



Digitaaliset menetelmät Suomen kirjallisuushistorian uudistajina

Suomen Akatemia (2022-2026)

DHL-FI
VERKOSTOTAPAAMINEN,
Turku 4.10.23

WP2: Datatiede

Leo Lahti, Julia Matveeva



Ohjelma

12.00–12.15: Leo Lahti: Tervetuloa ja esittäytymiskierros

12.15– 13.45: DHL-FI-konsortion osahankkeiden esittelyt (15 min. + 15 min. keskustelua/hanke):

- WP 2: Datatiede, TY, Leo Lahti
- WP 1: Kirjallisuushistoria, UEF, Kati Launis & Veli-Matti Pynttäri
- WP 3: Digitaaliset resurssit, Kansalliskirjasto, Osma Suominen

13.45–14.15 Kahvit

14.15– 15.45: Hanke-esittelyjä (15 min. + 15 min. keskustelua/hanke):

- Ilona Pikkanen: Constellations of Correspondence (CoCo) (<https://coco.rahtiapp.fi/>)
- Kati Kallio: Formulaic intertextuality, thematic networks and poetic variation across regional cultures of Finnish oral poetry (FILTER) (<https://blogs.helsinki.fi/filter-project/>)
- Mikko Tolonen: High Performance Computing for the Detection and Analysis of Historical Discourses (HPC-HD) (<https://hpc-hd.github.io/>)

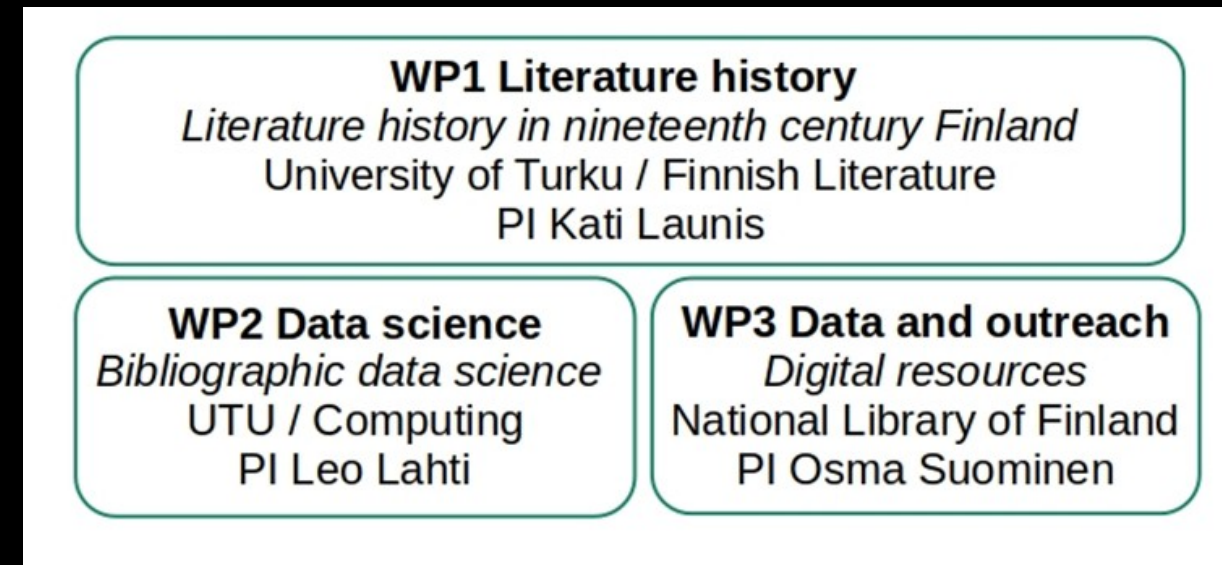
15.45 – : Kunkin paikalla olijan ”digikuulumisia” – eli vapaamuotoinen kierros siitä, millaisia hankkeita, hankeideoita, projekteja jne. kullakin on meneillään tai voisi tulevaisuudessa olla meneillään

