

The Zeeland Challenge

A benchmark test for nominal record linkage and life cycle reconstruction for the province of Zeeland, the Netherlands, 1796 – 1963, on the basis of the records of the vital registration



Version 2.2

Gerrit Bloothoof
Utrecht University, The Netherlands
g.bloothoof@uu.nl

Kees Mandemakers
International Institute of Social History, Amsterdam, The Netherlands
kma@iisg.nl



Contents

I	Introduction	2
II	Zeeland and the certificates	4
III	Data properties	6
IV	From vital registers to life cycles	11
V	Census information for Zeeland	12
VI	Proper names in Zeeland	13
VII	Presentation of results	15
VIII	Technical documentation	17
	VIII.1 Introduction	17
	VIII.2 REGISTRATIONS table	18
	VIII.3 PERSONS table	19
	VIII.4 LOCATIONS table	22

I Introduction

The identification of people in the past by computer analysis of digitized historical documents has been a research goal at many institutions around the world for several decades. Over the years, various methods for nominal record linkage were proposed to overcome the fuzziness of the historical data, and successes were reported. But whereas measures for edit distance are now widely used as part of complex linkage methods, few general purpose packages for record linkage have been developed. Methods were tuned to the specific properties of the data, which is understandable given the wide variety and quality in the historical data, both in time and among countries, from rich and precise to sparse and error prone.

To learn the strong and weak points in methods for record linkage and life cycle reconstruction, we now challenge scientists to test their approaches on a common data corpus and to share results. The corpus we have made available to this end consists of the key fields from the certificates of birth, marriage/divorce and death originating from the vital registration of the province of Zeeland in the Netherlands (LINKS_Zeeland_cleaned_2016_01). These 1.56 million certificates cover the period 1796 – 1963 (with some restrictions), are complete between 1811 and 1913, contain full names of individuals (in the Netherlands woman always keep their maiden name in formal registration), but as usual have fuzzy spelling and contain errors and some double (death) records.

Although there is no golden linking standard for the Zeeland data, we expect that this can be realized in the near future, also on the basis of a comparison of results of different linking methods. The Zeeland data are information-rich which gives a fair opportunity to attain agreement on many links.

The Zeeland data can be used for all kinds of linkage approaches, to name a few:

- pairwise linking of certificates (birth-marriage, marriage_child-marriage parents, birth-death, marriage-death) based on fields in common [can also be implemented as ego-based analysis, since the data are provided as records of individuals which are mentioned in certificates]
- pairwise linking of ego-based information without using additional information on parents or partners [which is a kind of equivalent of linking individuals across different censuses]
- group linking (graph linking), based on groups of family members across different certificate types (child+father+mother)
- life cycle reconstruction, combining all information on an individual (which may include various phases of analysis); an implementation of entity resolution
- deduplication (of the death records). When someone died in a different place than the residence, in both municipalities a death certificate may have been made.

In addition, the data can be used to study the effect of (intra- and inter-)name variation on linkage, but also on name encryption processes. Migration can be studied on the basis of places of birth, marriage and decease, while the use of geographic distance between locations in linkage decisions is a subject of interest as well.

The first proposal for this challenge was made at the workshop *Data Linkage: Techniques, Challenges and Applications* at the Isaac Newton Institute for Mathematical Sciences, Cambridge (UK) from September 12-16, 2016. Database construction was realized in the Catch-LINKS project.

II Zeeland and the certificates

Zeeland is the south-western province of the Netherlands and consists of a series of islands and the Dutch part of Flanders (Zeeuws-Vlaanderen). The islands were rather isolated until the flood of 1953 where after the Delta Works connected the islands and protected them against the North Sea.



Key data from historical certificates from the Vital Registration are distributed as *LINKS Zeeland Cleaned Dataset (Marriages, Births and Deaths), release 2016_01*, which includes 1,558,205 records of certificates. Overall, the data cover the period 1796 - 1963 but this varies per type of certificate. Due to privacy restrictions, birth certificates are available until 1913, marriage certificates until 1938 and death certificates until 1963. Whereas the start of the vital registration was in 1796, this only holds for Zeeuws-Vlaanderen (currently border area with Belgium) which became part of the French Empire in that year. For the islands (and the other parts of the Netherlands), the start was in 1811. In total there are 698,285 birth certificates, 192,231 marriage certificates, 1,690 divorce certificates and 665,999 death certificates. Figure 1 shows the number of certificates per main type and year.

The restrictions with respect to the time range of the certificates imply that a full life cycle reconstruction may be feasible for those born between 1811 and 1878, with a marriage before 1938 (maximum age 60) and death before 1963 (maximum age 85).

