

KANSALLISKIRJASTON AINEISTOT SUOMEN AKATEMIAN RAHOITAMASSA DIGITAALISEN HISTORIAN TUTKIMUKSEN (COMHIS) PROJEKTISSA

MIKKO TOLONEN

29.9.2016

Digitalia

Digitaalisen tiedonhallinnan
tutkimus- ja kehittämiskeskus



UNIVERSITY OF HELSINKI

Leverage from
the EU
2014–2020



Digital Humanities Academy Programme

COMPUTATIONAL HISTORY AND
THE TRANSFORMATION OF PUBLIC
DISCOURSE IN FINLAND, 1640–1910
(COMHIS)



UNIVERSITY OF HELSINKI

Computational History and the Transformation of Public Discourse in Finland, 1640–1910 (COMHIS)

Consortium partners:

- National Library of Finland, Centre for Preservation and Digitisation
- University of Helsinki, Faculty of Humanities
- University of Turku, Dept of Information Technology
- University of Turku, Dept of Cultural History



UNIVERSITY OF HELSINKI



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



Turun yliopisto
University of Turku

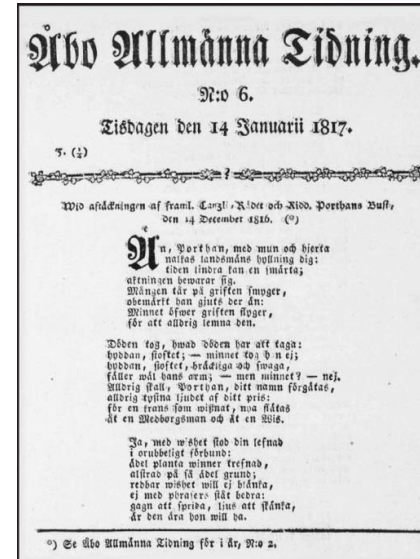
Research teams

- National Library of Finland: Kimmo Kettunen (PI), one post-doc
- University of Helsinki: Mikko Tolonen (PI), Leo Lahti, Jani Marjanen, Hege Roivainen
- University of Turku: Hannu Salmi (PI), Tapio Salakoski (PI), Heidi Hakkarainen, Asko Nivala, Heli Rantala



Objectives

Reassessing the scope, nature and transnational connections of public discourse in Finland 1640–1910.



Main resources:

- Fennica metadata (1640-1910)
- Finnish Newspapers (1771-1910)



UNIVERSITY OF HELSINKI

COMHIS Overview

Bibliographic metadata

Publishing trends and the development of public discourse

Full text analysis

Viral texts and social networks of Finnish newspaper publicity

Open source data analytics & methodologies



UNIVERSITY OF HELSINKI

QUALITY EXAMPLES OF HISTORICAL NEWSPAPER COLLECTIONS

- The British Library's 19th century collection has an estimated word accuracy of 78 %
- The estimated word accuracy (word recognition rate) of Digi is about 70-75 %
- These are quite low figures but realistic for OCR'd old newspaper collections



CONSEQUENCES

- Texts are hard to read for users
- Users may have difficulties in searching the collection
- Search results may be worse than expected anyhow
- Data mining and any further processing becomes more difficult

- Re-OCR'ing and post-correction of newspaper data needed, perhaps 80+ % word accuracy can be achieved



MITÄ VOIDAAN TEHDÄ METADATALLA?

