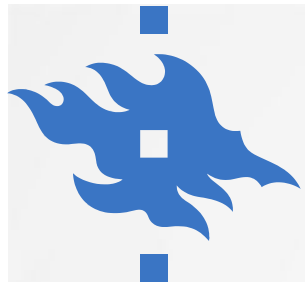




Digitaaliset ihmistieteet ja kulttuuriperintö

Eetu Mäkelä

Teknologiajohtaja, dataintensiivisen ihmistieteellisen tutkimuksen infrastruktuuri DARIAH-FI
Tutkimusohjelmajohtaja, Helsingin yliopiston humanistis-yhteiskuntatieteellinen instituutti
Ihmis- ja tietojenkäsittelytieteiden vuorovaikutuksen apulaisprofessori, Helsingin yliopisto



Sisältö

- FIN-CLARIAH ja DARIAH-FI
- Dataintensiivinen ihmistiede esimerkein
- Datakeskeisen tutkimuksen tarpeet aineistojen suhteen
- Takaisinkytkentä tutkimuksesta: alkuperäisten tietokantojen rikastus
- Finna ja FIN-CLARIAH

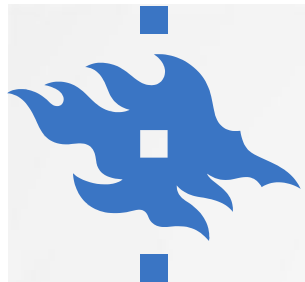




DARIAH-FI on uusi Suomen Akatemian rahoittama tutkimusinfrastruktuuri jonka missiona on tukea dataintensiivistä ihmistieteellistä tutkimusta

(DARIAH-FI + Kielivaraininfrastruktuuri FIN-CLARIN = FIN-CLARIAH)

D A R I A H - F I



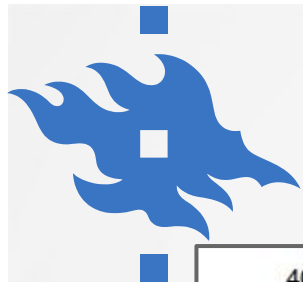
Laajojen data-aineistojen käyttö tutkimuksessa

1. laadullinen tutkimus: miten löydän isoista datalähteistä ne, jotka käsittelevät minua kiinnostavia asioita?
2. määrällinen tutkimus: data-aineistojen käyttö itsessään tilastollisiin analyyseihin

