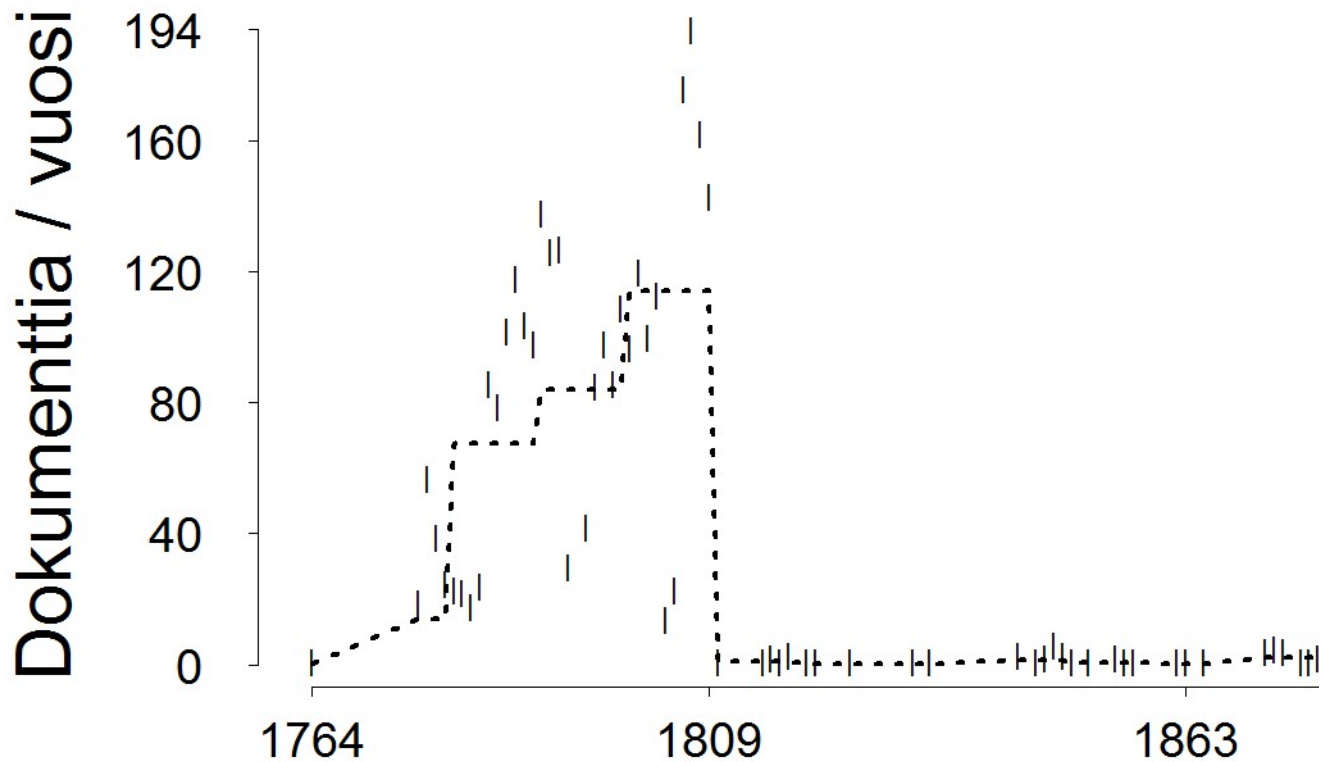


Kuva 1. Historia-aiheiset julkaisut Edinburghissa, 1540-1799. Kuvasta näkyy kuinka Edinburghissa julkaistuissa historia-aiheisissa dokumenteissa keskeisimmät yhteiskunnalliset muutokset vaikuttavat kirjastoluettelosta ilmeneviin julkaisumääriin vuosina 1540—1800. Näitä muutoksia, jotka näkyvät kaaviossa tummennettuina, ovat: Englannin sisällissota (1642—1651), Stuart-suvun paluu valtaistuimelle (1660), Mainio vallankumous (1688—1689), Ison-Britannian kuningaskunnan syntyminen eli unioni (1706—1707). Huomionarvoista on myös ettei Yhdysvaltojen itsenäisyys (1776) aiheuta samanlaista julkaisupiikkiä.

