KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

LONGPOP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Geocoding the Historic Sample of the Netherlands Database*

Diogo Paiva

Early-stage researcher of the LONGPOP project at the International Institute of Social History of the Koninklijke Nederlandse Akademie van Wetenschappen (KNAW)

Contents

* Thanks to Kees Mandemakers for its review and comments.

KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

LONGPOP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Summary

This paper presents the process of georeferencing addresses contained in the Historical Sample of the Netherlands (HSN). This is part of the Expected Result 11.4 - *Building algorithms to implement a GIS based address system for the two databases that are the most suited to start with, result: adding variables with GIS coordinates*.

The HSN is a historical database of Dutch individuals born between 1812 and 1922. It contains information on life events, genealogical pedigrees, occupations and residence of over 85,500 research persons. It is the goal of the HSN to establish the life courses of the research persons from the cradle to the grave. However, until recently, it lacked a proper implementation of a GIS. Earlier studies could only perform geographic analysis on a municipal level, not by applying spatial analysis *per se* but rather using municipalities as a categorical variable. Thus, it is the goal of this project to add geographic coordinates for each historical address present in the current public release of HSN[1] (almost 340,000 observed addresses).

Addresses were collected from the *bevolkingsregister* (population register) and later from the so called family cards. These historical records' main purpose was to register socio-biographic information on the household and the individual members. Given the secondary importance to addresses, these were recorded in a broad variety of content and structure, accentuated by the decentralized nature of the administrative procedure. Thus, the digitized addresses in the HSN vary widely in terms of precision (from door number in a street to a number in a municipality), structure (order of the elements) and abstraction (from defined structure to open area like fields outside of village).

To be able to proceed with the geocoding process, an algorithm was employed to decompose the addresses in their constituent elements (number, street, *wijk* or district, municipality). However, due to the nature of the historical sources, the data was only partially organized, and thus it was required to normalize it, before proceeding to georeferencing it.

Therefore, this project is divided into two phases: normalization and georeferencing. Once the data was properly normalized, geographic coordinates were given to the identified addresses, following a three level model: municipality, locality and street.

Goals

Add coordinates to each address, in a threefold system (street centroid[2], locality[3] and municipality[4])

[1] HSN Release Version 2010.01.
[2] I.e., the median point of the line representation of the street.
[3] A point coordinate representing the centre of the locality, frequently the main square or church.
[4] A point coordinate representing the administrative-political centre of the municipality, usually the locality where the *gementehuis* (town hall) is located.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Data & Resources

The georeference process used a table from HSN database named BEVADRES extracted from 2010 release (338,133 records)[5]. This table contains several fields, including:

- ID code for **research person** (IDNR)
- Number for **household** (HUISHNR)
- **Begin and end dates** of the research person residing in the address (ADRESDAT and ADUITDAT)
- **Street name**, if in the address (STRAAT)
- **House number**, if in the address (HUISNR)
- **Additional** information of **house number**, like letter or 'bis' (HUISNRTV)
- ***Wijk* name/number**, which is a subdivision of the municipality. Depending on the municipality, *wijk* can mean quarter/neighbourhood or district (WIJK)
- **House number** in the ***wijk***, i.e., if houses are numbered according to the *wijk* they belong to and not to the street (WIJKHSNR)
- **Additional** information of ***wijk* house number** (WIJKHSTV)
- **Name** of the **owner of the boarding house**, or head of household, in case the research person is a boarder (KOSTBAAS)
- ***Deelgemeente***, a division of the municipality (distinct from the *wijk*) (DEELGEMEENTE)
- **Municipality name** where the research person is registered (GEMEENTE)

Additionally, three other sets of data were gathered:

- Basisregistratie Adressen en Gebouwen (BAG), version of 2017[6]
- List of Dutch Municipalities (1812-2012), with coordinates for the municipal capital[7]
- List of Dutch Toponyms (1812-2012), with coordinates[8]

For the execution of this work, the following online resources were used for searching names of streets and/or localities:

- Geheugen van Drenthe, which contains information on places in the province of Drenthe
- Gemeentegeschiedenis.nl for obtaining historical municipal borders (using of Amsterdam Code) and defining moments when streets belonged to a given municipality
- Postcode.site, an alternative for indexed streets by municipality
- Wikipedia.nl, mainly for identifying unknown toponyms
- Adamlink, a very comprehensive site regarding historical streets in Amsterdam
- Stadsarchief Rotterdam's index of historical streets

[5] In the original table there are 338,776 records, but 643 were excluded for being outside of the geographic scope of the Netherlands.
[6] This file contains the official record of public spaces in the Netherlands (e.g. streets) associated with a postal code and the latter centroid coordinates.
[7] D. P. Huijsmans, *IISG-LINKS Dataset Historische Nederlands Toponiemen Spatio-Temporeel 1812-2012, Release 2013.2.*
[8] Idem.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

- Geonames.org, for toponyms search

The two main software used for this project were: **RStudio**, for the data wrangling and normalization; and **ArcGIS Pro**, for georeferencing images of old municipal maps from Kuyper's *Gemeente-Atlas*[9] to search localities and spatial data visualization

Methodology

Considering the available resources, namely a dataset with actual postal codes and public spaces in the Netherlands with a corresponding coordinate and a dataset of historical addresses already decomposed into its constituent elements (house number, street, *wijk* and municipality), a straightforward process of record linkage was defined. Fundamentally, the geocodification of the HSN would be achieved by assigning a modern postal code to each historical address. The granularity corresponds to the centroids of the postal codes polygons, which considering the nature of the historical data provides an adequate spatial precision.

