NameSampo: A Linked Open Data Infrastructure and Workbench for Toponomastic Research

Esko Ikkala
HELDIG – Helsinki Centre for Digital Humanities
University of Helsinki, Finland
Semantic Computing Research Group
Aalto University, Finland
esko.ikkala@aalto.fi

Jouni Tuominen
HELDIG – Helsinki Centre for Digital Humanities
University of Helsinki, Finland
Semantic Computing Research Group
Aalto University, Finland
jouni.tuominen@helsinki.fi

Jaakko Raunamaa
Department of Finnish Finno-Ugrian and Scandinavian Studies
University of Helsinki, Finland
jaakko.raunamaa@helsinki.fi

Tiina Aalto
Department of Finnish Finno-Ugrian and Scandinavian Studies
University of Helsinki, Finland
tiina.k.aalto@helsinki.fi

Terhi Ainiala
Department of Finnish Finno-Ugrian and Scandinavian Studies
University of Helsinki, Finland
terhi.ainiala@helsinki.fi

Helinä Uusitalo
Institute for the Languages of Finland
helina.uusitalo@kotus.fi

Eero Hyvönen
HELDIG – Helsinki Centre for Digital Humanities
University of Helsinki, Finland
Semantic Computing Research Group
Aalto University, Finland
eero.hyvonen@helsinki.fi

ABSTRACT

This paper presents a series of projects where one of the main sources for toponomastic research in Finland, the corpora of 2.7 million place names in the Names Archive database of the Institute for the Languages of Finland, was digitized, enriched, and published as Linked Open Data using a data processing pipeline. Utilizing the Linked Data infrastructure and various external data sources, a modern full-stack web application, NameSampo, was created in collaboration between toponomastic researchers and computer scientists for searching, analyzing, and visualizing digital toponomastic data sources.

CCS CONCEPTS

• **Information systems → Geographic information systems**; *Web interfaces*; *Resource Description Framework (RDF)*; *Search interfaces*; • **Applied computing → Digital libraries and archives**;

KEYWORDS

Onomastics, Toponomastics, Toponyms, Linked Data, RDF, Geographic Information Retrieval

ACM Reference Format:
Esko Ikkala, Jouni Tuominen, Jaakko Raunamaa, Tiina Aalto, Terhi Ainiala, Helinä Uusitalo, and Eero Hyvönen. 2018. NameSampo: A Linked Open Data Infrastructure and Workbench for Toponomastic Research. In *2nd ACM SIGSPATIAL Workshop on Geospatial Humanities (GeoHumanities'18), November 6, 2018, Seattle, WA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3282933.3282936

GeoHumanities'18, November 6, 2018, Seattle, WA, USA

ACM ISBN 978-1-4503-6032-6/18/11…$15.00
https://doi.org/10.1145/3282933.3282936

1 INTRODUCTION

Names are a linguistic universal and all known languages make use of names to identify individual people, places, and other referents. The study of names, known as 'onomastics' [6], is a multidisciplinary research field, but for a long time already, it has been profiled as a linguistic field of research [2]. This paper concentrates on Finnish onomastics and exclusively on toponomastics, the study of place names (see also [6]).

Toponomastics has played a significant role in Finnish onomastic research, even theoretically and methodologically. Since there has been, already from the mid-1900s, vast and comprehensive material of place names available, the research has been versatile and continuously nascent. These manual, non-digitized archives have served as a starting point especially in traditional toponomastics. However, the global megatrend of digitizing archival materials creates unprecedented possibilities for onomastic research. Electronic

access to archives means that data analytics, information visualization, and computational methods can be harnessed for answering new toponomastic research questions in the spirit of the Digital Humanities paradigm (cf. [4, 15]). In utilizing various archives or registers, we should bear in mind, however, that any place name collection can never include the whole onomasticon of the country nor even of some municipality or some other area [2]. Therefore digital tools for collecting, making comparisons, harmonizing, and further developing materials in various registers and archives are highly needed. Also gradual harmonization work of the official registers and archives is essential for ensuring interoperability and, e.g., federated search functionalities.

Research Questions and Goal One of the main sources for toponomastic research in Finland, the corpora of place names in the Names Archive database[1] of the Institute for the Languages of Finland has been recently digitized. Also a search interface solely for the Names Archive database was created by which places (i.e., place name records) based on their name, feature type, record collector, or historical municipality (parish) can be searched and viewed on a map. This is useful but not sufficient for a wide range of research needs in toponomastics. To answer toponomastic research questions in a digital environment, it should be possible to, among other things,

- combine data from multiple registers
- filter the *federated* search results by different properties (e.g., by place type or by broader area)
- produce circulation maps where a large number of map markers are rendered on a map without clustering
- create charts based on the search results
- use many kinds of historical and current background maps
- split the place names into compounds (if applicable)
- export the data for further analysis, e.g., with GIS or statistical software

In order to facilitate these requirements, the goal of the NameSampo project is to first convert, enrich and publish the digitized Names Archive as Linked Open Data [5]. After that, the focus is on creating a flexible, easy to use workbench of search, data analytic, and visualization tools by utilizing the Linked Data infrastructure, various external map and data sources, and a collaborative workflow between onomasticians and computer scientists. The intent is that a wide variety of toponomastic research questions can be addressed interactively using a web browser.