Data: bibliographic catalogues 1640-1828

Fennica Finnish National bibliography

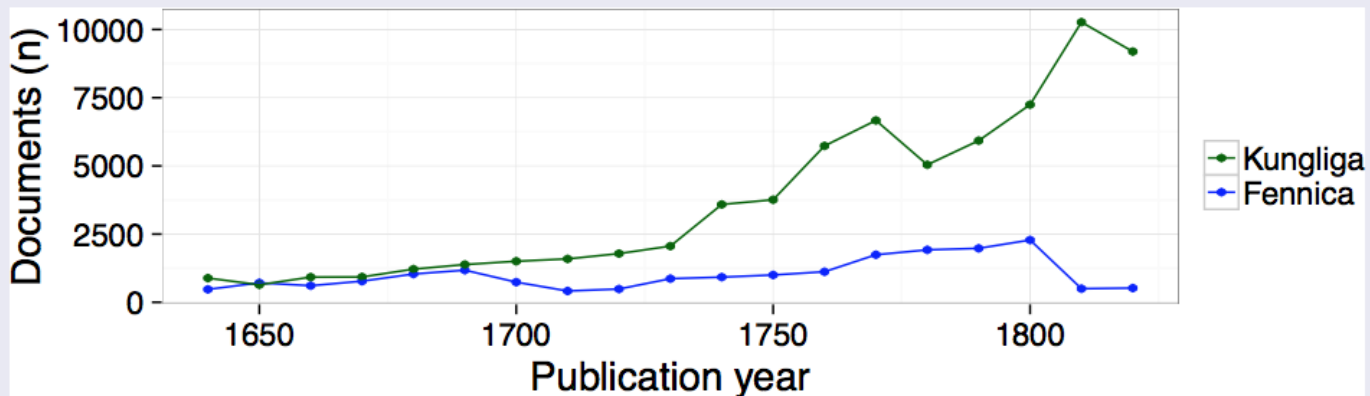
- ▶ 19241 documents 1640-1828

Kungliga Swedish National bibliography

- ▶ 69108 documents 1640-1828

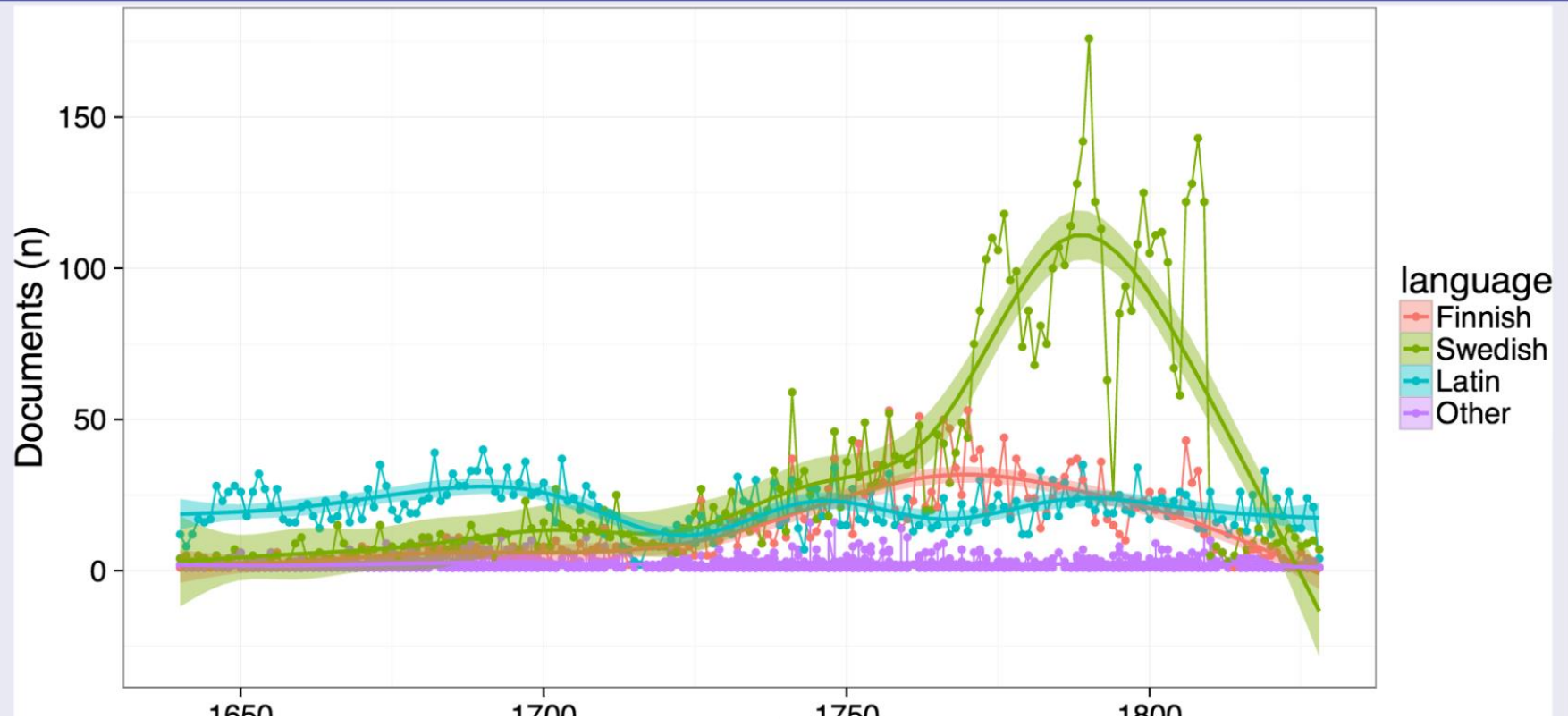
Why this project?

