

Chapter 2

Dutch Historical Toponyms in the Semantic Web

Ivo Zandhuis, Menno den Engelse and Edward Mac Gillavry

Abstract The standardisation of historical, geographical names or toponyms facilitates both the comparison and combination of datasets that have a spatial dimension. In a digital infrastructure, online web services can be provided that not only standardise the historical, geographic names in these datasets to a canonical name, but also enrich the dataset with geographical coordinates through a process of geocoding. Finally, these online web services can be used to create cartographic visualisations of the datasets. Since there were no convenient web services available for the standardisation of historical, geographic names in the Netherlands, the websites gemeentegeschiedenis.nl and histopo.nl were launched in 2013. Gemeentegeschiedenis.nl presents a uniquely identifiable web page (using a so-called “Uniform Resource Identifier” or URI) for every municipality in the Netherlands since 1812. The web pages provide relations between former and current municipalities and present maps of all the boundary changes over time. The website histopo.nl was launched to provide a disambiguation service for toponyms of settlements and their historical spelling variants. Both gemeentegeschiedenis.nl en histopo.nl present the information as web sites for regular visitors and as web services for computers to be incorporated in an online information system. The aims of the introduction of gemeentegeschiedenis.nl and histopo.nl are to demonstrate the advantages of a standardisation service for historical, geographic names on the one hand and building up our own expertise and experience on the other hand. In this chapter we present a description and evaluation of the information and centralised services presented on these websites and a description of potential services to create a more useful tool.

I. Zandhuis (✉) · M. den Engelse · E. Mac Gillavry
Rijksstraatweg 79, 2024 DB Haarlem, The Netherlands
e-mail: ivo@zandhuis.nl

M. den Engelse
e-mail: menno@islandsofmeaning.nl

E. Mac Gillavry
e-mail: edward@webmapper.net

2.1 Introduction

2.1.1 Problem

Almost every historical research question has a geographical dimension. When datasets are created, this geographical dimension is often reflected by the introduction of a field containing a geographical name.

When this geographical name or toponym is standardised, historians can compare their dataset with other datasets. This leads to additional information that can be used in the analysis. Standardised toponyms support the creation of an aggregated view on the data, for instance by means of geographically clustering records into larger geographical units. This enables the comparison with knowledge obtained from other resources. Visualisations of the dataset facilitate the historian in interpreting data.

In order to support the historian in the comparison or combination of datasets, and analyse it by aggregation or visualisation, at some point in time the historian wants to create a standardised form of the geographical name that is found in his source. This could be done at the moment of creation of the information, but maybe the standardisation is realised in a later stadium.

The construction, combination, comparison or visualisation of datasets can be realised with the historian's favourite data manipulation tool. The use of a centralised source of the standard in this tool prevents that an older version of inferior quality is used.

2.1.2 Presented Solution

An online web service combining and presenting all the geographical standards in a coherent and consistent way could prove a big help in the standardisation of a dataset and during the interpretation.

For the Dutch context we have introduced two websites: *gemeentegeshiedenis.nl* and *histopo.nl*. *Gemeentegeshiedenis.nl* presents information about municipalities in the Netherlands, hence the name “gemeentegeshiedenis”, meaning “municipal history”. We aimed to clearly distinguish between a platform for the administrative units and other types of *historical toponyms*, and therefore we introduced a second website called *histopo.nl* that primarily focuses on settlements.

2.1.3 Gemeentegeshiedenis.nl

Gemeentegeshiedenis.nl has two main characteristics.

First, the primary geographical entity is the second order administrative area, i.e. “municipality” in the Netherlands. *Gemeentegeshiedenis.nl* holds all consecutive

stages of the Dutch municipalities, from 1812 until now. The scope of geographical entities within *gemeentegeschiedenis.nl* is therefore strictly defined. Municipalities were instituted by law and geographic changes to their boundaries, mergers and separations all require changes in the law. Thus, the contents are more or less complete already and have been compiled based on datasets that have been released as open data or that have been provided by various parties to this project. Hence, the geographic representation of these administrative areas used in our service platform is a two-dimensional polygon instead of a one-dimensional point location.

The second aspect of this initiative is its ability to interrelate diverse, *de facto* standards for geographical names. A standardised name in one standard is related to the same name in another standard.

2.1.4 *Histopo.nl*

In contrast to municipalities, the history of settlements is not governed by law and their rise and fall often goes unrecorded. Because of their specific characteristics, we introduced the separate website *histopo.nl*.

Histopo.nl consists of a database with names for all kinds of historical settlements including administrative units, with all available spelling variants through time. The spelling variants resolve into a standardised, canonical name.