Results This paper presents an overview of the NameSampo project and system, with first results related to the vocabularies and ontologies used for modeling the underlying Names Archive data, data transformation into Linked Open Data, data service publication, user interface, and system architecture. The NameSampo user interface has not been officially published at the time of writing, but the development version link is available at the project webpage[2].

2 TOPONOMASTICS IN FINLAND AND RELATED WORK

The etymological research, which studies the origin of names, is the most traditional field in Finnish onomastics, likewise in other countries. Even though the first studies at the beginning of the 20th century could not yet make use of exhaustive place name collections but concentrated on single names instead, already the dissertation by Viljo Nissilä in 1939 [14] made a difference. Nissilä studied names of natural places associated with the Vuoksi River and besides etymological explanation of individual names, he made a classification of names according to their naming motivations.

Etymological research is closely connected to cultural-historical and settlement-historical research, where place names are used to study local history of some specific area. The lack of old historical documents in Finland makes place names especially significant historical source material. For example, Jouko Vahtola's dissertation [16] about the origins of the settlement in the Torne and Kemi River Valleys is very much based on systematic research of place names.

The study of loan names has been established as its own field in Finnish onomastics. Again, place names have been an important source when studying linguistic influences and contacts in the country. For example, in inland Finland place names show that the area of Saami speaking people has been much wider than it is today [1].

The place name collections have been of major relevance to the typological research which has aimed to construct a comprehensive picture of Finnish place names. The core issues have been the exploration of the structure of the names and naming systems as well as the examination of the principles of name giving and their semantic classification. Especially Eero Kiviniemi's work has been groundbreaking. In his dissertation [9], he analyzed the names including the Finnish first participle (*-vA*) and managed to show the strong model of analogy in the name giving. In order to be able to conduct this research, he scrutinized the whole manual Names Archive to find names including the first participle.

Moreover, Kiviniemi has made an investigation of lexemes or words included in place names [11], in which he presents the most frequent generic parts (e.g. *-järvi* in *Hauki/järvi* 'Pike/lake'), appellatival specifiers (e.g. *Hauki-* in *Hauki/järvi*) and the most common Finnish place names. This study has been made by measuring in millimeters the height of the manual place name slips in the Names Archive. Since the collections are arranged alphabetically according to the beginning of the name, the results concerning the generic parts can be characterized as more untrustworthy.

Kiviniemi was also influential on creating systematic categorization of topographic words based on a syntactic-semantic analysis model [10] which was originally created by Kurt Zilliacus [17]. This means that all the name parts used in place names are lexical-semantically and syntactic-semantically classified. Already before Kiviniemi's model researchers collecting place names added a description of the type of referent to the name slips of the Names Archive (e.g., place called *Alinenkorpi* 'low(er) forest' in Ahlainen municipality, is categorized as *pelto* 'field').

Antti Leino's doctoral thesis [12] is an good example of modern typological research which is based on using computational methodologies with digital place name data. Leino studied correlations between lake names (e.g., opposite pairs such as black - white as specific parts of lake names) by using a corpus of 58 000 lake names obtained from the National Land Survey of Finland's Place Name Register.

[1] https://www.kotus.fi/en/corpora_and_other_material/names_archive
[2] https://seco.cs.aalto.fi/projects/nimisampo

3 DATA SOURCES FOR TOPONOMASTICS

In this Section we present the main Finnish services in detail and compare the functionalities of different national and international services for place names.

3.1 The Names Archive of the Institute for the Languages of Finland

One of the Institute for the Languages of Finland's archives, the Names Archive contains valuable information for onomasticians and linguists, as well as for archaeologists, historians, folklorists, and natural scientists. The Names Archive's main collections consist of geographical place names which are mainly oral tradition stored by researchers and students with an academic linguistic background. The oldest collections are from the end of 1800. The archive contains approximately 2.7 million place names, organized by parish borders. Figure 1 illustrates the geographical distribution of the place names in Finland and beyond.

The place name information has been collected by using entry slips (depicted in Figure 2) with details like the place name in standard Finnish, the dialectal form and inflection of the name, location (parish, village, house, and coordinates), classification of the location, description of the location (if applicable), and any information about the name or location acquired by interviewing people. In addition to verbal descriptions of the locations, the names have been indicated on collection maps (approximately 10 000).

The corpora of place names in the Names Archive was converted into a digital archive in 2014–2017. The objective of the project was to convert the place name entry slips, collection maps, and the attribute and metadata related to them into digital format. That enables the use of Name Archive data together with the geographic and other data provided by other organizations. The actions of the digitization were

(1) scanning of the entry slips into digital format,
(2) conversion of the main information on the entry slips such as place name, geographical reference, parish, collector, collection year, and place type (yellow fields in Figure 2) into database format, and
(3) indicating the specific locations as coordinates.

