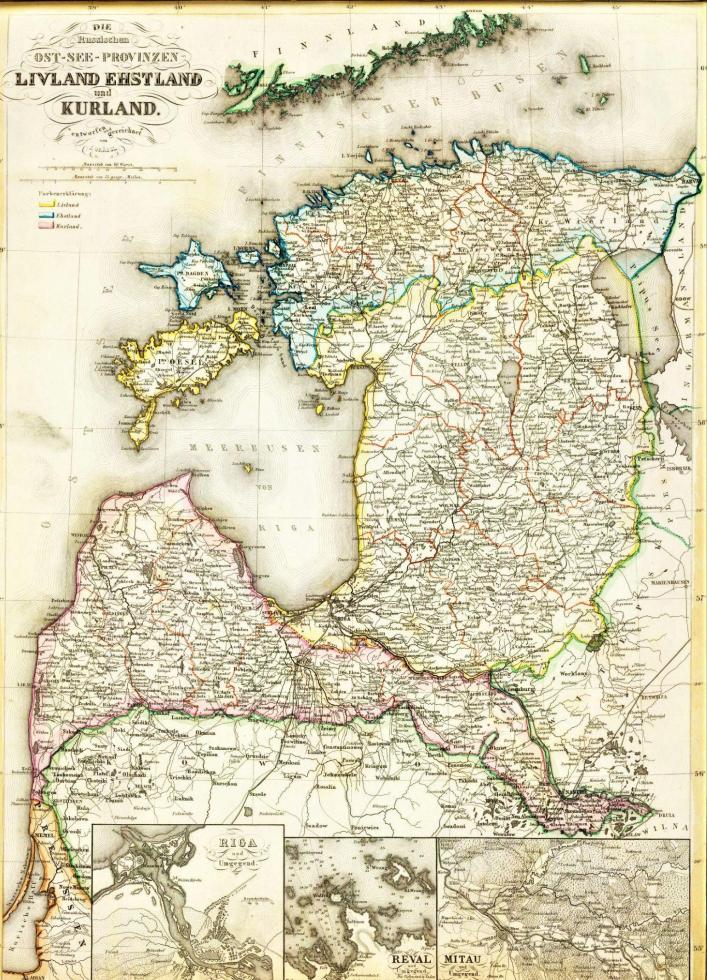


■	NONE (n = 31022)	■	manual (n = 15807)	■	poetry (n = 7884)	■	food (n = 2320)
■	fiction (n = 46596)	■	bureaucratic (n = 13074)	■	album (n = 4281)	■	dictionary (n = 1907)
■	OTHER (n = 7396)	■	online (n = 10890)	■	relig (n = 3749)	■	speeches (n = 1167)
■	scholarly (n = 19676)	■	letters (n = 9753)	■	travel (n = 2057)	■	calendar (n = 725)
■	education (n = 32370)	■	advertisement (n = 9798)	■	song (n = 2556)	■	games (n = 485)
■	youth (n = 18292)	■	biography (n = 9637)	■	drama (n = 2390)		

Case study: the emergence of the Estonian written language community 1800-1940.

# Crash course: social history of the Estonian language

- Very long (recorded) oral tradition
- Relatively new as a literary language
- Tied to social hierarchy:
  - From the Baltic crusades (13th century), German was established as the main language of commerce and culture in the region
  - Increasing overlap between linguistic and social divisions, which in turn reinforced each other
  - Using German was the prerequisite of almost any kind of social mobility
- (Slow) changes in the 19th century:
  - Liberation of peasants (≈ Estonians) from serfdom in 1816/19; additional social reforms during the 19th century, urbanization, industrial development
  - Enlighteners and estophiles among the Baltic German elite
  - National Awakening period (1850s-1870s)
  - Russification (1890s)
- “Coming of age”:
  - Modernist movements in the early 20th century
  - Official language of the Estonian Republic, est. 1918



The “Baltic provinces” of the Russian Empire, 1860  
(David Rumsey Map Collections)

# Changes to the language community 1800-1940

- From mostly oral and low prestige to a first written language community.
  - Literati focussed on Estonian community
- Mass literacy
  - 1800 - 66% could read, but it was mostly used for religious schooling
  - 1840s-1850s better reading comprehension and learning to write
  - 1897 - 91% read, 77% write
- Urbanization, newspapers, legal documents - written communication becomes common



Estonia & Livonia in the Russian Empire 1800

# European context

European written traditions are long, but mostly for the elite.

19c, mass literacy in some places. Huge educational reforms.

Estonian was in Scandinavian-German sphere of influence.

\* Measure not fully comparable

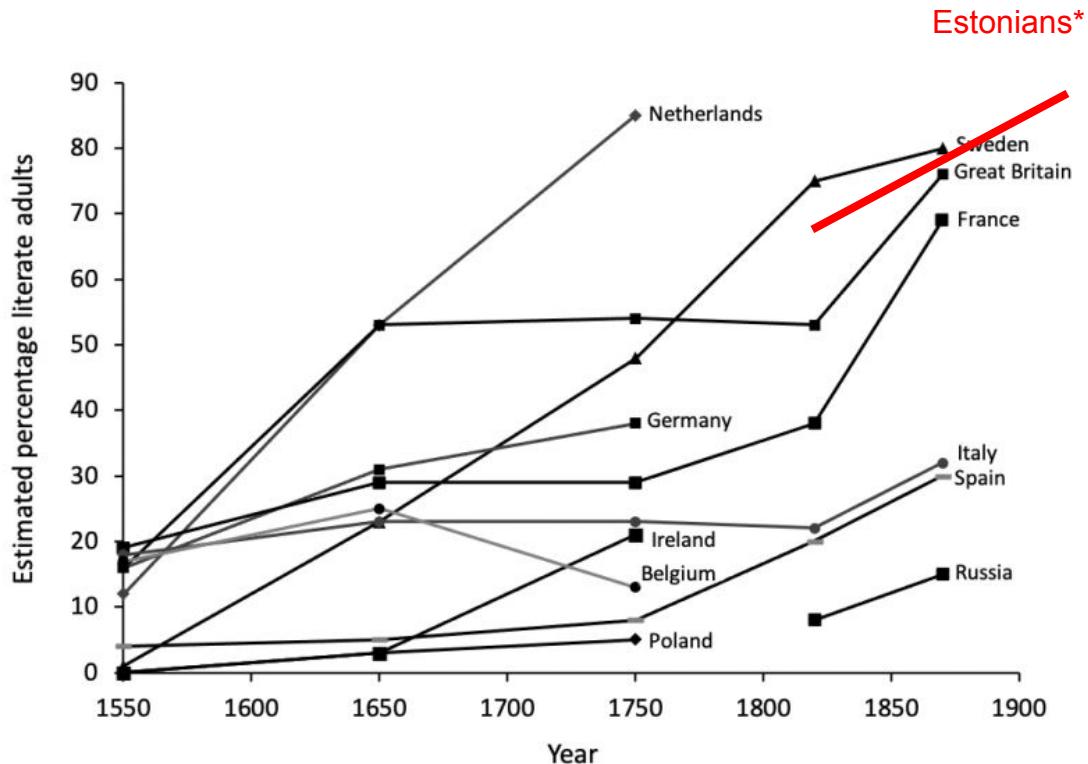


FIGURE P.1. Literacy rates for various European countries from 1550 to 1900. These estimates are based on book publishing data calibrated using more direct measures of literacy.<sup>8</sup>

Figure from Henrich 2020, data from Burigh & Zanden 2009

# The emergence of the community

How did the Estonian language community emerge from a German-dominated cultural area?

- Languages of publication and political events
- Influx of new authors
- Language choice for individuals
- Language choice in communities
- The role of schools for authors

# Estonian National Bibliography as a data source

Books published is a window to the language community

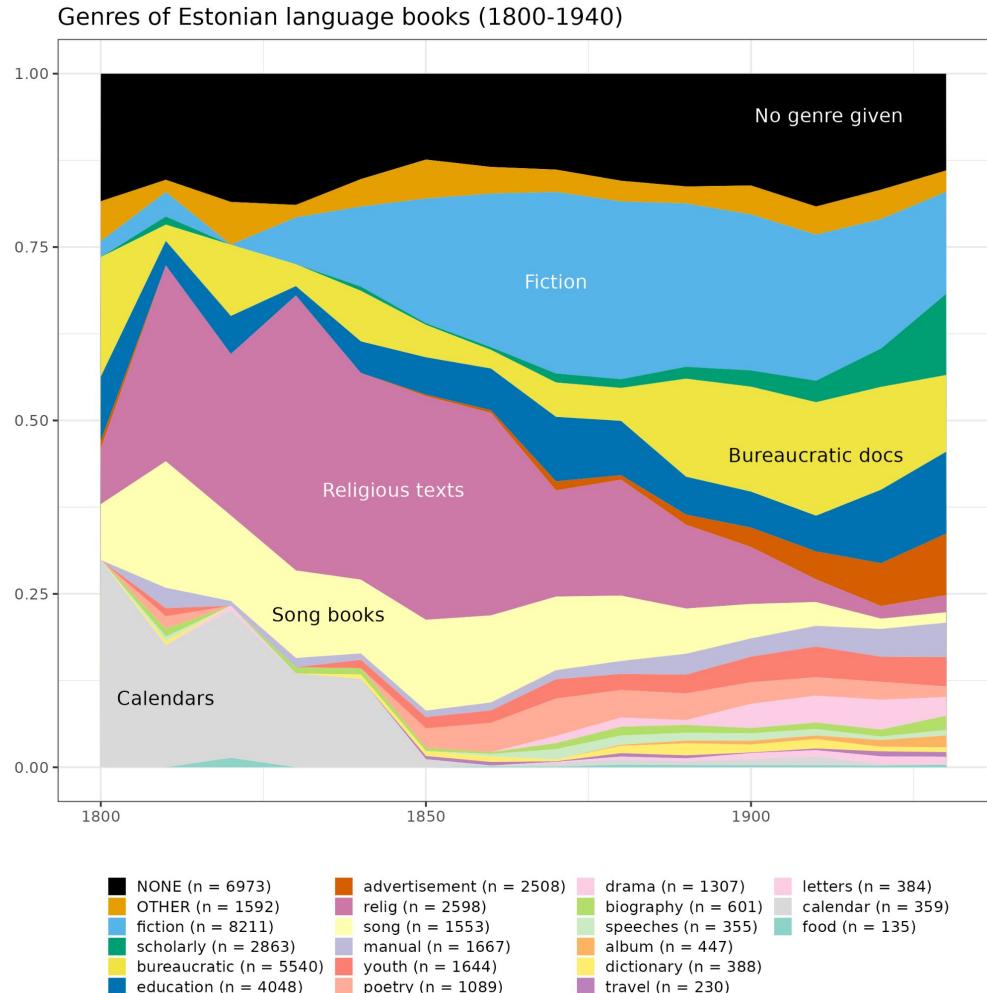
- Though not all of it. There's also newspapers, letters etc.

We can take books published in 1800-1940

- Includes books in Estonian and other languages
- Expected to have fairly good coverage of known books published
- Has data on authors and contributors to each book

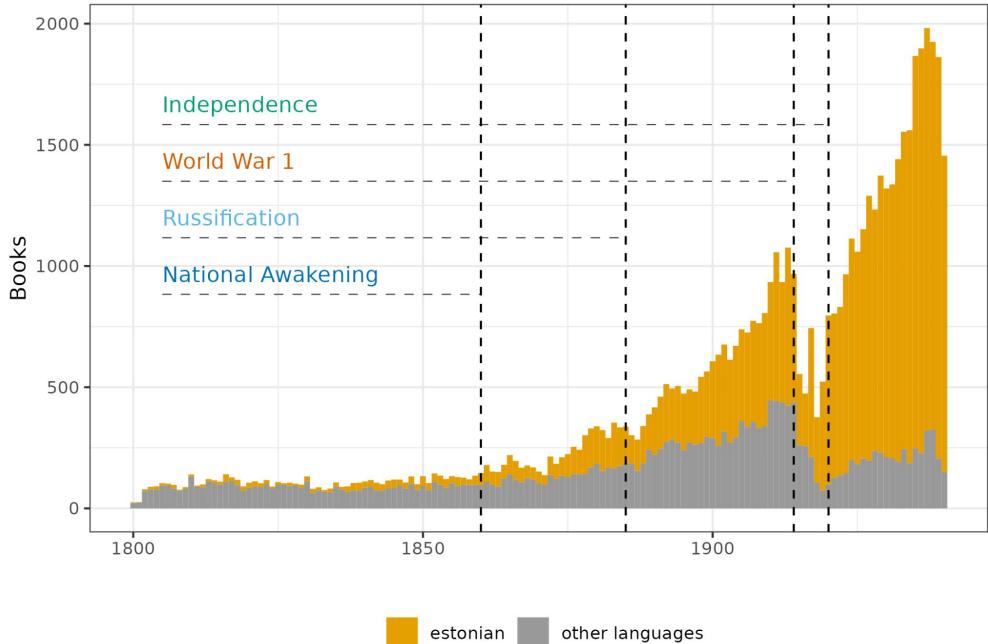
# Data

- 43,570 works
  - 25,986 Estonian (60%)
  - 17,588 other languages (40%)
- 9768 people
  - Authors, publishers, illustrators etc.
  - 5571 with birth dates (57%)
  - 711 multilingual authors (16%)
- Harmonized and cleaned author names and book metadata.
- Augmented metadata by detecting language from book titles.

