

ArchXai

AI ASSISTED INFORMATION REQUEST TOOLSET

FINDINGS & RESULTS - HANDLING INFORMATION REQUESTS IN ARCHIVES

Design process 2

Information request processes in 3 countries 3 - 7

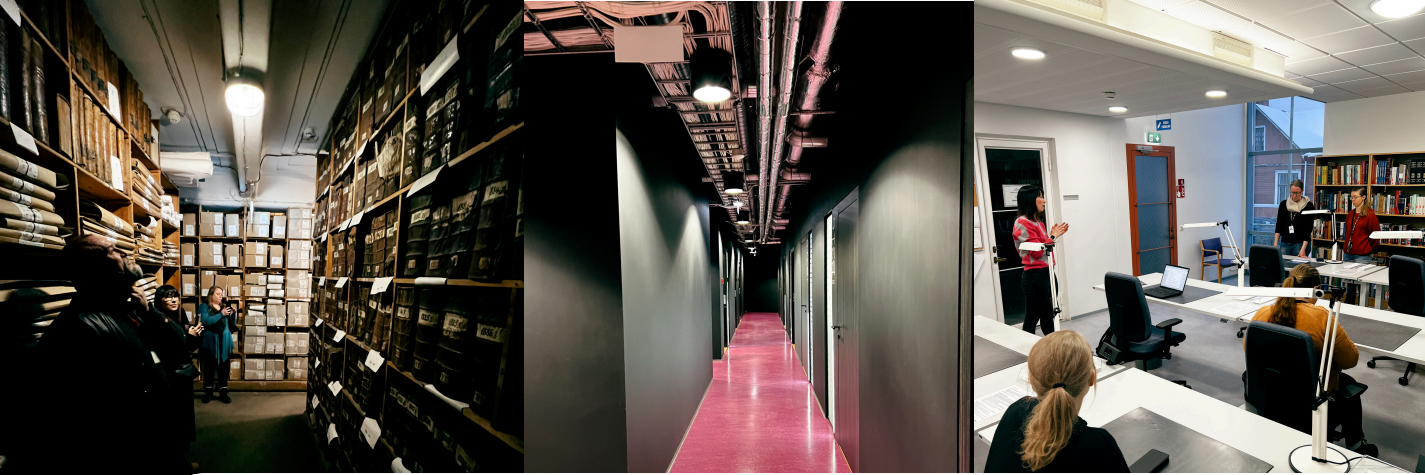
Customer segments 8

Bottlenecks 9 - 12

Needs for the Ai-tool 13

Fears about the Ai-tool 14 - 15

Questions about the Ai-tool 12



68 313
INFORMATION
REQUESTS IN 3
COUNTRIES (2025)



