

# Handwritten Text Recognition in the Archives

Maria Kallio  
Senior Research Officer  
[maria.kallio@arkisto.fi](mailto:maria.kallio@arkisto.fi)



THE NATIONAL ARCHIVES  
OF FINLAND

**READ**  
RECOGNITION & ENRICHMENT  
OF ARCHIVAL DOCUMENTS





- READ (Recognition and Enrichment of Archival Documents) 2016-2019
- Horizon 2020 Research and Innovation Infrastructures
- 14 partners
- The core mission is to make archival material more accessible through the development and dissemination of Automated Text Recognition (ATR)



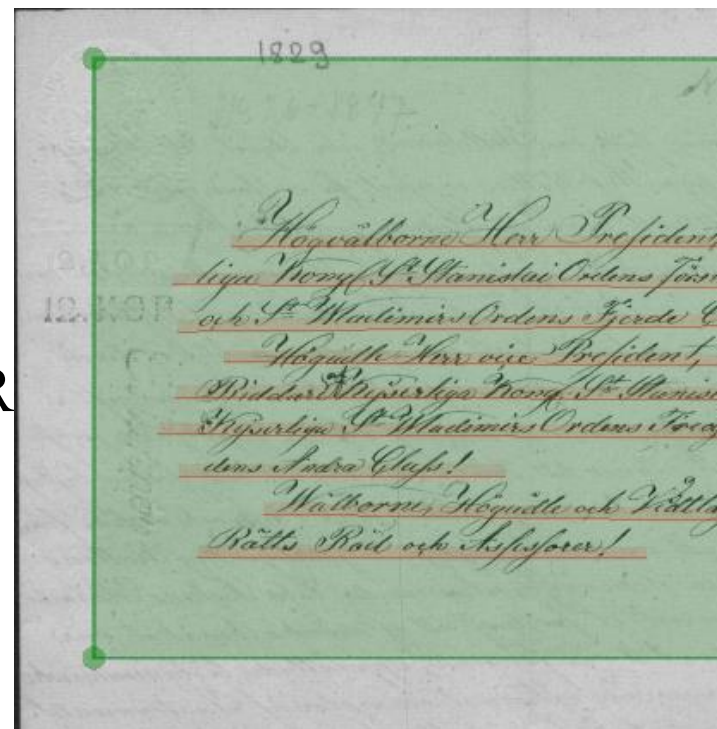
# New research infrastructure – Transkribus



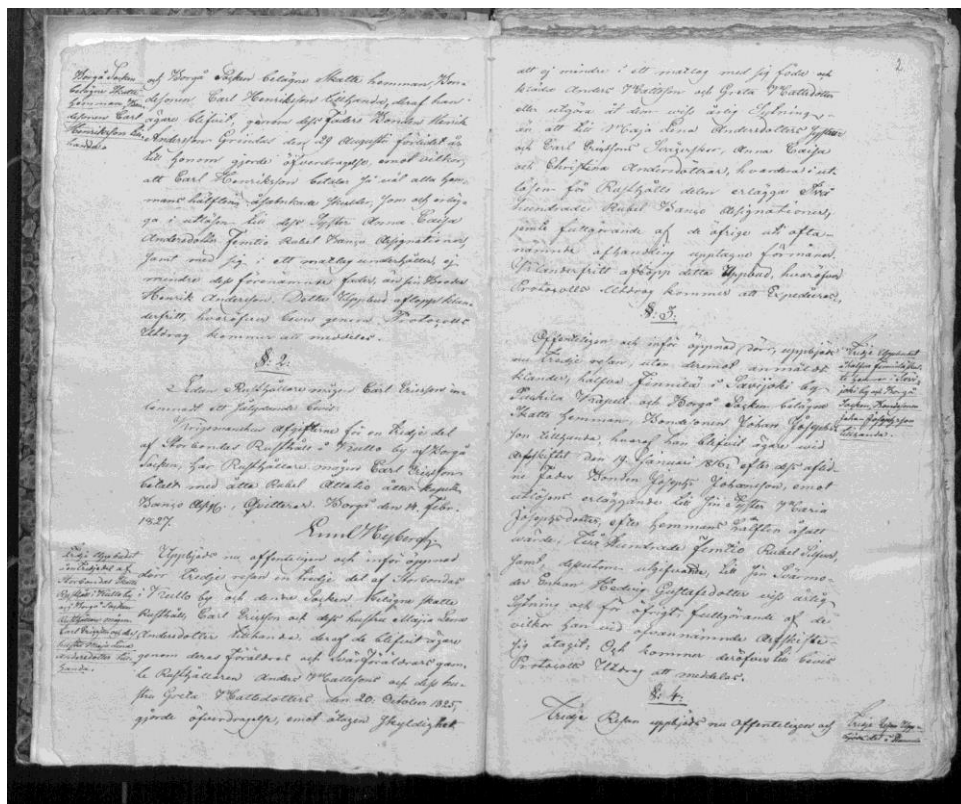
- Automated Text Recognition and other tools
- Enabling computers to automatically transcribe and search historical documents – of any date, language and layout!

# Automated Text Recognition (ATR)

- Machine learning (Neural networks)
- Learns to read specific styles of writing
- Needs to be trained by being shown document images and transcripts
- More training data → more accurate ATR
- Create a ATR model to transcribe and search a collection of documents



# Court Records 19th century



- Collection of 660 000 images
- Current HTR model CER 9%
- Most of the GT produced by volunteers



Med förestående afhandling förklarar jag  
mig till alla delar nöjd, samt förbinder  
mig äfven att uppfylla densamma uti  
alla delar, som hämnad försäkres. Ohkon som  
gfon.