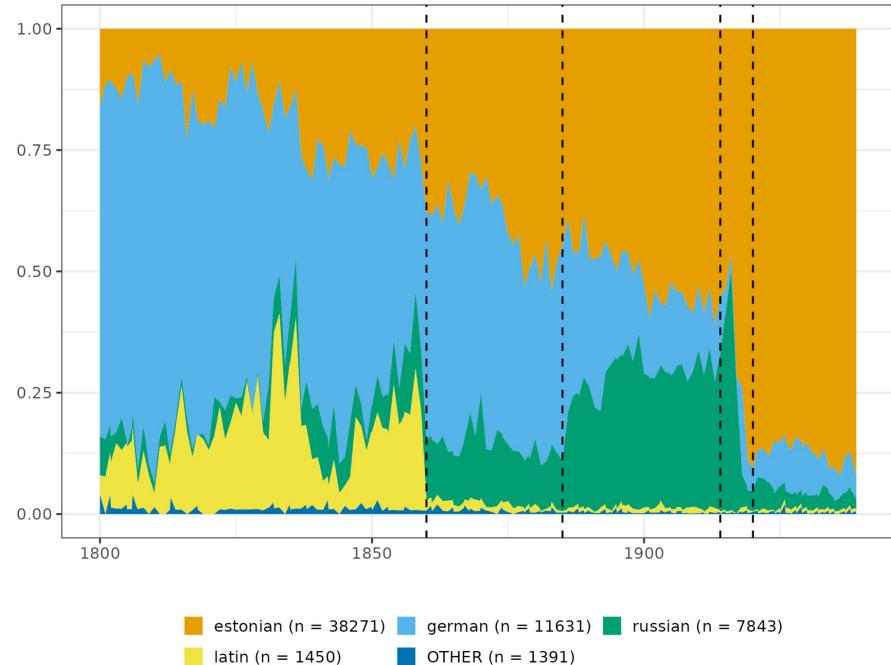


# Emergence of the Estonian literary community

Number of books in Estonian and other languages (1800-1940)



Share of language in books (1800-1940)



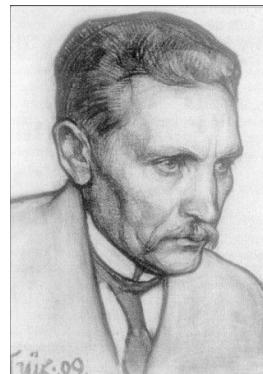
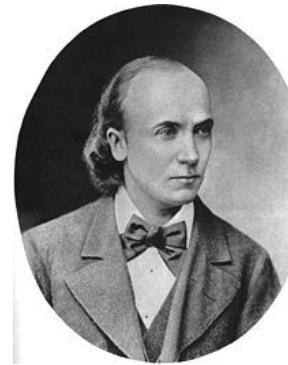
# Contributors on the books

Each book has a list of contributors

- Writers, illustrators, editors, publishers
- They are given as full name with years

We can look at who was involved and when

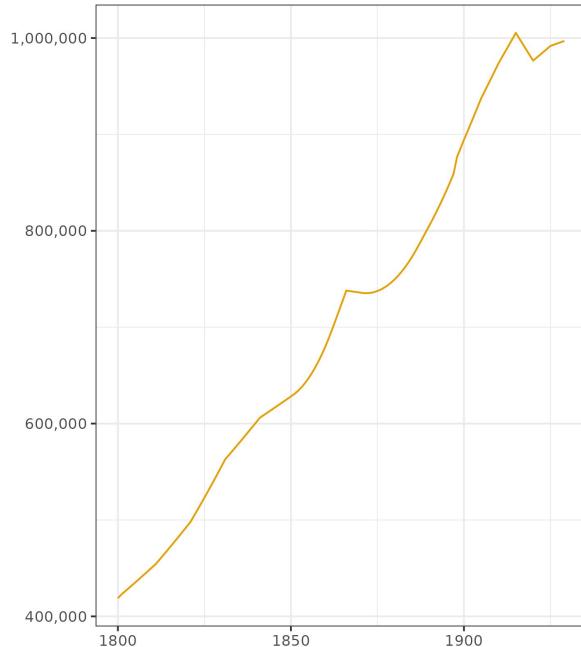
- E.g.
  - Friedrich Brandt (1830-1890), publisher (n = 206)
  - Matthias J. Eisen (1857-1934), folklorist (n = 327)
  - Johannes Käis (1885-1950), teacher (n = 242)
  - Richard Kivist (1888-1981), illustrator (n = 450)



# Population growth and publishing trends

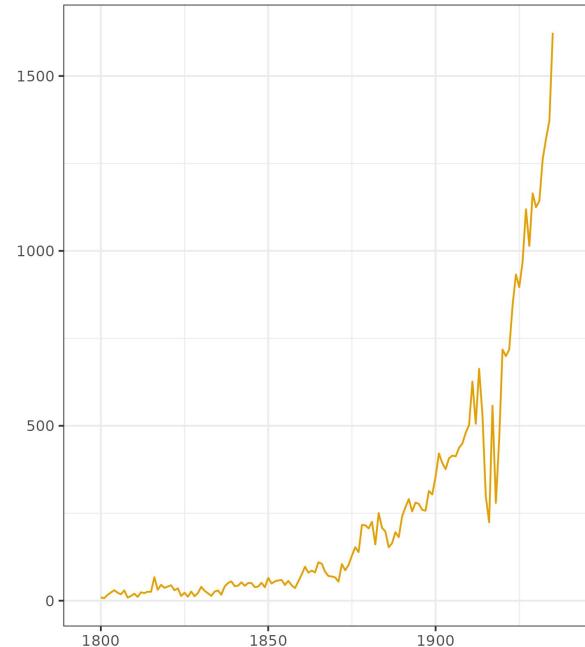
Estonian native speakers (1800-1930)

