What Linked Data can tell about Geographical Trends in Finnish Fiction Literature

– Using the BookSampo Knowledge Graph in Digital Humanities

Telma Peura[1,2], Petri Leskinen[1] & Eero Hyvönen[1,2]

[1] Semantic Computing Research Group (SeCo),
Aalto University, Finland;
[2] Helsinki Centre for Digital Humanities (HELDIG),
University of Helsinki, Finland

Introducing BookSampo (1)

Part of the Sampo series

- building **a national ontology** (FinnONTO) and infrastructure for **cultural heritage** (CH) data that can be accessed through a semantic portal (Hyvönen, 2020)
- open source
- based on semantic web, residing on a SPARQL endpoint
- unifying and connecting heterogenous data silos of CH data
- CultureSampo (2009), BookSampo (2011), WarSampo (2015), BiographySampo (2018).
- Three components of model building
    1) LD creation and publishing model
    2) an interface with multiple perspectives
    3) two-step filtering cycle

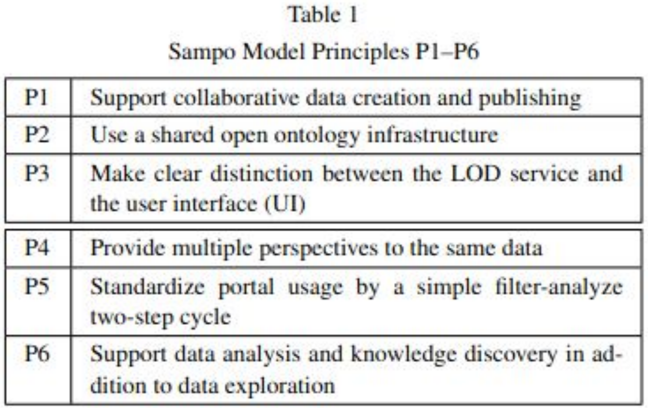


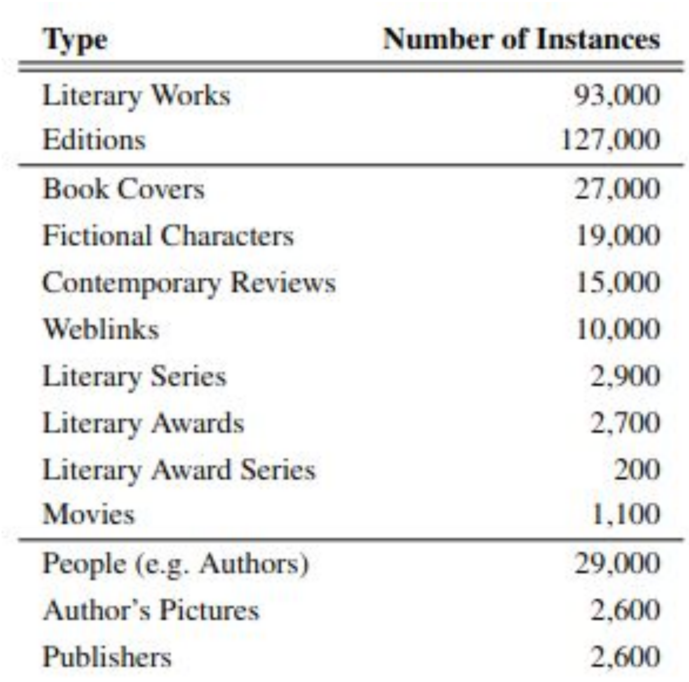
Table 1

Sampo Model Principles P1–P6

P1	Support collaborative data creation and publishing
P2	Use a shared open ontology infrastructure
P3	Make clear distinction between the LOD service and the user interface (UI)
P4	Provide multiple perspectives to the same data
P5	Standardize portal usage by a simple filter-analyze two-step cycle
P6	Support data analysis and knowledge discovery in addition to data exploration

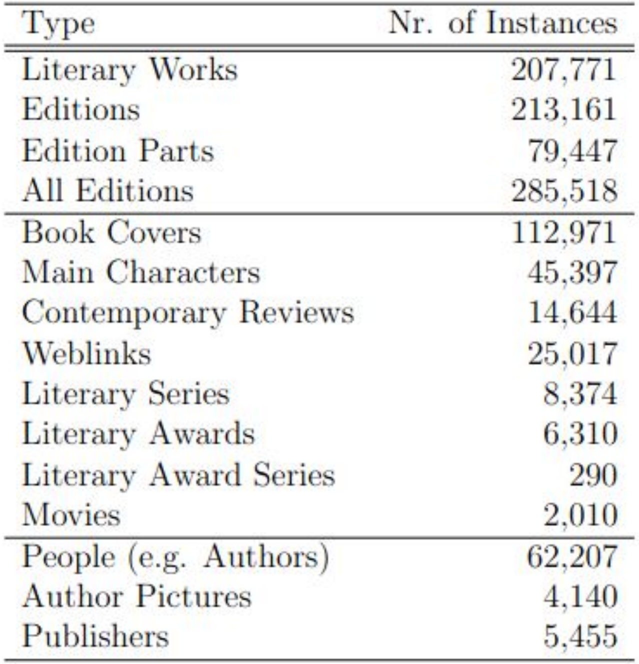
Hyvönen et al. (2022)

What does BookSampo contain?



Type	Number of Instances
Literary Works	93,000
Editions	127,000
Book Covers	27,000
Fictional Characters	19,000
Contemporary Reviews	15,000
Weblinks	10,000
Literary Series	2,900
Literary Awards	2,700
Literary Award Series	200
Movies	1,100
People (e.g. Authors)	29,000
Author's Pictures	2,600
Publishers	2,600

Mäkelä et al. (2013)



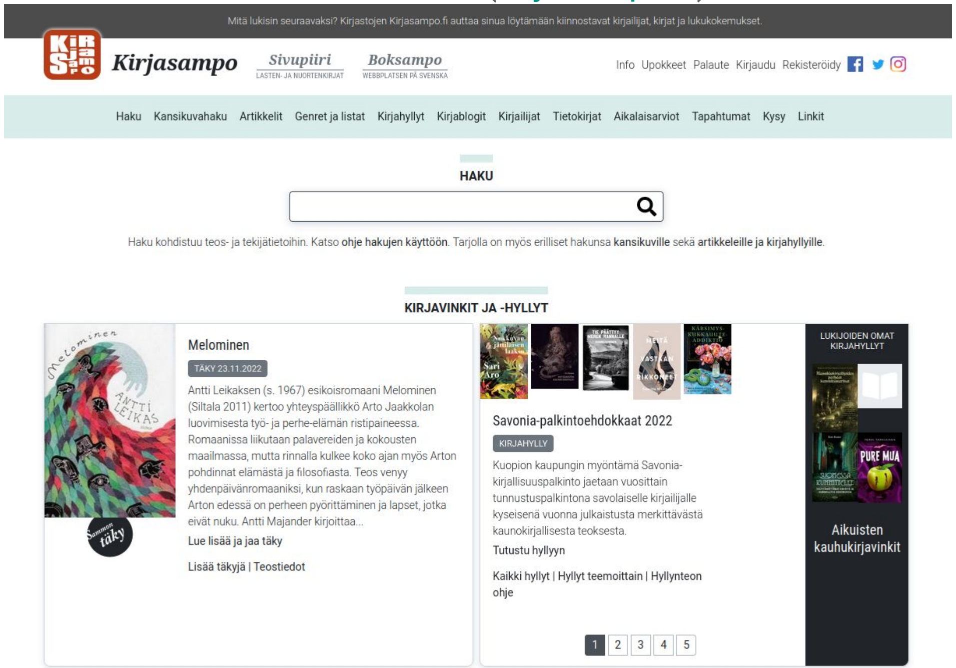
Type	Nr. of Instances
Literary Works	207,771
Editions	213,161
Edition Parts	79,447
All Editions	285,518
Book Covers	112,971
Main Characters	45,397
Contemporary Reviews	14,644
Weblinks	25,017
Literary Series	8,374
Literary Awards	6,310
Literary Award Series	290
Movies	2,010
People (e.g. Authors)	62,207
Author Pictures	4,140
Publishers	5,455