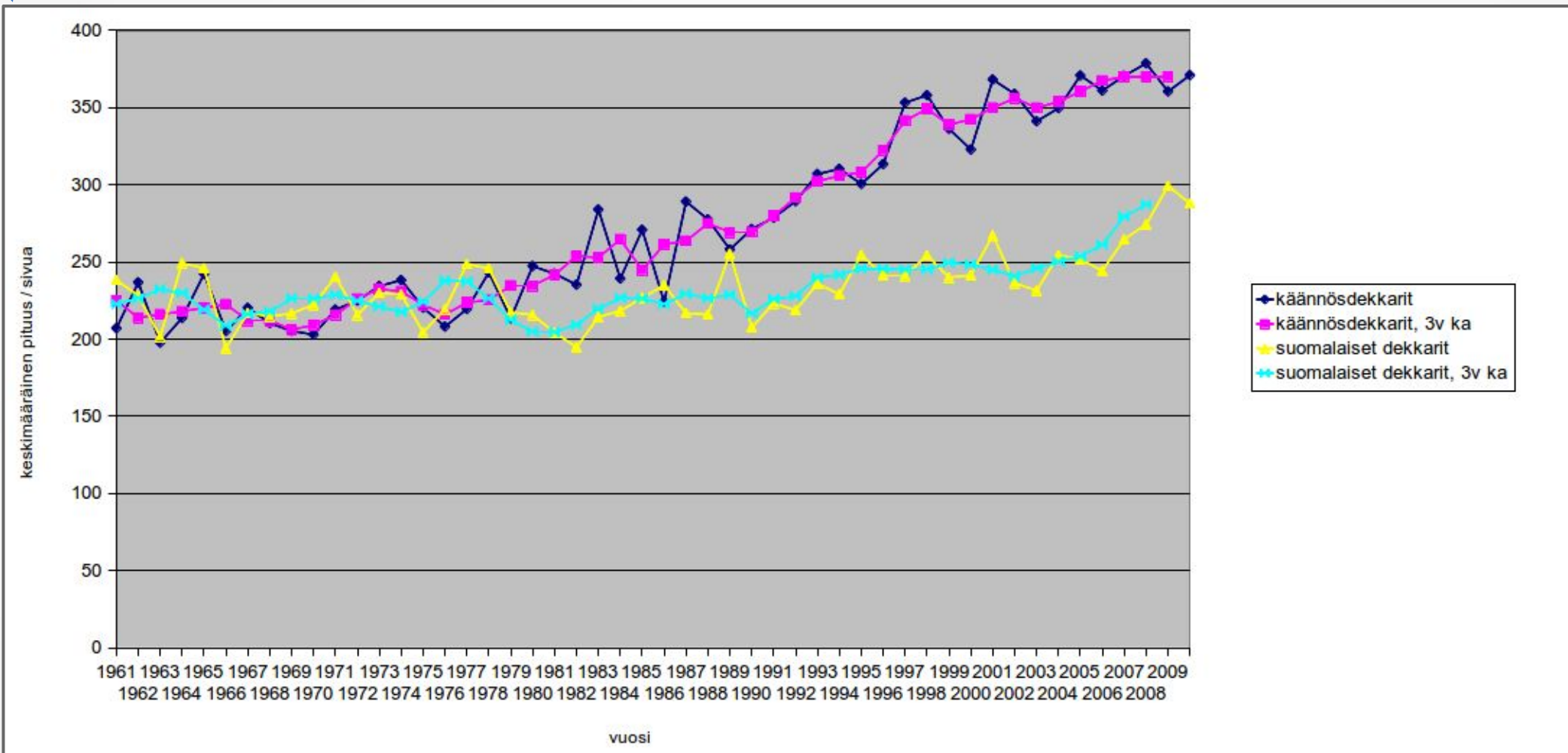


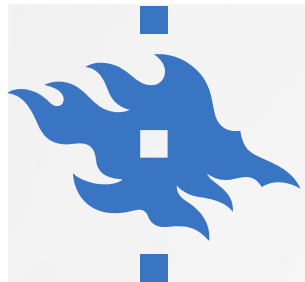


Dataintensiivinen ihmistiede esimerkein

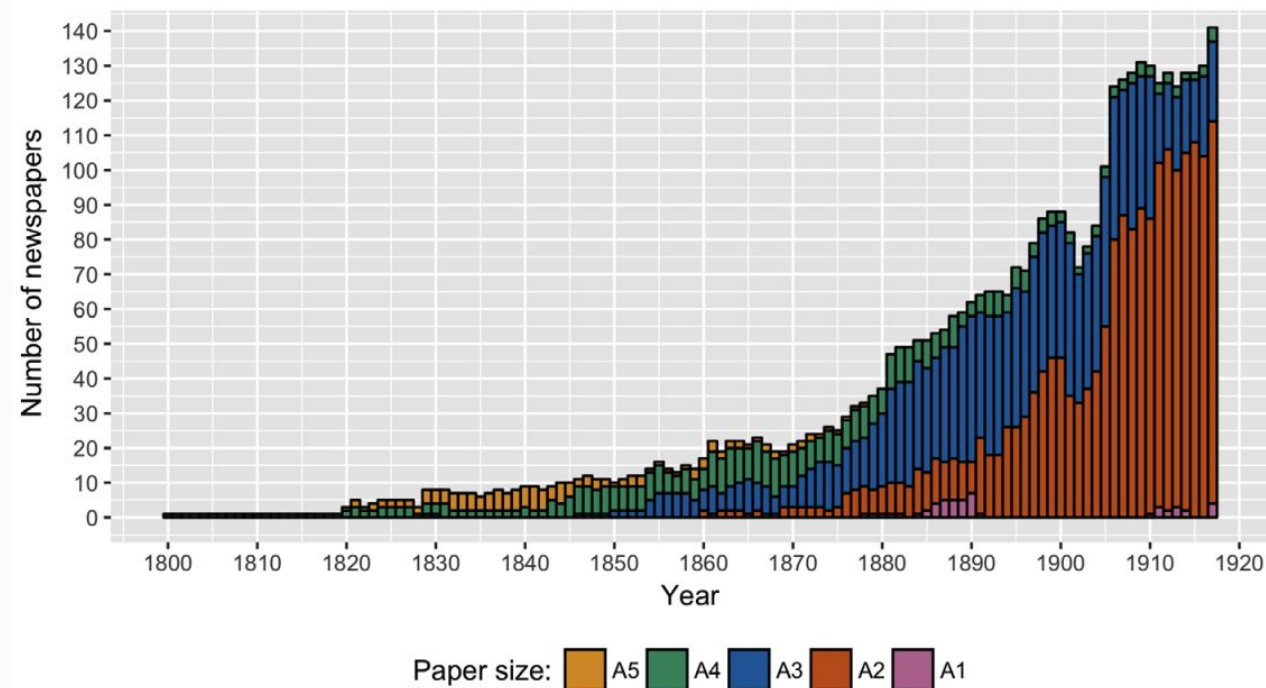
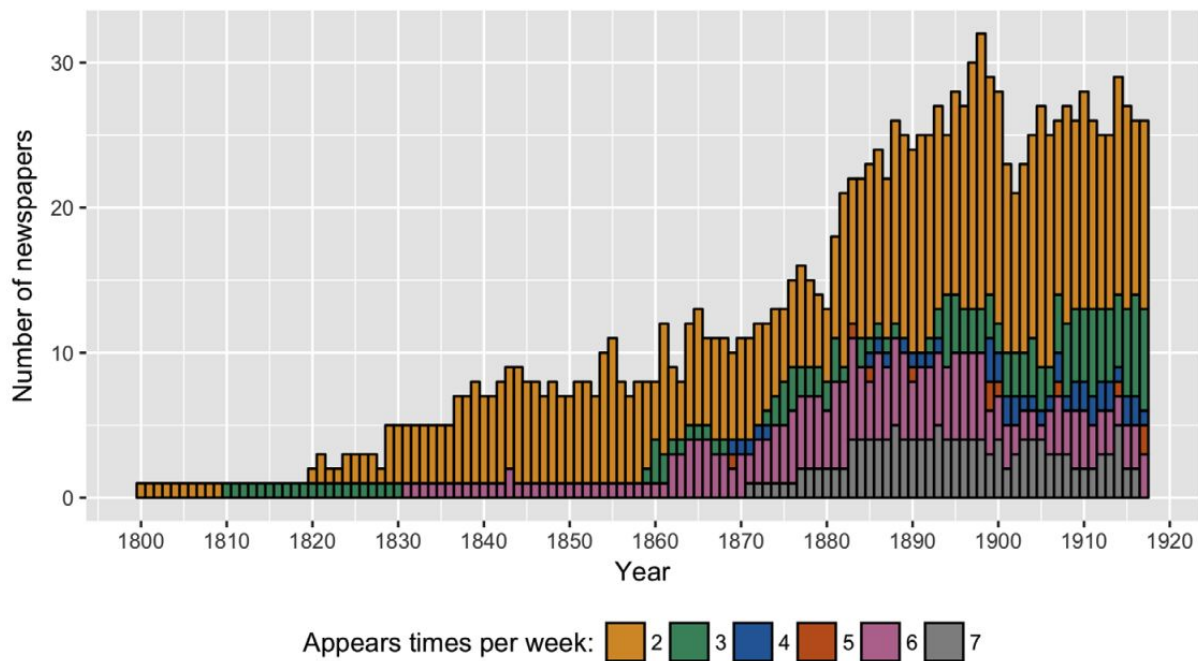


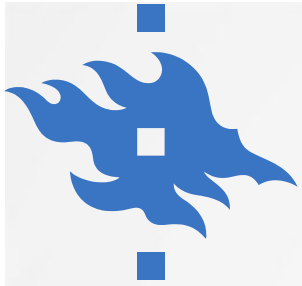
Dekkareiden pituuksien muutokset



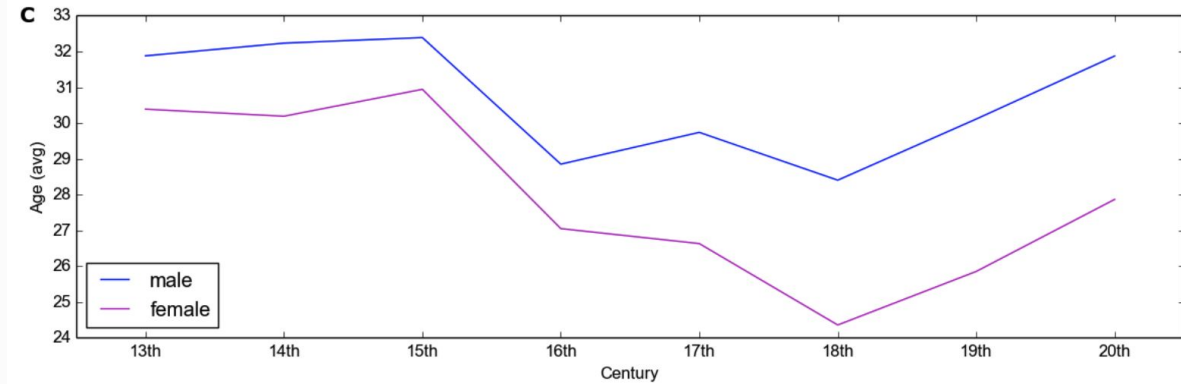
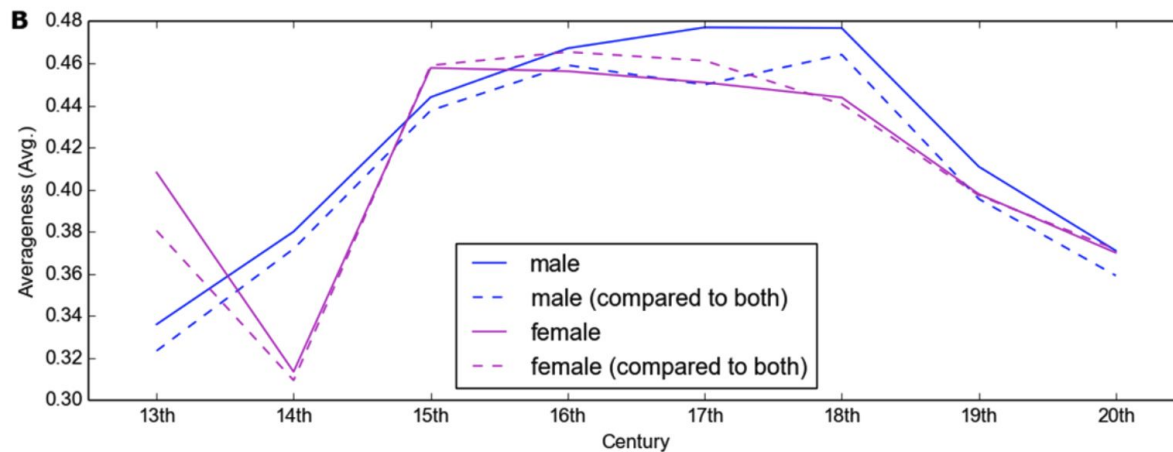
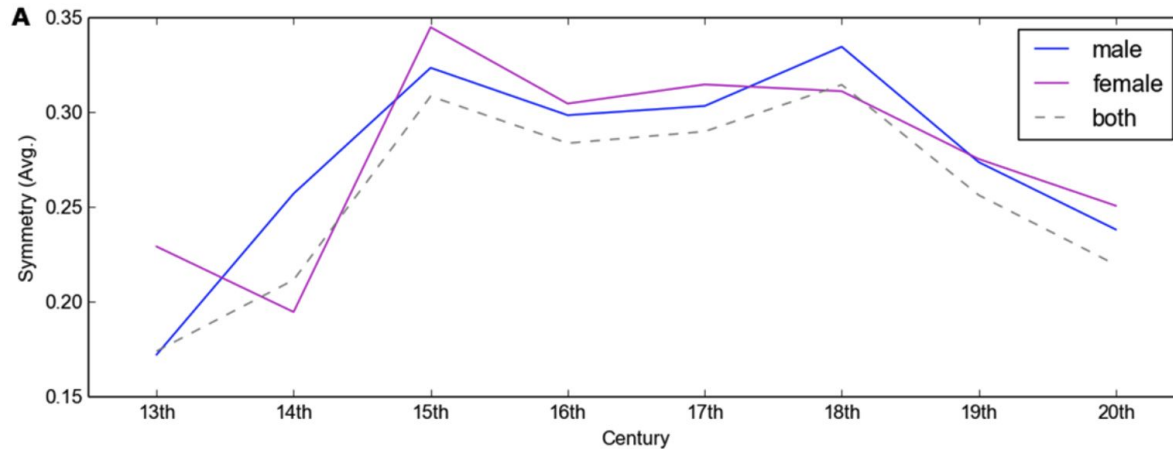


Suomalaisen sanomalehdistön kehitys



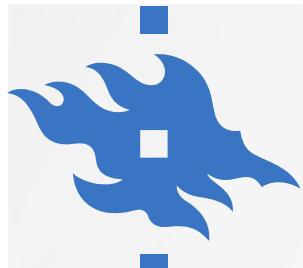


Kasvojen symmetria, keskivertoisuus ja ikä maalaustaiteessa läpi 700 vuoden



de la Rosa, Javier, and Juan-Luis Suárez. 2015. "A Quantitative Approach to Beauty. Perceived Attractiveness of Human Faces in World Painting". *International Journal for Digital Art History*, no. 1 (June). <https://doi.org/10.11588/dah.2015.1.21640>.





