

## Gegevensbestanden van de Volkstellingen 1795-1971

Vorbereitung van de digitale publicatie van de tellingen van 1795, 1829, 1839, 1849, 1859, 1930, 1960 en 1971

### Eindverslag

**Nederlands Instituut voor Wetenschappelijke Informatiediensten (NIWI-KNAW)  
in samenwerking met het  
Centraal Bureau voor de Statistiek (CBS)  
en de  
Historische Databank Nederlandse Gemeenten (HDNG)**

Michelle van den Berk, Peter Doorn en Luuk Schreven

Afdeling Geschiedenis  
Nederlands Instituut voor Wetenschappelijke Informatiediensten (NIWI-KNAW)  
Joan Muyskenweg 25  
Postbus 95110  
1090 HC Amsterdam  
T +31 20 462 8606 F +31 20 665 8013  
E [peter.doorn@niwi.knaw.nl](mailto:peter.doorn@niwi.knaw.nl) H [www.niwi.knaw.nl](http://www.niwi.knaw.nl)

### 1. Inleiding

Het NIWI-KNAW heeft in samenwerking met het Centraal Bureau voor de Statistiek (CBS) en de Stichting HDNG (Historische Databank Nederlandse Gemeenten, waarin participeren de KUN, UvA en het NIDI) het digitaliseringsproject 'Gegevensbestanden van de Volkstellingen 1795-1971' uitgevoerd. Het project bouwde voort op twee eerdere samenwerkingsprojecten tussen deze partijen en hing tevens samen met het door NWO gefinancierde project 'Life Courses in Context' (LCC) in samenwerking met het IISG (Historische Steekproef Nederlandse bevolking).

De KNAW-subsidie is gebruikt ter voorbereiding van de digitale publicatie van een selectie van de Volkstellingen 1795-1971. Het project bestond uit de volgende drie werkpakketten:

Werkpakket 1: Ontsluiting, elektronische publicatie en digitale archivering van de ingevoerde tellingen van 1795, 1829, 1839, 1849, 1859 en 1930. Dit werkpakket is uitgevoerd bij de KUN onder begeleiding van dr. Theo Engelen (KUN Geschiedenis); de beoogde uitvoerder was drs. Hans Heger. Ongelukkigerwijs is de heer Heger reeds in een vroeg stadium van het project uitgeschakeld door ziekte. In goed overleg met de Universiteit Nijmegen is een vervanger gevonden in de persoon van drs. Hennie Gooren. De controle van de telling van 1930 is uitgevoerd door dr. Tom Vreugdenhil van het CBS

Werkpakket 2: Voorbereiding van een digitale publicatie van de Volkstelling 1960. Dit werkpakket is uitgevoerd op het NIWI door drs. Michelle van den Berk onder begeleiding van dr. Peter Doorn (NIWI-KNAW).



Werkpakket 3: Toegankelijk maken en digitale archivering van de Volkstelling 1971. Dit werkpakket is uitgevoerd bij het CBS te Voorburg door drs. Luuk Schreven onder begeleiding van dr. Jacques van Maarseveen (CBS) en dr. Carel Harmsen.

De eindproducten van ieder werkpakket bestaan uit een bestand gereed voor import in een presentatiemedium én een beschrijving van het analyseproces.

Mede op basis van de in dit project verwerkte volkstellingen verscheen in november 2003 de publicatie *Nederland in verandering. Maatschappelijke ontwikkelingen in kaart gebracht, 1800-2000*, onder redactie van Erik Beekink, Onno Boonstra, Theo Engelen & Hans Knippenberg (Aksant, Amsterdam, 2003), ISBN: 90-5260-122-4, 185 blz. met cd-rom, € 20,-.

Over het project zijn verschillende lezingen en congresbijdragen gepresenteerd, zowel in het binnen- als in het buitenland. De belangrijkste voordrachten waren:

Luuk Schreven, Providing access to the Dutch population census of 1971, IASSIST 2003 Conference, May 27-30 2003, Ottawa.

Michelle van den Berk and Peter Doorn: A data archiving disaster: The Dutch population census of 1960. AHC-Conference, 6-9 August 2003, Tromsö.

Luuk Schreven: Providing access to the Dutch population census of 1971. AHC-Conference, 6-9 August 2003, Tromsö.

Peter Doorn and Kees Mandemakers: Life courses in context: a Collaboratory based on Dutch population registers and censuses in the 19th and 20th centuries. AHC-Conference, 6-9 August 2003, Tromsö.

