

Chapter 5

TOWARD DIGITAL SPATIALITY: RETHINKING THE WORLD'S LARGEST PLACE-NAME COLLECTION

Björn Karlsson, Kristina Neumüller,
and Elin Pihl*

THE pLaCe-nAMe CoLLeCTiON at the Institute for Language and Folklore (abbreviated in Swedish as Isof) in Uppsala consists of about 3.7 million cards with information about names from all over Sweden (and some from other parts of the world). It is, to our knowledge, the world's largest place-name collection. Thanks to digitization, which was already under way in the middle of the 1990s, the collection has been accessible online for over fifteen years. The development of the database has been compromised, however, by the analogue way of thinking about digital matters that characterized the initial digitization process. The relatively early digitization of the collection can thus be said to gradually have changed from an advantage to a significant hindrance, and we are now facing the challenge of turning the existing, rather unwieldy place-name database into a more useful and powerful resource, optimized for data-driven spatial research. The first steps have been taken by introducing a new data structure and the potential to geocode name data in the database, but there are many more steps to come. Both the past and the present are described in this chapter, together with an outline of our plans for the future.

The Development of the Place-Name Collection

It was in the early twentieth century that the first steps to create a place-name collection were taken. In 1902 a Royal Place-Name Committee was established, assigned to organize a systematic investigation of Swedish place-names. The results were to be published in a place-name series, with the Norwegian series *Norske Gaardnavne* serving as a model.¹ The goal was to describe and interpret all names within Swedish borders, with a detailed list of existing spellings and pronunciations for each name. In order to succeed in their task, the committee set about collecting information about Swedish place-names, in

* Björn Karlsson is a software developer, and Kristina Neumüller and Elin Pihl (<https://orcid.org/0000-0001-7990-0148>) are research archivists specializing in place-names at the Institute for Language and Folklore (Isof). Email: bjorn.karlsson@isof.se, kristina.neumuller@isof.se, and elin.pihl@isof.se.

1 Rygh, *Norske Gaardnavne*.

written historical sources and specialist literature and through the on-site interviewing of local informants, sometimes with a detailed map of the area. The name forms were written down in the Swedish dialect alphabet (Swedish: *Landsmålsalfabetet*), a very detailed phonetic notation system,² most often together with some kind of note on the location, and sometimes with other kinds of information about the name or the place.

In the 1920s the need to organize the increasing quantity of place-name material in a proper archive became obvious, and in 1928 the Swedish Place-Names Archives was founded. No one responsible for the archives at the time could possibly have imagined that, about eighty years later, the place-name information they were collecting and systematizing would be made available for the general public. The material in the archive was intended for specialists in place-names, familiar with its background, structure, purpose, and—not least—its many abbreviations. It was not considered a problem that an outsider would not easily be able to decipher the information in the place-name collection, nor understand the meaning of the abbreviations.

Over the years that followed, however, not only did the collected material continue to grow, but so did the interest in the material from others besides onomasticians.³ The place-name information turned out to be useful for philologists and other linguists, archaeologists, and researchers in agrarian history, as well as anyone interested in local history or genealogy. Moreover, access to place-name material was fundamental for people working with place-names at public authorities, such as Lantmäteriet (the Swedish mapping, cadastral, and land registration authority).

Hence, when an opportunity to finance the digitization of the place-name collection arose in the early 1990s, it was an easy decision to go ahead with the project. Deciding exactly how to carry out the process was not quite as easy, as no similar digitization project existed to serve as a model. A pilot project was therefore conducted, and after that the intricate procedure of scanning all the archive cards and registering their content was initiated. The material was then published gradually, with the last step being a rather time-consuming quality check. In 2008 the whole register was finally available online.

The First Attempt to Make the Collections Available Online

One of the main features of the first interface of the Swedish digital place-name register⁴ was to display the archive cards in the exact same order as found in the physical archive. In the archive, the material is organized parish by parish, following the administrative division that prevailed in Sweden in 1951. Within each parish there are four thematic sections (see Figure 5.1), and within each section the place-names are arranged

² See Lundell, “The Swedish Dialect Alphabet.”

³ Onomasticians are researchers within the field of names, personal names, place-names, and other names. Here, of course, the focus is place-names.

⁴ See www.isof.se/NAU-ortnamn/index.htm (accessed June 7, 2021).

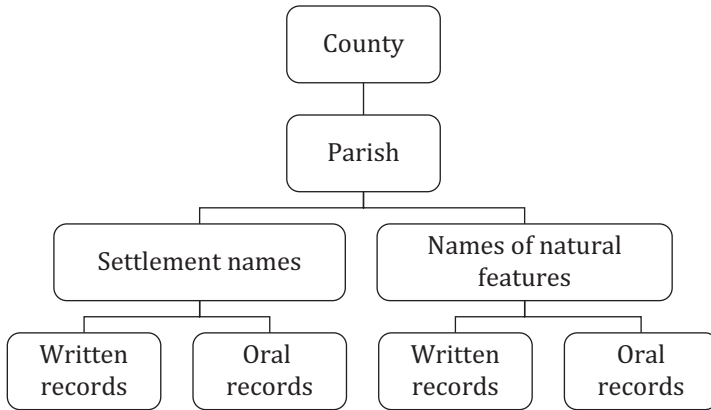


Figure 5.1. The structure of the Swedish place-name collection in the archive of Isof.

Image by Elin Pihl.

alphabetically. Further, within each place-name, the cards are sorted chronologically, displaying the oldest attestation first.

