European Union
European Regional Development Fund

Leverage from the EU 2014–2020

Fennica ja muut kirjastoluettelot avoimen tieteen lähteinä

Kirjastoverkkopäivät, 2015

Mikko Tolonen, Helsingin yliopisto

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Esityksen runko

- Tutkimusidea kirjastoluetteloille
- Kuvailutiedot, avoin tiede ja big datan ongelmat
- ”Historia” ja brittiläinen luettelointi 1470-1800
- Mitä voidaan sanoa yleisistä suomalaisista julkaisutrendeistä 1640-1917 Fennican pohjalta?

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

TUTKIMUSIDEA

Kirjastoluettelot: perinteinen tietokanta, uusi käyttö!

?
Ymmärretään tiedontuotantoa & kulttuurien vuorovaikutusta

!
Kvantitatiivinen lähestyminen laadullisiin kysymyksiin

AVOIN TIEDE

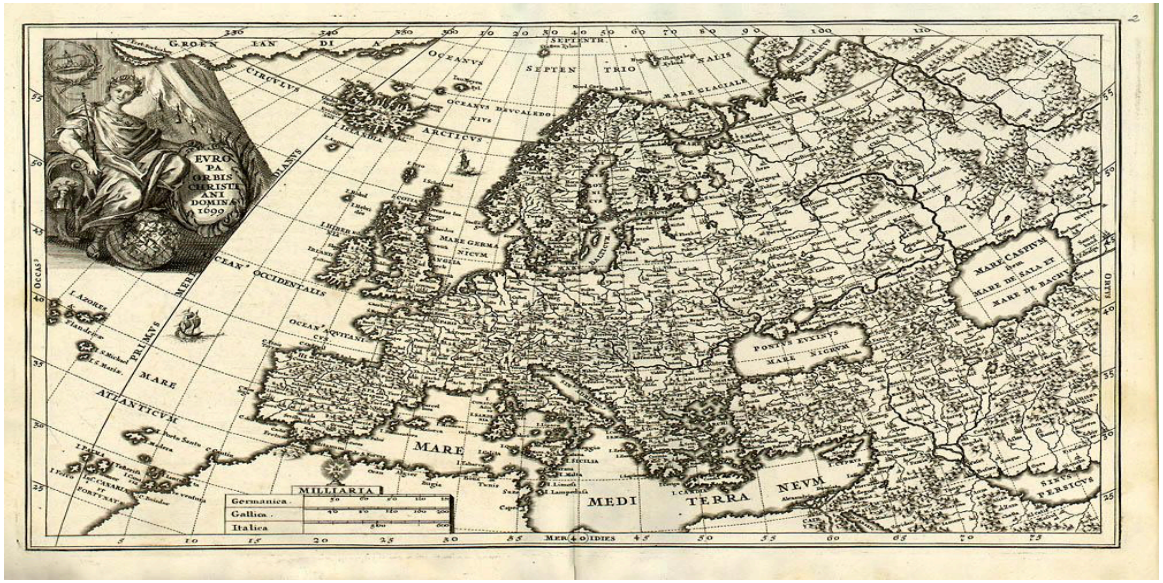
Läpinäkyvät metodit (R, Python, ..)

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

https://github.com/rOpenGov/estc/

rOpenGov



Mahdollisuudet: kokonaiskuva koko varhaismodernin ajan kirjatuotannosta + historian resonointi tiedontuotannossa

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

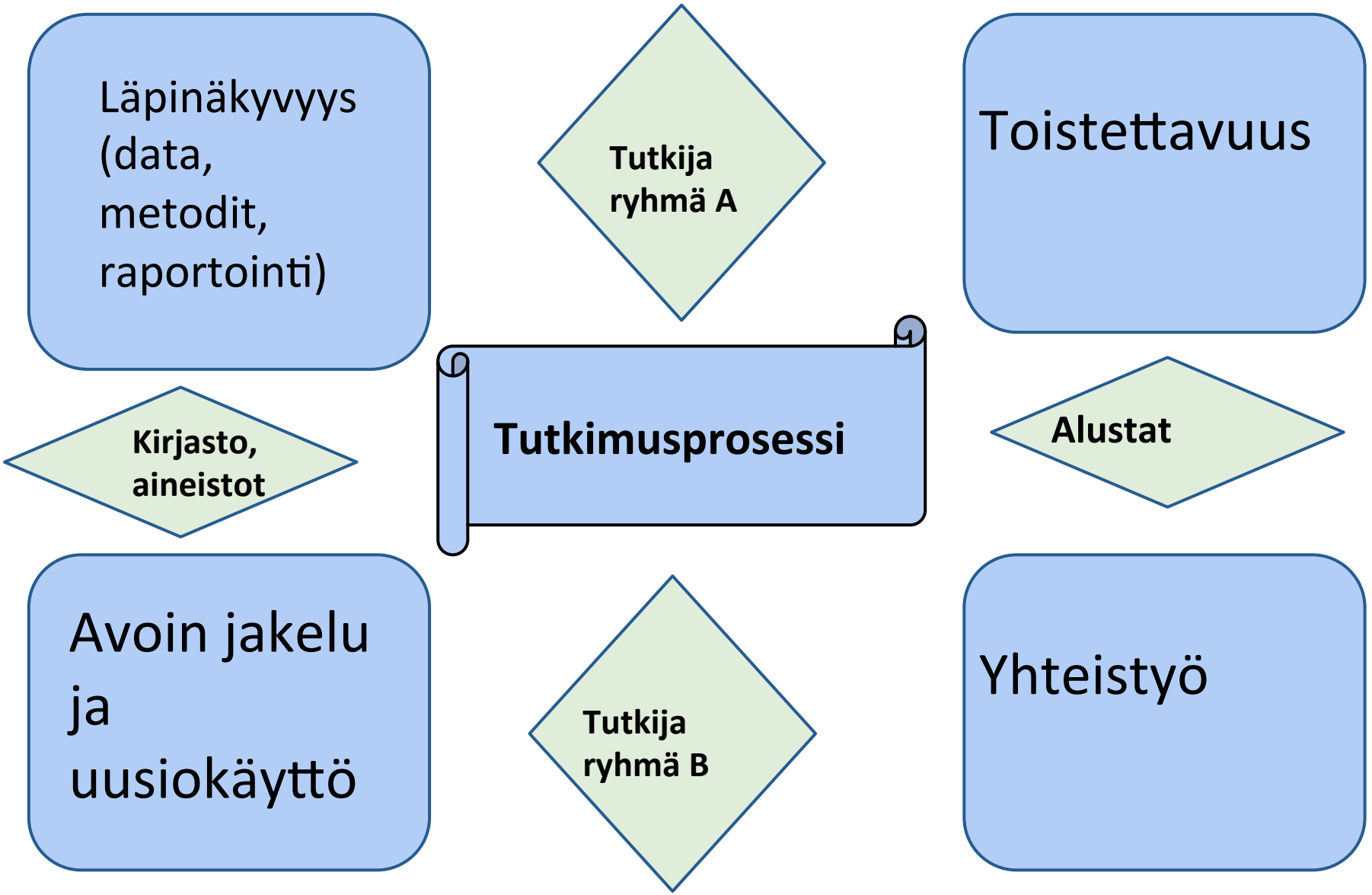
KUVAILUTIEDOT, AVOIN TIEDE JA BIG DATAN ONGELMAT

