

Digitaaliset menetelmät kotimaisen kirjallisuushistorian uudistajana (DHL-FI)

WP 3: Digitaaliset resurssit

Osma Suominen

4.10.2023

Konsortion jäsenet

1. Turun yliopisto / Tietotekniikan laitos
Konsortion PI: apulaisprofessori **Leo Lahti**
vastuualueet: Fennican analyysityökalut, datan käsittely ja julkaisu
2. Turun yliopisto / Historian, kulttuurin ja taiteiden tutkimuksen laitos
PI: dosentti **Kati Launis**
vastuualueet: kirjallisuudentutkimus, mm. temaattinen analyysi, lähiluku, teosluettelon laadinta
3. Kansalliskirjasto
PI: TkT, tietojärjestelmäasiantuntija **Osma Suominen**
vastuualueet: aineistojen tarjoaminen, Annif-työkalu, Fennican rikastaminen, datan julkaisu

Keskeiset aineistot

Data set	Size	Use in the project
Fennica national bibliography	ca. 60,000 records (out of 1M)	Main bibliographic analysis
Melinda union catalogue	ca. 60,000 records (out of 16M)	Complements Fennica with more detailed information from other bibliographic databases
Kanto national agent name database	ca. 200,000 records	Information on authors (including pseudonyms), other persons, groups and organizations referenced in the bibliographic metadata
Digitized fiction books from NLF	ca. 1,000	Meso-analysis: combining close reading with quantitative analysis, utilizing the Annif service of NLF

Mitä kaikkea
bibliografinen data
voisi meille kertoa?



Ilmari Grytan teos Lähde (2007)

Mikä oikeastaan on **kaunokirjallisuutta**?

Tarvitaan **operatiivinen määritelmä**, tapa erottaa kaunokirjallisuus muusta.

Apuna mm.

- Kansalliskokoelman hyllypaikka (signum)
- Aineistotyytit: vain monografiat ja sarjat; esim. kartat voi jättää pois
- UDK- ja YKL-luokitus
- sisällönkuvailu (YSO, KAUNO...) tai sen puute



Kuva: [Suppilo](#), Pielisen museo. CC BY 4.0

Teosluettelo 1800-luvun kaunokirjallisuudesta

Teosluetteloita on julkaistu ennenkin, mutta ne perustuvat tutkijoiden näkemyksiin, ei kirjastotietokantoihin.

Tunnistetaan bibliografisesta datasta kaunokirjallisuuden **ensipainokset** ja muokataan niiden pohjalta kattava **teosluettelo**.

Luettelo voidaan hyödyntää myös teoskuvailun kehittämiseen.



Kuva: Marko Oja

Kirjailijoiden tarkastelu

Kanto/Asteri-auktoriteettitietojen perusteella voidaan tutkia esim.

- pseudonyymien / oman nimen käyttöä
- naiskirjailijoiden osuutta
- kirjailijoiden ikää ja syntymäpaikkoja

KÄYTETTÄVÄ NIMENMUOTO	Onerva, L., 1882-1972 🔗
TYYPPI	henkilö
HENKILÖN TUNNISTE	Asteri ID: 000054289 🔗 ISNI: 0000 0000 8398 0278 🔗
SYNTYMÄAIKA	1882
KUOLINAIKA	1972
HENKILÖN VARIANTTINIMI	<i>Lehtinen, Hilja Onerva, 1882-1972</i> <i>Lehtinen, Onerva, 1882-1972</i> <i>L. Onerva, 1882-1972</i> <i>Madetoja, Hilja Onerva, 1882-1972</i> <i>Madetoja, Onerva, 1882-1972</i>
SYNTYMÄPAIKKA	Helsinki (yso-paikat)
KUOLINPAIKKA	Helsinki (yso-paikat)
HENKILÖN AMMATTI TAI TEHTÄVÄ	kirjailija (Metatietosanasto) runoilija (Metatietosanasto) sanoittaja (Metatietosanasto)
HENKILÖN TOIMINTA-ALA	kaunokirjallisuus (YSO) laulelmat (YSO) laulumusiikki (YSO)
HENKILÖN KÄYTTÄMÄ KIELI	suomi (Lexvo)
HUOMAUTUS HENKILÖSTÄ	Aiempi nimi vuosina 1905-1908: Hilja Streng Syntymänimi: Hilja Onerva Lehtinen Todellinen nimi: Hilja Onerva Lehtinen Todellinen nimi vuodesta 1913: Hilja Madetoja
TIEDONLÄHDE	Wikipedia, katsottu 8.5.2018 SKS. Biografiakeskus
URI	http://urn.fi/URN:NBN:fi:au:finaf:000054289 🔗

L. Onervan tiedot Kantossa:

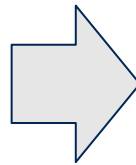
<http://urn.fi/URN:NBN:fi:au:finaf:000054289>



Temaattinen analyysi

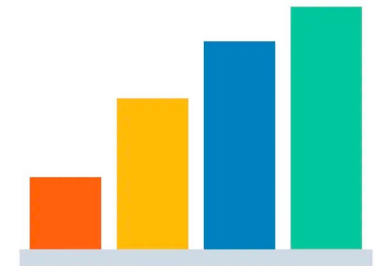
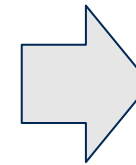


Digitoidut teokset



annif

Kansalliskirjastossa
kehitetty automaattisen
asiasanoituksen työkalu



Kaunokirjallisuuden
teemat

Datan korjausten vieminen Fennicaan

Bibliografisen datan käsittelyssä havaitaan varmasti ongelmia:

- virheelliset merkinnät (kielet, päivämäärät, koodikentät...),
- puutteelliset tekijätiedot,
- kirjoitusvirheet,
- tuplatietueet,
- kirjavat paikkakunta- ja julkaisijatiedot...