Even though the organization of the physical collection was copied, however, not all the information on the cards was transferred to digital text. Each physical card received its own post, hereafter called an “archive post.” The data transferred from the physical card to the archive post were as follows: the place-name (in a standardized modern form, as found in the records⁵ of Lantmäteriet); whether it classifies the name of a settlement or of a natural feature; whether it represents an older spelling or an oral record; what kind of locality the name denotes (and in this field many different values existed); and then the parish (Swedish: *socken*), district (Swedish: *härad* or equivalent), county (Swedish: *län*), province (Swedish: *landskap*), and country (with only Sweden, Åland,⁶ Finland, Estonia and “the rest of the world” as actual values). All archive posts were also labelled with a language—Swedish, Sami, or Finnish—but these labels are not

⁵ Not all names in the Place-Names Archives can be found in the registers of Lantmäteriet, but since the Spelling Reform in 1906 it was stated that the names should follow common spelling rules and that old spellings or decorative spellings should be avoided, and therefore all the place-names in the Place-Name Archives were standardized according to existing rules. Sometimes, however, there have been problems etymologizing the name, and therefore a problem deciding the correct spelling of the name. In other cases common spelling rules have changed over time, and the recordings from before that change have not been altered according to the new spelling, but new entries have. Sometimes the official name has been changed by Lantmäteriet. Thus, there are names that have been registered with different spellings. This is something we try to amend, but the changes have to be made manually, and therefore there are still some names that appear to be different place-names even though they are just different spellings of the same name.

⁶ Åland is not a country but an autonomous region of Finland. It has been treated as a country in the Swedish place-name collection, however.

Nr	Ortnamn	Lokal	Socken	Härad	Län
1	Uppsala	gd	Huddinge	Svartlösa hd	Stockholms
2	Uppsala	gd	Huddinge	Svartlösa hd	Stockholms
3	Uppsala	gd	Huddinge	Svartlösa hd	Stockholms
4	Uppsala	riven gd	Huddinge	Svartlösa hd	Stockholms
5	Uppsala	ägomark	Huddinge	Svartlösa hd	Stockholms
6	Uppsala	biskopsdöme	---	Allmänt	Uppsala
7	Uppsala	stift o. stad	Uppsala st.	---	Uppsala
8	Uppsala	stad o. stift	Uppsala st.	---	Uppsala
9	Uppsala	stad	Uppsala st.	---	Uppsala
10	Uppsala	stad	Uppsala st.	---	Uppsala
11	Uppsala	stad och stift	Uppsala st.	---	Uppsala
12	Uppsala	stad	Uppsala st.	---	Uppsala
13	?Uppsala(?)	stad	Uppsala st.	---	Uppsala
14	Uppsala	stad o. stift	Uppsala st.	---	Uppsala
15	Uppsala	stad och stift	Uppsala st.	---	Uppsala

Figure 5.2. Search results in the old version of the Swedish digital place-name register. Image by Björn Karlsson.

reliable, as many of the original Sami or Finnish names have been classified as Swedish.⁷ The archive posts that had references to literature were marked as containing an interpretation of the name, even though the post sometimes just refers to a description of the locality. All posts also were provided with a hidden identifier, a text string, in order to connect the scanned images of cards with the correct archive posts and to arrange the order of the digital archive posts; these identifiers are still used, but they are hidden from everyone except the software developers, and no new ones are created.

Each row in the digital table with archive posts represented one archive card in the physical collection; see Figure 5.2. If both sides of the card were used, each side had its own row in the table with archive posts. The physical card was visible in the form of a PDF copy (stored in a file system outside the database) whenever the user clicked on a name. Thus, the datings of the written records, the different spellings of the name, the pronunciation, and any other information on the physical card could be read by users, but no such information was transferred into the database.

The lack of complete digitized data caused problems updating the chronological order within each place-name when new information was being added. The only way not to disrupt the structure was to insert each new post manually, one by one. To import files with several hundred posts was thus out of the question. As a result, while the physical archive was being updated, the digital register paradoxically was not, and with every year that passed the divergence between the two grew bigger and bigger. The information that was added to the physical collection primarily concerned references

⁷ These classifications have been transferred to the new digital register since they at least give a hint of the possible different origins of place-names in Sweden. If we decide to change the labels, the changes will have to be made manually by an expert in Sami and Finnish, and we have so far not been able to prioritize this step.

to recently published literature. The absence of these references in the digital register meant that it could no longer be relied on for keeping up to date with current research. The users of the digital version were therefore put at an unfortunate disadvantage.

The need for a restructuring was not only due to the need for data updates but also, as shown, because of how the existing data were organized. Which of the posts actually denote the same locality could not be deduced from the data. When searching for the place-name "Uppsala," over 5,000 hits would be retrieved—a result that revealed the total number of archive posts for "Uppsala" in the database (or cards in the archive), but not how many different places named "Uppsala" there are. The hits could be sorted by parish and thus give a clue as to which posts denoted the same locality, but, on the whole, the system could not be said to be very practical.

Since it was evident that the existing flat structure would need to be replaced, the question arose how best to reorganize the digital material in a manner that would preserve all archive information without being conservative in its structure. It was clear that the best way to proceed was to cut the digital version loose from being an exact mirroring of the physical one and allow it to develop a structure of its own. The new structure had to facilitate maintaining, updating, and displaying the data, and it also needed to be forward-looking in order to pave the way for future developments and additions.

Second Attempt to Make the Collections (More) Available

What was missing in the first version of the Swedish digital place-name register was a hierarchical structure in which all archive posts concerned with one and the same place-name were bundled together under a common heading. The ability to single out each unique place-name, instead of each and every archive post concerning the same place-name, would simplify the search for relevant names and places, and also make it possible to connect the place-name (and hence all the archive posts related to that name) to point coordinates or a geographic area (a polygon).

Nevertheless, the data at hand, as described above, consisted only of a few fields and was not substantial enough, and so it was not possible to immediately implement the desired structure and identify which archive posts denoted the same place-names and therefore should be kept together. The standardized name form and the parish information helped make an initial connection, but the unique place-name could not be based on the geographical location only, since there are sometimes homonymous place-names within the same parish.⁸ A rough division in names of settlements and of natural features made it possible to split up homonymous names denoting widely different localities—e.g., "Lillsjön" in the parish of Grums, Värmland, which is to be regarded as two unique names, since the one "Lillsjön" denotes a lake and the other "Lillsjön" denotes a settlement (even though it is clear that the settlement name emanates from the lake

⁸ See Gammeltoft, this volume.

name). But this was not enough for a significant number of homonymous names, since the different names often belonged to the same group of names of settlements or of natural features. It became apparent that it was crucial to discern the kind of locality that the names are denoting, and the only information to help us along was the data field containing information about locality. As mentioned, this field contained numerous different words and abbreviations—e.g., the Swedish word *gård*, pl. *gårdar* (“farm”), could be rendered as *gård*, *g*, *g.*, *gd*, *gårdar*, *gdr*, *gdar*, *gd:ar*, *f.d. gd*, *nu försvunnen gård*, *förr gd*, etc.