2.1.5 *Purposes of the Websites*

By inter-relating these standards we introduce the historical dimension into the international standardisation of contemporary geographical names, combining the best of both worlds. On the one hand, the websites we developed aim to present information about the historical development of municipalities and settlements for a broader public. On the other hand, the websites could facilitate researchers in the humanities by means of more elaborate search and standardisation functionality. Therefore, the websites provide geographical services for cultural heritage institutes and research in the humanities dealing with historical toponyms.

Thus, the websites are able to assist historians who aim to standardise the historical toponyms in their data set. Besides that, the website can play a role in more elaborate data entry projects and is able to standardise historical toponyms semi-automatically. After creation, the data can be used for instance by converting one coding standard for historical toponyms into another. Or data related to municipalities or settlements can be mapped geographically using thematic mapping techniques in order to reveal spatio-temporal patterns in datasets.

In this chapter, we describe the online geographical services we provide. We evaluate the usage and the quality and quantity of data and services. After that we elaborate on the potential of the platform we anticipate.

2.2 Prior Work and Development Elsewhere

The importance of the standardisation of historical toponyms is broadly recognised. There are various (online) resources for historical place names. In this chapter we focus on a relevant selection. For a broader discussion of resources for historical toponyms, see Southall et al. (2011).

2.2.1 *Standards for Toponyms in Present Time*

The Getty Thesaurus of Geographic Names,¹ generally referred to as “TGN”, is a structured vocabulary currently containing over two million names and other information about places. These places include not only administrative, political entities (e.g. cities, nations), but also physical features (e.g. mountains, rivers). Both current and historical places are included. The temporal coverage of the TGN ranges from prehistory to the present and the scope is global. While many records in TGN include geographic coordinates, these coordinates are approximate and are intended for reference only. These geographic coordinates in TGN typically represent a single point, corresponding to a point in or near the centre of the inhabited place, political entity, or physical feature.

A widely adopted open dataset for geographical entities is the Geonames Geographical Database.² The data set has been created through crowd-sourcing and contains all kinds of geographical entities, each with a unique identifier. One category of entities is the “administrative division”. The second order administrative division is equivalent to the municipality in the Amsterdam Code. However, there are very few historical municipalities available in the data set.

Closely related to Geonames is Wikipedia,³ in the sense that collecting and maintaining information about toponyms in Wikipedia is also a community effort. There are numerous web pages for geographical entities, amongst others municipalities. Due to its nature, Wikipedia does not provide any indication about the completeness and accuracy of the contents. Most data in Wikipedia is also available as Linked Data via the Semantic Web portal Dbpedia.⁴

¹<http://www.getty.edu/research/tools/vocabularies/tgn/>

²<http://www.geonames.org>

³<http://www.wikipedia.org>

⁴<http://www.dbpedia.org>

2.2.2 *Websites with Historical Toponyms*

Various initiatives are developed to augment the contemporary view on toponyms with a historical dimension.

The community-built gazetteer of ancient places Pleiades⁵ for instance, gives scholars, students and enthusiasts worldwide the ability to use, create, share, and map historical geographic information primarily about the ancient world.

In the United Kingdom the main entry into historical toponyms is The Historical Gazetteer of England's Place-Names.⁶ All British toponyms surveyed by the English Place-name Society are uniquely identified with a URI and described in terms of their relation with an upper geographic division. Future objective is to provide the data not only in a web page but in a machine-readable format as well.

For Danish toponyms the Digitalt atlas over Danmarks historisk-administrative geografi (DigDag)⁷ is constructed with presentation of the corresponding geographic polygons. Searching a place and its upper geographic divisions is enabled through a click on the map. Web services are available to obtain data about toponyms, including the geographical polygons.

2.2.3 *Dutch Toponyms*

Before the launch of gemeentegeschiedenis.nl and histopo.nl there have been various lists, “gazetteers”, available for standardising Dutch historical toponyms, but these were not available online and were scattered across various cultural heritage institutes. Thus, an important aspect of this initiative is its ability to interrelate these diverse, de facto standards:

On gemeentegeschiedenis.nl for municipalities:

- CBS code: assigned by Statistics Netherlands, the Dutch national statistical office
- Amsterdam code: assigned by Van der Meer and Boonstra as part of the Historical Geographic Information System project (Van der Meer and Boonstra 2011).
- GeoNames code: an online, crowd-sourced database of over 10 million toponyms

On histopo.nl for settlements:

- GeoNames code, again

⁵<http://pleiades.stoa.org>

⁶<http://www.placenames.org.uk>

⁷<http://www.digdag.dk>

- Kloeke code: assigned by the Meertens Institute for research and documentation of Dutch language and culture (Kloeke 1926).