August 2022

Introducing BookSampo (2)

- a collaboration between FInnish public libraries and SW researchers
    - started in 2007
    - published 2011
- original data came from a dump from the Helsinki metropolitan area libraries
- new material automatically converted every night from BTJ Finland Ltd (Mäkelä et al. 2013)
- Kaunokki ontology for fiction literature
- Maintained by a team of librarians
    - improves the quality of the metadata
- contains practically all literature of Finnish public libraries (Mäkelä et al. 2011)
    - a good data source for DH research

Current interface (kirjasampo.fi):


Kirjasampo
Sivupiiri
LASTEN- JA NUORTENKIRJAT
Boksampo
WEBBPLATSEN PÅ SVENSKA
Info Upokkeet Palaute Kirjaudu Rekisteröidy
Haku Kansikuvahaku Artikkelit Genret ja listat Kirjahyllyt Kirjablogit Kirjailijat Tietokirjat Aikalaisarviot Tapahtumat Kysy Linkit
HAKU
KIRJAVINKIT JA -HYLLYT
TÄKY 23.11.2022
KIRJAHYLLY
LUKIJOIDEN OMAT KIRJAHYLLYT
1 2 3 4 5

Mitä lukisin seuraavaksi? Kirjastojen Kirjasampo.fi auttaa sinua löytämään kiinnostavat kirjailijat, kirjat ja lukukokemukset.

Haku kohdistuu teos- ja tekijätietoihin. Katso **ohje hakujen käyttöön**. Tarjolla on myös erilliset hakunsa **kansikuville** sekä **artikkeleille ja kirjahyllyille**.

Melominen

Antti Leikaksen (s. 1967) esikoisromaani Melominen (Siltala 2011) kertoo yhteyspäällikkö Arto Jaakkolan luovimisesta työ- ja perhe-elämän ristipaineessa. Romaanissa liikutaan palavereiden ja kokousten maailmassa, mutta rinnalla kulkee koko ajan myös Arton pohdinnat elämästä ja filosofiasta. Teos venyy yhdenpäivänromaaniksi, kun raskaan työpäivän jälkeen Arton edessä on perheen pyörittäminen ja lapset, jotka eivät nuku. Antti Majander kirjoittaa...

Lue lisää ja jaa täky

Lisää täkyjä | Teostiedot

Savonia-palkintoehdokkaat 2022

Kuopion kaupungin myöntämä Savonia-kirjallisuuspalkinto jaetaan vuosittain tunnustuspalkintona savolaiselle kirjailijalle kyseisenä vuonna julkaistusta merkittävästä kaunokirjallisesta teoksesta.

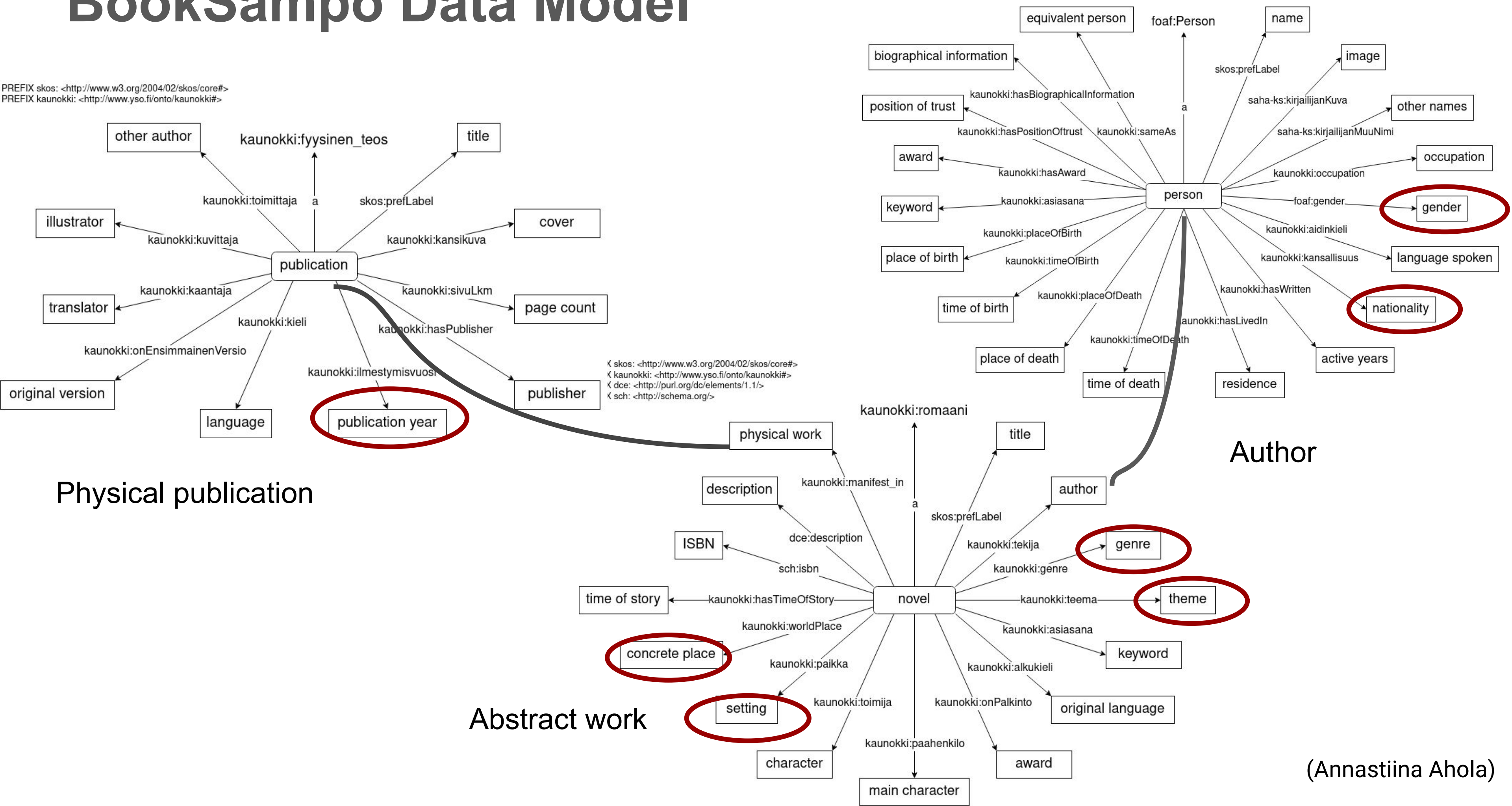
Tutustu hyllyyn

Kaikki hyllyt | Hyllyt teemoittain | Hyllynteon ohje

Aikuisten kauhukirjavinkit

BookSampo Data Model

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX kaunokki: <http://www.yso.fi/onto/kaunokki#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX saha-ks: <http://seco.tkk.fi/saha3/kirjasampo/>



PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX kaunokki: <http://www.yso.fi/onto/kaunokki#>

publication
a → kaunokki:fyysinen_teos
kaunokki:toimittaja → other author
skos:prefLabel → title
kaunokki:kuvittaja → illustrator
kaunokki:kansikuva → cover
kaunokki:kaantaja → translator
kaunokki:sivuLkm → page count
kaunokki:onEnsimmainenVersio → original version
kaunokki:kieli → language
kaunokki:ilmestymisvuosi → publication year
kaunokki:hasPublisher → publisher

Physical publication

person
a → foaf:Person
kaunokki:sameAs → equivalent person
skos:prefLabel → name
kaunokki:hasBiographicalInformation → biographical information
saha-ks:kirjailijanKuva → image
kaunokki:hasPositionOftrust → position of trust
saha-ks:kirjailijanMuuNimi → other names
kaunokki:hasAward → award
kaunokki:occupation → occupation
kaunokki:asiasana → keyword
foaf:gender → gender
kaunokki:placeOfBirth → place of birth
kaunokki:aidinkieli → language spoken
kaunokki:timeOfBirth → time of birth
kaunokki:kansallisuus → nationality
kaunokki:placeOfDeath → place of death
kaunokki:hasWritten → active years
kaunokki:timeOfDeath → time of death
kaunokki:hasLivedIn → residence

Author

skos: <http://www.w3.org/2004/02/skos/core#>
kaunokki: <http://www.yso.fi/onto/kaunokki#>
dce: <http://purl.org/dc/elements/1.1/>
sch: <http://schema.org/>

novel
a → kaunokki:romaani
kaunokki:manifest_in → physical work
skos:prefLabel → title
dce:description → description
kaunokki:tekija → author
sch:isbn → ISBN
kaunokki:genre → genre
kaunokki:hasTimeOfStory → time of story
kaunokki:teema → theme
kaunokki:worldPlace → concrete place
kaunokki:asiasana → keyword
kaunokki:paikka → setting
kaunokki:alkukieli → original language
kaunokki:toimija → character
kaunokki:onPalkinto → award
kaunokki:paahenkilo → main character

Abstract work

(Annastiina Ahola)

Case Study: Geographical diversity in novels

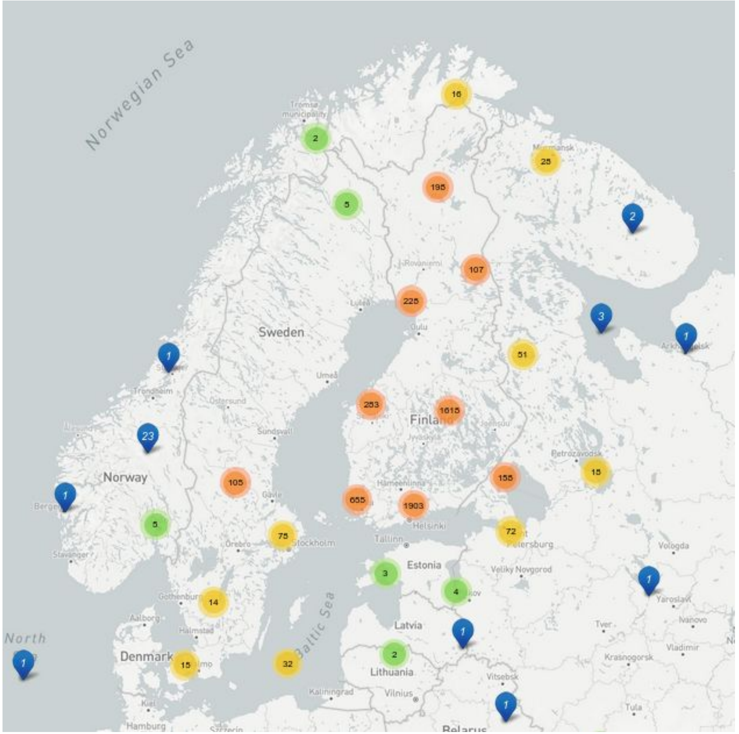
Distant Reading (Moretti, 2013)

- adopting a quantitative perspective to literature
- capturing large patterns in the literary production

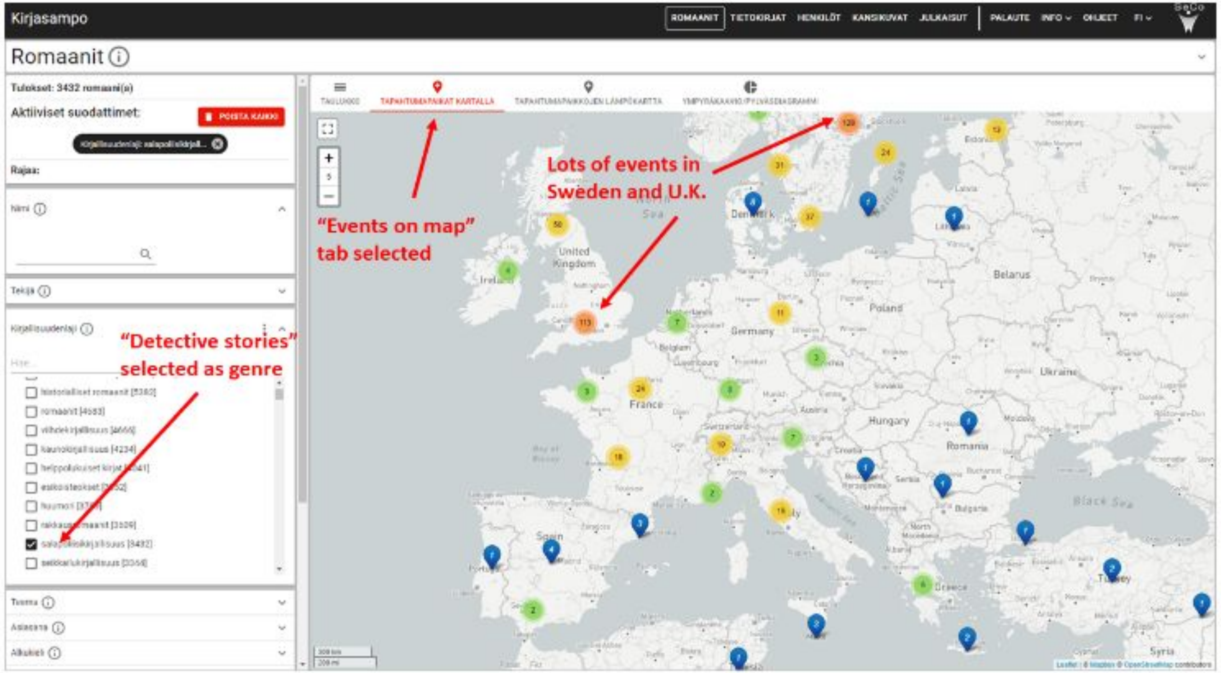
Literary Geography (e.g. Evans & Wilkens, 2018)

- in terms of author nationalities and languages
  - extra-textual
- in terms of novel settings
  - contents

How has the geography of novels published in Finnish developed in the past 50 years?