The record linkage between modern postal codes coordinates and historical addresses are but a final step in a larger workflow of digitization and georeferencing of the historical addresses present in the Dutch population registers (Figure 1). Therefore, the success of linkage is highly dependable on the performance of the previous steps, thus creating the first challenges that we faced. The system of data entry of the HSN relies on a series of checks and revisions after the manual data entry of records that takes human transcription error to a minimum. Nonetheless, some mistakes persist and are passed on to database. Moreover, data entry in the HSN follows a logic of literal transcription and minimal interpretation from the person entering the record. Although it helps reduce transcription errors, it mirrors the spelling variety present in the sources. It became evident that a pre-stage standardization was needed.

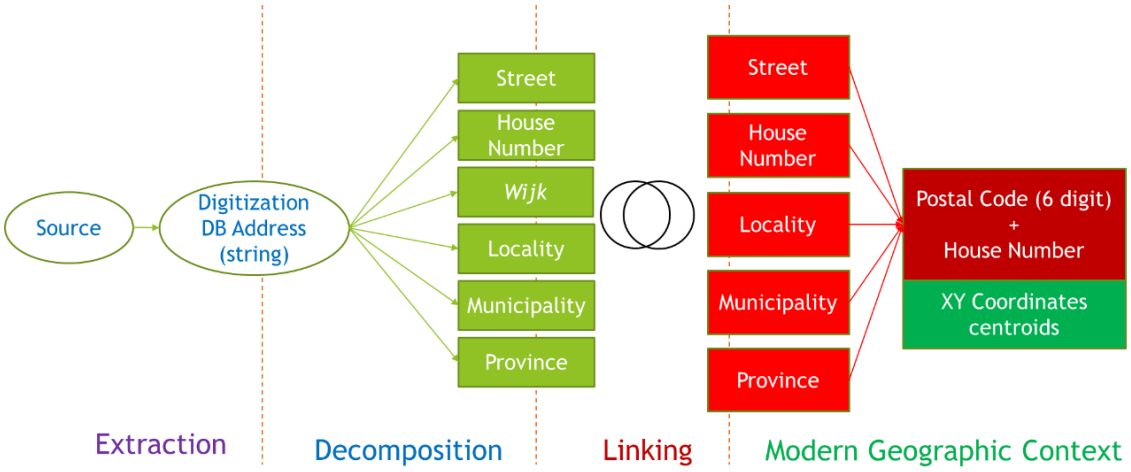


Figure 1 - Methodology of Georeferencing the HSN

[9] Kuyper, J (c. 1868), *Gemeente-atlas van Nederland: naar officieele bronnen bewerkt*, Leeuwarden. It is a collection of maps for the existing Dutch municipalities in the 1860 and 1870s.

KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

LONGPOP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Since the transcribed addresses were recorded as literal text strings, additional to the spelling variation, the structures of the addresses were also copied in their original format. This created a structural issue when decomposing the addresses into its elements. Since for most of the period covered by the sources used in the HSN there was not a single norm followed by all municipalities regarding the addressing system, a variety of addressing formats and systems were recorded. The decomposition of addresses was obtained through an algorithm that could not efficiently take into consideration all the diversity in the recorded addresses. This resulted in part of the data being imperfectly arranged. E.g., when the decomposition algorithm considers "A Alexanderplein 21" it can decompose the address as "A" (*wijk*) + "Alexanderplein" (street) + "21" (house number). But when the same address was recorded slightly different as "Alexanderplein 21 A", it becomes very challenging to properly distinguish the structure of house number "21"+ wijk "A" from just a house number with an extra letter "21A". Regional and local differences conditioned the organization of administrative space and, consequently, of addressing. Figure 2 shows differences in the schemes of addresses that vary if the municipality is located in an urban or a rural setting. It is noticeable that the *wijk* in urban municipalities refers to a subdivision of the locality (quarters/neighbourhoods) and localities coincide with municipality (e.g. Amsterdam or Leiden), whilst in rural municipalities the *wijk* is usually a large area that includes one or more localities, streets are named after the locality (e.g. the locality of Kleine Dorp has 3 streets and all are named Kleine Dorp). In less dense municipalities addresses can be composed of only the municipality and a house number, i.e., house numbering system is for the whole municipality, instead of numbering by small sections (like streets or *wijken*).

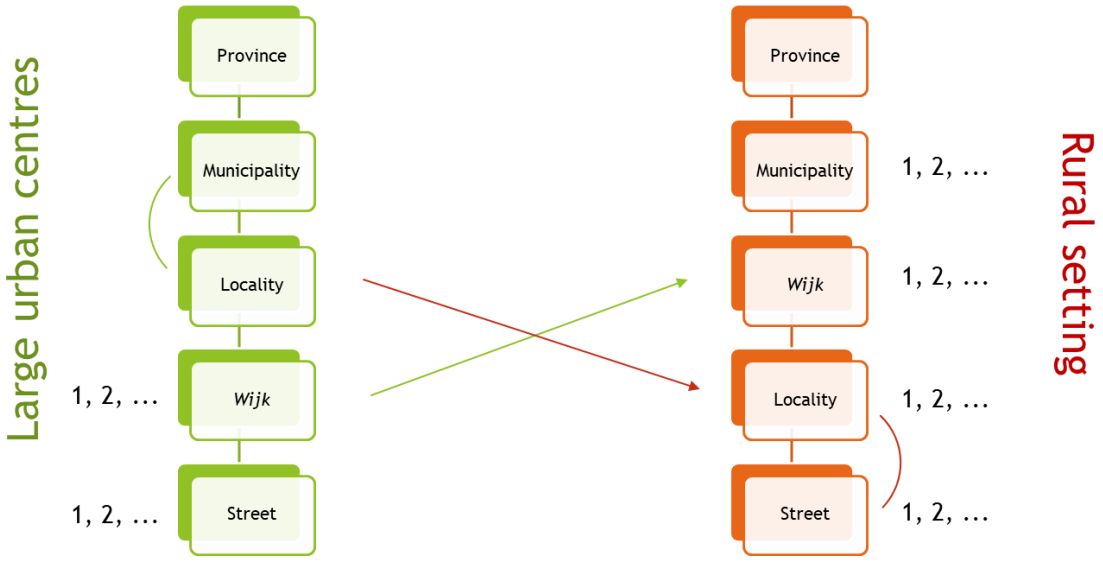


Figure 2 - Different addressing schemes

Besides spelling variation and address system/structure variation which impact data entry and the decomposition algorithm, a third challenge arises from the nature of the sources and the dimension of historical time on administrative processes. The Dutch population registers

