



Virtual Lab at the National Library of Estonia

Peeter Tinitis, Oct 6, 2022

DH Estonia 2022

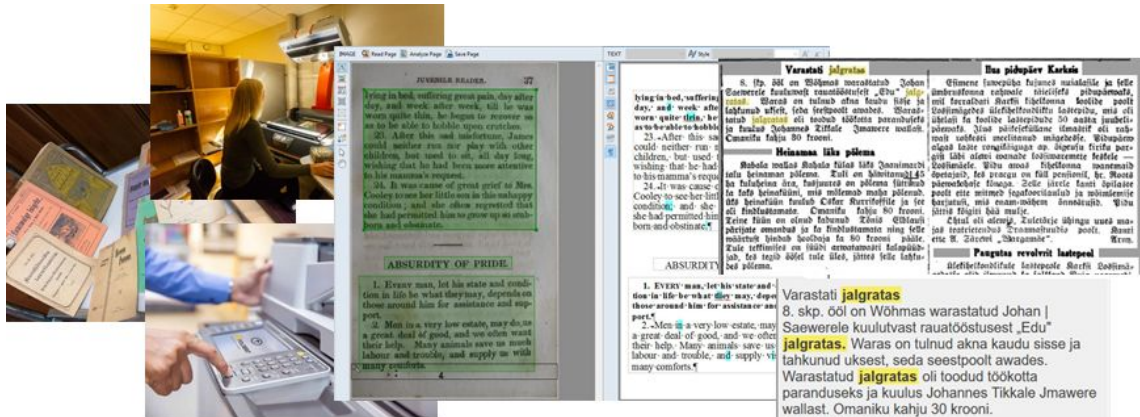
Libraries have large digital collections

Long time effort in digitization

- Focus on preservation
- Interfaces for reading

New frontiers in usage

- Collections as data
- Creative reuse



Example: Texts as data

RaRa newspapers
& periodicals

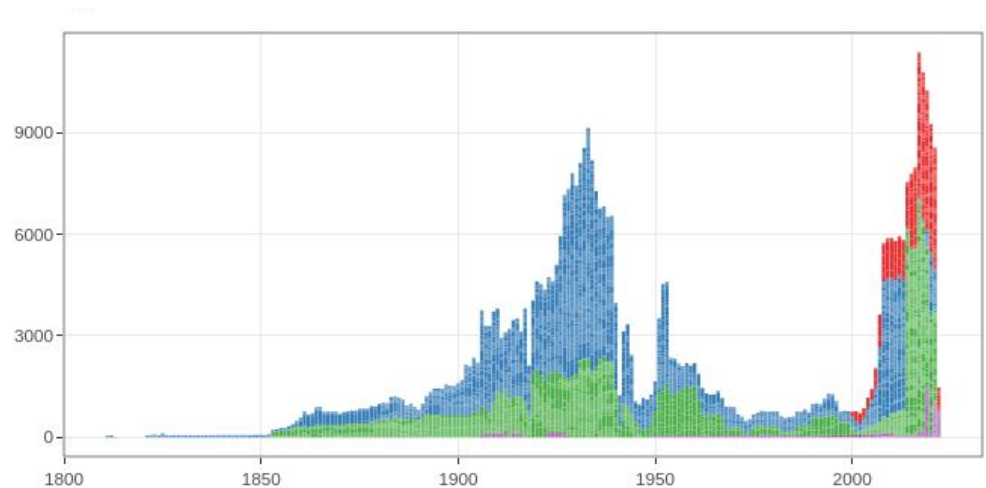
~9.1 M articles

~4.7 M pages

~465 K issues

~2601 publications

(+20-30% in last 2 years)





