

1800-luvun kirjeenvaihdon laajat ja syvät verkostot -tutkimushankkeen aloitusseminaari 17.9.2021

Kirjesampo – historialliset kirjeet semanttisessa webissä: työkalu ja sen soveltaminen Republic of Letters -aineistoihin 1500–1800

Prof. Eero Hyvönen, Director

Aalto University & University of Helsinki (HELDIG)

Semantic Computing Research Group (SeCo)

<http://seco.cs.aalto.fi/u/eahyvone/>

Contents

1. Background (2014-2018):

- Republic of Letters Research in Aalto & HELDIG in 2014-2021

2. LetterSampo (2019-):

- Application to Republic of Letters (RofL) datasets 1500-1800
- Video demonstration

3. Network analyses: (2020-)

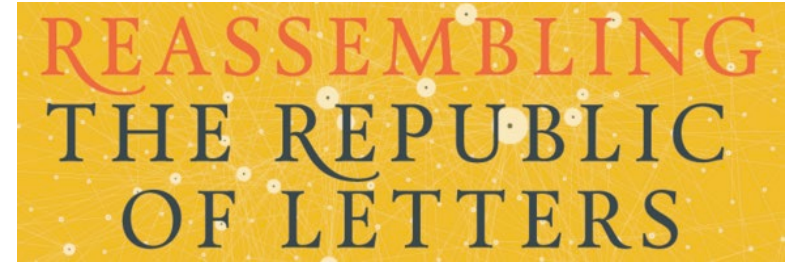
- Comparing RofL with modern communications networks





1. Background (2014-2018) – Research on Republic of Letters

Background (2014-2018)



- Reassembling the Republic of Letters 1500-1800, 2014-2018
 - EU COST Action lead by the University of Oxford
 - Over 30 countries involved
 - <http://www.republicofletters.net/>
- Mapping the Republic of Letters project
 - Collaborations & researcher exchange with Stanford University
- Cultures of Knowledge Project at Oxford, Phase III, 2015-2017
 - Developing the Early Modern Letters Online (EMLO) service
 - Aalto as subcontractor, focus on a Linked Data model
 - Funded by Mellon Foundation, USA
 - <http://www.culturesofknowledge.org/>