Kuvailutietojen monikäyttöisyys

- Kvantitatiivinen kehys laadulliselle tutkimukselle
- Tiedontuotannon ymmärtäminen
- Julkaisijat ja niiden verkostot (visualisointi)
- Julkaisupaikat, ”cultural transfer”
- ”Historian”, ”filosofian”, ”uskonnon” analysointi (ei genreinä per se, mutta alakategorioina)
- Yksittäiset kirjoittajat ja näiden vertailu

Avoimen tieteen periaatteet

- Ei ainoastaan avoin julkaisu – koko tutkimusprosessi on keskeinen ja sen tulisi olla mahdollisimman avoin.
- Metodit, tutkimus, data ja tulokset kaikki avoimiksi.
- Läpinäkyvyys, toistettavuus, informaali yhteistyö, uudet aloitteet
- Pääsy “raakadataan” on institutionaalinen kysymys



Ajatus avoimen tieteen ekosysteemistä humanistisessa tutkimuksessa

Big data lähestyminen ja sen ongelmat

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Viesti

- Big data ja muilla uusilla “Digital Humanities” – lähestymistavoilla voidaan tuottaa lisäarvoa, mutta pääpaino oltava laadussa ja avoimuudessa ketjun kaikissa osissa:
- Luettelointityö (kirjastoluetteloiden tapauksessa)
- Datan avaaminen
- Datan siivoaminen tutkimuksessa
- Tutkimusdatan avoimuus
- Tutkimuksen uusiokäyttö / sen mahdollistaminen

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Kuvailutietojen nykytilasta

- **Kysymys mitä ei enää haluta luetteloida**
- *Dimensiot (esim. sivunumerot), Fennican puutteet – MIKSI?*
- **Yleinen valitus Marcin sotkuisuudesta**
- *Meille tutkijoille ei ongelma*
- **Versionhallinnan mahdollisuudet**
- *Github yms. versionhallinta harvoin käytössä, mahdollisuus kaksisuuntaiseen vuorovaikutukseen*
- **Kokonaisten kokoelmien kuvailun globaalit periaatteet**
- *Kansallisbibliografiat, miksi ei yhtenäisiä standardeja mitä näihin kuuluu ja mitä ei?*

Datan laadun todellisuus (Fennica)

- "","publisher","n"
- "1","G. W. Londicer",1829
- "2",NA,1463
- "3","J. C. Frenckell",1316
- "4","Frenckellianis",1185
- "5","C. A. Londicer",484
- "6","Johan Winter",375
- "7","Johan Larsson Wall",276
- "8","Peter Hansson",262
- "9","Joh. Christoph. Frenckell",248
- "10","debat Petrus Wald. Acad. typogr",213

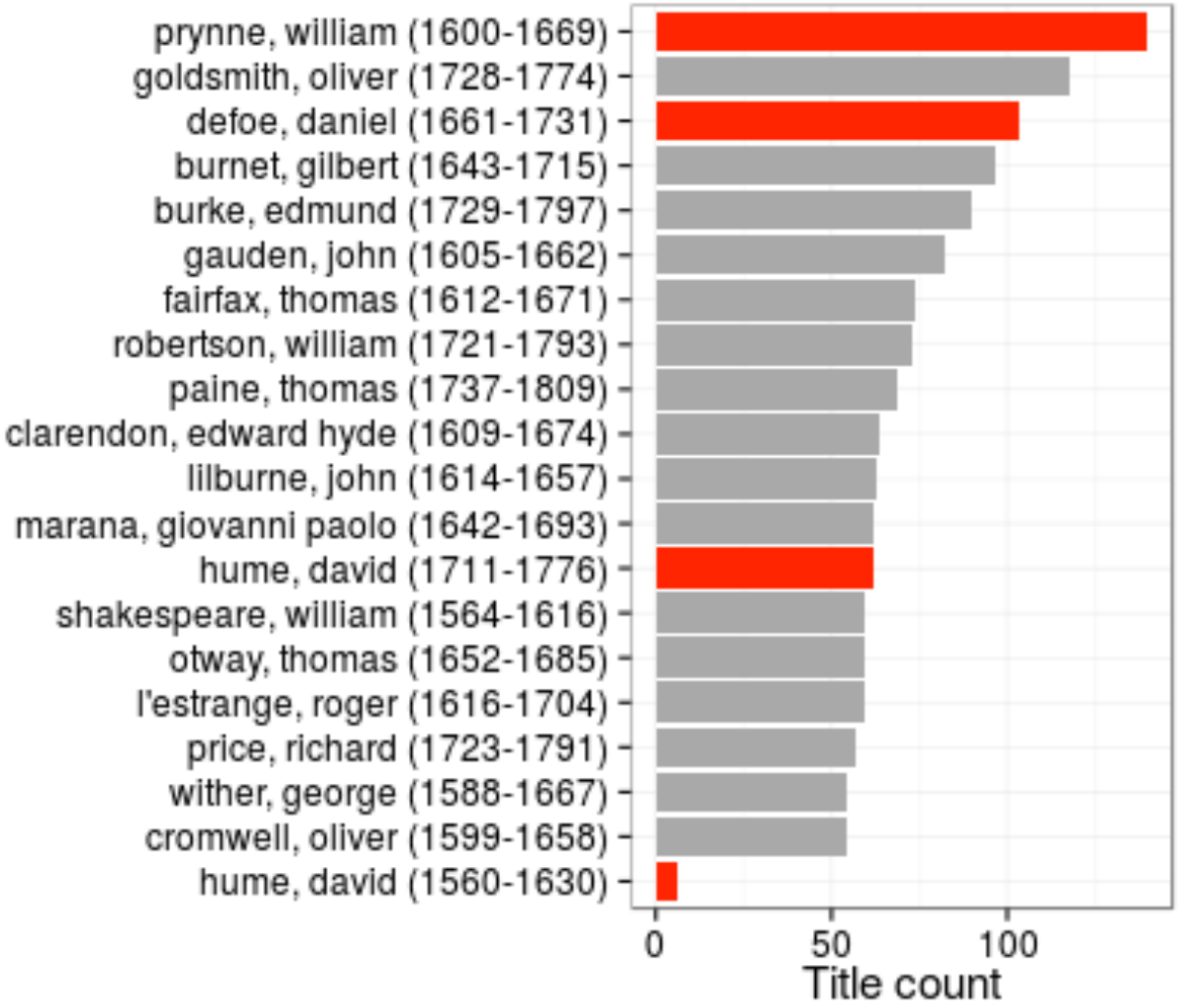
Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

PILOTTITUTKIMUS: ”HISTORIA” TILASTOLLISESTI BRITTILÄISEN ENGLISH SHORT-TITLE CATALOGUE, ESTC KATALOGIN PERUSTEELLA

Kuka kirjoitti historiaa Britanniassa?