BookSampo for research



Tulokset: 3432 romaani(a)

Aktiiviset suodattimet:

Kirjallisuudenlaji: salapoliisikirjall...

Rajaa:

- □ historialliset romaanit [5380]
- □ romaanit [4583]
- □ viihdekirjallisuus [4666]
- □ kaunokirjallisuus [4234]
- □ helppolukuiset kirjat [4041]
- □ esikoisteokset [3952]
- □ huumori [3700]
- □ rakkausromaanit [3509]
- ☑ salapoliisikirjallisuus [3432]
- □ seikkailukirjallisuus [3356]

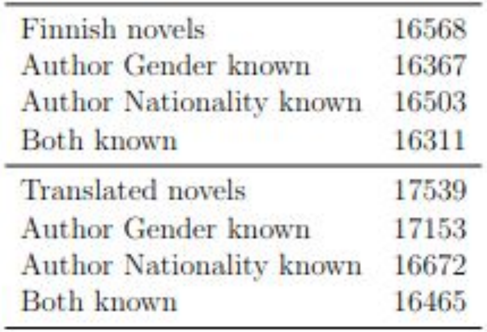
Example: detective stories on a map (Hyvönen et al., 2022)

Observations on data quality

Lack of coherence:

- over 600 URIs for genres, and over 400 themes
  - without a clear hierarchy
- “everything that helps in searching”
- e.g. How to group war literature?
  - 2nd WW ⇒ wars (events) ⇒ <u>societal events</u>
  - hostilities (warfare) ⇒ <u>societal events</u> ⇒ events
  - wartime ⇒ societal periods of time ⇒ time periods
  - the cold war ⇒ mutual action ⇒ action

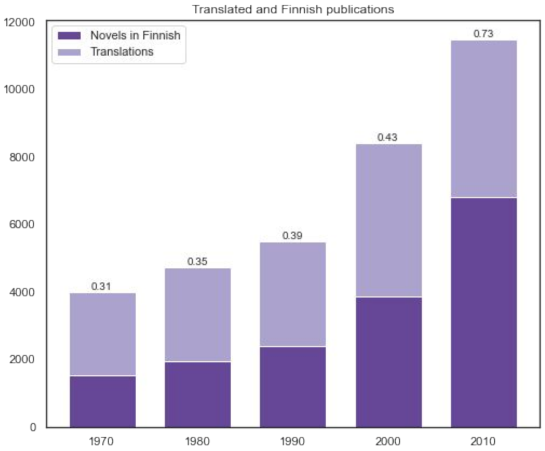
Data preprocessing and methods

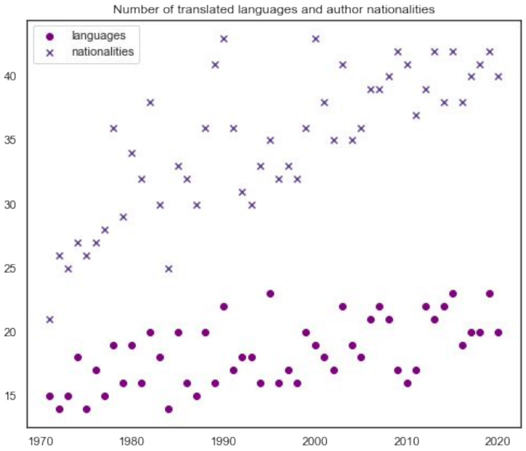


Finnish novels	16568
Author Gender known	16367
Author Nationality known	16503
Both known	16311
Translated novels	17539
Author Gender known	17153
Author Nationality known	16672
Both known	16465

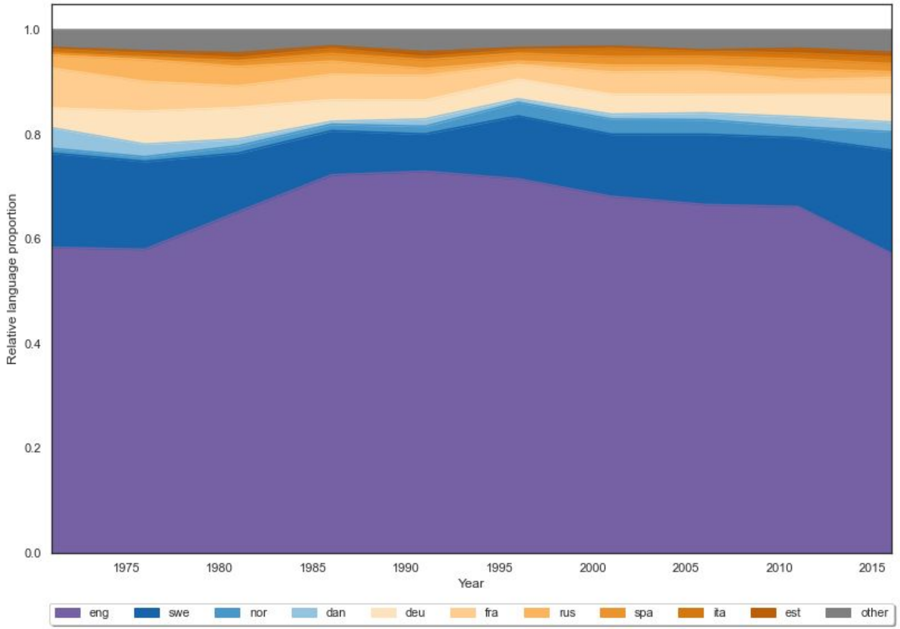
SPARQL-querying the BookSampo SW and constructing necessary subgraphs

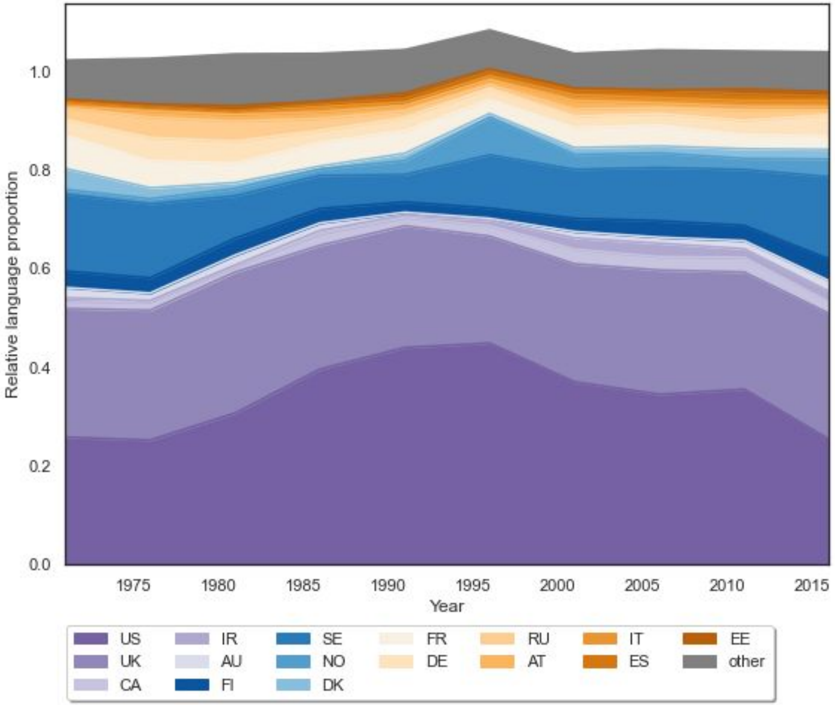
- constructing a sub-graph with the wanted attributes
  - dealing with annotation errors and irregularities to simplify the query process
  - e.g. Scottish, English, Welsh ⇒ Brits; Ålandians, Swedish-speaking Finns ⇒ Finns
  - Sami people, Finnish Swedes
  - removing duplicated nationalities (several URIs)
  - unification of the language codes
    - most based on lexvo, some irregularities
- Jupyter Notebooks
  - SPARQLWrapper, RDFLib
- Filtering on publication year (1971-2020)





