

**Aanvraag voor een subsidie uit het digitaliseringsfonds KNAW
ter voorbereiding van de
digitale publicatie van een selectie van de**

Gegevensbestanden van de Volkstellingen 1795-1971

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

Aanvrager:

Nederlands Instituut voor Wetenschappelijke Informatiediensten (NIWI-KNAW)
Joan Muyskenweg 25
Postbus 95110
1090 HC Amsterdam

contactpersoon:

dr. P.K. Doorn (Hoofd Afdeling Geschiedenis)
T +31 20 462 8606 F +31 20 665 8013
E peter.doorn@niwi.knaw.nl H www.niwi.knaw.nl

1. Doel en korte omschrijving van het project

Het NIWI-KNAW stelt voor om 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) een digitaliseringsproject uit te voeren 'Gegevensbestanden van de Volkstellingen 1795-1971'. Het project bouwt voort op twee eerdere samenwerkingsprojecten tussen deze partijen en dient tevens ter voorbereiding op een bij NWO ingediende aanvraag in het kader van de investeringen in grote apparatuur in samenwerking met het IISG (Historische Steekproef Nederlandse bevolking). De KNAW-subsidie zal worden gebruikt ter voorbereiding van de digitale publicatie van een selectie van de Volkstellingen 1795-1971.

Het project zal bestaan uit de volgende drie werkpakketten:

1. Ontsluiting, elektronische publicatie en digitale archivering van de ingevoerde tellingen van 1795, 1829, 1839, 1849, 1859 en 1930.
2. Voorbereiding van een digitale publicatie van de Volkstelling 1960.
3. Toegankelijk maken en digitale archivering van de Volkstelling 1971.

Nadat de drie werkpakketten zijn gerealiseerd, ontstaat enerzijds een reeks van gedigitaliseerde tellingen, die alle toegankelijk zullen zijn in StatLine: 1795, 1849, 1859, 1899, 1930, (en in een



latere fase) 1960 en 1971. Hiermee wordt een uiterst waardevol instrument voor het historisch en sociaal-wetenschappelijk onderzoek in Nederland gecreëerd.

Anderzijds ontstaat een zo compleet mogelijk geïntegreerd bestand van historische gegevens op gemeentelijk niveau van 1795 tot ca. 1995. Behalve de genoemde volkstellingen worden diverse reeds bestaande databestanden hierin geïntegreerd, zoals de Historisch Ecologische Databank van de UVA en de Hofstee-bestanden van het NIDI. Een historisch coördinatenbestand om de gegevens op gemeenteniveau cartografisch weer te geven is beschikbaar (NLKAART).

IWI en het gebiedsbestuur maatschappijwetenschappen van NWO financierden een eerder uitgevoerd project 'Digitalisering Volkstellingen 1795-1971' van NIWI-KNAW en CBS. Het HDNG-project is eerder gesubsidieerd door het Gebiedsbestuur Geesteswetenschappen van NWO. Het CBS draagt op diverse manieren bij aan het voorgestelde project: door het beschikbaar stellen van het reeds ingevoerde materiaal (de data-entry is reeds door het CBS bekostigd) en het beschikbaar stellen van CBS-StatLine software voor de ontsluiting van het materiaal. Bovendien zijn enkele medewerkers van het CBS, die ook in het eerdere project 'digitalisering volkstellingen 1795-1971' een centrale rol vervulden, graag bereid om in het vervolgproject te figureren (i.c. de heren Jacques van Maarseveen, Jan Jonker en Tom Vreugdenhil). Het CBS stelt hen en/of enkele andere medewerkers voor een deel van hun tijd in de gelegenheid om in een nieuw samenwerkingsproject te participeren.

De omvang van het project beslaat 28 arbeidsmaanden en de doorlooptijd is anderhalf jaar (18 maanden). In het project worden drie werkpakketten onderscheiden.