Residing in modern Estonia



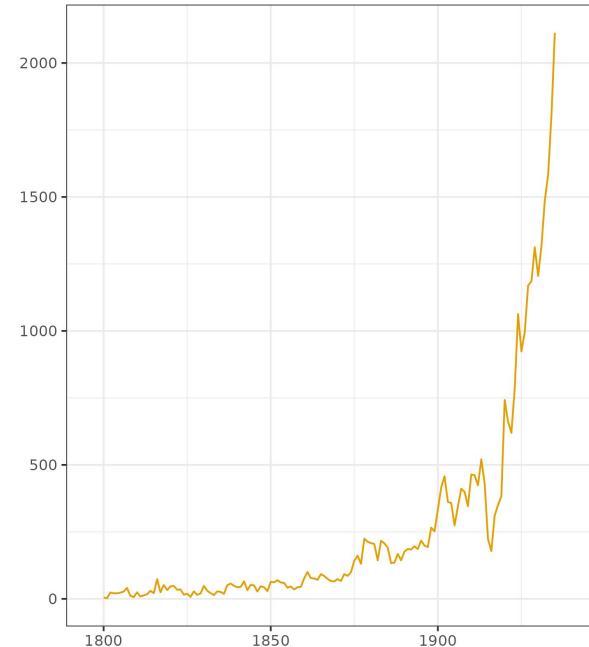
Estonian language books

Per million native speakers



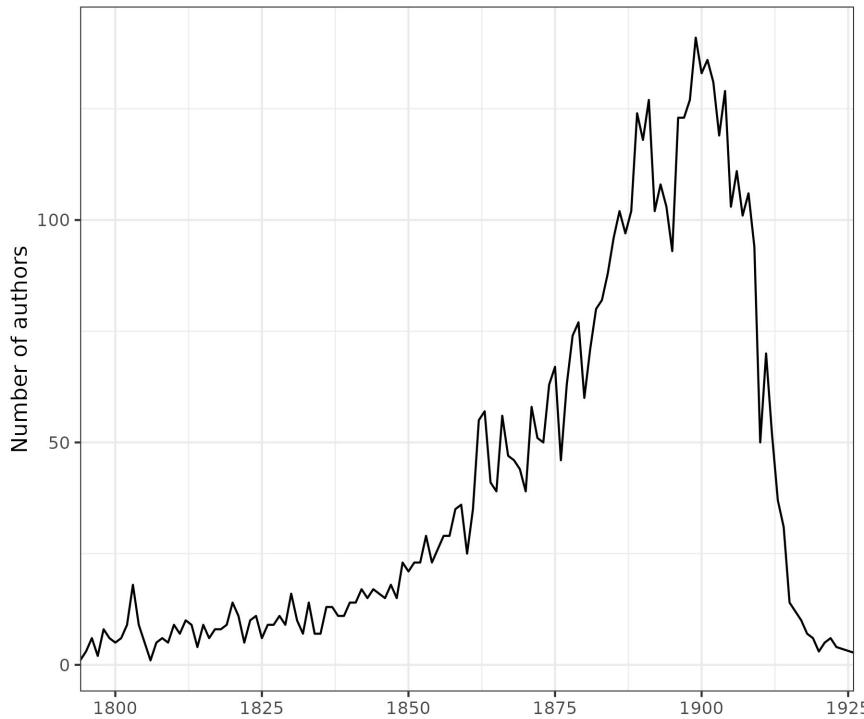
Estonian literary actors

Per million native speakers

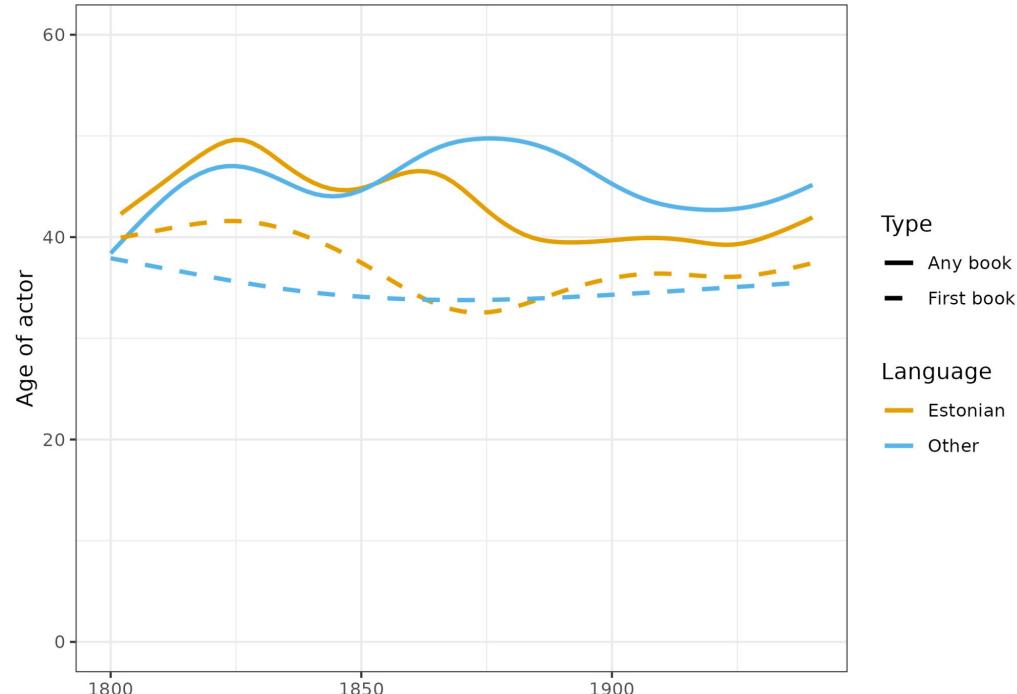


# New people to written language community

Birthyears of actors active in 1800-1940

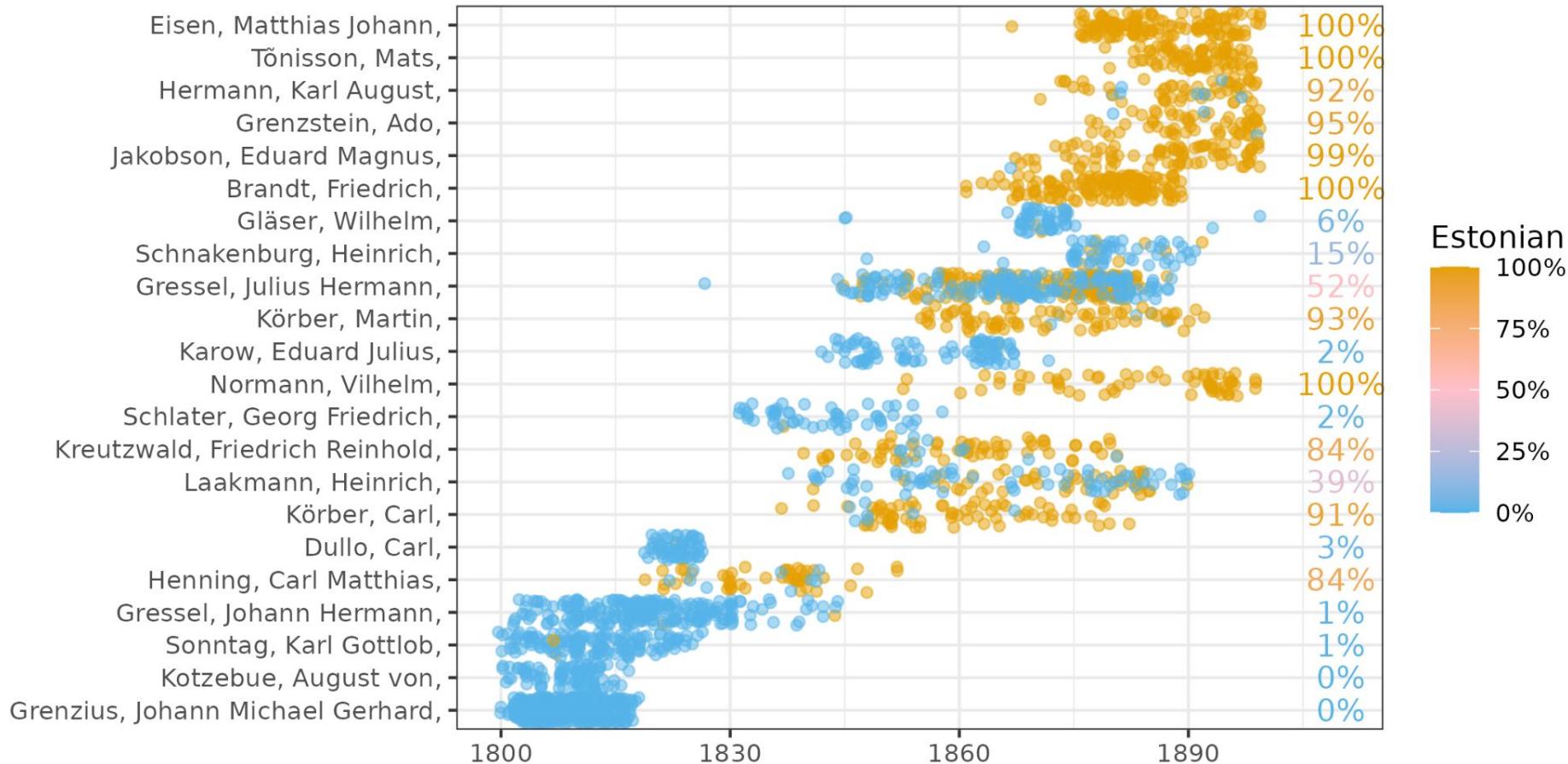


Age of average active living actor by language (1800-1940)  
Across all books (—) and for first books published (- - -)



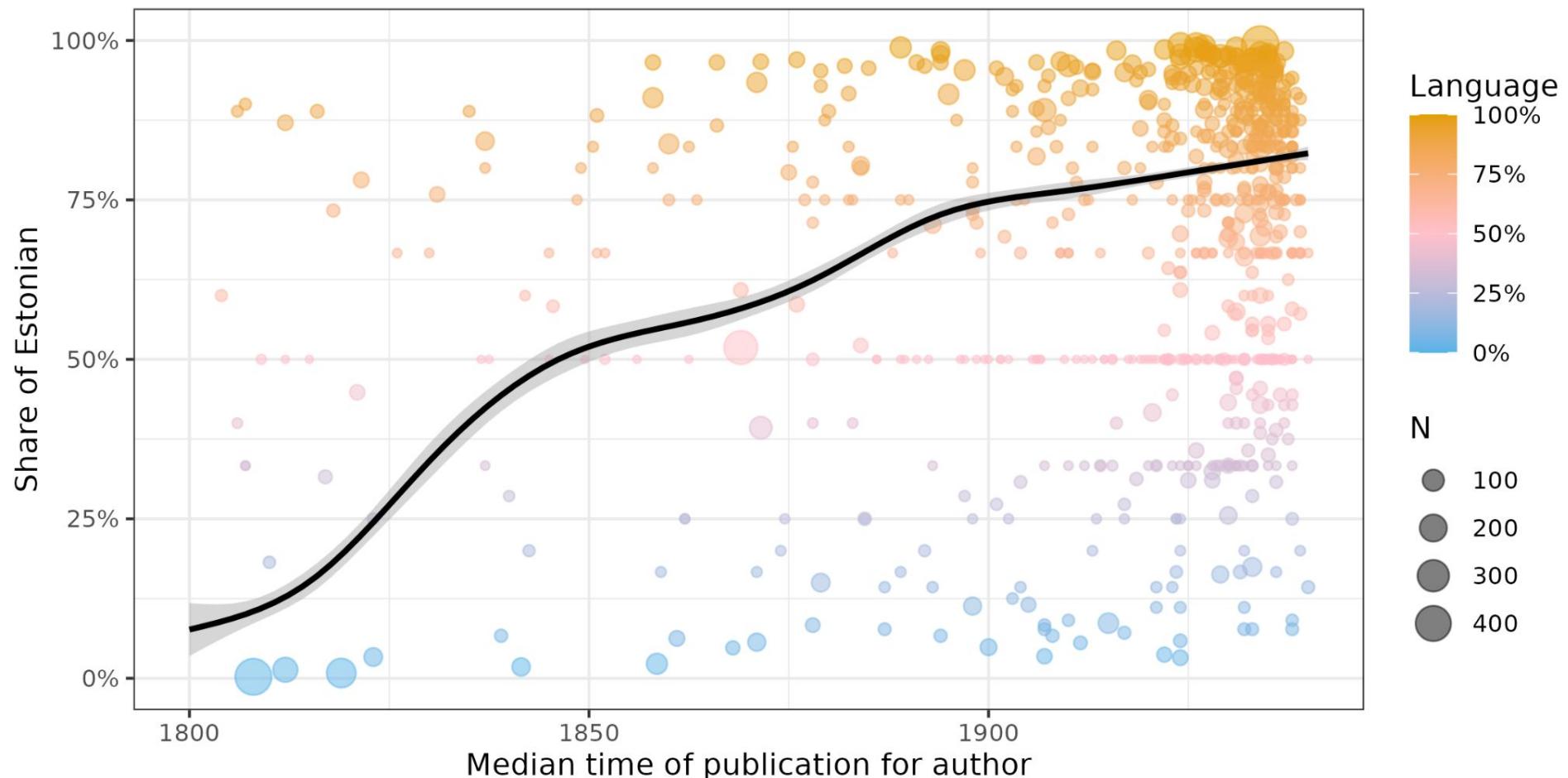
# Language of books during actor's life

Actors with > 50 works, lifetime share on the right



# Language choice for multilingual actors (n = 711)

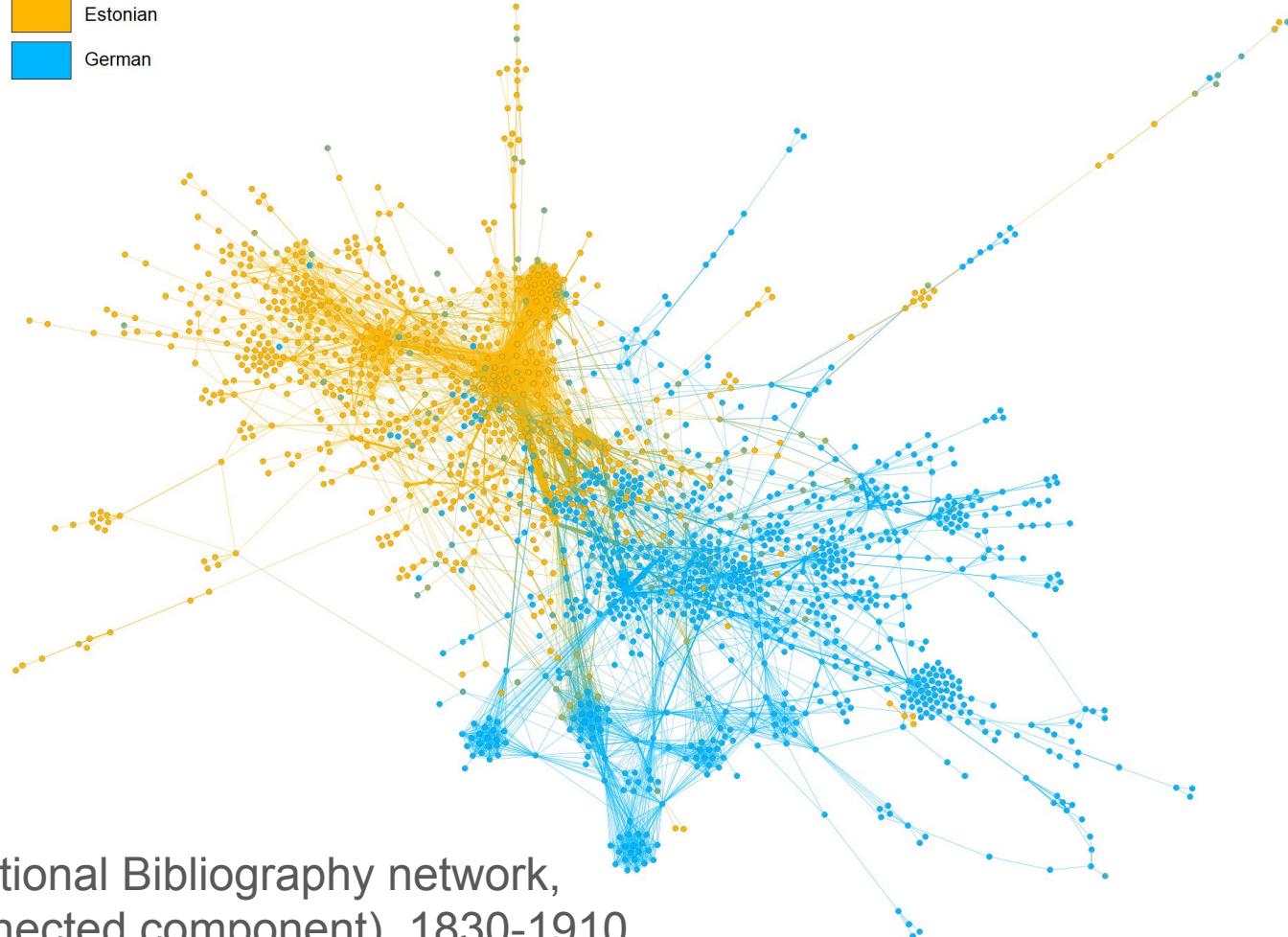
Lifetime share of language for actor vs median time of publication



# Networks

- The National Bibliography data can also be viewed as a literary network, where nodes are actors and edges are works (e.g. Hill *et al* 2019)
- Shows the relationships between authors, publishers, illustrators, editors and others
- Can be constructed diachronically and be coded to represent attributes like language