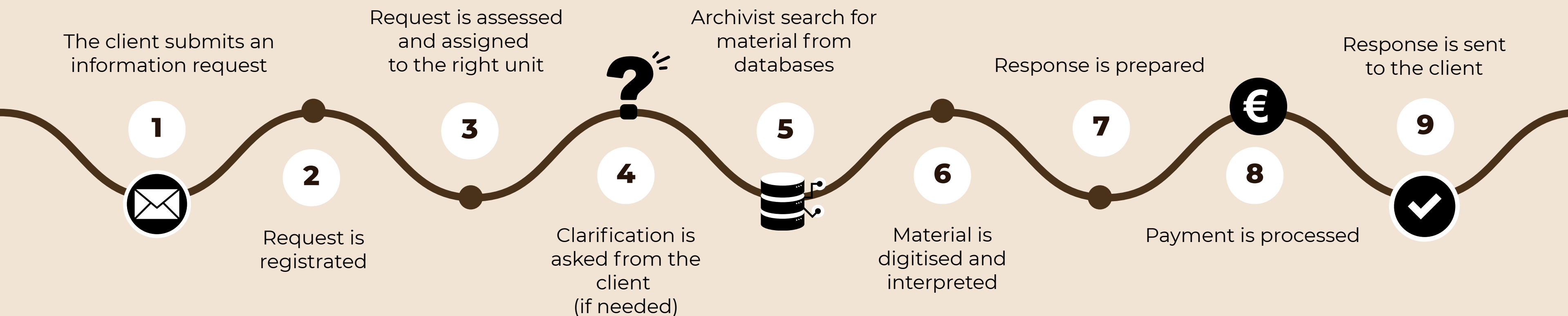
206
ARCHIVISTS
WORKING IN 3
COUNTRIES

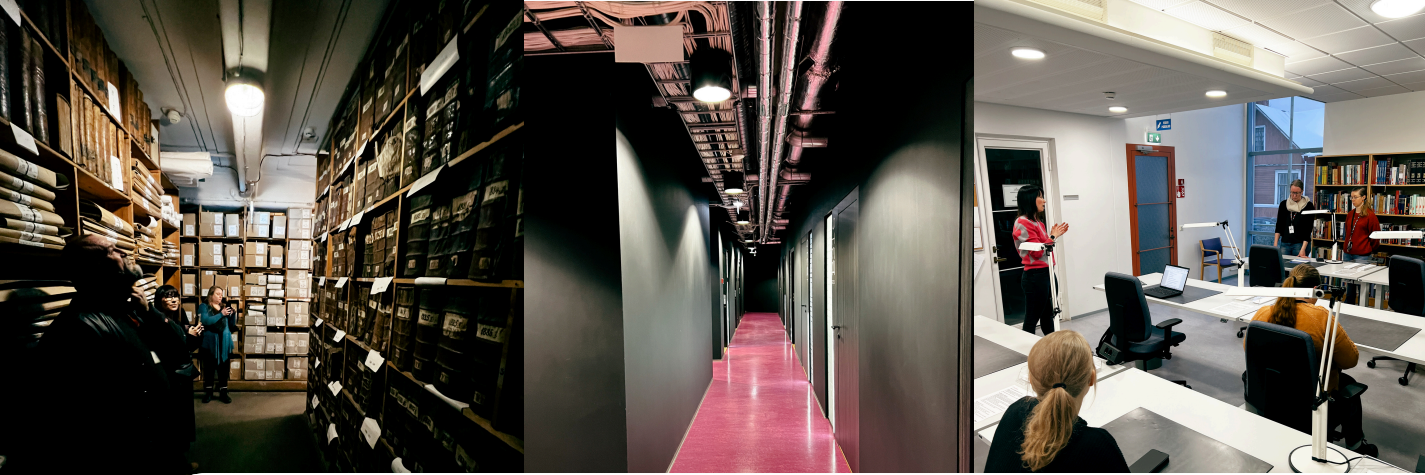


in 5 to 30 days
RESPONSE TO
THE CUSTOMER

Information request timeline

Estonia, Latvia, Finland





68 313

INFORMATION
REQUESTS IN 3
COUNTRIES (2025)



206

ARCHIVISTS
WORKING IN 3
COUNTRIES



in 5 to 30 days

RESPONSE TO
THE CUSTOMER

Information request bottlenecks

Estonia, Latvia, Finland



The client submits an information request

1

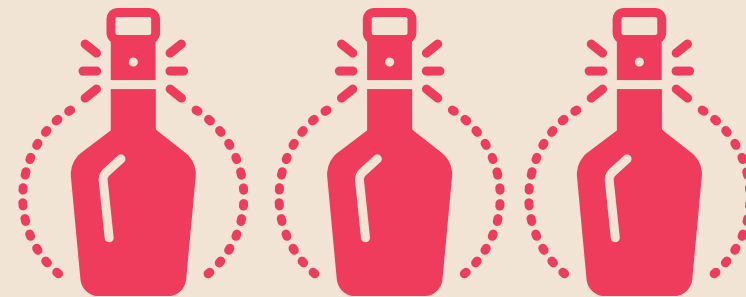


Request is registered

2

Request is assessed and assigned to the right unit

3



Archivist search for material from databases

5



Material is digitised and interpreted

6

Response is prepared

7

Payment is processed

8



Response is sent to the client

9



Shared Customer Segments



GENEALOGICAL & HISTORICAL RESEARCHERS

- birth records (time of birth)
- family history, citizenship, and related information

Clients often lack the necessary background information to make an information request.

