

# OCTAVO: Analysing Early Modern Public Communication

## What

New understanding of the development of public discourse and knowledge production in early modern Europe.

## How

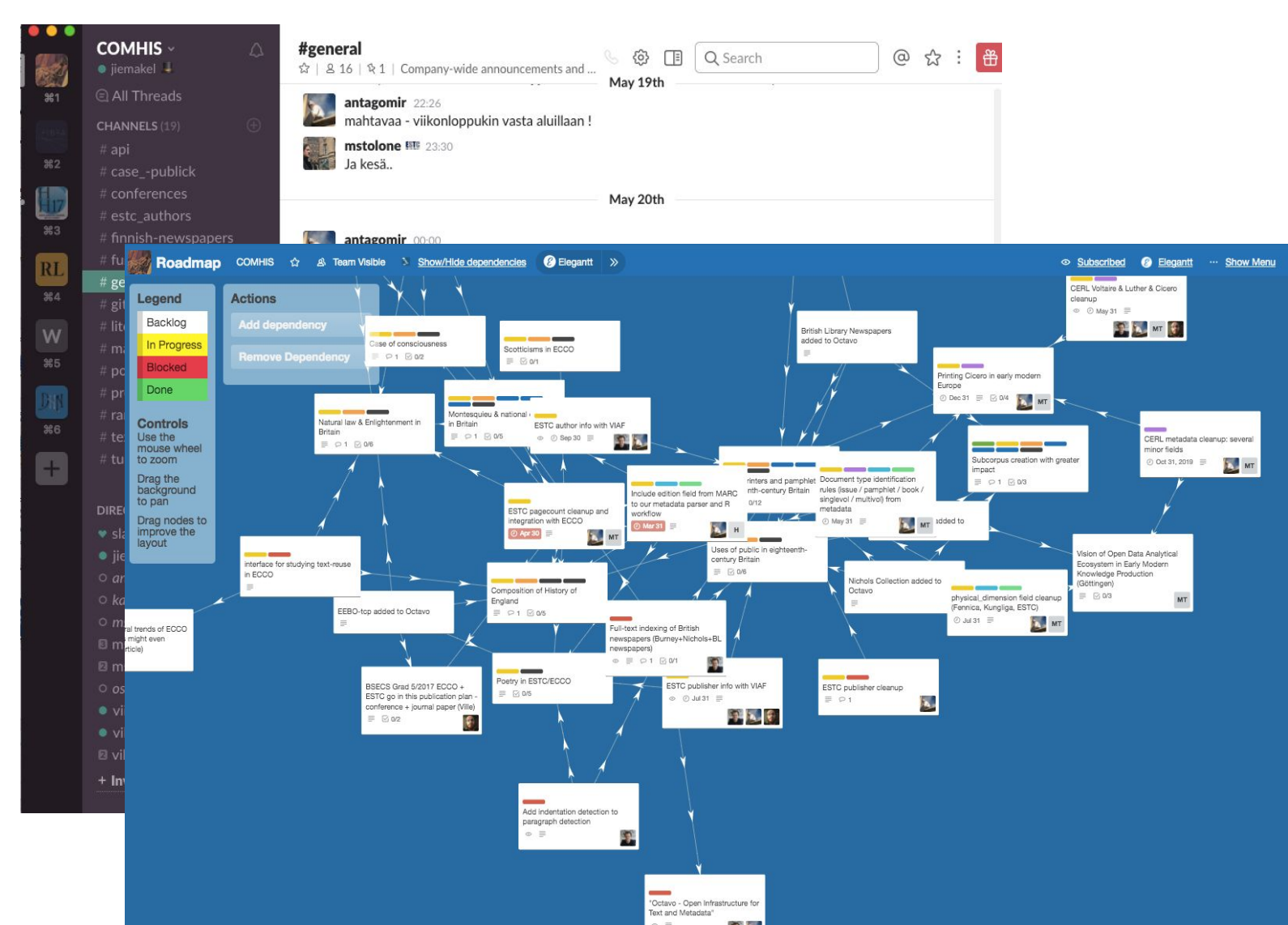
Building an analysis ecosystem Octavo based on metadata and full-text collections (Britain, Sweden, Finland, and CERL) to model Europe.

## Why

Big data approach to early modern resources is impossible without systematic cleaning and integration of metadata to full-text collections.

## COMHIS: Loose collective with aligned interests

The group is comprised of people from multiple fields whose research agendas are aligned: *computer scientists* who research open workflows, algorithms, and interfaces for humanities text and metadata; *linguists* who explore the relationship between words and concepts; and *historians* who are interested in conceptual and historical processes, who analyse data and results, and who frame needs for research method development.

