

NameSampo

A Linked Open Data Infrastructure and Workbench for Toponomastic Research

October 23, 2018

HELDIG Digital Humanities Summit 2018

Esko Ikkala (HELDIG, Aalto), Jouni Tuominen (HELDIG), Eero Hyvönen (HELDIG, Aalto)

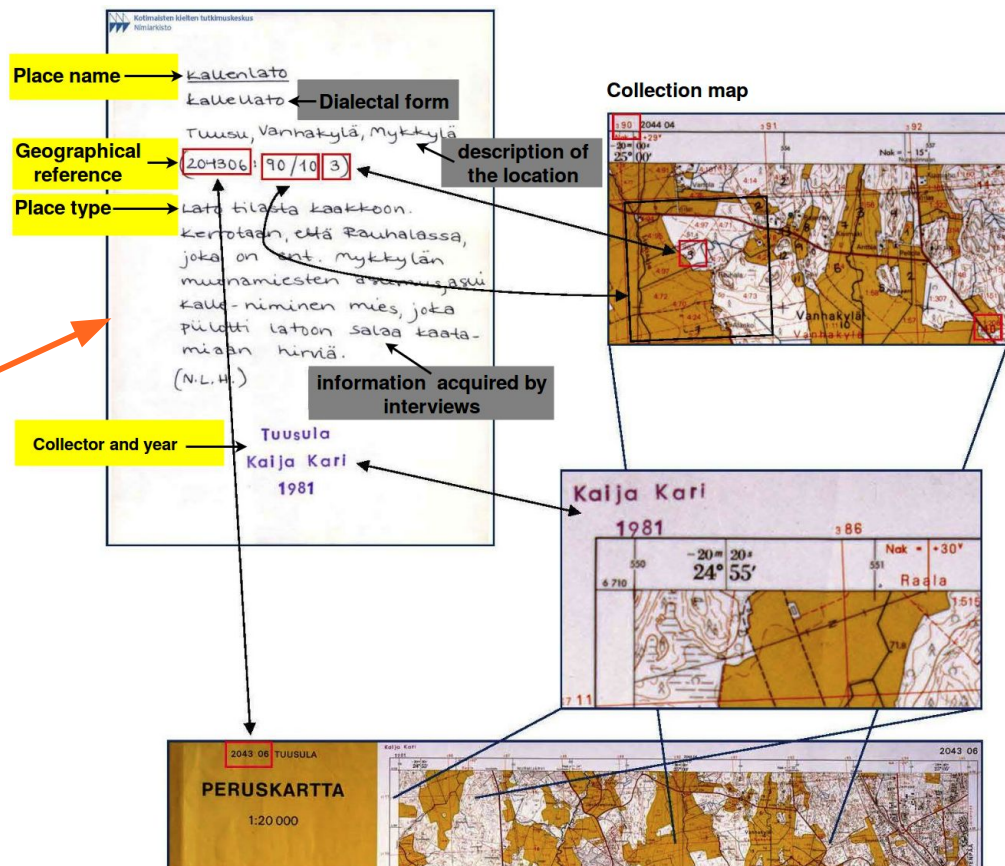
Jaakko Raunamaa (UH), Tiina Aalto (UH), Terhi Ainiala (UH), Helinä Uusitalo (Kotus)

Outline

- **Introduction**
 - Names Archive of the Institute for Languages of Finland
 - Digitisation of the Names Archive
- **Research questions**
 - Toponomastic research in a digital environment
- **NameSampo infrastructure for toponomastic research**
- **Examples of using NameSampo**

Names archive

- Owned by the Institute for Languages of Finland
- Collection started in the beginning of 20th century
- ~ 2.7 million entry slips of toponymic information



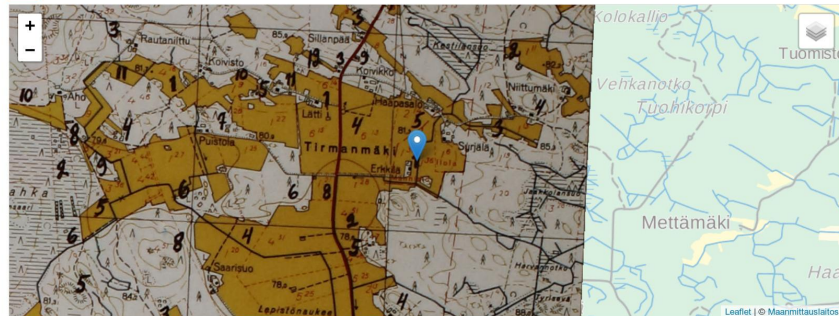
Digitisation of the Names Archive

- The corpora of place names in the Names Archive was converted into a digital archive in 2014–2017
- Search interface: nimiarkisto.fi