The platform for the digital archive was created in cooperation with CSC – IT Center for Science[3]. The project group was assisted by a steering group and a support group representing several of the Institute's cooperation partners.

The objective of the digitization is to improve the usability of the existing data and to offer an electronic environment for enriching and supplementing the data. In the initial stage, the digitization project focused on names in Finnish, but the idea is to include the collection of place names in Swedish and Saami. When the corpora have been converted into digital format, the data can be further enriched and supplemented. Citizens can provide information about whether the names are still used today or have been replaced by new ones, whether a place exists anymore, or what new names have been created. They can also provide information about new names that have emerged and supplement background information about places and names.

[3]https://www.csc.fi/

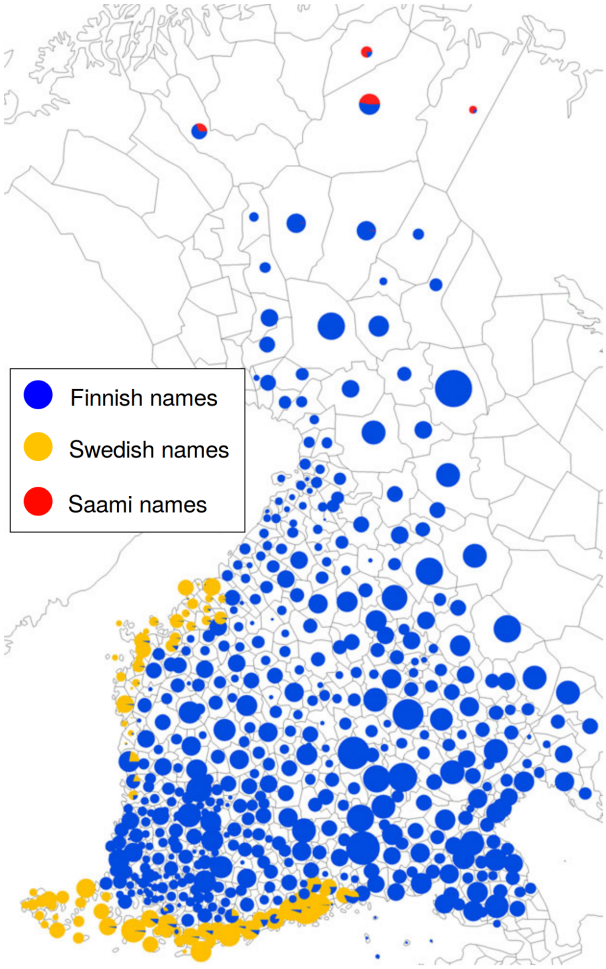

Figure 1: The map describes the numbers of Finnish, Swedish, and Saami place names in the Names Archive. The collections have been organized based on the parish borders in 1938. The size of the circle represents the number of place names in a parish. The typical number is 5 000–6 000 place names per parish but some parishes include tens of thousands of place names.

3.2 Place Name Register of the National Land Survey of Finland

Another widely used Finnish place name source is the Place Name Register[4], maintained by the National Survey of Finland, which contains approximately 800 000 place names. The Place Name Register is provided as Geography Markup Language (GML) files or through a Web Feature Service (WFS) API, but it is not available as tabular data. However, a Linked Data conversion[5] has been published in connection with another project.

The main difference between the Place Name Register and the Names Archive is the sheer amount of names. The Place Name

[4]https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/product-descriptions/geographic-names
[5]http://www.ldf.fi/dataset/pnr/index.html

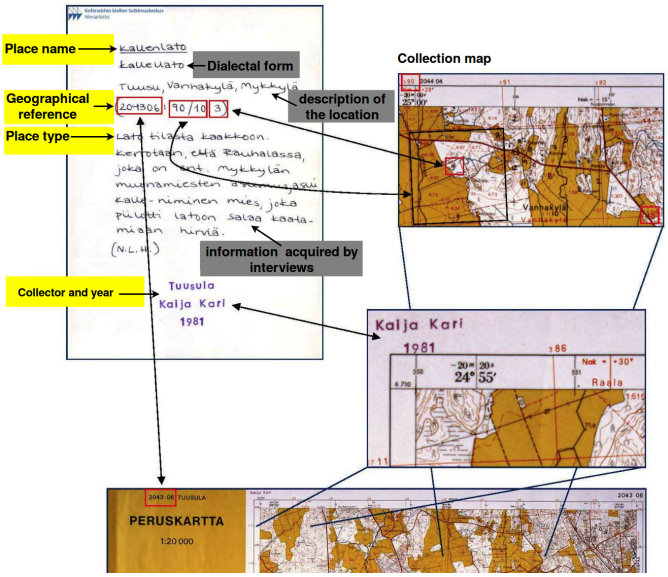

Figure 2: Conversion of the main information on the digitized Names Archive entry slips into database format

Register contains names of significant places found in the current map of Finland, whereas the Names Archive has many more microtoponyms (names only used in small user circles) and chronological variety as well. Names in the Place Name Register are also standardized according to existing praxis which means that it does not depict similarly dialectical differences as the Names Archive does. Noteworthy is also that many names in the official register are censored, and for example in Kirkkonummi municipality the lake name *Kolmperä* ('three' + 'back'; 'butt'; 'tail') has—according to the Names Archive's name slips and old maps—been originally *Kolmperse* ('three' + 'ass').