2. Kader van het project

2.1. *Voorgeschiedenis digitalisering Volkstellingen*

Van 1997 tot 1999 hebben het CBS en het Nederlands Instituut voor Wetenschappelijke Informatiediensten (NIWI-KNAW) gezamenlijk een project uitgevoerd onder de titel 'Digitalisering Nederlandse Volkstellingen 1795-1971'. Dit project is succesvol afgerond en heeft geresulteerd in de productie van twee sets Cd-rom's, een gezamenlijke Website, een wetenschappelijk symposium en een boek met de resultaten van het symposium.¹ In de pers is het project lovend besproken. Ook door onderzoekers is enthousiast gereageerd op de resultaten.

In het project zijn alle gepubliceerde volkstellingsboeken (ca. 200 delen met in totaal ca. 42.500 pagina's) in de vorm van digitale afbeeldingen (images) ontsloten. Van de volkstelling van 1899 (incl. beroeps- en woningtelling, 26 delen, ca. 10.000 pagina's) is een volledig doorzoekbare en analyseerbare database gemaakt (ontsloten en toegankelijk via CBS StatLine).

De gegevens zijn zowel op CD-ROM als op Internet raadpleegbaar (zie www.volkstelling.nl en www.cbs.nl). De digitale archivering van de elektronische volkstellingsgegevens vindt plaats door de NIWI-data-archieven (Steinmetzarchief en NHDA). Er wordt nauw samengewerkt met het Wetenschappelijk Statistisch Agentschap van NWO bij het verschaffen van toegang tot digitale CBS-bestanden. Overleg tussen NIWI-KNAW, WSA en CBS is gaande over de mogelijkheden om deze samenwerking nog uit te breiden op het gebied van de digitale archivering van de historische CBS-bestanden.

De Volkstellingen van 1960 en 1971 bestaan nog in digitale vorm, maar niet op een dusdanige wijze dat deze voor onderzoekers toegankelijk zijn. De Volkstellingsboeken van de jaren 1795, 1849, 1859 en 1930 zijn ingetypd, maar deze ruwe bestanden zijn zonder toelichting en documentatie onbruikbaar voor onderzoek. Tijdens het project is ook nagegaan of en op welke wijze de handgeschreven transparanten (en daarvan gemaakte lichtdrukken), die uiterst

¹ J.G.S.J. van maarseveen en P.K. Doorn, red., *Nederland een eeuw geleden geteld: een terugblik op de samenleving rond 1900* (Amsterdam, 2001).

gedetailleerde informatie geven van de tellingen van na de tweede wereldoorlog, gedigitaliseerd kunnen worden.

2.2. *Voorgeschiedenis Historische Databank Nederlandse Gemeenten (HDNG)*

In 1999 is een samenwerkingsproject van start gegaan met als partners de Stichting HDNG en het CBS, met als doel om statistische gegevens uit een groot aantal bronnen en databases met betrekking tot Nederlandse gemeenten over de laatste twee eeuwen bijeen te brengen. Aan het project nemen de volgende onderzoekers deel:

dr. Hans Knippenberg (UVA) – Historisch Ecologische Databank
dr. Theo Engelen (KUN)
dr. François Hendrickx (KUN)
drs. Hans Heger (KUN)
dr. Onno Boonstra (KUN) – Historisch-Kartografisch coördinatenbestand NLKAART
drs. Erik Beekink (NIDI) – Historisch-demografische Hofstee-bestanden (afkomstig van LUW)

Van de zijde van het CBS nemen deel aan het project:

dr. Tom Vreugdenhil
dr. Jacques van Maarseveen
drs. Jan Jonker

In de afgelopen periode zijn door HDNG-medewerker H. Heger de data van de volkstellingen 1829 en 1839 ingetoetst. Daarna is hij belast met de controle, de analyse en het schrijven van een notenapparaat. Intoetsfouten zijn verbeterd en zogenaamde 'bronfouten' zijn opgespoord, geanalyseerd en verwerkt in het notenapparaat. De beschrijving van de controle en analyse levert met name voor de volkstelling van 1839 nuttige inzichten op.

Sinds april 2000 werkt Heger intensief samen met de heren J. Jonker en T. Vreugdenhil van het CBS. In overleg met en onder begeleiding van de genoemde medewerkers van het CBS is besloten om te beginnen met de bewerking van de omvangrijke telling van 1849.