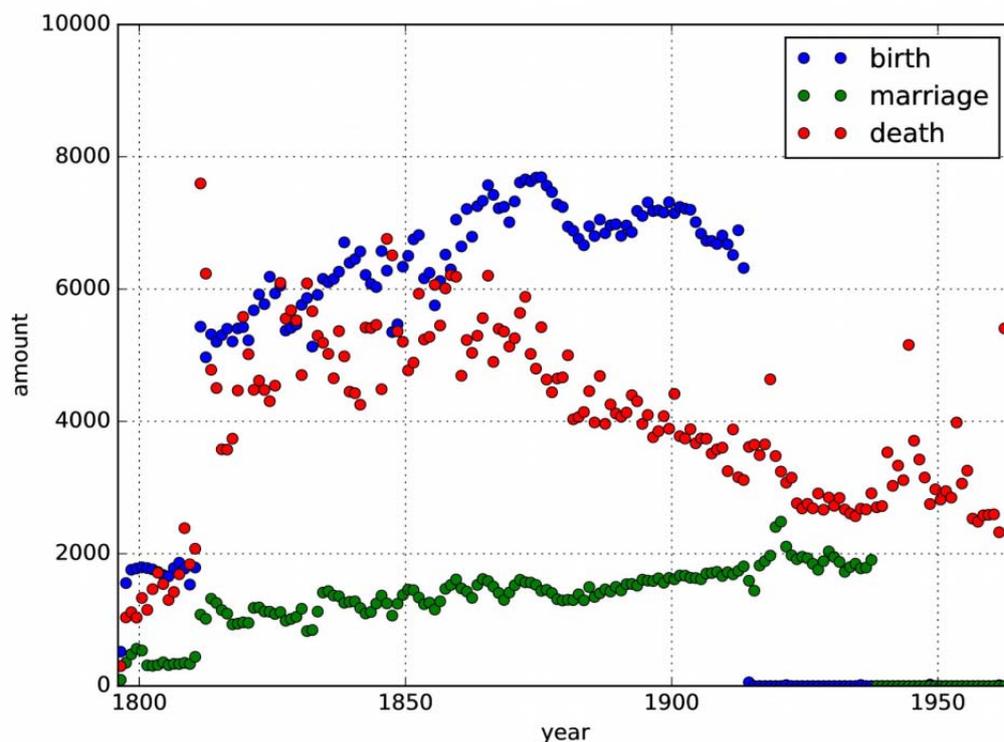


Figure 1: Number of certificates per type and year between 1796 and 1963.

The certificates originate from the *Zeeuws Archive* in Middelburg, and can be accessed¹ at <http://www.wiewaswie.nl/en/home/> where digitized key data are presented, while many records include a scan from the original certificate (see examples in figures 2-4). This website is managed by the *Dutch Centre for Family History* (CBG) in The Hague. In 2013 the digital key data were made available to the *International Institute of Social History* (IISH) in Amsterdam which produced a clean version in a MySQL database with three tables (with registrations, persons, locations, see section), which for this challenge are made available as three .csv (or MS Access) files.

At the IISH, the data were cleaned to some extent. Family names were scanned for prefixes ('*de Jonge*') and titles ('*Meester Jonge*'). These elements were removed and stored in dedicated fields. Titles were left out in this release, since they were too sparse to be helpful in linking. Proper names were not standardized. Some certificates contain remarks or notes. This information was scanned and parsed for viable information which then was moved into its respective fields. In case gender was not explicitly mentioned, this was derived from the role of an individual in the certificate, if possible.

¹ The access to the information at this site is free, but the search function is restricted to a single person, which is sufficient in many cases. A more refined search requires payment.

The cleaned version of the data is available from <https://socialhistory.org/en/hsn/> . A password is needed and can be obtained after requesting and returning a contract (on data protection) to Kees Mandemakers at the IISH (kma@iisg.nl). The licence contract includes the title by which the dataset should be cited.



To the health of the young bridal couple in ZEELAND

III Data properties

In order to get a feel of the data, we first present originals from a pre-printed birth, marriage and death certificate (of Catharina Vermeulen), which are shown together with a translation in figure 2 (birth certificate), figure 3 (marriage certificate of third marriage, and figure 4 (death certificate). Hand-written sections in the certificates are highlighted in yellow, while the digitized key data is highlighted in blue. Note that not all relevant information has been digitized (not all occupations, or previous partners). The digitization process, by volunteers, started in the early nineties in the last century, and rules on what and how to digitize were not limited.

	N. 105 Catharina Vermeulen.	<p>Heden den <u>breiden</u> van de maand <u>December</u> des jaars duizend acht honderd <u>twaanveertig</u>, verscheen voor mij <u>Beuzenreder</u>, Ambtenaar van den Burgerlijken Stand van de <u>stad IJzendijke</u>, Provincie Zeeland, <u>Petrus Vermeulen</u>, oud <u>veertig</u> jaren, van beroep <u>hengstherder</u>, wonende te <u>IJzendijke</u>, dewelke mij heeft verklaard, dat op den <u>eersten</u> van de maand <u>December</u> des jaars achttien honderd <u>twaanveertig</u>, des <u>avonds</u> te <u>zes</u> ure, binnen deze <u>gemeente</u>, in het huis wijk <u>buur</u>, nummer <u>twee honderd twee en vyftig</u>, is <u>geboren</u> een kind van de <u>vrouwelike</u> <u>kinne</u>, van hem <u>afgevoerd</u>, en van <u>Maria Tromp</u> <u>zonder</u> <u>bevoeg.</u> <u>zijn</u> <u>huusvrouw</u>, en aan welk <u>kind</u> hij verklaart te <u>geven</u> de <u>naam</u> van <u>Catharina</u>.</p> <p>welke aangifte en verklaring zijn geschied in bijwezen van <u>Jan Wouda</u> <u>Wouda</u>, oud <u>veertien en vyftig</u> jaren, van beroep <u>van</u> <u>Peter Geansen</u>, oud <u>veertien en vyftig</u> jaren, van beroep <u>slager</u>, wonende te <u>IJzendijke</u>, als getuigen daartoe door den aangever gekozen. En is hiervan aanstonds opgemaakt deze akte, die na gedane voorlezing aan den aangever en de ge- tuigen, door hen en mij is geteekend.</p> <p><u>P. Vermeulen</u> <u>Jan Wouda</u></p> <p><u>Procurator</u></p>
--	---------------------------------------	---