All these standards for the Dutch context are existing systems that all have their quality and have been used in datasets in the past.

2.3 Functionality

Gemeentegeschiedenis.nl and histopo.nl provide various functions in order to help historians and cultural heritage experts to create and use better geographical data.

2.3.1 *Geographical Standardisation with a Historical Dimension*

The comparison of historical data related to municipalities over time and space is a complex matter due to the heterogeneity of historical toponyms. Historical documents that describe people, economic activity or political votes contain references to the contemporary administrative areas. In many instances, the place of birth (actually, the municipality where the birth was registered) recorded in historical documents presents a peculiar anachronism as it reflects the administrative subdivision at the time of birth, not at the time at which the place of birth was recorded.

Digitising these historical sources, historians not only have to copy the geographical description as precisely as possible as it was recorded in the original source, but also have to interpret the original and provide a standardised spelling of this geographical description. At a later stage this facilitates statistical analysis or spatial analysis using a Geographic Information System (GIS). Within the context of the dataset, this standardised form must be unambiguous and unique. Drawing this standardised spelling from a widely adopted thesaurus, historians are then able to combine and contrast their own datasets with other datasets that use the same thesaurus.

2.3.2 *Combining Existing Sources on gemeentegeschiedenis.nl*

The main source for gemeentegeschiedenis.nl is the thesaurus of Dutch municipalities, generally referred to as the Amsterdam Code (Van der Meer and Boonstra 2011). Although previously published in a book and available under an open license as a downloadable PDF file, there was no online resource in a machine-readable format. In the Amsterdam Code, identifiers are reused over time:

a municipality that only changed name therefore kept the same identifier. When two municipalities were combined into one, the Amsterdam Code of the municipality with the largest number of inhabitants was inherited by the newly formed municipality. This implies that the municipalities are comparable through time, but the codes are not unique. They are only unique in combination with a time-stamp indicating the period of the existence of a municipality.

Each of the municipalities on gemeentegeschiedenis.nl is further identified by their CBS code. This code is assigned and maintained by Statistics Netherlands, the Dutch national statistical office. The historical dimension is limited, but the code is used in most of the currently available datasets, which enables the temporal comparison between the current situation and the past. While the CBS codes do not cover municipalities before 1830, the Amsterdam Code encompasses a broader, temporal range as it identifies municipalities from as early as 1812.

GeoNames contains and classifies various types of toponyms. The geographical coordinates, stored with all the names, enabled us to combine the municipality with the place names that are used for an inhabited place which is not an administrative unit. As a result of this spatial combination, the web pages provide lists of all known settlements in a municipality.

2.3.3 *Combining Existing Sources on histopo.nl*

GeoNames mainly contains names for settlements with an administrative function. These are the basis for combining information on settlements on histopo.nl.

The Kloeke Code for settlements, assigned by the Meertens Institute,⁸ is used in the research and documentation of geographical patterns in Dutch language and culture (Kloeke 1926). Since these settlements are no official administrative areas, disambiguation is ensured by combining the Kloeke code identifier with the coordinates of the settlements. Furthermore, Kloeke codes are not available for all settlements in the Netherlands.

On histopo.nl two additional lists of names of settlements are incorporated. The first list is constructed by Simon Hart and published by the City Archives in Amsterdam.⁹ Hart made a list of all the toponyms he found in sources in the Amsterdam City Archives and resolved them to contemporary names of settlements. A similar list was constructed with the names found in the conscription registries from the 1830s until around 1930s.¹⁰ A selection of these registries throughout the Netherlands was digitised and made available through data-entry. Toponyms were organised and resolved to a standardised name.

⁸<http://www.meertens.knaw.nl>

⁹http://stadsarchief.amsterdam.nl/archieven/hulp_bij_onderzoek/herkomstonderzoek/ (in Dutch).

¹⁰<http://militieregisters.nl/en/>

2.3.4 Access to Information

On gemeentegeshiedenis.nl, the Dutch municipalities have their own web page. Each of these web pages contains a set of maps that do not only reflect the changes in the delineation of the municipalities, but also the mergers of municipalities over time (Boonstra 1992). Furthermore, all settlements that are located within the boundaries of a municipality are listed together with links to other online resources that describe that particular municipality or settlement. Visitors to the website can learn about the history of the Dutch municipalities and researchers can obtain information to assist them in standardising the administrative areas in their datasets.

Each of these web pages is available at a short and readable URL that also acts as a Unique Resource Identifier (URI) (Berners-Lee et al. 1998). For example, the former municipality of Schoten is available at and is identified with the URI <http://www.gemeentegeshiedenis.nl/gemeentenaam/Schoten>.