De begeleiding van het project was in handen van een begeleidingscommissie, samengesteld uit vertegenwoordigers van de participerende instellingen. Deze commissie bestond uit de volgende personen:

dr. Carel Harmsen (CBS)

dr. Tom Vreugdenhil (CBS)

dr. Jacques van Maarseveen (CBS)

dr. Peter Doorn (NIWI-KNAW)

dr. Theo Engelen (KUN)

De financiële rapportage en verantwoording is in verband met de invoering van een nieuw tijdschrijfsysteem bij het NIWI nog niet geheel afgerond en zal zo spoedig mogelijk afzonderlijk worden ingediend.

## 2. Werkpakket 1 “Oude tellingen”

### **Deliverables:**

Bestanden met gecontroleerde en geanalyseerde data die gereed zijn om in te brengen in CBS StatLine.

Rapporten met beschrijvingen van het analyseproces bij de diverse volkstellingen.

Heger en Gooren hebben de tellingen van 1795-1859 gecontroleerd en gecorrigeerd en enkele ontbrekende delen van tellingen zijn ingevoerd. Een uitvoeriger verslag van de verrichte werkzaamheden is afzonderlijk beschikbaar. Samengevat bestonden de belangrijkste werkzaamheden in dit werkpakket uit het volgende:



**Volkstelling 1795**

De diverse tabellen en lijsten zijn in de onderhavige digitale bron integraal opgenomen en in Excel gecontroleerd. Bovendien zijn de gegevens over inwonertallen in de diverse lijsten onderling vergeleken. De totalen zijn herberekend en ook weer onderling vergeleken. Voor de provincies Noord- en Zuid-Holland zijn de gegevens uit deze bron ook nog vergeleken met "Het Volk van Holland"<sup>1</sup>. Bij geconstateerde fouten is het juiste gegeven vermeld. In de toegevoegde kolom met opmerkingen is aangegeven welke wijziging er heeft plaats gehad.

**Volkstelling 1830**

De resultaten van de volkstelling van 1830 zijn niet in hun geheel, maar in gedeelten en op verschillende momenten in de 19<sup>e</sup> eeuw gepubliceerd. De Volkstelling van 1830 zoals in deze digitale bron gepresenteerd, is derhalve gebaseerd op meerdere gedrukte bronnen. Per bestand is vermeld welke publicatie de grondslag vormde. De telling is ingevoerd door Hans Heger en Henny Gooren, en vervolgens gecontroleerd in Excel op data-entryfouten en bronfouten.

**Volkstelling 1840**

De algemene volkstelling van 1840, zoals gepresenteerd in deze digitale bron, is gebaseerd op de gedrukte uitgave van 1841, en op de aanvullingen en correcties die als bijlage van de telling van 1859 zijn gepubliceerd. Deze telling is ingevoerd door Hans Heger en Henny Gooren, en vervolgens gecontroleerd in Excel op data-entryfouten en bronfouten.

**Volkstelling 1849**

De volkstelling van 1849 is in 1852 gepubliceerd, in 4 banden. Per provincie en voor het Rijk heeft de VT1849 de volgende indeling: kleine tabellen, 1<sup>e</sup> deel, 2<sup>e</sup> deel, 3<sup>e</sup> deel, 4<sup>e</sup> deel. De gegevens zoals in de digitale bron gepresenteerd, zijn gebaseerd op deze publicatie. De oorspronkelijke opzet is gehandhaafd. Wel zijn er enkele kolommen toegevoegd die het gebruik kunnen vergemakkelijken. In de kolommen noten en opmerkingen zijn de noten uit de gedrukte bron opgenomen, alsmede de errata die in de diverse delen achteraf zijn opgenomen. Daarnaast zijn er twee kolommen opgenomen met data-entry fouten en bronfouten.

**Volkstelling 1859**

De volkstelling van 1859 is gepubliceerd in 1863 in 5 banden. Zij heeft een geheel andere opzet dan die van 1849. De indeling is sterker gedifferentieerd. Deel 1 is onderverdeeld in een 1<sup>e</sup> tot en met 5<sup>e</sup> gedeelte, waarbij enkele gedeelten weer gescheiden zijn in A, B, en C. De telling is tevens onderverdeeld in een feitelijke en een werkelijke bevolking. De controles op de ingevoerde volkstelling zijn uitgevoerd in Excel. Een deel van de controle van 1859 is niet geheel gereedgekomen. Besloten is om de resterende controles uit te voeren in het kader van het project Life Courses in Context.