First, we had to identify which words were used when denoting a certain locality. With every difference in spelling counted, the number of different words, abbreviations, etc., reached over 85,000 varieties. Unfortunately, the most common words were “non-existing” and “locality missing”—i.e., where there was either no definition of locality or an uncertainty about the locality (marked with a question mark), or where the persons transcribing the card had missed filling in the information in the right field. Next in line was the word *åker* (“field,” “arable land”), demonstrating how many of the names in the archive actually describe non-settlements. When the most common varieties were identified, they were clustered together under a few catch-all terms—i.e., “area” (Swedish: *område*),⁹ “artifact” (Swedish: *anläggning*),¹⁰ “ground” (Swedish: *terräng*), “land” (Swedish: *ägomark*),¹¹ “name element” (Swedish: *ortnamnselement*), “people” (Swedish: *personer*),¹² “settlement” (Swedish: *bebyggelse*), “water” (Swedish: *vatten*), and “watercourses” (Swedish: *vattendrag*). Being able to make this distinction, we could match the posts that had the same name, within the same parish and the same catch-all term for locality, and trust that most of the posts were bundled under the correct place-name. Some names came to be bundled together even though they denoted different places, however, and some were separated even though they indeed denoted the same place. This is something we will not be able to solve automatically but must attend to manually.

Having linked most of the posts concerning the same place-name, we finally reached the hierarchy we wanted. In the new digital register the top level containing the unique place-name is called *Ortnamn* (“Place-Name”) and all the archive posts belonging to the same place-name are found in the sub-level *Arkivposter* (“Archive Posts”). To ensure users can get an overview of relevant place-names, the locality type and parish (and higher administrative units) are displayed together with the place-name in the search results in the web application *Ortnamnsregistret* (“Place-Name Register”); see Figure 5.3. Users start by searching for a place-name (or parts of a place-name) and then go on to

⁹ Containing names denoting (parts of) parishes, counties, neighbourhoods, municipalities, etc.

¹⁰ Containing names denoting bridges, roads, ditches, gates, lighthouses, etc.

¹¹ The difference between “ground” and “land” is often confusing, but as a guideline “land” incorporates those names designating lands used in farming while “ground” tends to contain names of natural features.

¹² Even though the archive is a place-names archive it still holds local names for people connected to a place, and certain names emanating from places but considered personal names.

The screenshot shows the Ortnamnsregistret website interface. At the top, there is a header with the site logo, the name 'Ortnamnsregistret', a contact email 'namn@isof.se', and a user icon. Below the header is a search bar with the title 'Sök ortnamn'. It includes a dropdown for 'Anpassa sökning' (set to 'Exakt'), a text input field containing 'Uppsala', a 'Sök' button, and a help icon. Below the search bar, there are tabs for 'Ortnamn' and 'Arkivposter'. The main content area displays 'Antal sökträffar: 77' and a 'Sida 1' button. A list of search results follows, each showing a place name and its administrative region: 'Uppsala, bebyggelse, Ytterby sn, Inlands Södre hd, Göteborgs och Bohus län, Bohuslän', 'Uppsala, terräng, Dalskogs sn, Nordals hd, Älvsborgs län, Dalsland', 'Uppsala, ägomark, Ockelbo sn, Gästriklands v. d:a tg, Gävleborgs län, Gästrikland', 'Uppsala, terräng, Ovensjö sn, Gästriklands v. d:a tg, Gävleborgs län, Gästrikland', and 'Uppsala, bebyggelse, Hammars sn, Sundbo hd, Örebro län, Närke'.

Figure 5.3. The new Swedish digital place-name register, displaying the results of the simple search for “Uppsala.” Image by Björn Karlsson.

view the archive posts that belong to the chosen place-name.¹³ The archive posts are listed according to information type: “written record” (Swedish: *belägg*)—i.e., older spellings of the name in handwritten sources;¹⁴ “oral record” (Swedish: *uppteckning*)—i.e., the pronunciation of the name written in Swedish dialect alphabet; and “references” (Swedish: *litteraturhänvisning*) to literature that in some way deals with the name or place. Information about place-names and archive posts are continually being added to the digital place-name collection, but, due to lack of resources and of space,¹⁵ no more cards are added to the physical place-name archive.

The new Swedish digital place-name register was first tested within Isöf’s own staff, and with a few handpicked experts. After the comments and suggestions had been taken into consideration and applied, we finally launched a beta version in March 2020,

¹³ It is also possible to view all the place-names belonging to a certain parish, but then some adjustments to the search fields are necessary.

¹⁴ There is an ongoing project to display all the written records in chronological order, displaying the year of the record together with each archive post.

¹⁵ And also a belief that it is unnecessary and uneconomical to keep double versions of the same information.

accompanied by a REST (REpresentational State Transfer) web service.¹⁶ What is much appreciated in the new version is that, instead of the over 5,000 hits shown in the old register (exhibiting how many cards with that name that existed in the archive), a search for “Uppsala” in the new database results in seventy-seven matches corresponding to the number of locations carrying this name that are recorded in the place-name collection. In other words, each of these seventy-seven places bears the same name, and therefore the result shows seventy-seven place-names (i.e., one place equals one place-name, even though the names of two or more places may be identical). Linked to each place-name are archive posts showing every instance (recorded in the collection) of the specific name used for the specific place.

The How and Why of the New Digital Register

The new Swedish digital place-name register has more components than the first one: the publicly available web application that makes it possible to search the database; the editing tool that makes it possible to edit existing archive posts and add new place-names and/or archive posts; and a REST web service. The data are stored in a MariaDB database and Elasticsearch is used for searching. Figure 5.4 illustrates interrelations in the system. In the following section we describe the system in more detail.

MariaDB Database

The MariaDB database is a digital copy of the physical place-names collection when it comes to the card register, but with an architecture and relations stretching far beyond the potentials of the physical archive; see Figure 5.5. All data in the Swedish digital place-name register are in Swedish only.