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

started mid-19th century consisting on a double page form recording the household members and associated information (including place of residence), with all changes being also recorded. This later evolved to a family card system at the beginning of the 20th century until the 1930s when a personal card was introduced. It was a de-centralized system, with each municipality having the responsibility for the bookkeeping. This leads to a degree of diversification of the forms, ways of recording, linguistic differences, personal styles and local specificities that were gradually reduced in the 1900s with introduction of the new systems.

For tackling these challenges, a standardization and normalization process was designed. This is an intermediary step between address decomposition and record linkage with the BAG file (modern postal code coordinates). The fields to be standardized are: **STRAAT**, **HUISNR**, **HUISNRTV**, **WIJK**, **WIJKHSNR**, **WIJKHSTV** and **DEELGEMENTE**. Concerning fields related to house numbering, the main issue was misplaced values, i.e. street names or *wijken* wrongly placed by the decomposition algorithm. Regarding the other fields (streets, *wijken* and *deelgemeenten*), although a fair amount of misplacement was identified, the main issue was the lack of standardization of names. The street names was the most problematic field with a higher degree of entropy and therefore the first efforts were focused on the field STRAAT (Table 1).

Field	*Unique values*	*Unique values (by municipality)*[10]
STRAAT	64,432	85,173
HUISNR	3,325	48,624
HUISNRTV	5,057	9,487
WIJK	3,885	12,056
WIJKHSNR	2,292	44,004
WIJKHSTV	1,085	4,920
DEELGEMEENTE	2,362	2,702
Values to Review	**82,780**	**206,966**

Table 1 - Values for standardize/normalize

The process to deal with street name values considers two main stages: (1) definition and conversion to a standard and (2) identification and typifying the value (Table 2). For the first stage, the standard is the official name (and spelling) present in the BAG file. The second stage verifies if the street can be found today, and if it exists it typifies it based on the suffix (e.g. –straat, -weg, -gracht). A list of possible outcomes is provided in Attachment 1.

STRAAT (original value)	*straat_std (standardized value)*	*Type (identified and typified)*
2e Atjehstr.	Tweede Atjehstraat	straat
Prinsegr	Prinsengracht	gracht

[10] I.e. summing unique values counts per each municipality. This allows duplicates in the case that same values were recorded in different municipalities. For example, the value "dorpstraat" is counted once in *Unique values*, but 76 in *Unique values (by municipality)* as 76 different municipalities recorded a Dorpstraat.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

K. Houtstr	Korte Houtstraat	not found

Table 2 - Examples of street standardization process

As this process of standardizing values and typifying them has a large human input, especially at the beginning of the project, a set of rules and decision making diagram was established early on to prevent deviations (Attachments 2 and 3). Dealing with the whole set of street names took more than 500 hours of work, spread over several months. Without these rules it would be increasingly hard to make the same decisions in similar cases as time (and knowledge of the data) progressed. Because some values in **STRAAT** are misplaced, a new standardized variable was created for localities: **woonplaats_std**. After all street names were processed, the other fields were standardized following a similar standardizing scheme with a parallel process of reassigning misplaced values (Table 3).

STRAAT	*WIJK*	*DEEELGEM*	*straat_std*	*wijk_std*	*woonplaats_std*
Mijnden A				A	Mijnden
		Oud. Aa			Oud Aa
J.v/d Doesstr.			Jacob van der Doesstraat		
D	Kruidenierstraat		Kruidenierstraat	D	

Table 3 - Examples of standardized addresses

Due to the lack of a dictionary or conversion file to assign validated standards, the process was initially fully manual coding using R. Every municipality was coded with the Amsterdam Code that served as parameter to create subsets of street names to standardize. After the initial provinces were processed (Drenthe, Flevoland – only Urk – and Friesland) a small dictionary was obtained. The subsequent provinces were standardized by a semi-automatic process. A sub-routine was designed to convert original street names into known standards and afterwards verify if the standardized form was present in the BAG file. If this failed, then a manual standardization was necessary. The introduction of this subroutine accelerated the process without compromising the rigour of verifying official names. The workflow of both manual and semi-automatic processes are illustrated in the diagrams of Attachment 4.

Once the process of standardization and normalization is concluded, the data is prepared to be linked with BAG file and thus for georeferencing the historical addresses present in the HSN. Because different address systems were used along the period of the HSN, a triple system of coordinates was created: street, locality and municipality.

Street coordinates (**s_lon**, **s_lat**) were linked with the HSN addresses through the BAG file. Given the historical nature of the addresses of the HSN, a simple process of record linkage would fail substantially. Therefore, a variable matching criteria was used to improve the efficiency of the linkage. Street coordinates are linked in a four turn process. For each turn, a subset of addresses that failed matching is produced and only those are tried with new criteria. Firstly, addresses from HSN are linked with BAG file using street + house number + municipality. Secondly, street + locality + municipality are the criteria. Thirdly, to deal with

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

historical municipalities that were annexed into others, old municipality as locality + street + current municipality are used. Finally, just street + municipality are connected.