[Main page](#)
[Satunnainen paikannimi](#)
[Etsi paikannimeä](#)
[Ohjeet](#)
[Tietoja Nimiarkistosta](#)
[Anna palautetta](#)

[Page](#) [Discussion](#)



Erkkilä (Kotus, 1980)

Merkinnän tiedot

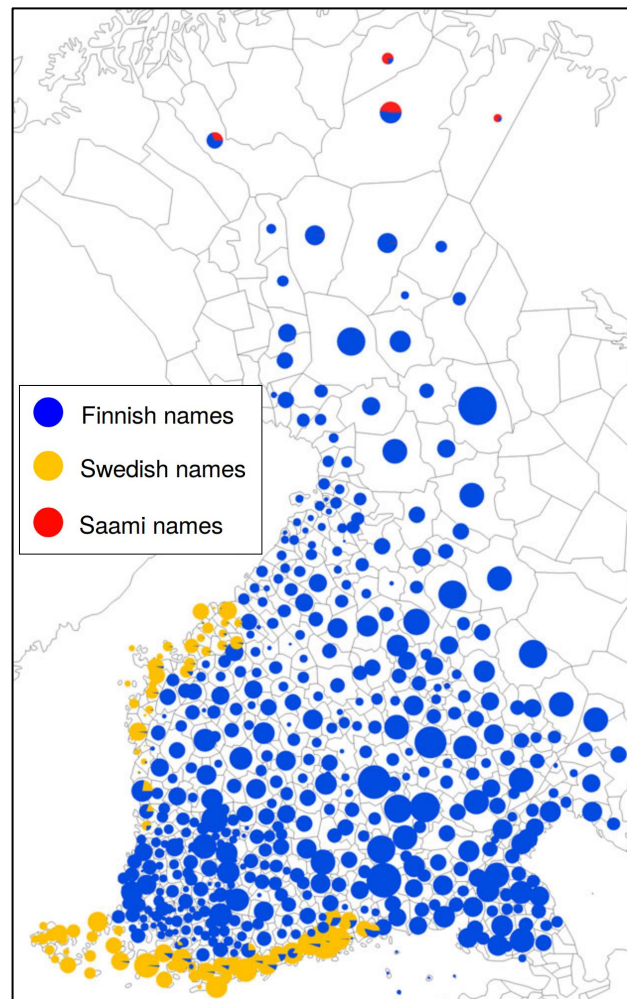
Nimi: Erkkilä
Nimenlaji: paikannimi
Paikanlaji: tila

Sijaintitiedot

Keruukartan ruutu: 39 / 68
Pistesijaintimerkintä: 1
Karttalehden tunnus: 1564
Koordinaatit: 61°1'10.996964"N, 22°52'15.327113"E (WGS 84)
Paikkannuksen tarkkuus: Kerääjä on sijoittanut merkinnän tähän pisteeseen.

Names archive

Language	Territory	Name entries	Scanned	Available at nimiarkisto.fi
Finnish Swedish	Finland	2 300 000	yes	yes
Finnish	Karelia	326 000	yes	yes
Saami	Finland, Petsamo, Swedish and Norwegian Lapland	33 000	yes	no
Finnic	Sweden, Norway, East Karelia, Viena Karelia, Aunus Karelia, Vepsian territory, Ingria, Estonia	91 000	yes	no



Research Questions and Goal

To answer toponomastic research questions in a digital environment, one should be able to:

- combine data from multiple registers
- filter the search results (e.g., by place type or by broader area)
- split place names into an appellative specifier and a generic part
- produce distribution maps without clustering the markers
- create charts based on the search results
- use many kinds of historical and current background maps
- export the data for further analysis, e.g., with GIS or statistical software

Ontology of Geographic Features

- Names Archive (*Institute for Languages of Finland*)
 - 2 300 000 place entries
 - **nearly 15 000 place types**, no hierarchy
- The Place Name Register (*National Land Survey of Finland*)
 - 800 000 place entries
 - **153 place types** with four-level hierarchy

▶ Develop an ontology for classification of geographic features together with onomastic researchers