Researcher needs

Texts



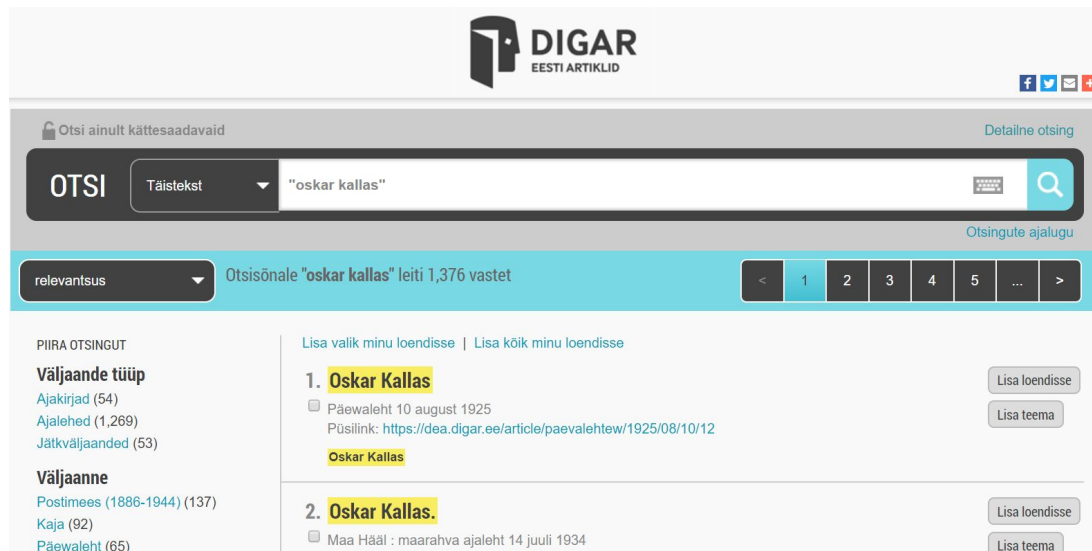
Metadata



Tools



Approaches to collections: Search engine



The screenshot displays the DIGAR search engine interface. At the top, the DIGAR logo (EESTI ARTIKLID) is visible alongside social media icons for Facebook, Twitter, and Email. Below the logo, a search bar contains the text "oskar kallas" and a search button. The search results are displayed in a list format, with the first result being "1. Oskar Kallas" from a newspaper article dated August 10, 1925. The second result is "2. Oskar Kallas." from a newspaper article dated July 14, 1934. The interface also includes a navigation bar with page numbers (1, 2, 3, 4, 5, ...) and a filter dropdown menu set to "relevantsus".

OTSI Täistekst "oskar kallas" Otsingute ajalugu

Otsi ainult kättesaadavaid Detalne otsing

relevantsus Otsisõnale "oskar kallas" leiti 1,376 vastet

PIIRA OTSINGUT

Väljaande tüüp

Ajakirjad (54)

Ajalehed (1,269)

Jätkväljaanded (53)

Väljaanne

Postimees (1886-1944) (137)

Kaja (92)


Päevaleht (65)

Lisa valik minu loendisse | Lisa kõik minu loendisse

- Oskar Kallas** Lisa loendisse
Päevaleht 10 august 1925
Püsiliik: <https://dea.digar.ee/article/paevalehtew/1925/08/10/12>
Oskar Kallas Lisa teema
- Oskar Kallas.** Lisa loendisse
Maa Hääl : maarahva ajaleht 14 juuli 1934 Lisa teema



More complex queries?



The screenshot shows the DIGAR search interface. At the top right is the DIGAR logo (EESTI ARTIKLID) and social media icons for Facebook, Twitter, Email, and a plus sign. Below the logo is a search bar with the text "names that cooccur with oskar kallas". To the left of the search bar is a dropdown menu labeled "Täistekst" and a button labeled "OTSI". To the right of the search bar is a search icon and a keyboard icon. Below the search bar is a button labeled "Otsingute ajalugu". Above the search bar is a lock icon and the text "Otsi ainult kättesaadavaid". To the right of the search bar is a link labeled "Detailne otsing". Below the search bar is a teal banner with the text "Otsisõnale names that cooccur with oskar kallas leiti 0 vastet".