Today, the **second** of the month **December** of the year **eighteen hundred forty two**, appeared for me **mayor**, registrar of the Vital Registration of the **town of IJzendijke**, province of Zeeland, **Petrus Vermeulen**, **forty** years old, with the occupation of **stallion breeder**, living in **IJzendijke**, who has certified to me that on the **first** of the month of **December** at **six** o'clock in the **evening**, in this municipality [of **IJzendijke**], in the **outlying** district, number **two hundred fifty six**, is **born** a child of **female** gender, **of him** **registrant**, and, **of Maria Tromp**, without occupation, **his house wife**, and to which he **declares to give the name of Catharina**. (plus two witnesses with age, occupation and residence)

Figure 2: Pre-printed birth certificate of Catharina Vermeulen (1842) with handwritten information highlighted in translation. Blue highlighted information has been digitized.

700

AKTEN VAN HUWELIJKS-VOLTREKKING.

ZEELAND
75
C.
ALSO OPCEENT

N ^o 123 25 juli 1888 Hubregtsen Pieter in Catharina Vermeulen	Heden den vijf en twintigsten van de maand Juli des jaars een duizend acht honderd acht en tachtig, verschenen voor mij Burgemeester, Ambtenaar van den Burgerlijken stand van de gemeente Oostburg, provincie Zeeland, in het openbaar binnen het gemeentehuis Pieter Hubregtsen oud <u>29</u> jaren, geboren te Oostburg, van beroep <u>veldarbeider</u> , wonende te Oostburg, van staat ongehuwd, meerderjarige zoon van Abraham Hubregtsen, overleden te Oostburg in van Pieterella Geraards, zonder beroep, wonende te Oostburg, in der tijd echtscheiden ter eene; en Catharina Vermeulen oud <u>47</u> jaren, geboren te IJzendijke, van beroep <u>widweuw</u> , wonende te Oostburg, van staat <u>Widweuw</u> van van Livinus Scheerens, overleden te Oostburg, meerderjarige dochter van Petrus Vermeulen in van Maria Tromp, beiden overleden te Waterland Kerkste, in der tijd echtscheiden ter andere zijde;
--	--

Today, the **twenty fifth** of the month of **July** of the year **eighteen hundred eighty eight**, appeared for me **Mayor**, registrar of the Vital Registration of the municipality of **Oostburg**, province of Zeeland, **Pieter Hubregtsen**, **twenty nine** years old, born in **Oostburg**, with the occupation of **land worker**, living in **Oostburg**, unmarried, son of **majority of age** of **Abraham Hubregtsen**, deceased in Oostburg and of **Pieterella Geraards**, without occupation, living in Oostburg, formerly spouses at one side; and **Catharina Vermeulen**, **forty seven** years of age, born in **IJzendijke**, with the occupation of **labour woman**, living in **Oostburg**, **widow of Livinus Scheerens**, deceased in **Oostburg**, daughter of **majority of age** of **Petrus Vermeulen** and of **Maria Tromp**, both deceased in **Waterland Church**, formerly spouses at the other side. (plus four witnesses with age, occupation and residence)

Figure 3: The first part of the certificate of the third marriage of Catharina Vermeulen (1842-1911). Highlighting as in figure 2.

<p>No. 1</p> <p>afgedrukt door de byzondere raad van de gemeente in den deslieden dag el van de staande akte</p> <p>P. Houbregtsen W. Verplanke</p> <p>de hulle</p>	<p>Heden <u>elf Januari</u> negentienhonderd elf, verschenen voor mij, Ambtenaar van den burgerlijken stand der gemeente OOSTBURG:</p>
	<p><u>Pieter Houbregtsen</u> oud <u>twee en vijftig</u> jaren, <u>veldarbeider, man van de</u> <u>overledene</u> wonende te <u>Oostburg</u></p>
	<p>en <u>Willem Verplanke</u> oud <u>een en vijftig</u> jaren, <u>gemeenteboede</u> wonnende te <u>Oostburg</u></p>
	<p>die verklaarden dat op <u>tien Januari</u> deses jaars des <u>vóór</u> middags te <u>elf</u> ure in de gemeente Oostburg <u>aan de Vliet</u> <u>straat</u> is overleden:</p>
	<p><u>Catharina Vermeulen</u> veldarbeidster, oud <u>acht en zestig</u> jaren</p>
	<p>gehoren te <u>Izendijke</u> en wonende te <u>Oostburg</u>, huisvrouw van <u>Pieter Houbregtsen</u> waarmede, toen weduwe van <u>Izaak Herman</u> en van <u>Levinus Scheerens</u>, dochter van <u>Petrus Vermeulen</u> en <u>Maria Tromp</u>, beiden overleden</p>
	<p>Waarvan akte, welke overeenkomstig de wet is voorgelezen.</p>
	<p><u>P. Houbregtsen</u> <u>W. Verplanke</u></p>
	<p><u>J. H. J.</u></p>