**Volkstelling 1930**

De volkstelling 1930 is gecontroleerd door Tom Vreugdenhil van het CBS. De publicaties van deze telling betaan uit tien delen. De data-entry bestanden van de eerste 9 delen, voornamelijk bestaande uit tabellen, zijn omgezet naar Excel spreadsheets, met daarin per tabel, staat of lijst een werkblad. Per (papier) deel is één Excel spreadsheet gemaakt. Deel 10, dat uit allerlei lijsten en classificaties bestaat, is omgezet in een Access-database. De diverse lijsten zijn eenduidig en onderling consistent gemaakt; daarbij is een aantal fouten verbeterd. Aan de items zijn, waar nodig, codes toegevoegd om er codelijsten voor

---

<sup>1</sup> Rapport van de commissie tot de telling en verdeling van het volk van Holland (Den Haag 1796)

StatLine mee te kunnen maken. Een en ander is in twee documenten beschreven (één voor de conversie naar Excel en één voor de codelijsten).

De controle van deze telling is nog niet geheel gereed, onder andere omdat gebleken is dat een deel van de gegevens ontbreekt. De ontbrekende tabellen (gepubliceerd op ca. 140 pagina's) zullen ingevoerd worden in het kader van het LCC-project.

### 3. Werkpakket 2: Volkstelling 1960

#### **Deliverables:**

Nauwkeurige inventarisatie van de lacunes en afwijkingen in de digitale volkstelling van 1960 ten opzichte van de gepubliceerde boeken en gelichtdrukte tabellen.

Projectplan voor de aanvulling en correctie van de gegevens op basis van de lichtdrukken met kostenraming.

Beschrijving van publicatieformaat voor Volkstelling 1960.

Ten behoeve van de reconstructie en ontsluiting van de volkstelling van 1960 op basis van de in het Steinmetz Archief aanwezige dataset p0061 zijn de volgende activiteiten verricht (een uitvoeriger rapport over dit werkpakket is afzonderlijk beschikbaar):

#### **3.1. Inventarisatie missende en dubbele records:**

Deze inventarisatie is uitgevoerd door een vergelijking te maken van de aantallen mannen en vrouwen per gemeente uit de gepubliceerde tabellen van het CBS met die in dataset p0061 van het Steinmetz Archief.

Uit documentatie van het Steinmetz Archief was reeds eerder gebleken dat er grote hoeveelheden records misten. Aan de andere kant bevatten sommige gemeenten ook veel te veel records. Door deze vergelijkingen werd duidelijk om hoeveel gemeenten het exact ging en om welke aantallen records, zowel absoluut als relatief. Op deze getallen zijn de aanbevelingen voor reconstructie uit het rapport gebaseerd.

Vervolgens is een studie gemaakt van de bij de *data cleaning* door het Steinmetz Archief afgewezen records en de records uit de ASCII bestanden (gebaseerd op eertijds ingelezen ponskaarten *zonder data cleaning*). Bij het aanmaken van dataset p0061 (een SPSS bestand) heeft het Steinmetz Archief een aantal regels voor *data cleaning* toegepast. Een aantal uitgesloten records is weliswaar vervuild, maar bevat wel degelijk waardevolle informatie, en zou alsnog in de dataset opgenomen moeten worden. Overigens werd hierbij ook duidelijk dat de data cleaning en conversie niet geheel correct zijn verlopen: dataset p0061 bevat wel degelijk coderingen, die niet in het codeboek voorkomen en gecorrigeerd moeten worden. Uit de ASCII bestanden, waarin de records nog op de originele volgorde staan (d.w.z. zoals de ponskaarten zijn ingelezen), werd duidelijk welke records dubbel in het bestand zitten en verwijderd kunnen worden.

#### **3.2. Data invoer voor missende gemeenten en gemeenten met ernstige afwijkingen in de aantallen mannen en/of vrouwen**

Van sommige gemeenten ontbreken zoveel records (soms zelfs alle) dat het zinvol is om aanvullende informatie te leveren. Deze kan worden gecreëerd uit de handgeschreven tabellen in het archief en bibliotheek van het CBS (deze zgn. losbladige 'transparanten' bevinden zich in het archief; daarvan zijn lichtdrukken gemaakt, die zich in gebonden vorm in de bibliotheek bevinden). Op grond van analyse van de ontbrekende gegevens is voorgesteld om de geaggregeerde gegevens in te voeren voor de gemeenten waarbij meer dan 5% of meer dan 1000 mannen of vrouwen ontbreken. Het wordt niet rendabel geacht om ook de geaggregeerde tabellen in te voeren van de ca. 500 gemeenten waarbij tussen de 1 en 1000 records (of minder dan 5% van het totaal aantal mannen of vrouwen) missen. Op basis van een proef met twee gemeenten is gebleken dat data-entry van de geaggregeerde tabellen



van de ca. 30 betreffende gemeenten haalbaar was in combinatie met het Life Courses in Context project. Deze handgeschreven tabellen van het CBS zijn inmiddels door het NIWI handmatig ingevoerd.