Yleisimmät auktorit (otsikoiden perusteella)

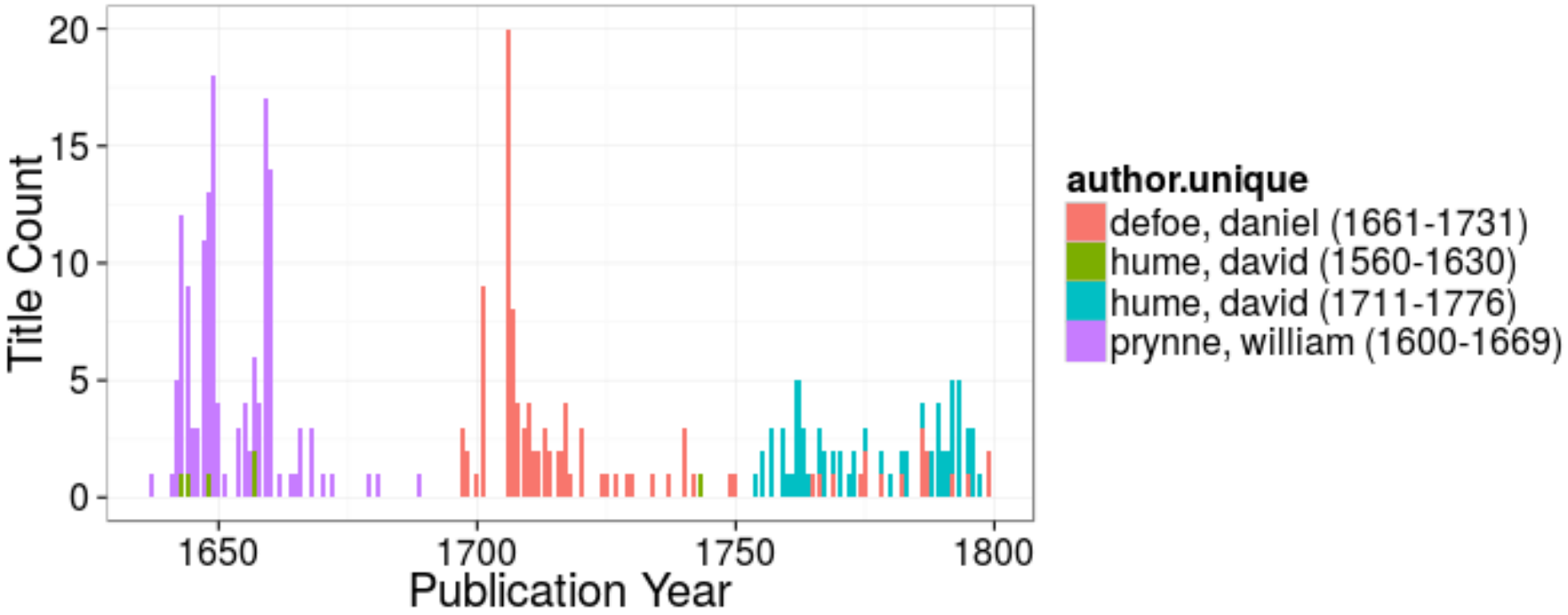


Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Kuka kirjoitti historiaa?

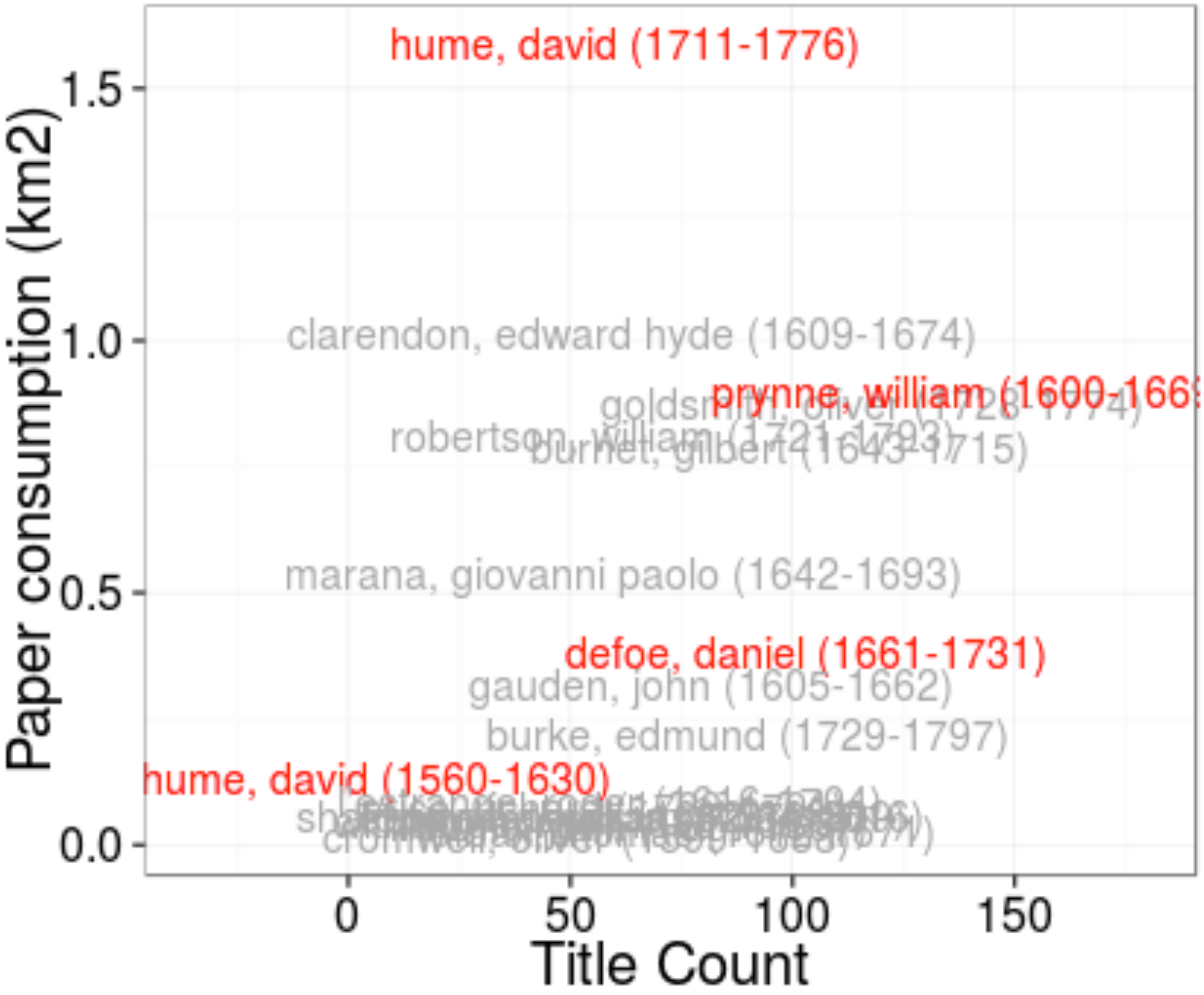
Julkaisujen määrät aikajanalla: William Prynne, Daniel Defoe, David Hume (1558–1629) sekä David Hume (1711-1776)



Kuka kirjoitti historiaa?