Today, the **eleventh** of the month **January** of the year **nineteen hundred eleven**, appeared for me, Registrar of the vital registration of the municipality of OOSTBURG: **Pieter Houbregtsen**, **fifty two** years old, with the occupation of **field worker, husband of the deceased**, living in Oostburg, and **Willem Verplanke**, **fifty one** years old, beadle, living in Oostburg who certified that on the **tenth** of the month of **January** in this year, at **eleven** o'clock in the **morning**, in this municipality of Oostburg in the **Vlietstraat** is deceased **Catharina Vermeulen**, field worker, **sixty eight** years old, born in **Izendijke** and living in Oostburg, house wife of **Pieter Houbregtsen**, mentioned before, and formerly **widow of Izaak Herman and of Levinus Scheerens**, daughter of **Petrus Vermeulen** and **Maria Tromp**, both deceased.

Figure 4: Death certificate of Catharina Vermeulen (1842-1911) with translation and highlighting as in figure 2.

In table 1 the presence of information on ego (the child in a birth certificate, bride or groom in a marriage certificate, or deceased in a death certificate), father, mother or partner in a certificate is indicated. Because pre-printed sheets were used, this key information is usually completely available and digitized. There may be more information which is not digitized but can be viewed from the scans of the original certificates at www.wiewaswie.nl and may be used for clerical review (for those with a good command of Dutch). For instance, when the father is registrant at the birth of children, information on his age and occupation will be present. Also information on witnesses was not digitized, although at marriage this usually concerns valuable information on close

relatives (with age, occupation, residence, relationship). Residence (of key persons) may have been mentioned in the certificate, but is not digitized.

<i>field</i>	<i>Birth ego</i>	<i>Marriage ego</i>	<i>Decease ego</i>	<i>B/M/D children</i>	<i>Decease partner(s)</i>
<i>Ego full name</i>	x	x	x	x	(x)
<i>Ego age</i>	<i>birth date</i>	x	x		
<i>Ego place of birth</i>	x	x	x		
<i>Mother full name</i>	x	x	x		
<i>Father full name</i>	(x)	(x)	(x)		
<i>Partner(s) full name</i>		x	x (<i>last partner</i>)		
<i>Ego occupation</i>		x	x		

Table 1: Digitized key information on ego, mother and father at various events. Ego may not be mentioned at decease of a remarried partner. The father is not always known. Occupation is not always mentioned or is not digitized.

Each certificate has its own information content, but can also of different quality. The information in a marriage certificate is the most consistent because the bride and groom had to show an extract of their birth certificate. On the other hand, the information in a death certificate can be imprecise, erroneous or incomplete, for instance when someone died as widow of old age and neighbours with little knowledge about her and her family did the registration.

The digitized data from the certificates are organized in three tables. One table REGISTRATIONS contains the information that exceeds the individual level, such as municipality, date and type of registration. The contents of a certificate is divided over records, one for each actor, and stored in the PERSONS table. For a birth certificate these are the born child (ego), the mother and the father, resulting in three records in PERSONS. As locations are coded in both the REGISTRATIONS and PERSONS table, the LOCATIONS table gives the location name and code. Technical documentation on the three tables is given in section VIII.

IV From vital registers to life cycles

The same person will have different roles across birth, marriage and death certificates. As child or parent in a birth certificate, as bride, groom or parent in a marriage certificate or as deceased, parent or partner in a death certificate. It is a challenge to link and integrate this information into one life cycle.

An example of a life cycle reconstruction is shown in figure 5 for *Catharina Vermeulen* of whom some certificates are shown in figures 2-4. As can be concluded from table 1, the reconstruction could have been enriched with places of birth, marriage and decease, occupations, and the names of the parents of partners. Note *Catharina's* role as child, bride and mother.

1842: Catharina Vermeulen is born.

1872: Catharina Vermeulen, age 29, marries Izaak Herman, age 32.

1872: child Maria Herman is born.

1874: child Tannetje Herman is born.

1875: child Tannetje Herman, age 0, dies.

1876: child Tannetje Herman is born.

1876: child Tannetje Herman, age 0, dies.

1877: child Catharina Herman is born.

1878: child Petrus Herman is born.

1881: partner Izaak Herman, age 41, dies.

1888: child Maria Herman, age 15, dies.

1900: child Petrus Herman, age 22, marries Suzanna Beerens, age 20.

1905: child Catharina Herman, age 27, marries Marinus Moes, age 23.

1939: child Petrus Herman, age 61, dies.

1881: Catharina Vermeulen, age 38, marries Livinus Scheerens, age 52.

1882: child Abraham Scheerens is born.

1883: partner Livinus Scheerens, age 54, dies.

1903: child Abraham Scheerens, age 20, marries Adriana Simpelaar, age 19.

1928: child Abraham Scheerens, age 45, marries Catholijntje Cornelis, age 39.

1954: child Abraham Scheerens, age 71, dies.

1888: Catharina Vermeulen, age 45, marries Pieter Hubregtsen, age 29.

1911: Catharina Vermeulen, age 68, dies.