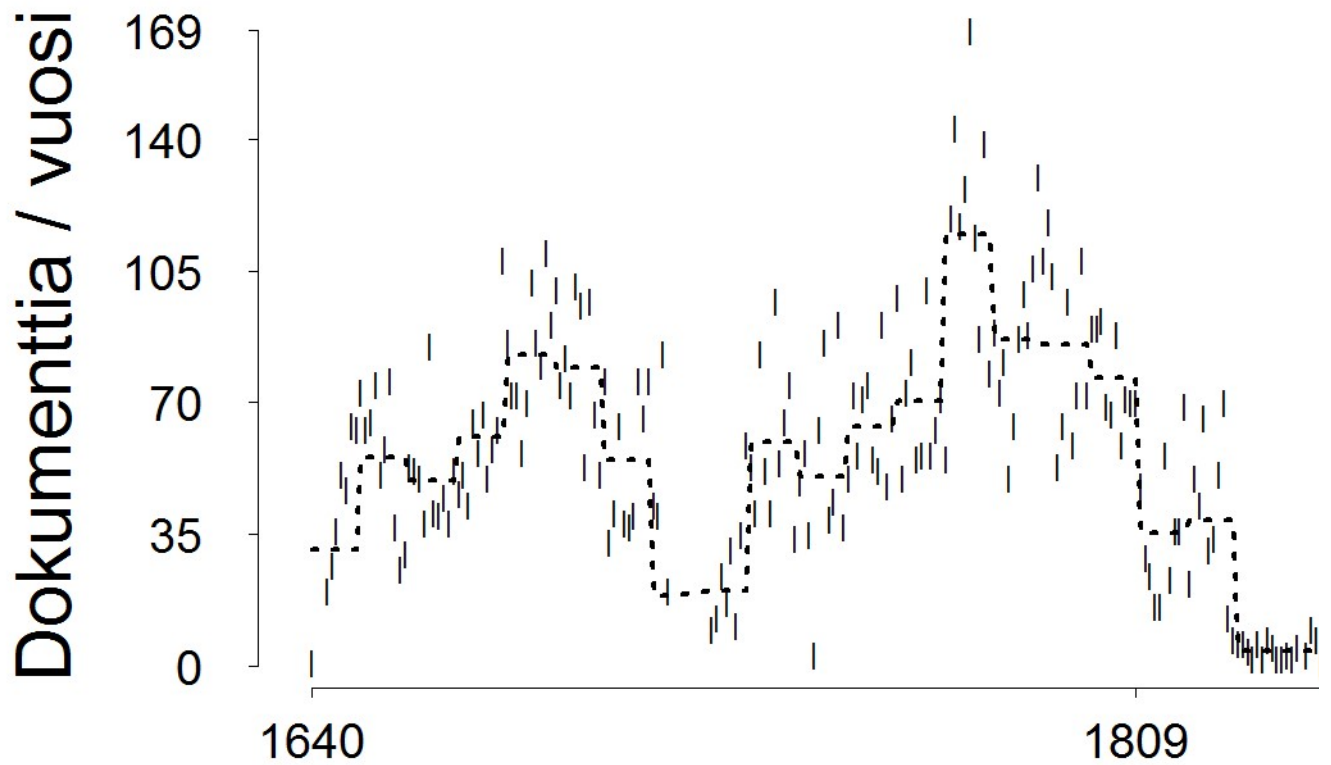
Kotimaisen historiantutkimuksen kannalta kiintoisan esimerkin tarjoaa tieteellisten laskentakirjastojen laadinta suomalaisen kansallisbibliografian Fennican tutkimiseen, vaikka samalla tuleekin tiedostaa — ja korjata — Fennican nykyiset puutteet. Aineiston alustava analyysi osoittaa esimerkiksi suoraan historiallisiin tapahtumiin yhteytyvät tiedontuotannon huiput ja aallonpohjat Vaasassa, Helsingissä, Turussa ja Suomessa vuosina 1764-1917 (Kuvat 2-5).