For every available standard a separate web page is constructed to enable users to resolve a code defined according to a specific standard. The Amsterdam Code 10382 identifies the former municipality of Schoten, which is learned by following the URI <http://www.gemeentegeshiedenis.nl/amco/10382>. The same goes for the CBS code 1173: <http://www.gemeentegeshiedenis.nl/cbscode/1173> (Fig. 2.1)

There are also URIs that provide a list of municipalities in one province, for instance: <http://www.gemeentegeshiedenis.nl/provincie/Noord-Holland>

2.3.5 Providing the Data

Using a process referred to as “content negotiation” a client application obtains the information available in the appropriate data format. In the case of a web browser this is HTML. By manipulating the *accept-header* of the client application requesting the URI RDF/XML or JSON is provided. Presenting these formats in a

Fig. 2.1 Map of the municipality of Schoten in 1885 as presented on <http://www.gemeentegeshiedenis.nl/gemeentenaam/Schoten>



web browser can be forced by using a different URI. This is good practice and performed by Dbpedia as well. For RDF-XML this is <http://www.gemeentegeschiedenis.nl/gemeentenaam/rdfxml/Schoten> and for JSON this is <http://www.gemeentegeschiedenis.nl/gemeentenaam/json/Schoten>. Notice that in this last format extra information is available in GeoJSON (Butler et al. 2008) about the geographic polygon(s) of the consecutive stages through time.

Corresponding URIs for information concerning one province are: <http://www.gemeentegeschiedenis.nl/provincie/rdfxml/Noord-Holland> and <http://www.gemeentegeschiedenis.nl/provincie/json/Noord-Holland>. As an extra service the overview of municipalities in one province is distributed in CSV-format: <http://www.gemeentegeschiedenis.nl/provincie/csv/Noord-Holland>.

2.3.6 *Providing the GIS Data*

Maps presented on the HTML-pages are based on existing files containing the geometries of municipalities through the years (Boonstra 1992). These shape-files are imported into a PostgreSQL database,¹¹ with a PostGIS extension.¹² In this environment geometries can be combined with data. A MapServer¹³ is used to enable serving this information on the World Wide Web. This server application provides the standardised interfaces Web Map Service (WMS) (OGC 2009) and Web Feature Service (WFS) (OGC 2014).

The WMS interface provides, based on a query, a map in a specified image format like PNG. This technique is used to present the images on the HTML-pages of the website.

The WFS interface provides, again based on a query, geometries and requested features encoded in a standardised format like GeoJSON (Butler et al. 2008).

2.3.7 *Search User Interface*

At the moment, the website provides a simple search on the names of municipalities and toponyms. Users can enter a search term and various names of municipalities and settlements are returned. If any of the codes (Amsterdamse Code, CBS-code) is provided instead, the matching municipality or settlement is returned. For every municipality, the search function returns the web page with information about the right information.

¹¹<http://www.postgresql.org>

¹²<http://www.postgis.org>

¹³<http://www.mapserver.org>

The user interface provides widgets that enable historians to narrow down the search result set by selecting a year, a region, and/or a standard. Thus, the responsibility remains with the historians for the correct interpretation of the historical source and for the selection of the right standard and standardised form of the geographical description.

2.3.8 Search API

The search query entered by a user on *gemeentegeshiedenis.nl*, is handled by an Application Programming Interface (API) serviced on *histopo.nl*.¹⁴

A user searching for information about a place name does not know whether he is searching for a municipality beforehand, let alone that he used the right spelling. By using the API of *histopo.nl* as a service on *gemeentegeshiedenis.nl*, users searching for a place name on *gemeentegeshiedenis* can be offered alternatives if no municipality is retrieved.

2.3.9 Architectural Overview

In the paragraphs above, we described the various techniques we used to realise the service platform. This could be visualised in the following architectural overview (Fig. 2.2).

2.4 Evaluation

After two years of development and preliminary use of *gemeentegeshiedenis.nl* and *histopo.nl* we experienced solutions and problems in four areas: the quality and quantity of the data, organisational issues and the quantity of the services. But first we elaborate on the usage of the platform in various situations at this moment.

2.4.1 Usage

After introduction of the platform, various projects were interested in using the data-services we provide.

