

Online presentatie van gegevens over voornamen en familienamen

Dr Gerrit Bloothoof

DOC Namen, Meertens Instituut KNAW, Amsterdam

6 februari 2009 [met aanvullingen 17 december 2009 | juni 2010]

Het Meertens Instituut van de Koninklijke Nederlandse Akademie voor Wetenschappen kent een Documentatie en onderzoekcentrum (DOC) Namen. Dit geldt als het Nederlandse expertisecentrum op naamkundig gebied. Een van de taken is publieksvoorlichting over persoonsnamen. Daarvoor kan momenteel op het internet de Nederlandse Voornamen Databank (NVD) en de Nederlandse Familienamen Databank (NFD) geraadpleegd worden. Beide databanken zijn momenteel beperkt in omvang en gegevenspresentatie.

In 2006 heeft het Meertens Instituut een selectie uit de Gemeentelijke Basisadministratie gekregen met betrekking tot de voornamen van alle Nederlanders, en in 2007 is hetzelfde gebeurd voor de familienamen. Daarover is het College Bescherming Persoonsgegevens geïnformeerd.

We zijn voornemens om uit deze GBA bestanden een aantal geaggregeerde gegevens over voornamen en familienamen online te presenteren. Dit voornemen wordt ondersteund door de Nederlandse Vereniging voor Burgerzaken. Omdat daarbij, met betrekking tot te geven van frequenties en informatie over geografische spreiding, aspecten van privacy aan de orde zijn willen wij daarover aan het CBP een toelichting geven.

I. Voornamen, familienamen en privacy

De voornaam wordt meestal als minder gevoelige informatie beschouwd dan de familienaam omdat de voornaam zelden gebruikt wordt als kernveld in andere informatiebronnen. Voor de familienaam ligt dat anders.

We moeten in het algemeen de vraag stellen of privacy in het geding is wanneer:

1. De frequentie van een voor- of familienaam in Nederland wordt gegeven.
2. De frequentie van een voornaam per geboortjaar wordt gegeven.
3. De frequentie van een voor- of familienaam per gemeente in Nederland wordt gegeven.

Bij hoogfrequente gegevens op enig aggregatieniveau bieden de bovengenoemde presenties geen informatie die op persoonlijk niveau als bezwaarlijk kan worden ervaren. Bij kleine aantallen, of zelfs unieke gevallen, kan dat anders zijn.

Ad 1. De kennis dat een persoon met een zeldzame voor- of achternaam in Nederland woont, kan soms ongewenst zijn. Wij hanteren daarom een ondergrens van 5 bij de presentie van frequenties over Nederland¹. Anderzijds is het in het kader van voornaamgeving en familienaamwijziging voor ambtenaren van de burgerlijke stand of rechters van belang om te weten of de naam in Nederland al bestaat. Bij naamfrequenties lager dan 5 wordt daarom wel vermeld dat de naam voorkomt, maar wordt de frequentie zelf niet gegeven.

¹ In Noorwegen www.ssb.no/english/subjects/00/navn_en/ wordt een grens van 3 gehanteerd voor voornamen, familienamen en de combinatie van beide; in Frankrijk www.meilleursprenoms.com/stats/top_english.htm een grens van 5 voor voornamen.

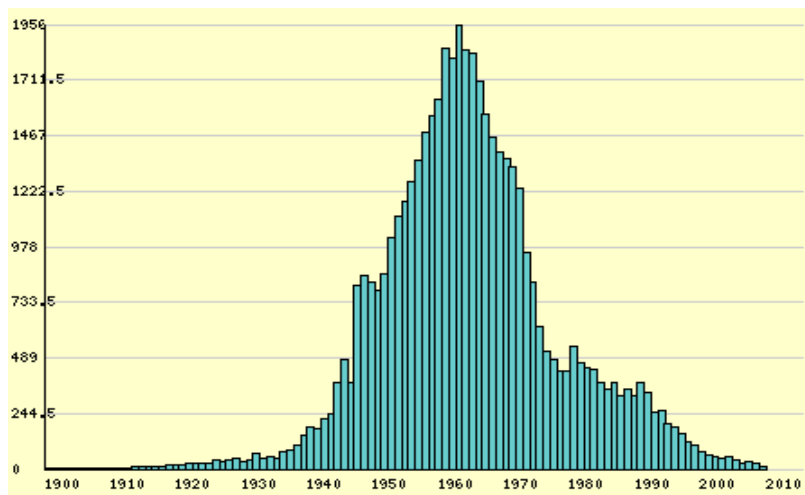
Voorbeeld

De familienaam *Schotanus* komt 1420 keer in Nederland voor.