CultureSampo

Finnish Culture on the Semantic Web 2.0
(the portal is only partially available in English, sorry)

[Instructions](#) [Main Page](#) [About](#) [Feedback](#)

Language: suomi svenska English



Quick Search: organized [Map](#) [Browse and Search](#) [Relation Search](#) [Search and Organize](#) [Collections](#) [Finnish History](#) [Skills](#) [Biographies](#) [Kalevala](#) [Karelia](#)

Constrain objects: Tell me about which and

View constrained objects:

- 1878: painting: Kaarle Herttua
- 1879: painting: Karl Emanuel
- 1883: painting: Taiteilijan isän
- 1886: drawing: Isä lukee pikku
- 1887-1889: painting:
- 1891: painting: Juhani Ahon
- 1892-1894: painting: Gunnar
- 1893: painting: Herra Ahngerin
- 1893: painting: Herra Ahngerin muotokuva
- 1896: photograph:
- 1897: painting: Ferdinand von
- 1899: graphic art: Isä
- 1902: photograph:
- 1905: drawing: Mai
- 1906: photograph:
- 1906: photograph:

1893: painting: Herra Ahngerin muotokuva

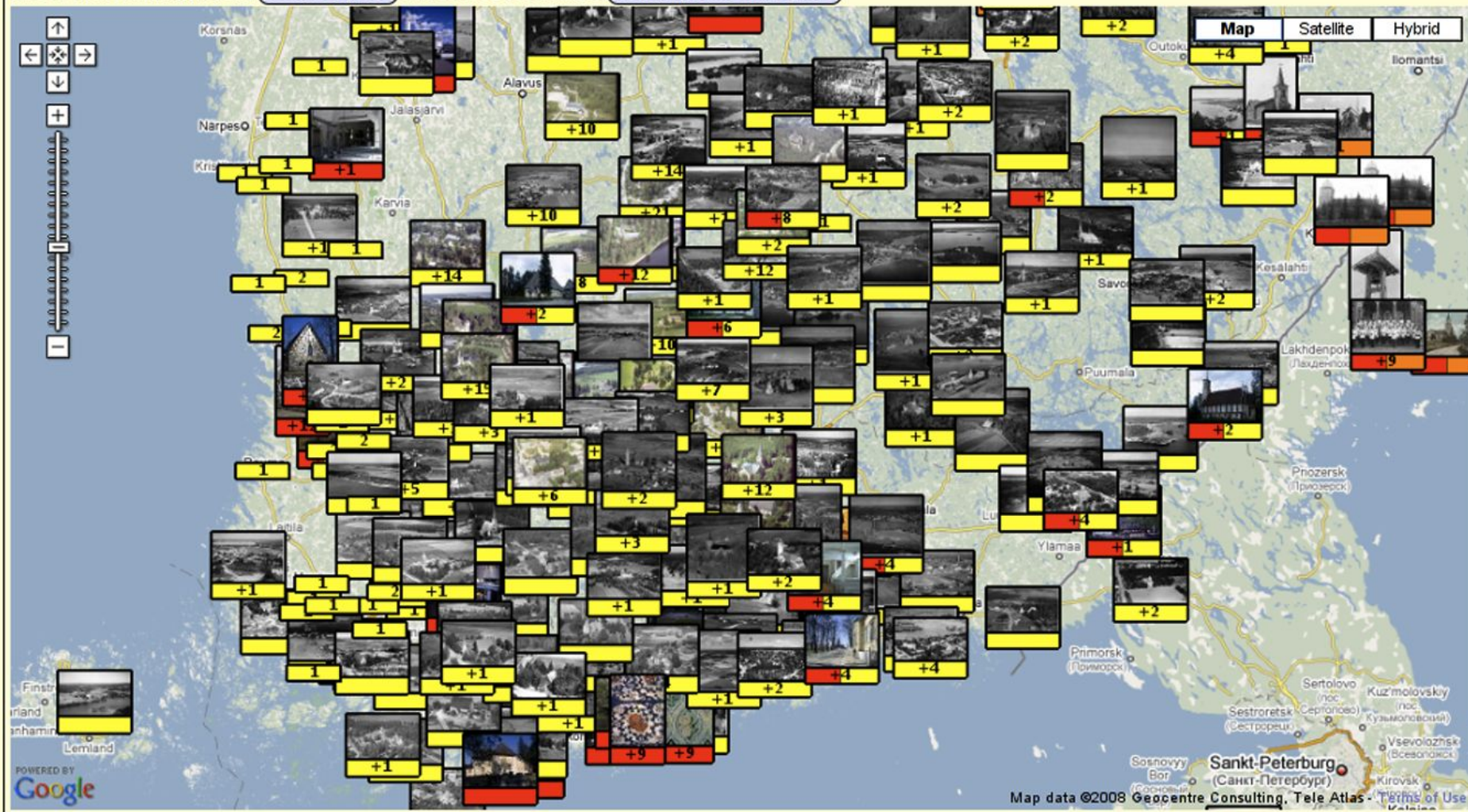
Sat, 31 Dec 1892
22:00:00 GMT
Sat, 31 Dec 1892
22:00:00 GMT



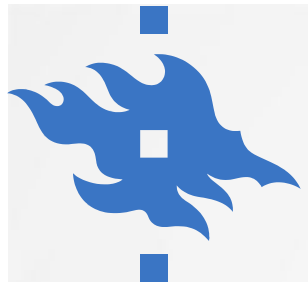
Constrain objects: Tell me about which and

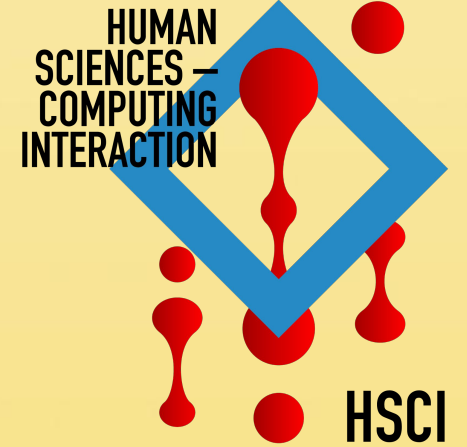
View constrained objects: [organized](#) [as a list](#) [on a map](#) [on a time line](#)

placed on the map by property and colored according to



POWERED BY Google
ominaisuuksilla
■ place of subject
■ place of manufacture ■ place of photographing





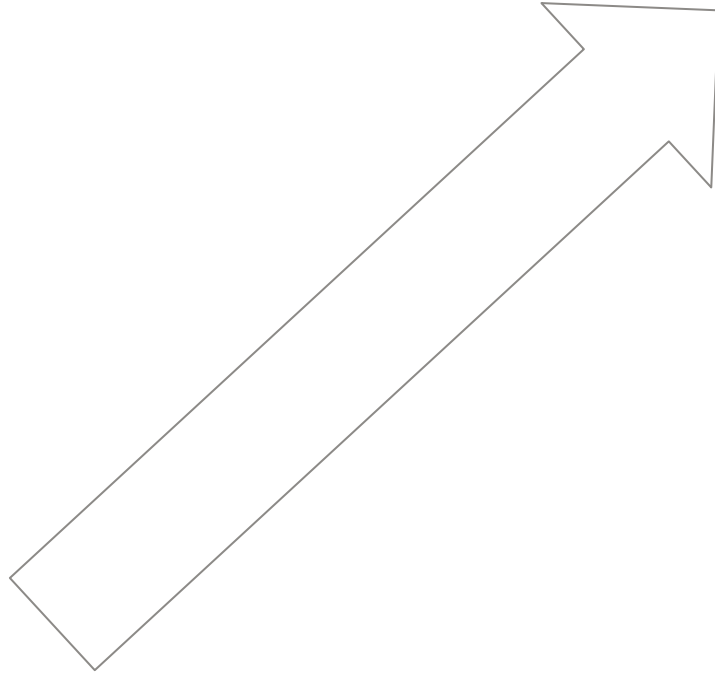
Datakeskeisen tutkimuksen tarpeet aineistojen suhteen: 1. ymmärrys aineiston rajoista

Contents of Europeana

Country	Items
Netherlands	5,207,709
Germany	4,802,046
Spain	3,645,926
Sweden	3,184,576
France	3,179,520
Italy	2,444,697
United Kingdom	2,354,445
Austria	1,873,554
Poland	1,690,506
Norway	1,500,039
Belgium	1,132,309
Europe	864,806
Denmark	793,823
Hungary	393,018
Greece	392,456
Czech Republic	360,326
Slovenia	359,369
Iceland	347,966
Estonia	251,858
Finland	225,973

Country	Items
Netherlands	5,207,709
Germany	4,802,046
Spain	3,645,926
Sweden	3,184,576
France	3,179,520
Italy	2,444,697
United Kingdom	2,354,445
Austria	1,873,554
Poland	1,690,506
Norway	1,500,039
Belgium	1,132,309
Europe	864,806
Denmark	793,823
Hungary	393,018
Greece	392,456
Czech Republic	360,326
Slovenia	359,369
Iceland	347,966
Estonia	251,858
Finland	225,973

???