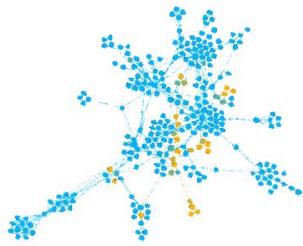
 Estonian  
 German



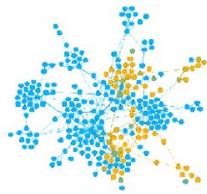
Estonian National Bibliography network,  
(largest connected component), 1830-1910

## Largest connected components

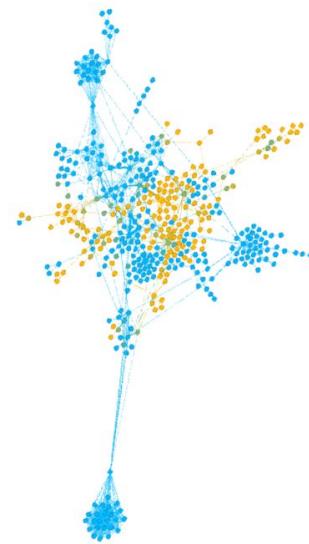
1830-1849



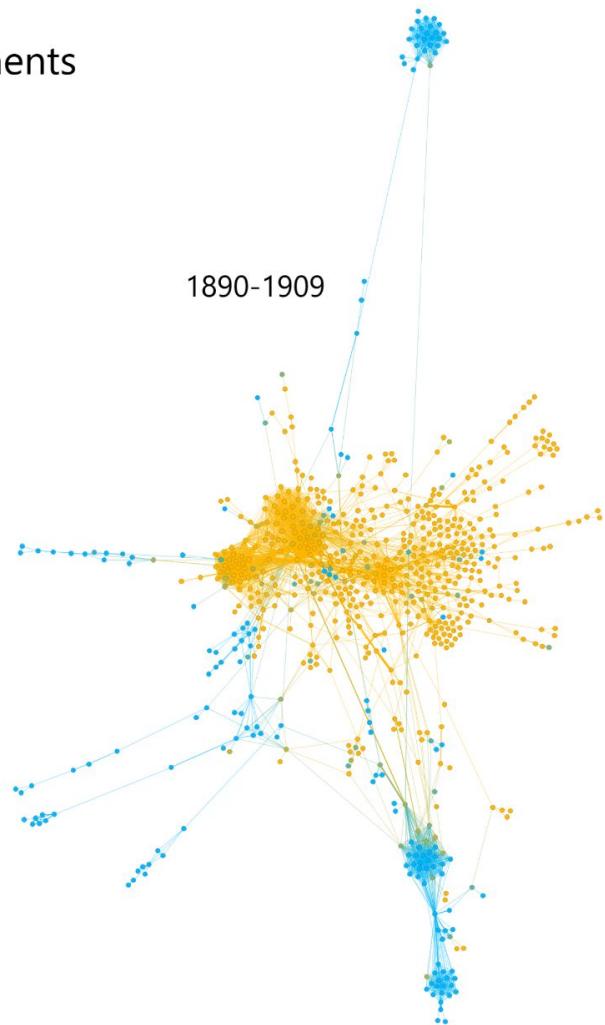
1850-1869



1870-1889



1890-1909

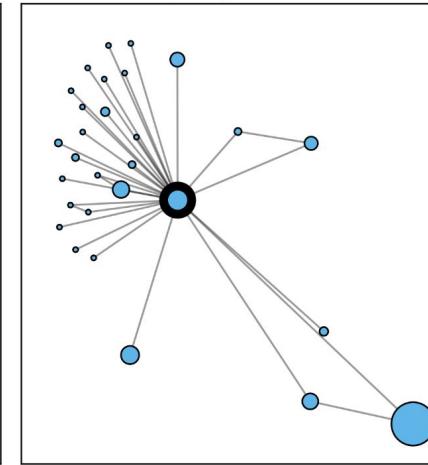
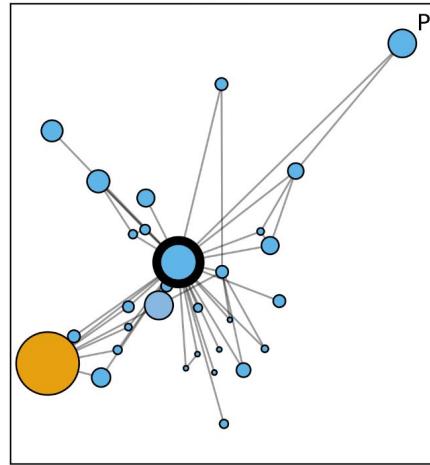
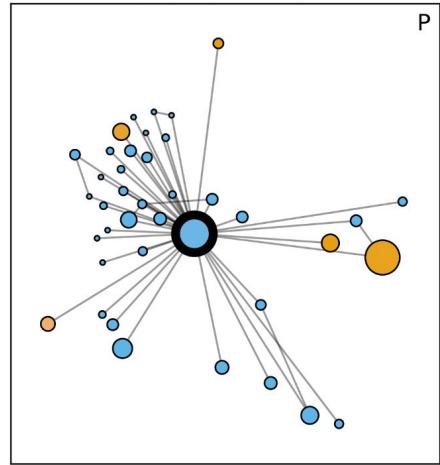


# Ego networks (radius=1) of most central nodes by betweenness, 1850-1869

1. Karow, Eduard Julius

2. Kluge, Franz

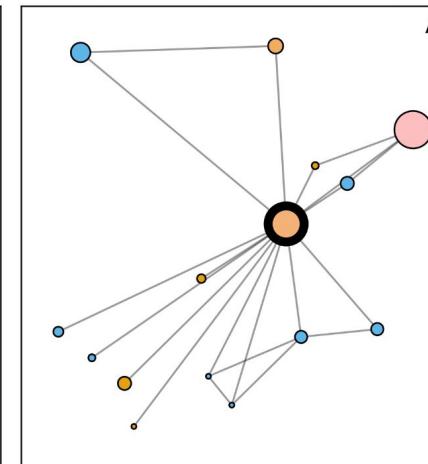
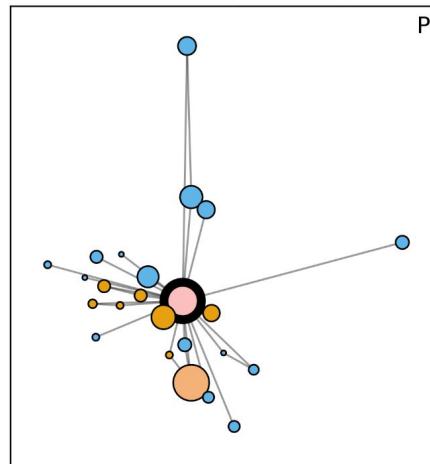
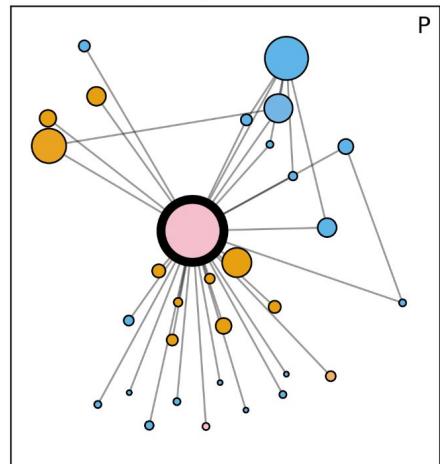
3. Schulz, Carl Anton



4. Gressel, Julius Hermann

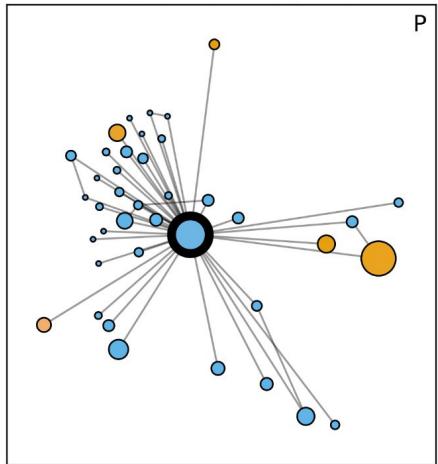
5. Laakmann, Heinrich

6. Kreutzwald, Friedrich Reinhold

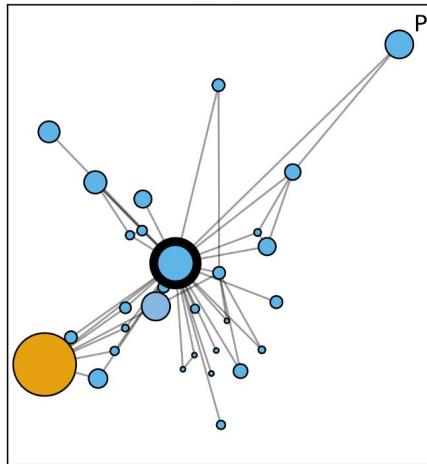


# Ego networks (radius=1) of most central nodes by betweenness, 1850-1869

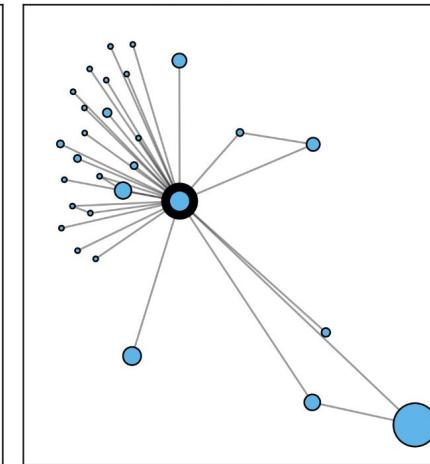
1. Karow, Eduard Julius



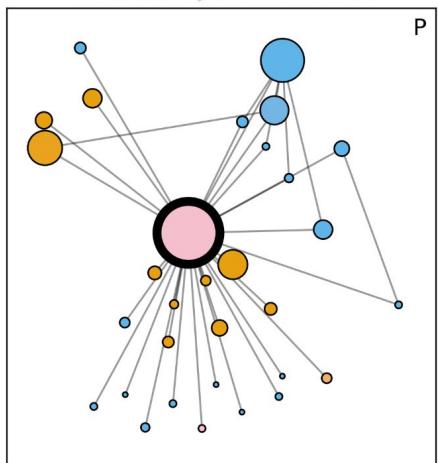
2. Kluge, Franz



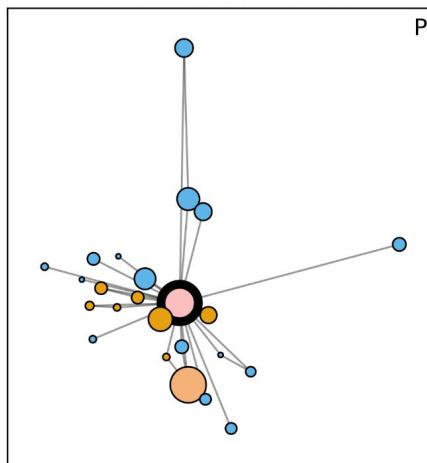
3. Schulz, Carl Anton



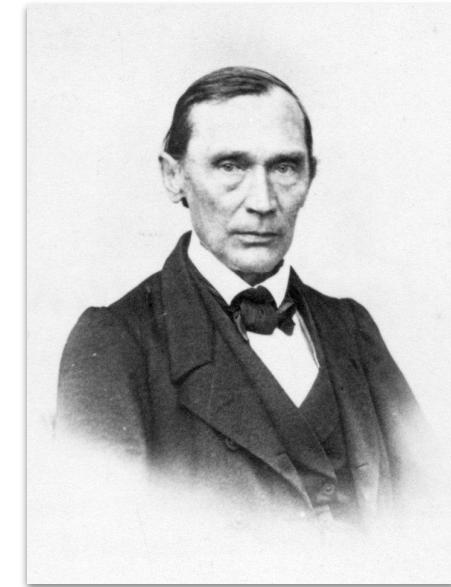
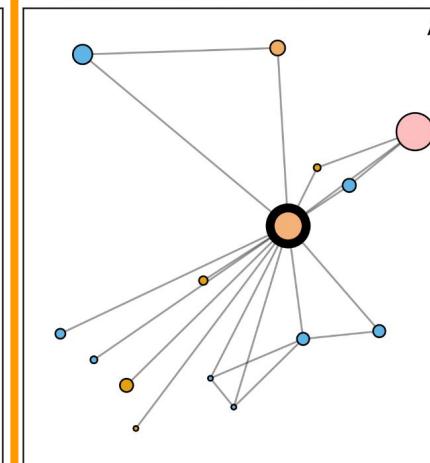
4. Gressel, Julius Hermann