## Julkaisutoiminta Vaasassa



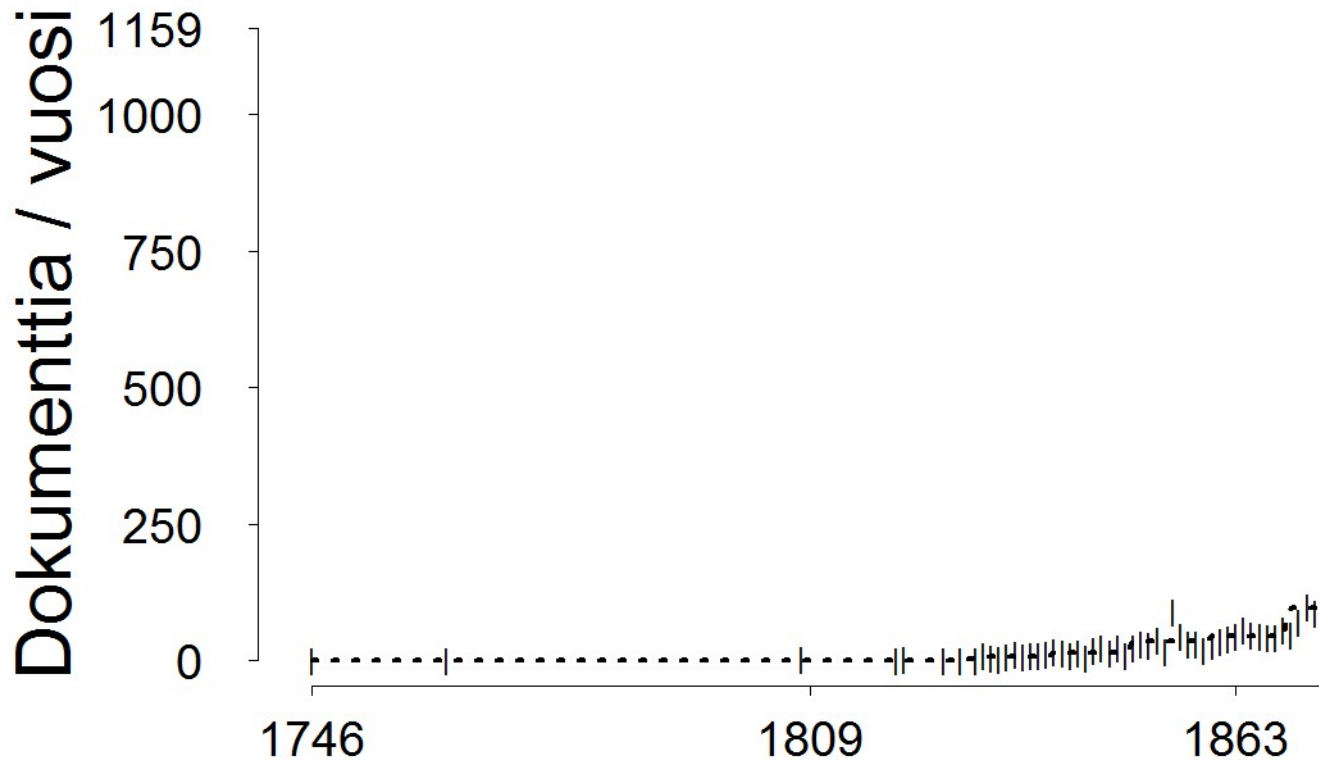
Kuva 2. Julkaisutoiminta Vaasassa 1764-1917. Oheinen kuva näyttää viittaavan tiedontuotannon romahdukseen Vaasassa välittömästi Venäjän vallan alussa<sup>[18]</sup>. Tämä saattaa osittain selittyä sillä, että painotoimintaa harjoitti V näyttää myös siltä, että hoviokuden asiakirjojen julkaisuoperaate ja painaminen muuttui Venäjän vallan myötä. Tämä kuvaa yksinkertaisella tavalla Suomen historiassa tapahtuneita julkaisuuteen ja julkiseen keskusteluun liittyviä kehitys painotuotannon muodossa näyttää olevan tilastollisesta näkökulmasta Venäjän vallan aikana hitaampaa kuin mitä yliopiston siirron ja Helsingin kehittämisen merkitystä korostanut historiantutkimus on aiemmin esittänyt.<sup>[19]</sup> systemaattisia keisarikunnan aikaisen painotuotannon liittyviä puutteita kirjastoluettelossa (on tunnettua, että Fennica ei ole täydellinen, vaan luetteloinnista puuttuu runsaasti erityisesti 1800-luvun ruotsinkielistä tuotantoa). On k tavan tunnistaa tiedontuotannon historiaan liittyviä trendejä. Lopulliset johtopäätökset tulee kuitenkin varmistaa myös muilla keinoin. Yksi syy miksi tahdomme tässä julkaisussa tuoda nämä vielä vajavaiset tilastolliset näkökulmat e kansallisbibliografian täydentymistä enää seuraaviksi vuosikymmeniksi (ks. esim. huomio Jyrki Hakapää, Kirjan tie lukijalle. Kirjakauppojen väkintuminen Suomessa 1740-1960, SKS, 2008, s. 380), vaan tämä tärkeä luettelointityö