- ▶ Visible roots of the diversity of “Finnishness” reflected in knowledge production, 1640-1828



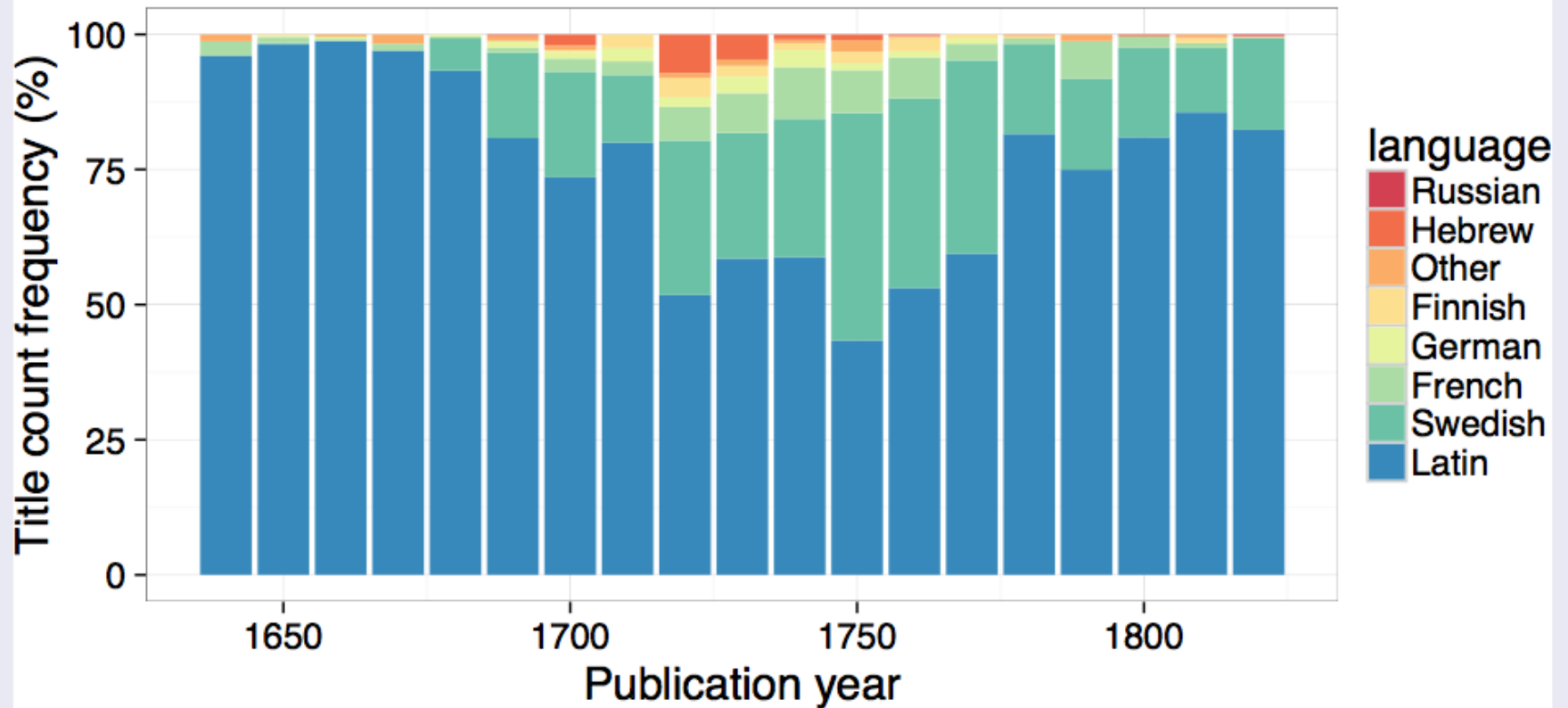