These information requests are often very broad and time-consuming. They require extensive communication with the client and rely on the archivist's tacit knowledge.



ADMINISTRATIVE AND LEGAL RECORDS CLIENTS

- buildings and properties
- inventory of the estate
- court documents

Clients usually know exactly what they are looking for, making these requests easier and faster for the archivist to process.

These clients are often regular and experienced archive users, e.g. public authority.



The Bottlenecks

CUSTOMERS KNOWLEDGE ABOUT MAKING THE INFORMATION REQUEST IS WEAK

- The clients do not know what kind of information Archives have.
 - *Customers are unfamiliar with the correct archive or unit, and requests are often sent to the wrong unit.*
- Customers do not read the instructions.
- Requests submitted are often too broad - the customers are unable to narrow down the information request and this requires additional question.
 - *Common question: "I want something about my grandfather" – this is typical and too large request to handle.*
- Customers do not understand the archival logics and assume that everything can be found by providing just one name.
 - *There are hundreds of people with a same name: the client doesn't provide additional information such birth date and place.*
 - *Information requests often lack the essential details needed for the search (e.g. first name, surname, maiden name, date/place of birth).*
- The terminology used in application forms is difficult for customers to understand.
 - *Customers do not realise which relevant information they may already have, and therefore do not provide it.*
- Chatbots designed to guide customers do not function properly.