In Figure 5.5, each row in the table “PlaceName” represents a unique place-name. The goal is that there should not be two or more rows that represent the same place-name, but that goal has not yet been fulfilled. It is not easy to extract the unique set of place-names from 3.7 million archive posts. The physical card uses the official spelling of the place-name, but, since the collections have been added to for almost a century and some official names have changed, there are still some place-names that have two or more different spellings (and therefore are two or more different place-names until archivists manually bring them together to one row).

A place-name is usually related to a specific parish, but some place-names are not. For example, the name of a district is not related to a specific parish, which is only logical, since the district is found at a higher administrative level and consists of many parishes. Those place-names that are not related to a specific parish are, however, related to a row in the table “District;” see Figure 5.5. Today over 90 percent of the place-names have the related parish as a coordinate source. This means that the coordinates shown for the place-name are the same as the coordinates for the parish that the place-name

16 See <https://PlaceNameService.isof.se> (accessed June 7, 2021).

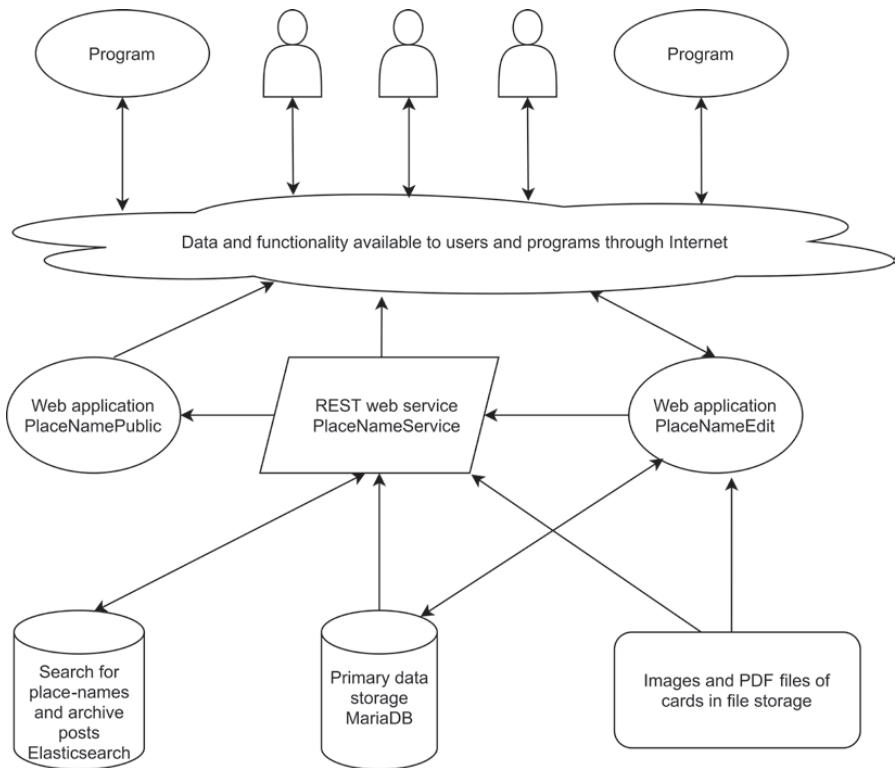


Figure 5.4. System description. Image by Björn Karlsson.

belongs to. The church is often used as a position for parishes, but sometimes it is a geographically central position. The table “PlaceName” has a string with the field “type,” which holds information about the kind of locality the name belongs to—e.g., “settlement” or “watercourse.”

Most rows in the table “ArchivePost” (see Figure 5.5) are related to a physical card. If both sides of a card have been used, two rows exist in the table, one for each side of the card. New archive posts that are added to the place-name collection do not have a corresponding card, which of course also means that there is no image or PDF file of a card. Each archive post is related to one place-name. Each archive post is also related to exactly one so-called project. The word “project” is used in the editing tool PlaceNameEdit (see below) for clearly defined and delimited information sources. For example, if we do an on-site interview in a parish, that would be a project, and, if we were to make excerpts from an old source that has not yet been added to the register, that would be another project. In addition, each book or volume, etc., that we extract examples from becomes a project of its own. It is only when a project is published in the editing tool that the information is available to external users. Each project contains only

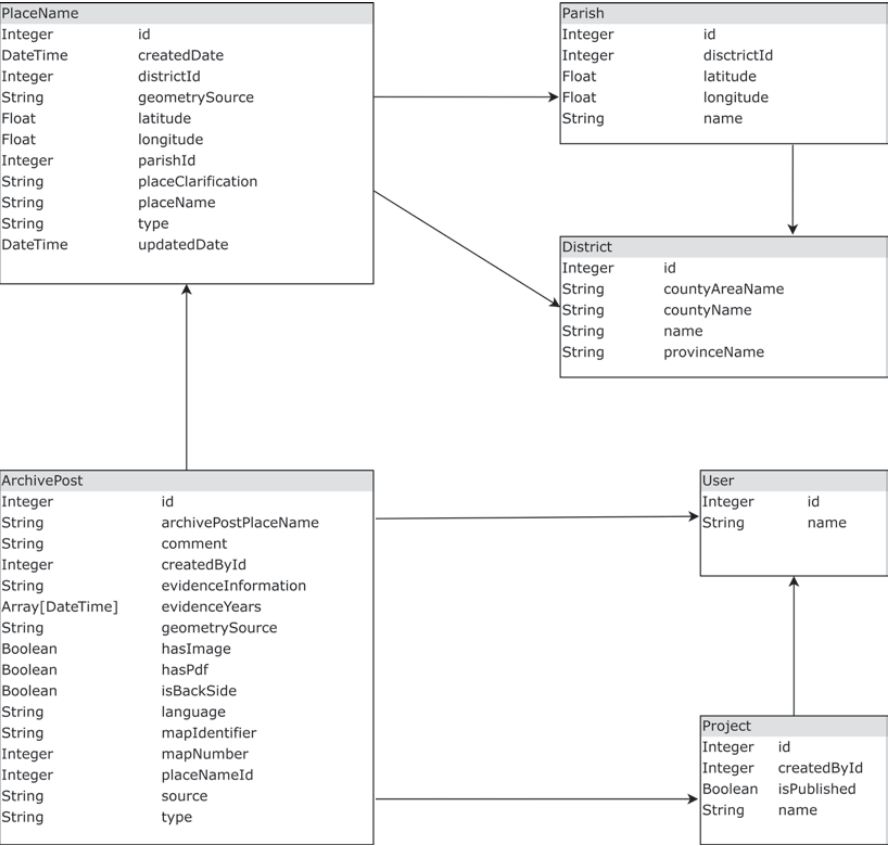


Figure 5.5. The design of the Swedish digital place-name register. Image by Björn Karlsson.

archive posts of the same type. There are, as mentioned above, three types of archive posts: written records, oral records, and references.