ASIASANAT FENNICASSA

Topic richness per language (publications in Fennica)



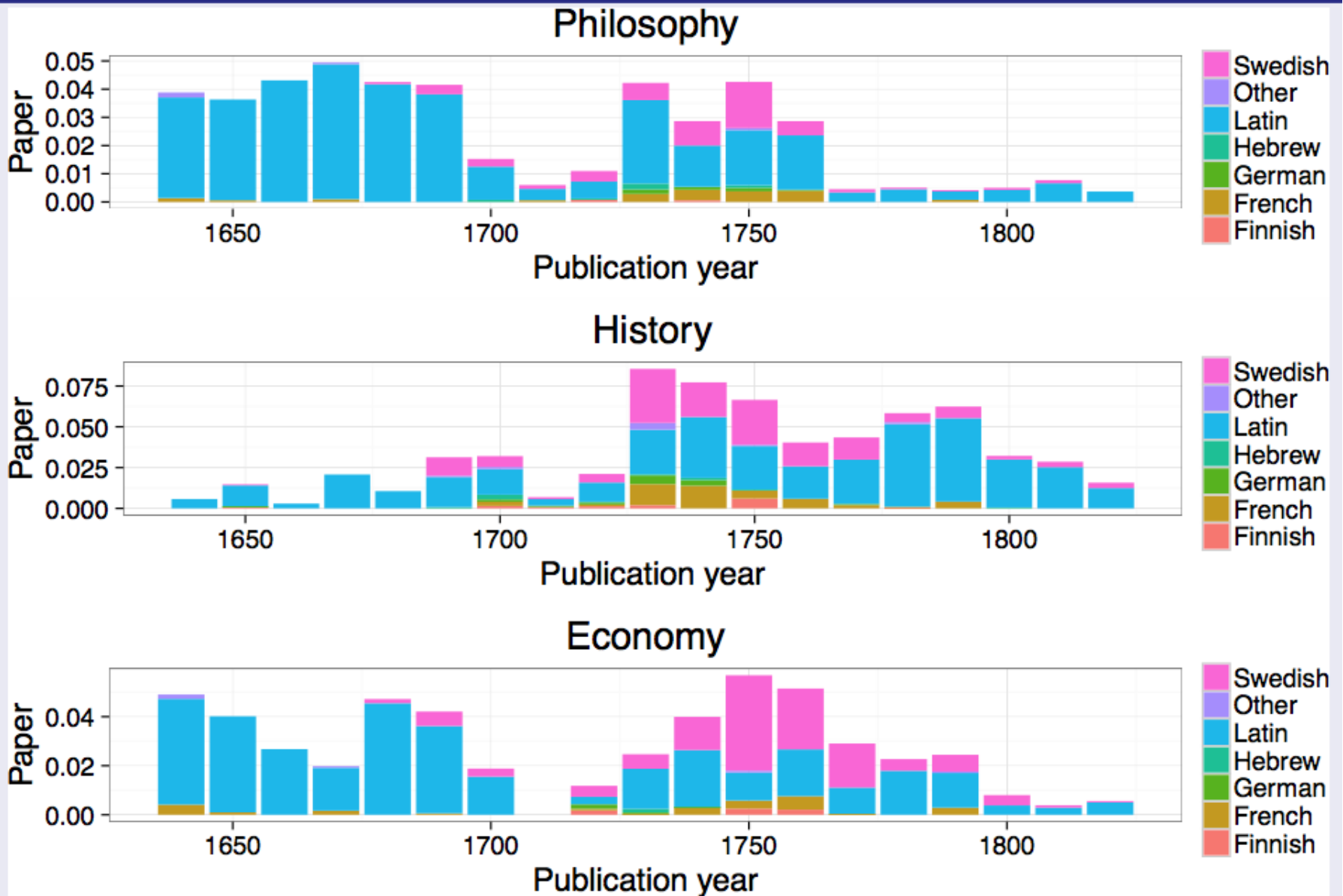
Dissertation languages in Turku 1640-1828