De data-archieven van het NIWI-KNAW zijn bij het project betrokken voor de digitale duurzaamheid van de bestanden die in het project tot stand komen.

3. **Wetenschappelijke en innovatieve waarde van het project**

Nationale volkstellingen behoren tot de meest elementaire informatiebronnen over de toestand in een land. Naast de omvang van de bevolkingsgrootte bevat de VT doorgaans informatie over de structurele kenmerken van een land, zoals leeftijd, geslacht, burgerlijke staat, levensbeschouwing, huishoudenssituatie, beroepswerkzaamheid en nationaliteit. In diverse jaren is de VT gecombineerd gehouden met een beroepstelling (BRT) en een woningtelling (WT).

De eerste algemene volkstelling in Nederland vond plaats in 1795 onder de Bataafse Republiek. Sedert 1829 is er sprake van periodieke tellingen, die eens in de 10 jaar werden gehouden. De telling van 1940 werd in verband met de oorlog uitgesteld tot 1947. Na 1971 is in Nederland geen VT meer gehouden als gevolg van de waargenomen non-respons van ca. 25 % bij de proeftellingen voor de voorgenomen Volkstelling 1981.

De digitale Volkstellingen zullen een belangrijk instrument voor historisch en sociaal-wetenschappelijk onderzoek vormen. In tal van landen worden of zijn (nationale) historische censusprojecten uitgevoerd, zoals in de VS, Engeland, Ierland, Frankrijk, Noorwegen, Denemarken, Duitsland, Rusland en Oostenrijk. In diverse gevallen zijn deze projecten gebaseerd op het oorspronkelijke basismateriaal, waardoor databases op individueel niveau zijn opgebouwd.

Afgezien van individuele gegevens ('microdata') vormt het gemeentelijk niveau waarschijnlijk het meest door onderzoekers gebruikte niveau voor de analyse van sociaal-economische, culturele en

demografische gegevens over het Nederlandse verleden. Reeds vele jaren is gewerkt aan en wordt gebruik gemaakt van bijvoorbeeld de historisch-ecologische databank (HED), die in de HDNG wordt geïntegreerd. Bij de HED gaat het namelijk om gegevens op gemeenteniveau. Dit geldt evenzeer van de bij het NIDI berustende gegevens uit het Hofstee-archief, dat historisch-demografisch materiaal uit de afgelopen ruim anderhalve eeuw bevat gebaseerd op gemeentelijke bevolkingsgegevens. De bundeling van HED en Hofstee-archief, aangevuld met nog niet eerder digitaal beschikbare volkstellingsgegevens, zal een gemeentelijke databank opleveren, die uitnodigt tot onderzoek naar nog niet eerder gelegde verbanden. Deze databank zal voor onderzoekers van verschillende disciplines (historici, geografen, sociologen, demografen, epidemiologen) van grote waarde zijn, juist door de koppeling en integratie van gegevens uit uiteenlopende bronnen en gegevensbestanden.

Uit raadplegingsgegevens van de website van het project 'Digitalisering Volkstellingen 1795-1971' (ruim 200.000 bezoeken in 2000) en verkoopcijfers van de geproduceerde CD-ROM's (ruim 500 sets CD-ROM's verkocht) is bekend dat deze gegevens niet alleen een uiterst belangrijk instrument voor wetenschappelijk onderzoek opleveren, maar dat deze tevens tot de verbeelding van het publiek in het algemeen spreken. Te denken valt aan de volgende doelgroepen:

- Amateur-historici geïnteresseerd in lokale geschiedenis, heemkundige kringen e.d.
- Provinciale en lokale overheden (kwantitatieve gegevens over het verleden van de gemeente).
- Middelbaar en hoger onderwijs.

Archieven en bibliotheken treden dikwijls op als intermediair bij het aanbieden van de gegevens.