Coordinates for localities (**p_lon**, **p_lat**) and municipalities (**g_lon**, **g_lat**) were obtained by linking the **woonplaats_std** with the Huijsmans' file. For those localities that are inexistent in this file, the coordinates were obtained through the georeferencing of historical maps of municipalities compiled by J. Kuyper and locating the missing localities, using ArcGIS Pro. Finally, a composite coordinate (**n_lon**, **n_lat**) is created of the most precise location for each addresses. These values are the street coordinates unless these are unknown in which case the locality coordinates are taken and if they are also unknown then the municipality coordinate are used.

Results

At the end of the georeferencing process every address was assigned at least one set of coordinates, from the centre of the municipality. They were obtained using the Huijsmans file that includes time stamps for the municipalities. But a few (0.1%) failed linkage between the municipality referred in the HSN and the dates of the municipalities in the Huijsmans file, which had to be manually solved.

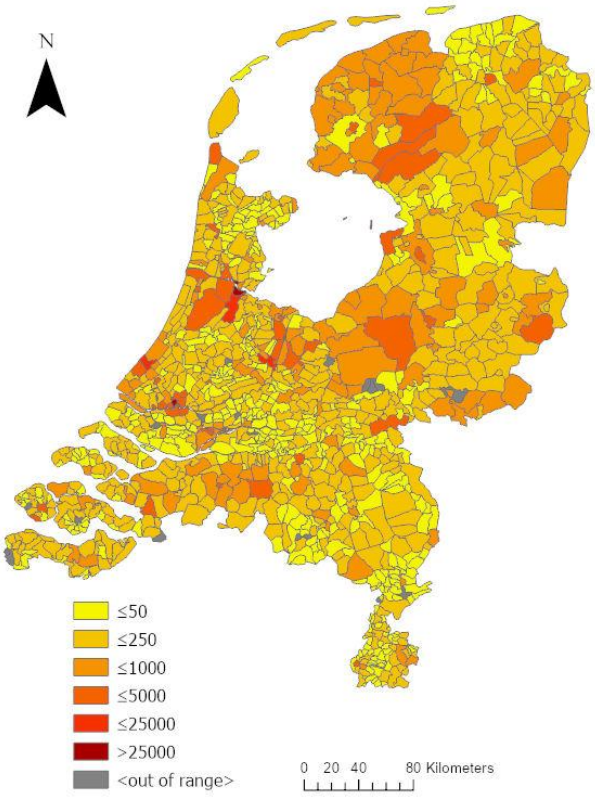


Figure 3 - Distribution of HSN's addresses (municipal borders based of Kuyper Atlas – 2nd half of 19th century)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

As detailed in the methodology, assigning coordinates at a street level was the largest task and included a lengthy process of data correction and standardization. Given the amount of time available to finish the georeferencing (roughly 2 years of non-exclusive work time) it was possible to revise the data in detail. However, some limitations still exist, especially regarding archival research that would prolong this task far beyond the end of the project time. Also, the use of timestamped house number systems could not be done. Although the database contains house numbers (and *wijk* house numbers) georeferencing address by using them meant that would be necessary to define for each municipality the sequence of house numbering systems in use per historical period. This demands heavy archival work that is outside the scope of this project. Therefore, the main focus was to properly identify streets and approximate house numbers to today's system and in case of failure disregard house numbers and assign a median point in the street.

Early on a strategic decision had to be made regarding the available time and resources, the goals of this project and achieving the best results possible. The granularity was then set as a centroid of the section of a street, rather than houses. This made house numbers less relevant but identifying streets correctly (in order to successfully link with the BAG file) became fundamental. Table 4 shows the distribution per province of the values that underwent the process of standardization.

Province	Values in STRAAT	%
Drenthe	1,200	1.4
Flevoland[11]	4	0.01
Friesland	3,828	4.5
Gelderland	4,648	5.4
Groningen	2,493	2.9
Limburg	3,793	4.5
Noord-Brabant	6,724	7.9
Noord-Holland	20,989	24.6
Overijssel	4,265	5.0
Utrecht	6,763	7.9
Zeeland	5,064	6.0
Zuid-Holland	25,402	29.8
	85,173	

Table 4 - Values to standardize per province

At the beginning, the provinces of Drenthe and Friesland (along with the municipality of Urk nowadays in Flevoland – 5.5% of total values to revise) were used to estimate how long the street standardization process would take. Being initially a manual process, with some training the average values reviewed passed from 33 to 63 per hour. By this rate it would take more than a year of exclusive work to finish it (~280 working days). In order to reduce the amount of time needed to finish this process, subroutines were introduced for Gelderland to complement the manual work. This would be a cumulative process as these subroutines use a dictionary

[11] Flevoland was created in 1986 and from the municipalities present in the HSN the former island of Urk is the only currently in its limits.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

that grew each time a municipality was processed. After Gelderland and Groningen were revised (8.3%), the estimate for the duration dropped significantly to ~150 working days, considering an increased processing rate of around 100 streets per hour. Finally, the whole process went through October 2017 to August 2018, only taking the correspondent of 87 full working days (Figure 4).