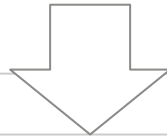


keyword	items	translation
metsätalous	12922	forest economy
puut	7078	trees
puutavara	6622	woodstuff
metsät	6440	forests
uitto	6342	log driving
talvi	5857	winter
kuljetus	5577	transport
miehet	5514	men
työvälineet	4679	tools
metsänhoito	4619	forestry
rakennukset	4544	buildings
lumi	4504	snow

???

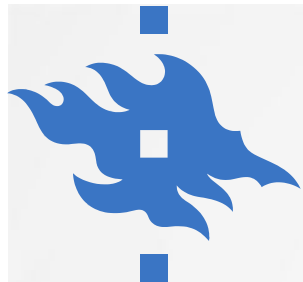
Country	Items
Netherlands	5,207,709
Germany	4,802,046
Spain	3,645,926
Sweden	3,184,576
France	3,179,520
Italy	2,444,697
United Kingdom	2,354,445
Austria	1,873,554
Poland	1,690,506
Norway	1,500,039
Belgium	1,132,309
Europe	864,806
Denmark	793,823
Hungary	393,018
Greece	392,456
Czech Republic	360,326
Slovenia	359,369
Iceland	347,966
Estonia	251,858
Finland	225,973

Lusto - The Finnish Forest Museum



collection	IMAGE	TEXT	SOUND	TOTAL
2021008_Ag_FI_NDL_justo	70,869			70,869
9200301_Ag_EU_TEL_a0611_Newspapers_Finland		48,394		48,394
2021002_Ag_FI_NDL_fng	11,428			11,380
2021007_Ag_FI_NDL_slis	37	2,711	183	2,931
2021003_Ag_FI_NDL_fragmenta	1,619			1,619
2021001_Ag_FI_NDL_Raita_collection			494	494
92080_Ag_FL_TEL_a0486_ETravelFinland		113		113
total	83,953	51,218	677	135,800

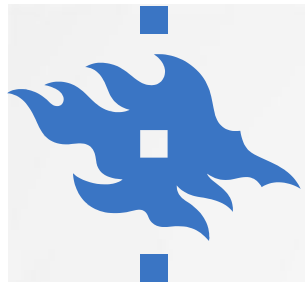
keyword	items	translation
metsätalous	12922	forest economy
puut	7078	trees
puutavara	6622	woodstuff
metsät	6440	forests
uitto	6342	log driving
talvi	5857	winter
kuljetus	5577	transport
miehet	5514	men
työvälineet	4679	tools
metsänhoito	4619	forestry
rakennukset	4544	buildings
lumi	4504	snow



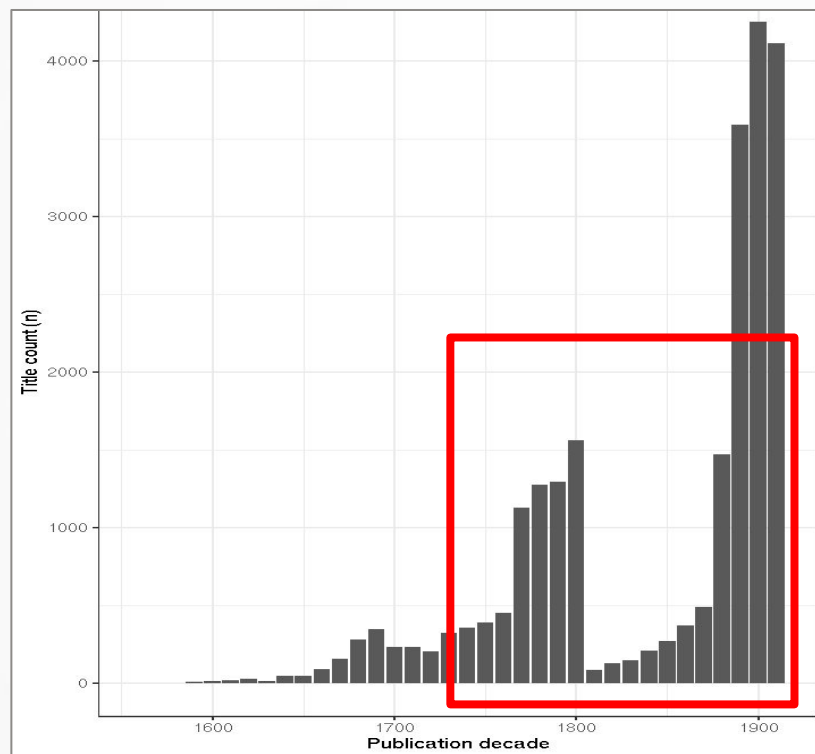
Finnan top avainsanat

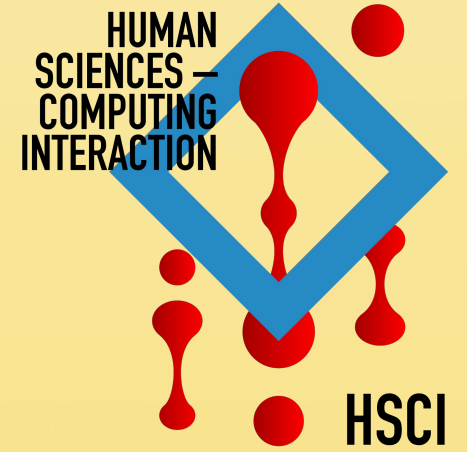
314,969 Suomi	62,501 Venäjä	47,696 rakentaminen	40,043 Maalöydöt
238,606 historia	61,916 Arkeologiset löydöt	47,439 puurakennukset	39,753 talvi
181,875 Finland	61,915 Kaivauslöydöt	47,388 vaatteet	39,168 metsätalous
160,014 Helsinki	61,508 1900-luku	47,331 autot	39,028 Sverige
113,945 Yhdysvallat	60,336 personhistoria	45,918 lapset	38,688 1970-luku
107,163 naiset	58,815 1800-luku	45,643 Iso-Britannia	38,171 författare
104,305 historia.	58,807 kirjailijat	44,140 1990-luku	37,875 Suomi.
103,809 arkistokuva	56,975 yritykset	43,661 Tampere	37,860 kehittäminen
89,754 rakennukset	56,597 Ruotsi	41,899 koulutus	37,849 Uleåborg
85,194 henkilöhistoria	54,907 lapset (ikäryhmät)	41,184 Företagsekonomi, förvaltning och marknadsföring	37,500 2000-2009
82,964 miehet	54,336 nuoret	41,085 Business Management, Administration and Marketing	37,361 1980-luku
69,107 arkkitehtuuri	49,399 artikkelit (julkaisut)	40,891 Saksa	37,259 matkailu
67,278 kartat	48,938 Oulu		37,133 asuinrakennukset
62,635 rock	47,719 tutkimus		