In informatie-technologisch opzicht is het project noch voor het NIWI-KNAW, noch voor de partners duidelijk innovatief. De projectpartners maken gebruik van technologieën die reeds zijn ontwikkeld, en waarmee zij reeds ervaring hebben opgedaan. Het innovatieve element is geheel gelegen in het beschikbaar maken van voor diverse vakgebieden relevante *content* op een gebruiksvriendelijke manier. De gemakkelijke toegang tot de informatie die beschikbaar komt, zal wel een stimulans betekenen voor innoverend onderzoek op deze vakgebieden.

4. Projectbeschrijving: werkpakketten en producten (*deliverables*)

4.1. Werkpakket 1: Ontsluiting, elektronische publicatie en digitale archivering van de ingevoerde tellingen van 1795, 1849, 1859 en 1930

Werkpakketleider

HDNG (in dezen vertegenwoordigd door de KUN, dr. Th. Engelen), in samenwerking met het CBS.

Doel

De gemeentelijke gegevens zullen worden opgenomen in de HDNG; de gegevens zullen tevens als elektronische publicatie per Volkstellingsjaar toegankelijk worden gemaakt. Hiervoor zal gebruik worden gemaakt van CBS-StatLine. Het materiaal dient te worden gecontroleerd en waar nodig gecorrigeerd, gedocumenteerd en geconverteerd naar StatLine.

Deliverables

Het werkpakket leidt tot twee concrete en welomschreven eindproducten:

- bestanden met gecontroleerde en geanalyseerde data die gereed zijn om in te brengen in een presentatiemedium (CBS StatLine).
- rapporten met beschrijvingen van het analyseproces bij de diverse volkstellingen.

Beschikbaarheid materiaal

Bij het CBS berusten de reeds digitaal ingetoetste uitkomsten van de volkstellingen van 1795, 1849, 1859 en 1930. Het betreft hier de data in de meest ruwe vorm, zonder enige meta-informatie en documentatie. Hoe basaal de huidige vorm van opslag ook moge zijn, de data zijn in potentie van grote waarde voor onderzoekers uit velerlei disciplines. De gegevens zijn in hun huidige ruwe vorm echter niet beschikbaar en niet bruikbaar voor onderzoekers. Om de gegevens tot vrucht te brengen in presentabele en verantwoorde databanken en elektronisch te publiceren, is in de eerste plaats een grondige controle, analyse, correctie en documentatie van de data noodzakelijk.

Controles en consistentie-checks

De ruwe data dienen te worden onderworpen aan allerlei controles (bijvoorbeeld controles van horizontale en verticale totalen en subtotalen, checks aan de bron) en analyses van de gevonden fouten. Vanzelfsprekend zal van deze werkzaamheden per volkstelling verantwoording worden afgelegd in een nauwkeurige beschrijving van het analyseproces.

Elektronische publicatie en digitale duurzaamheid

De databestanden zullen elektronisch worden gepubliceerd met behulp van de door het CBS ontwikkelde StatLine-programmatuur. De gemeentelijke gegevens uit de volkstellingen zullen tevens worden opgenomen in de Historische Databank Nederlandse Gemeenten. Alle gegevens zullen vervolgens elektronisch worden gearchiveerd in het Nederlands Historisch Data Archief van het NIWI-KNAW. Hiermee wordt de digitaal duurzame toegankelijkheid van de data, onafhankelijk van presentatievorm, software of hardwareplatform, gegarandeerd.

Personele inzet en doorlooptijd

Voor alle voorbereidende werkzaamheden voor de elektronische publicatie en digitale archivering van de volkstellingen van vóór 1960 zal gebruik worden gemaakt van de expertise van drs. H. Heger, onder begeleiding van dr. O. Boonstra en dr. Th. Engelen (KUN) alsmede van drs. J. Jonker en drs. T. Vreugdenhil (CBS). De werkzaamheden zullen een doorlooptijd van achttien maanden hebben en een gemiddelde personele inzet van drie werkdagen per week. De begrote kosten (personeel) bedragen Kfl. 95

Fasering

Op basis van de ervaringen van het CBS met de uitgave van de Volkstelling 1899 en de controles en annotatie van de tellingen van 1829, 1839 en (deels) 1849 is de fasering van de werkzaamheden als volgt gepland:

- Maand 1-4: afronding bewerking van volkstelling 1849.
- Maand 5-10: bewerking volkstelling 1859.
- Maand 11-12: bewerking van de relatief kleine volkstelling 1795.
- Maand 13-18: bewerking volkstelling 1930. Deze telling is in vergelijking met de negentiende-eeuwse voorgangers erg omvangrijk en complex, maar de uitvoerder kan profiteren van reeds uitgevoerd verkennend onderzoek door medewerkers van het CBS.