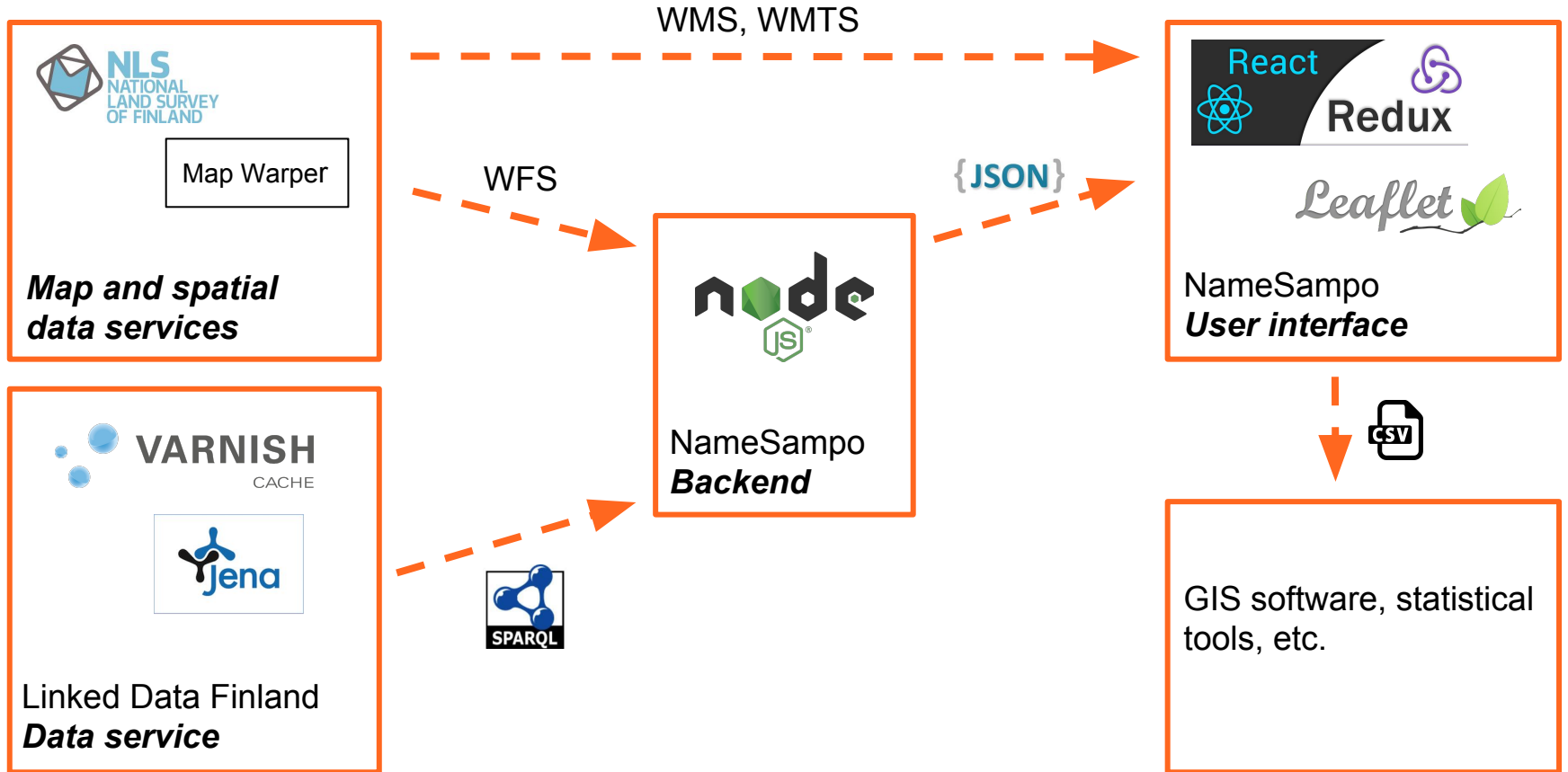


A word cloud of Finnish place names ending in '-mäki'. The words are arranged in a roughly circular pattern. The most prominent word is 'harjumäki' in red. Other visible words include 'metsämäki' (blue), 'kokonpolttomäki' (orange), 'mäntymäki' (orange), 'niittymäki' (orange), 'peltomäki' (orange), 'ahomäki' (orange), 'kalliomäki' (orange), 'kuoppamäki' (blue), and 'tervanpolttomäki' (orange).



MÄKI 'hill'

NameSampo infrastructure



Using NameSampo

- A specific name type can easily be visualized and studied through the NameSampo user interface.
- An example on the right shows results of search expression **pask(a)** which is one of the most common Finnish profanities. Altogether, there are 4674 name entries found including the name element *pask(a)**.