Stanford
University





REASSEMBLING
THE REPUBLIC OF LETTERS
IN THE DIGITAL AGE

Standards, Systems, Scholarship

Edited by Howard Hotson and Thomas Wallnig



Göttingen University Press

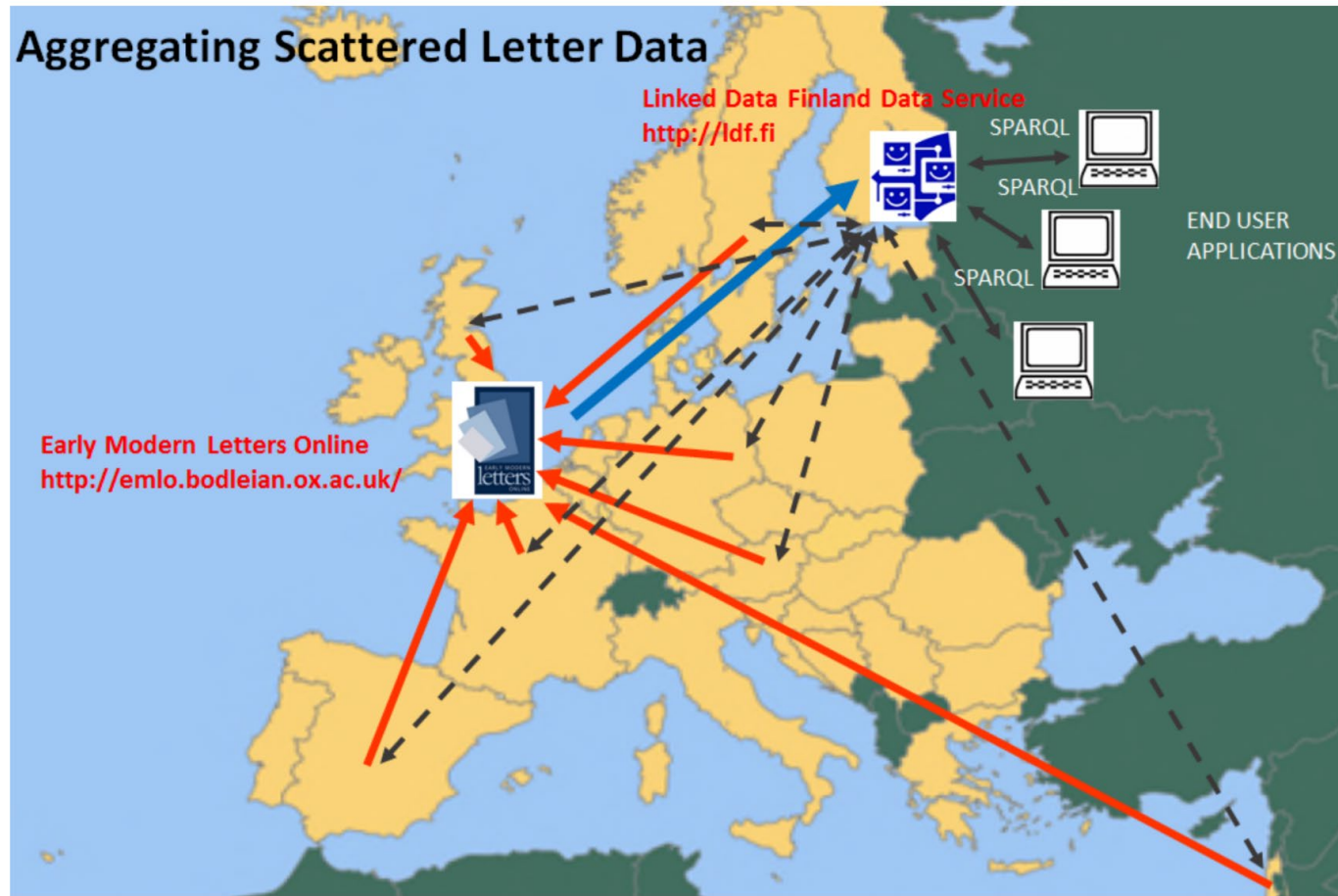


Free online version available:
<https://doi.org/10.17875/gup2019-1146>

Aalto Research Team:
Eero Hyvönen (leader)
Eetu Mäkelä
Jouni Tuominen



Our Contribution: Using Linked Data & Ontologies



Results:

- EMLO Data as Linked Data on the Linked Data Finland platform
- Tools for creating and harmonizing the data
- First data-analyses tested

Challenges:

- Interoperability: aligning & harmonizing the data
- Incompleteness & uncertainty of historical data
- Modelling prosopographical data of correspondents
- The EMLO data is not openly available for research!

Fig. 1. Overview of the Linked Data approach in creating and aggregating distributed epistolary data.

Why Linked Open Data

- Enriching everybody's data collaboratively from separate silos
 - Everybody wins by collaboration!
- Creating Findable, Accessible, Interoperable, Re-usable data
 - The value of data increases!
- Creating more intelligent applications for the public, curators, and researchers
 - The machine “understands” linked data!



FAIR



<https://www.go-fair.org/fair-principles/>



LetterSampo Project in Aalto/HELDIG 2018-2021

- Using the “Sampo model” for studying epistolary linked data
 - Petri Leskinen joins the Aalto team
 - Develop the LetterSampo-portal that can be used without programming skills
- Applying Network Analysis methods to study the data
 - New collaborations with Complex Systems group at Aalto
 - Prof. Mikko Kivelä, Javier Urena-Carrion
- International collaborations for getting new open datasets for research
 - Prof. Charles van den Heuvel
 - Huygens Institute, Netherlands & University of Amsterdam
 - Provided the **CKCC corpus (20 000+ letters)**
 - Dr. Stefan Dumont
 - Berlin-Brandenburg Academy of Sciences and Humanities, Germany
 - Provided the **correspSearch dataset (150 000+ letters)**
- CoCo project starting in 2021 of Finnish epistolary data (Academy of Finland)
 - Dr. Ilona Pikkanen (SKS, leader), Dr. Mikko Koho (HELDIG), Dr. Jouni Tuominen (Aalto)

<https://seco.cs.aalto.fi/projects/rrl/>



Semantic Computing Research Group (SeCo)
Making computers and the web more intelligent and interoperable!



HOME PAGE

MISSION

NEWS

PROJECTS

EVENTS

PEOPLE

PARTNERS

CAREERS

TEACHING

CONTACT

PUBLICATIONS

APPLICATIONS

ONTOLOGIES

LINKED DATA

INFRA FOR DATA

SERVICES

TOOLS

SOFTWARE

DOWNLOAD

Wikipedia: *Semantic web* *suomeksi*

Google this site

Search!