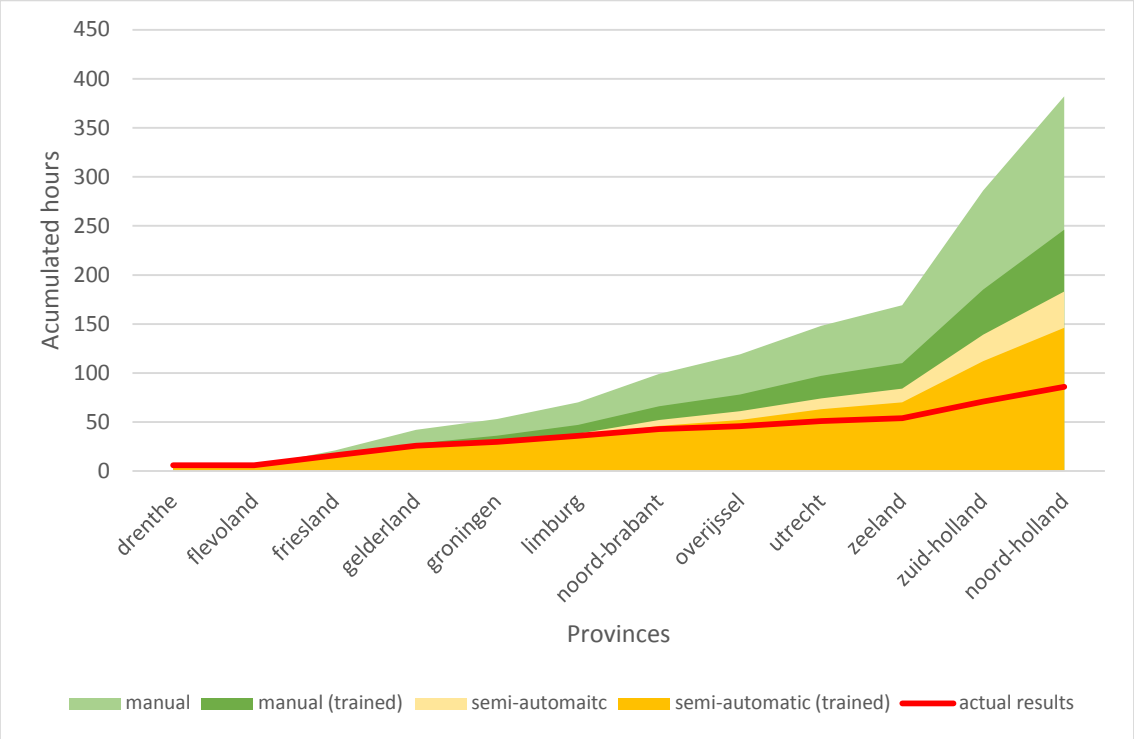


Figure 4 - Estimated hours and actual progress for standardizing streets

The rate of processing rose as the dictionary grew and for the largest provinces (Zuid- and Noor-Holland) 240 streets/hour were processed. In this way, it was possible to reduce the amount of street names roughly to almost a third. For the remaining fields, the process of standardization and normalization was quick, as there are less values to be revised. WIJKHSNR increased has some values were misplaced in **STRAAT** (Table 5).

Field	*Unique values (by municipality)*	*Standardized values (by municipality)*
STRAAT	85,173	32,794
HUISNR	48,624	45,377
HUISNRTV	9,487	9,480
WIJK	12,056	10,367
WIJKHSNR	44,004	44,335
WIJKHSTV	4,920	4,899
DEELGEMEENTE	2,702	2,207
Total	**206,966**	**149,459**

Table 5 - Standardization Process Results

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Regarding street identification (before processing the other fields), it was possible to identify and typify 64% of the names in **STRAAT**, either as *weg* (55%), "found, but undefined" (6%) or *plaats* (3%). For the remainder street names values 6% were coded as without street address and 30% were streets names that could not be found. Most frequent reasons for the latter include: street renaming; street ceased to exist; street exists but does not belong to the same municipality or its successor; names were incorrectly entered in the database; names do not refer to streets or places, but to relative locations like *kom* (town centre) or *veen* (field) or polder. Given that in later stages of the process some misplaced street values were rearranged, the identified streets rose to 66.4%. Nonetheless, identification rates vary significantly between municipalities. The best results were obtained in the provinces of Noord-Holland (73.9%), Zuid-Holland (69.4%) and Limburg (68.3%), while Friesland (34.3%), Zeeland (46.9%) and Drenthe (52.5%) present the worst, without considering Flevoland that only refers to Urk. In parallel to worst results is lower proportion of addresses systems based on streets, with the exception of Zeeland where although most addresses are at the street level (70.4%) the successes was well below the average. An explanation for this may be the higher level of modern urban development and its destructive consequences in municipalities that were mostly rural in the 19th century, in comparison with more developed municipalities that change less and where street systems are more stable. Finally, localities were scarcely indicated (4.8% of total addresses) in the sources. Maybe this was because it usually was the municipality centre and thus was deemed unnecessary to explicitly mention the locality. Nevertheless, for those localities mentioned, 97.8% were properly identified (Table 6).

Province	*Streets*	*Identified streets*	*Localities*	*Identified Localities*
Drenthe	2,075	1,090	1,680	1,662
Flevoland	5	0	0	0
Friesland	9,992	3,428	3,918	3,891
Gelderland	9,600	5,503	1,991	1,975
Groningen	4,516	2,895	962	962
Limburg	7,965	5,437	1,006	963
Noord-Brabant	13,229	7,775	2,151	1,998
Noord-Holland	64,949	48,011	1,592	1,520
Overijssel	9,198	5,946	672	659
Utrecht	18,477	12,172	648	648
Zeeland	9,746	4,569	421	418
Zuid-Holland	77,669	54,151	1,488	1,465
Total	**227,421**	**150,977**	**16,529**	**16,161**

Table 6 - Identification results

In the final step, the georeferencing of coordinates via BAG file, the success varied as the result of proper identification of streets and linking them with the BAG file. The high dependency on the identification of streets resulted in georeferencing 44.7% of all addresses at street level (66.4% of addresses with street). The use of localities' coordinates was possible for 4.3% of the addresses (Figure 5). Thus, currently 51% of the addresses were only georeferenced at the level of municipality. The ability to assign street coordinates was hindered by the fact that

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

addresses either do not have any street mentioned or have street names that are no longer existent.