Loppukeskustelu

18.00 Illallinen Tårget-ravintolassa

Digital History for Literature in Finland - konsortio (DHL-FI)

- WP 1: Kirjallisuustieteet, UEF: Kati Launis, Aino Mäkikalli, Viola Parente-Čapková & Veli-Matti Pynttäre
- WP 2: Datatiede, TY: Leo Lahti (konsortion vastuullinen johtaja), Julia Matveeva, Pyy Kantanen
- WP 3: Kansalliskirjasto: Osma Suominen, Pasi Koste, Katerina Sornova, Juha Rautiainen, Juho Inkinen & Nicholas Volk
- Hankekumppanit: Mikko Tolonen (HY), Ilona Pikkanen (SKS) ja Veronika Laippala (TY)
- Ohjausryhmä: Åsa Arping, Katherine Bode, Elsi Hyttinen, Kati Kallio, Minna Kaukonen, Päivi Lappalainen, Hannu Salmi



Tavoite:

1) Etsiä/louhia bibliografinen metatieto ”kaikesta” Suomessa 1800-luvulla (1809-1917) julkaistusta kaunokirjallisuudesta Kansalliskirjaston Fennica-julkaisutietokannasta

-> tuottaa uutta tietoa Suomen kirjallisuushistoriasta digitaalisten menetelmien ja aineistojen avulla

2) Kehittää avoimia digitaalisia menetelmiä ja aineistoja tutkimusyhteisön ja suuren yleisön käyttöön



Suomen kansallisbibliografia Fennica

- Tietovaranto, jonka avulla pääsemme ”pinnanaliseen” tietoon
- Yli miljoona kirja- ja monografiatietuetta (painettua ja sähköistä), n. 73 000 jatkuvaa julkaisua (lehteä tai sarjaa), n. 52 000 karttaa; karttuu jatkuvasti
- Tietoa kirjoista 1488 ->, lehdistä 1771 ->, sarjajulkaisuista, kartoista, audiovisuaalisesta ja elektronisesta aineistosta. Painotuotteet tutkimuksen fokuksena.
- Turun yliopisto ja Helsinki Computational History Group (Leo Lahti & Mikko Tolonen) ovat aloittaneet Fennican kuvailutietojen rikastamisen osana aiempaa tutkimusta ja kansallisen FIN-CLARIAH-tiekarttainfrastruktuurin kehitystyötä (SA)

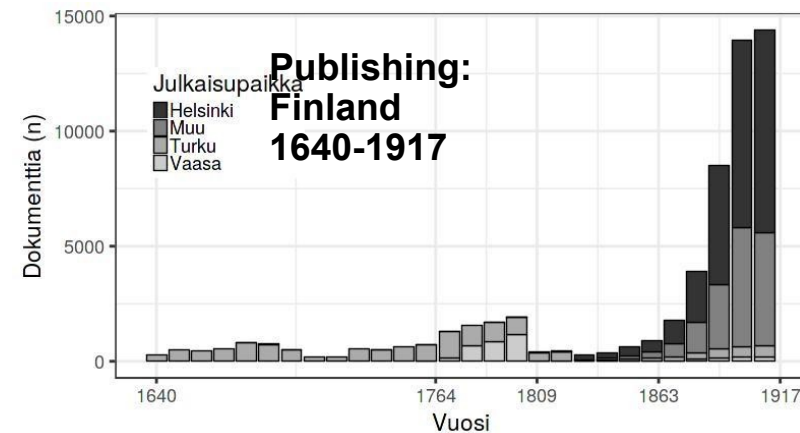
HISTORICAL METHODS

<https://doi.org/10.1080/01615440.2018.1526657>



A Quantitative Approach to Book-Printing in Sweden and Finland, 1640–1828

Mikko Tolonen^a , Leo Lahti^b , Hege Roivainen^a , and Jani Marjanen^{a,*} 



Merkintöjen vertailukelpoisuus?