Fennica - Suomen kansallisbibliografia



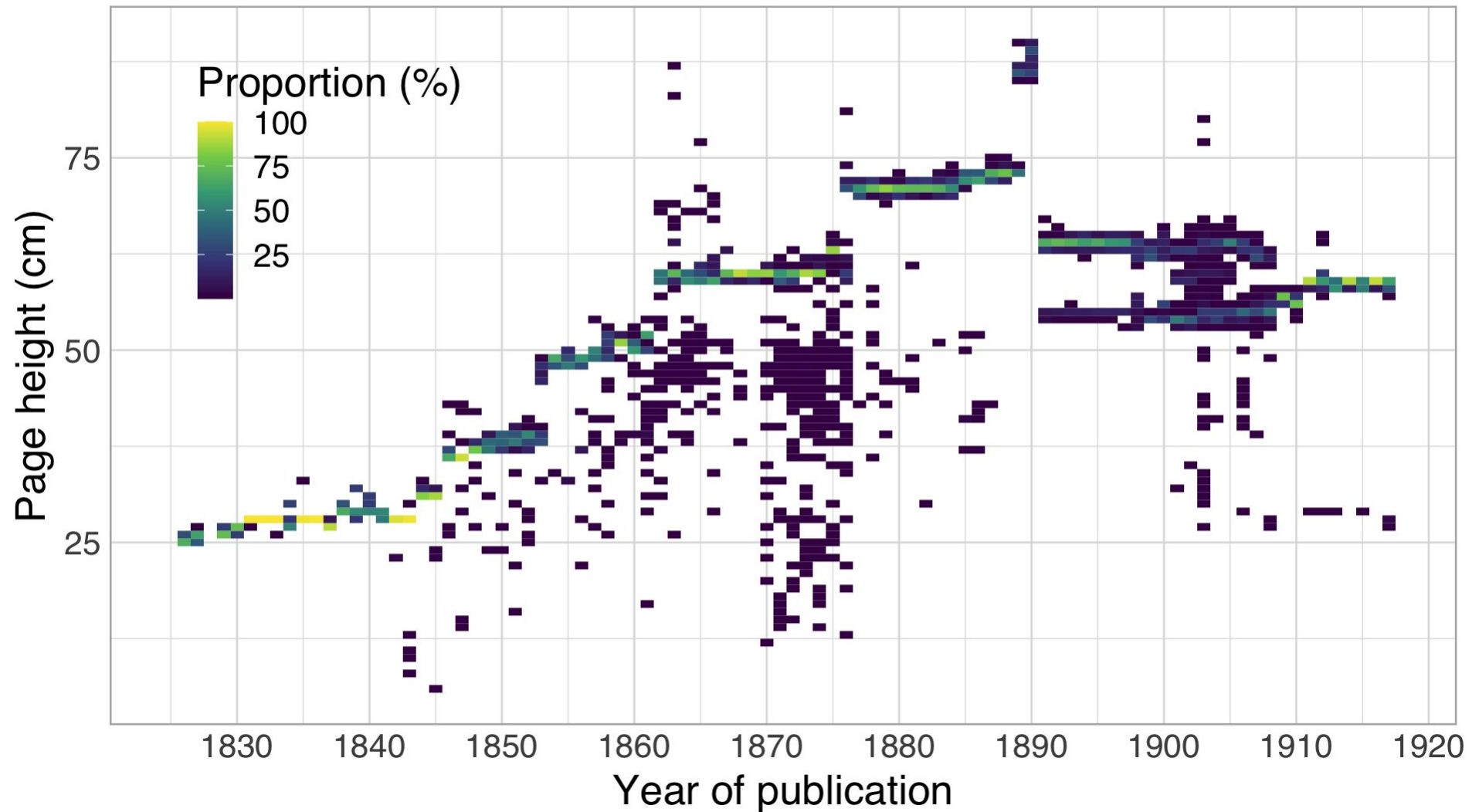


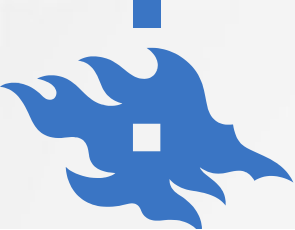
**Datakeskeisen tutkimuksen
tarpeet aineistojen suhteen:**
1. ymmärrys aineiston rajoista
2. rikkain muoto aineistosta



“Åbo Underrättelser, sivukoko: A3”

Proportion of pages within height brackets by year





“Åbo Underrättelser, sivukoko: A3”



Åbo Underrättelser

nro 101, 16.4.1885
Turku, Förlags Ab Sydvästkusten



Åbo Underrättelser

nro 83, 27.3.1885
Turku, Förlags Ab Sydvästkusten



Åbo Underrättelser

nro 59, 2.3.1885
Turku, Förlags Ab Sydvästkusten



Åbo Underrättelser

nro 60, 3.3.1885
Turku, Förlags Ab Sydvästkusten



Åbo Underrättelser

nro 291, 26.10.1885
Turku, Förlags Ab Sydvästkusten



Åbo Underrättelser

nro 333, 7.12.1885
Turku, Förlags Ab Sydvästkusten



Åbo Underrättelser

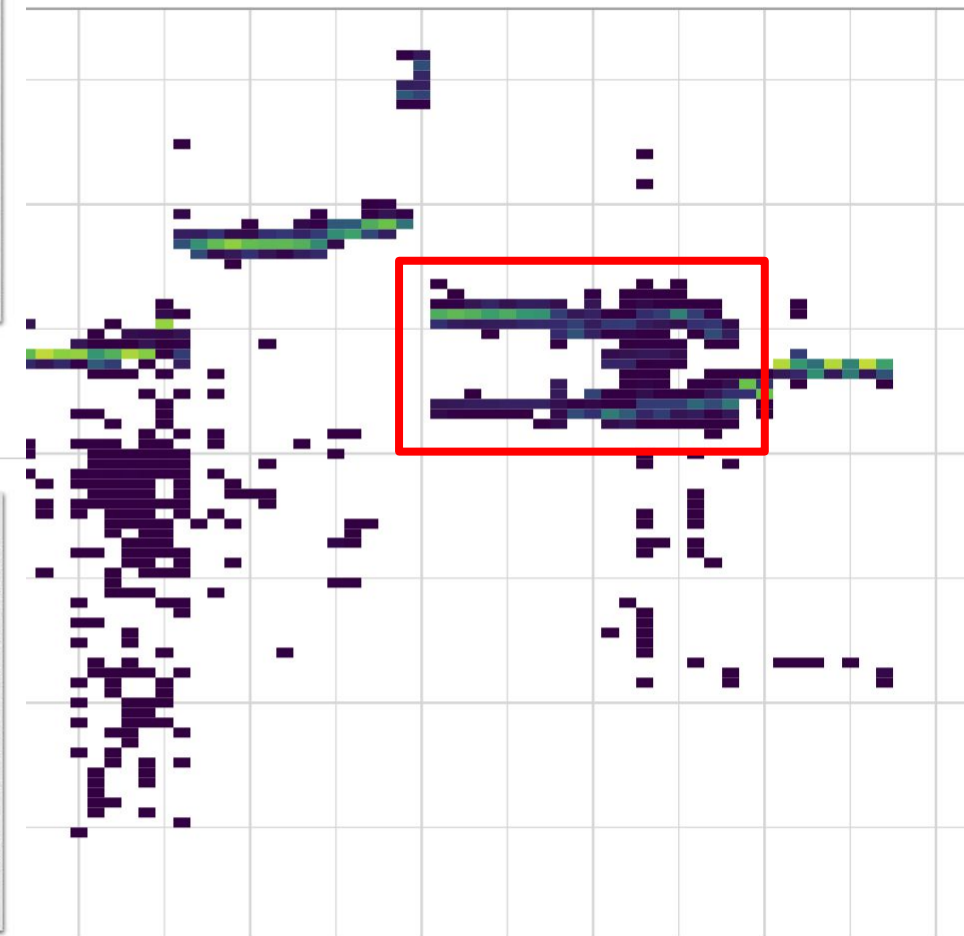
nro 210, 6.8.1885
Turku, Förlags Ab Sydvästkusten