Julkaisujen määrän ja paperin kulutuksen vertaus

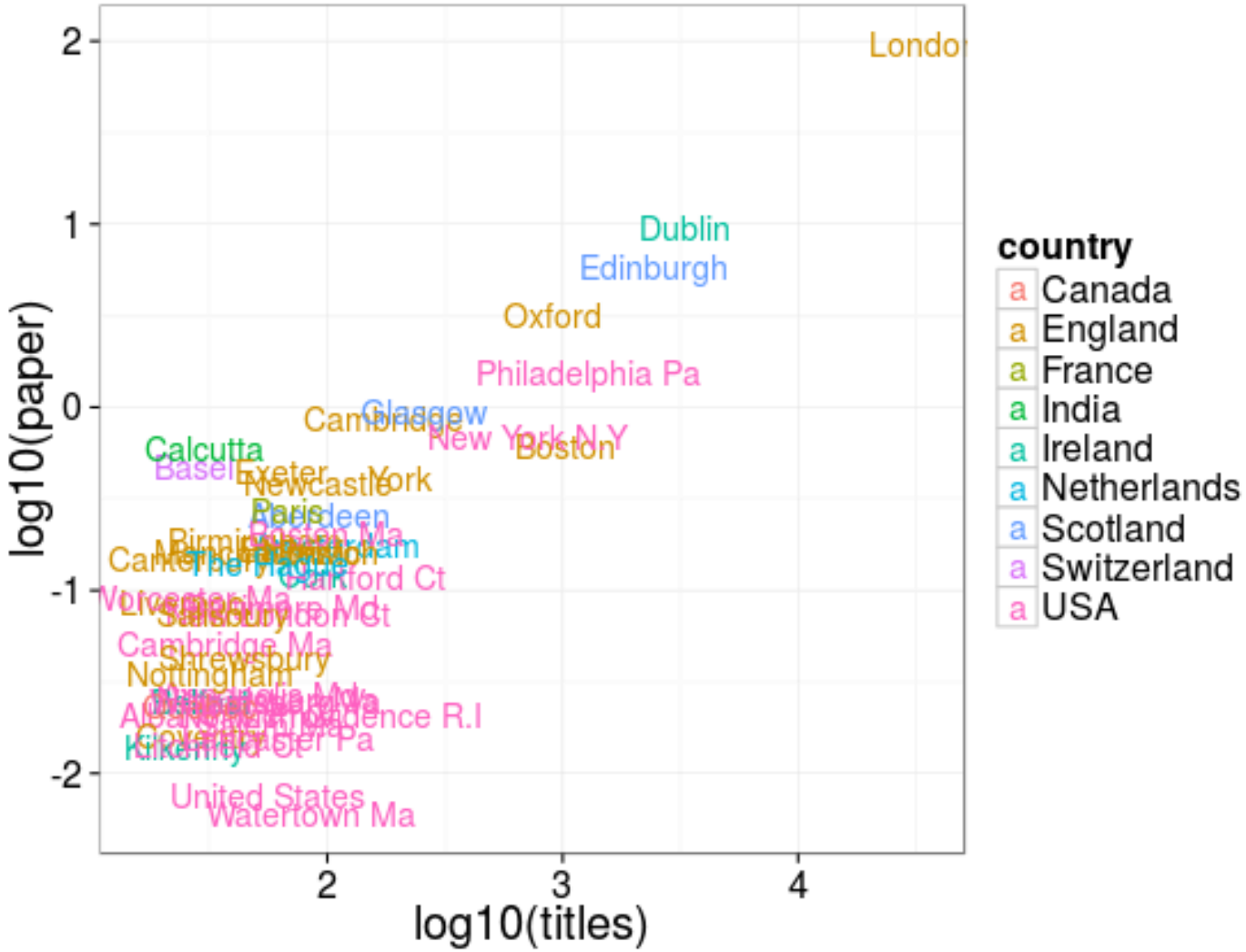
(yleisimmät auktorit)




Kuka kirjoitti historiaa?
Missä historiaa julkaistiin?
Miten julkaisemin en muuttui?

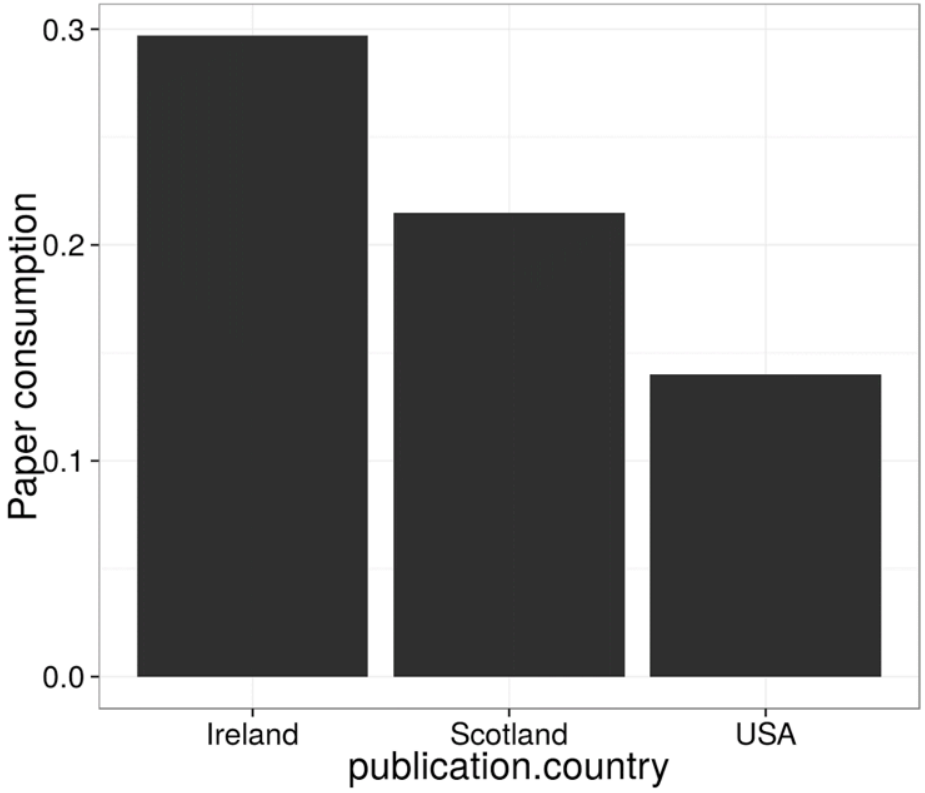
Missä historiaa julkaistiin?