Interfaces & libraries

Delpher

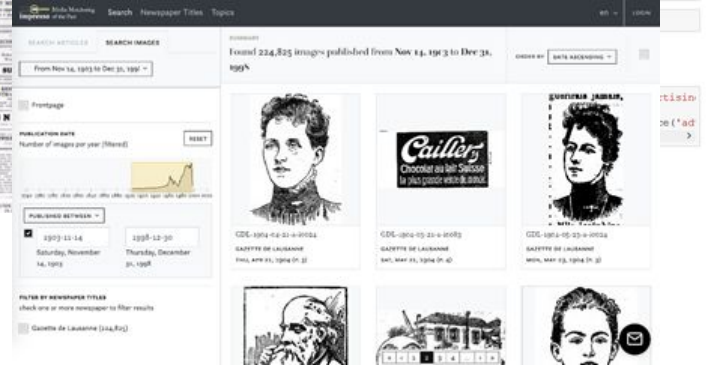


Show when the articles were published

```
In [ ]: alt.Chart(df).mark_line().encode(
  x='year(date):T',
  y='count()',
  tooltip=[alt.Tooltip('year(date):T', title='Year'), alt.Tooltip('cou
).properties(width=600)
< >
```

Find the longest article

```
In [ ]: # Which is the longest article(s)?
df[df['words'] == df['words'].max()]
```



Open Science movement

FAIR data

- Not just open, but findable+usable

Open Science Movement

- FAIR in science

Make analyses transparent,
interoperable, reusable



(Heunis 2020)



Researcher needs

Texts



Metadata



Tools



GLAM Labs

GLAM Labs community
(galleries, libraries, archives, museums)

Creative uses of data.

Computational access to digital collections



Virtual Lab at RaRa

Working towards from data to use

- Access points
- Data enrichment
- Case studies

Learning from international examples:

Creative Europe: Open Digital Libraries (with Dutch Royal Library and Austrian National Library)

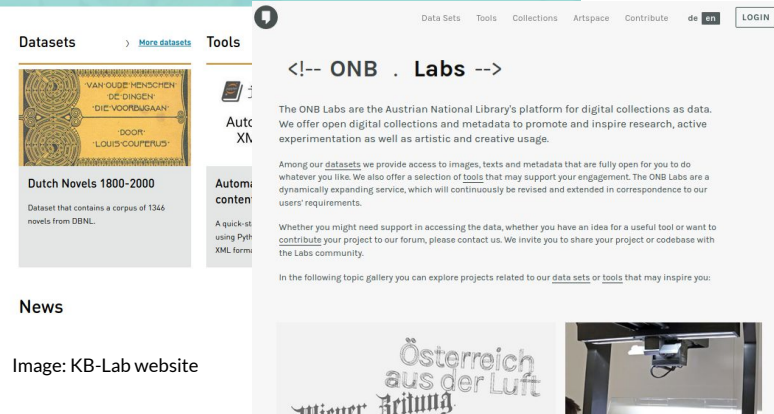
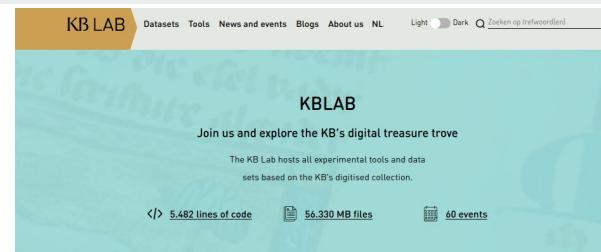