¹⁴The API is documented on <http://api.histopo.nl/docs/>

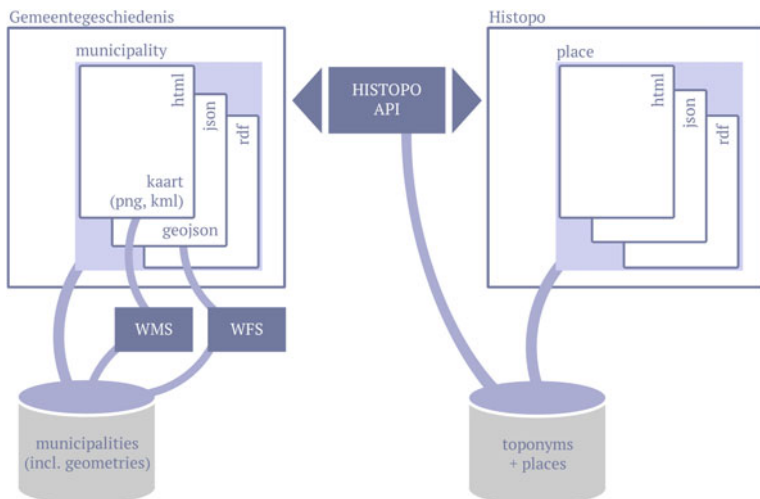


Fig. 2.2 Architectural overview of the websites *gemeentegeschiedenis.nl* and *histopo.nl*

A website with information about archival institutions in the Netherlands¹⁵ uses *gemeentegeschiedenis.nl* to refer to archival holdings. An archival institution provides access to archival collections concerning a particular (former) municipality. By combining the archival institution and the names of former municipalities users can discover which institution is holding the archive of their interest.

In the CEDAR-project¹⁶ (Merono 2013) historical Dutch census-data, which was originally available as scans and in spreadsheets, is converted into Linked Open Data. References to municipalities are common and made with the Amsterdam Code. They experimented with the usage of URIs of *gemeentegeschiedenis.nl*. By relating the Census data to the municipalities in *gemeentegeschiedenis.nl*, data in both datasets can be combined. Furthermore future datasets with the same standardised toponym can be combined as well.

As part of the LINKS-project¹⁷ a dataset was constructed with historical Dutch toponyms. The relation between settlements and municipalities available on *gemeentegeschiedenis.nl* was included (Huijsmans 2013). Unfortunately, new knowledge that is incorporated in our databases must lead to a new release of Huijsmans dataset.

¹⁵<http://www.archiefwiki.org>

¹⁶<http://www.cedar-project.nl>

¹⁷<http://www.iisg.nl/hsn/news/links-project-nl.php>

2.4.2 *Data Quality*

Combining the various datasets resulted in some surprising errors and inconsistencies in the data. In the presentation of the data these errors are extra visible. For instance, the municipalities of Oosterwolde in the province of Friesland and Oosterwolde in the province of Gelderland somehow ended up in one map presented on the web page of Oosterwolde[Ge].¹⁸

The link between a toponym, provided by GeoNames, and a municipality is calculated from the maps: if the coordinates of an inhabited place are within the boundaries of the shape of a municipality a link is constructed. The precision of the maps, however, turned out to be insufficient to construct a trustworthy link. In the calculations an inaccuracy of at most 1 km should be taken into account. The maps should be replaced to improve this function.

It is strange to use a specific dataset oriented to municipalities, when toponyms are handled that are older than 1812 or in situations where standardisation to municipalities is not relevant. Standardising historical toponyms in general should be handled by a separate dataset and service which we introduce on the website histopo.nl.

We have already learned that people are very eager to help improve the information on gemeentegeschiedenis.nl when it is incorrect, particularly the maps. Additional functionality for visitors to provide feedback on the information, or even to help us correct it, could prove to be very helpful. New information can be added as well. New names, with different spellings could be provided by historians doing research in a particular source, with unknown names.

2.4.3 *Data Quantity*

Another area of future growth of our service platform lies in the addition of new names of settlements and their spellings variants. Of particular interest are Dutch exonyms for foreign settlements—e.g. the Dutch toponym “Jarmuiden” refers to the British port of Yarmouth—and foreign exonyms for Dutch settlements, e.g. French toponym “Nimegue” refers to the Dutch city “Nijmegen”. This enables historians to use historical sources from all around the world to combine information about the Netherlands. We made a start with this in developing histopo.nl.

The service platform could also be extended to smaller geographic entities, like neighbourhoods, streets or even individual houses. That way, we gradually create a historical gazetteer or geocoder. We hope and aim that these services will enable

¹⁸http://www.gemeentegeschiedenis.nl/gemeentenaam/Oosterwolde_Ge and http://www.gemeentegeschiedenis.nl/gemeentenaam/Oosterwolde_Fr

historians to interpret, enrich and visualise their historical data in new and innovative ways. Moreover, the standardisation services provide them with a means to actually publish their historical data as linked open data.