REASSEMBLING
THE REPUBLIC
OF LETTERS

From Reassembling the Republic of Letters to LetterSampo - Letters on the Semantic Web

Background - Republic of Letters on the Semantic Web

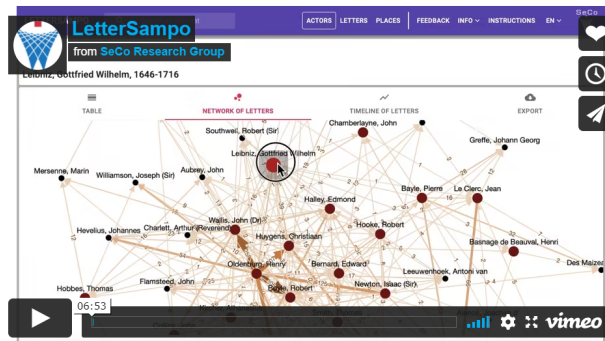
SeCo participated in the Digital Humanities EU COST action [Reassembling the Republic of Letters 1500-1800](#), 2014-2018, led by prof. Howard Hotson (University of Oxford) and Dr. Thomas Wallnig (Universität Wien). Eero Hyvönen (Aalto University and University of Helsinki (HELDIG)) led the [Work Group 2 People and Networks](#) in the initiative involving over 30 countries. Related to this topic, there was also the project [Cultures of Knowledge](#), phase III (2015-2017), at Oxford where SeCo collaborated with Oxford and Stanford University with the goal of designing a Linked (Open) Data infrastructure, linked data service, and tooling for the underlying humanist scholarly community, based on the data harvested in the [Early Modern Letters Online \(EMLO\) database](#) and portal.

Next Step: LetterSampo - Letters on the Semantic Web

Inspired by the international collaborations on the Republic of Letters and the Bio CRM data model, SeCo group moved on to designing and implementing a prototype framework and semantic portal demonstrator that supports Digital Humanities research on epistolary data. For this purpose, a new LetterSampo project was started at Aalto university and HELDIG. Our goal is to develop a prototype called "LetterSampo - Letters on the Semantic Web", based on the ["Sampo" model](#) and ["Sampo-UJ" framework](#). LetterSampo is a new member in the Sampo series of semantic portals, a sister to, e.g., the [Mapping Manuscript Migrations service](#) of medieval and Renaissance manuscripts, and [BiographySampo](#) for biographical and prosopographical research. LetterSampo can be used for analysing epistolary data in different data repositories in different countries, or for aggregated international datasets, such as Republic of Letters Data. LetterSampo project is an independent new initiative in SeCo group funded by Aalto and the HELDIG Centre.

We welcome collaborations with researchers in humanities regarding epistolary data, and have started to collaborate with the network scientists prof. Mikko Kivelä and Javier Urena Carrion of Aalto University, with the goal of applying their temporal network analysis tooling in epistolary data. Collaborations with the Dutch researchers developing the [ePistolarium system](#) and German researchers working on the [correspSearch system](#) are underway.

For a presentation and demonstration of the "LetterSampo - Historical Letters on the Semantic Web" system, see the video below:



[From Republic of Letters 1500-1800 to LetterSampo from SeCo Research Group on Vimeo.](#)

Contact Persons

[Prof. Eero Hyvönen](#), Aalto University and University of Helsinki

[Dr. Jouni Tuominen](#), Aalto University and University of Helsinki

[Petri Leskinen](#), Aalto University

A painting of a historical interior scene, possibly a study or a library. Several people are seated at a long table, engaged in reading or writing. The room is filled with bookshelves and various objects, creating a sense of a busy, intellectual environment. The lighting is warm, highlighting the figures and the details of the room.