3.3 Comparison of National and International Registries and Services for Place Names

In the beginning of the 20th century, as the impact of nationalism was strong, it was fairly popular in many European countries and regions to collect place names into books and archives. There are not, however, many archives whose material is digitized. Another problem is that the level of digitization between different archives is very heterogeneous. For example, all the 3.7 million name slips in the Swedish Place Name Register[6] are digitized but do not have coordinates to locate them on a map.

Text- or map-based search is the standard way for a human user to access digital place name collections but due to various data models and user interface designs, it is not possible to utilize the collections in a uniform way. Also for Digital Humanities research, the availability and functionalities of an Application Programming Interface (API) or a documented data dump are crucial for programmatically accessing the data.

Creating and integrating geographical gazetteers in spatial humanities has become an active area of research and development.[3] Several national and international place name registers are available for electronic access, for example Getty Thesaurus of Geographic Names[7], GeoNames[8], Canadian Geographical Names Data Base[9], The Historical Gazetteer of England's Place-Names [10], Gazetteer for Scotland[11], The Placenames Database of Ireland[12], Dutch service Histopo.nl[13], and Danish service DigDag[14]. Table 1 presents a comparison of these digital services and their features for working with place names.

4 NAMESAMPO WORKBENCH FOR TOPONOMASTICS RESEARCH

In this Section, we present the main phases of the NameSampo project, analyze the data models and technical architecture of the NameSampo system, and give practical examples of how the workbench facilitates toponomastic research.

4.1 Harmonizing Data Model and Ontologies for Place Name Registers and Place Types

Based on our earlier research on creating the Finnish Ontology Service for Historical Places and Maps[15] [7], a simple harmonizing data model was used for the Linked Data conversion of the digitized Names Archive. The namespace prefixes used with the classes and properties are presented in Table 2, and the core properties of the main class *hipla:Place* are presented in Table 3. Also a number of ad hoc properties were used to model information that is found solely in one registry (e.g. collection map information in the Names Archive). The Linked Data conversion of the Place Name Register uses an almost similar data model.

Table 2: Namespace prefixes

Prefix	Namespace
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
skos	http://www.w3.org/2004/02/skos/core#
schema	http://schema.org/
gs	http://www.opengis.net/ont/geosparql#
wgs84	http://www.w3.org/2003/01/geo/wgs84_pos#
xsd	http://www.w3.org/2001/XMLSchema#
hipla	http://ldf.fi/schema/hipla/
pnr	http://ldf.fi/schema/pnr/
na	http://ldf.fi/schema/names-archive/

For modeling the hierarchy of place types, the hierarchy used in the Place Name Register (prefix *pnr:*) formed a natural basis. The place types (prefix *na:*) used by the Names Archive collectors were

[6] http://www.sprakochfolkminnen.se/sprak/namn/ortnamn/ortnamnsregistret/

[7] http://www.getty.edu/research/tools/vocabularies/tgn/index.html
[8] http://www.geonames.org/
[9] http://www4.rncan.gc.ca/search-place-names/search?lang=en
[10] http://www.placenames.org.uk
[11] http://www.scottish-places.info
[12] https://www.logainm.ie/en/
[13] http://histopo.nl
[14] http://www.digdag.dk
[15] http://hipla.fi

Table 1: Comparison of place name services

Service	Text search functionalities	Map features	API
Getty Thesaurus of Geographic Names	basic text search	no	SPARQL
GeoNames	text search with filters	no	RESTful
Canadian Geographical Names Data Base	text search with filters	no	RESTful
The Historical Gazetteer of England's Place-Names	text search with filters	map view for one place name at a time	no
Gazetteer for Scotland	text search with filters	map search	no
The Placenames Database of Ireland	text search with autocomplete	map search and browsing	XML
Dutch service Histopo.nl	basic text search	no	no
Danish service DigDag	text search with filters	map search	WMS, WFS

placed at the lowest hierarchy level because, in general, they are more elaborate than the Place Name Register's place types.

Thus, the hierarchical data model for the place type ontology was constructed from the following six classes with *rdfs:subClassOf* relations. The number of instances is shown in parenthesis:

gs:Feature
  pnr:PlaceTypeTheme (9)
    pnr:PlaceTypeGroup (30)
      pnr:PlaceTypeSubgroup (59)
        pnr:PlaceType (153)
          na:PlaceType (14 572)

4.2 Converting the Names Archive into Linked Data

The Names Archive data was provided as a CSV table, where each row represents a digitized place name entry slip and coordinates are expressed in the Finnish ETRS89-TM35FIN coordinate system. For ensuring interoperability with other systems, the coordinates were transformed into World Geodetic System (WGS84) with a PHP script[16].

