



Centraal Bureau voor de Statistiek

Centrum voor Beleidsstatistiek

**Documentatierapport Verzekerde Personen en
Dienstverbanden Wet Administratieve
Lastenverlichting en Vereenvoudiging in Sociale
verzekeringswetten, voorlopige versie (VPD-
WALVISvlp) 2006-20080103V2**

VIDIV-nummer: 2008-231-MCB

Datum: 6 oktober 2008

Opgesteld door: Ronald Blokzijl

Bronvermelding

Publicatie van uitkomsten geschiedt door het onderzoeksbureau of de opdrachtgever op eigen titel. Verwijzing naar het CBS betreft uitsluitend het gebruik van de microdatabestanden. Dat wordt als volgt geformuleerd:

“Eigen berekening [*naam onderzoeksbureau, c.q. opdrachtgever*] op basis van bij het CBS beschikbaar gestelde microdatabestanden betreffende verzekerde personen en dienstverbanden volgens de Wet Administratieve Lastenverlichting en Vereenvoudiging in Sociale verzekeringswetten (WALVIS) 2006 tot en met maart 2008 (voorlopige versie)V2, welke afkomstig zijn van het Uitvoeringsinstituut Werknemers Verzekeringen (UWV).”

Versiegeschiedenis

Versie 2

081006 VPD-WALVISvlp 2006-20080103V2

080804 GBA-selectiebestand VPD-WALVISvlp 2006-20080103V2

- Voor versie 2 is het loonheffingnummer versleuteld.

Versie 1

080730 VPD-WALVISvlp 2006-20080103V1

080804 GBA-selectiebestand VPD-WALVISvlp 2006-20080103V1

De gebruiker dient rekening te houden met de volgende *do's & don'ts*:

- Dit microdatabestand is een voorlopige versie. De definitieve versie wordt in oktober verwacht en zal dan zijn aangevuld met een bedrijfsindicator (BEID). De inhoud van dit rapport zal bovendien worden aangevuld en uitgebreid..
 - Uitsluitend de combinatie van RIN met SRTNUM 'R' of 'S' identificeert een persoon. Zie blz. 10 voor een beschrijving.
-

Inhoudsopgave

Bronvermelding	2
Versiegeschiedenis.....	3
1. Inleiding	7
2. Inhoud van de bestanden.....	8
2.1 Structuur van de bestanden.....	8
2.2 Overzicht van variabelen	8
2.3 Algemene toelichting.....	9
3. Beschrijving van de variabelen	10

1. Inleiding

Bij het Centrum voor Beleidsstatistiek (CvB) worden microdatabestanden van het Centraal Bureau voor de Statistiek (CBS) beschikbaar gesteld voor gebruik door externe onderzoekers. Van alle personen die in het hier beschreven microdatabestand voorkomen, zijn gegevens uit de Gemeentelijke basisadministratie (GBA) geselecteerd. Deze selectie wordt het GBA-selectiebestand genoemd. Dit documentatierapport beschrijft de inhoud en structuur van het microdatabestand '081006 VPD-WALVISvlp 2006-20080103V2'. Het complete GBA-bestand wordt in een apart documentatierapport beschreven.

Hoofdstuk 2 beschrijft de structuur van het microdatabestand, geeft een opsomming van de variabelen en bevat een toelichting op het bestand met daarbij een indicatie van de kwaliteit. In hoofdstuk 3 wordt per variabele meer informatie gegeven. Voor de categoriale variabelen worden alle mogelijke scores en hun betekenis opgesomd.

Indien gewenst, kan een bijlage geleverd worden met daarin de frequentieverdelingen van de categoriale variabelen: '081006 Bijlage 1 Frequentietellingen VPD-WALVISvlp 2006-20080103V2'. In deze bijlage worden van de (semi-) continue variabelen de eerste en laatste vijf waarden gegeven.