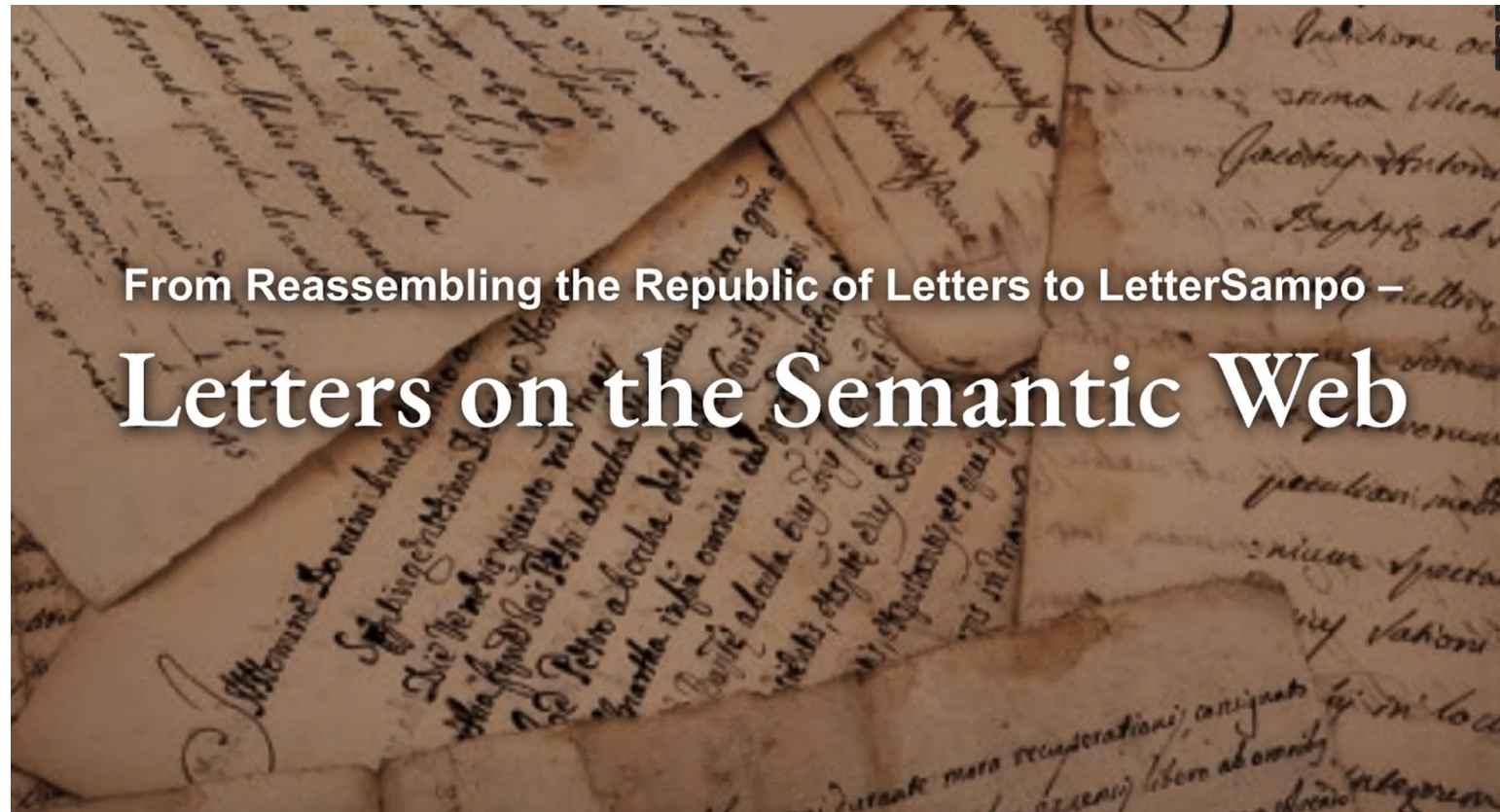
2. LetterSampo (2019-)