Unbalanced representation of languages and nationalities





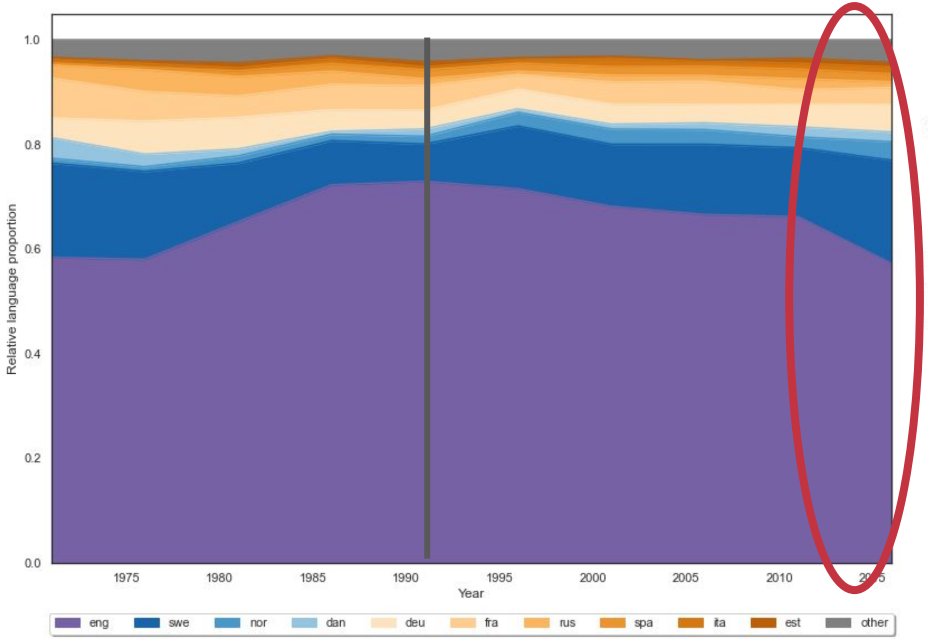
Aalto University
School of Science

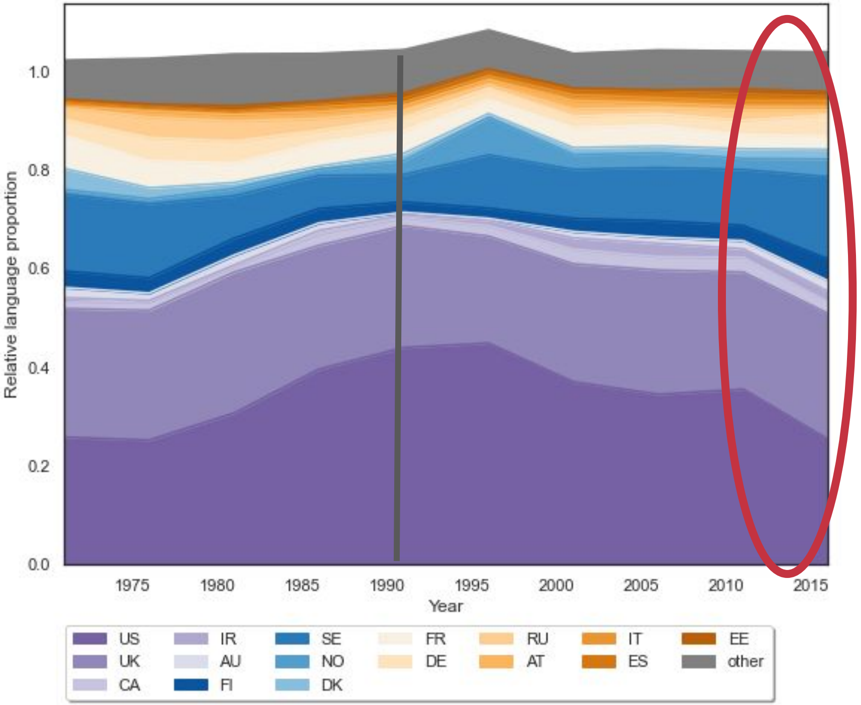
UNIVERSITY OF HELSINKI

SeCo

HELDIG
Helsinki Centre for Digital Humanities

Unbalanced representation of languages and nationalities





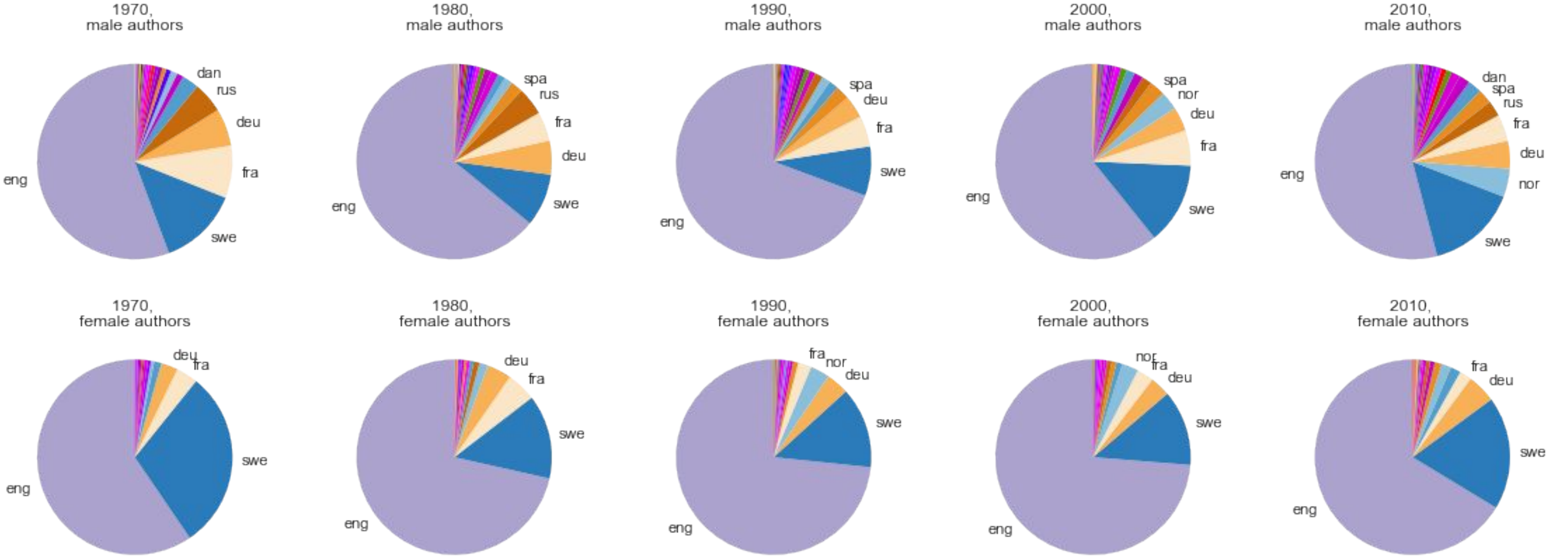
Aalto University
School of Science

UNIVERSITY OF HELSINKI

SeCo

HELDIG
Helsinki Centre for Digital Humanities