Åbo Underrättelser

nro 97, 12.4.1885
Turku, Förlags Ab Sydvästkusten

eight brackets by year



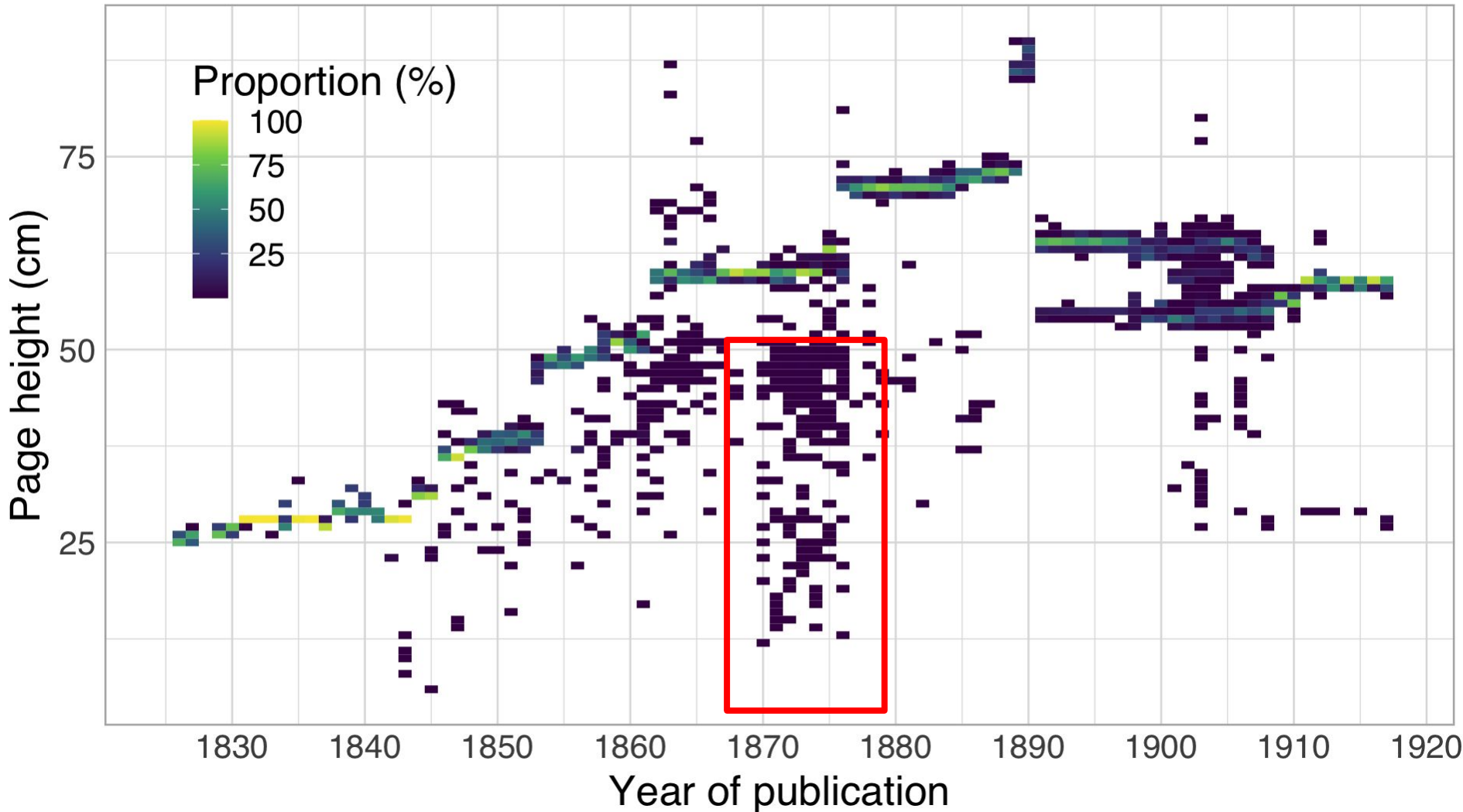
1870 1880 1890 1900 1910 1920

of publication



“Åbo Underrättelser, sivukoko: A3”

Proportion of pages within height brackets by year



Päivän kysymyksiä.

Kirjottanut
O. M. R.

Järvi, Uuden Suomen kirjapaino Suojatie,
Cia Oksanen, 1906.

Päivän kysymyksiä.

Sinä päivänä, jota nyt elämme, olemme luomassa uutta Suomea. Suomi ei ole ainoastaan saanut tahtain taikii entiset perustuslainsäädännökseen, se on jättänyt saanut uusia vapauksia, jotka tulevat kantamaan vielä arvaamattomia hedelmiä. Sitä miikka, suurta poliittista juur-lakkowittoa, jolloin Suomen uudestaantyminen tapahtui, ei saata yksikään unohtaa, ken waan silloin muhana oli. Silloin tapahtui hki niistä

Sähkösanoma „Kokkolalle“.

Ei sovintoa — Sota tulossa!

Pietari 7 p. Helmik. Hallituksen Sanansaattaja ilmoittaa että Japanin lähettiläs on antanut nootin, jonka sisältö on että Japani lakkauttaa enemmän keskustelut ja kutsuu lähetyksensä Pietarista. Sen jälkeen antoi Keisari käskyn että Wenäjän lähetyksen on paikalla lähteminen Tokiosta.

Uutistoimisto.

Hinta 5 penniä

Lisälehti HELSINGIN SANOMILLE

Helmikuun 13 p. 1907

Taaskin kamala murhatyö Wiipurissa.

Kasjanyöstö?

Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

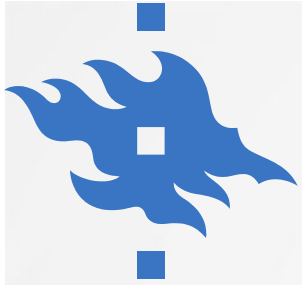
Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

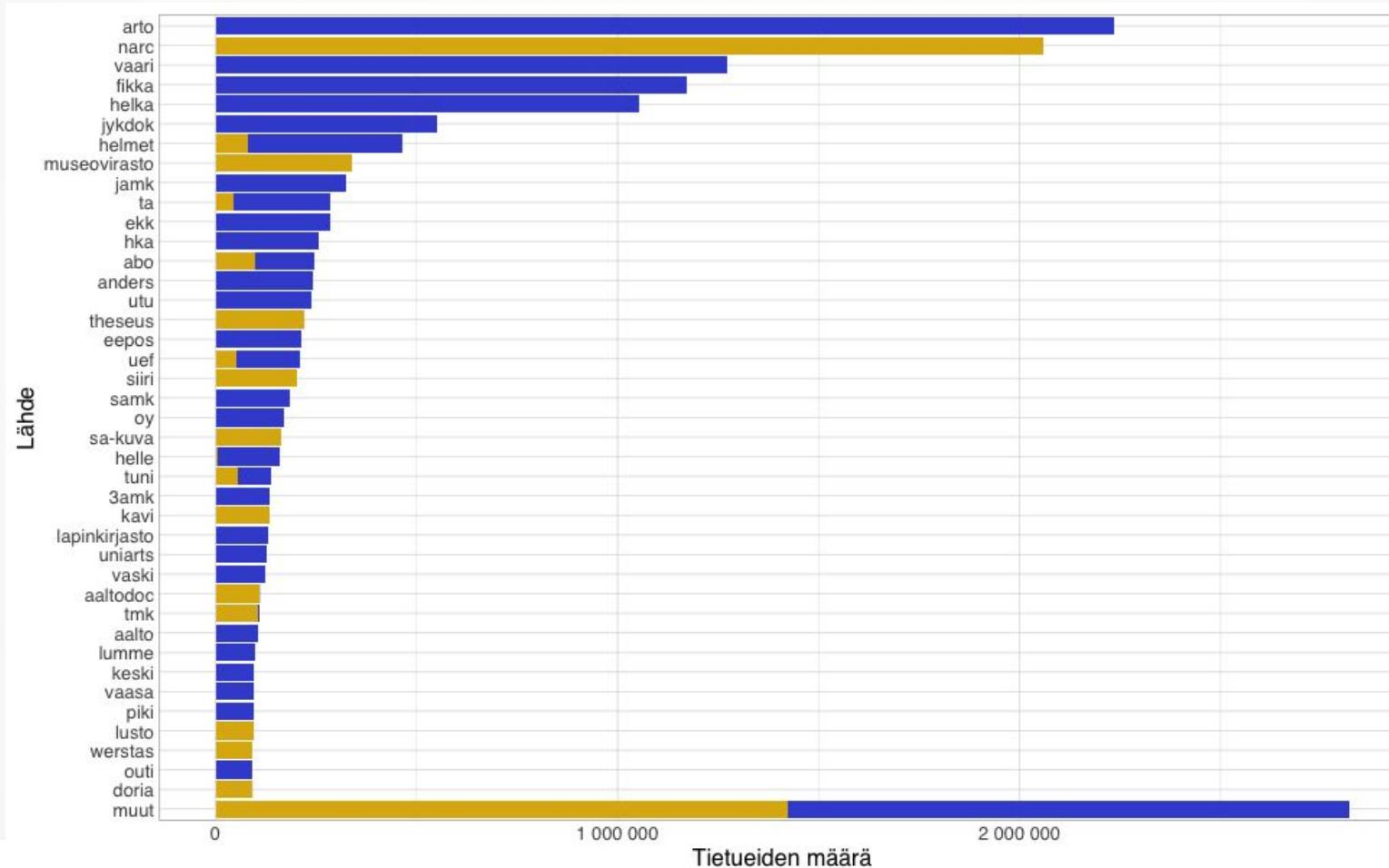
Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

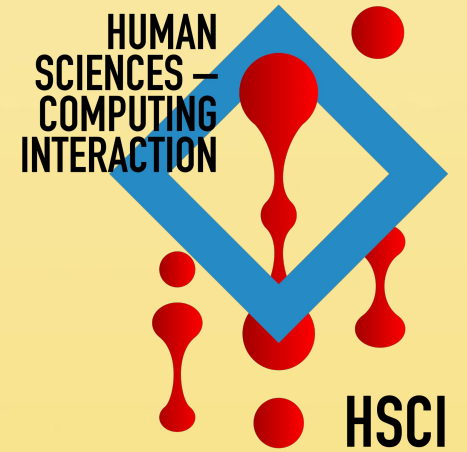
Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...