– Historical Letters on the Semantic Web



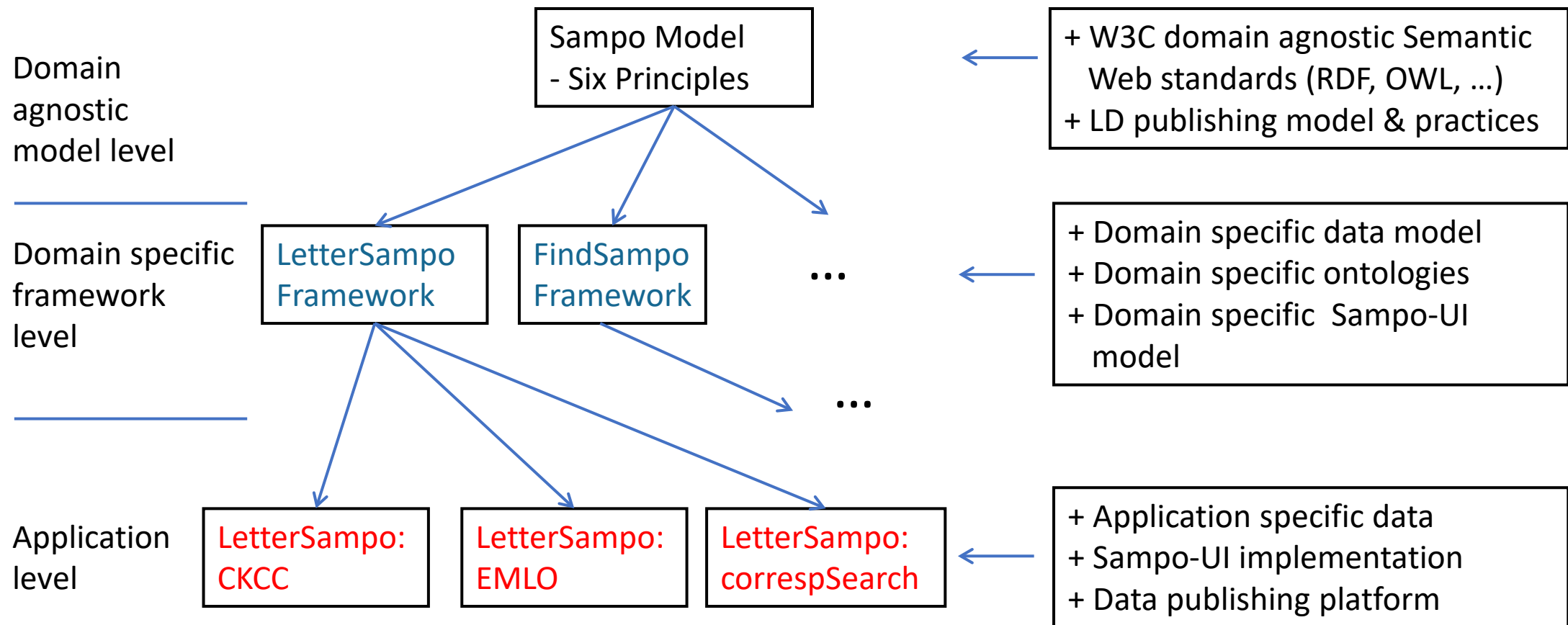
Video illustrating how the LetterSampo Framework is used based on the EMLO data about 160 000+ letters of the Republic of Letters:

<https://vimeo.com/461293952>



Sampo Model -> LetterSampo Framework

Reusing the same general framework/software for different letter datasets!



LetterSampo – Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data of the Republic of Letters

Eero Hyvönen^{1,2}[0000–0003–1695–5840], Petri Leskinen²[0000–0003–2327–6942], and
Jouni Tuominen^{2,3}[0000–0003–4789–5676]

¹ Semantic Computing Research Group (SeCo), Aalto University, Finland

² HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
<http://seco.cs.aalto.fi>, <http://heldig.fi>



<https://seco.cs.aalto.fi/publications/2020/hyvonen-et-al-lettersampo-2020.pdf>

Abstract. Epistolary data about historical letters is typically distributed in different archives depending on where the letters were sent to and received, and the data are represented using local heterogeneous data models and different natural languages. To study such letter data from a global perspective, the heterogeneous, distributed data in siloes need to be aggregated and harmonized into larger services where local metadata can enrich each other to complete missing information in the local collections. This paper presents a framework, based on the so-called Sampo model, for representing, publishing, and using epistolary data as Linked Data on the Semantic Web for Digital Humanities research. To test the framework we discuss its merits and challenges by presenting a demonstrational implementation called LetterSampo and by applying it to the epistolary CKCC dataset of ca. 20 000 letters of the Huygens Institute related to the Republic of Letters (RofL). The framework has also been applied to the correspSearch dataset of the Berlin-Brandenburg Academy of Sciences and Humanities, and was originally developed using the Early Modern Letters Online (EMLO) data of the University of Oxford. The novelty of the LetterSampo framework is based on using Linked Data and SPARQL querying integrated with faceted search for filtering out data of interest in flexible ways from multiple perspectives. The filtered data can be visualized and analyzed by seamlessly integrated data analytic tools for DH research on the portal without programming skills. For customized analyses, the underlying SPARQL endpoint can be used with modest knowledge about SPARQL querying and programming.