2.4.4 *Organisational Issues*

At this moment the project is not financially supported. This endangers future development and continuity of the system. If the usefulness is recognised by a broader audience in humanities and cultural heritage and long term continuity is desirable, financial and organisational support is needed. This support can be found in a national organisation taking responsibility for this task. Another scenario might be that we develop a more commercially oriented business model, in order to provide the necessary financial means.

Not all the datasets we used have a clear copyright statement. In all cases we got informal, personal approval for reusing the available dataset. This means we are unable to provide a clear copyright statement on our dataset as well.

2.4.5 *Service Quantity*

We envision that the historical databases containing toponyms and the geographic representation of Dutch municipalities with various types and links between the entries should be available to the general public in a variety of ways.

Historians can use the search interface to standardise their dataset, but have to search for every entry one by one. Therefore, extra functionality is needed to improve the usability and applicability of the websites.

2.5 Potential

We envision that we (or others) can develop all kinds of services based upon the data-services provided by gemeentegeschiedenis.nl and histopo.nl. We distinguish two types of future functionality: services that assist historians to standardise the geographical entities in their datasets and services that assist them to process their standardised datasets. The remainder of this chapter is used to explain and describe these potential services.

In order to obtain a dataset with standardised forms of the original geographical entity in the historical document, historians should add at least one additional field in their records. (A field with extra information, like the standardised toponym, could be added as well). The original records contain the contemporary and authentic description of the geographical name that needs standardising. In the

additional field, the service will store the interpretation of this original description by means of the right URI referencing the toponym. The main advantage of using a service platform approach is that new knowledge about toponyms and relations is continuously incorporated into the databases. Therefore, the geocoding success rates of the service platforms will improve over time without additional efforts on the part of the historians.

Datasets that contain standardised geographic names can be processed more consistently. For example, one could create a service that converts the standardised geographic names in a dataset into another standard. Furthermore, historians might want to visualise their data set on a map in order to identify geographic patterns.

2.5.1 Indexing or Large-Scale Data Entry Projects

Adding a standardised geographical code or name can be performed at data entry. In that case it would be convenient to incorporate the search into the user interface of the data entry software. In the Netherlands an interesting example where this service could come in handy is <http://www.velehanden.nl>. On this crowd-sourcing platform, archival services can start a data entry project on a specific archival source.

For developing this service the API can be used. A more general service would be a SPARQL-endpoint. A SPARQL-endpoint can be used to formulate a specific query that is not provided by the API. That way the current crude Semantic Web functionality of RDF-XML export is extended.

2.5.2 Automatic Standardisation

Additionally, the API can be used to provide automatic standardisation. Therefore, the result set should be organised according to a certain relevance function: the standardised form with the highest probability will be presented as first entry in the result set. The automatic detection can be improved by two extra variables that provide information about the context of the source:

1. place: the location where the source was created
2. target area: the area where the result must be located
3. time: the period in which the source was created and therefore the period in which the toponym or spelling variant is used

Results are selected that are located in the requested geographical target area and existing in the requested time period, prior to the end date of the period in which the source was created.

The order in which the result set is presented could be determined as follows. Of course, the name with the smallest amount of differences with the search term is

highest on the result list. To start with, this difference can be expressed in for instance the Levenshtein distance (Levenshtein 1966). When more results are delivered, the geographical unit that has the smallest geographic distance to the place of origin is higher on the result list.

An extra service could enable historians to upload a CSV file. As they indicate which column contains the geographical name to be standardised and the standard in which it is to be standardised, the service platform could return a CSV file that now contains an extra column that holds the automatically generated standardisation. In this case, only the result with the highest probability is returned.

2.5.3 Converting from One Standard into Another

Combining various datasets with differently standardised geographic names can be very complex. A service enabling historians to convert from one standard into another could be very useful. Historians would upload a CSV file and indicate the column that contains the standardised name and select the standard and period in which it is to be converted. A new CSV file is automatically generated that contains an extra column holding the additional standardised value.

Historians have to be aware that the different standards are not equivalent: a settlement that is converted into a municipality results in the loss of precision of the geographical location. However, the service platform enables historians to convert whatever they like (Fig. 2.3).

A scholar could, for instance, want to connect her dataset containing GeoNames ids with Kloeke-referenced linguistic data. In order to do so, she can use the API to find the Kloeke code for every GeoName-id in her dataset. All the records in her dataset are augmented with an extra field containing the connected Kloeke code. If she doesn't want to use the API, she should be able to download a list combining the two standards. This list can be used as table in a database.

2.5.4 Enrichment with Geographic Coordinates

For visualisation purposes the standard codes can be converted into geographic coordinates: in a GIS these coordinates can be used to draw a settlement or municipality on a map, together with the information related to the settlement or municipality. For every geographic entity the service platform is able to provide the geographic coordinates. Uploading a file containing multiple geographic entities, the standard forms of the toponyms are then subsequently enriched with the coordinates of their geographic representation in batch.