5. Laakmann, Heinrich



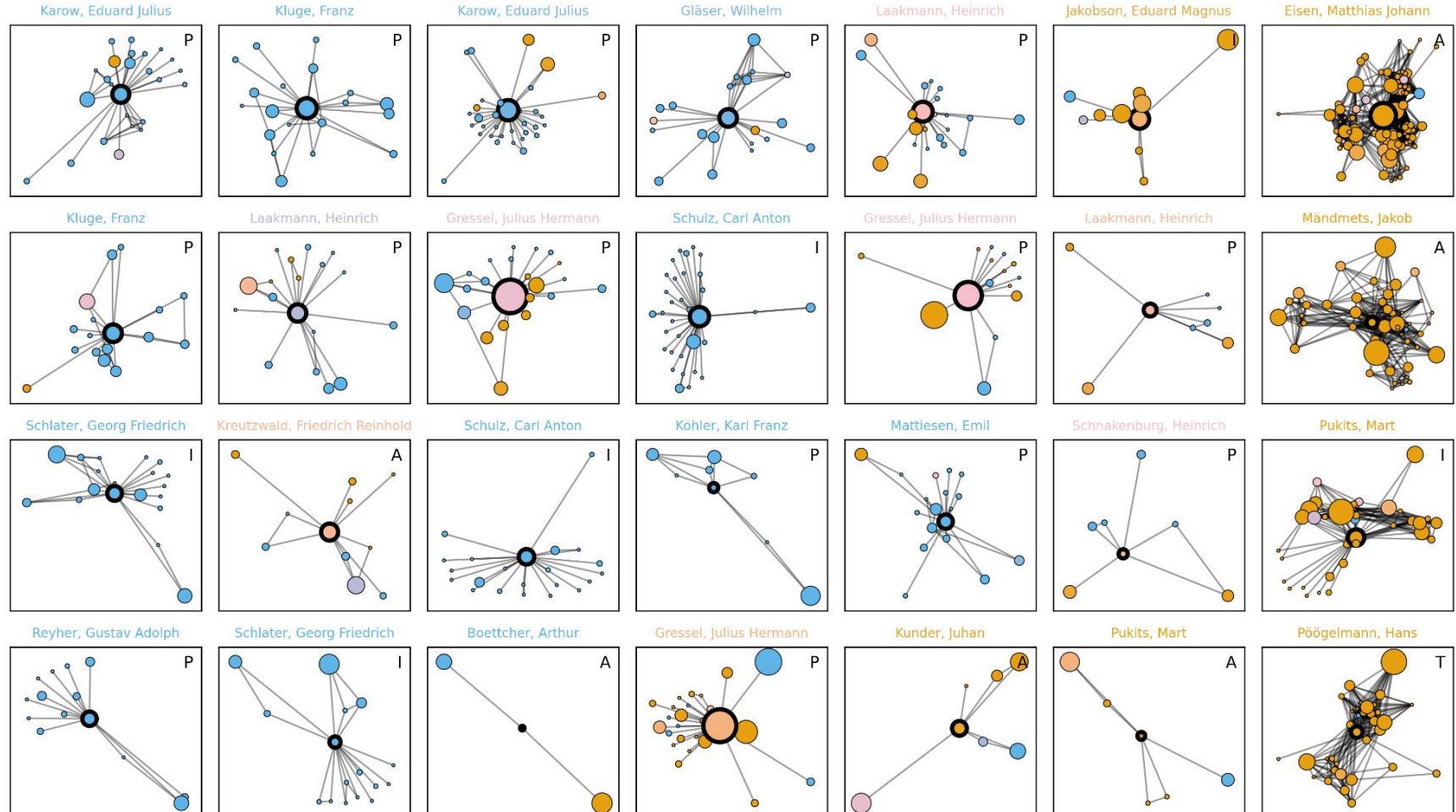
6. Kreutzwald, Friedrich Reinhold



- Leading figure in the early National Awakening
- Author of the Estonian national epic “Kalevipoeg”
- Published mainly in Estonian, but also German and other languages

# Ego networks (radius=1) of most central nodes, 1840-1910

betweenness centrality rank



1840-1849

1850-1859

1860-1869

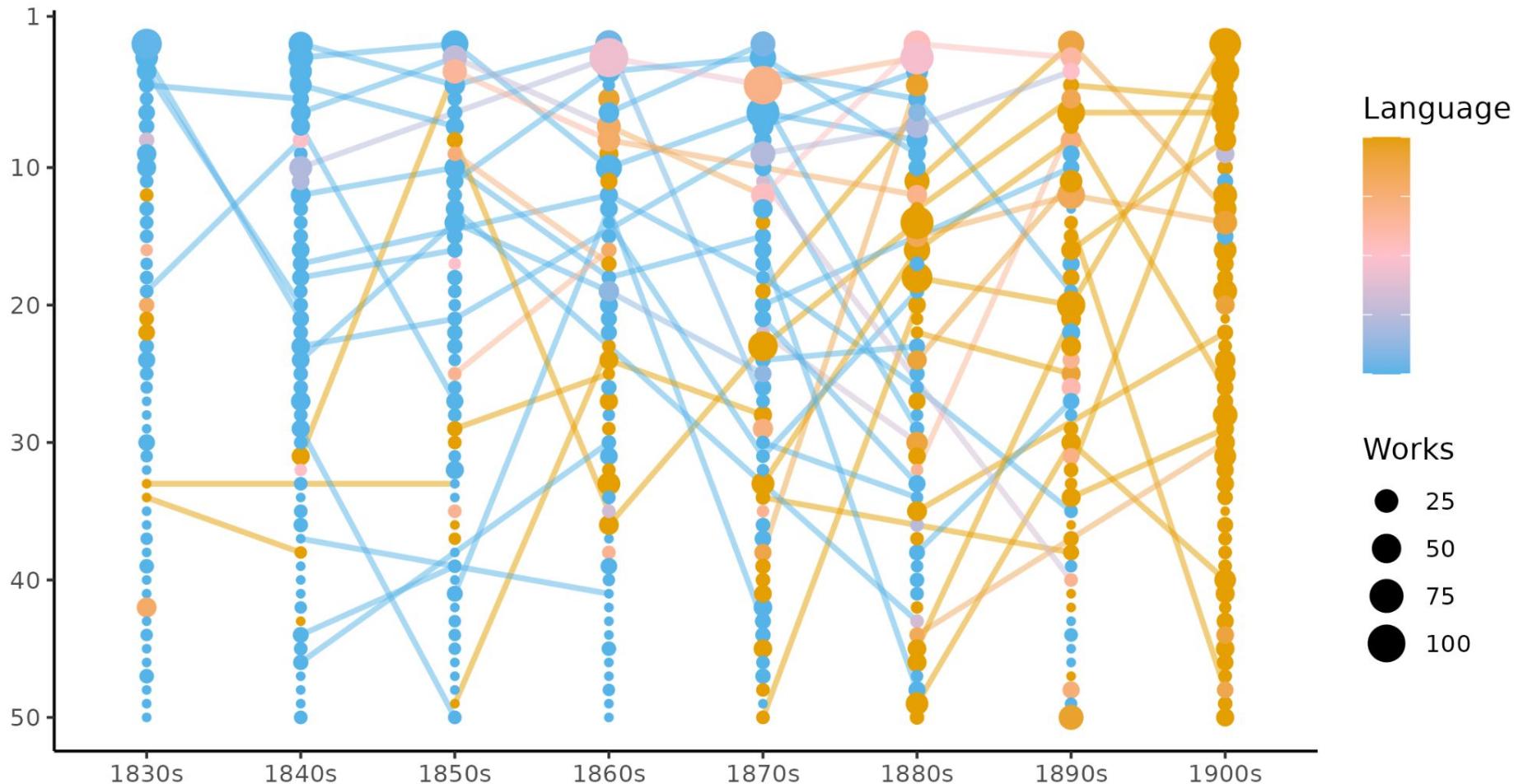
1870-1879

1880-1889

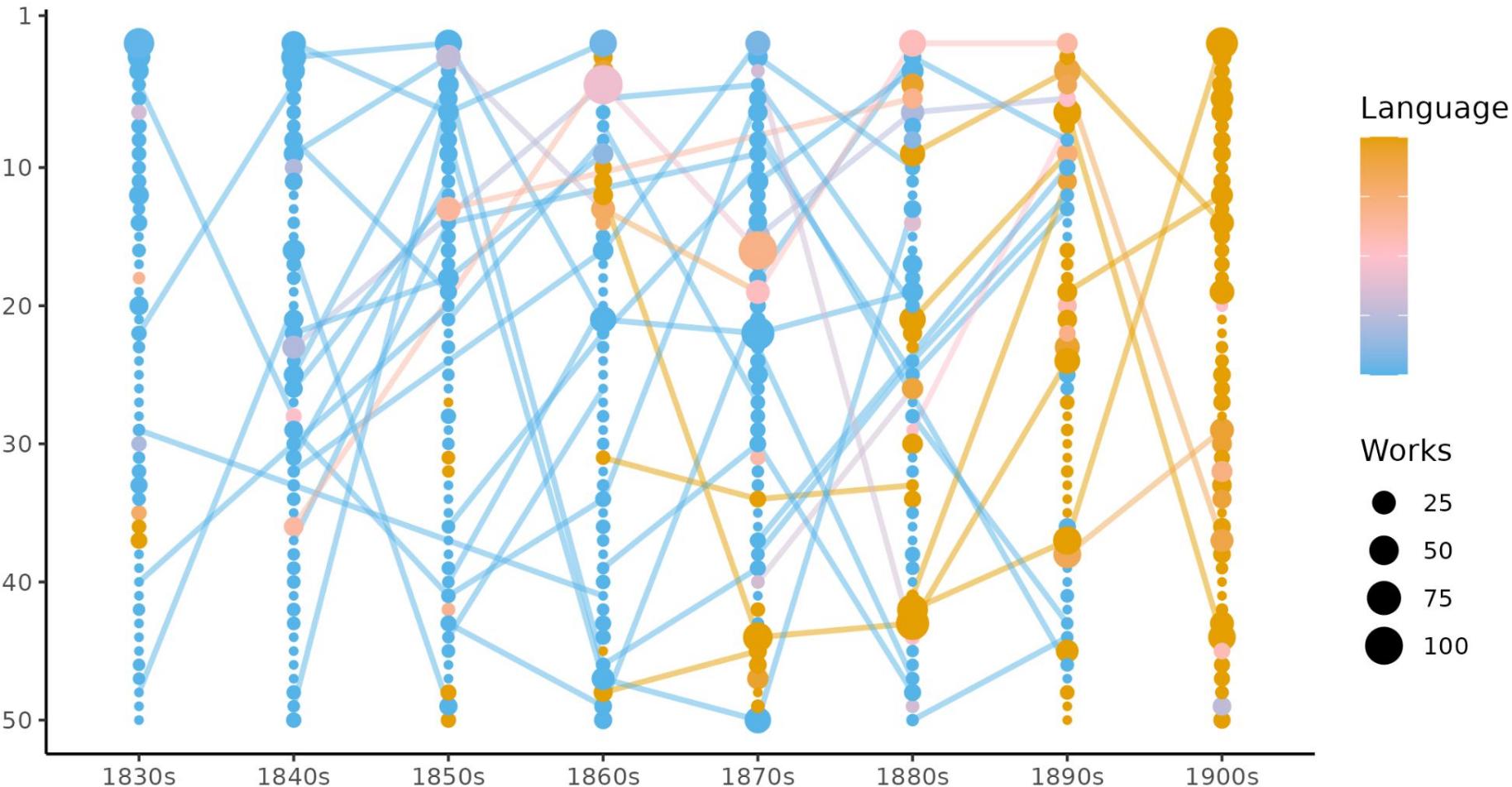
1890-1899

1900-1909

# Top 50 most central nodes by decade, ranked by betweenness



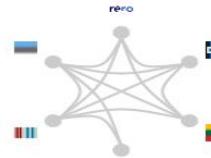
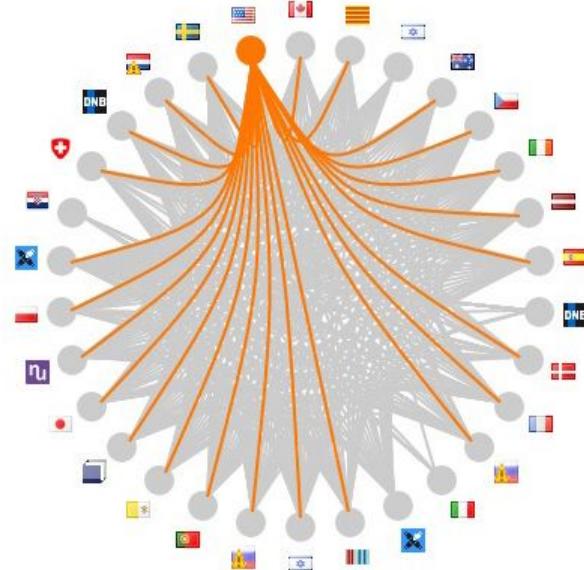
# Top 50 most central nodes by decade, ranked by closeness