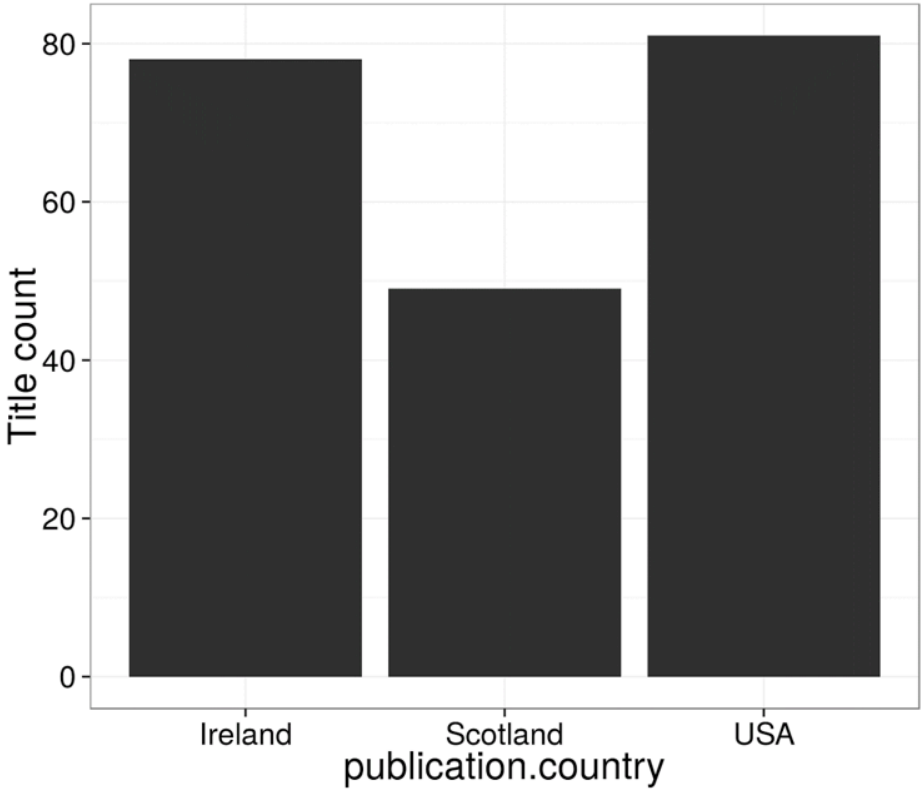
Julkaisumäärien ja paperinkulutuksen vertailua



Missä historiaa julkaistiin?