Dissertation languages (Fennica)



UNIVERSITY OF HELSINKI

Printing by topics in different languages in Turku



Viral Texts and Social Networks of Finnish Public Discourse in Newspapers and Journals 1771–1910

- Developing text reuse detection and identifying cross-border flows between languages (Finnish/Swedish)
- Virality of newspaper and journal discourse in nineteenth-century Finland: cultural rhizomes and social networks



Suomenmaan Yleisölle, suurimmalla kunnioituffella.

Muutamia vuosia tätä ennen kiihuhti laifensää-
tyinen Suomenmaan yleisö, Suomalaisen Kirjalli-
suuden Seuran lehoituffesta kansallistuntoa kunnioittamalla yksimielishyhdellä auttamaan muistopatsaan toimeen saantia maamme entiseen yä- ja yliopisto-kaupunkiin jalolle opettajalle, muinaisaikain tutkijalle ja isänmaan ystävälle **Henrik Gabriel Porthan**, sekä uskoi Kirjallisuuden seuralle huolta pitää sen hanffimifesta. Sentähden on Kirjallisuuden Seura toimittanut kuuwapatfaan, jonka kuwanweistäjä Carl Sjöstrand on Romaesfa muodostanut ja kunnikaallisen walinyajan esimies Münchenissä, Ferdinand von Miller, pronssi-wasfesta walaanut, pystytettäväksi Turussa sille paikalle, joka tälle kuwapatfaalle siellä oli walittu ja suosiollifesti annettu. Musfiwi ja muut paikalla tarpeellifet tehtävät ovat myös siihen määrään jo ennättäneet, että koko työ woipi walmiiksi joutua tulewan syyskuun alussa. Suurella ilomiehellä saa Kirjallisuuden Senra siis tämän kautta ilmoittaa, että muistopatsaan juhlallinen paljastaminen on woitu määrätä perjantailfisi, mainitun Syyskuun 9:tfi päiwätfi, jolloin juhlapuheita tulewat pitämään Kirjallisuuden Seuran kunnia-esimies kanslianeuwos Pönnrot suometsi ja profesfori Cagnæus ruotsifisi sekä Seuran sihteeri profesfori Ahlqwist lulemaan tätfi juhlatfi telemäksjä suomenkielisen runoelman. Pähempiä määrähkfiä juhlallisuudesta annetaan siinä juhlamenojen ofoituffessa, jonka Turkuun asetettu juhlafoomitea sitä warten tulee toimittamaan.

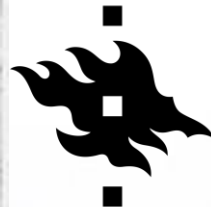
Kirjallisuuden Seura, jolla on aikomus aikansa antaa tarkka tilintefo tätä kuwapatfasta warten ko'otuiista rahoista, toiwoo luottamuffella että Suomen Yleisö, wäeffäästi koloutumalla tähän muistopatsaan paljastusjuhlaan, wieläfin ofoittaa suurella innolla suosiwanfa ensimmäistä kansallis-muistopatsasta Suomesfa. Helsingistä, 1 p:nä Elokuuta 1864.

Suomalaisen Kirjallisuuden Seuran puolesta:
Gab. Rein. M. Aflander.
Sihteerin sijaisena:
Carl Gust. Borg.

Pyhdetään maamme muuttakin suomalaisia sanomalehtiä palstoihinsa ottamaan tämä ilmoitus.

Example Cluster

Suometar 5 August 1864



RSITY OF HELSINKI

Suomenmaan Yleisölle,

suurimmalla kunnioituffella.