# Linked biographies

Many contributors have a link to VIAF

Some more data in other datasets



## Julius Aamisepp (Q12365553)

No description defined

In more languages		Configure	edit
Language	Label	Description	Also known as
English	Julius Aamisepp	No description defined	
Estonian	Julius Aamisepp	No description defined	Julius Aamisepp
Russian	Аамисепп, Юлий Сильмозин	No description defined	
Finnish	No label defined	No description defined	

All entered languages

## Statements

instance of	human	edit
-------------	-------	------

# Linked biographies

- Gathered birth places from various datasets
- Geo-coded them for coordinates



001	ERRRja11124581	(VIAF cluster)
003	ERRR	
008	990322n(jad)nmaan     a	
035	ta (ERRR)a11124581	
040	ta ERRR ja est to: ErRR ja E/TUR	
043	Ic ee	
100	1 ta Aamisepp, Julius, zd 1883-1950.	
400	1 ta Aamisepp, Julius.	
500	1 ta Asmussen, Olof Kjet., zd 1883-1950.	
670	ta EE, 1. kd., 1986, EE, 14. kd., 2000.	
680	21 Eesti põlumajandusteadlane, sordiarendaja, põlumajandusdr (1947).	



Artikkel | Arutelu

Sisse logimata | Arutelu | Kaastöö | Loo kontoo | Logi in | Otsi Vikiweedist | Esileht

## Julius Aamisepp

Julius Aamisepp (1. september (vk) 20. august) 1883 Harjumaa, Kloostri vald, Karlepa, Karlepa-Tõnu talu – 19. jaanuar 1950 Jõgeva) oli eesti sordiarendaja.

## Eesti biograafiline andmebaas ISIK

avaleheküla - nimekirja ja otseai - asemikohad [110] - asustused [470] - ajakud [448] - kaastöö väljaannetele [1133] - nimeenrekujud [925] - pseudonüümid [917] - sulgulised [209] - sunnikohad [622] - surmakohad [186] - tegevusala [299] - lihendid ja allikad - admin

### AAMISEPP, JULIUS

Nimi  
Nime normitud kuju Aamisepp, Julius.

#### Eludaatuminid

Sünnaaeg 1883-09-01  
Sünnikoh Karilepa-Tõnu, Kloostri v. Harjumaal.

#### Surmaaeg

1950-01-19  
Surmakoh Jõgeva

#### Märkus

maetud Tartu Raadi kalmistule

#### Üldandmed

Tegevusalda sordiarendaja

#### Biograafilised andmed

1. IX 1883 Harjumaa Kloostri valda Karlepa-Tõnu.  
Tulenev sõnast Silm A., Liisa Vrager, Ae 1917 Anna Maria Volmer, Ants (1918-  
vt), Valve Jaagus (1920, vt), Ilmar (1927-92, vt). [https://et.wikipedia.org/w/index.php?title=Julius\\_Aamisepp&oldid=34925772](#).  
Parakuks 1920. aastal töötas ta Eesti Riigivabariigi riigiteenistuses, mungavastustatud  
1905-06. Ürituslikku tööd suurte tulvaväsimete ajal 1914-16. Eesti Sõjaväelaste  
Kesküüdroo juhatuse liige, Harju maakonnaniinukogu liige ja sekkretär 1917-20, 1918 Eesti  
diviinide ülem, 1919-20. Riigivabariigi riigiteenistuses 1920-21, ENSV TA korrespondentti 1946,  
UN (1946-50), Vaikne riigiteenistus 1947, ENSV TA korrespondentti 1946-50, ENSV TA korrespondentti 1946-  
50, Surmud 19.1.1950 Jõgeva, maetud Tartu Raadi kalmistule. EAV, EES, EBLU, EAT2, ENE1,  
ENE2, EABL, VJ, ENE2(14), ETeadBL - ISOTAMM 2

### Julius Aamisepp (Q1965555)

No description defined

+ in more languages

Language	Label	Description	Also known as
English	Julius Aamisepp	No description defined	
Estonian	Julius Aamisepp	No description defined	Julius Aamisepp
Russian	Аамисепп, Юль	No description defined	
Finnish	Julius Aamisepp	No label defined	No description defined

All entered languages

#### Statements

instance of human

Amis

Aamisepp, Julius

männlich

Pressearchiv des Herder-Instituts Marburg

Wikipedia (Stand: 25.09.2018): [https://en.wikipedia.org/wiki/Julius\\_Aamisepp](https://en.wikipedia.org/wiki/Julius_Aamisepp)

Lebensdaten: 1883-1950

Estland (XA-EE)

räfischer

Geburtsort: Karlepa (Gem. Lääne-Harju)

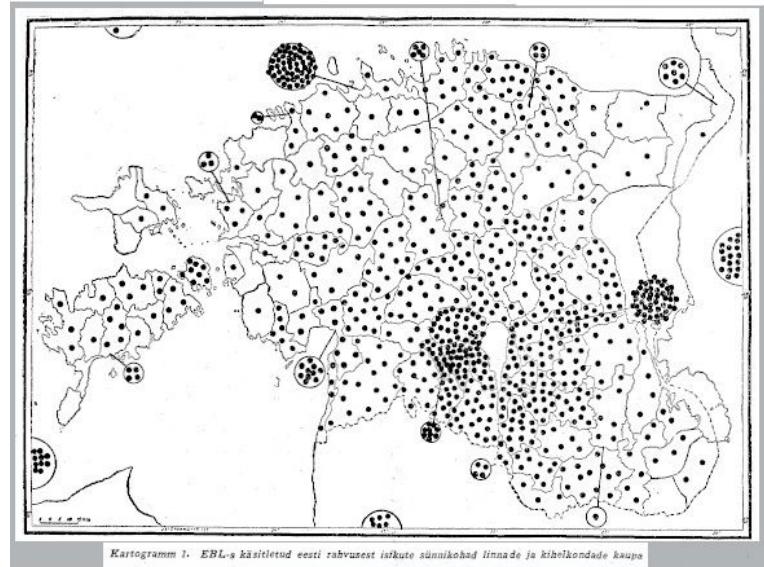
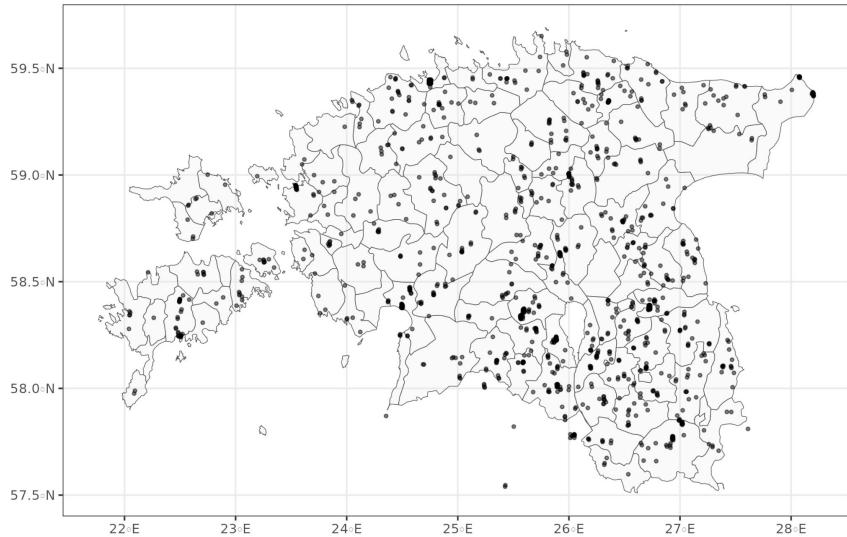
Sterbeort: Jõgeva

Agrarwissenschaftler

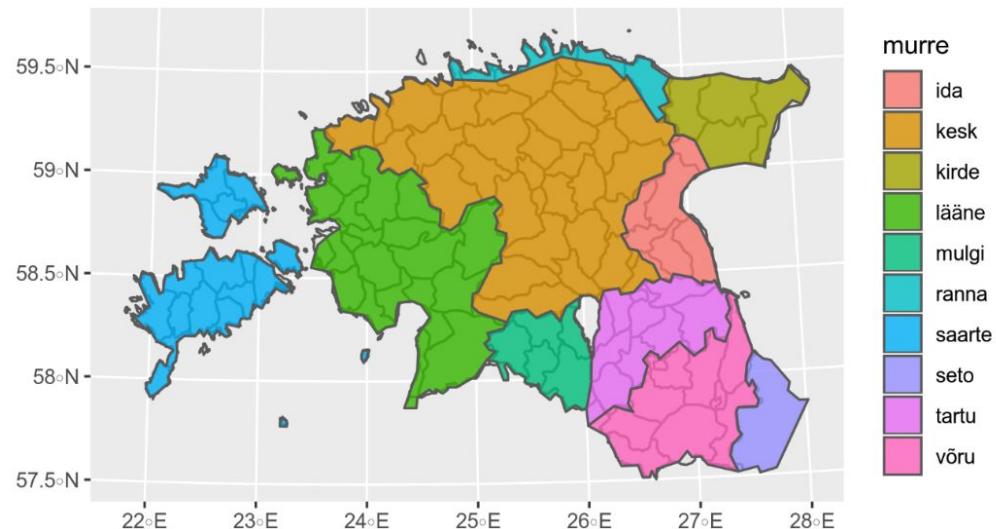
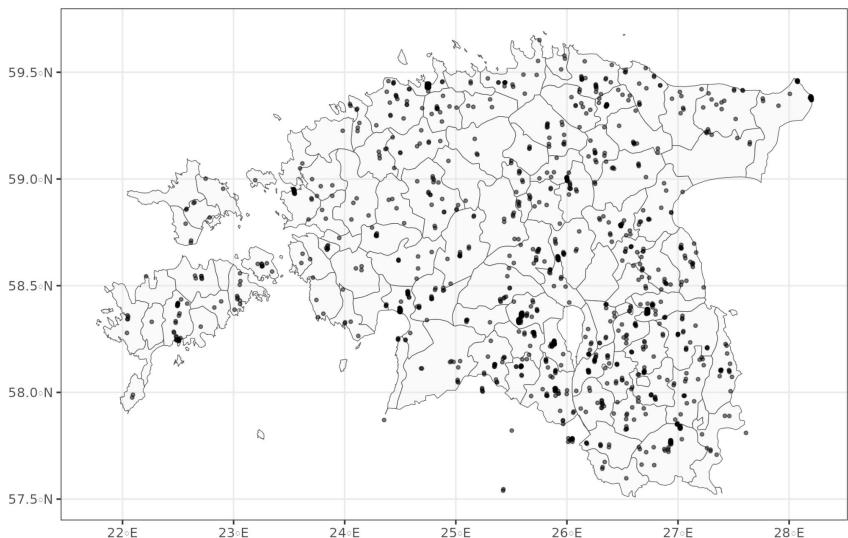
Dozent