Each parish is related to a single district. The table “District,” as shown in Figure 5.5, contains information about districts, counties and provinces in Sweden.¹⁷

Additionally, Isot uses so-called history tables for the most important data types and those data types that change often. Each history table is related to exactly one of the standard tables and has the same columns as the standard table plus some extra columns that contain extra information about changes. A history table contains the complete history of changes (including current values) that the standard tables have

¹⁷ It also contains information about countries or areas other than Sweden. These countries and areas (except for Sweden) have no information in the fields “name,” “countyName,” and “provinceName” in the table “District.”

had over time. They are very useful for amending data-processing errors. The technique trigger, which is available in most SQL databases, is used to update history tables with changes. An advantage of using triggers is that all changes are caught in the history tables, and not just changes that are made by programs such as PlaceNameEdit (see below) but also changes that are made by software developers using SQL statements directly to the database. The updating of history tables should be as simple and fast as possible, and we therefore try to avoid the use of foreign keys or indexes in history tables, which could cause slow database changes.

Currently, much of the information in the database is identified by a serial number that ranges from one and upward for each data type. This is not an ideal identification for data, and it should be improved to at least some kind of globally unique identifier in the future.

Elasticsearch Database

We have two different databases: MariaDB and the search database Elasticsearch.¹⁸ This takes extra resources but has been judged to be worthwhile because of the characteristics that Elasticsearch has. It is fast and has many different search alternatives, which we use in the new Swedish digital place-name register.¹⁹

Elasticsearch is not an SQL database in which data is stored in fixed tables with fixed columns. Instead, data are stored as documents, and JSON is used for the input and output of documents. A “document” is similar to a row in a table in an SQL database. In Elasticsearch, documents representing the same type of data (for example, place-names) are similar to a table in an SQL database. Documents contain fields where each field has a name and a value. The value is of a certain data type, such as text, numbers, Boolean value, etc. One field in a document is similar to the value of one column in one row in an SQL database. The main difference is that documents are more flexible. Elasticsearch does not complain if you add a new field to a document, and it is common that fields contain structured data with its own fields.

Two types of documents are stored in Elasticsearch: place-names and archive posts. The reason for this duplication of data is that Elasticsearch has search functionality that is not found in MariaDB. Both types of documents have extra information added in Elasticsearch compared to the information stored in MariaDB for the same data type. The reason for this is that we have decided to make each document in Elasticsearch self-explanatory without the need of extra information. This makes it easier and faster to return requested data from Elasticsearch without the need to aggregate information from several data types. Some fields in the place-name and archive post documents

18 It is built on top of Apache Lucene, which handles the data storage.

19 Elasticsearch is also fault-tolerant, distributed, and scalable, but these characteristics are not used today. Currently, Elasticsearch is installed on only one server with no redundancy. If usage of the Swedish place-name register grows in the future, however, we are well prepared to increase capacities.

have been added only to be able to apply specific search criteria. This information is not returned out of PlaceNameService (see below). One example of this is that place-name documents in Elasticsearch contain a field with information about the types of archive posts that belong to the place-name.

It is possible to store documents in Elasticsearch without specifying the data type that exists in the fields in the document. Elasticsearch tries to infer the data type for a new field from the first JSON document containing that field that is inserted into Elasticsearch. In reality, the data type is often defined in advance in order to get the desired behaviour from Elasticsearch. For example, both sorting and searching in Swedish alphabetical order require that the user specify the data type for text fields in a specific way. Default language for text fields in Elasticsearch is English, which does not have the Swedish characters *å*, *ä*, and *ö*. If support is needed for other languages than English in Elasticsearch (version 7.X), the user has to install the Elasticsearch plugin *analysis-icu* on the computer where Elasticsearch is installed.

The Public Application

Most of the information²⁰ in the place-name database is publicly available through an online application called Ortnamnsregistret (Place-Name Register).²¹ PlaceNamePublic is the informal name used during system development. To avoid any misunderstandings we will henceforth use the name PlaceNamePublic when describing the new digital public register.

In PlaceNamePublic, the same searches made in the first digital register are possible, but the new hierarchical structure is visible in the search result, and some new features are available. As with the first digital register, the user can search in the whole of the database for a certain place-name. The user can also refine the search through different search criteria if the area and the type of information are known. The search will be conducted both within place-names and archive posts, making it possible to find alternative names and older official names that are not in use today.

To start with, a simple search box is visible; see Figure 5.6. There is a blank field into which the user types a name or part of a name or any combination of letters (where the Swedish word “Sökord” [“key word”] is visible before entering any letters). The letters the user types become a search string that is matched against the place-names in the database. Already at this stage, by using the drop-down menu to the left, search strings can be directed to match with eight different alternatives: “Börjar med” (“Begins with”), “Innehåller” (“Contains”), “Slutar med” (“Ends with”), “Liknande” (“Similar”), “Exakt” (“Exact”), “Jokertecken” (“Wildcard search”), “Id för ortnamn” (“Place-name ID”), and “Id för arkivpost” (“Archive post ID”). The “Similar” search makes it possible to search for

²⁰ Projects in the editing tool are not publicly available until they are approved by a place-name expert. There are also some posts in the database with more information, but these are very few and have been regarded as unnecessary to present in the public application.

²¹ See <https://ortnamnsregistret.isof.se> (accessed June 7, 2021).

Ortnamnsregistret

Kontakt: namn@isof.se

Sök ortnamn

Anpassa sökning Ange ortnamn eller ortnamnselement

Innehåller ▼ Sökord Sök ?

