COMPUTATIONAL HISTORY AND THE TRANSFORMATION OF PUBLIC DISCOURSE IN FINLAND, 1640–1910 (COMHIS)
Computational History and the Transformation of Public Discourse in Finland, 1640–1910 (COMHIS)

Consortium partners:
• National Library of Finland, Centre for Preservation and Digitisation
• University of Helsinki, Faculty of Humanities
• University of Turku, Dept of Information Technology
• University of Turku, Dept of Cultural History
Research teams

• National Library of Finland: Kimmo Kettunen (PI), one post-doc
• University of Helsinki: Mikko Tolonen (PI), Leo Lahti, Jani Marjanen, Hege Roivainen
• University of Turku: Hannu Salmi (PI), Tapio Salakoski (PI), Heidi Hakkarainen, Asko Nivala, Heli Rantala
Objectives

Reassessing the scope, nature and transnational connections of public discourse in Finland 1640–1910.

Complementary approaches:
- Library catalogue metadata
- Full text-mining
- All the digitized Finnish newspapers and journals published before 1910.
COMHIS Overview

**Bibliographic metadata**
Publishing trends and the development of public discourse

**Full text analysis**
Viral texts and social networks of Finnish newspaper publicity

*Open source data analytics & methodologies*

- Pioneer transparent and reproducible data analytics in the digital humanities
- Showcase the vast opportunities of quantitative analysis of digitized materials in the reinterpretation of key questions in historiography.
Quality examples of historical newspaper collections

- The British Library’s 19th century collection has an estimated word accuracy of 78 %
- The estimated word accuracy (word recognition rate) of Digi is about 70-75 %
- These are quite low figures but realistic for OCRRed old newspaper collections
Consequences

- Texts are hard to read for users
- Users may have difficulties in searching the collection
- Search results may be worse than expected anyhow
- Data mining and any further processing becomes more difficult
- Re-OCRing and post-correction of newspaper data needed, perhaps 80+ % word accuracy can be achieved
Publishing Trends
and the Development of Public Discourse

• Large-scale analysis of library catalogue metadata collections

• Intellectual geography and transcending of national borders
Cicero vs. Luther

Title count timeline for the top authors

Publication Decade

Author
- Cicero, Marcus Tullius (65-NA)
- Luther, Martin
- Luther, Martin (1483-1546)
Death of Turku (or/and mistakes in catalogue)?

Finnish publications in Turku, 1488-1828: title count

Title Count

Publication Year
Viral Texts and Social Networks of Finnish Public Discourse in Newspapers and Journals 1771–1910

• Developing text reuse detection and identifying cross-border flows between languages (Finnish/Swedish)

• Virality of newspaper and journal discourse in nineteenth-century Finland: cultural rhizomes and social networks
Example Cluster

Suometar 5 August 1864
Example Cluster

Suometar 5 August 1864

Reprinted six times:

Päivätär 6 August 1864
Mikkelin Wiikko-Sanomia 11 August 1864
Sanomia Turusta 12 August 1864
Tähti 12 August 1864
Hämäläinen 19 August 1864
Oulun Wiikko-Sanomia 20 August 1864
Dedicated open source ecosystems for the digital humanities

- Balance automatization & customization
- Open data analysis tools (R / Python / ...)
- Reproducible notebooks (Rmarkdown / iPython)
- Transparent workflows
- Best practices from computational sciences

https://github.com/rOpenGov/fennica
Cooperation

• Digital Humanities Centre at University College London
• NULab for Texts, Maps and Networks at Northeastern University, Boston
• Open Knowledge Finland ry.