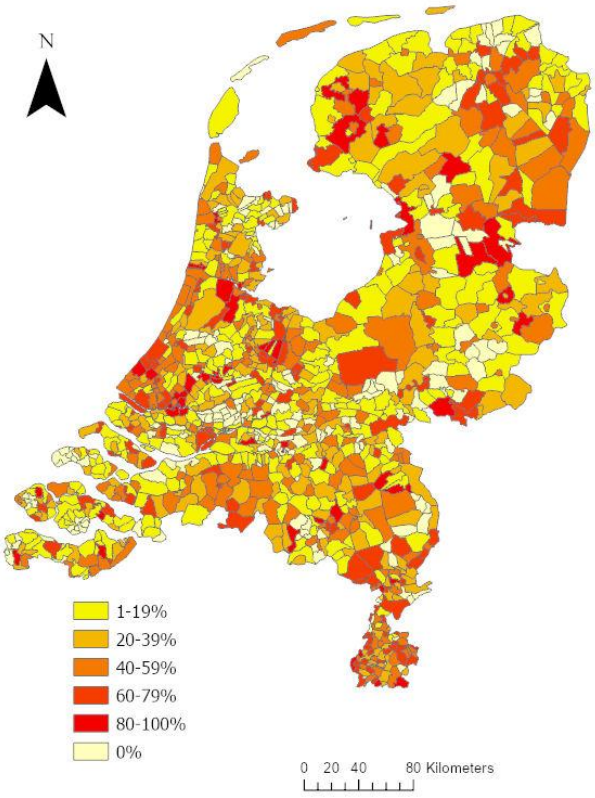


Figure 5 - Success in assigning coordinates for streets and/or localities

In Figure 6, the final output is presented where all addresses were given, at least, a point coordinate derived from the street (**s_lat, s_lon**), locality (**p_lat, p_lon**) and municipality (**g_lat, g_lon**) and the composite nearest point coordinate (**n_lat, n_lon**). This model presents some flexibility for spatial analysis, however it produces an undesirable effect of "spatial heaping". Proportional to the amount of addresses referenced by municipal coordinates is the amount of addresses georeferenced in the same exact location, despite in reality referring to different places.

Further research, namely archival research, might increase the amount of street and localities coordinates versus municipal georeferencing, improving the depth of the HSN GIS. This research should yield interesting results for the historical streets, identifying simple cases of name changing and location of former streets. In addition, as HSN is further developed, a new set of addresses will be available, that can be standardized using the large dictionary created for this project, obtaining faster results while supplying new data itself.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

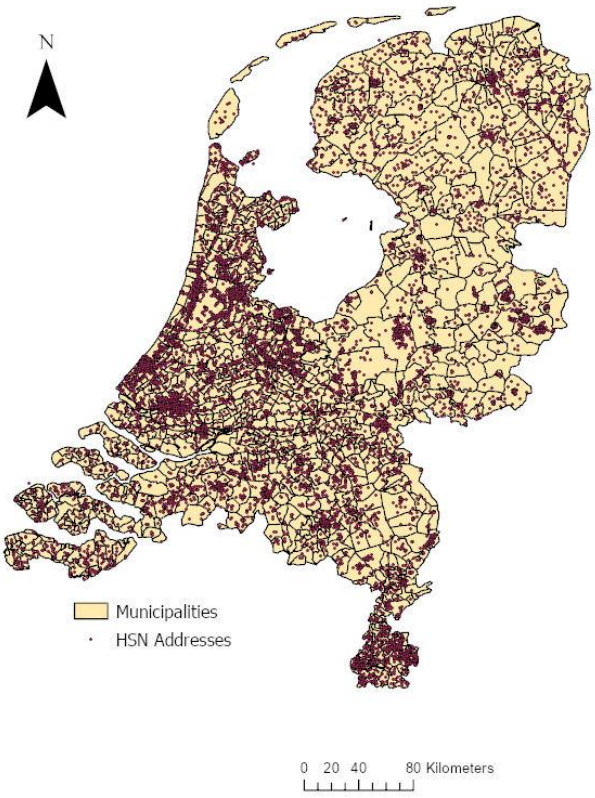


Figure 6 – Distribution of HSN Addresses

KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

LONGPOP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Attachment 1 – Possible outcomes from Identification