Börjar med

Innehåller m myndigheten Isof > Tillgänglighetsredogörelse >

Slutar med

Liknande

Exakt

Jokertecken

Id för ortnamn

Id för arkivpost

Figure 5.6. The simple search entry function in the Swedish digital place-name register.
Image by Björn Karlsson.

place-names that are difficult to spell or that have changed over the years. For example, if the user searches for “Jöteborg,” the search result will also show “Göteborg.” In “Wildcard search” the character “?” matches exactly one character and the character “*” matches zero or more characters. “Wildcard search” can be used to get statistics about how common different name parts or name elements are within place-names. For example, “Wildcard search” with the search string “*sjö*berg*” (*sjö* is the Swedish word for “lake” and *berg* is the Swedish word for “mountain”) will show how many place-names include the elements *sjö* and *berg* regardless of whether the names have any letters before *sjö* or in between *sjö* and *berg* (e.g., a genitive *s* or a plural *ar*) or if the names end with a simple *berg* or with a definite article, etc. The search for “Place-name ID” and “Archive post ID” makes it possible to search for an exact hit—i.e., when making a reference to a specific place-name or a specific archive post.

In order to refine search results, users can press the “plus” sign or click on the link that reads “Fler sökalternativ” (“More options”) at the right side of the web page. Then the search can be limited geographically (by choosing parish, district, county, province, and country); by language (but, as explained above, a major portion of the names will be counted as Swedish even though they might have a different origin); by name type (where only the choice between “settlement names,” “nature names” [i.e., names of natural features], “name elements,” and “personal denotations” can be made); and, finally, by the type of archive posts that are connected to the place-name (the options being “written records,” “oral records,” or “references”). If the user selects a specific type of archive post,

this means that the place-name in the search result has at least one archive post of this type. Any combination of these search criteria can be used when searching for place-names.

It is also possible to leave the initial search box empty, unfold the expanded search box, and search for all the names in a certain parish, or all the names of natural features that have archive cards that are oral records, and so on.

When a place-name has been entered into the search field, the search may be adjusted in line with the desired criteria, and the related place-names are found in a list below the search. Since the search is based on the hierarchy described in the previous section, the user gets a more precise idea of how many places, in Sweden, are known by the same name. Moreover, all the catch-all terms for locality and the locating administrative units are visible in the search results. It is also possible to download the search result as an Excel file containing information about locality (both the new catch-all term and the original designation from the first digital version), the parish and the other levels of administrative units and country, and language, as well as a unique ID number referring to the place-name (not the archive posts).

Each unique place-name is found in a single row, and, by clicking on the name, the user can view all the archive posts belonging to the chosen place-name. If the user simply wants to view more detailed information on the chosen place-name, the user can use the “plus” sign to the right in the same row.

The archive posts are displayed in a predefined order in case there are posts of all three types, the order being: “written records,” “references,” and “oral records.” The written records are sorted by the year when the source was written. In case there are any spelling differences, the archive posts are sorted alphabetically.

One of the new features that has been appreciated and utilized by our users is the function to send comments about specific archive posts to us. When viewing a certain archive card the user can choose to push the button that reads “? Frågor och synpunkter” (“Questions and remarks”), and then a fill-in form appears into which the user can type a question or a comment (which is mandatory), an email address (optional), and a phone number (also optional). The form contains information about the actual archive post. The staff working with the register (and questions about place-names) receive an automatically generated email with all information needed to respond to the comments or simply correct mistakes.

We provide documentation about the PlaceNamePublic resource, such as how PlaceNamePublic works, with a description of how to search for names and how to process the information on the archive cards, as well as on PlaceNamePublic’s structure and background. There are also some shorter help texts found next to most search criteria. The documentation and web application interface are available only in Swedish.

Internal Web Application for Data Maintenance

In the first digital register it was not possible to add information that would break up the structure of the already existing posts. The ability to add information is crucial for remaining an attractive source of knowledge, however. So, when given the chance to remake the whole system of the digital place-names register, the idea of an independent

editing tool emerged, and hence the internal web application PlaceNameEdit was born. In the application the user can edit information about place-names, archive posts, parishes, districts, and projects (where the new information can be defined to a certain resource), as well as manage the users of the application itself. The tool is developed for place-name experts within Isof, which means that no software-developing skills are needed to add or edit information. The web application is developed by Isof's software developers, and most of the web application's user interface was generated with the Django Baker tool.

Changes in place-names and archive posts are collected within projects (see the "MariaDB Database" section above). Information belonging to a project may be publicly visible or not depending on the current status for the project; the status is changed by the place-name experts. This means that it is possible to publish a set of updates at the same time, but it also means that others can check the added information before it is made publicly available.

REST Web Service

PlaceNamePublic receives all place-name information through a REST web service called PlaceNameService. PlaceNamePublic or any future client applications have no knowledge of where and how data are stored, which is a desirable feature as it makes it easier to change data storage with little or no impact on client applications. PlaceNameService makes place-name information available to anyone who wants to include the data in their own application.

Swagger has been implemented in PlaceNameService.²² Swagger extends the functionality of a web service and makes it possible to test the web service in a web browser and also view information about the web service REST API. REST web services normally return data in JSON format or resources such as images or documents in their formats. Documentation about PlaceNameService, http addresses, and parameters can be found with the help of Swagger. Returned data are not yet documented in Swagger but can be viewed by testing the web service functionality with Swagger.

There is much resemblance between PlaceNamePublic's functions and PlaceNameService's functions, but with the latter there is also the possibility to receive information about types of data that can be used for searching for place-names, the status for PlaceNameService, the OpenAPI Specification schema for PlaceNameService,²³ and copying information about place-names and archive posts from MariaDB to Elasticsearch (this last functionality is only for internal usage at Isof and is not described in Swagger).

²² Enter <https://PlaceNameService.isof.se> in a web browser to view the Swagger interface of PlaceNameService (accessed June 7, 2021).