The Bottlenecks

ARCHIVISTS POSSESS TACIT KNOWLEDGE THAT DEVELOPS PROGRESSIVELY THROUGH PROFESSIONAL EXPERIENCE AND EXPERTISE

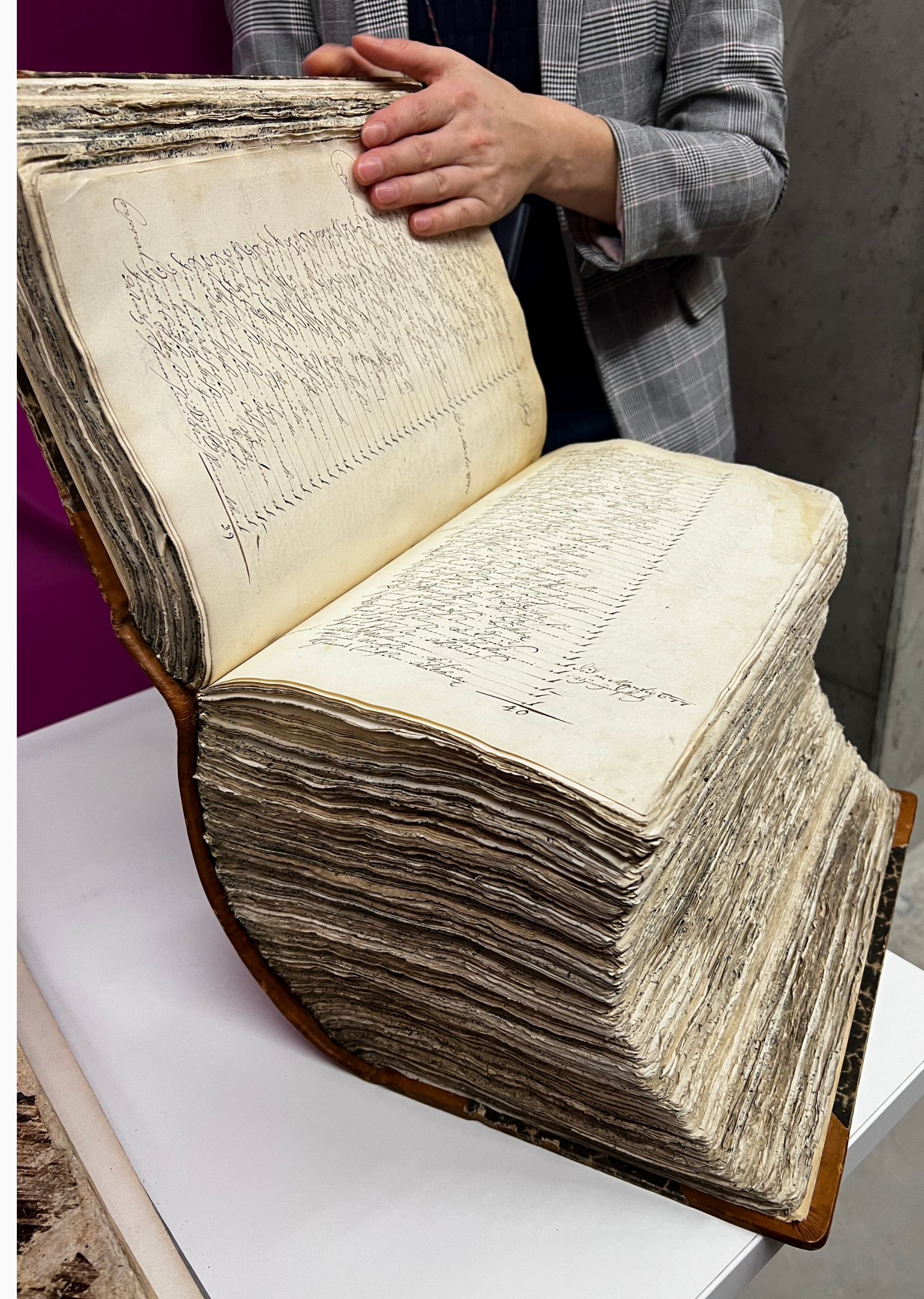
- Tacit knowledge is difficult to describe or document systematically.
- The work of an archival specialist is not taught at university.
- Knowledge from experienced colleagues is frequently consulted (“borrowing brains”).
- Experts use their tacit knowledge to choose the right database for information requests; each database has its own specialists.
- “Same name can be written in many different ways, and you just have to know.”
- “Texts might be illegible.” Tacit knowledge is pivotal when interpreting the old texts and translations.
 - *For example at the National Archives of Finland, there are only a few individuals who are able to read Old Russian. Latvia, there are relatively few historians who are both interested in and capable of conducting research on materials from the medieval period and the era of the Russian Empire. In Latvia there are also dedicated specialists for the Soviet period.*



The Bottlenecks

THE DATABASES ARE DISTRIBUTED ACROSS DIFFERENT ARCHIVAL INSTITUTIONS.