Frenckell-julkaisijat Fennicassa

[J. C. Frenckell]

typis Frenckellianis

Frenckell

[J. C. Frenckells änka]

tryckt hos Johan Christopher Frenckell

impressit Joh. Christoph. Frenckell

[Frenckell]

typis Frenckelliorum

in officina Frenckelliana

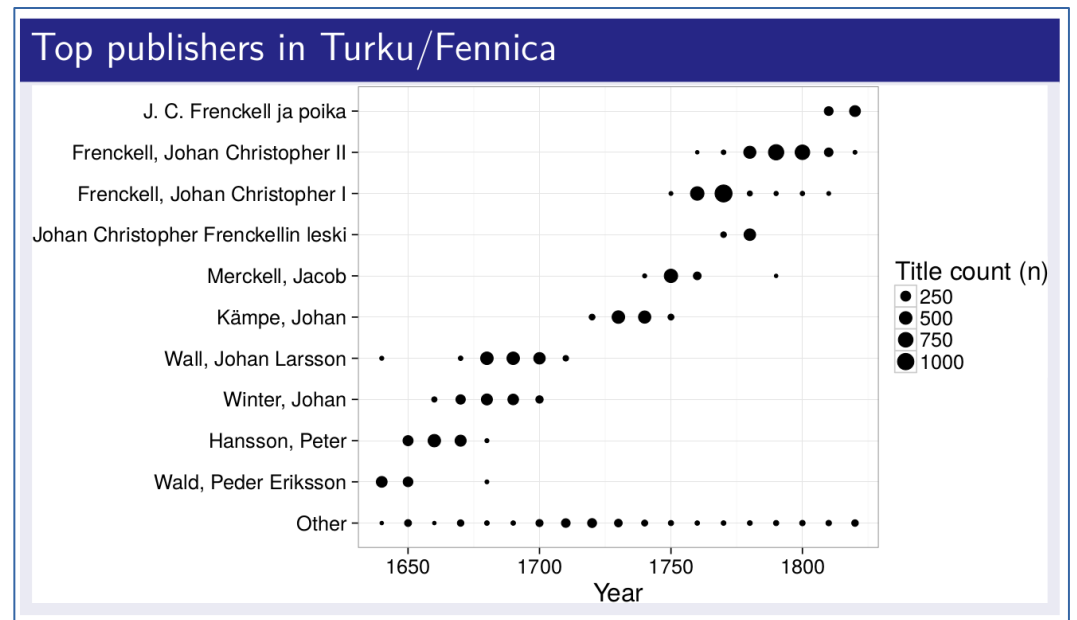
tryckt i Frenckellska boktryckeriet

wastuudest prändäTTY Frenckellin kirja-prändisä

.. 484 kirjoitusmuotoa ..

Julkaisijat

- 13 345 julkaisijaa
- <10% harmonisoituja nimiä?



Paikkatiedon visualisointi

R-CMD-check **passing** repo status **Active** codecov **63%** chat **on gitter** Watchers **29** Stars **175** Follow @ropengov **562**
CRAN **1.0.0** downloads **163** < downloads **163/month**