Suomenlehti on ollut...
Suomenlehti on ollut...
Suomenlehti on ollut...



Finnan alkuperäistietueiden saatavuus lähteittäin





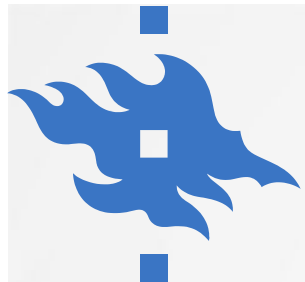
Takaisinkytkentä tutkimuksesta: alkuperäisten tietokantojen rikastus



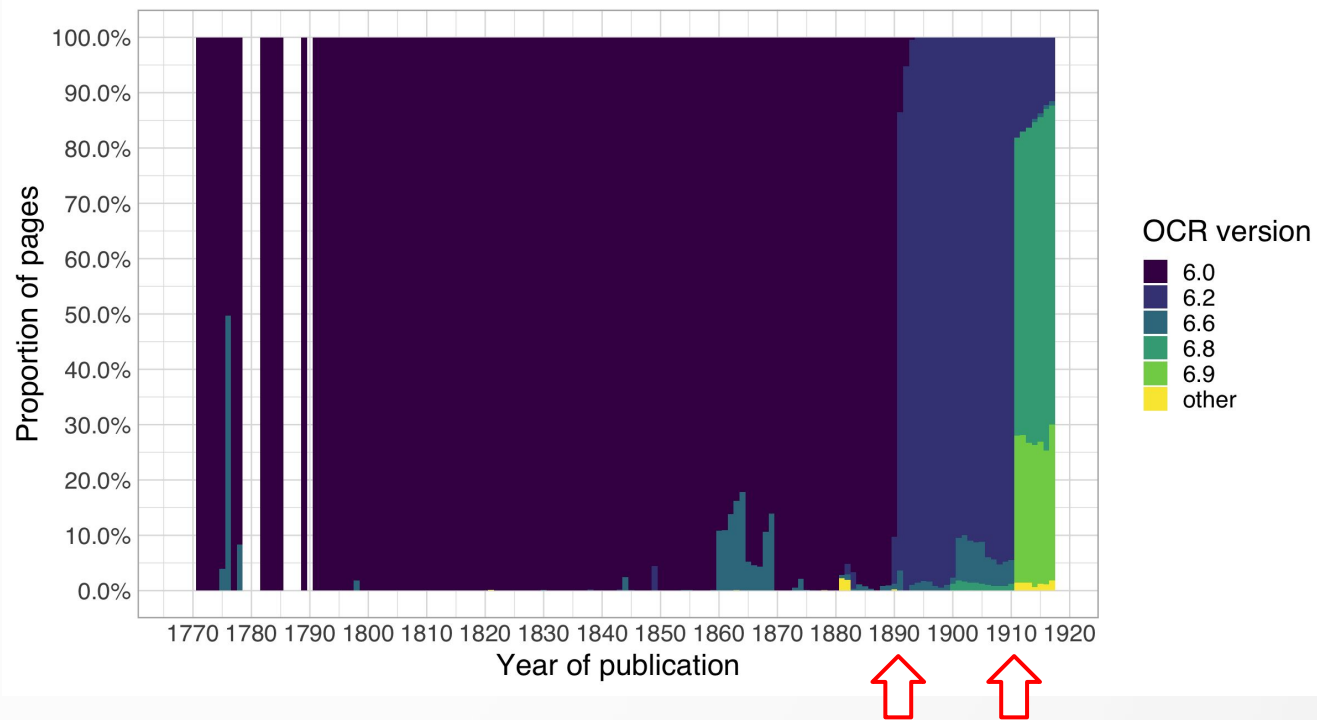
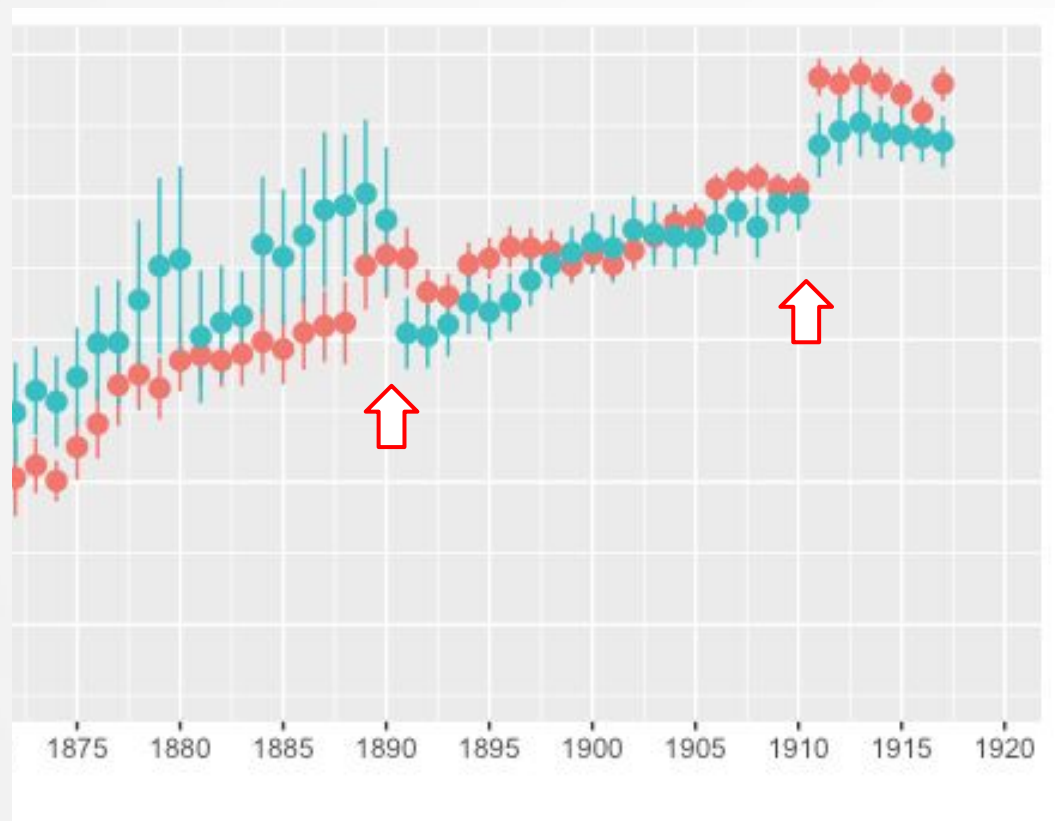
Metatietojen harmonisointi ja rikastus