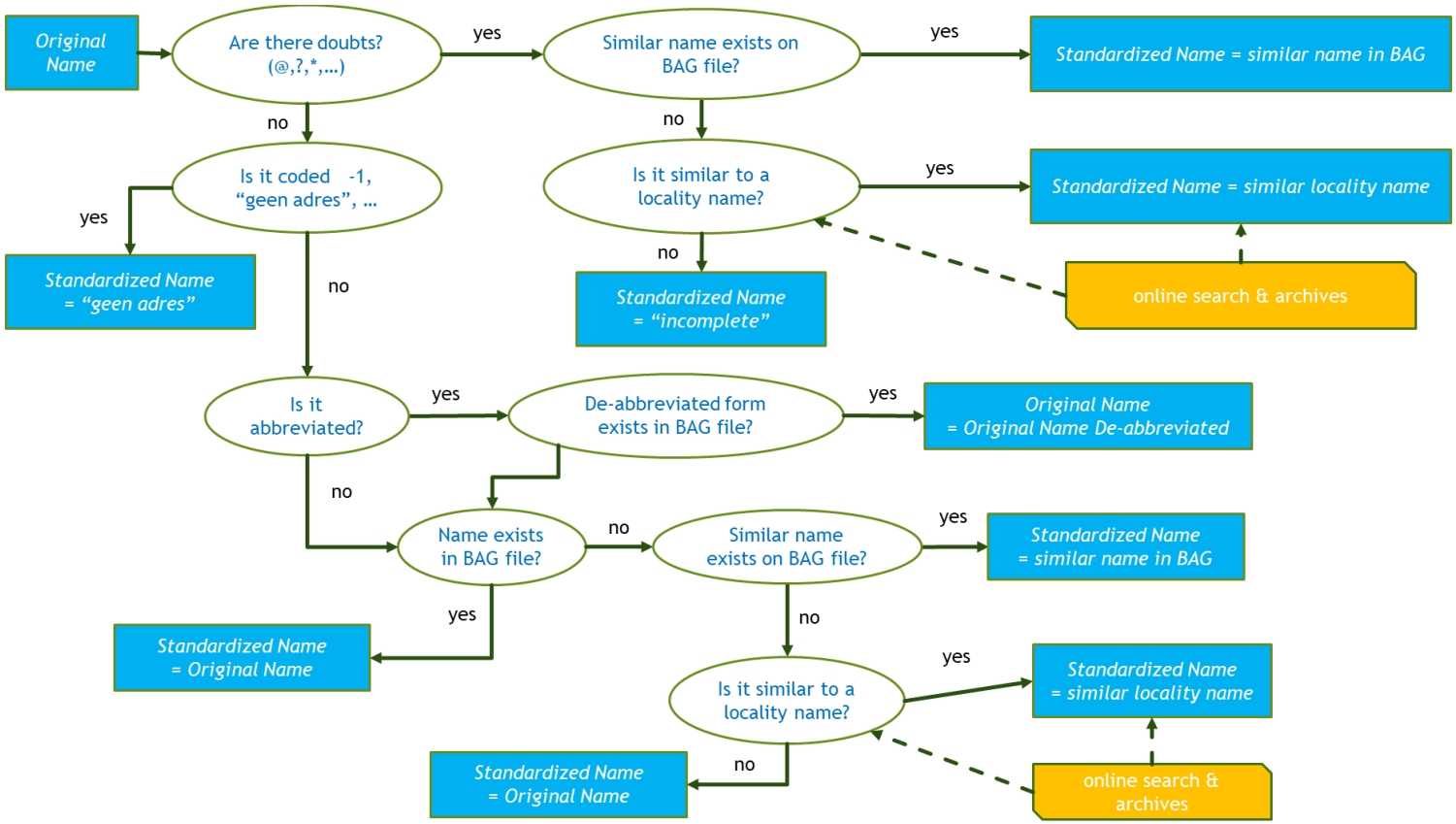
Category	Type	Description
Geen Adres	Geen Adres	If there is not information on address (e.g. only municipality)
Not found	Not found	When the address exists but could not be found
Found, but undefined	Found, but undefined	When address info is not enough to determine a type
Plaats	Buurt / Buurtschap	Neighbourhood
	Volksbuurt	
	Dorp	Village
	Dijkdorp	
	Dorpje	
	Dorpskern	
	Engdorp	
	Esdorp	
	Forensedorp	
	Kanaaldorp	
	Kerkdorp	
	Lintdorp	
	Nooddorp	
	Ontiginningsdorp	
	Plattelandsdorp	
	Streekdorp	
	Terdorp	
	Veendorp	
	Vestingdorp	
	Wegdorp	
	Wierdedorp	
	Hoofdplaats	Capital of municipality
	Gehucht	Hamlet
	Esgeucht	
	Veengebied	Area
	Gebied	
	Kanaal	Canal
	Kern / Kernen	Residential centre
	Landgoed	Estate
	Park	Park
	Plaats	Locality
	Polder	Polder
	Stad	Town / City
	Stadsdeel	City district
	Strafkolonie	Penal Colony
	Veenkolonie	Peat Colony

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

	Wijk	District
	Volkswijk	
	Woonwijk / Woonwijken	
	Woonplaats	Locality
	Zone	Zone
Weg	Baan	Road
	Brink	
	Dijk	Dike
	Gang	Passage
	Herdgang	
	Gracht	Canal
	Haven	Port
	Hoek	Corner
	Hof	Court
	Kade	Quay
	Laan	Avenue
	Markt	Market
	Pad	Path
	Plein	Square
	Poort	Gate
	Singel	Boulevard
	Steeg	Alley
	Straat	Street
	Streek	
	Vaart	Canal
	Wal	Wall
	Weg	Road

LONGPOP
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Attachment 2 – Standardization Diagram