Figure 5: Example of a life cycle reconstruction of Catharina Vermeulen (1842-1911) based on certificates where she was mentioned, sectioned by marriage. The death of her daughter Catharina is not present (possibly because after 1963), and the death of the third husband in 1928 is not mentioned because Catharina was not mentioned in his death certificate (he remarried after her death).

V Census information for Zeeland

Since 1795, 1830 and every decade later, censuses were organised in the Netherlands. The aggregated results can be downloaded from <http://www.volkstellingen.nl>. Information on individuals has not been preserved. The counts (see figure 6) give an impression of the steady growth of the population of Zeeland and are therefore indicative for the number of individuals that could be identified from the vital register.

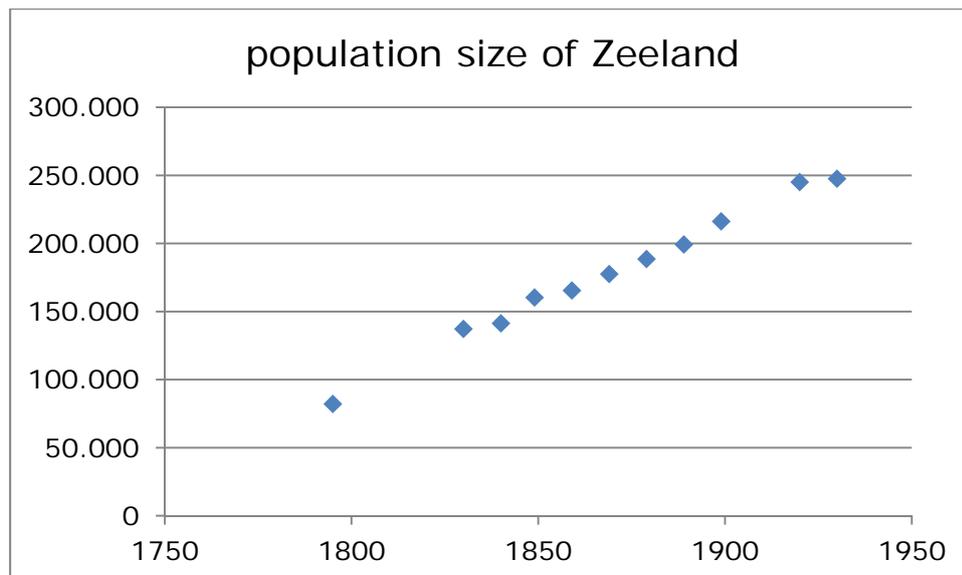


Figure 6. Population size of Zeeland as derived from censuses. The census of 1795 did not include Zeeuws-Vlaanderen, while the census of 1909 is not yet digitized.

From 1859 onwards, the censuses also provide information on the place of birth of the inhabitants, aggregated for 1) the same as the present residence, 2) elsewhere in the province of Zeeland, 3) elsewhere in the Netherlands, 4) in Belgium and other countries. These counts (see table 2) give an impression of the migration to the province, and the percentage of people for which at least no certificate of birth will be found in the vital records of Zeeland (between 7 and 9%). No information is available on the percentage of the population which left the province in the decade. A first estimate could be an equivalence to the figures for in-migration, which may imply that for about another 8 % of the population only a birth certificate and possibly a marriage certificate is present in Zeeland (while they die and/or marry outside the province).

	same municipality	Zeeland	the Netherlands	Belgium
1859	66.3	92.2	97.0	99.9
1869	66.0	92.4	96.8	99.8
1879	67.8	93.7	97.2	99.8
1889	66.6	93.0	96.8	99.3
1899	66.3	93.2	97.2	99.6
1920	64.1	91.2	96.6	
1930	62.6	91.2	97.1	

Table 2. Cumulative percentage of place of birth for the population of Zeeland (no details for Belgium after 1920).

VI Proper names in Zeeland

It may be helpful to give some background on proper names in Zeeland. As has been mentioned before, both men and women keep their first name and family name throughout life in all official registrations in the Netherlands. Thus, women do not change name at marriage (although in daily life they often do). Nevertheless, spelling can vary considerably across certificates.

Although many took a hereditary family name in the 17th century already, after 1811 this was prescribed by French legislation (already after 1796 in the south of the country, including Zeeuws-Flanders). The use of patronyms was abandoned after 1811. A family name can have a prefix as a separate particle (*van Asten, de Vries, van den Berg*). Sometimes, however, the prefix and surname are glued (*Demeulemeester* versus *de Meulemeester*), which is more a Flemish habit (also compare US: Vanderbilt > van de Bilt (De Bilt is a Dutch place name)). As in the Zeeland data prefix and family name are in separate fields this may cause a complication in linking. Noble names may consist of a number of elements (*de Vos van Steenwijk*, which is parsed with prefix *de* and family name *Vos van Steenwijk*).

Most first names are of Germanic or Christian origin, and are Dutch forms of common Western names. Names of women are often derived from male names with a diminutive: *Jan* > *Jantje*. French names are quite common as well, especially in higher social layers. Dutch and French forms may both have been used for the same person (*Willem* – *Guillaume*). In contrast to Protestants, Catholics usually chose the Latinized form of a first name: *Gerrit* > *Gerardus*. The number of first names given to a child increases over time, as is shown for Zeeland in figure 7. More names increase linking likelihood but the order and presence of multiple names may change between certificates.