Actor ⓘ

Descartes, René, 1596-1650



TABLE



LETTERS



NETWORK OF LETTERS

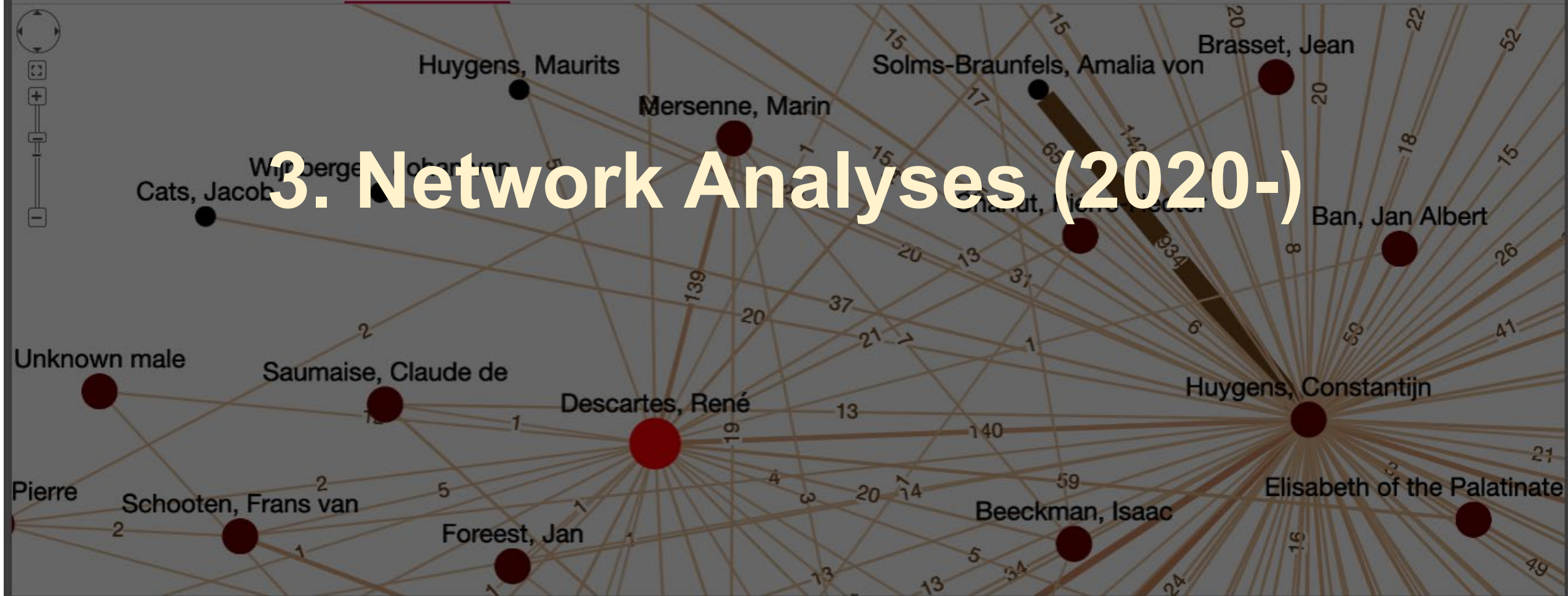


TIMELINE OF LETTERS



EXPORT

3. Network Analyses (2020-)



Actor ⓘ

Descartes, René, 1596-1650



TABLE



LETTERS



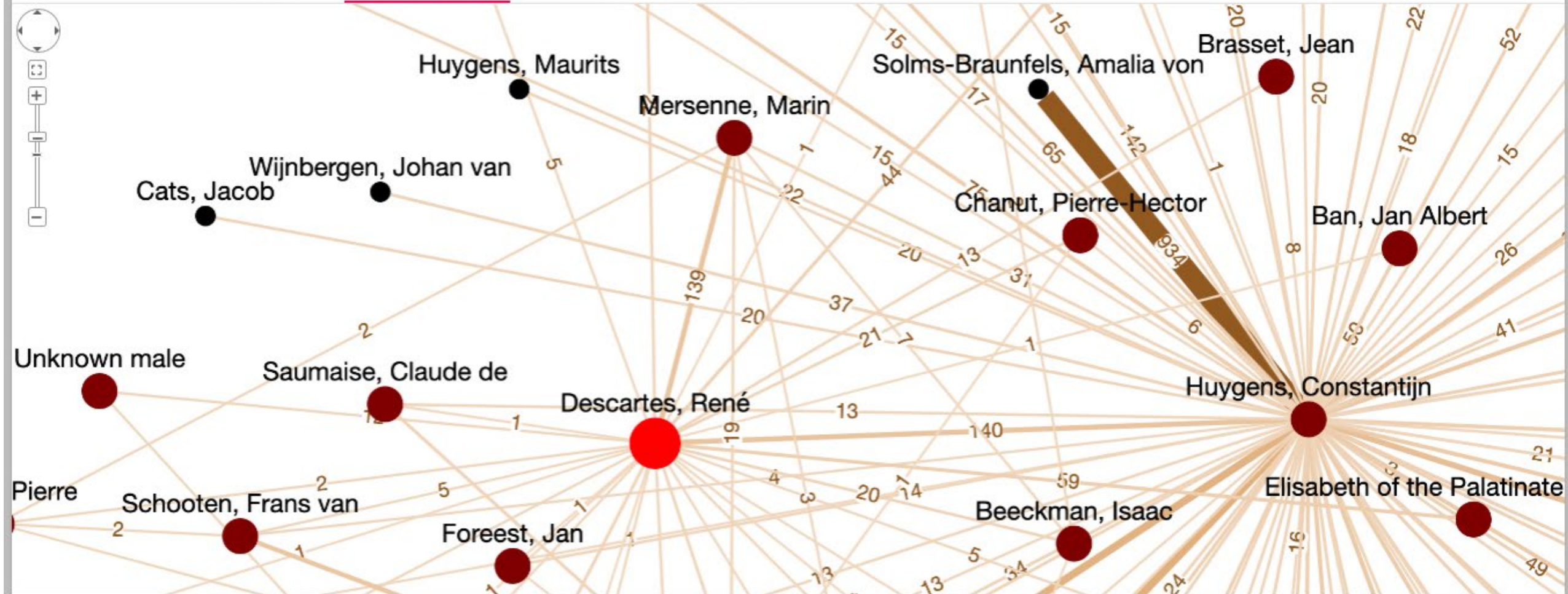
NETWORK OF LETTERS



TIMELINE OF LETTERS



EXPORT



Communication Now and Then: Analyzing the Republic of Letters as a Communication Network

Javier Ureña-Carrion¹[0000-0003-2763-9747], Petri Leskinen²[0000-0003-2327-6942], Jouni Tuominen^{2,3}[0000-0003-4789-5676], Charles van den Heuvel⁴[0000-0001-9638-400X], Eero Hyvönen^{2,3}[0000-0003-1695-5840], and Mikko Kivelä¹[0000-0003-2049-1954]

¹ Complex Systems Group, Aalto University, Finland

² Semantic Computing Research Group (SeCo), Aalto University, Finland
<https://seco.cs.aalto.fi>

³ HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
<https://heldig.fi>

⁴ Huygens Institute for the History of the Netherlands: Amsterdam, NL

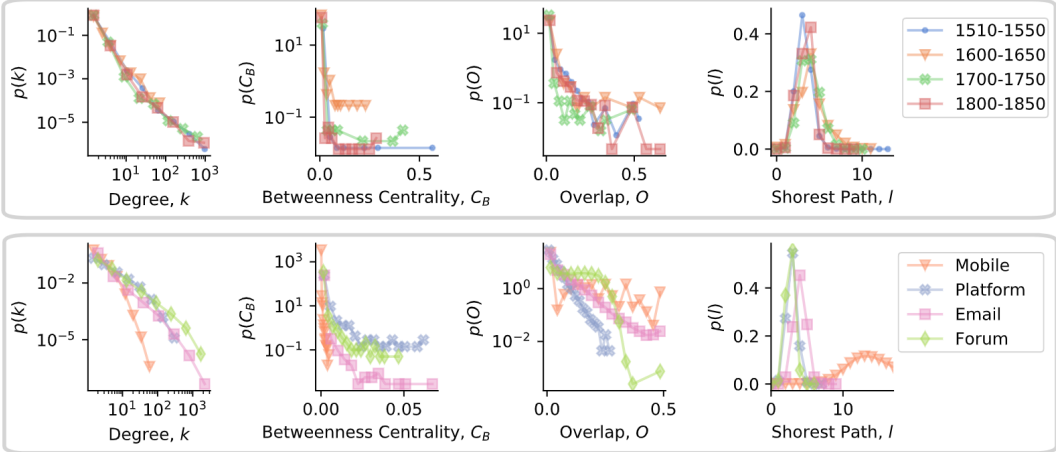
Abstract. Huge advances in understanding patterns of human communication, and the underlying social networks where it takes place, have been made recently using massive automatically recorded data sets from digital communication, such as emails and phone calls. However, it is not clear to what extent these results on human behaviour are artefacts of contemporary communication technology and culture and if the fundamental patterns in communication have changed over the history. This paper presents an analysis of historical epistolary metadata with the aim of comparing the underlying historical communication patterns with those of contemporary communication. Our work uses a new epistolary dataset containing metadata on hundreds of thousands of letters sent between the 16th and 19th centuries. The analyses indicate striking resemblances between contemporary and epistolary communication network patterns, including dyadic interactions and ego-level behaviour. Despite these positive findings, certain aspects of the letter datasets are insufficient to corroborate other similarities or differences for these communication networks.