Paperinkulutus ja julkaisumäärät Irlannissa, Skotlannissa ja USAssa



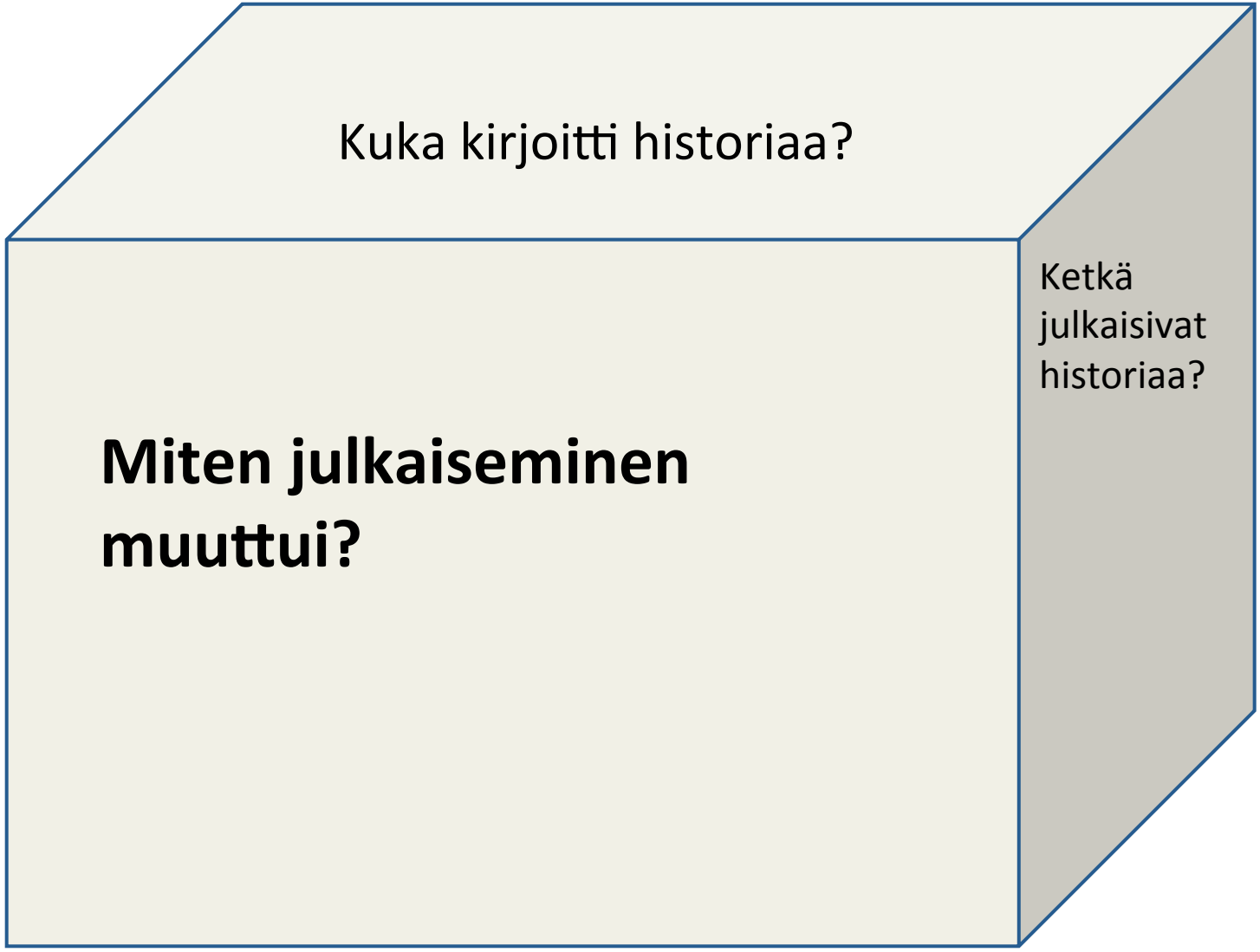


Paperinkulutus

Julkaisumäärät

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

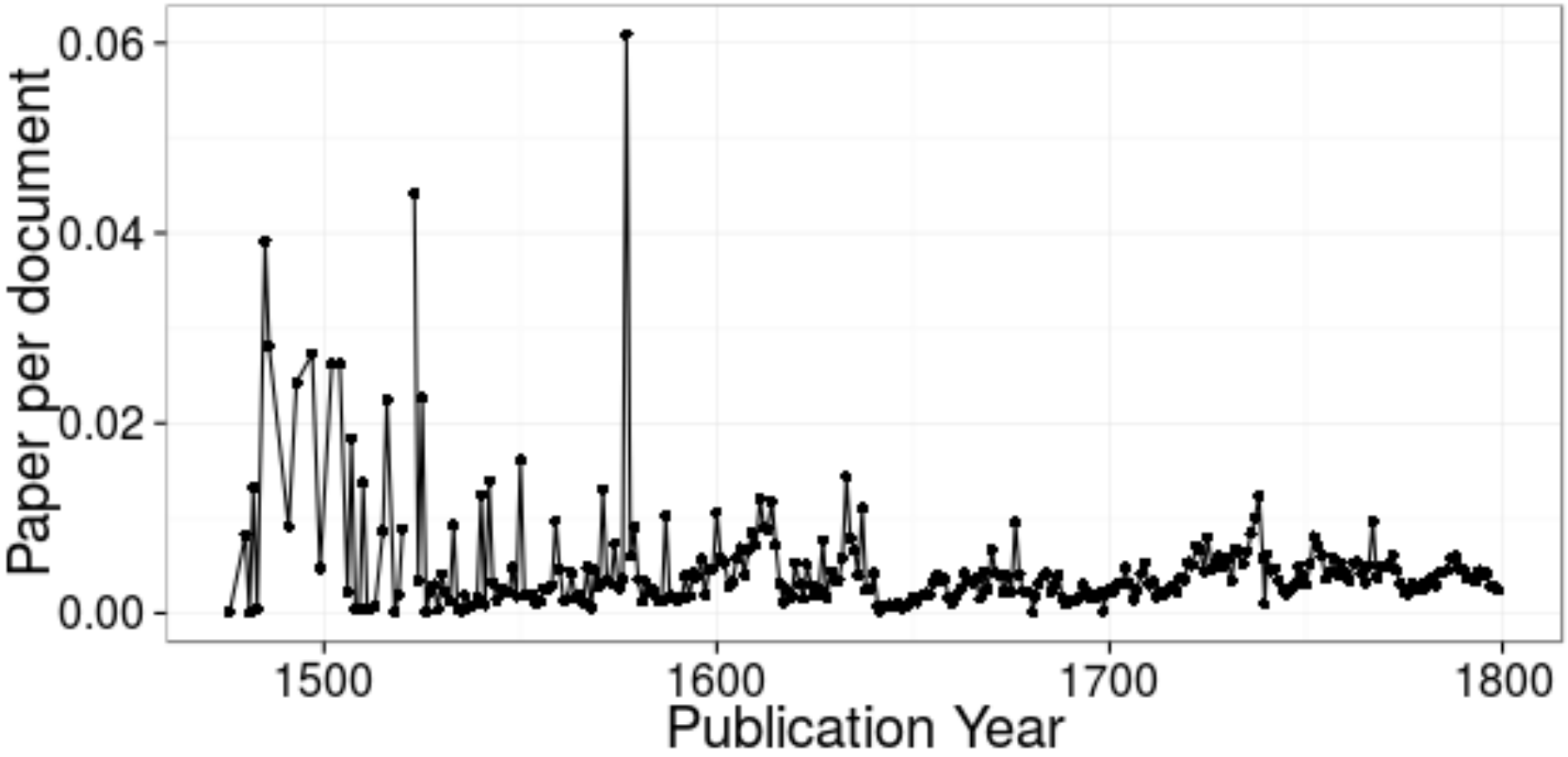
UNIVERSITY OF HELSINKI



UNIVERSITY OF HELSINKI

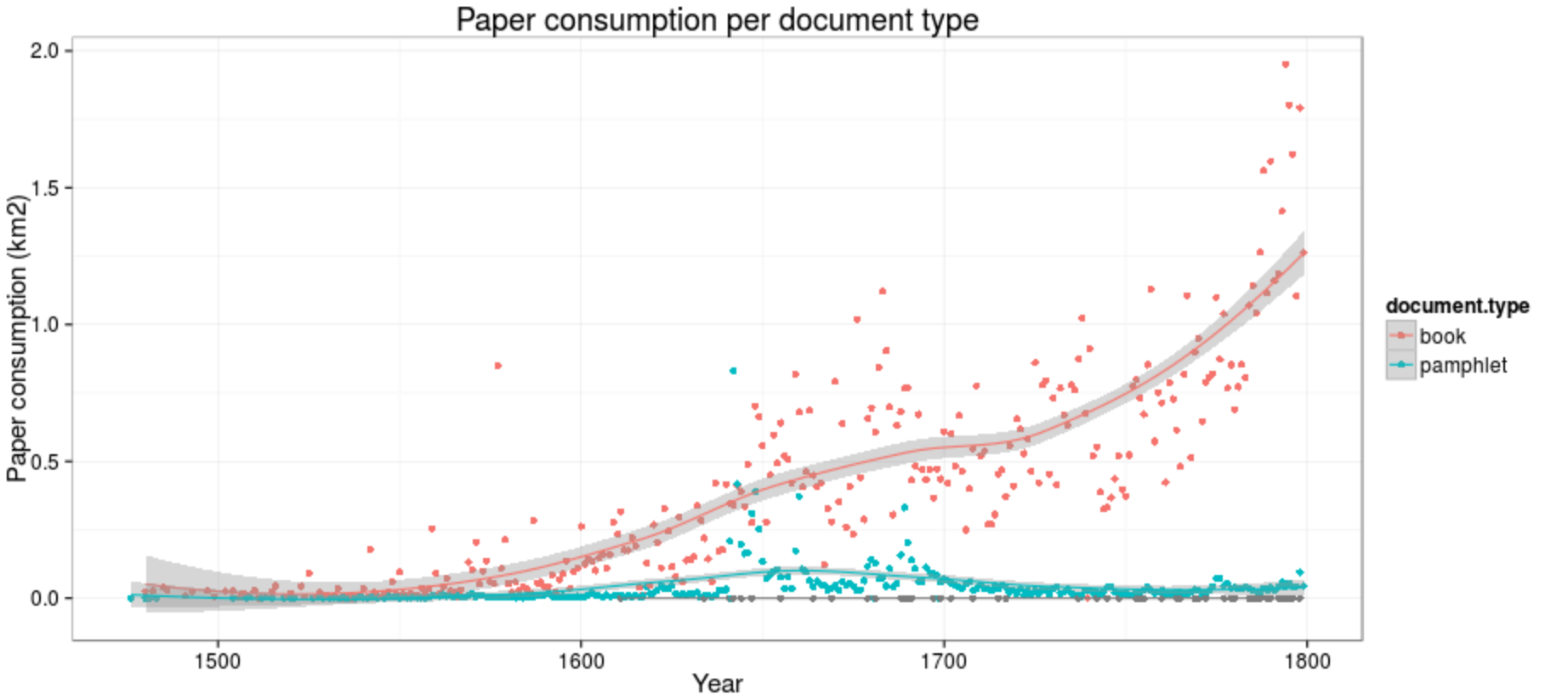
Miten julkaiseminen muuttui?

Keskimääräinen paperinkulutus per dokumentti



Miten julkaiseminen muuttui?