Then the CSV rows were converted to Linked Data according to the data model presented in Section 4.1 by a Python data processing pipeline[17] using RDFLib[18]. To facilitate toponomastic analysis, an open source linguistic compound splitter[19] was added to the data processing pipeline. The compound splitter makes it possible to split a place name into an appellative specifier and a generic part.

Finally, the data was published at the Linked Data Finland platform[20] (LDF.fi) [8]. The LDF.fi portal is powered by a combination of Fuseki SPARQL servers[21] running in Docker containers[22] for storing the primary data and a Varnish Cache web application accelerator[23] for routing URIs and content negation.

[16] http://www.loukko.net/koord_proj/
[17] https://github.com/SemanticComputing/kotus-names-archive-csv2rdf
[18] https://rdflib.readthedocs.io/en/stable/
[19] https://github.com/tsnaomi/finnsyll
[20] This linked data service with various datasets, tools, and learning materials is available at http://ldf.fi. The SPARQL endpoint for NameSampo data is http://ldf.fi/kotus-digital-names-archive/sparql.
[21] https://jena.apache.org/documentation/fuseki2/
[22] https://www.docker.com/
[23] https://varnish-cache.org/

4.3 Creating an Ontology for Classification of Geographic Features

An initial observation was that the place names of the newly digitized Names Archive were categorized into nearly 15 000 distinct place types, with no hierarchy amongst them. As presented in Section 4.1, the Place Name Register maintained by the National Land Survey of Finland includes a four-level hierarchy for place types, with 153 place types at the lowest hierarchy level, which made the comparison of these two data sources ineffective. To facilitate the interoperability of these two major data sources, an ontology for classification of geographic features was developed by onomastic researchers.

The starting point was a spreadsheet equipped with the official place type classes from the Place Name Register and another spreadsheet with all distinct place types from the Names Archive. Then, one by one, the Names Archive's place types were placed by an onomastician under the corresponding place types of the Place Name Register classification.

For example, when defining a hill, the Names Archive collectors have used over one thousand different expressions. You can find expressions for cliffs, rocks, ridges, and all places of data which are defined somehow to be higher than their surroundings. All these expressions were placed under the corresponding Place Name Registry class 'hill', 'a place that is higher than its surrounding area'.

The combining of place types was often not straightforward. For example, the Place Name Registry classification made for maps and official use does not make a difference between tar pits, archaeological sites, meeting points, places somehow connected to hunting or fishing, and so on. Places of this kind have been named because, for some reason, they are known or meaningful for the name givers. The only suitable Place Name Register class for these is 'other place on the terrain', but this class is clearly too vague.

Also in some cases the Place Name Register classes are too close to each other by their meanings. For example, when dealing with classes of sites which are bordered by waterways, one must decide if a Names Archive place type should be placed in the class 'group of islands or islets' or 'island or islet' or 'stone in water or group of stones in water' or 'shallows'. Also place type expressions, such as 'a south side of an island', are problematic; this does not mean an island itself but a specific place on an island, and the actual place type remains unclear.

Finally, the resulting spreadsheet was converted into an ontology by the Python data processing pipeline introduced in Section 4.2.

Table 3: Properties of the class *hipla:Place*, the superclass of all places

Property	Identifier	Cardinality	Range	Value
Preferred name in use	skos:prefLabel	1	Literal	string@languge
Preferred place type	rdf:type	0..1	rdfs:Class	URI
Preferred part of relation	gs:sfWithin	0..1	rdfs:Class	URI
Latitude	wgs84:lat	0..1	Literal	Decimal degrees
Longitude	wgs84:long	0..1	Literal	Decimal degrees
Polygon of the place	schema:polygon	0..1	Literal	String "x1,y1 x2,y2 ... xn,yn"

For corrections and updates, the class hierarchy can be edited in the spreadsheet and the conversion process can be run again.

4.4 Using NameSampo for Toponomastic Research

The NameSampo workbench depicted in Figure 3 offers multiple functions for producing map and statistic views of a single place name, groups of place names, and name types. The search results can be filtered by the class of the geographic feature (e.g., only show results belonging to the class 'field'). At the time of writing the Place Name Register place type and the Names Archive place type are shown in separate columns, but in the next version of the user interface they will be combined into a single column. The place type ontology described in 4.3 provides the hierarchy for the place types, so it is possible to add a hierarchical filter to this unified place type column.

The place names can also be classified by their appellative or generic part. Summaries and distributions of feature types can be made, as well as circulation maps of certain place name types. With the Linked Data approach it is also possible to use multiple place name collections as a source. From the menu button in the top left corner of the user interface the user can select the data sources, select from three options for rendering markers (clustered / individual / heatmap), and export the results as a CSV file for further analysis.

For studying the evolution of place names, NameSampo includes a variety of background maps and historical map layers from the 19th century Senate atlas to current topographical map of National Land Survey of Finland. They which provide an option to study the historical changes that happened in the surroundings of a place. The background map and map layers can be selected from the icon at the top right corner of the map view. The opacity of the map layers can be controlled from the bottom right corner of the map view.