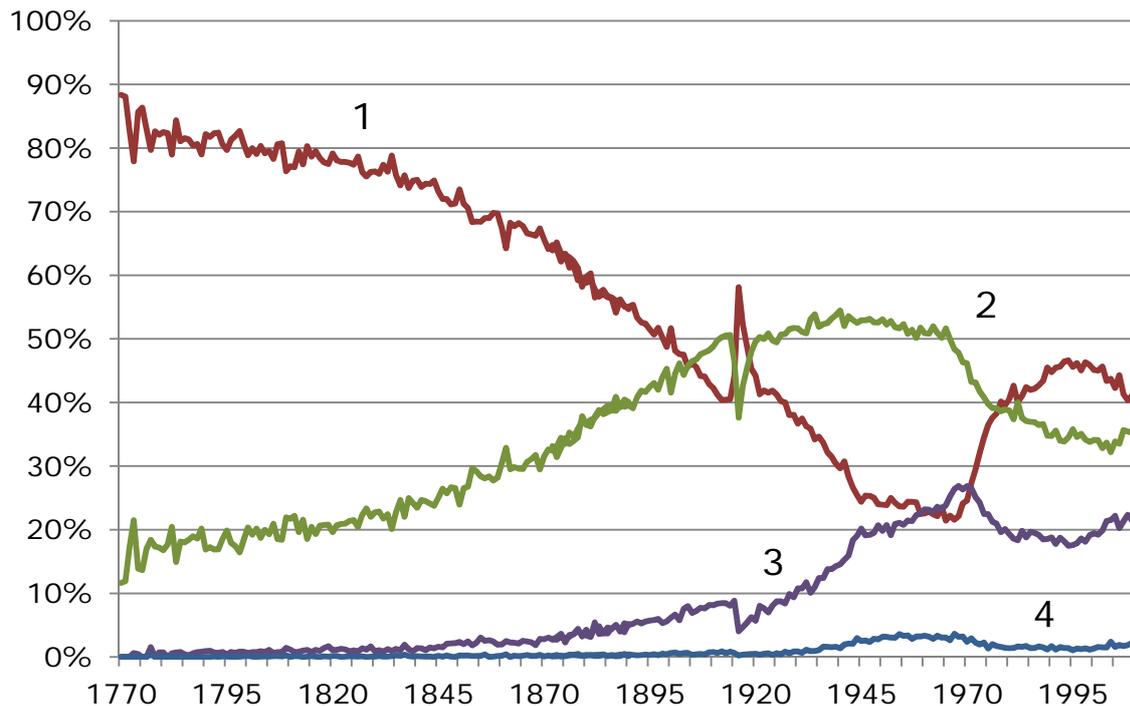


Figure 7. Number of first names (per born child) in the province of Zeeland (1770 -2014)². There is no great difference between boys and girls. The peak around 1916 is due to a proposed first name tax, which never came into effect.

The distribution of first names (and surnames as well) follows Zipf's law, as is found for many name sets in various era's and countries. For the Netherlands (derived from all marriage certificates from the 19th century), figure 8 shows the relationship between the frequency of a male first name on the x-axis and the raw and cumulative number of names with that frequency on the y-axis. The latter relationship is approximately linear on a double logarithmic scale, denoting a power law. The 100 most popular names (at right hand side of figure 7) account for 77% of all names (tokens).

² Bloothoof, G. and Onland, D. (2016), Multiple first names in the Netherlands (1760-2014), *Names*, 64:1, 3-18, DOI: 10.1080/00277738.2016.1118860