A Gender Bias in Language Diversity



Aalto University
School of Science

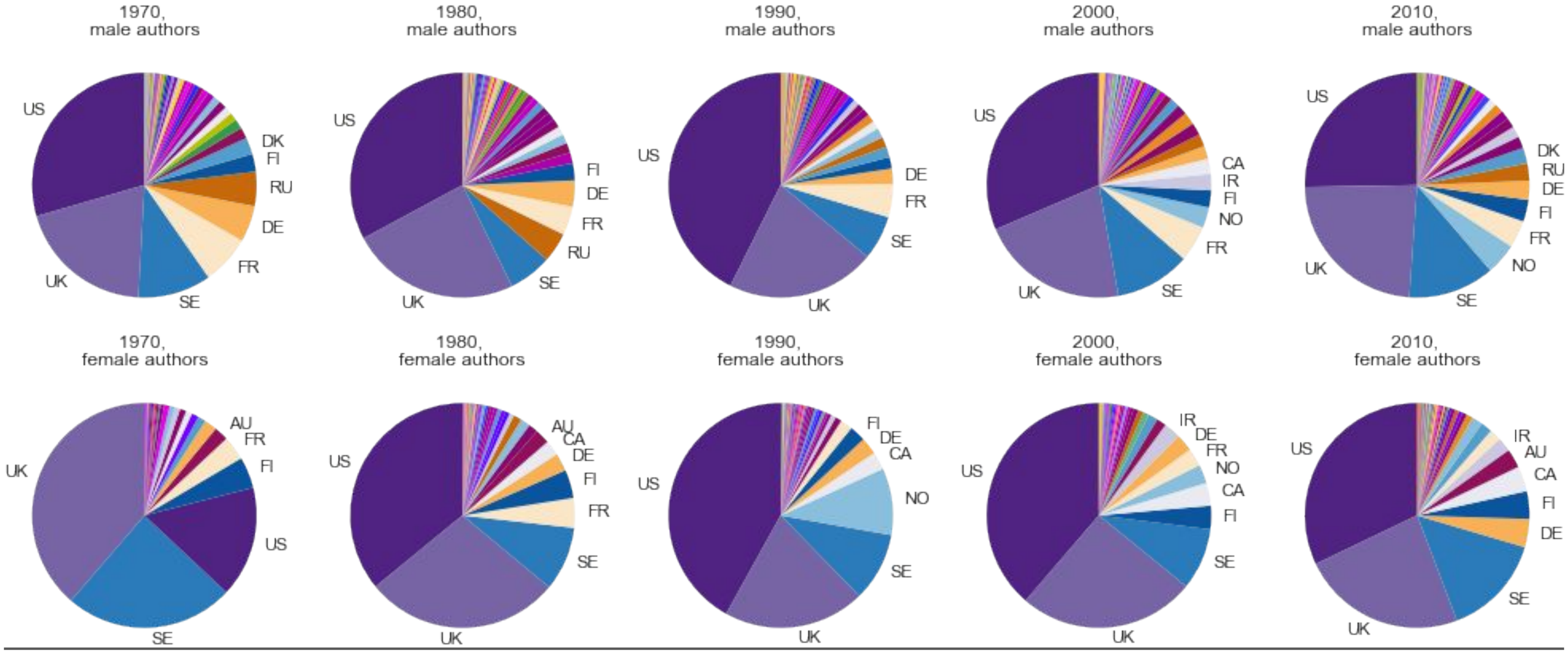
UNIVERSITY OF HELSINKI

KIRJASTOT.fi

SeCo

HELDIG
Helsinki Centre for Digital Humanities

A similar pattern in nationalities



Aalto University
School of Science

UNIVERSITY OF HELSINKI

KIRJASTOT.fi

Kirjasampo

SeCo

HELDIG
Helsinki Centre for Digital Humanities

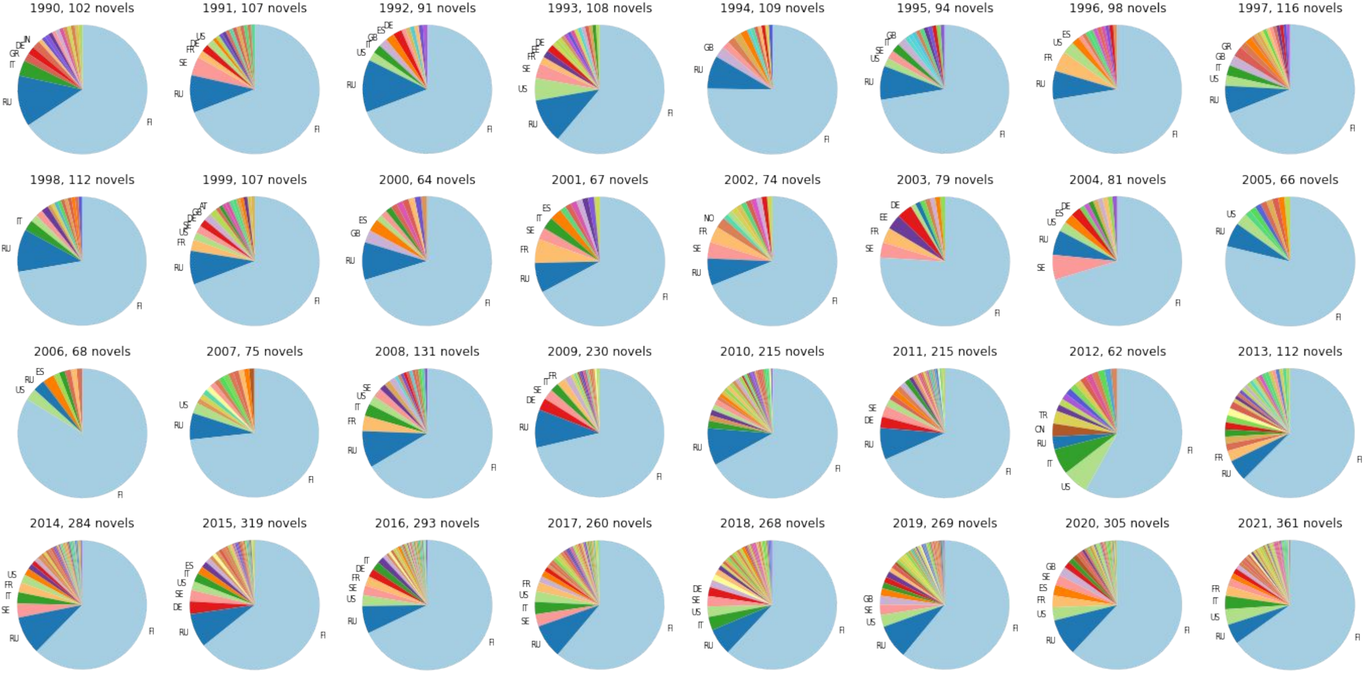
Geography *in* the novels written by Finnish authors