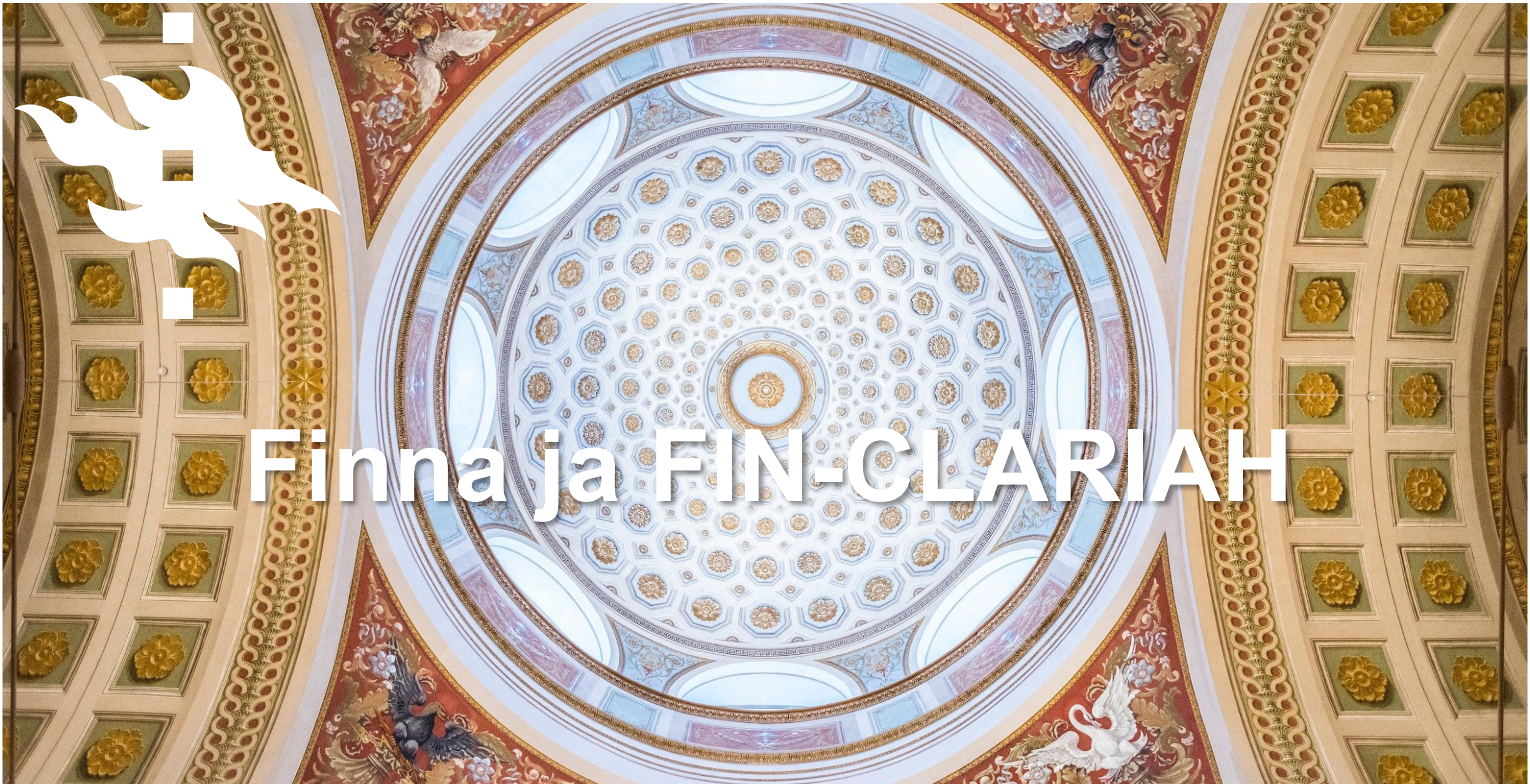
- “[4],vii-xii,[4],222p.,plate” → 240 pages
- Nimivaihtelu
 - Akseli Gallen-Kallela
 - Gallén-Kállela, Akseli
 - Akseli Valdemar Gallen-Kallela
 - Gallen, Aksel'
 - Gallen, Axel
 - Gallén, Axel Waldemar
 - Kalela, Akselis Valdemaras Galenas-
 - Гален-Калела, А. 1865-1931 Аксэли
 - + ~130 muuta
- Genrerikastus
- Uustranskriptio modernein tekstintunnistusmenetelmin
- ...



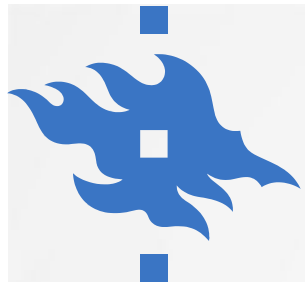


Virheiden löytäminen ja korjaaminen





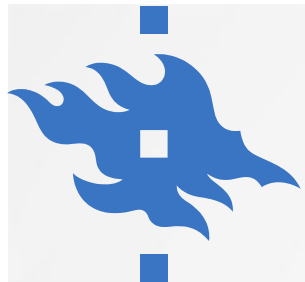
Finna ja FIN-CLARIAH



Finna ja FIN-CLARIAH

- Tutkijat tarvitsevat aineiston mahdollisimman rikkaassa muodossa, ovat valmiita näkemään vaivaa sen käytön eteen
 - Mieluummin sotkuista ja alkuperäistä kuin typistettyä.
- Mahdollisimman kattava ymmärrys aineiston rajoista ja puutteista tärkeää.
 - Tutkijat ovat kuitenkin valmiita näkemään vaivaa tämänkin eteen, eivätkä (yleensä) tee typeryyksiä.
 - Jos mahdollista, ylläpitäkää jotain kuvausta aineistostanne, mutta älkää pelätkö antaa välittömästi sotkuista ja alkuperäistä.

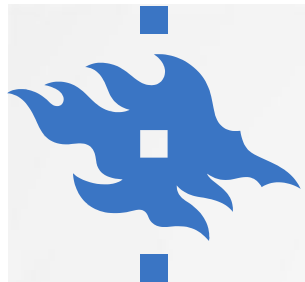




Finna ja FIN-CLARIAH

- Finna voi toimia aineistojen välityskanavana tutkijoille.
- Olemme kiinnostuneita sekä metadatatista että primääriaineistoista
- Pääseekö Finnan kautta konsistentisti haravoimaan erilaisia primääriaineistoja?
- Olennaista saada kustakin aineistosta mahdollisimman hyvin tietää mitä se edustaa, miten se on kuvailtu jne (Mutta! Paljon mieluummin sotkuinen aineisto josta ei tiedä kuin ei aineistoa!)

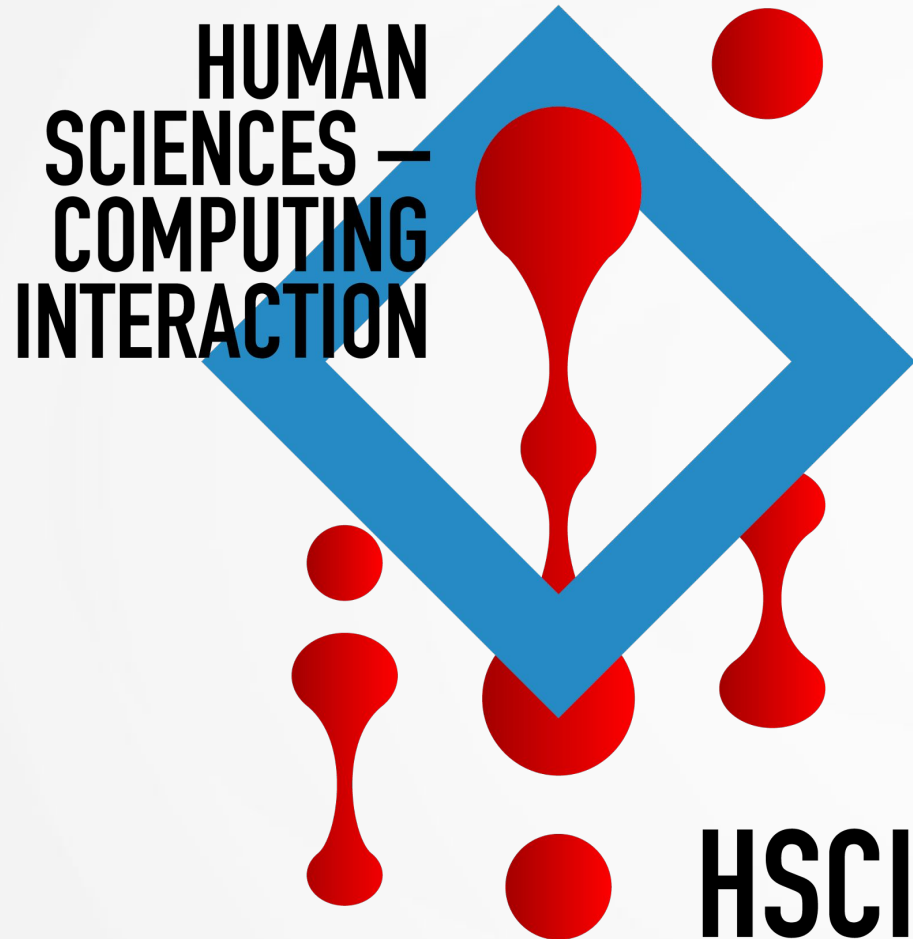
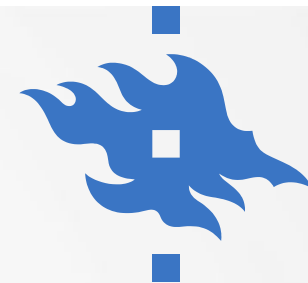




Finna ja FIN-CLARIAH

- Työnjako: FIN-CLARIAH suunniteltu niin, että kulttuuriperintölaitokset voivat vain luovuttaa aineistonsa infrastruktuurille, infrastruktuuri hoitaa tutkimuskäytön tuen (koska tutkimuskäytöllä hyvin erityiset tarpeet, ei järkeä yrittää tukea osana “yleispalvelua”)
- DARIAH-FI:n nykyvaiheen työpaketeissa mukana myös kulttuuriperintöorganisaatioita:
 - WP3.1 Kansalliskirjaston isojen aineistojen siirto tutkimusympäristöön
 - WP3.2 Kansallisarkiston massadigitointiaineistojen siirto tutkimusympäristöön
 - (+ WP4.3 Aineistojen vinoumien selvitys ja tutkimuskäytettävyyden tuki)
- DARIAH-FI:n seuraavassa vaiheessa suunniteltuna fokusalueena rakenteiset kulttuuriperintöaineistot





<http://heldig.fi/hsci>

Eetu Mäkelä

eetu.makela@helsinki.fi

<http://iki.fi/eetu.makela>

[@jiemakel](#)