²³ OpenAPI Specification of PlaceNameService can be retrieved in two different formats. OpenAPI Specification in JSON format is retrieved by the web address <https://PlaceNameService.isof.se/openapi-schema-json> (accessed June 7, 2021). OpenAPI Specification in YAML format is retrieved by the web address <https://PlaceNameService.isof.se/openapi-schema-yaml> (accessed June 7, 2021). OpenAPI Specifications can be used by software development tools to automatically generate client side code in different programming languages. This makes it easier to construct software that communicates with REST web services.

PlaceNameService is currently available to anyone without requiring any identification of the user of the web service. This may change if needed in the future. The documentation and REST web service interface are in English.

Even though we do not know what will happen in the future, we can be sure that things will change. Therefore, PlaceNameService uses two different techniques that will help in the future to handle changes in the web service functionality and data.

PlaceNameService has a version number in the web address; the “v1” part in the web address <https://PlaceNameService.isof.se/v1/geo-reference/provinces> is the version number. If we need to make changes (that are not compatible with version “v1”) to the web service API, a new version “v2” will be created. Versions “v1” and “v2” would both be supported by PlaceNameService during a transitional period until all programs that use version “v1” have been updated to version “v2.”

Most of the returned data in the JSON format have a field named “fields” that currently contains an empty array. This array may in the future be populated with objects whereby each instance represents one value associated with the type of data that it belongs to. Adding fields to the JSON data object directly on the highest level will result in problems for client applications to PlaceNameService that use hard-typed programming languages.

Most returned data contain a field named “identifier” that currently has no meaningful value. This field should in the future be used for some kind of globally unique identifier for each type of data. Due to a lack of resources it has not yet been realized, but it is a prioritized next step.

System and Data Management

Changes in place-name-related information are normally made by place-name experts who use the web application PlaceNameEdit. Occasionally software developers help the place-name experts to make the same change to a large dataset directly into the database by using SQL statements.

One of the place-name experts is designated to communicate with the software developers about improvements to existing functionality and the development of new functionality in the place-name register system. This place-name expert gathers responses from public users and place-name experts alike and makes decisions about the changes that should be implemented next.

All information (data in databases, software and files showing cards) in the Swedish digital place-name register is copied once a day to a backup. This ensures that the register can be restored in case of hardware or software problems.

Software changes all the time. Most software is updated at least a couple of times each year. It is the software developers’ and system administrators’ responsibility to periodically update the software that is used to construct the programs.²⁴ If software is not

24 Only those techniques that are most relevant in this context are listed here. All these techniques except for Red Hat are free to use without costs: Ubuntu 20.X, a Linux operating system on

updated, security risks and maintenance problems will increase with time. Sometimes updates of external software result in changes in the code of Isof's applications.

Moving forward

Even though we have accomplished a major improvement to the Swedish digital place-name register, there are still many more changes we wish to make, both on the technical side and from a user (internal and external alike) point of view.

One of the prioritized next steps is to include another place-name collection within Isof, specialized in Scanian place-names, the so-called Skånsk ortnamnsdatabas (Scanian place-name database).²⁵ The physical archive is located in Lund and was digitized at a later stage. Some lessons had been learned from the earlier digitization project, and more information was transferred from the physical card to the digital system. This causes some problems, however, when trying to merge the two databases. The Scanian place-name collection has information about approximately 800,000 archive posts.²⁶ A large quantity of place-name material from the Scanian database is not found in the general register, but there are also posts in the larger register that are not found in the Scanian database, mostly recently added references to literature. We still have not resolved how to overcome these challenges, but our aim is to incorporate the Scanian database into the place-name register so that users will no longer need to search in two different web applications.

The new register makes it possible to add more exact coordinates and polygons, and we hope to be able to add a map to make all the names' distribution visible. There are developments within Isof in fields other than place-names that make use of a standardized map, and we are exploring the possibilities of developing this map to be used for our register as well. There are steps that need to be taken before we can show the names on a map, however, such as where to place the name considering that

development computer; Red Hat 7, a Linux operating system, on test and production servers; PyCharm Community Edition, a development tool, on development computer; Python 3.X, a programming language; Django 3.2.X, a framework for developing Python programs; Django Baker, a tool that automatically generates a web application, has created much of the functionality in PlaceNameEdit; input to Django Baker is a description of the tables in the database; Django REST framework, a tool for developing REST web services with Python, is used in PlaceNameService; Django REST Swagger, a tool for implementing Swagger in REST web services developed with Django, is used in PlaceNameService; Python package elasticsearch is used in PlaceNameService; the Python package Pandas is used in PlaceNamePublic to create Excel files; Elasticsearch, a fast and flexible search tool where data are stored as documents; MariaDB, a SQL database, on test and production servers; Gogs, a version control system that is locally installed and based on Git; HTML, CSS, and JavaScript are the basic techniques for building web applications; jQuery is used to add functionality to web applications; and Bootstrap 4.X is used to make PlaceNamePublic responsive to different screen sizes and to add visual elements to the web application.

25 Skånsk ortnamnsdatabas, <https://www3.isof.se/beb> (accessed June 1, 2021).

26 And these 800,000 posts are just the digitized ones; approximately 200,000 archive posts have not yet been digitized.

names sometimes “move,” accounting for the fact that named places grow in different directions during the time we have recordings, or discovering how to automatically use other sources that already have assigned coordinates.

Another improvement we wish to make is to digitize more of the information on the physical card. With more information digitized, new searches can be done, and other ways to arrange the material may help discover new findings. One way of digitizing the physical cards could be to register data by using optical character recognition (OCR) and handwritten text recognition (HTR) technologies. A fair share of the cards are machine-written, but they are still outnumbered by the many handwritten ones. There are tools such as Transkribus²⁷ that can be trained to transcribe handwritten text, but the cards are written by many different hands, and the amount of text on the cards is sparse—not nearly enough to train a tool such as Transkribus. Even if it was possible to gather enough text by the same hand, the fact is that the cards were scanned at only 200 dpi, making it harder for the OCR or HTR technologies to actually decipher the text. Another reason for adding more of the information found on the cards is that people in general are becoming less and less used to reading handwritten texts, and even relatively neat and legible handwriting will no doubt prove difficult to decipher for many users in the future.