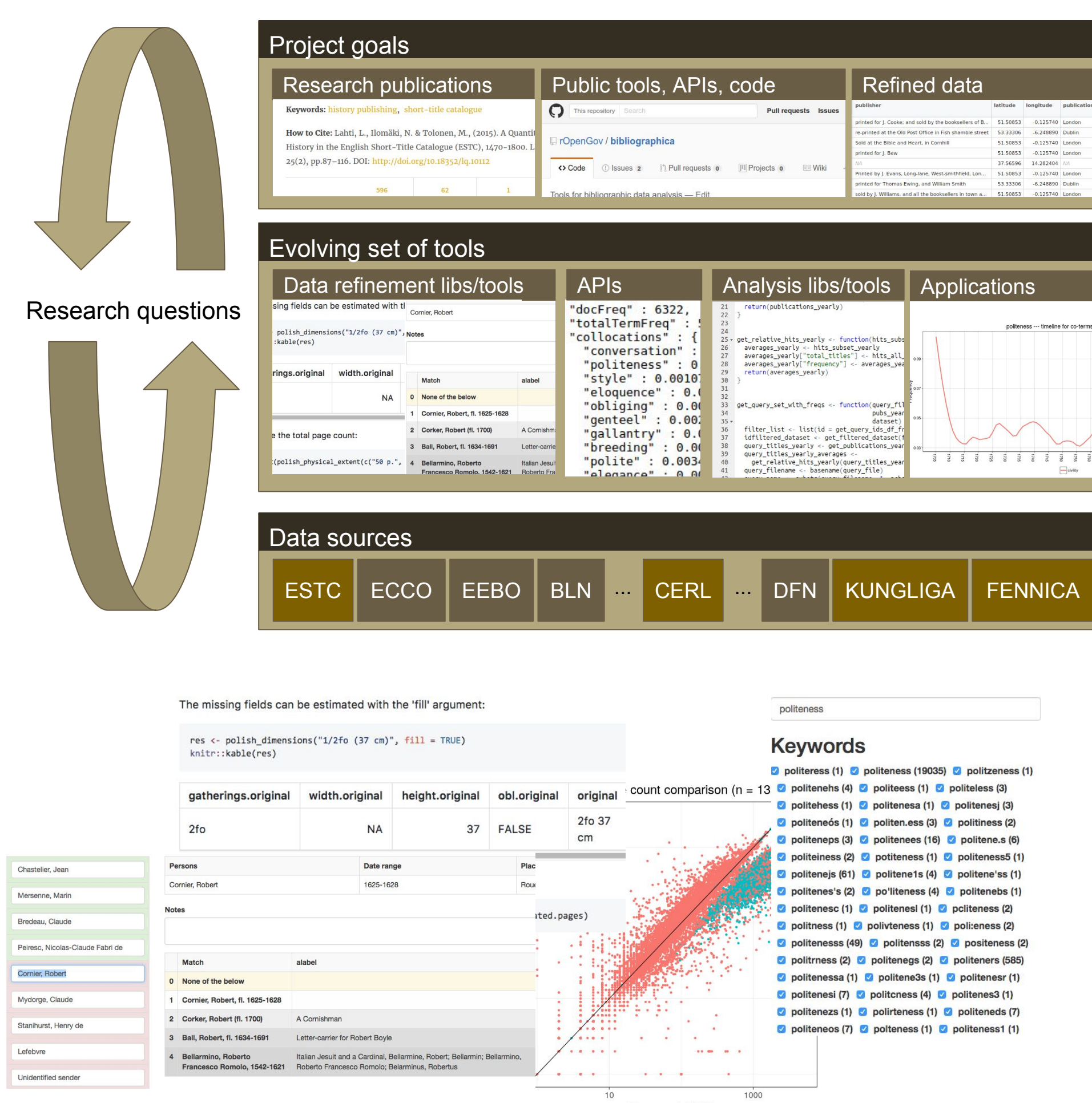


## OCTAVO: Shared open infrastructure

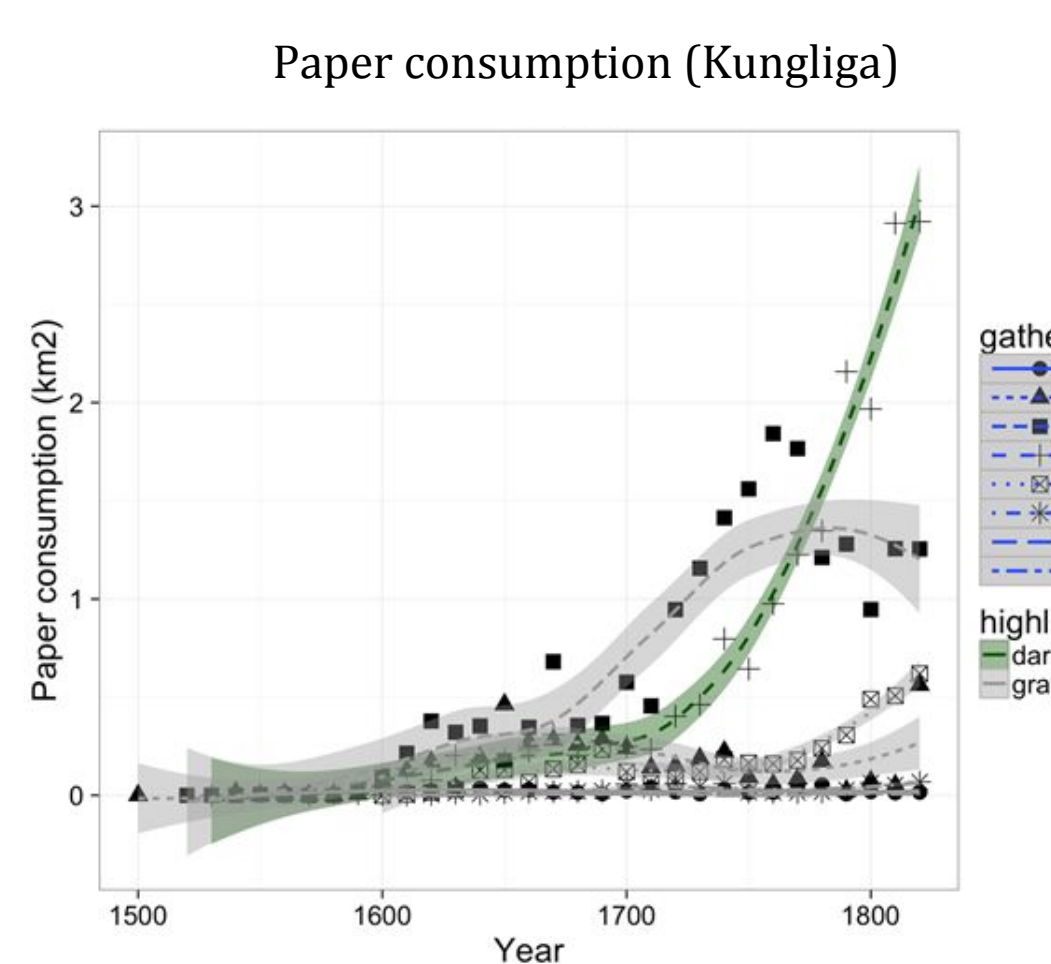
**Text and metadata:** Crossing from content to metadata and back allows accessing the contexts of a discourse. It is a way to ensure that many editions of a work do not bias content analysis and to analyse the material dimensions of a discussion.

**Cleaning up data:** 80 % of statistical analysis is tidying up the data. This is often neglected yet implicitly assumed.

**Creating infrastructure:** Functionalities are developed both as programming libraries and as end-user interfaces.



## The rise of the octavo-sized book



**Methods and data:** National library catalogues (ESTC, Sweden, Finland) contain rich metadata and follow a regular system of data collection. After rigorously cleaning the data, it is possible to compare how book sizes developed in the early modern period.

**Argument:** The rise of the octavo format during the eighteenth century relates to technological advancements and changes in reading habits. The pocket-sized octavo book allowed for introspective reading as opposed to reading out loud from a pulpit. At the same time, specialisation of public forums and ways of conceptualising public discourse also shifted.