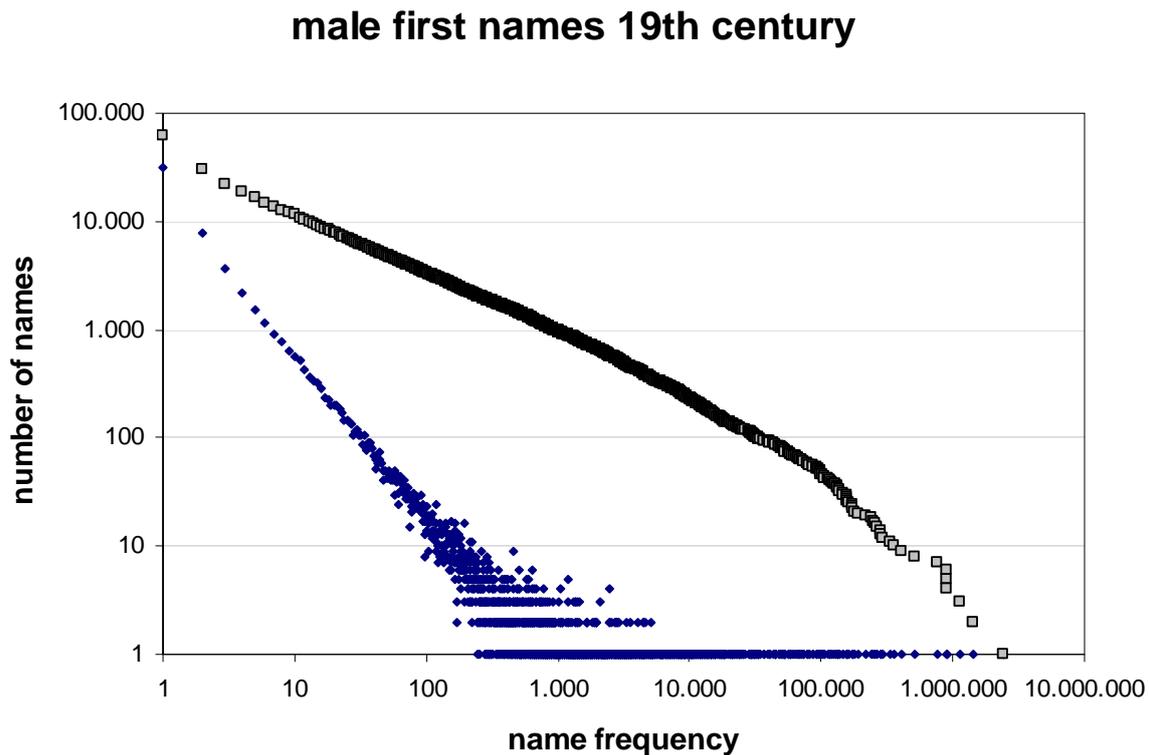


Figure 8. Number of male first names with a certain frequency. Raw data (blue diamonds) and cumulative data (grey squares). The latter demonstrates Zipf's law.

VII Presentation of results

For this benchmark it is essential that a comparison of results across different methods is possible. To understand results, the method used should be documented in such a way that replication by others is an option.

Results should be presented in such a way that comparison across methods is possible. They could be presented at an aggregated level by number and percentage of links found, but also at an individual level as a list of `<id_person>` which refer to that person. As identifier for the individual the identifier of the oldest available record could be used (ideally the id related to birth certificate). The result then becomes a list of lists of all matched `<id_person>`.

If record linkage is done at the certificate level by matching across two certificate types (any combination of certificates on the basis of the names of the parental or bridal couple), results could be presented by the number or percentage of pairs of matched certificates. Explicitly the results could be presented as a list of tuples of `<id_register>`.

Results could also focus on the number of identified individuals and their full lifecycle (i.e. birth, marriage(s), death (of partners), (birth, marriage and death) of children). They could be presented as lists of person ID's originating from consistent events in which an individual may have participated: as child, bride or groom, parent (at birth, marriage and death of children), partner and deceased. Together with their role in the event (as child, bride or groom, parent, partner).

During the processing of the data, double records may be encountered (mainly for certificates of decease). Deduplication results may be presented by listing corresponding the certificate's <register_id>. Deduplication in itself may be the goal of research as well, of course.

Detailed statistics are welcomed, as well as examples of successes, problems and failures.

Whereas the data have been cleaned and standardized to some extent, no standardization of proper names is provided. We welcome results in this area, which could be based on some error distance, but could also be derived as result from the linkage procedure itself.

[We could add our name standards and other enhancements (names of parents or partner explicitly in all ego records) to the release as version 2016_2]

VIII Technical documentation

VIII.1 Introduction

The dataset LINKS_Cleaned_Zeeland_2016_1 consists of three tables: REGISTRATIONS, PERSONS and LOCATIONS. The table REGISTRATIONS consists of the general data from a certificate registration which exceed the individual level; the PERSONS table includes all individuals mentioned in a certificate in separate records. For example, a birth certificate generates three person records: one for the child, one for the mother and one for the father (each with its own <id_person>). These three records share the serial number of the registration <id_registration>, and have a field <role> coding the role in the event, by which the relations between persons is not lost.

In addition there is a LOCATIONS table which provides the standardized names of the locality and/or municipality, province and/or region and country. In the other tables places are represented by code only.

The following number of registrations are included

Birth certificates	698,285
Marriage certificates	192,231
Divorce certificates	1,690
Death certificates	665,999

The HISCO codes of the occupational titles and other codes such as HISCLASS, HISCAM and SOCPO can be derived from the following release:

Kees Mandemakers, Sanne Muurling, Ineke Maas, Bart Van de Putte, Richard L. Zijdeman, Paul Lambert, Marco H.D. van Leeuwen, Frans van Poppel and Andrew Miles, *HSN standardized, HISCO-coded and classified occupational titles, release 2013.01* (IISG Amsterdam 2013). This release can be downloaded from <https://socialhistory.org/en/hsn/hsn-standardized-hisco-coded-and-classified-occupational-titles-release-201301>

The LATLON codes of the locations and other codes such as the Amsterdam code can be derived from the following release: D.P. Huijsmans, *IISG-LINKS Dataset Historische Nederlandse Toponiemen Spatio-Temporeel 1812-2012, Release 2013.2*. This release can be downloaded from

<https://socialhistory.org/en/hsn/historical-dutch-toponyms-spatio-temporal-1812-2012>

In the following description 'N' indicates a numerical value and 'T' indicates a text field.

VIII.2 REGISTRATIONS table

The table REGISTRATIONS consists of general data from a certificate registration which exceed the individual level, such as the date and place of marriage in a marriage certificate. All data in this table are standardized/cleaned with the exception of the field `registration_seq`.