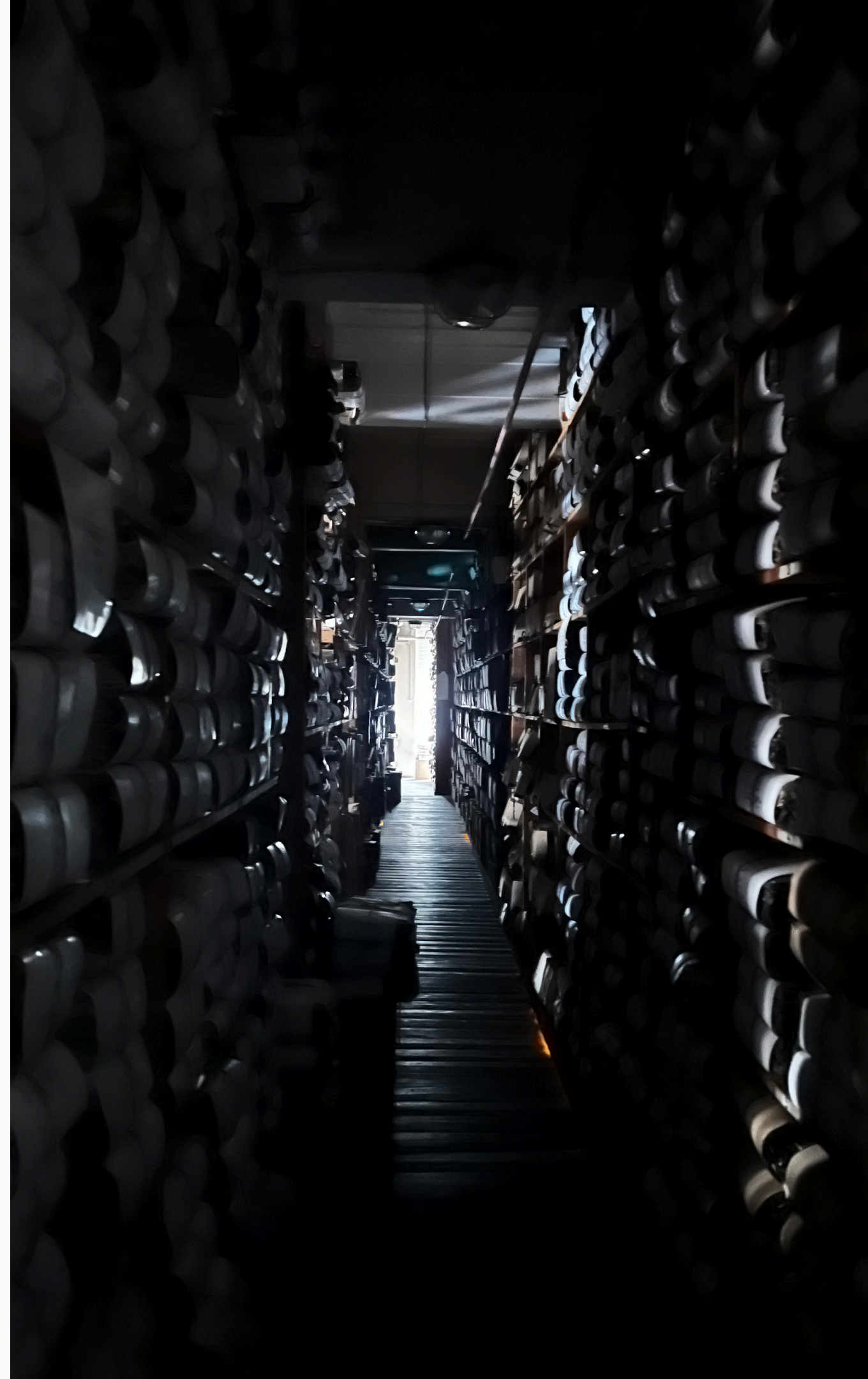
- The databases are fragmented, and locating relevant information within them relies on the professional expertise of the archivist.
- In each country databases utilized are for example public national land survey authority, National Libraries, the population register, burial register etc. In addition historical maps, churchbooks and registers maintained by volunteers.
- There are also databases that can be used by the citizens for free. Databases, their contents, and their intended uses are not readily understandable, accessible, or engaging to a broader public audience.



The Bottlenecks

THE INTERPRETATION OF HISTORICAL TEXTS IN RUSSIAN AND GERMAN.

- In each archive there are specialists for different languages.
- Information requests typically involve materials in multiple languages.
- At the National Archives of Finland, there are only a few individuals who are able to read old Russian.
- Historical German-language texts are common in Latvia and in Estonia. It is also not possible to work in an archive without proficiency in Russian.
- Usually the requests involve multiple languages.
- Historical Russian and German place names and their variants are difficult. Same name can be written in different ways in different times.
- Texts might be illegible and tacit knowledge is pivotal when interpreting the old texts and translations.