De familienaam *Adiba* komt minder dan 5 keer in Nederland voor.

De familienaam *Oeteldonk* komt niet voor in Nederland.

Ad 2. De uitsplitsing van voornaamfrequentie naar geboortjaar, is een presentatieverfijning van de totale frequentie in Nederland. Ook hier wordt de grens van 5 per geboortinterval aangehouden². Dat wil zeggen dat wanneer de frequentie per jaar onder 5 ligt, de frequentie van naburige jaren bijeen wordt genomen tot de grens van 5 is gehaald (vervolgens wordt gemiddeld over de betrokken jaren). Naast absolute aantallen kan ook de relatieve frequentie worden getoond, onder dezelfde procedure.

Voorbeeld

Histogram van de populariteit van *Yvonne*. Tussen 1911 en 1916 zijn de gegevens gemiddeld.

Ad 3. Door de frequentie per gemeente te geven, en via een kleurschaal in een kaart te presenteren, wordt een indruk verkregen van de verspreiding van een voornaam of familienaam over Nederland. In overeenstemming met het voorgaande wordt de frequentie niet gegeven wanneer die in Nederland of in een gemeente lager is dan 5. Door inkleuring wordt wel zichtbaar gemaakt in welke gemeenten een familienaam voorkomt, ook als de frequentie lager is dan 5. Voor zeldzame namen, in het bijzonder wanneer die in kleine gemeenten voorkomen, kan dat mogelijk identificatie van individuele personen vergemakkelijken. Daarom worden in zulke gevallen niet voor alle gemeenten informatie getoond.

Daarvoor is een procedure ontworpen, waarbij in eerste instantie werd uitgegaan van een kans op het vinden van een persoon die kleiner is dan 1:500.000. In de praktijk bleek dat tot een aantal ongewenste effecten te leiden. Het was bezwaarlijk dat de plaats waar de meeste naamdragers wonen voor kleine families dan meestal niet werd getoond, omdat daar de directe kans om iemand te vinden te grootst is.

² De Sociale Verzekeringsbank (www.svb.nl) publiceert jaarlijks de babynamen met aantal (ook 1); België (http://www.statbel.fgov.be/figures/d22a_nl.asp) hanteert voor voornamen per jaar een ondergrens van 2

Daarom is vrij snel na lancering van de Nederlandse Familienamen Bank de procedure als volgt aangepast (17 december 2009).

Bij een beslissing om informatie over het voorkomen van naamdragers in een gemeente al dan niet te tonen moet met de onderstaande factoren en aannamen rekening worden gehouden:

- (1) de kans om iemand te vinden is omgekeerd evenredig met het totaal aantal naamdragers van de familienaam: N
> hoe groter het aantal naamdragers, hoe moeilijker het is om een bepaalde persoon daaruit te traceren
- (2) de kans om iemand in een gemeente te vinden is omgekeerd evenredig met het inwoneraantal van die gemeente: I_g
> hoe meer inwoners een gemeente heeft, hoe moeilijk het zal zijn daar iemand te vinden
- (3) de kans om iemand in een gemeente te vinden is evenredig met het aantal naamdragers in die gemeente: n_g
> hoe meer naamdragers in een gemeente wonen, hoe groter de kans is dat een gezochte persoon daarbij zit
- (4) de kans dat iemand geheim wenst te blijven is omgekeerd evenredig met het aantal naamdragers in een gemeente: n_g
> hier gaat het om een aanname over het begrip 'onbekend willen blijven'. De gedachte is dat hoe minder naamdragers in een gemeente wonen, hoe groter de kans is dat daarbij iemand zit die dat geheim wil houden.

Bovenstaande kansen kunnen vermenigvuldigd worden (waarbij er nog een constante factor aan wordt toegevoegd):

P = de vindkans in een gemeente voor iemand die geheim wenst te blijven

$$P = \text{constante} * (1/N) * (1/I_g) * (n_g) * (1/n_g)$$

$$= \text{constante} * 1/(N * I_g)$$

P moet onder een bepaalde waarde blijven. Daar kiezen we $1/K$ voor, wat voorgesteld kan worden als de kans om iemand met een bepaalde familienaam te vinden, die dat niet wil, in een stad met K inwoners.

$$P = \text{constante} * 1/(N * I_g) < 1 / K$$

waaruit volgt

$$I_g > \text{constante} * K/N$$

Omdat we K vrij kunnen kiezen, kunnen we de constante factor daar in opnemen, daarom

$$I_g > K/N$$

Dit betekent dat we voor een bepaalde familienamen alle gemeenten kunnen tonen die aan deze eis voldoen.