*= feces, muck; dirt

The screenshot displays the NameSampo interface. At the top, the search bar contains the query '*pask*'. Below the search bar is a table of results with columns: Name, Modifier, Base, PNR type, NA type, Area, Year, and Source. The table lists various names and their associated metadata. To the right of the table is a map of Finland showing a heatmap of the search results, with red and yellow areas indicating higher concentrations. Below the map is a pie chart titled '4674 results for the query '*pask*'' showing the distribution of name types. The legend for the pie chart is as follows:

Names archive place type	Count
Lampi	807
Undefined	691
Pelto	273
Saari	213
Lahti	205
Suo	204
Mäki	195
Niemi	127

At the bottom of the page, there are logos for Aalto University School of Science, University of Helsinki, HELIGI (Helsinki Centre for Digital Language Studies), and the Institute for the Languages of Finland (Kotipaisten Kielten Keskus).

Using NameSampo

Automatic compound splitting

NameSampo

Search place names

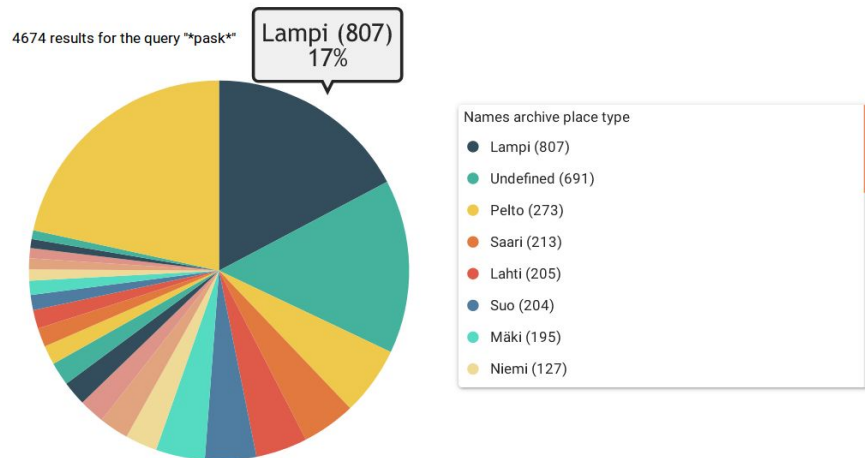
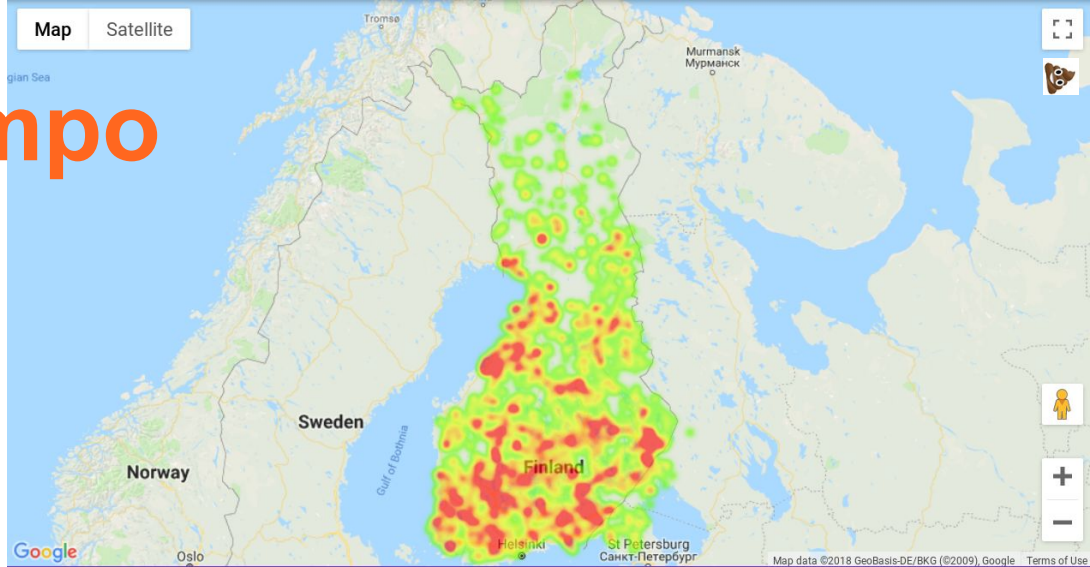
Name	Modifier	Base	PNR type	NA type	Area	Year	Source
Paskaharju	Paska	harju	Kohouma	Mäki	Alajärvi	1956	NA
Paskosalmi	Pasko	salmi	Suo tai kosteik...	Suo	Alajärvi	1965	NA
Ohrapaskanneva	Ohra	paskannev	Suo tai kosteik...	Neva	Alajärvi	1956	NA
Ohrapaskanneva	Ohra	paskannev	Suo tai kosteik...	Neva	Alajärvi	1963	NA
Paskoharjunmäki	Paskoharjun	mäki	Kohouma	Mäki	Alajärvi	1956	NA
Paskonevankangas	Paskonevan	kangas	Metsäalue	Metsä	Alajärvi	1965	NA
Paskonnevansaari	Paskonnevan	saari	Metsäalue	Metsä	Alajärvi	1956	NA
Paskamäki	Paska	mäki	Kohouma	Mäki	Alastaro	1989	NA
Hevosenspikapuu	Hevosenspaska	puu	Undefined	Undefined	Alastaro	1974	NA
Paskaranniittu	Paskarannan	niittu	Niitty tai laidun	Niitty	Alastaro	1974	NA
Paskuntöykkä	Paskun	töykkä	Kohouma	Mäki	Alastaro	1978	NA
Pasko			Muu maastop...	Maa	Alatornio	1968	NA
Paskajänkkä			Suo tai kosteik...	Suo	Alatornio	1955	NA
Paskalahti	Paska	lahti	Järven osa	Lahti	Alatornio	1956	NA
Paskonsaari	Paskon	saari	Saari tai luoto	Saari	Alatornio	1957	NA