In alle vier gevallen bestaan de eindproducten uit een bestand gereed voor import in een presentatiemedium én een beschrijving van het analyseproces.

4.2. Werkpakket 2: Voorbereiding van een elektronische publicatie van de Volkstelling 1960

Werkpakketleider

NIWI-KNAW (dr. P.K. Doorn); in samenwerking met het CBS.

Doel

Onderzoeken en voorbereiden van de mogelijkheden om de Volkstelling van 1960 digitaal te restaureren en publiceren.

Beschikbaarheid materiaal

Een digitale versie van de Volkstelling van 1960, afkomstig van het CBS, is aanwezig in het Steinmetzarchief, onderdeel van het NIWI-KNAW. Eerder onderzoek heeft uitgewezen dat dit digitale materiaal lacunes bevat en soms afwijkingen vertoont ten opzichte van de gepubliceerde tellingen. Op grond van een verkenning lijkt het echter mogelijk om de lacunes aan te vullen en te corrigeren met gegevens uit het CBS-archief (de zogenaamde transparanten en lichtdrukken). Nader onderzoek moet worden of het publicatieformaat bij dit papieren formaat kan aansluiten: de gegevens zouden in dat geval op laag aggregatieniveau toegankelijk gemaakt worden. Dit is het schaalniveau waarop de gegevens altijd al beschikbaar zijn geweest voor onderzoekers in de bibliotheek van het CBS. StatLine kan dienen als software voor de ontsluiting.

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

Werkwijze en tijdsraming

- Inventarisatie en beschrijving van de tabellen in de lichtdrukken (1 maand)
- Berekening van tabellen vergelijkbaar met die in de lichtdrukken (1 maand)
- Vergelijking van gelichtdrukte met gegenereerde tabellen en inventarisatie van lacunes en afwijkingen (6 maanden)
- Onderzoeken van oorzaken van lacunes en afwijkingen (3 maanden)
- Vaststellen van mogelijkheden om lacunes aan te vullen (1 maand)
- Vaststellen van mogelijkheden om correcties in digitale bestanden aan te brengen (2 maanden)
- Tijdsraming van aanvullingen en verbeteringen (2 maanden)
- Maken van een opzet voor de digitale publicatie van de gegevens in StatLine (2 maanden)

Personele inzet

Uitgegaan wordt van een personele inzet van gemiddeld 0,6 FTE gedurende 18 maanden van een sociaal-economisch historicus met voldoende statistische ervaring. Begeleiding vindt plaats door dr. P.K. Doorn en drs. H.D. Tjalsma van het NIWI-KNAW en dr. J. van Maarseveen en drs. J. Jonker van het CBS. De begrote kosten bedragen Kfl. 95

4.3. Werkpakket 3: Toegankelijk maken en digitale archivering van de Volkstelling 1971

Werkpakketleider

CBS (dr. J.G.S.J van Maarseveen)

Doel

Toegankelijkstelling en digitale archivering van de Volkstelling 1971

Omschrijving werkpakket

De digitale bestanden van de laatste Volkstelling zijn gedurende het eerder uitgevoerde project opgespoord en geconverteerd naar moderne media. De telling is nu aanwezig in het bestandsopslagsysteem van het CBS, maar deze is niet toegankelijk voor onderzoekers. In verband met de privacy van de toenmalige respondenten dient de wijze van toegankelijkstelling zorgvuldig te worden afgewogen. Een op het eerste gezicht voor de hand liggende mogelijkheid zou zijn om de gegevens te aggregeren tot hetzelfde niveau als van de print-outs, die in de