Image: KB-Lab website

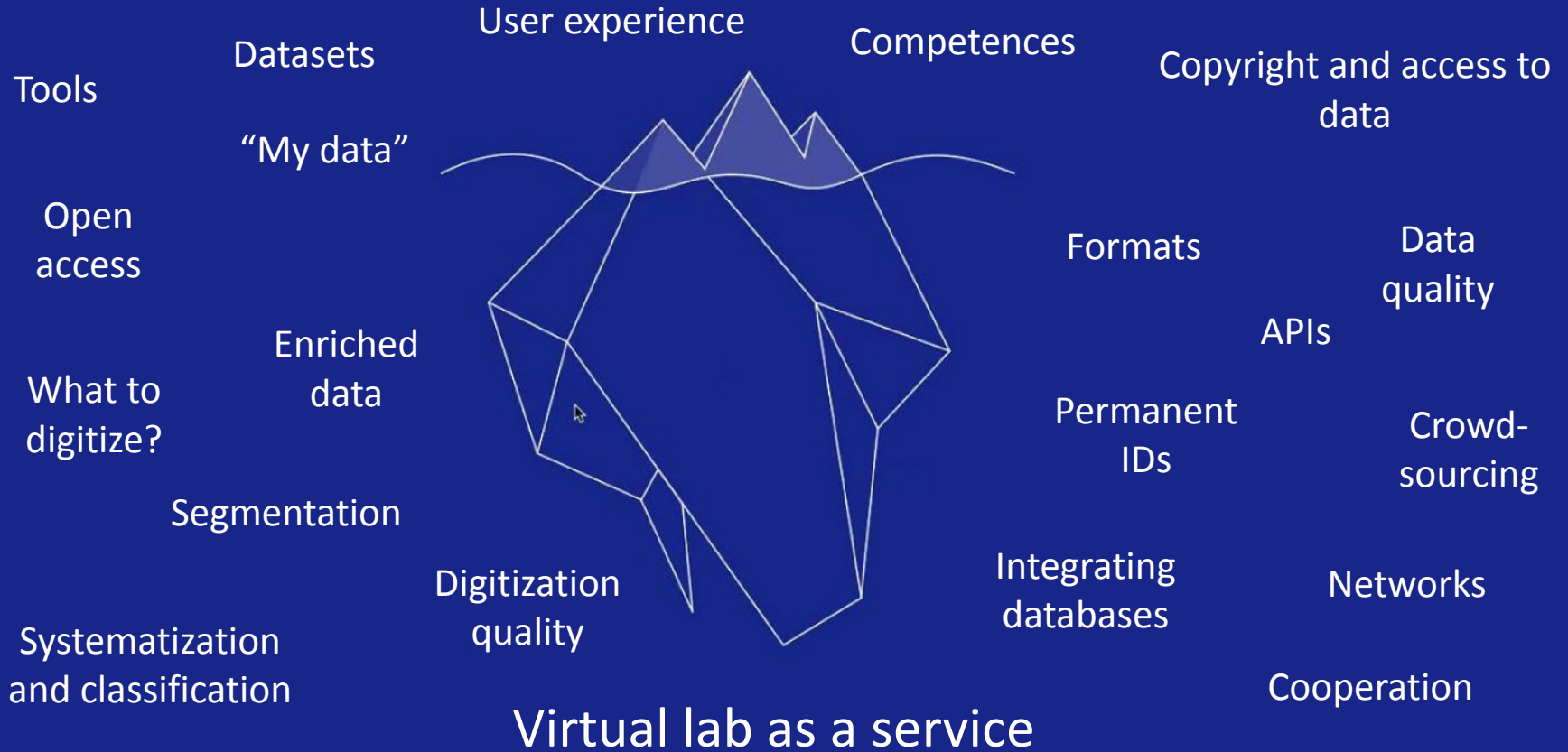
Image: ÖNB-Lab website



Steps on the way

- Making the lab (last 2 years)
 - Service design (2021)
 - Mapping the needs - interviews with representative users and stakeholders
 - Reflecting and designing the plans on the basis of this
 - Legal analysis (2022)
 - What can we do in which limits
 - Platform (2022)
 - Updated website that caters for this (data, tools, case studies)
 - Migrating and making (2022 onwards)
 - Datasets and tools

Virtual lab as a platform





Data available, data planned

Estonian National Bibliography (enriched publication metadata, people and organizations)

Digital archive text collection - metadata, fulltext access, ngrams (periodicals, books)

Thematic collections (e.g. images of postcards, parliamentary collections etc)

Goal: multiple formats where possible



Tools available, tools planned

Comfortable access to full texts and metadata (jupyter notebooks)

National Bibliography metadata explorer (point & click interface)

Ngram search on newspapers (like google ngrams)