Onder deze condities blijkt $K = 200.000$ een redelijke keuze.

$N < 5$	wordt nooit in kaart getoond
$N = 5$	$I_g > 40.000$ (naamdragers worden alleen getoond in gemeenten met meer dan 40.000 inwoners)
$N = 10$	$I_g > 20.000$
$N = 20$	$I_g > 10.000$ (veel gemeenten)
$N > 100$	$I_g > 2.000$ (vrijwel overal)

In enkele gevallen kan het voorkomen dat relatief veel naamdragers in een kleine gemeente niet getoond worden. We gaan er echter van uit dat wanneer er meer dan 10 naamdragers in een gemeente zijn, er van een wens tot geheim blijven geen sprake zal zijn. De kans is immers ook zeer groot dat bijvoorbeeld één van de naamdragers in het telefoonboek staat. Deze naamdragers worden daarom altijd getoond.

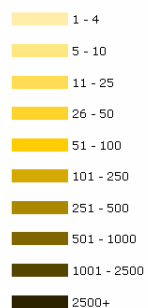
Indien naamdragers niet in alle gemeenten worden getoond komt bij de kaart te staan: "Ter bescherming van de privacy wordt alleen aanwezigheid in gemeenten met meer dan $\langle I_g \rangle$ inwoners getoond." (I_g wordt naar beneden afgerond op hele duizendtallen)

Aanpassing juni 2010

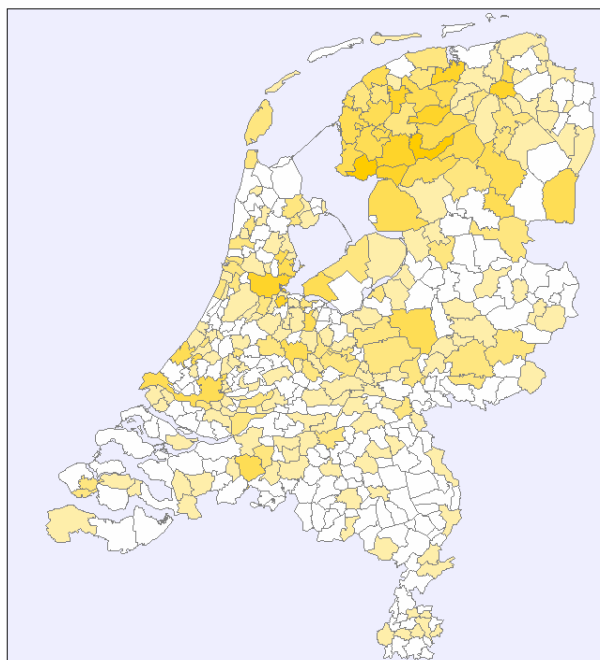
Voor de familienamen is gecontroleerd of een naam in het telefoonboek van 2007 (GBA selectiejaar) voorkwam in een bepaalde gemeente. Indien dat het geval is wordt de naam altijd in die gemeente getoond, ongeacht de uitkomst van voorgaande berekeningen. Als dat niet het geval is dan wordt de tekst bij de kaart "Ter bescherming van de privacy wordt alleen aanwezigheid getoond in gemeenten met meer dan $\langle I_g \rangle$ inwoners of wanneer de naam in het telefoonboek voorkomt."

Voorbeelden

Schotanus : 1.420 personen

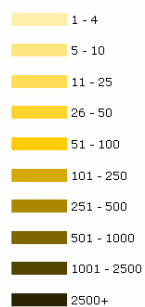


Plaats muispijltje op een gemeente voor het exacte aantal (indien meer dan 4)



De verspreiding van de familienaam *Schotanus*

Bloothoof : 19 personen



Plaats muispijltje op een gemeente voor het exacte aantal (indien meer dan 4)

Ter bescherming van de privacy wordt alleen aanwezigheid getoond in gemeenten met meer dan 10.000 inwoners of wanneer de naam in het telefoonboek voorkomt.



De verspreiding van de familienaam *Bloothoof*, met privacy beperking.