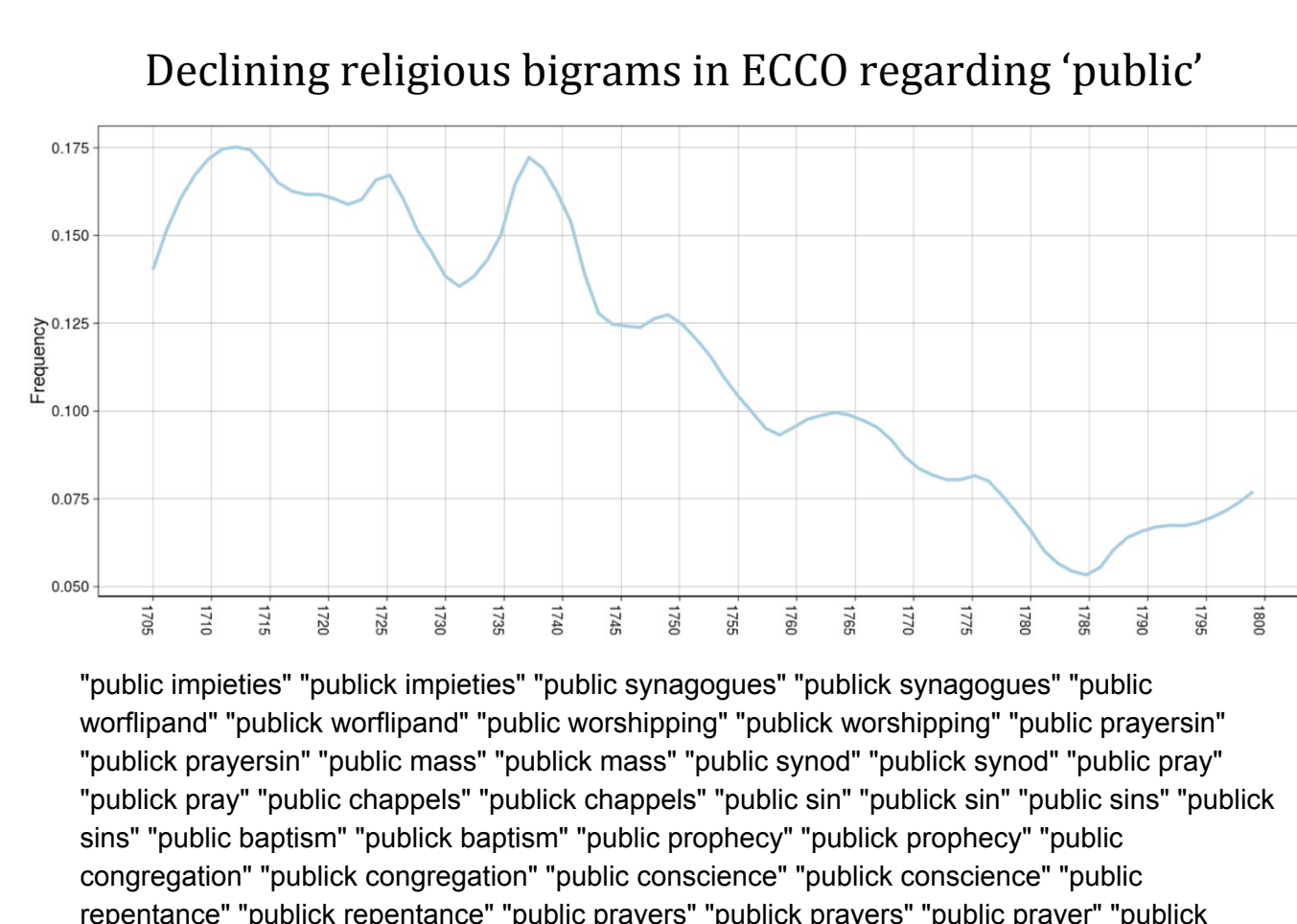
## Metadata analysis of publications



**Methods and data:** This statistical analysis combines metadata for books and Finnish-language newspapers between 1820 and 1910 to compare top publication places.

**Argument:** The nineteenth century was a period of strong growth in publishing in Finland. A city-by-city comparison shows that the growth in newspaper publishing happened largely through the spread of newspapers geographically. Book production, however, remained predominantly in Helsinki. Production of books and newspapers became two separate operations with different agents, practices, and audiences.