100	1 ta Aamisepp, Julius zd 1883-1950
375	21 ta Aamisepp, Julius
500	1 ta Aamisepp, Julius.
549	1 ta Aamisepp, Julius.
550	1 ta Aamisepp, Julius.
551	1 ta Aamisepp, Julius.
552	1 ta Aamisepp, Julius.
553	1 ta Aamisepp, Julius.
554	1 ta Aamisepp, Julius.
555	1 ta Aamisepp, Julius.
556	1 ta Aamisepp, Julius.
557	1 ta Aamisepp, Julius.
558	1 ta Aamisepp, Julius.
559	1 ta Aamisepp, Julius.
560	1 ta Aamisepp, Julius.
561	1 ta Aamisepp, Julius.
562	1 ta Aamisepp, Julius.
563	1 ta Aamisepp, Julius.
564	1 ta Aamisepp, Julius.
565	1 ta Aamisepp, Julius.
566	1 ta Aamisepp, Julius.
567	1 ta Aamisepp, Julius.
568	1 ta Aamisepp, Julius.
569	1 ta Aamisepp, Julius.
570	1 ta Aamisepp, Julius.
571	1 ta Aamisepp, Julius.
572	1 ta Aamisepp, Julius.
573	1 ta Aamisepp, Julius.
574	1 ta Aamisepp, Julius.
575	1 ta Aamisepp, Julius.
576	1 ta Aamisepp, Julius.
577	1 ta Aamisepp, Julius.
578	1 ta Aamisepp, Julius.
579	1 ta Aamisepp, Julius.
580	1 ta Aamisepp, Julius.
581	1 ta Aamisepp, Julius.
582	1 ta Aamisepp, Julius.
583	1 ta Aamisepp, Julius.
584	1 ta Aamisepp, Julius.
585	1 ta Aamisepp, Julius.
586	1 ta Aamisepp, Julius.
587	1 ta Aamisepp, Julius.
588	1 ta Aamisepp, Julius.
589	1 ta Aamisepp, Julius.
590	1 ta Aamisepp, Julius.
591	1 ta Aamisepp, Julius.
592	1 ta Aamisepp, Julius.
593	1 ta Aamisepp, Julius.
594	1 ta Aamisepp, Julius.
595	1 ta Aamisepp, Julius.
596	1 ta Aamisepp, Julius.
597	1 ta Aamisepp, Julius.
598	1 ta Aamisepp, Julius.
599	1 ta Aamisepp, Julius.
600	1 ta Aamisepp, Julius.
601	1 ta Aamisepp, Julius.
602	1 ta Aamisepp, Julius.
603	1 ta Aamisepp, Julius.
604	1 ta Aamisepp, Julius.
605	1 ta Aamisepp, Julius.
606	1 ta Aamisepp, Julius.
607	1 ta Aamisepp, Julius.
608	1 ta Aamisepp, Julius.
609	1 ta Aamisepp, Julius.
610	1 ta Aamisepp, Julius.
611	1 ta Aamisepp, Julius.
612	1 ta Aamisepp, Julius.
613	1 ta Aamisepp, Julius.
614	1 ta Aamisepp, Julius.
615	1 ta Aamisepp, Julius.
616	1 ta Aamisepp, Julius.
617	1 ta Aamisepp, Julius.
618	1 ta Aamisepp, Julius.
619	1 ta Aamisepp, Julius.
620	1 ta Aamisepp, Julius.
621	1 ta Aamisepp, Julius.
622	1 ta Aamisepp, Julius.
623	1 ta Aamisepp, Julius.
624	1 ta Aamisepp, Julius.
625	1 ta Aamisepp, Julius.
626	1 ta Aamisepp, Julius.
627	1 ta Aamisepp, Julius.
628	1 ta Aamisepp, Julius.
629	1 ta Aamisepp, Julius.
630	1 ta Aamisepp, Julius.
631	1 ta Aamisepp, Julius.
632	1 ta Aamisepp, Julius.
633	1 ta Aamisepp, Julius.
634	1 ta Aamisepp, Julius.
635	1 ta Aamisepp, Julius.
636	1 ta Aamisepp, Julius.
637	1 ta Aamisepp, Julius.
638	1 ta Aamisepp, Julius.
639	1 ta Aamisepp, Julius.
640	1 ta Aamisepp, Julius.
641	1 ta Aamisepp, Julius.
642	1 ta Aamisepp, Julius.
643	1 ta Aamisepp, Julius.
644	1 ta Aamisepp, Julius.
645	1 ta Aamisepp, Julius.
646	1 ta Aamisepp, Julius.
647	1 ta Aamisepp, Julius.
648	1 ta Aamisepp, Julius.
649	1 ta Aamisepp, Julius.
650	1 ta Aamisepp, Julius.
651	1 ta Aamisepp, Julius.
652	1 ta Aamisepp, Julius.
653	1 ta Aamisepp, Julius.
654	1 ta Aamisepp, Julius.
655	1 ta Aamisepp, Julius.
656	1 ta Aamisepp, Julius.
657	1 ta Aamisepp, Julius.
658	1 ta Aamisepp, Julius.
659	1 ta Aamisepp, Julius.
660	1 ta Aamisepp, Julius.
661	1 ta Aamisepp, Julius.
662	1 ta Aamisepp, Julius.
663	1 ta Aamisepp, Julius.
664	1 ta Aamisepp, Julius.
665	1 ta Aamisepp, Julius.
666	1 ta Aamisepp, Julius.
667	1 ta Aamisepp, Julius.
668	1 ta Aamisepp, Julius.
669	1 ta Aamisepp, Julius.
670	1 ta Aamisepp, Julius.
671	1 ta Aamisepp, Julius.
672	1 ta Aamisepp, Julius.
673	1 ta Aamisepp, Julius.
674	1 ta Aamisepp, Julius.
675	1 ta Aamisepp, Julius.
676	1 ta Aamisepp, Julius.
677	1 ta Aamisepp, Julius.
678	1 ta Aamisepp, Julius.
679	1 ta Aamisepp, Julius.
680	1 ta Aamisepp, Julius.
681	1 ta Aamisepp, Julius.
682	1 ta Aamisepp, Julius.
683	1 ta Aamisepp, Julius.
684	1 ta Aamisepp, Julius.
685	1 ta Aamisepp, Julius.
686	1 ta Aamisepp, Julius.
687	1 ta Aamisepp, Julius.
688	1 ta Aamisepp, Julius.
689	1 ta Aamisepp, Julius.
690	1 ta Aamisepp, Julius.
691	1 ta Aamisepp, Julius.
692	1 ta Aamisepp, Julius.
693	1 ta Aamisepp, Julius.
694	1 ta Aamisepp, Julius.
695	1 ta Aamisepp, Julius.
696	1 ta Aamisepp, Julius.
697	1 ta Aamisepp, Julius.
698	1 ta Aamisepp, Julius.
699	1 ta Aamisepp, Julius.
700	1 ta Aamisepp, Julius.
701	1 ta Aamisepp, Julius.
702	1 ta Aamisepp, Julius.
703	1 ta Aamisepp, Julius.
704	1 ta Aamisepp, Julius.
705	1 ta Aamisepp, Julius.
706	1 ta Aamisepp, Julius.
707	1 ta Aamisepp, Julius.
708	1 ta Aamisepp, Julius.
709	1 ta Aamisepp, Julius.
710	1 ta Aamisepp, Julius.
711	1 ta Aamisepp, Julius.
712	1 ta Aamisepp, Julius.
713	1 ta Aamisepp, Julius.
714	1 ta Aamisepp, Julius.
715	1 ta Aamisepp, Julius.
716	1 ta Aamisepp, Julius.
717	1 ta Aamisepp, Julius.
718	1 ta Aamisepp, Julius.
719	1 ta Aamisepp, Julius.
720	1 ta Aamisepp, Julius.
721	1 ta Aamisepp, Julius.
722	1 ta Aamisepp, Julius.
723	1 ta Aamisepp, Julius.
724	1 ta Aamisepp, Julius.
725	1 ta Aamisepp, Julius.
726	1 ta Aamisepp, Julius.
727	1 ta Aamisepp, Julius.
728	1 ta Aamisepp, Julius.
729	1 ta Aamisepp, Julius.
730	1 ta Aamisepp, Julius.
731	1 ta Aamisepp, Julius.
732	1 ta Aamisepp, Julius.
733	1 ta Aamisepp, Julius.
734	1 ta Aamisepp, Julius.
735	1 ta Aamisepp, Julius.
736	1 ta Aamisepp, Julius.
737	1 ta Aamisepp, Julius.
738	1 ta Aamisepp, Julius.
739	1 ta Aamisepp, Julius.
740	1 ta Aamisepp, Julius.
741	1 ta Aamisepp, Julius.
742	1 ta Aamisepp, Julius.
743	1 ta Aamisepp, Julius.
744	1 ta Aamisepp, Julius.
745	1 ta Aamisepp, Julius.
746	1 ta Aamisepp, Julius.
747	1 ta Aamisepp, Julius.
748	1 ta Aamisepp, Julius.
749	1 ta Aamisepp, Julius.
750	1 ta Aamisepp, Julius.
751	1 ta Aamisepp, Julius.
752	1 ta Aamisepp, Julius.
753	1 ta Aamisepp, Julius.
754	1 ta Aamisepp, Julius.
755	1 ta Aamisepp, Julius.
756	1 ta Aamisepp, Julius.
757	1 ta Aamisepp, Julius.
758	1 ta Aamisepp, Julius.
759	1 ta Aamisepp, Julius.
760	1 ta Aamisepp, Julius.
761	1 ta Aamisepp, Julius.
762	1 ta Aamisepp, Julius.
763	1 ta Aamisepp, Julius.
764	1 ta Aamisepp, Julius.
765	1 ta Aamisepp, Julius.
766	1 ta Aamisepp, Julius.
767	1 ta Aamisepp, Julius.
768	1 ta Aamisepp, Julius.
769	1 ta Aamisepp, Julius.
770	1 ta Aamisepp, Julius.
771	1 ta Aamisepp, Julius.
772	1 ta Aamisepp, Julius.
773	1 ta Aamisepp, Julius.
774	1 ta Aamisepp, Julius.
775	1 ta Aamisepp, Julius.
776	1 ta Aamisepp, Julius.
777	1 ta Aamisepp, Julius.
778	1 ta Aamisepp, Julius.
779	1 ta Aamisepp, Julius.
780	1 ta Aamisepp, Julius.
781	1 ta Aamisepp, Julius.
782	1 ta Aamisepp, Julius.
783	1 ta Aamisepp, Julius.
784	1 ta Aamisepp, Julius.
785	1 ta Aamisepp, Julius.
786	1 ta Aamisepp, Julius.
787	1 ta Aamisepp, Julius.
788	1 ta Aamisepp, Julius.
789	1 ta Aamisepp, Julius.
790	1 ta Aamisepp, Julius.
791	1 ta Aamisepp, Julius.
792	1 ta Aamisepp, Julius.
793	1 ta Aamisepp, Julius.
794	1 ta Aamisepp, Julius.
795	1 ta Aamisepp, Julius.
796	1 ta Aamisepp, Julius.
797	1 ta Aamisepp, Julius.
798	1 ta Aamisepp, Julius.
799	1 ta Aamisepp, Julius.
800	1 ta Aamisepp, Julius.
801	1 ta Aamisepp, Julius.
802	1 ta Aamisepp, Julius.
803	1 ta Aamisepp, Julius.
804	1 ta Aamisepp, Julius.
805	1 ta Aamisepp, Julius.
806	1 ta Aamisepp, Julius.
807	1 ta Aamisepp, Julius.
808	1 ta Aamisepp, Julius.
809	1 ta Aamisepp, Julius.
810	1 ta Aamisepp, Julius.
811	1 ta Aamisepp, Julius.
812	1 ta Aamisepp, Julius.
813	1 ta Aamisepp, Julius.
814	1 ta Aamisepp, Julius.
815	1 ta Aamisepp, Julius.
816	1 ta Aamisepp, Julius.
817	1 ta Aamisepp, Julius.
818	1 ta Aamisepp, Julius.
819	1 ta Aamisepp, Julius.
820	1 ta Aamisepp, Julius.
821	1 ta Aamisepp, Julius.
822	1 ta Aamisepp, Julius.
823	1 ta Aamisepp, Julius.
824	1 ta Aamisepp, Julius.
825	1 ta Aamisepp, Julius.
826	1 ta Aamisepp, Julius.
827	1 ta Aamisepp, Julius.
828	1 ta Aamisepp, Julius.
829	1 ta Aamisepp, Julius.
830	1 ta Aamisepp, Julius.
831	1 ta Aamisepp, Julius.
832	1 ta Aamisepp, Julius.
833	1 ta Aamisepp, Julius.
834	1 ta Aamisepp, Julius.
835	1 ta Aamisepp, Julius.
836	1 ta Aamisepp, Julius.
837	1 ta Aamisepp, Julius.
838	1 ta Aamisepp, Julius.
839	1 ta Aamisepp, Julius.
840	1 ta Aamisepp, Julius.
841	1 ta Aamisepp, Julius.
842	1 ta Aamisepp, Julius.
843	1 ta Aamisepp, Julius.
844	1 ta Aamisepp, Julius.
845	1 ta Aamisepp, Julius.
846	1 ta Aamisepp, Julius.
847	1 ta Aamisepp, Julius.
848	1 ta Aamisepp, Julius.
849	1 ta Aamisepp, Julius.
850	1 ta Aamisepp, Julius.
851	1 ta Aamisepp, Julius.
852	1 ta Aamisepp, Julius.
853	1 ta Aamisepp, Julius.
854	1 ta Aamisepp, Julius.
855	1 ta Aamisepp, Julius.
856	1 ta Aamisepp, Julius.
857	1 ta Aamisepp, Julius.
858	1 ta Aamisepp, Julius.
859	1 ta Aamisepp, Julius.
860	1 ta Aamisepp, Julius.
861	1 ta Aamisepp, Julius.
862	1 ta Aamisepp, Julius.
863	1 ta Aamisepp, Julius.
864	1 ta Aamisepp, Julius.
865	1 ta Aamisepp, Julius.
866	1 ta Aamisepp, Julius.
867	1 ta Aamisepp, Julius.
868	1 ta Aamisepp, Julius.
869	1 ta Aamisepp, Julius.
870	1 ta Aamisepp, Julius.
871	1 ta Aamisepp, Julius.
872	1 ta Aamisepp, Julius.
873	1 ta Aamisepp, Julius.
874	1 ta Aamisepp, Julius.
875	1 ta Aamisepp, Julius.
876	1 ta Aamisepp, Julius.
877	1 ta Aamisepp, Julius.
878	1 ta Aamisepp, Julius.
879	1 ta Aamisepp, Julius.
880	1 ta Aamisepp, Julius.
881	1 ta Aamisepp, Julius.
882	1 ta Aamisepp, Julius.
883	1 ta Aamisepp, Julius.
884	1 ta Aamisepp, Julius.
885	1 ta Aamisepp, Julius