The screenshot in Figure 4 describes how a specific name type can easily be visualized and studied through the NameSampo user interface by a toponomastician. The example shows the geographical distribution of place names starting with *Pakosauna-* (*Pako* 'refuge, escape' + *sauna* 'sauna; place with fireplace and cover') and their area of concentration in the region of Ostrobothnia (in Finnish 'Pohjanmaa') as a heatmap on the left. When we look closer at the individual markers on the right side of Figure 4, we can see that names starting *Pakosauna-* are far from the bigger settlements. It can therefore be assumed that they are really meant for escaping and hiding. Historical sources reveal that the period of the so called isovifha (in English 'the great hatred') during the the Great Northern War (1700–1721) was especially harsh for Ostrobotnia and attacks of Russian cossacks forced many people to flee from their homes [13]. In addition, the Names Archive collectors have added notes about the oral tradition which suggest that most of the *Pakosauna-* names are from the turbulent times of the 18th century.

4.5 NameSampo System Architecture

The general architecture of NameSampo is presented in Figure 5. The system consists of a NodeJS[24] backend build with Express framework[25] (in the middle) and a frontend based on React[26] and Redux[27] (on the right). The Linked Data Finland platform used for the public open linked data service is shown on the left. An instance[28] of MapWarper[29] (on the left) was created for aligning the historical maps and for using them in NameSampo. When designing the architecture, the main goal of the backend was to ease the combining of attribute data from multiple SPARQL endpoints and raster data from various spatial data sources into a React frontend.

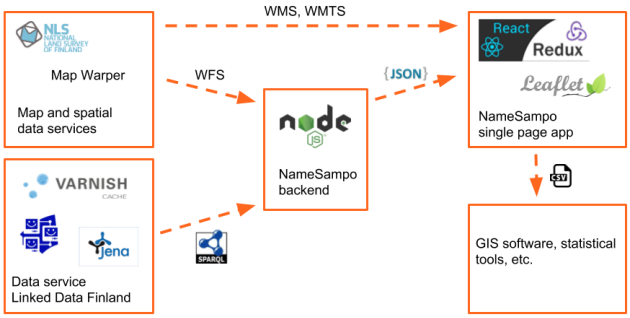

Figure 5: NameSampo system architecture

The architecture of the NameSampo user interface is presented in more detail in Figure 6. The search results provided by the backend are maintained in the application state, which enables efficient sorting and filtering of thousands of result rows and various dynamic visualizations based on the properties of the results. The application consists of four main React components with the following functionalities.

[24] https://nodejs.org/en/
[25] https://expressjs.com/
[26] https://reactjs.org/
[27] https://redux.js.org/
[28] http://mapwarper.onki.fi
[29] https://mapwarper.net/