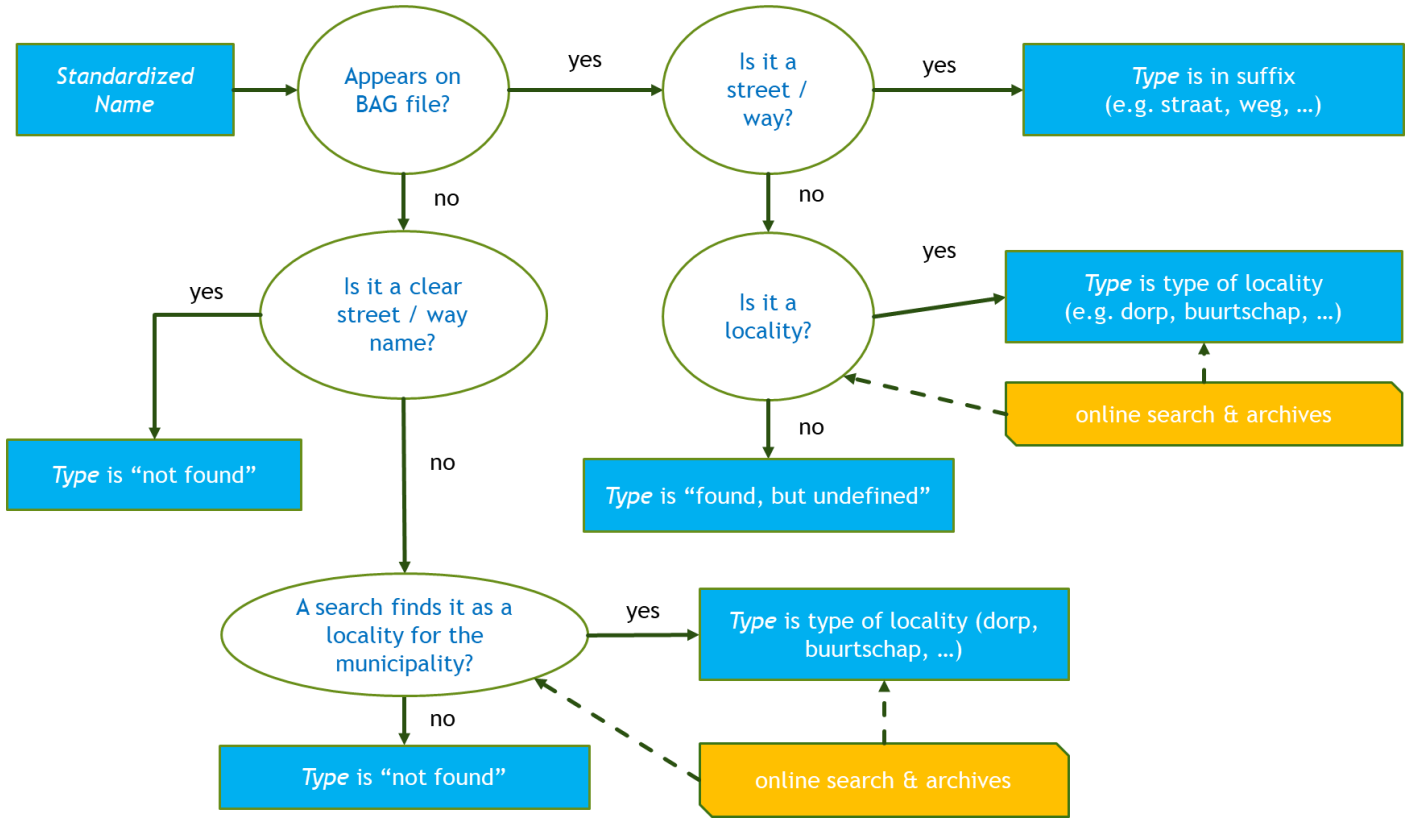


KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

LONGPOP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Attachment 3 – Identification Diagram



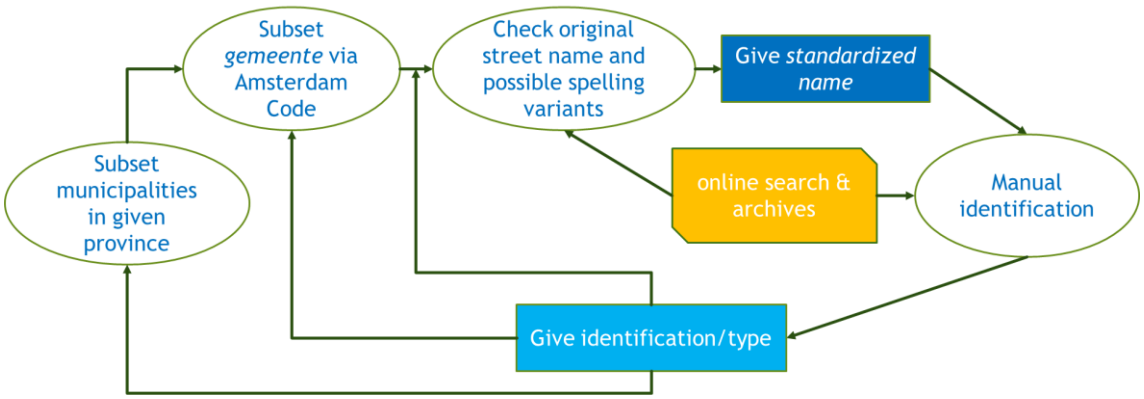
KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

LONGPOP

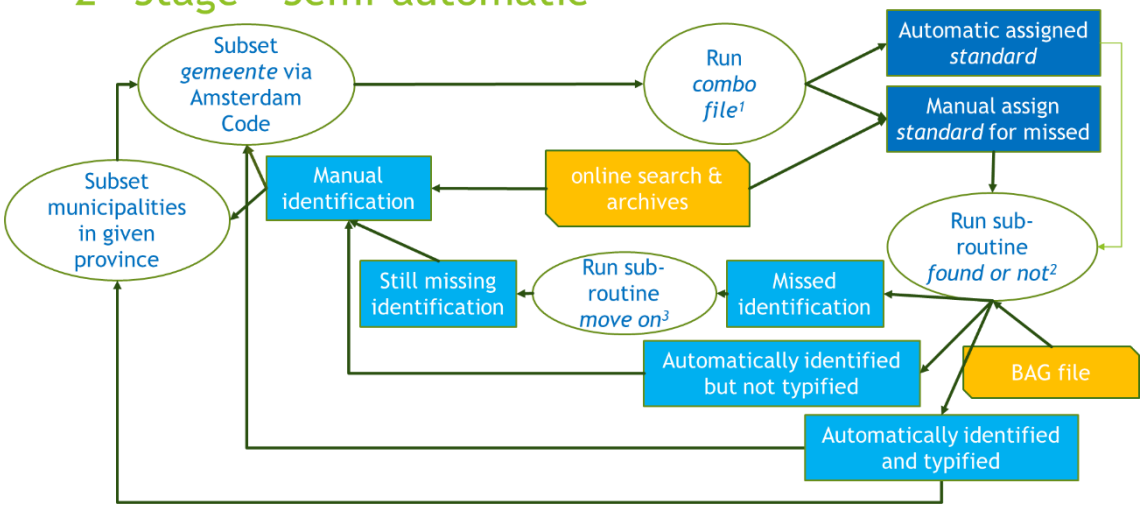
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Attachment 4 – Workflows

Workflow Conceptual Diagram 1st Stage - Manual



Workflow Conceptual Diagram 2st Stage – Semi-automatic



1 – Combo File: uses dictionary to assign known standards

2 – Found or not: to identify and/or typify street names

3 – Move on: to quickly typify common missed identification