Needed Key Features for AI-tool



Text recognition OCR + HTR

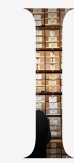
The tool must accurately interpret both printed (OCR) and handwritten text (HTR), which is essential for historical materials with varying name forms.

"The challenge does not lie in recognizing the text itself, but rather in being able to interpret and translate it accurately."

"Text recognition has long been anticipated ..."

"Digital catalogues and the integration of optical character recognition (OCR) into them."

"The recognition of text in digital materials. It must be taken into account that a name may appear in different documents in varying forms, as personal names have historically been subject to change."



Indexed documents

Extensive, consistent indexing makes large digitized collections easier to search, filter, and navigate.

"Indexed documents: the greater the number of indexes available, the better!"

"It would be highly beneficial to have a substantial amount of digitized material, as well as a dedicated facility responsible for producing copys."



Databases

Unified access to multiple archival databases reduces dependency on expert knowledge and simplifies information retrieval.

"To get all the databases to use at once!"

"A new archivist does not know which database to use."

"It is difficult to conceptualize what kind of tool could be developed, given the multitude of databases involved and the knowledge that currently resides with archival professionals themselves."



Tacit knowledge

The system should capture archivists' implicit expertise to support tasks like interpreting context or conducting genealogical research.

"A chatbot designed to capture and convey tacit knowledge would be highly valuable. However, the question remains as to who would input the knowledge and on what criteria."

"An intelligent AI system could provide substantial support, for example in genealogical investigations."

"Earlier information requests could also be examined to uncover similar patterns..."



Chatbot for the customer

An AI chatbot guides users through multiple databases and helps them find relevant information without requiring deep local knowledge.

"Efforts have been made to produce instruction materials; however, locating relevant information continues to require extensive local knowledge."

"Customers don't know how to search for information in archives."

"Archives use terminology that can be difficult to understand."

Fears about Ai-tool

- AI's ability to interpret historical texts accurately, including context, time periods, and illegible materials.
- How reliable AI is in multilingual processing, especially for Russian and German texts and historical name variants.
- AI's capacity to replace or sufficiently support archivists' tacit knowledge and professional judgement.
- Practical limitations related to fragmented databases and restricted access to key authoritative data.
- Uncertainty about the overall efficiency and cost-effectiveness of AI compared to human expertise.



Examples of Open Databases with Digitized Historical materials

FINLAND

- Astia – published digitized archival materials
- Search for Finnish Court Records
- Karelia Database
- “Soviet Union as a Fate” Database
- Nuohtti – 30,000 digitized Sámi documents and photos
- WarSampo
- WarVictimSampo 1914–1922
- Database of Finnish Medieval Documents
- Heraldic Database
- Sotamuistosampo – Veterans’ stories

ESTONIA

- SAAGA – Digitized church records, census lists, maps, civil records, remove maps.
- AIS (Archival Information System), – Central database for all archival holdings
- Arkaader (Estonian Film Heritage), –Digitized feature films, documentaries, animation, newsreels.
- Meediateek – audiovisual collections
- Maps database
- Vallakohtud – court records crowdsourcing initiative
- Vabadussõda – war of independence crowdsourcing initiative

LATVIA

- Raduraksti – digitized church registers, census records, passports, university student lists, etc.
- Diaspora – digitized card indexes and other materials relating to the Second World War period and refugees from Latvia.
- See, Hear Latvia! – digitized photographs, audio recordings, feature films, documentaries, newsreels, etc.
- KGB Documents – digitized documents on Soviet-era repression and the activities of the KGB.
- Digital Library of Latvia – digitized materials from various cultural heritage institutions in Latvia, including some collections of the National Archives (e.g. minutes of the Riga Town Magistracy)