2. Inhoud van de bestanden

2.1 Structuur van de bestanden

Microdatabestand

Een record in het microdatabestand bevat gegevens over een werknemer, te weten loonheffingnummer, begindatum dienstverband en einddatum dienstverband. Het bestand bevat RIN en SRTNUM als sleutel waarmee gekoppeld kan worden met andere bestanden.

GBA-selectiebestand

Het bijbehorende GBA-selectiebestand bevat persoons- en huishoudensgegevens van alle personen die in het microdatabestand voorkomen en ook geregistreerd staan in de GBA. De persoonsgegevens bestrijken een periode van 1 januari 1995 tot 1 januari 2006. De huishoudensgegevens worden door het CBS berekend voor de stand op 1 januari. In het GBA-selectiebestand zijn huishoudensgegevens beschikbaar van 1998 tot en met 2006.

Zowel het microdatabestand als het GBA-selectiebestand is gesorteerd op respectievelijk SRTNUM en RIN. Met behulp van deze twee variabelen als koppelsleutel kan de gebruiker persoons- en huishoudensgegevens uit het GBA-selectiebestand toevoegen aan het microdatabestand, zodat deze gegevens in het onderzoek meegenomen kunnen worden.

Om RIN-nummers toe te kennen aan het microdatabestand is gebruik gemaakt van de meest recente versie van de GBA. Hierin is slechts een beperkt aantal variabelen aanwezig, daarom wordt dit bestand niet gebruikt voor het GBA-selectiebestand. Het GBA-selectiebestand maakt gebruik van gegevens tot 1 januari 2006. Hierdoor kunnen personen die na 1 januari 2006 naar Nederland zijn geïmmigreerd, niet worden teruggevonden in het GBA-selectiebestand.

2.2 Overzicht van variabelen

Onderstaand volgt een lijst van alle variabelen in het microdatabestand. Zie hoofdstuk 3 voor een nadere toelichting op de variabelen in het microdatabestand.

<i>Nummer</i>	<i>Variabele & Label</i>	<i>Format</i>	<i>Lengte</i>	<i>Decimalen</i>
1	SRTNUM Soort identificatiecode	Tekst	1	0
2	RIN Identificatienummer	Getal	9	0
3	LOFI_VERSL Versleuteld loonheffingnummer	Tekst	12	0
4	BEGINDAT Begindatum dienstverband (jjjjmdd)	Getal	8	0
5	EINDDAT Einddatum dienstverband (jjjjmdd)	Getal	8	0

2.3 Algemene toelichting

Populatieafbakening

Voor deze voorlopige versie van het bestand is nog geen informatie samengesteld.

Kwaliteit

Het microdatabestand is geen publicatiebestand. Dit betekent dat voor veel variabelen door het CBS weinig of geen controles en correcties hebben plaatsgevonden. Op de variabelen die wel door het CBS zijn bekeken, zijn op grond van de beschikbare informatie correcties uitgevoerd. Deze correcties zijn als zodanig in het microdatabestand overgenomen.

Omdat het geen officieel vastgesteld publicatiebestand betreft, heeft dit tot gevolg dat de bestanden niet noodzakelijk exact overeenkomen met Statline, de elektronische databank van het CBS (statline.cbs.nl).

Een goede stelregel om in eerste instantie te kunnen inschatten of een variabele in het microdatabestand van goede kwaliteit is, is te kijken naar Statline. Als een kenmerk is opgenomen in Statline, is dat over het algemeen van goede kwaliteit. Als een kenmerk niet is opgenomen, dan is de kans groot dat deze variabele van matige of slechte kwaliteit is.

De gegevens op Statline kunt u vinden onder Arbeid en sociale zekerheid → Arbeidsmarkt → Banen en lonen.

3. Beschrijving van de variabelen

In dit hoofdstuk wordt per variabele meer informatie gegeven. Voor de categoriale variabelen worden alle mogelijke scores en hun betekenis opgesomd.