N <code>id_registration</code>	Primary key
N <code>registration_maintype</code>	Main type of the registration as delivered by the archive (the next variable <i>registration_type</i> gives an overview of all possibilities). Values: 1 certificate of birth 2 certificate of marriage/divorce 3 certificate of death
T <code>registration_type</code>	Type of the registration (standardized). Values: g birth certificate h marriage certificate o death certificate s divorce certificate
T <code>extract</code>	Is the registration an extract (summary) of another one? Values: y yes n no u unknown
N <code>registration_location_no</code>	Identifying number of the location of registration; secondary key to the table LOCATIONS.
N <code>registration_day</code>	Day number of the registration
N <code>registration_month</code>	Month number of the registration
N <code>registration_year</code>	Year number of the registration
N <code>registration_flag</code>	Date of the registration may be derived from the date of the event of the registration (not all archives have delivered date of registration).

Values:

- 1 Derived date
- NULL Original date

T registration_seq Sequential number of the registration in the source (this may be an alpha numerical value)

VIII.3 PERSONS table

The table PERSONS contains all appearances of persons. In general every birth certificate generates records for three persons (newborn child, mother and father), a marriage certificate generates minimally six person records (bride, groom, parents of bride and groom and sometimes also the names of former spouses) and a death certificate generate three or four person records (deceased, father, mother and possibly a spouse). All data in this table are standardized/cleaned with the exception of <familyname> and <firstname>.

N id_person Primary key

N id_registration Secondary key to the table REGISTRATIONS

N registration_maintype Maintype of registration.
Values:

- 1 Birth certificates
- 2 Marriage certificates (including divorces)
- 3 Death certificates

T firstname Firstname (including multi-part firstnames)

T prefix Prefix family name

T familyname Family name

T sex Sex is also included if there is only indirect evidence from the role in the registration (such as mother, bride).
Values:

- male
- female
- unknown
- NULL (to be interpreted as unknown)

T civil_status Civil status is included a) if it is explicitly mentioned in the sources, b) from the scanning of

	the remarks.
	Values:
	divorced
	divorced or widowed
	married
	unknown
	unmarried
	widow
	widowed
	widower
N role	Code for role or function of a person in a certificate such as child, father, mother, bride.
	Values:
	1 Newborn child
	2 Mother
	3 Father
	4 Bride
	5 Mother bride
	6 Father bride
	7 Bridegroom
	8 Mother bridegroom
	9 Father bridegroom
	10 Deceased
	11 Partner deceased
T occupation	Occupational title reported at the event
N age_day	Age in days beyond weeks
N age_week	Age in weeks beyond months
N age_month	Age in months beyond years
N age_year	Age in years
T birth_date	Original date of birth (if presented in a single field)
N birth_day	Day number of birth
N birth_month	Month number of birth
N birth_year	Year number of birth
N birth_date_flag	Quality of the date.
	Values:
	0 No date of birth (default-value)
	1 Valid date of birth from source
	2 No date of birth; date of registration used

	as proxy
	3 No valid date of birth; date of registration used as proxy
N birth_location	Identifying number of the location of birth; secondary key to the table LOCATIONS.
T death	Is the person alive or death at the date of the registration? Values: a person is alive n not alive or unknown The value 'a' is deduced from role 1 (birth by definition alive), or role 4 (bride), 7 (bridegroom) or 10 (deceased) with an age in days, weeks, months or years and/or having an occupational title.
T stillbirth	In two ways information of 'lifeless reported' (not always 'stillbirths') was identified. Values: y in the firstname one can find terms such as 'levenloos' y-r lifeless-reported was retrieved from the remarks in the registration
T death_date	Original date of death (if presented in a single field)
N death_day	Day number of death
N death_month	Month number of death
N death_year	Year number of death
N death_date_flag	Quality of the data. Values: 0 No date of death (default-value) 1 Valid date of death from source 2 No date of death; date of registration used as proxy 3 No valid date of death; date of registration used as proxy

N death_location	Identifying number of the location of death; secondary key to the table LOCATIONS.
T mar_date	Original date of the marriage (if presented in a single field)
N mar_day	Day number of the marriage
N mar_month	Month number of the marriage
N mar_year	Year number of the marriage
N mar_date_flag	Quality of the marriage date. Values: 0 No date of marriage (default-value) 1 Valid date of marriage from source 2 No date of marriage; date of registration used as proxy 3 No valid date of marriage; date of registration used as proxy
N mar_location	Identifying number of the location of marriage; secondary key to the table LOCATIONS.

VIII.4 LOCATIONS table

The table LOCATIONS contains the standardized form of the locations that show up in the civil certificates. Location numbers in the tables REGISTRATIONS and PERSONS refer to the field id_location in this table.

A Dutch municipality may consist of several locations, however quite often location and municipality are the same. A Dutch municipality belongs to a province and the whole country has eleven provinces. Zeeland is one of them. For locations outside the Netherland the difference between location and municipality is less clear and provinces are not always recognizable. Usually we used the broader category of region to define entities such as counties in the UK, Kreise in Germany, states in the USA, islands in the Indonesian archipelago, etc.

N location_no	Primary key
T location	Name of the location
T municipality	Name of the municipality in which the location is located (location and municipality could be the same).

T province	Name of the province in which the municipality is located.
T region	Name of the region (or state) in which the location is located.
T country	Name of the country in which the location, municipality, province or region is located.