# Mapped birthplaces of the contributors

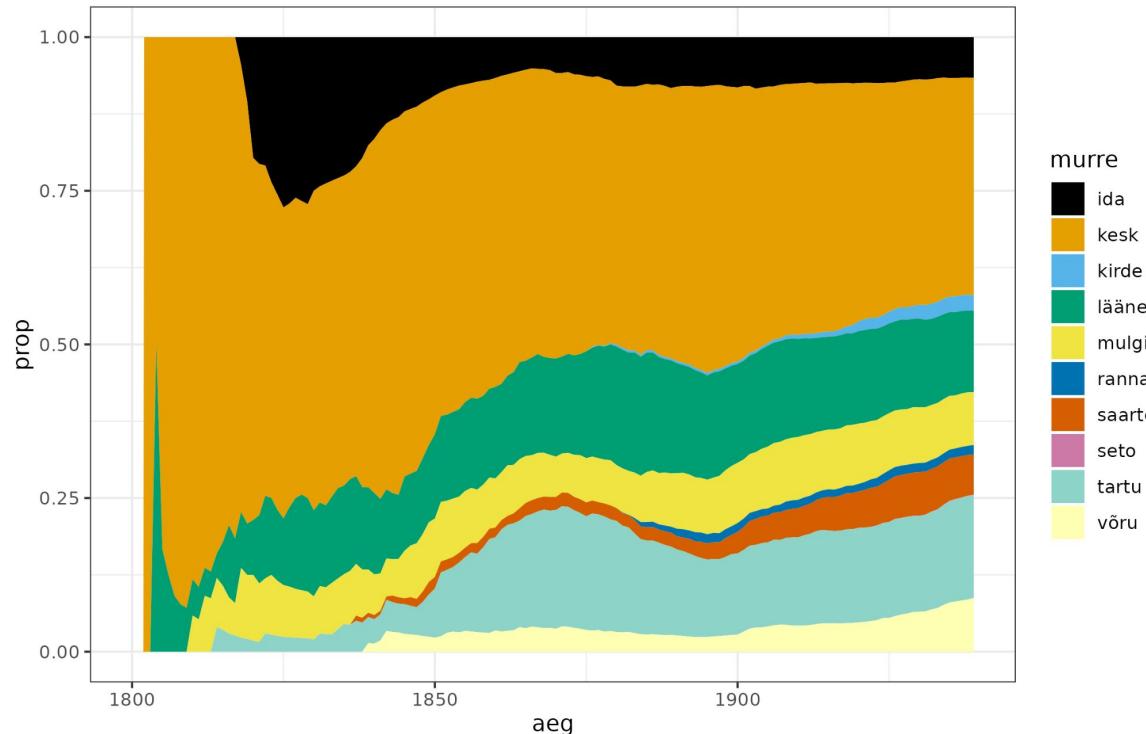


# Mapped birthplaces and dialects



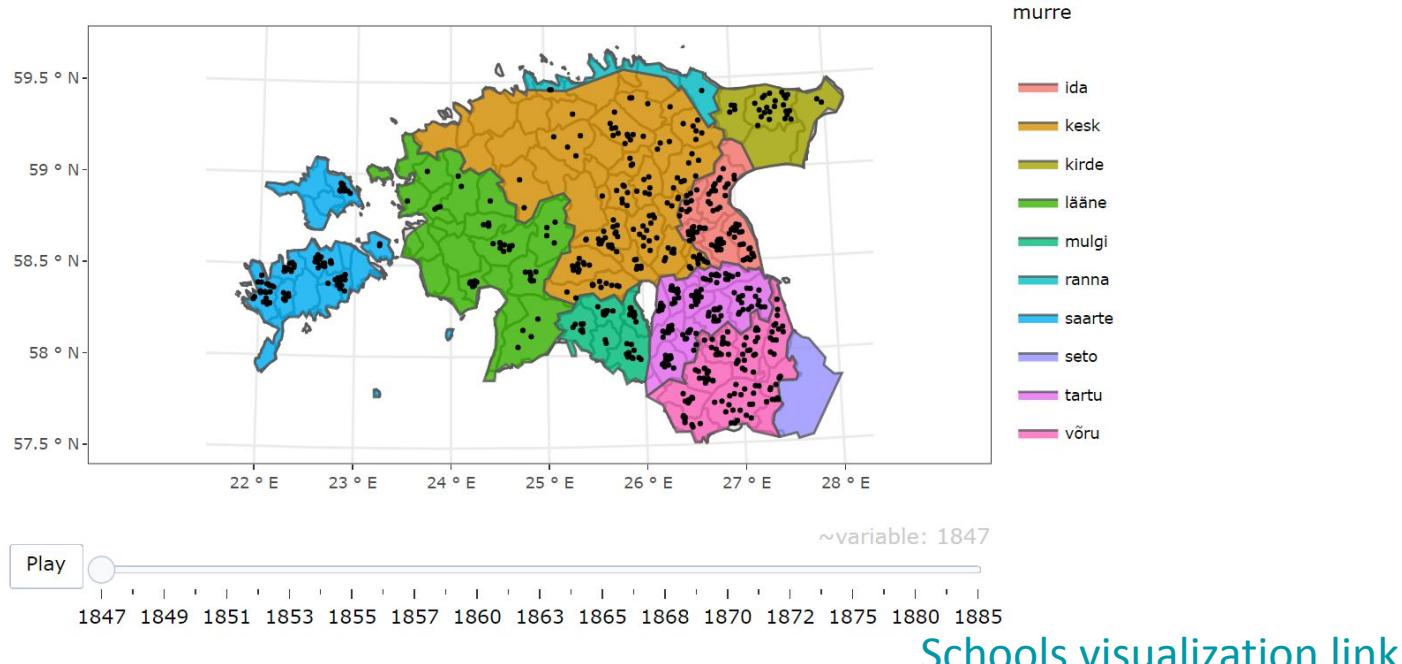
# Share of dialects in the dataset

## Cumulative share of different dialects of Estonian 1800-1940



# Public schools in 19c

- Are founded from 1830s to 1880s



# Estonia & Livonia

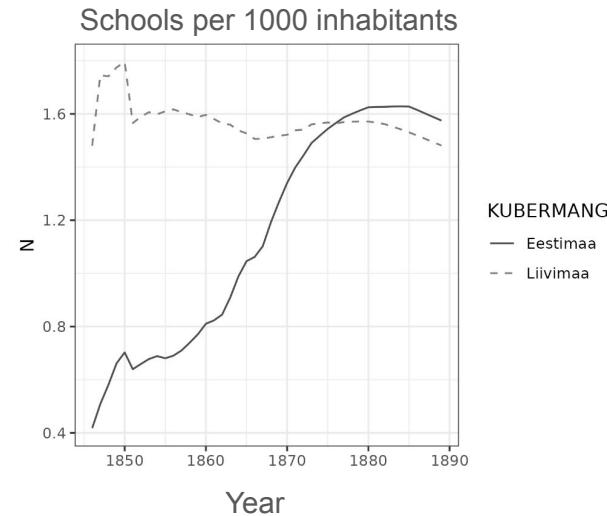
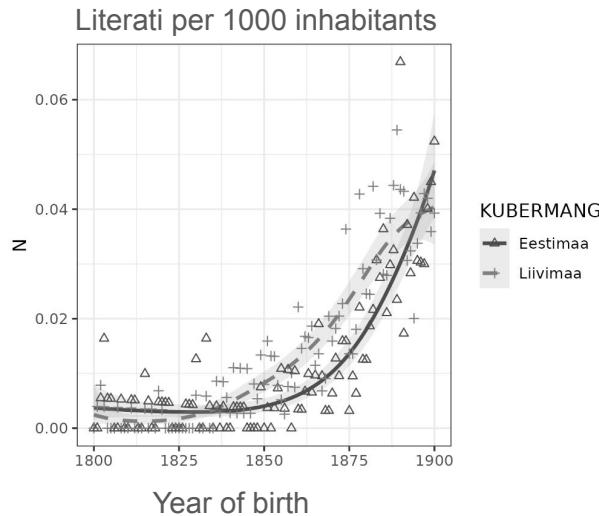
- Administrative rules differ
- Livonia was founding schools quicker, and peasants grew quicker in wealth
- Influence on written language



# Origins of the contributors

Odds of a child born to join the literati in their lifetime (~3.3 kids per 1000)

- 1850
  - 1:800 Livonia
  - 1:1600 Estonia
- 1875
  - 1:400 Livonia
  - 1:800 Estonia
- 1900
  - 1:200 Livonia
  - 1:200 Estonia



# Results

## Research questions

- We can see shifts in language use as response to major historical events
- The 19th century Estonian literary community is shaped by young new actors
- The Estonian community initially emerges *from within* the existing Baltic German dominated network - not alongside it
- Schools played an important role in supporting it

# Adapting bibliographic data for research use

Has a strong research potential

- Needs harmonization, processing
- Benefits hugely from linking to other datasets
- Needs overviews of what is in the data, and what is the coverage
- We're building a curated dataset with this in mind

# Adapting digital archives for research use

Keep the researcher in mind

- Access
- Overviews and contents
- Tools to explore
- Case studies, case studies, case studies
- 
- ... repeat

# Thank you!

Team at the University of Tartu



UNIVERSITY OF TARTU

Centre for Digital Humanities  
and Information Society

Team in the National Library of Estonia



# Literature

- Leo Lahti, Jani Marjanen, Hege Roivainen & Mikko Tolonen (2019) Bibliographic Data Science and the History of the Book (c. 1500–1800), Cataloging & Classification Quarterly, 57:1, 5-23, DOI: [10.1080/01639374.2018.1543747](https://doi.org/10.1080/01639374.2018.1543747)
- Mark J Hill, Ville Vaara, Tanja Säily, Leo Lahti & Mikko Tolonen (2019) Reconstructing Intellectual Networks: From the ESTC's bibliographic metadata to historical material. In: Proceedings of the Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 4th Conference.