APIs, SPARQL etc

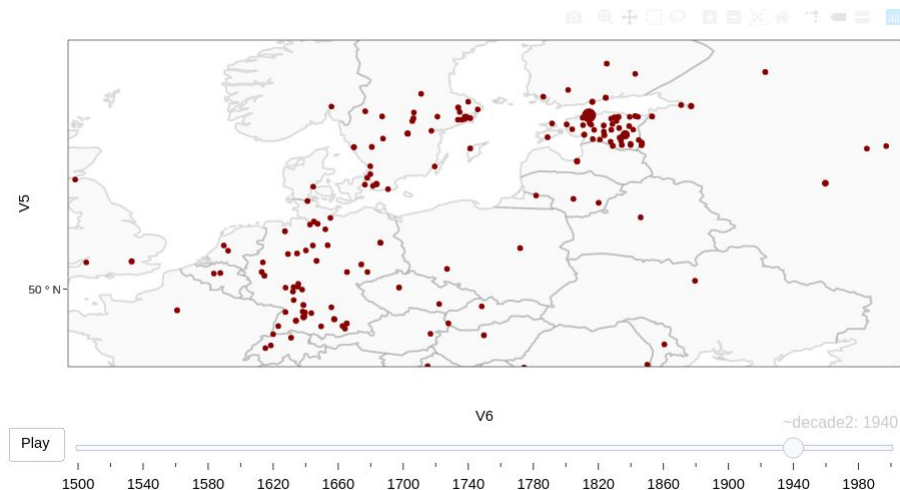
Bibliographic metadata explorer (work in progress)

Explore aspects of the bibliographic data

- Here, enriched with geoinfo
- But also just explore the contents

Get a better understanding of

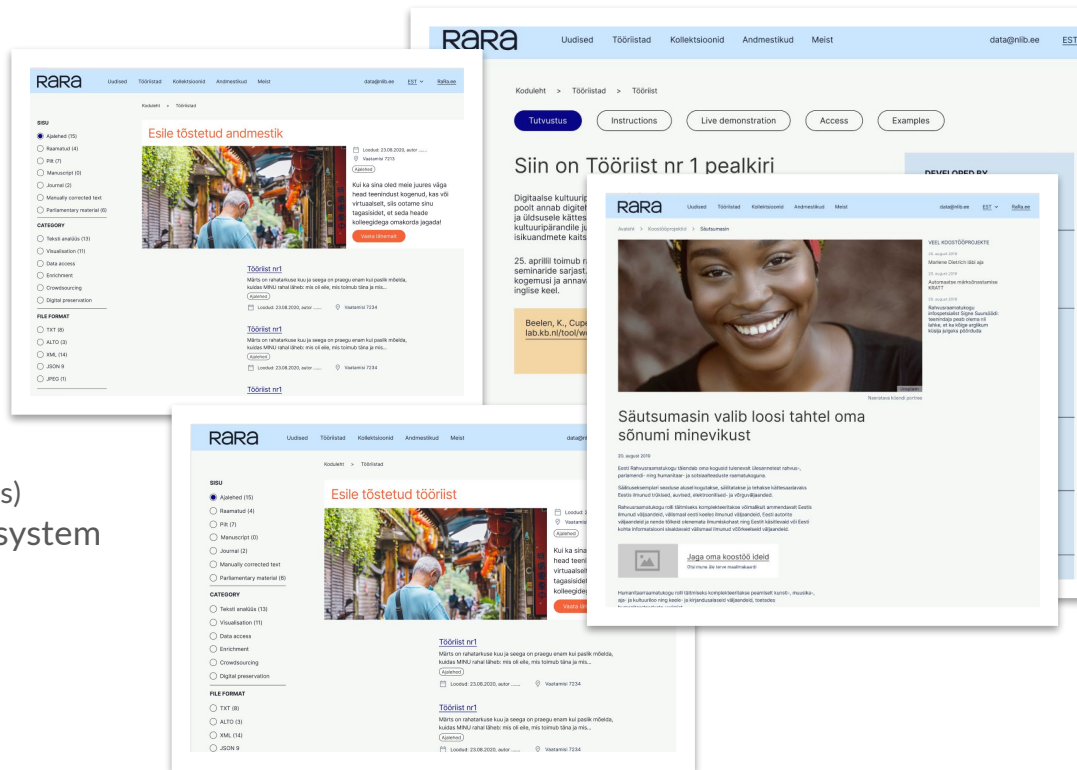
- Dataset (gaps and biases)
- Cultural history



Virtual lab

Data, tools, case studies

- Website due to release in 2022
- Finding creative uses
 - Mapping what's being done
 - Encouraging use (scholarships, prizes)
- Getting the work done back into the system
 - Derived & enriched text
 - Algorithms and tools made





A call

If you want to help! If you see what you like or want to show how to do better.

Talk to me after or e-mail at data@nlib.ee.

We may have a job for you. :)



Thank you

From the team at the National Library of Estonia

& only possible through the work of generations of librarians