Havainnoista laaditaan raportit ja ne pyritään korjaamaan alkuperäisiin Fennica/Melinda-tietueisiin.



[Wite-Out-korjauslakka.](#)

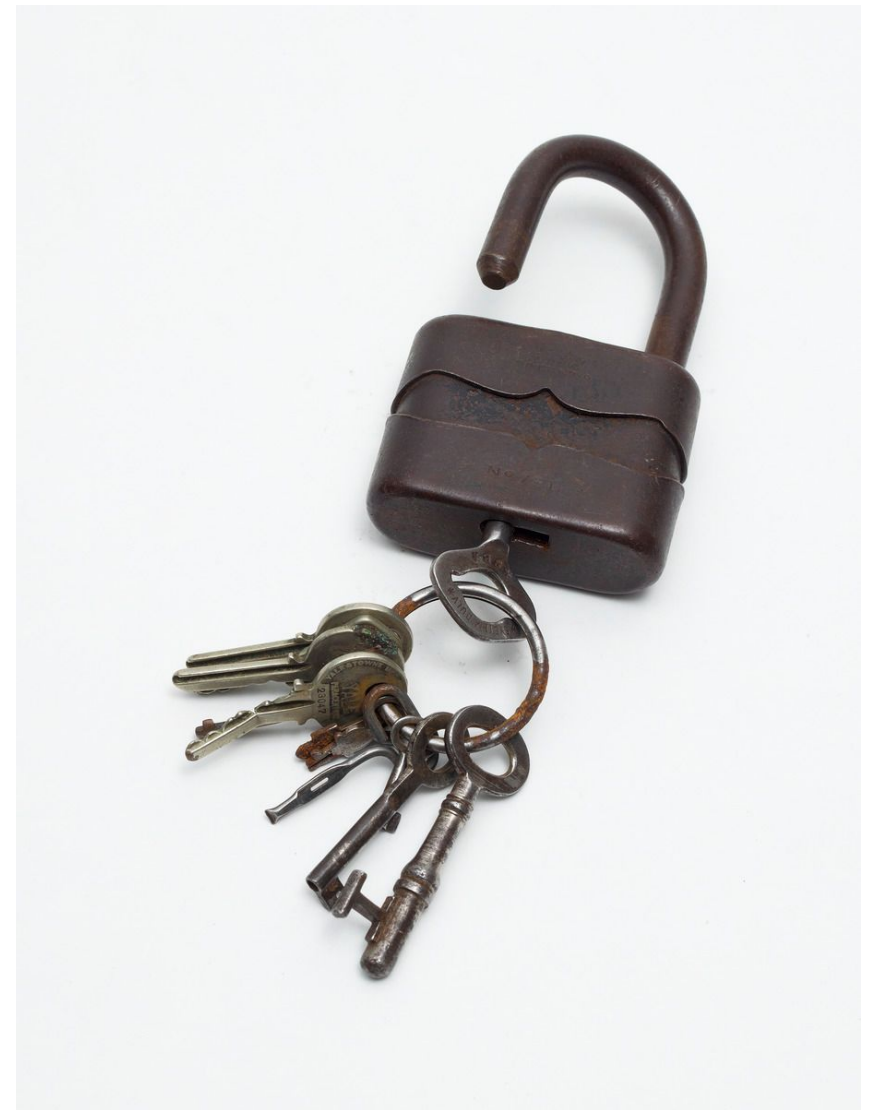
Suomen kansallismuseo,
Vankilamuseon kokoelma.
CC BY 4.0

Aineistojen avoimuus

Hankkeessa käytetään jo valmiiksi **avoimia digitaalisia aineistoja**.

Harmonisoitu metadata julkaistaan **avoimen datan pakettina**.

Hankkeen tuloksena syntyy myös **teosluettelo** sekä mahdollisia muita avoimia data-aineistoja kuten koneoppimismalleja.



[Riippulukko ja avaimet](#)

Suomen kansallismuseo, kansatieteelliset kokoelmat
CC BY 4.0

Kansalliskirjaston asiantuntijoiden työtä

- **Koordinointi** (Osma Suominen)
 - yhteydenpito tutkijoihin, pienimuotoinen tekninen apu
- **Taustatuntemus Fennicasta** (Pasi Koste, Satu Niininen, Jussi-Pekka Hakkarainen)
 - Satu Niininen toimittanut tietuelistauksia Fikka-järjestelmästä (sis. Fennican signumit)
 - Veli-Matin vierailu ja tutustuminen aineistoihin 30.8.2023
- **Apua Kanton hyödyntämiseen** (Katerina Sornova, Martin Engberg)
 - valmius auttaa Kanton/Asterin kanssa
- **Digin aineistot käyttöön** (Juha Rautiainen, Erno Liukkonen)
 - valmius auttaa digitoitujen aineistopakettien käytössä
- **Annifin käyttö digiaineistoille** (Juho Inkinen)
 - Annifin soveltamista kaunokirjallisuuden automaattiseen sisällönkuvailuun
- **Melindan/Fennican täydentäminen** (Nicholas Volk)
 - valmius tehdä korjailuja Fennican tietueisiin virhelistojen perusteella
- **Viestintätiimi**
 - hankkeen [verkkosivut](#) kolmella kielellä Kansalliskirjasto.fi:n alla



Kansalliskirjaston DHL-FI-työ vuonna 2024?

- tarjoamme aineistoja ja tukea: Fennica, Kanto, digitoidut aineistot...
- teemme korjailuja tietueisiin, kun saamme virhelistoja
- kehitämme automaattista sisällönkuvailua
- ...mitä muuta?



Kiitos!

osma.suominen@helsinki.fi