Aside from the information on the physical cards, the archive also contains other kinds of documents that have so far not been made publicly available. Perhaps the most important of these are maps documenting the locations of recorded place-names. Also very useful are the compilations of settlement names found in the province records and the land registers found at the Swedish National Archives, covering the period between the early sixteenth century and the late nineteenth century. We are still exploring ways to make these assets digitally available, but our hope is that they can be connected and be searchable from the place-name register.

On the line of more information, we would like to enhance the Swagger interface so that the users may find explanations as to the kind of information the returned data consist of. Today users can find information about how data are retrieved, but no explanations about the returned data are given.

To reduce complexity in the system design and decrease maintenance costs, we would like to use only Elasticsearch for data storage. Likewise, PlaceNameEdit should use only PlaceNameService for all data handling, including access to images and PDF files. This would result in fewer dependencies between system components. Fewer dependencies imply easier administration and easier further development of the system. The editing tool should also be made more powerful and flexible so that the users can easily carry out most of the changes that they need to do themselves.

On a smaller scale, we aim to create a definition of the geographical regions that is consistent with the book *Årsbok för Sveriges kommuner 34*, published in 1951.²⁸ Today there are some differences between the book and the definitions of geographical regions

27 See <https://readcoop.eu/transkribus/?sc=Transkribus> (accessed June 7, 2021).

28 Statistiska centralbyrån, *Årsbok för Sveriges kommuner 34*.

in the Swedish place-name collection. When we have a correct definition we would like to move on to create exact coordinates or polygons for all the place-names. This would enable us to show the names on maps and generate a better understanding of the distribution of different names.

There is definitely a need to add some kind of globally unique identifiers to all pieces of non-trivial data. This should have been done before the introduction of the new place-name register, but time was the limiting factor. Another useful step would be to include the Swedish place-name collection into the linked open data (LOD) structure. In the REST web service, we plan to add support for place-name information downloading. This should make it possible for others to create their own copies of the Swedish place-name collection.

In the public web application PlaceNamePublic, the needs of place-name experts and public users decide the functionality that should be added next. The development project for the first version of PlaceNamePublic did not have enough resources to perform extensive user tests, but user tests are one of the best ways to improve the usability of software, and we hope to be able to perform such tests in the future. There is also a need for more automatic tests within the REST web service PlaceNameService.

As has been seen, we would like to meet our users' expectations. The register does not contain the one thing that many users expect to find, namely etymologies and interpretations of place-names. There are indeed references to advanced onomastic literature, where many etymologies can be found, but these are often not easily accessible to the public, and some references are written with so many abbreviations that the reference is intelligible only to the onomasticians at Isolf. There is an onomastic series at Isolf, however, called *Sveriges ortnamn* ("Place-Names of Sweden"),²⁹ that strives to give a possible interpretation to all the names in our place-name collection. This is of course a huge task, and perhaps a mission impossible, but some volumes have been produced covering different parts of Sweden. In the interdisciplinary project TORA we tried to digitize one of the volumes in the series.³⁰ We learned much, and hopefully we can use this experience and make a connection between the onomastic series and the digital place-names register so that interpretations can be found in PlaceNamePublic.

There is also a need to enrich our collection. Today we cannot enrol students and the like to gather place-names for us, but have to rely on local heritage societies and other laymen interested in place-names to gather more information on site or from older, not previously extracted, records. We therefore want to create a web application that makes it possible for anyone to contribute to the Swedish place-name collection. Isolf already has a web application³¹ on which crowdsourcing is used to transcribe pieces of folklore on cards into digital information. One possibility is to use the same web application, but with other fields, to collect more information from the physical place-name cards into

29 Institutet för språk och folkminnen, "Sveriges ortnamn."

30 See Karsvall, this volume.

31 The web application is called Folke and can be reached at <https://sok.folke.isolf.se>.

the corresponding archive posts. The crowdsourcing contribution can either improve the quality of existing data or add new place-name data. Hopefully, it will also be possible to add new types of content, such as pictures of the place in question, or sound files displaying the modern and local pronunciation. The information must be delivered in a controlled manner; however, otherwise the data quality may decline.

It is to be hoped that all these suggestions—and more—will be carried out, and users of the digital place-name register will be able to search and find much more information than was possible previously, delivered in a more structured and intelligible way, and also to download and create their own databases and presentations of place-names. Moreover, with a crowdsourcing tool, we hope that the public will want to participate in creating a richer and more reliable place-name collection for future onomasticians and lay users.

Bibliography

- Lundell, Johan August. "The Swedish Dialect Alphabet." *Studia Neophilologica* 1 (1928): 1–17.
- Ortnamnsregistret (Swedish Digital Place-Name Register), first version. www4.isof.se/NAU-ortnamn/index.htm (accessed June 7, 2021).
- Ortnamnsregistret (Swedish Digital Place-Name Register), second version. <https://ortnamnsregistret.isof.se> (accessed June 7, 2021).
- PlaceNameService. <https://PlaceNameService.isof.se> (accessed June 7, 2021).
- Rygh, Oluf. *Norske Gaardnavne: Oplysninger samlede til Brug ved Matrikelens Revision efter offentlig Foranstaltning udg. med tilføjede Forklaringer. Forord og Indledning*. 19 vols. Kristiania (Oslo): I Kommission hos Cammermeyer, 1887–1936.
- Skånsk ortnamnsdatabas. www3.isof.se/beb (accessed June 7, 2021).
- Statistiska centralbyrån. *Årsbok för Sveriges kommuner 34*. Stockholm: Kungl. boktryckeriet P. A. Norstedt & söner, 1951.
- Sveriges ortnamn*. Series A, 77 vols. Uppsala: Institutet för språk och folkminnen, 1906–. www.isof.se/sprak/namn/ortnamn/sveriges-ortnamn.html (accessed June 7, 2021) [earlier: På offentligt uppdrag utg. av Kungl. ortnamnskommissionen. *Sveriges ortnamn*. Uppsala: Lundequistska bokh].
- TORA – Topografiskt register på Riksarkivet (Topographical Register at the Swedish National Archives). <https://riksarkivet.se/tora> and <https://tora.entryscape.net/wp> (both accessed June 7, 2021).
- Transkribus. <https://readcoop.eu/transkribus/?sc=Transkribus> (accessed June 7, 2021).