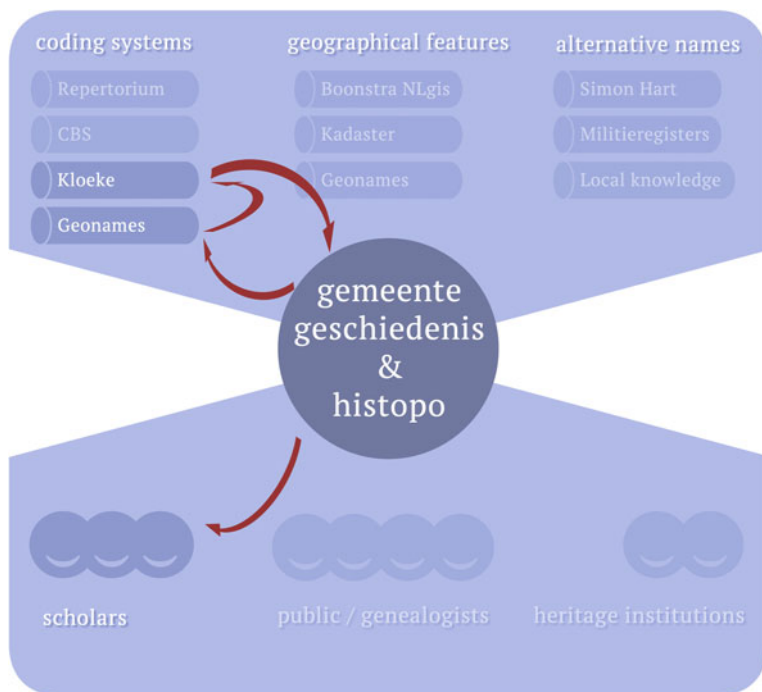


Fig. 2.3 Visual representation of the example of a scholar combining two coding systems for geographical names

2.5.5 Visualising Information

One of the many functions a GIS can perform is the visualisation of geographic datasets. However, the operation of a GIS system can be fairly cumbersome for the uninitiated. A future functionality of a service platform to automatically generate a map based on historical datasets that have been uploaded could be very useful, either to draw the main conclusions in a historical research project, or to assist in the decision whether to involve professional mapping expertise (Fig. 2.4).

If, for instance, a scholar wants to visualise historic data on municipalities in the Dutch province of Limburg, he should be able to download the right selection of geometries of municipalities from our platforms. These geometries can be used to construct a map with the information contained in his dataset. In the dataset a type of municipality encoding, like the Amsterdam Code or the CBS Code, must be included. If this is not the case, the platform could be used to standardise the geographic fields in the dataset first.

Historians would upload their datasets containing both the unique identifiers for the geographic entities, i.e. the Amsterdam Code, CBS code or Kloeke code, and the corresponding attribute values to be represented on the map. Once the specific

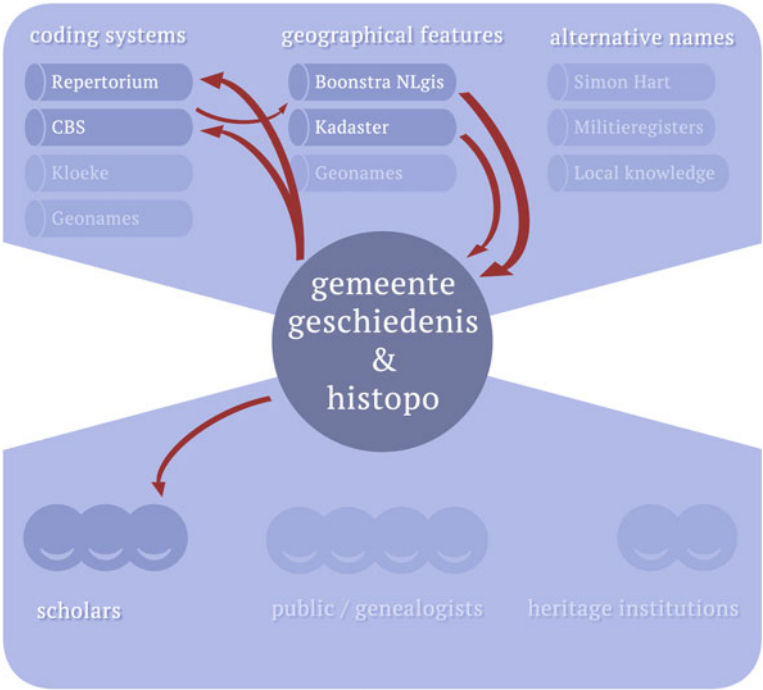


Fig. 2.4 Visual representation of the example of a scholar using the geometries of municipalities

year for the dataset has been selected, the values will be matched to the contemporary geographic delineations of the municipalities. Depending on the nature of the values the appropriate map type is created to visualise the data set geographically.