Muutamia vuosia tätä ennen kiiruhti laicensää-
tyinen Suomenmaan yleisö, Suomalaisen Kirjalli-
suuden Seuran lehoituffesta kansallistuntoa kunnioittamalla yksimielishyhdellä auttamaan muistopatsaan toimeen saantia maamme entiseen pää- ja yliopisto-kaupunkiin jalolle opettajalle, muinaisain tutkijalle ja isänmaan ystävälle **Henrik Gabriel Porthan**, sekä uskoi Kirjallisuuden seuralle huolta pitää sen hanffimisesta. Sentähden on Kirjallisuuden Seura toimittanut kuvapatsaan, jonka kuvanveistäjä Carl Sjöstrand on Romaassa muodostanut ja kunnikaallisen valinpojain esimies Münchenissä, Ferdinand von Miller, pronssi-wasfesta walanut, pystytettäväksi Turussa sille paikalle, jota tälle kuvapatsaalle siellä oli walittu ja suosiollisesti annettu. Musfivi ja muut paikalla tarpeelliset tehtävät ovat myös siihen määrään jo ennättäneet, että koko työ woipi walmiiksi joutua tulewan syyskuun alussa. Suurella ilomiehellä saa Kirjallisuuden Seura siis tämän kautta ilmoittaa, että muistopatsaan juhlassa wäljastaminen on woitu määrätä perjantaiksi, mainitun Syyskuun 9:ksi päiväksi, jolloin juhlapuheita tulewat pitämään Kirjallisuuden Seuran kunnia-esimies kanslianeuwoos Pönnrot suomelksi ja profesfori Cygnæus ruotsiksi sekä Seuran sihteeri profesfori Ahlqwist lukemaan tässi juhlassa tekemänsä suomenkielisen runoelman. Tähempä määrähffisiä juhlassa annetaan siinä juhlamenojen osoituffessa, jonka Turkuun asetettu juhlasomitea sitä warten tulee toimittamaan.

Kirjallisuuden Seura, jolla on aifomus aifonansa antaa tarkka tilintelo tätä kuvapatsasta warten ko'otuiata rahoiata, toiwoo luottamuffella että Suomen Yleisö, wäelkkäästi lofoutumalla tähän muistopatsaan wäljastusjuhlaan, wieläfin osoittaa suurella innolla suostuwan ensimmäistä kansallis-muistopatsasta Suomesa. Helsingistä, 1 p:nä Elokuuta 1864.

Suomalaisen Kirjallisuuden Seuran puolesta:

Gab. Klein.

M. Aflander.

Example Cluster

Suometar 5 August 1864

Reprinted six times:

Päivätär 6 August 1864

Mikkelin Wiikko-Sanomia 11 August 1864

Sanomia Turusta 12 August 1864

Tähti 12 August 1864

Hämäläinen 19 August 1864

Oulun Wiikko-Sanomia 20 August 1864



UNIVERSITY OF HELSINKI

TUTKIMUKSEN KIERTOKULKU

- Big data ja muilla uusilla “Digital Humanities” – lähestymistavoilla voidaan tuottaa lisäarvoa, mutta pääpaino oltava laadussa ja avoimuudessa ketjun kaikissa osissa:
 - Aineiston tuottamisen perustyö (esim. metadatan tuottaminen, OCR laatu)
 - Datan avaaminen
 - Datan siivoaminen tutkimuksessa
 - Tutkimusdatan avoimuus
 - Tutkimuksen uusiokäyttö / sen mahdollistaminen
- Kaikkien toimijoiden merkitys!



HELDIG.FI / DH VERKOSTO

STRATEGY NETWORK RESEARCH ▾ TEACHING PUBLICATIONS EVENTS ▾ INFRASTRUCTURE JOBS Q
CONTACT

- Kickoff tapahtuma 5.10.2016

- Avoimen tieteen filosofia ja historia konferenssi 31.11-1.12.2016

- Tervetuloa!



HELSINKI DIGITAL HUMANITIES

HELDIG: Helsinki Centre for Digital Humanities



UNIVERSITY OF HELSINKI