Pamfletit verrattuna kirjoihin

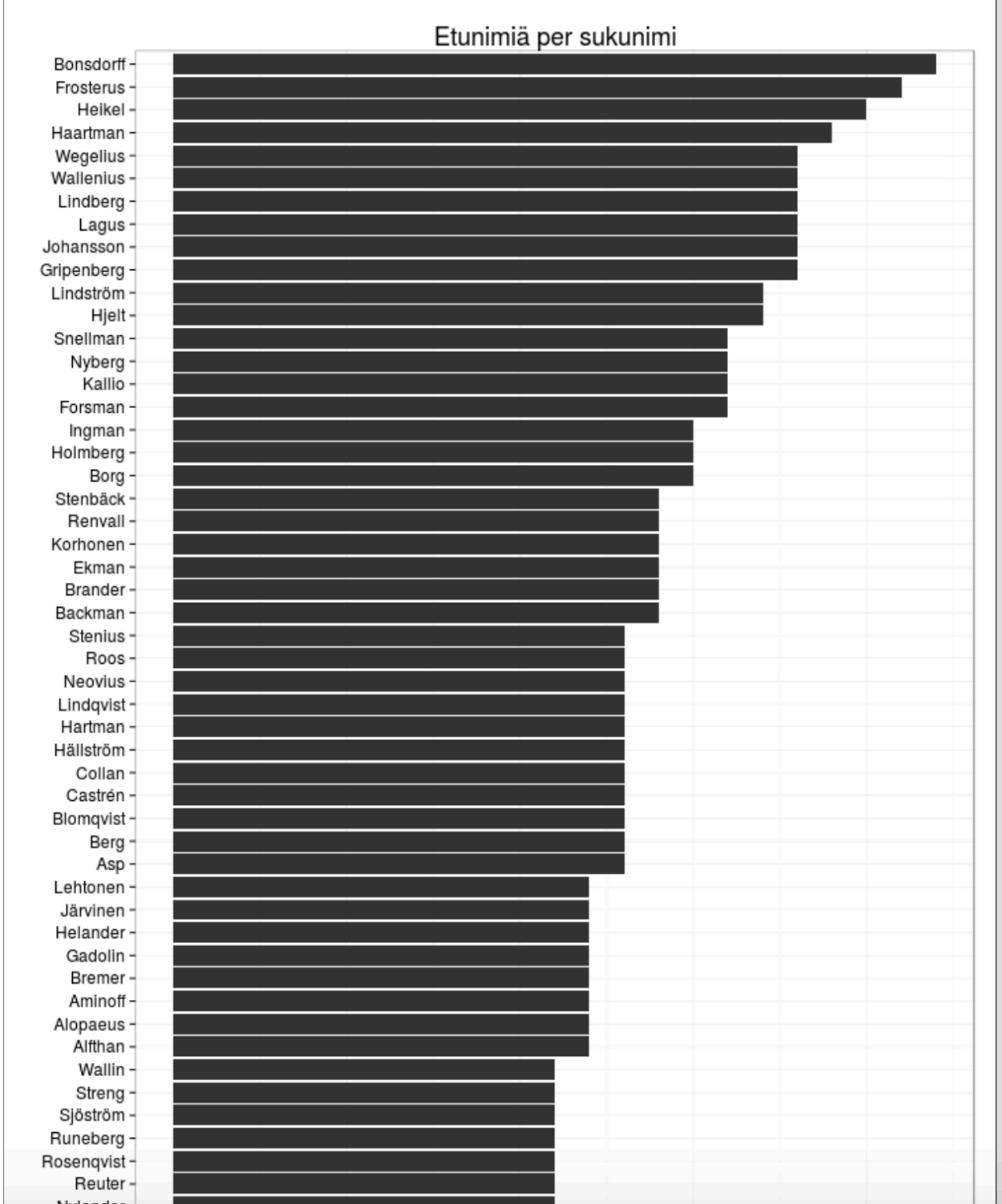


Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Esimerkki 2

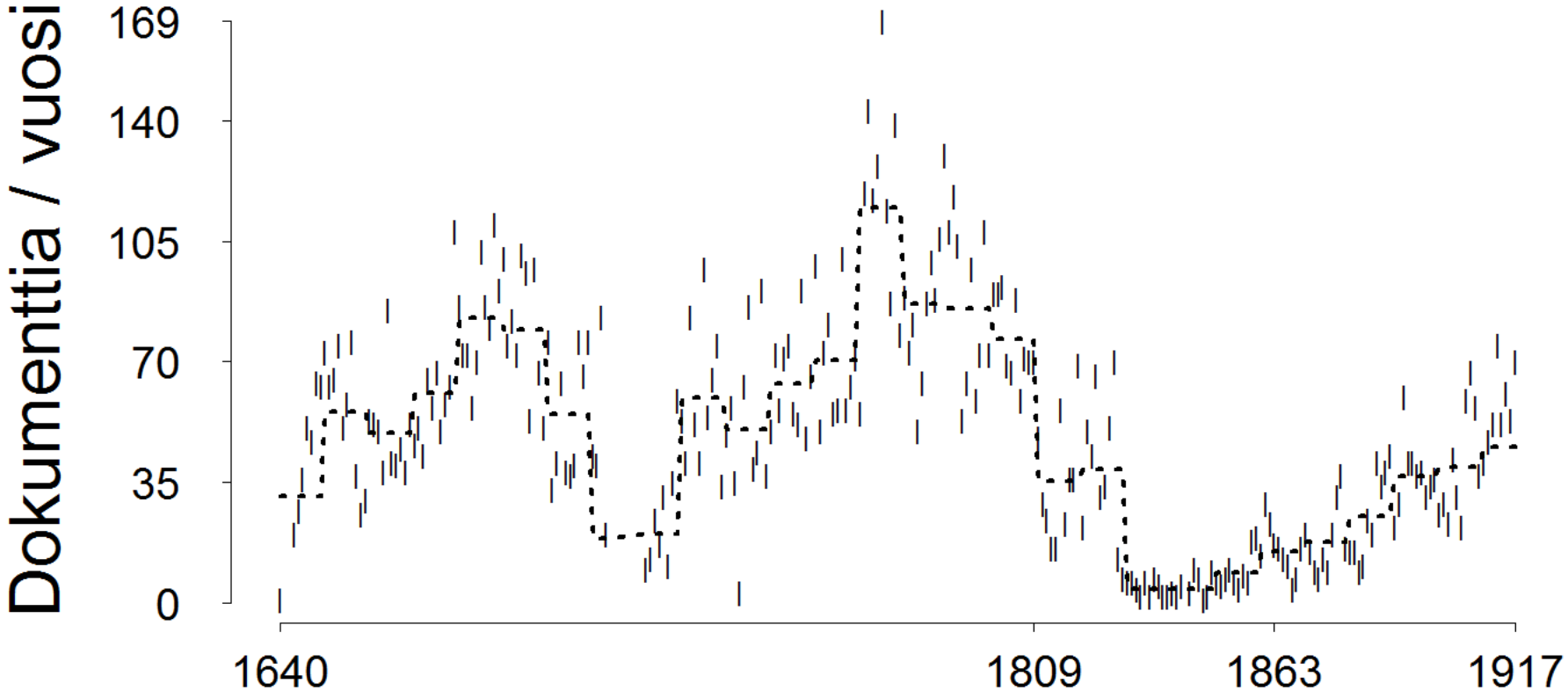
TIEDONTUOTANTO SUOMESSA 1640-1917 FENNICAN PERUSTEELLA


Etunimiä per sukunimi
Bonsdorff
Frosterus
Heikel
Haartman
Wegelius
Wallenius
Lindberg
Lagus
Johansson
Gripenberg
Lindström
Hjelt
Snellman
Nyberg
Kallio
Forsman
Ingman
Holmberg
Borg
Stenbäck
Renvall
Korhonen
Ekman
Brander
Backman
Stenius
Roos
Neovius
Lindqvist
Hartman
Hällström
Collan
Castrén
Blomqvist
Berg
Asp
Lehtonen
Järvinen
Helander
Gadolin
Bremer
Aminoff
Alopaeus
Alfthan
Wallin
Streng
Sjöström
Runeberg
Rosenqvist
Reuter