Harmonized place types

Sort and filter search results

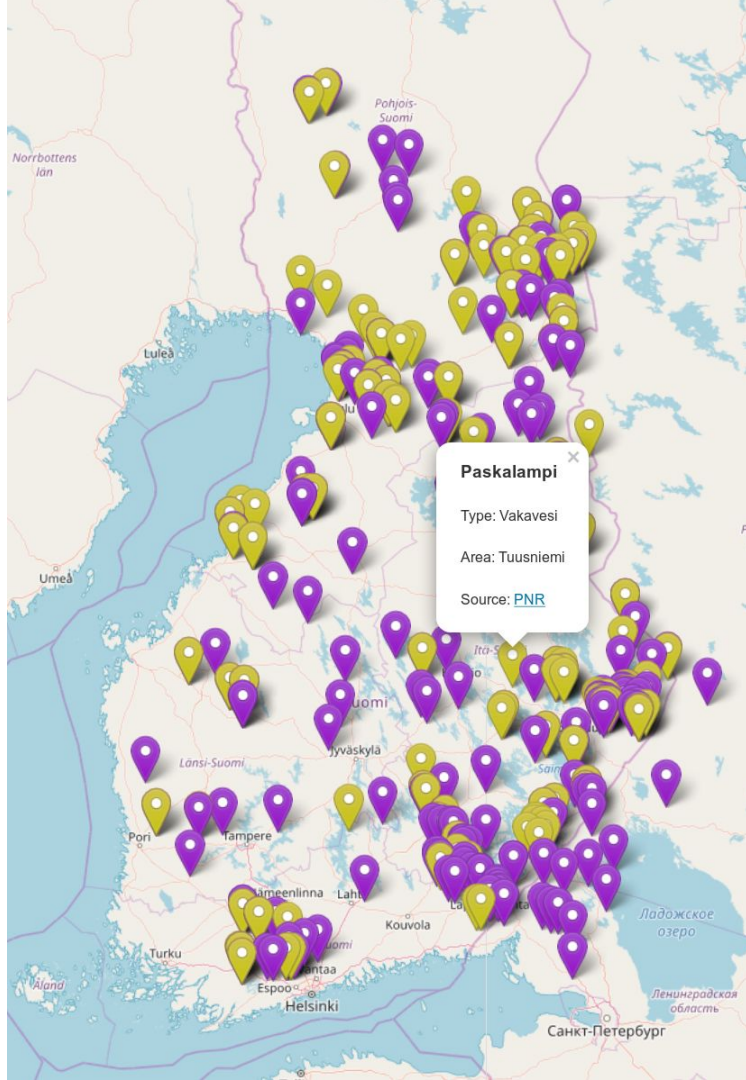
Using NameSampo

- It is easy to notice that the feature type “lampi” (‘pond, pool, waterhole, small lake’) is the most common. Clearly, the name element *pask(a)* has been connected to small and insignificant waters. Many of the ponds are located in swampy areas where the element *pask(a)* is probably describing the quality of the water as well.



Using NameSampo

- Search expression
 - *paskalampi OR paskalammi*
 - 473 results
- Comparison of
 - Place Name Register (yellow)
 - Names Archive (violet)



Thank you!

- **Project website: seco.cs.aalto.fi/projects/nimisampo**
- **Workshop article accepted in ACM SIGSPATIAL 2018, Seattle, USA**
 - November 6th
 - bgmartins.github.io/sigspatial-geohumanities/
- **nimiarkisto.fi** (published in 2017)
- **nimisampo.fi** (under development, publication in the winter 2018-2019)