# Julkaisutoiminta Turussa 1



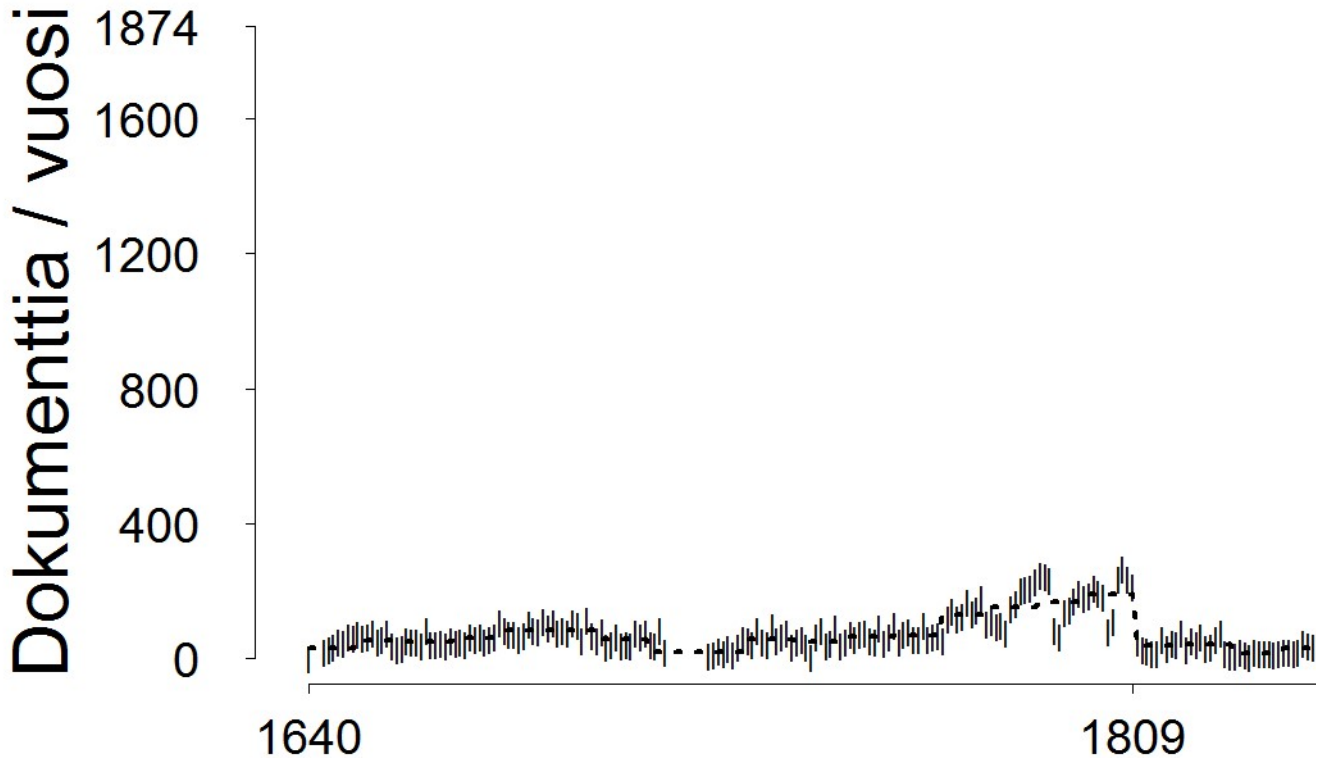
Kuva 3. Julkaisutoiminta Turussa 1640-1917. 1700-luvun kehitys näyttäisi seuraavan eurooppalaista kasvavaa trendiä, kirjapainotoiminnan kehitys 1820-luvulta lähtien on taas nykyisten luettelointitietojen valossa hidasta.