1 SRTNUM

Soort identificatiecode

F	Origineel foutief sofinummer/burgerservicenummer
R	RIN-persoon (persoon is gevonden in GBA)
S	Versleuteld sofinummer/burgerservicenummer

Bronbestanden die aan het CBS geleverd worden, bevatten doorgaans als persoonsidentificator het sofinummer. Het sofinummer wordt als zeer identificerend beschouwd en daarom vervangt het CBS dit nummer door een intern persoonsidentificatienummer, het betekenisloze RIN-nummer.

Een microdatabestand is voorzien van de door het CBS toegevoegde variabelen soort identificatiecode (SRTNUM) en identificatienummer (RIN), die samen het sofinummer/burgerservicenummer vervangen. Records met een sofinummer/burgerservicenummer dat is teruggevonden in de GBA, zijn voorzien van een geldig RIN-nummer. Deze records hebben de waarde 'R' op SRTNUM gekregen.

De overige records bevatten een sofinummer/burgerservicenummer dat niet is teruggevonden in de GBA. Afhankelijk van de situatie hebben zij de waarde 'F' of 'S' op SRTNUM gekregen. Indien het sofinummer/burgerservicenummer ongeldig is, dat wil zeggen niet voldoet aan de 11-toets, dan wordt het sofinummer/burgerservicenummer onveranderd overgenomen in het veld RIN en krijgt SRTNUM de waarde 'F'. Vooral de waarden '0' en '999999999' komen vaak voor als ongeldig sofinummer/burgerservicenummer. Indien het sofinummer/burgerservicenummer geldig is maar niet wordt teruggevonden in de GBA, dan wordt het sofinummer/burgerservicenummer versleuteld via een vaste formule en krijgt SRTNUM de waarde 'S'.

De combinatie van SRTNUM 'R' en RIN identificeert een uniek persoon. Evenzo kan worden aangenomen dat personen met SRTNUM 'S' een unieke waarde op RIN hebben gekregen. De meest voorkomende reden dat zo'n persoon wel een geldig sofinummer/burgerservicenummer heeft maar niet in de GBA staat ingeschreven, is namelijk dat hij of zij in het buitenland woont. Door immigratie kan een persoon het ene jaar niet en het volgende jaar wel in de GBA staan ingeschreven. Records met SRTNUM 'F' duiden op een ongeldig sofinummer/burgerservicenummer; met deze records zijn geen personen te identificeren. Records met SRTNUM 'F' zijn daarom niet geschikt voor het koppelen aan andere bestanden of aggregeren op persoonsniveau.

Als een microdatabestand meerdere records per persoon bevat, dan kan voor SRTNUM 'R' en SRTNUM 'S' worden aangenomen dat records met eenzelfde RIN-nummer dezelfde persoon aanduiden. Binnen één bestand is het geen probleem om records met SRTNUM 'R' of 'S' te aggregeren op persoonsniveau. Records met SRTNUM 'R' of 'S' kunnen ook worden gekoppeld aan records in een ander microdatabestand met dezelfde waarden op SRTNUM en RIN: deze

records hebben betrekking op dezelfde persoon. De gebruiker kan met behulp van SRTNUM en RIN ook gegevens uit het GBA-selectiebestand toevoegen aan het microdatabestand.

Er moet echter een kanttekening worden gemaakt bij het koppelen van verschillende microdatabestanden. Een persoon kan namelijk in het ene bestand de waarde 'R' op SRTNUM hebben en in het andere bestand de waarde 'S', en vice versa. Deze situatie kan zich voordoen als een persoon ten tijde van het ene bestand wél in de GBA stond ingeschreven en ten tijde van het andere bestand niet. Omdat de records in dat geval ook een ander RIN-nummer hebben gekregen, zijn zij niet aan elkaar te koppelen.

2 RIN

Identificatienummer

Persoonsidentificatienummer, ter vervanging van het sofinummer/
burgerservicenummer. Zie toelichting bij SRTNUM.

3 LOFI_VERSL

Versleuteld Loonheffingnummer

4 BEGINDAT

Begindatum dienstverband (jjjjmmdd)

5 EINDDAT

Einddatum dienstverband (jjjjmmdd)