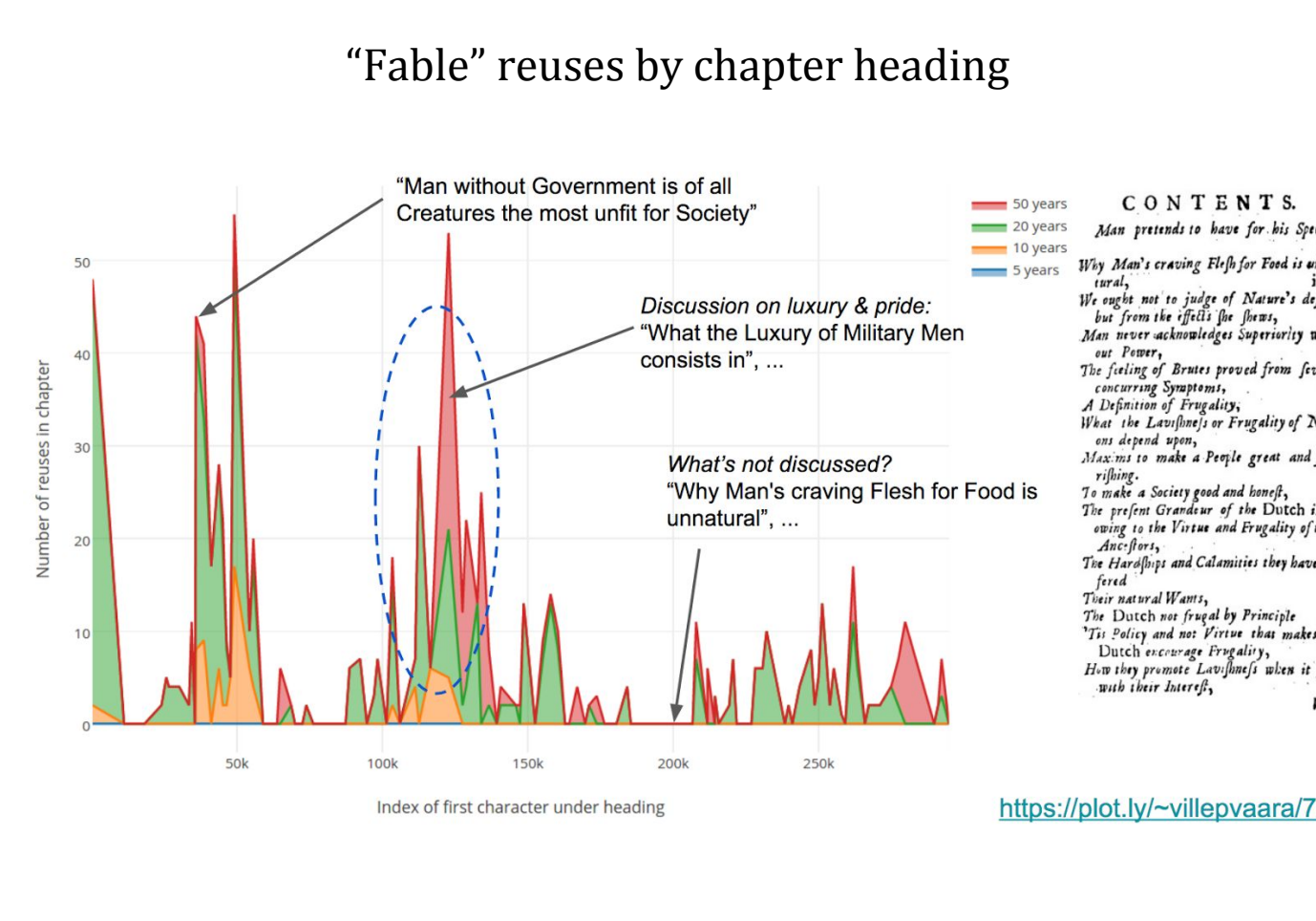
## Tracing the public sphere



**Methods and data:** Study of the distributional semantics of bigrams relating to 'public' in ECCO. A combination with ESTC metadata allows for subcorpus analysis.

**Argument:** Studying the use of bigrams relating to 'public' qualifies earlier notions of a changing vocabulary regarding the public sphere in eighteenth-century Europe. We may distinguish the conceptual change that took place by looking at which bigrams became more frequent and which declined in use. As a telling example, religious bigrams became overshadowed by e.g. political bigrams such as 'public opinion' or 'public spirit'.

## Text reuse of Mandeville's *Fable*



**Methods and data:** This analysis charts the reuse of Mandeville's *Fable of Bees* (1714). The data has been produced by running ECCO through BLAST software which enables working with the data despite OCR errors. The results have been enriched with ESTC metadata.

**Argument:** Text reuse tells us about reactions to a publication and both the sources and the influence of a work. In the case of Mandeville, we can see that certain chapters of the *Fable* were hotly debated, while others do not seem to have occasioned public discussion. This analysis also corroborates the claim that discussion of the *Fable* only really began after the expanded edition in 1723.



[comhis.github.io](https://comhis.github.io)

Antti Kanner, Jani Marjanen, Ville Vaara, Hege Roivainen, Viivi Lähteenoja, Laura Tarkka-Robinson, Eetu Mäkelä, Leo Lahti and Mikko Tolonen