### 3.3. Publicatieformaat

Op grond van discussies met historici wordt aanbevolen om de VT 1960 op individueel niveau beschikbaar te stellen. De ca. 300.000 ontbrekende records, die grotendeels betrekking hebben op ca. 30 gemeenten, worden door historici voor lief genomen. Voor deze gemeenten kunnen de geaggregeerde tabellen worden aangeboden.

De wijze van toegang tot de individuele data kan op dezelfde wijze plaatsvinden als bij de Volkstelling 1971, namelijk via *remote execution* op een CBS-server (zie hierna). De geaggregeerde tabellen kunnen op dezelfde wijze worden ontsloten als de volkstellingstabellen voor de andere volkstellingsjaren, namelijk in StatLine en als Excel-spreadsheets.

## 4. Werkpakket 3: Volkstelling 1971

### **Deliverables:**

On-site access tot de Volkstelling van 1971 voor onderzoekers.

Volledig gedocumenteerde digitale Volkstelling van 1971 ter digitale archivering.

De gegevensbestanden van de volkstelling van 1971 zijn in het digitale CBS archief opgeslagen. Binnen het derde werkpakket zijn deze bestanden gecontroleerd op volledigheid, gedocumenteerd en worden ze binnenkort ontsloten.

De bestanden zoals ze werden aangetroffen waren niet compleet. Van één Nederlandse gemeente ('s-Hertogenbosch) ontbrak de helft van het aantal inwoners. Door terug te grijpen op oudere bestanden die opnieuw geconverteerd moesten worden is het gelukt om de ontbrekende gegevens opnieuw aan de bestanden toe te voegen. Om de bestanden helemaal te completeren zijn ook de ontbrekende labels van enkele variabelen aan de bestanden toegevoegd.

Toen de bestanden volledig gereconstrueerd waren, is begonnen aan de samenstelling van een omvangrijk documentatierapport. Dit rapport is als bijlage bij dit eindverslag toegevoegd. Het beschrijft de structuur van de databestanden en geeft voor een groot aantal variabelen een toelichting op de wijze waarop deze tot stand zijn gekomen en een definitie. Er zijn overzichten opgenomen van de variabelen, de voorkomende waarden en van frequentietellingen over deze variabelen.

Het derde onderdeel van het werkpakket richtte zich op de ontsluiting van de gegevens. In nauw overleg met het CBS is besloten om de gegevens niet alleen on-site voor onderzoekers beschikbaar te stellen, maar ze ook via *remote-execution* aan te bieden. Het eerste betekent dat onderzoekers naar het CBS moeten afreizen om de gegevens lokaal te analyseren. De methode van *remote-execution* richt zich op de analyse op afstand en werkt via e-mail. Een onderzoeker kan een analyse opdracht in SPSS syntax doorsturen naar het CBS alwaar deze op de data wordt gedraaid.

Om een werkende analyse opdracht te formuleren is het noodzakelijk dat onderzoekers zich in de structuur en de inhoud van de bestanden kunnen verdiepen. Daartoe dient niet alleen het documentatierapport, ook een proefbestand is van belang. Het proefbestand behorende bij de gegevensbestanden van de volkstelling 1971 bestaat uit een kleine (0,01%) steekproef uit de gegevensbestanden van de volkstelling. Een rapportage over de opzet van het proefbestand is als bijlage bij deze rapportage gevoegd.

Binnenkort wordt er een experiment met een beperkt aantal onderzoekers opgezet om deze manier van ontsluiting te testen. Ook worden geaggregeerde tabellen uit de



volkstellinggegevens gegenereerd voor ontsluiting via CBS-Statline en de website [www.volkstelling.nl](http://www.volkstelling.nl).

**Afzonderlijke rapporten als bijlagen:**

**Werkpakket 1:**

Hennie Gooren, Hans Heger en Tom Vreugdenhil: Controle en Correctie van de tellingen van 1795, 1830, 1840, 1849, 1859 en 1930

**Werkpakket 2:**

Michelle van den Berk & Peter Doorn: The reconstruction of the digital Dutch Population Census of 1960

**Werkpakket 3:**

Luuk Schreven: Documentatierapport volkstelling 1971