Erik Johanson Suojala  
Kunde som  
(barn.)

Till vittnen

Erik Mattsson Alutia. Fredrik Carlsson Mikkola  
(barn.) Kunde ifrån Ohkon (barn.)

- 23 Mmed förestående afhandling förklarar jag ↵
- 24 nig till alla delar nöjd, samt förbinder ↵
- 25 nig äfven att uppfylla densamma uti ↵
- 26 alla delar, som hämnad försäkres. Ohkon som ↵
- 27 gfon ↵

Keyword Spotting Results

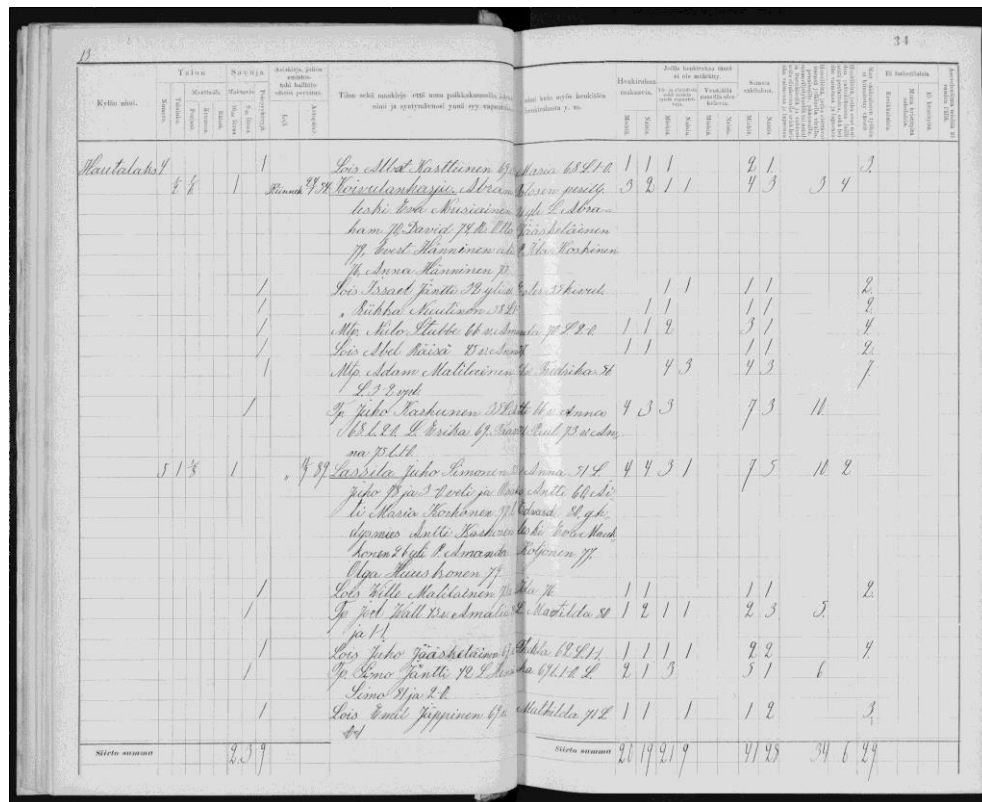
"förmyndare" (519 hits) "äktenskap" (147 hits)

Confidence	Page Nr.	Line transcription	Previ...
0.2496	1041	Ftörmyndares	
0.2494	156	manna lagens städganeden bom förupudarenädd	
0.2486	863	feor gndare for sin onm yndig	
0.2482	163	de mot Bonden Nikadenmn Neloson Is Bruuni, Loeyrdare för	
0.2452	447	di förnni ndae b,fåttni p skilas	
0.2446	483	älle Carn till färmyndaren :	
0.2439	802	söärmyndne ör sknna dredrika Alianders datten	
0.2437	363	ganden bim foyrdalevåt, till efterrattelse.	
0.2429	816	Lovisa Borgström sig vara villig att attråda förmynder.	
0.2380	479	Förmyndares	
0.2371	1026	gens talganden om örmyndarevart, gy=	
0.2351	867	salorm vodon soig dare sor b alde	
0.2287	274	gjens tlgmden oom förvmjndarvar till o=	
0.2285	942	å igende sörnymdarno tillförörtrande	
0.2244	449	fönngndare or bernälde ngndirg utse och	
0.2215	449	sördngndare sön	
0.2198	354	tillförännåd sönmjrdere för sormessevage Aen=	
0.2190	812	Hemmala sig villig att ifrågakamne förmyndlr	

Preview

Close

# Poll Tax Records 19th century



The image shows an open historical poll tax record book. The left page is numbered 33 and the right page is numbered 34. The book contains handwritten entries in Finnish, organized into columns. The columns include names of individuals or households, their addresses, and numerical values representing tax amounts. The entries are written in a cursive script typical of the 19th century. The book is bound in the middle, and the pages are slightly aged.

- Collection of 1,5 million images
- In house GT production
- Layout analysis and document understanding → digital register and research data



# Our goals

- To provide 19th century Court Records in searchable and computer readable form by the end of the project
- To enable new versatile possibilities for research by giving our customers the chance to download all the produced data for their own use
- To integrate text recognition in to our digital services on large scale



# For more information:

**Transkribus website:** <https://transkribus.eu/>

**READ website:** <https://read.transkribus.eu/>

**Contact:** [maria.kallio@arkisto.fi](mailto:maria.kallio@arkisto.fi)  
or [email@transkribus.eu](mailto:email@transkribus.eu)

**Twitter:** @Transkribus