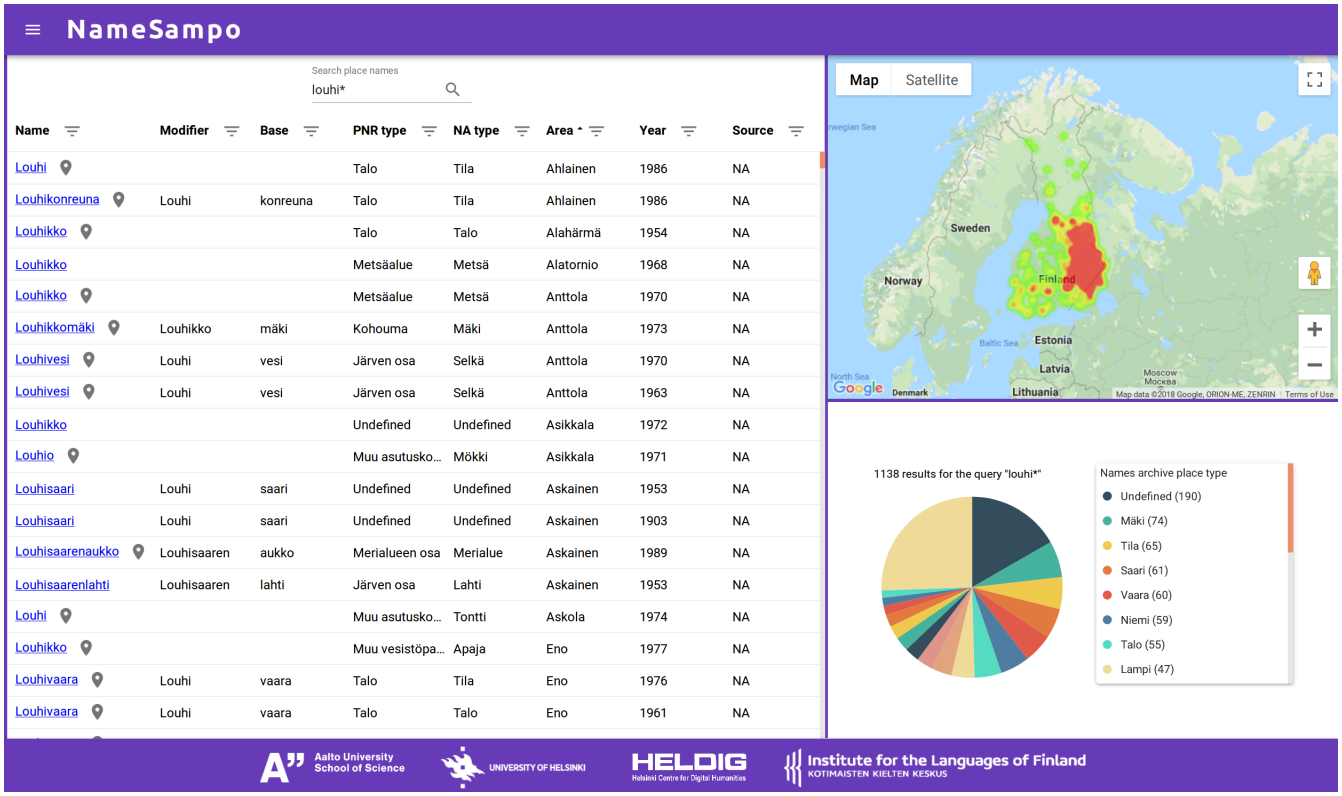

Figure 3: Main functions of the NameSampo workbench

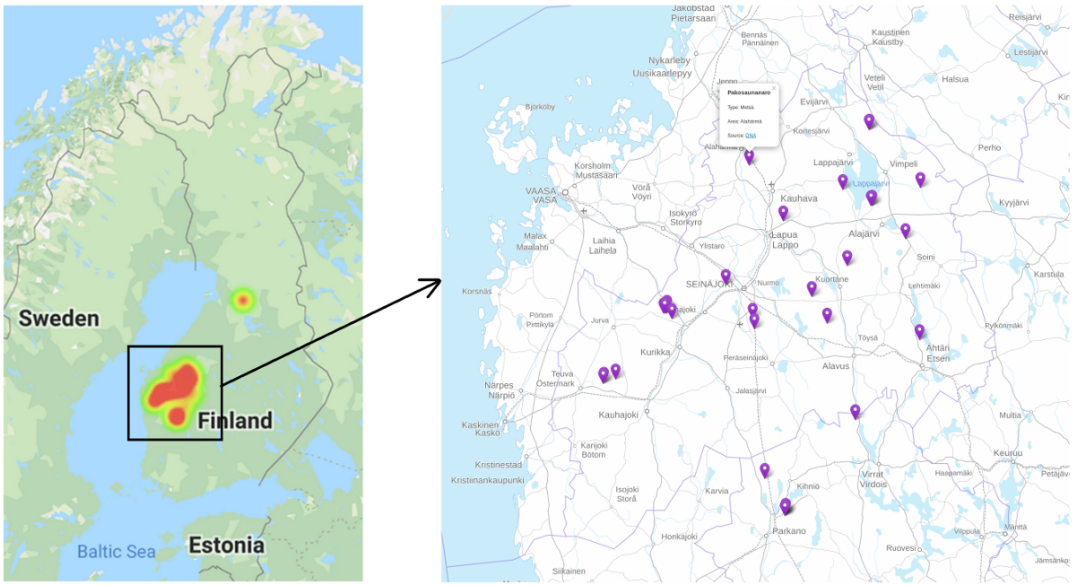

Figure 4: A heatmap and the exact locations of the place names starting with *Pakosauna*- ("Escape sauna")

Input component

- searches place names with one query term
- offers optional support for automatic suggestions and wildcard queries

Result table component

- built with React Virtualized[30] for efficiently rendering large amount of tabular data
- orders the results by column, ascending/descending
- filters the results by column values
- exports the results as a CSV file for further analysis

Result map component

- uses alternatively Google Maps JavaScript API[31] or Leaflet[32]
- visualizes the search results on a map
- colors map markers based on their data source
- offers three options for rendering markers: clustered / individual / heatmap
- supports various background maps and map layers

Result statistics component

- built with Victory[33] charting components
- generates distributions over result properties

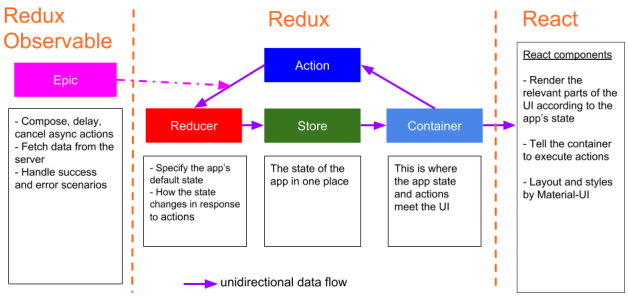

Figure 6: NameSampo user interface architecture

5 DISCUSSION

The main aim of the NameSampo project—from a humanist point of view—is to create new and more effective tools for toponomastic researchers to process the Names Archive data and combine it with other data sources. NameSampo meets these expectations and shows how beneficial the cooperation between computer scientists and humanists can be. Despite some semantic challenges, creating the ontology for classification of geographical features enables effective federated search functionalities, result filtering, and statistics based on the combined Names Archive and Place Name Register data. Also the automatic splitting of place names into the appellative specifier and the generic part opens up new possibilities for studying the characteristics of place names.