Dataset	Content	Observation Period	Source
Platform	Wall-posting on Facebook in New Orleans, where a directed link is created if a user posts on another user's wall. Data obtained from crawling public profiles in January 2009, each link tracks activity for one year after the first communication event. Total of ~ 41K nodes, ~ 183.5K edges and 854.6 ~K wall posts.	1 year	[50]
Email	Dataset of emails between employees of Enron Corporation, obtained as part of an investigation into the Corporation's collapse. Sampling might focus on 150 users. Total of ~ 20K users, 297.5K edges and ~ 1.1M emails.	1998-2004	[51]
Forum	Activity on a movie forum, where a contact is created if a user comments on a thread posted by another user. Total of ~ 6K users, ~ 138.1K edges and ~ 1.4M thread comments.	7 years	[52]
Mobile	Call Detail Records of mobile phone calls from an operator in a European country in 2007, where the operator had approximately 20% market share. We used a reduced version of the full dataset, where we focused on a region of the country. The small version of the dataset has ~ 174K nodes, ~ 190K edges and ~ 6.8M phone calls.	4 months	[3]

vs. correspSearch letters

TABLE 1: Datasets analyzed and compared in this paper

“The analyses indicate striking resemblances between contemporary and epistolary communication network patterns, including dyadic interactions and ego-level behaviour.”



Conclusions

- Lots of historical epistolary data has been collected
- Letters are a window to studying the history and networks of people
 - Using metadata only can be useful without letter texts
- Epistolary should be opened also as data for DH researchers
- Data harvesting and aggregation from different data owners needed
 - A nice domain for transnational studies
- Challenges related to incompleteness and uncertainty of data
 - Data literacy is needed!
- Great that Finnish data & analyses will be created in CoCo!



More Information – Questions?

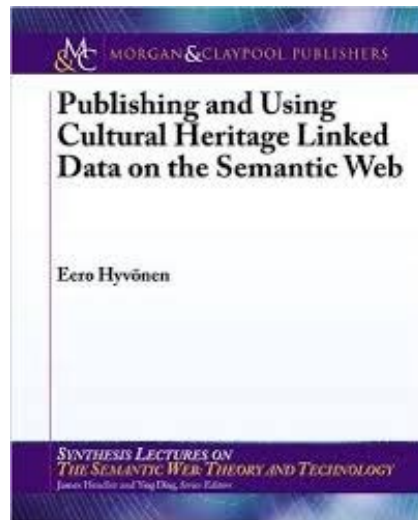
From Republic of Letters to Lettersampo

<https://seco.cs.aalto.fi/projects/rrl/>

Sampo Model & Portal Series:

<https://seco.cs.aalto.fi/applications/sampo/>

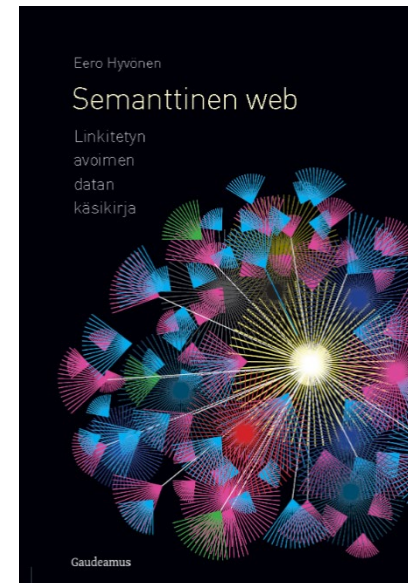
Text books about Linked Data, CH, and SW:



In English

2012

<https://www.amazon.com/Publishing-Cultural-Heritage-Synthesis-Technology/dp/1608459977>



In Finnish

2018

<https://www.gaudamus.fi/semanttinen-web/>