bibliotheek van het CBS toegankelijk zijn. Mocht hiertoe besloten worden dan kan de digitale publicatie plaatsvinden met behulp van StatLine. Gezien de vele print-outs is dit echter een erg bewerkingintensieve, tijdrovende en derhalve kostbare aangelegenheid. Een ander belangrijk nadeel van deze publicatievorm is dat onderzoekers steeds weer nieuwe wensen kunnen formuleren, waaraan niet met deze digitale publicatie kan worden voldaan maar wel met behulp van de oorspronkelijke databestanden. Gezien het gebrek aan capaciteit bij het CBS voor dit soort extra werkzaamheden, mag verwacht worden dat aan dit soort extra wensen niet kan worden voldaan. Daarom gaat de voorkeur uit naar een andere werkwijze waarbij de onderzoeker direct gebruik kan maken van het basismateriaal zelf.

Bij deze tweede mogelijkheid gaat het op kortere termijn om vormen van 'on-site' werken, op wat langere termijn om 'rekenen op afstand' ('remote access' genoemd), met behulp van de oorspronkelijke databestanden, met waarborging uiteraard van de statistische beveiligingseisen. Deze innovatieve werkwijze komt zoveel mogelijk tegemoet aan de specifieke wensen van onderzoekers. Om dit zelfstandige gebruik van de databestanden door de onderzoeker mogelijk te maken dient het CBS een gedetailleerde beschrijving (meta-informatie) op te stellen van de opgeslagen databestanden. Behalve deze uitvoerige documentatie dient het CBS tevens, in overleg met externe onderzoekers, een proefbestand samen te stellen waarmee de onderzoekers zich vertrouwd kunnen maken met de opbouw en structuur van de databestanden. De digitale archivering kan plaatsvinden bij het CBS zelf of in de data-archieven van het NIWI-KNAW.

Deliverables:

- On-site access tot de Volkstelling van 1971 voor onderzoekers
- Volledig gedocumenteerde digitale Volkstelling van 1971 ter digitale archivering

Personele inzet:

Uitgegaan wordt van een personele inzet van 0,75 FTE op jaarbasis, bij een doorlooptijd van 9 à 12 maanden. Begrote kosten Kfl. 80

5. Projectorganisatie

De werkpakketleiders (Doorn, Engelen en van Maarseveen) voeren de dagelijkse leiding van de werkzaamheden die in het betreffende werkpakket worden uitgevoerd. De werkpakketleiders vormen een stuurgroep en dragen gezamenlijk de verantwoordelijkheid van de coördinatie tussen de werkpakketten.

Voor de begeleiding van het project wordt een begeleidingscommissie samengesteld bestaande uit vertegenwoordigers van de participerende instellingen. Deze commissie bestaat uit de volgende personen:

drs. J. Jonker (CBS)
dr. T. Vreugdenhil (CBS)
dr. J. van Maarseveen (CBS)
drs. E. Beekink (NIDI)
dr. P. Doorn (NIWI-KNAW)
dr. H. Knippenberg (UvA)
dr. Th. Engelen (KUN)
dr. F. Hendrickx (KUN)

De samenwerking tussen de projectpartners wordt contractueel vastgelegd. In het contract worden geregeld:

- Het beheer van de financiën. Het NIWI-KNAW treedt op als penvoerder ('coördinerend partner').

- Wanneer welke producten door welke partner worden geleverd.
- De verantwoordelijkheden voor de data-distributie.
- De verdeling van de materiële en immateriële opbrengsten (*credits*).
- Projectleiding en rol van de begeleidingscommissie.

6. Samenvattende begroting

Werkpakket 1:

Salariëring van één medewerker
(18 maanden, gem. 0,6 aanstelling, inschaling 10-8) Kfl. 95

Werkpakket 2:

Salariëring van één medewerker
(18 maanden, gem. 0,6 aanstelling, inschaling 9 of 10) Kfl. 95

Werkpakket 3:

Salariëring van één medewerker
(18 maanden, gem. 0,5 aanstelling, inschaling 9 of 10) Kfl. 80

Totaal Kfl. 270

De overige benodigde materiële kosten worden gedragen door de participerende instellingen.