Only about ⅓ of yearly publications is annotated for the geographical setting

Grouping place names by country and year

Russia the 2nd most popular country in almost every year

The proportion of Finland slightly decreasing



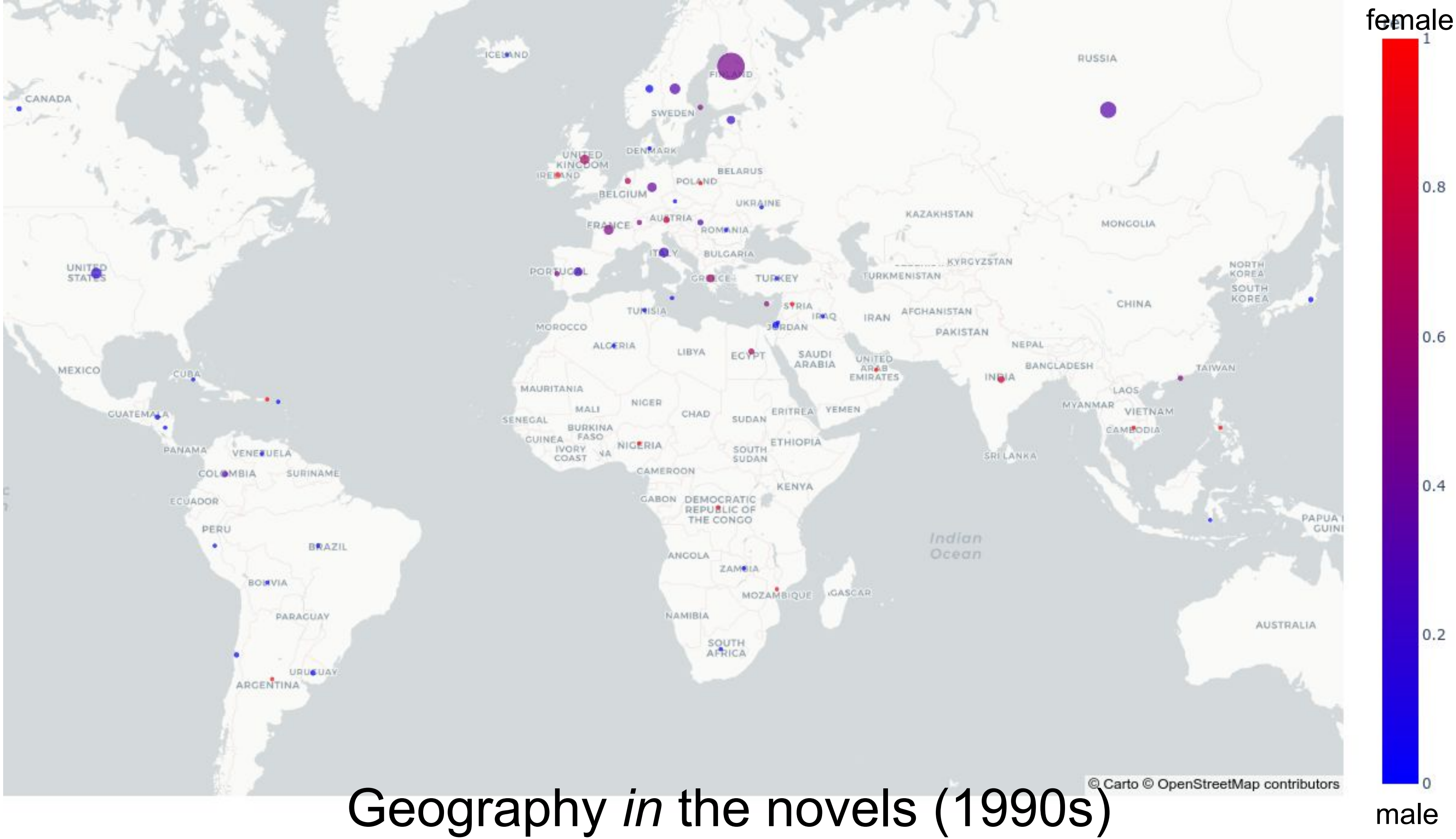
Aalto University
School of Science

UNIVERSITY OF HELSINKI

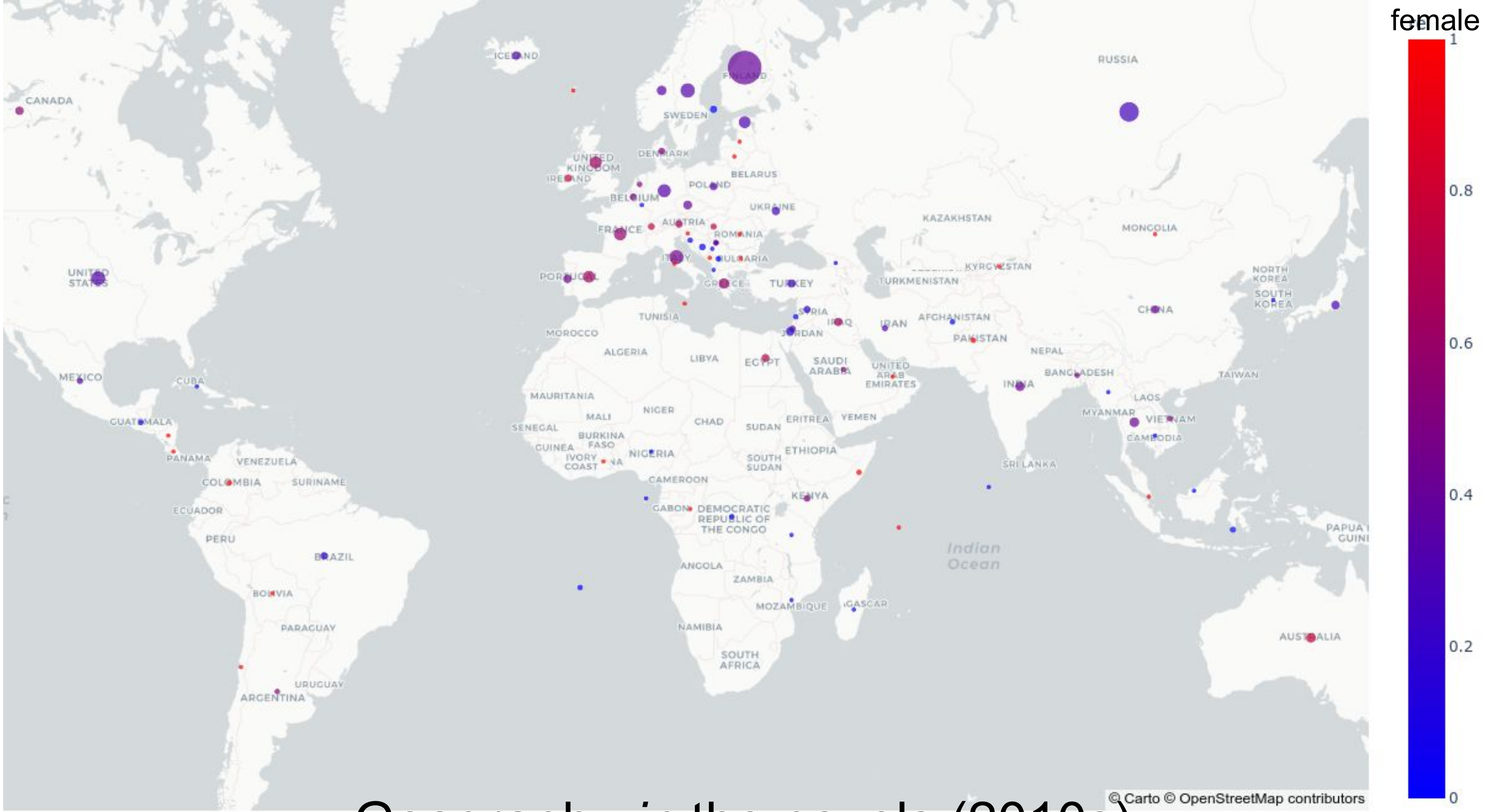
KIRJASTOT.fi

SeCo

HELDIG
Helsinki Centre for Digital Humanities



Geography *in* the novels (1990s)



Geography *in* the novels (2010s)

What do authors write about when they write about Russia?

Clustering based on genre and theme keywords

1)	biographical, historical war novels, memories	48/96/144
2)	war literature, societal novels	5/110/117
3)	personal development, relationships, moving, psychological	42/30/72
4)	thrillers, crime fiction, espionage	6/52/61
5)	women, historical romance	23/5/28

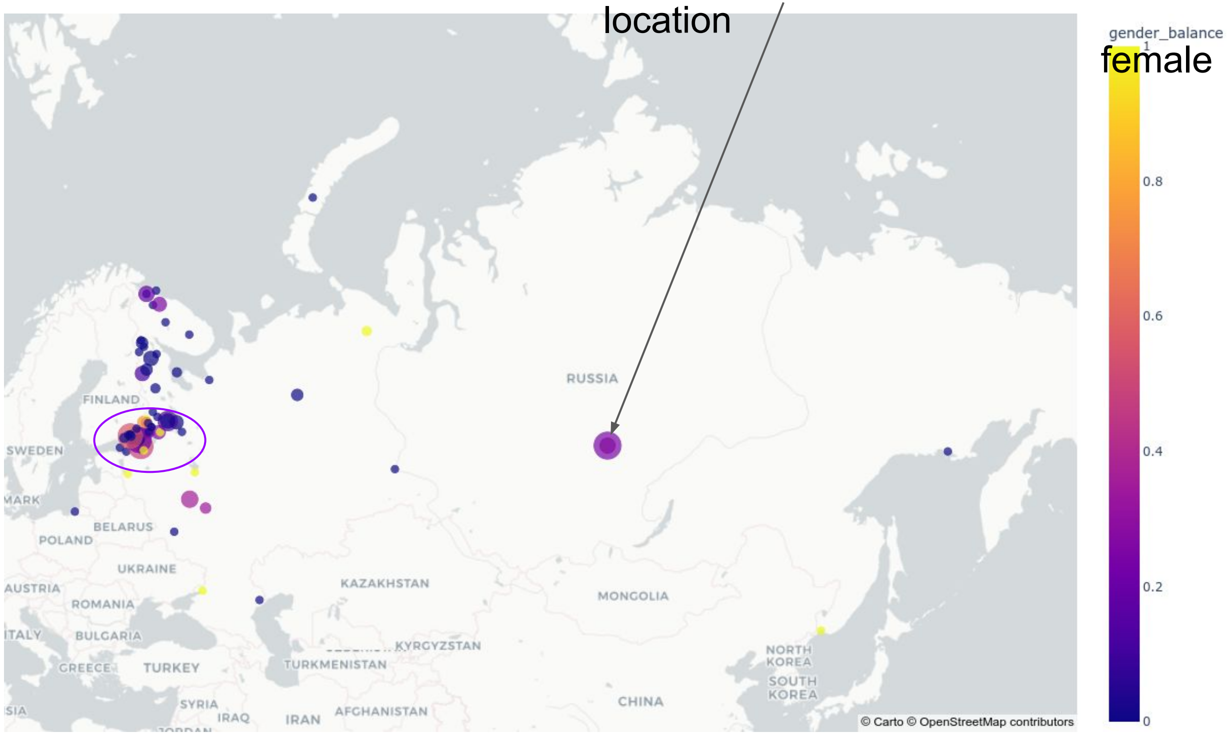
⇒ Female and male authors have different perspectives on Russia in the context of novels

On the map

Novels about Russia actually tell about Finland.

Focus on Karelia.

Confirms the previous claim (Hallila & Hägg, 2007) that Finns are interested in historical, national topics.



The country of Russia as the annotated place, not a specific location

female

male

Aalto University School of Science

UNIVERSITY OF HELSINKI

KIRJASTOT.fi

SeCo

HELDIG Helsinki Centre for Digital Humanities

Conclusions and Future Directions

The BookSampo Knowledge Graph can be used in DH research, but it requires careful preprocessing and compromising on detailed/specific claims.