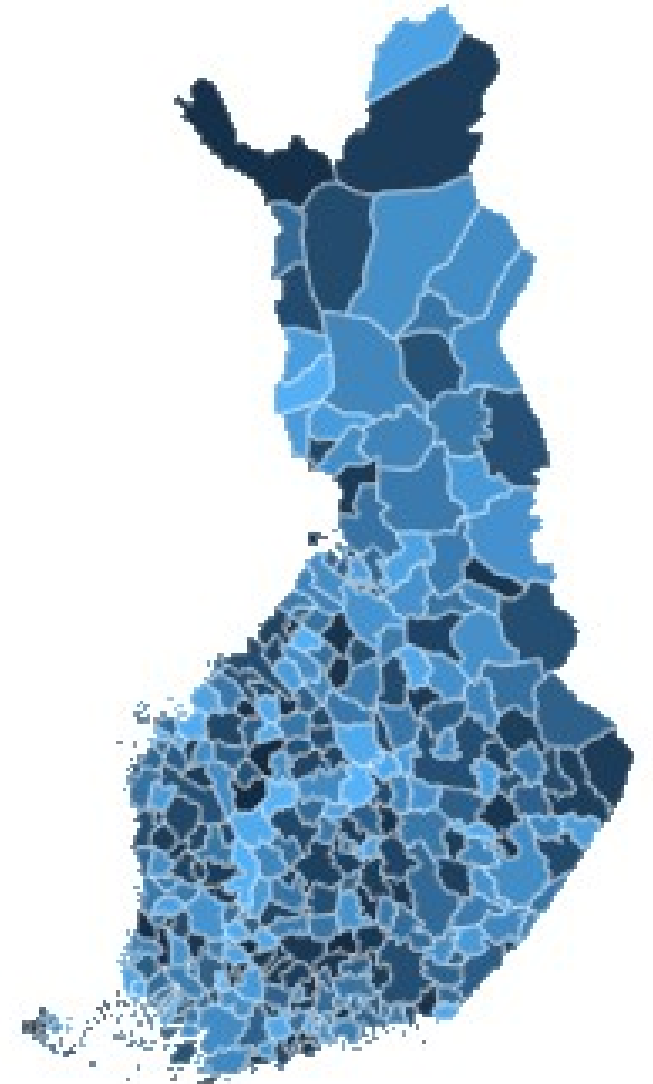
geofi - Access Finnish Geospatial Data

Access Finnish Geospatial Data.

Installation and use

```
# install from CRAN
install.packages("geofi")

# Install development version from GitHub
remotes::install_github("ropengov/geofi")
```