# Julkaisutoiminta Helsingissä



Kuva 4. Julkaisutoiminta Helsingissä 1746-1917. Visuaalinen esitys Helsingissä tapahtuneesta kirjapainotoiminnan kehityksestä viittaa siihen, että jos Fennican luettelointiin on ollenkaan luottamista, niin ennakosensuurin osittais painotoiminnan kehityslähtee käyntiin vasta huomattavasti myöhemmin. Yliopiston siirtäminen Helsinkiin näyttää olleen julkaisutoiminnan kannalta, jos ei lamauttava, niin ainakin hidastava toimenpide.

# Julkaisutoiminta Suomessa



Kuva 5. Julkaisutoiminta Suomessa 1640-1917 kertoo siitä, että 1700-luvun lopun ruotsalainen kehitys näyttää samalta kuin yleinen eurooppalainen painotoiminnan nousu. Vuoden 1809 jälkeen toiminta Suomen alt painotoiminnassa nousee samalle tasolle kuin Ruotsin vallan aikana. Tämä kuva korjaantuu ainakin osittain kansallisbiografian täydentyessä.

Kartoittamalla lisäksi vaikka ruotsalaisiin tai venäläisiin paikannimiin tehtävien viittausten määrää ja laatua suomalaisissa sanomalehdissä, voimme saada uutta tietoa suomalaisen julkisen keskustelun rakenteesta ja kehityksestä historian eri vaiheissa.<sup>[20]</sup> Samaa voidaan tehdä paikallisessa mittakaavassa tutkimalla viittauksia naapurikuntien ja suurempien kaupunkien välillä.<sup>[21]</sup> Kansallisbiografian historiallisten henkilöiden nimien paikantaminen ja kartoitus voi myös osaltaan kehittää suomalaisen julkisen sfäärin tutkimusta. Näin uudet informaatiotekniikan menetelmät voivat antaa aatehistorian tutkimuksen käyttöön uudenlaisia tutkimusmateriaalia ja tutkimuskeittäviä näkökulmia.