Novels seem to **become more diverse** both in terms of extra-textual indices and textual settings.

Bias in the annotations?

- Geographical setting annotated only in a third of the novels
- **When is the geographical setting seen important?**
  - E.g. genre literature
  - **When should it be annotated?** How fine-grained?

Insights from this exploratory research can help to improve the annotation practices further.

Collaboration between DH and SW researchers, and librarians is needed.

References

Evans, E., & Wilkens, M. (2018). Nation, ethnicity, and the geography of british fiction, 1880-1940. *Journal of Cultural Analytics*, *3*(2), 11037.

Hallila, M., & Hägg, S. (2007). history and historiography in contemporary Finnish Novel. *AVAIN-Kirjallisuudentutkimuksen aikakauslehti*, (4), 74-80.

Hyvönen, E. (2020). "Sampo" Model and Semantic Portals for Digital Humanities on the Semantic Web. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)*. CEUR-WS. org.

Hyvönen, E., Ahola, A., & Ikkala, E. (2022). BookSampo Fiction Literature Knowledge Graph Revisited: Building a Faceted Search Interface with Seamlessly Integrated Data-Analytic Tools. In *International Conference on Theory and Practice of Digital Libraries* (pp. 506-511). Springer, Cham.

Moretti, F. (2013). *Distant reading*. Verso Books.

Mäkelä, E., Hypén, K., & Hyvönen, E. (2013). Fiction literature as linked open data—The BookSampo dataset. *Semantic Web*, *4*(3), 299-306.

Thank you! Questions?