Fennica: avoin data & menetelmät

<http://data.nationallibrary.fi/>

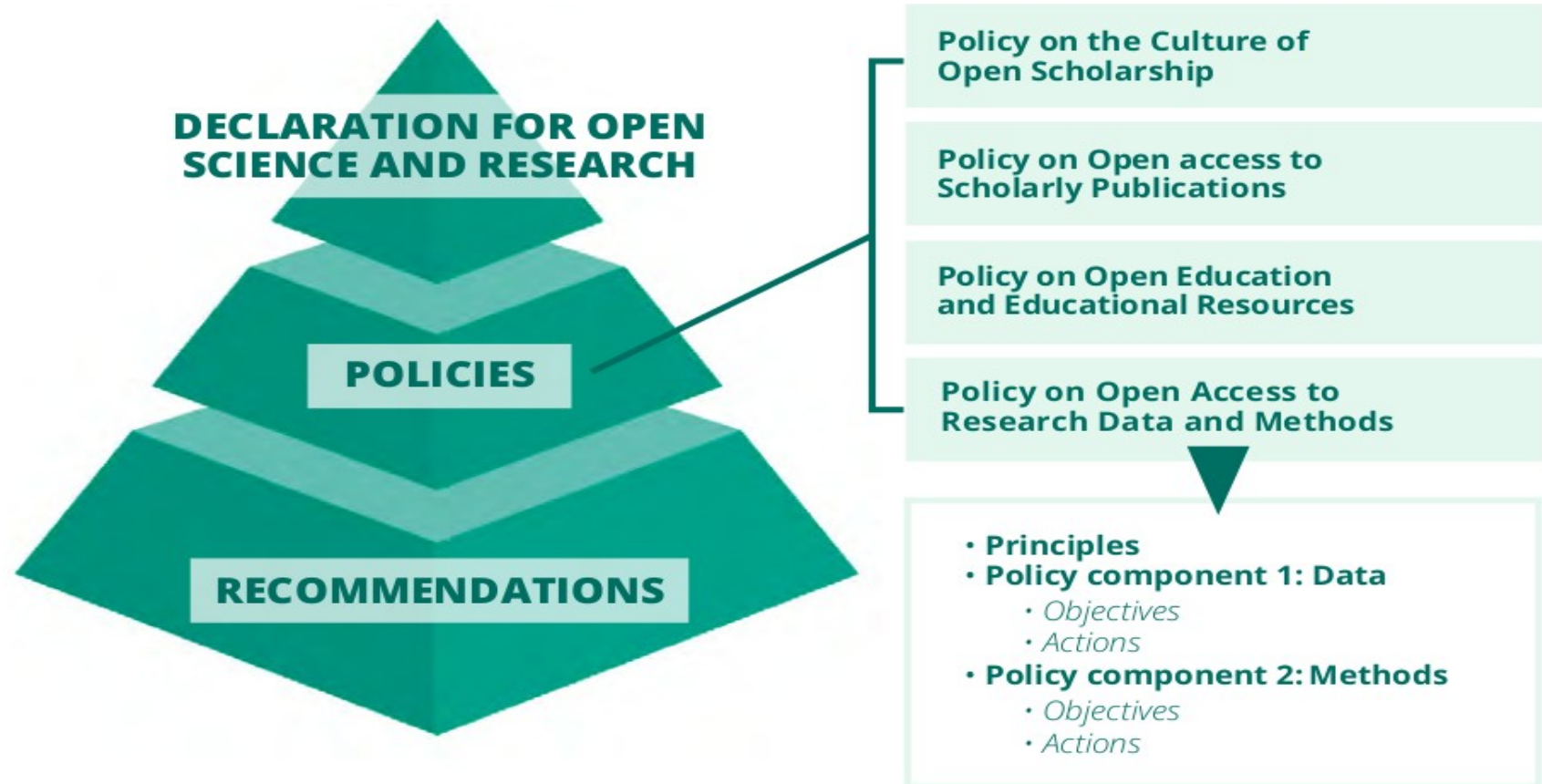
Kehitetään avoimia digitaalisia menetelmiä ja tutkimusaineistoja tutkimusyhteisön käyttöön.


- Aineisto rikastetaan, analysoidaan ja raportoidaan toistettavasti.
- Kehitetään erityisesti bibliografisten kuvailutietojen louhintaan ja analysointiin soveltuvia välineitä, jotka mahdollistavat kirjallisuushistorian datalähtöisen tutkimuksen.
- **Myöhemmin** kokonaisuutta täydennetään **lehtiaineistojen sisältämällä kaunokirjallisuudella**. Haasteena fiktion erottelu muusta kirjallisuudesta.
- Tutkimuskäyttöä varten jalostetut kuvailutiedot jaetaan avoimesti,
- Kytkeytyy työhön kansainvälisillä aineistoilla (mm. British Library ESTC; Mikko Tolonen & Leo Lahti).

National policy work for open science and research

Strategic principles in national policy work

- Responsible openness
- Access to sustainable infra & services
- Researcher merit & careers