In addition to the automatic splitting of place names, the multiple background maps and historical map layers, the distributions over result properties, and the exporting of the federated result data as CSV, are functionalities that are not available in other national or international services for place names. Naturally, the NameSampo user interface could have many other features. It would be practical to have, for example, a possibility to compile and filter the results of different searches as one result set and display those on a map with customized markers. Functionalities for handling the duplicate entries of the same place name and location should also be developed, because the Names Archive data contain multiple place name entries of the same location, which have been collected in different years.

In general, combining data from multiple archives and collections poses several challenges, such as the usage of different data models, different coordinate systems, duplicate entries, and different categorizations. The Linked Data approach offers solutions for many of these problems but as the unification of the Names Archive and the Place Name Register place types illustrates, it may require considerable manual work by domain experts to harmonize the values of merely a single property (e.g. place type) in two different databases.

All in all, the NameSampo system demonstrates how the usability of digitized data can be enhanced. These kind of projects are needed in order to make sure that digitized archives are not just digital copies of the original archives, but that they are also used and developed.

[30] https://github.com/bvaughn/react-virtualized
[31] https://developers.google.com/maps/documentation/javascript/
[32] https://leafletjs.com/
[33] https://formidable.com/open-source/victory/

ACKNOWLEDGMENTS

The authors wish to acknowledge *Toni Suutari* for discussions, support, and providing the Names Archive data. Thanks to *Institute for the Languages of Finland* and *University of Helsinki Future Fund* for funding, and *CSC – IT Center for Science, Finland* for computational resources.

REFERENCES

[1] Ante Aikio. 2007. The Study of Saami Substrate Toponyms in Finland. In *Etymologie, Entlehnungen und Entwicklungen. Festschrift für Jorma Koivulehto zum 70. Geburtstag*, Irma Hyvärinen, Petri Kallio, and Jarmo Korhonen (Eds.). Uusfilologinen yhdistys ry., Helsinki, Finland, 5–34.
[2] Terhi Ainiala, Minna Saarelma, and Paula Sjöblom. 2012. *Names in Focus: An Introduction to Finnish Onomastics.* Finnish Literature Society, Helsinki, Finland.
[3] Merrick Lex Berman, Ruth Mostern, and Humphrey Southall (Eds.). 2016. *Placing names. Enriching and integrating gazetteers.* Indiana University Press.
[4] Eileen Gardiner and Ronald G. Musto. 2015. *The Digital Humanities: A Primer for Students and Scholars.* Cambridge University Press.
[5] Tom Heath and Christian Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space (1st edition).* Morgan & Claypool.
[6] Carole Hough. 2016. Introduction. In *The Oxford Handbook of Names and Naming*, Carole Hough (Ed.). Oxford University Press, Oxford, 1–13.
[7] Eero Hyvönen, Esko Ikkala, and Jouni Tuominen. 2016. Linked Data Brokering Service for Historical Places and Maps. In *Proceedings of the 1st Workshop on Humanities in the Semantic Web (WHiSe).* CEUR Workshop Proceedings, 39–52. Vol 1608.
[8] Eero Hyvönen, Jouni Tuominen, Miika Alonen, and Eetu Mäkelä. 2014. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In *Proceedings of ESWC 2014 Demo and Poster Papers.* Springer-Verlag.
[9] Eero Kiviniemi. 1971. *Suomen Partisiippinimistöä: Ensimmäisen Partisiipin Sisältävät Henkilön- Ja Paikannimet.* Ph.D. Dissertation. Finnish Literature Society, Helsinki, Finland.
[10] Eero Kiviniemi. 1975. *Paikannimien rakennetyypeistä.* Finnish Literature Society, Helsinki, Finland.
[11] Eero Kiviniemi. 1990. *Perustietoa paikannimistä.* Finnish Literature Society, Helsinki, Finland.
[12] Antti Leino. 2007. *On toponymic constructions: As an alternative to naming patterns in describing Finnish lake names.* Ph.D. Dissertation. Finnish Literature Society, Helsinki, Finland.
[13] Armas Luukko. 1967. *Suomen historia 1617–1721.* WSOY, Porvoo, Finland.

[14] Viljo Nissilä. 1939. *Vuoksen paikannimistö: 1.* Ph.D. Dissertation. Finnish Literature Society, Helsinki, Finland.
[15] Simon Taylor. 2016. Methodologies in Place-name Research. In *The Oxford Handbook of Names and Naming*, Carole Hough (Ed.). Oxford University Press, Oxford, 69–86.
[16] Jouko Vahtola. 1980. *Tornionjoki- ja Kemijokilaakson asutuksen synty: nimistötieteellinen ja historiallinen tutkimus.* Ph.D. Dissertation. Pohjois-Suomen historiallinen yhdistys, Rovaniemi, Finland.
[17] Kurt Zilliacus. 1966. *Ortnamnen i Houtskär.* The Society of Swedish Literature in Finland, Helsinki.