Kanttekeningen

- In de kaartpresentatie ziet de gebruiker door met de muis op een gemeente te gaan staan, zowel de naam van de gemeente als de (relatieve) frequentie (mits ≥ 5)
- De verspreidingskaart van familienamen is gebaseerd op de *woonplaats* in 2007, die van de voornamen op de *geboorteplaats*. Dat maakt het vinden van een persoon die zowel een zeldzame voornaam als een zeldzame familienaam heeft, door het combineren van de beide kaarten aanzienlijk minder kansrijk.
- In het Repertorium Nederlandse Familienamen (Van Gorcum 1962-1980), die gebaseerd is op de volkstelling van 1947, worden per gemeente alle familienamen genoemd, met frequentie tot en met 1.

II. Belangen

Voornamen

Bij publicatie van de Nederlandse voornamenvoorraad is er een groot maatschappelijk belang in verband met de keuze van voornamen door ouders. Een ambtenaar van de burgerlijke stand moet beslissen of een door ouders gekozen voornaam in overeenstemming is met de naamwet. De ambtenaar zou dan moeten weten welke voornamen in Nederland al eerder geaccepteerd zijn. Dat kan de consistentie in besluitvorming sterk bevorderen.

Ook voor ouders kan het veel onzekerheid over toelaatbaarheid van een voornaam wegnemen indien zij over dezelfde kennis zouden beschikken. Het kan als een vorm van openbaarheid van bestuur worden beschouwd dat ouders inzage moeten kunnen hebben in beslissingen die door de ambtenaren van de burgerlijke stand in het verleden zijn genomen.

De Nederlandse Vereniging voor Burgerzaken en het DOC Namen zijn daarom overeengekomen dat het goed zou zijn om via het internet voor elke voornaam informatie te krijgen of deze *onder Nederlands recht* eerder aan een kind is gegeven.

De daarbij te verstrekken gegevens zijn:

De enkelvoudige naam, frequentie als eerste naam voor mannen, frequentie als volgnaam voor mannen, frequentie als eerste naam voor vrouwen, frequentie als volgnaam voor vrouwen.

Toelichting

(1) Met een enkelvoudige voornaam wordt bedoeld een naam zonder spaties, eventueel wel met koppelteken of apostrof. Goedkeuring van een samengestelde voornaam zal vrijwel altijd neerkomen op goedkeuring van de afzonderlijke enkelvoudige voornamen.

(2) Veel voornamen zijn indicatief voor het geslacht van de drager (in Duitsland is dit zelfs verplicht). Maar er zijn ook veel namen die zowel aan jongens als meisjes worden gegeven. Bij een volgnaam, waarbij vaak van vernoeming sprake is, mag er eigenlijk geen discussie rond geslacht ontstaan (bij 1/3 van zowel mannen als vrouwen met een derde of vierde voornaam is die *Maria*). Voor de eerste voornaam, die meestal ook de roepnaam is, kan er mogelijk sprake zijn van ongepastheid, bijvoorbeeld *Roosje* voor een jongen. Daarom wordt er een onderscheid gemaakt in frequenties naar geslacht, en eerste naam of volgnaam.

Voorbeeld:

Jan	man	vrouw
eerste naam	216.706	14
volgnaam	204.585	611

Aadam	man	vrouw
eerste naam	<5	-
volgnaam	-	-

Een verspreidingskaart van een voornaam heeft vooral een cultureel doel. Zo'n kaart geeft inzicht in regionale voorkeuren, en kan daarbij ouders behulpzaam zijn bij het maken van een keuze.

Familienamen

Een bestaande familienaam mag niet als voornaam worden gekozen, tenzij de naam als voornaam gebruikelijk is. Ook mag bij naamsverandering niet voor een bestaande familienaam worden gekozen. In deze gevallen is het voor betrokkenen, ambtenaren van de burgerlijkerlijke stand en rechters, van belang om de huidige familienamenvoorraad van Nederland te kennen. Publiek maken van deze namen is daarom een maatschappelijk belang.

Kaarten met de verspreiding van de familienamen dienen vooral een cultureel doel: ze geven informatie over regionale identiteit en kunnen historici en genealogen behulpzaam zijn. Familienamen kunnen als deel van ons erfgoed worden beschouwd.

Verspreidingskaarten zijn ook via www.familienaam.nl beschikbaar. Deze zijn gebaseerd op de familienamen van telefoonabonnees in 1993. Deze gegevens zijn echter incompleet en vertonen lacunes bij een aantal gemeenten. De verspreiding van familienamen is ook af te leiden uit het Repertorium Nederlandse Familienamen (Van Gorcum 1962-1980), die gebaseerd is op de volkstelling van 1947, maar ook die informatie heeft zijn beperkingen en is, hoewel relevant, gedateerd.