Alkuperäiset kuvailutietueet eivät sellaisenaan sovellu tilastolliseen analyysiin, vaan ne pitää ensin jalostaa sopivaan muotoon. Tämä sisältää mm. maantieteellisten ja henkilönimien erilaisten kirjoitusmuotojen yhtenäistämistä, samannimisten kirjoittajien erottelua, kirjoitusvirheiden käsitteilyä, tai kirjojen sivumäärän arviointia bibliografisen standardimerkintöjen pohjalta. Tietoja voidaan myös täydentää ja rikastaa muiden lähdeaineistojen nojalla. Kirjoittajista voidaan esimerkiksi selvittää puuttuvat elinvuodet, sukupuoli, muuta taustatietoa, painopaikkojen perusteella painomaa ja niin edelleen. Valtaosa työstä voidaan ja tulee automatisoida, kun tutkittavien aineistojen määrät kasvavat kymmeniin tai satoihin tuhansiin dokumentteihin. Samalla analyysin tilastollinen voima kasvaa ja satunnaisien virheiden vaikutus johtopäätöksiin häviää. Kaikkea ei voida kuitenkaan automatisoida. Juuri tästä syystä joustavaan lähdekoodiin perustuvat räätälöidyt laskentakirjastot tarjoavat erinomaisia mahdollisuuksia alan tutkimukselle verrattuna valmiisiin ohjelmistokokonaisuuksiin, jotka usein olettavat että aineisto on jo valmiiksi saatettu tilasto-ohjelmiston vaatimaan muotoon. Tämä on merkittävä puute, sillä käytännön kokemuksen perusteella aineistojen siirtäminen ja laadunvarmistus ohjelmistojen ja tilastollisten analyysien vaatimaan muotoon on yksi työläimmistä ja eniten aikaa vievistä työvaiheista. Tieteellisen ohjelmoinnin hyödyntäminen auttaa automatisoimaan kaiken mikä automatisoitavissa on. Tieteellinen laskentakirjasto tarjoaa nimensä mukaisesti kokoelman hyvin dokumentoituja menetelmiä tutkimuskohteen erityispiirteet huomioivien tietojenkäsitteilytehtäviin. Laskentakirjastot auttavat automatisoimaan merkittäviä osa aineistojen tilastollisista analyysistä, mutta valmisohjelmistoista poiketen antavat samalla tutkijalle merkittävästi laajemmat vapaudet yhdistellä ja täydentää menetelmiä halumallaan tavalla, tutkimuskohteen erityispiirteet huomioiden. Tämä tarjoaa tutkijalle sujuvan pääsyn sekä aineistojen jalostamiseen että tilastolliseen analyysiin ja visualisointien kannalta keskeisiin työkaluihin. Laskentakirjastoja ylläpidetään usein avoimen lähdekoodin palveluissa, joihin koko työvirta voidaan toteuttaa läpinäkyvällä ja toistettavalla tavalla, tieteellisen avoimuuden periaatteita kunnioittaen (ks. kuva 6).

## Tiedonlouhinnan vaihtoehtoja aatehistorian näkökulmasta

### Laajojen aineistojen tekstilouhinta

- **Tavoite:** ymmärtää käsitteellistä muutosta ja kielen käyttöä
- **Aineistot:** tekstitietokannat (ECCO, EEBO, digitoidut sanomalehdet jne.)
- **Potentiaali:** kvantitatiivinen evidenssi laadullisen tutkimuksen tukena ja uudet havainnot; etäluentaa
- **Käytäntö:** alkuperäisaineistot harvoin avoimesti saatavilla tai sidottu rajoitettuihin käyttöliittymiin
- **Skaalautuvuus:** rajattu; data on kooltaan massiivista ja voi vaatia huomattavaa jalostamista; aineistosta ja menetelmistä riippuvainen;
- **Metodologinen näkökulma:** aatehistorian sovelluksia toistaiseksi hyvin vähän

### Kuvailutiedot tilastollisena työkaluna

- **Tavoite:** ymmärtää yleisiä trendejä historiallisten ilmiöiden dokumentoinnissa
- **Aineistot:** kirjastot ja arkistot tulvivat erilaisia kuvailutietokantoja (metadata)
- **Potentiaali:** tutkimuskäyttö aliarvioitu
- **Käytäntö:** saatavuudessa ongelmia, mutta vähemmän kuin tekstiaineistojen kohdalla
- **Skaalautuvuus:** erinomainen; data on rakenteista ja sen koko on rajattu
- **Metodologinen näkökulma:** tarjoaa selkeän lähtökohdan uusien käytäntöjen soveltamisessa humanistiseen tutkimukseen.