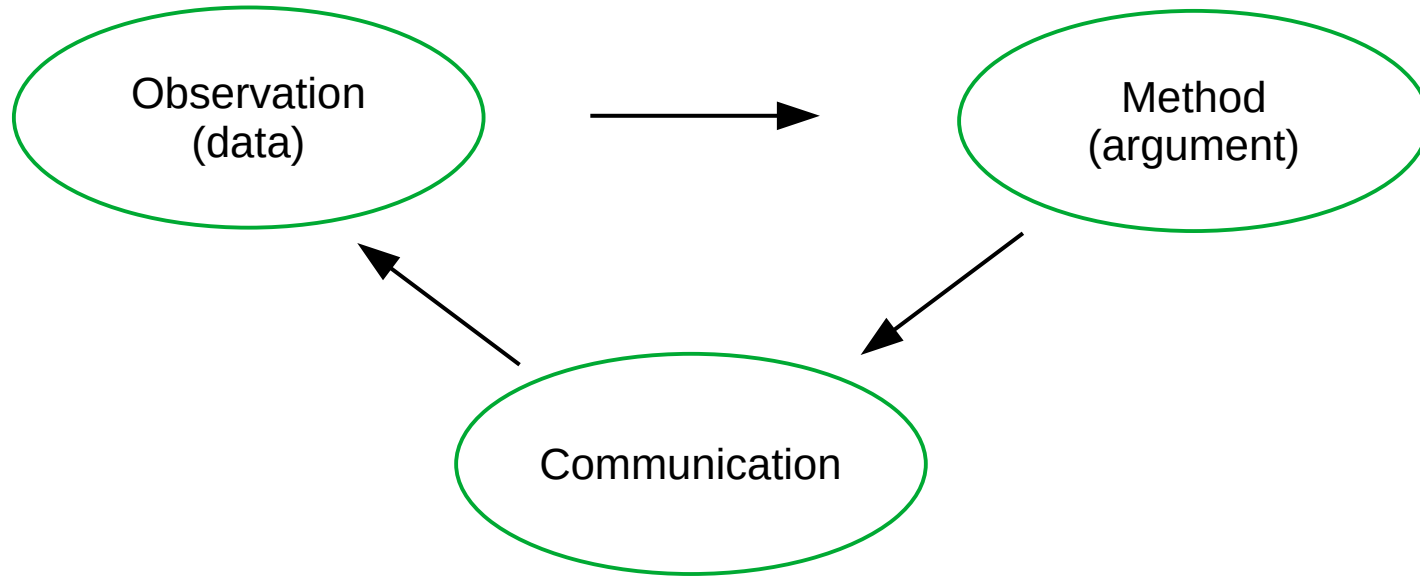
RESPONSIBLE RESEARCH SERIES 4:2023

OPEN RESEARCH DATA AND METHODS. NATIONAL POLICY AND EXECUTIVE PLAN BY THE HIGHER EDUCATION AND RESEARCH COMMUNITY FOR 2021–2025.
Policy components 1 (Open access to research data) and 2 (Open access to research methods and infrastructures)

Responsible research series 4:2023
5th volume
AUTHOR: Open Science Coordination in Finland, Federation of Finnish Learned Societies
PUBLISHER: The Committee for Public Information (TJNK) and Federation of Finnish Learned Societies (TSV)
Helsinki, 2023
ISSN: 2670-062X
ISBN: 978-952-5995-87-9
DOI: <https://doi.org/10.23847/tsv.669>
Accepted on 14 April 2021 and 30 January 2023 by the National Open Science and Research Steering Group.

- Drawn up by the Finnish research community / work groups 2019-2023 (30+ members from various organizations)
- Builds on national and international work and dialogue; public commenting round
- Coordinated by National Coordination for Open Science and Research (Federation of Finnish Learned Societies)

Research process & life cycle



*“as open as possible,
as closed as necessary”*

Vuosi 1

Toteutettu

- Kokonaisuus siirretty CSC:n palvelimille
- Koodien rakennetta parannettu; tukee jatkokehitystä
- Laajennus aiemmasta aikarajauksesta (n. 1500-1900) koko Fennicaan; >1M tietuetta
- Automaattisesti generoitava sivusto laadittu tukemaan datan jalostusta (Rahti)

Tekeillä

- Selvitetään aiemmin käsin laadittujen kirjallisuuslistojen tuottamista automaattisin rajauksin
- Selvitetään ongelmallisten tietojen tarkistamista Kansalliskirjaston kanssa
- Muokataan datan jalostussivua julkaisuvalmiiseen muotoon (disclaimerit sivun tarkoituksesta ym)

Jatko

- Jalostustyön ensimmäinen versio valmiiksi
- Aineiston jatkojalostus ja rikastus Kansalliskirjaston kanssa
- Tutkimuskäytön vahvistaminen