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

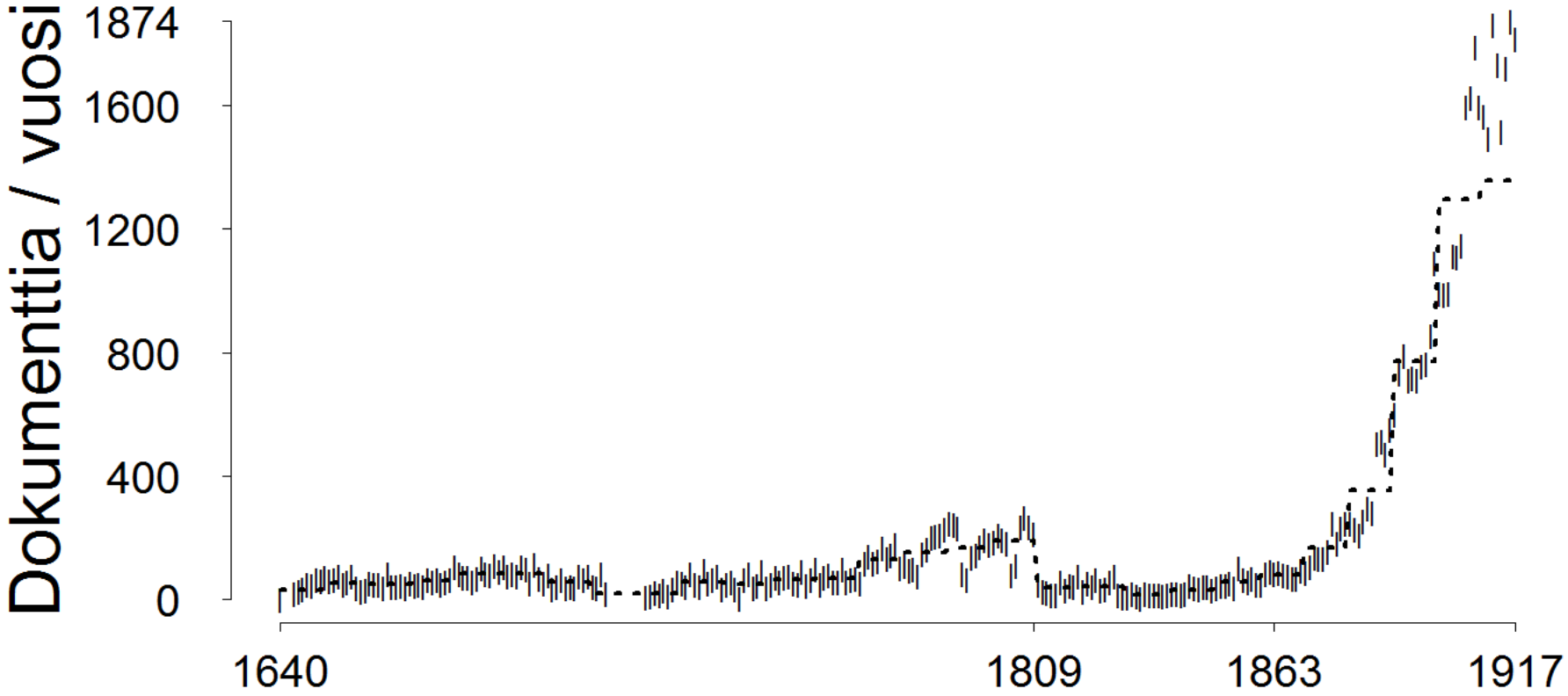
Julkaisutoiminta Turussa 1640-1917



Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Julkaisutoiminta Suomessa 1640-1917



Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

Mitä voidaan sanoa suomalaisesta tiedontuotannosta Fennican esikäsittelyn perusteella?

- Ilman parempaa tietoa, näyttäisi olevan silmiinpistävää romahdus 1809 jälkeen.
- Helsingin rooli 1800-luvulla sivistyksessä ehkä erilainen kuin kuva jonka saa lukemalla esim. Klingeä.
- Mutta, keskeisin selittävä tekijä puutteet luetteloinnissa: Kansalliskirjastossa tiedetään että ruotsinkielisen materiaalin osalta puutteet suuria.
- Silti, 1820-luvun loppupuolella tai edes 1850-luvulla luulisi julkaisutoiminnan nousevan. Mitään viitteitä siitä ei ole.
- Joka tapauksessa, Fennicaan ei voi näiltä osin luottaa ennen kuin aineistoa täydennetään = ei voi vetää näitä johtopäätöksiä.

Avoimen tieteen kiertokulku Fennican tapauksessa

- Perusluettelointi kuntoon – Kansalliskirjasto!
- Kuvaustietojen siivoaminen, rikastuttaminen, yhdistely ja analysointi – Fennica tutkimusryhmä + muut
- Tutkimusdatan avaaminen
- Tutkimusdatan uusiokäyttö – muut tutkijat, sekä aineistonhaltijat – Kansalliskirjasto!

UNIVERSITY OF HELSINKI

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

LOPUKSI

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Digitaaliset aineistot Suomessa

- Aineistoja on jo hyvin esim. aatehistorian tutkimuksen kannalta
- Fennica (1483-1917)
- Erityinen painopiste: digitoidut (kaikki) suomalaiset sanomalehdet 1770-1910
- Calonius-Naumannin kokoelma
- Turun väitöskirjat ym.

Digitalia
Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus

UNIVERSITY OF HELSINKI

Laajojen historiallisten aineistojen tutkijakäyttö

- Innovatiivinen tutkimus on usein ruohonjuuritasolta nousevaa.
- Kehitystä ei tapahdu jos meillä on erikseen työkalujen tuottajat ja tutkijat.
- Aito yhteistyö aineistojen haltijoiden (kirjastojen) ja tutkijoiden kanssa välttämätöntä
- Avoimuus sekä raaka- että tutkimusdatan kanssa on tie eteenpäin.

YLIMÄÄRÄISIÄ DIOJA

Entä sitten kun aineistoon itseensä ei voi täysin luottaa?

- Koskaan ei voi tietää voiko aineistoon luottaa tai sillä tehdä tulkintoja laajoista trendeistä ennen kuin aletaan sen soveltaminen (siksi ”julkaise ajoissa ja usein” käyttäen versionhallintaa). Tämä on erilainen tapa ajatella suhteessa perinteiseen näkemykseen historiallisesta lähdekritiikistä. #AVOINTIEDE
- Aineistoja tulee täydentää, tarvitaan riittävä massa aineistoa että voidaan tehdä big data –tutkimusta
- Pitää muistaa, että aineiston ei täydy oltava täydellistä että se olisi käyttökelpoista. Tästä huolimatta aineiston putsaus ennen analyysia vie n. 70% ajasta.