In case the dataset contains nominal values, a chorochromatic map is created. Municipalities that have the same nominal value are filled with the same colour. An example is a map coloured-coded according to the largest political party per municipality. If the data set contains ordinal or relative values per municipality, a choropleth map is created. Municipalities with increasing values are filled with colours of increasing saturation or lightness. An example is a map that shows the relative growth of the population per municipality. Finally, if the data set contains absolute values, a proportional symbol map is created. An example is a map that shows increasing number of conscripts per municipality using increasing sizes of bars, squares or circles.

The resulting map could be easily downloaded in Scalable Vector Graphics (SVG) or PDF formats for inclusion in scientific publications or presentations. Interactive versions of the maps could be embedded in third-party websites using a JavaScript-based mapping API that would request map images from the service platform.

2.6 Conclusion

This chapter describes the current and future functionality of two websites we envision to be useful for historians with interest in a geographical component in their research. We evaluated the current functionality and identified aspects that can be elaborated. A large amount of these aspects cover the augmentation and improvement of the data itself. Considering the amount of unique visitors and the length of their stay, the website satisfies a need. Also, the amount of feedback and exposure we experienced, as well as projects that used some of our functionality suggest that an historical gazetteer for the Netherlands is much appreciated.

Hopefully, future services will be implemented for our service platform in the near future. Historical research and access to cultural heritage would improve. A financially viable organisation of the initiative is essential.

Acknowledgment We would like to thank Onno Boonstra (Radboud University Nijmegen) who applied for financial support in a “Klein Data Project” (Small Data Project) at DANS-KNAW for further development of some of the web services described.

References

- URLs checked March 20, 2015.
- Berners-Lee, T., et al. (1998). *Uniform resource identifiers (URI): Generic syntax*, <http://www.ietf.org/rfc/rfc2396.txt>.
- Boonstra, O. (1992). NLKAART. A dynamic map of the Netherlands, 1830–1980. In H. J. Smets (ed.), *Histoire et Informatique V. Montpellier* (pp. 315–324). The shape files of this research are deposited at DANS: <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:43063>.
- Butler, H., et al. (2008). *The GeoJSON format specification*, <http://geojson.org/geojson-spec.html>.
- Huijsmans, D. P. (2013). *IISG-LINKS Data-set Historische Nederlandse Toponiemen Spatio-Temporeel 1812-2012*, Release 2013.2 <http://www.iisg.nl/hsn/data/place-names.html>.
- Kloeke, G. G. (1926). De totstandkoming van de “kaart van het Nederlandsche taalgebied ten behoeve van het dialectgeografisch onderzoek” In Grootaers, L., & G. G. Kloeke, *Handleiding bij het Noord- en Zuid-Nederlandsch Dialectonderzoek*, (pp. 57–65) Den Haag : Martinus Nijhoff <http://www.meertens.knaw.nl/projecten/mand/LITkloeketotstandkoming.html>.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10, 707–710.
- Meer, A. v.d., & Boonstra, O. (2011). *Repertorium van Nederlandse gemeenten vanaf 1812*. Den Haag : DANS (Data Archiving and Networked Services) <http://www.dans.knaw.nl/nl/over/organisatie-beleid/publicaties/DANSrepertoriumnederlandsegemeenten2011.pdf>.
- Meroño-Peñuela, A., Guéret, C., Ashkpour, A., & Scharnhorst, A. (2013). *Publishing, harmonizing and consuming census data: The CEDAR project*. Open data on the web workshop, world wide web consortium, Open Data Institute, Open Knowledge Foundation, 23–24 April 2013, Google Campus, Shoreditch.
- OGC (Open Geospatial Consortium). (2009). *Web map service implementation specification—Version 1.3.0* http://portal.opengeospatial.org/files/?artifact_id=14416.

- OGC (Open Geospatial Consortium). (2014). *Web feature service 2.0 interface standard—Version 2.0.2* <http://docs.opengeospatial.org/is/09-025r2/09-025r2.html>.
- Southall, H., Mostern, R., & Berman, M. L. (2011). On historical gazetteers. *International Journal of Humanities and Arts Computing*, 5(2), 127–145.

Population Reconstruction

Bloothoof, G.; Christen, P.; Mandemakers, K.;
Schraagen, M. (Eds.)

2015, X, 302 p. 55 illus., Hardcover

ISBN: 978-3-319-19883-5