

Het datawarehouse opent zijn deuren

Op 16 september 2001 kwam de Gebruikersgroep rond het Datawarehouse Arbeidsmarktgegevens (DWH) voor een derde keer samen.¹ Enkele weken voordien werden de langverwachte tabellen van de eerste basistoepassing naar de geïnteresseerden toegestuurd.² De eerste oefeningen konden dan ook worden voorgesteld. Vergelijkende analyses met andere gegevensbronnen geven bemoedigende resultaten rond de meerwaarde die het Datawarehouse biedt aan de bestaande statistische bronnen.

De verdeling over arbeidsmarktposities

Om het geheugen een beetje op te frissen en eventuele nieuwkomers in de gebruikersgroep wegwijs te maken, beginnen we met een voorstelling van de ruggengraat van het Datawarehouse Arbeidsmarktgegevens: de nomenclatuur van arbeidsmarktposities. De nomenclatuur kent een opsplitsing tot op 5 posities, omwille van de leesbaarheid wordt tabel 1 beperkt tot 3 posities.

Op het *eerste niveau* wordt de populatie in het Datawarehouse opgesplitst naar een werkende, een werkzoekende en een niet-beroepsactieve groep. Verder is er ook een groep 'onbekend', waarin zich personen bevinden die wel in het Datawarehouse gekend zijn, maar niet voor wat betreft hun arbeidsmarktsituatie, zoals bijvoorbeeld huisvrouwen of kinderen.³ Tabel 1 illustreert de verdeling van de 18-64-jarigen over deze 4 groepen. Het *tweede niveau* binnen de nomenclatuur geeft meer informatie over het statuut van de werkenden (loontrekkend, zelfstandig, helper). Voor personen die een loontrekkende job met een zelfstandigenstatuut combineren werd een aparte categorie gecreëerd. Nog meer detail wordt verkregen op het *derde niveau*. Loontrekkenden worden ingedeeld naargelang ze 1 of meer-

dere jobs uitoefenen, zelfstandigen en helpers kunnen onder een statuut in hoofdberoep, in bijberoep of na pensioenleeftijd vallen. Ook de werkzoekenden worden ingedeeld op basis van het regime waarin ze een uitkering ontvangen: werkzoekend na voltijdse tewerkstelling, werkzoekend na studies, werkzoekend na een vrijwillig deeltijdse job.

Binnen de groep van de niet-beroepsactieven kent het Datawarehouse drie categorieën: voltijds bruggepensioneerden, voltijds loopbaanonderbrekers en vrijgestelde werkzoekenden.

Een eerste blik op de cijfers naast de nomenclatuurposities doet bij sommigen misschien de wenkbrauwen fronsen. De aantallen liggen inderdaad lager dan wat wordt bekomen op basis van de gewoonlijk gebruikte bronnen. De oorzaken hiervan werden al in eerdere artikels in OVER•WERK besproken.⁴ Samengevat: in het werkende segment slaagt het Datawarehouse er dankzij een vergaande integratie van de bestanden in om dubbeltellingen, veroorzaakt door personen met meerdere jobs, uit te sluiten. In het werkloze segment zijn enkel de uitkeringsgerechtigde werklozen gekend. Andere categorieën van de nwwz, zoals vrij ingeschrevenen of schoolverlaters in wachttijd, zijn totnogtoe niet in het Datawarehouse gekend.⁵

Vergelijkende analyses

DWH versus Steunpunt WAV-raming

Een eerste analyse vergelijkt de werkzaamheidsgraad (aandeel werkenden in de bevolking op ar-

beidsleeftijd) op basis van het Datawarehouse met de *raming die het Steunpunt WAV maakt op basis van administratieve bronnen* voor de berekening van haar subregionale arbeidsmarkt cijfers.⁶ De oefening betreft het tweede kwartaal 1998, de bevol-

king op arbeidsleeftijd wordt afgebakend als de bevolking tussen 18 en 64 jaar. Telkens wordt de werkzaamheidsgraad berekend voor België als geheel en voor de 3 gewesten afzonderlijk.

Tabel 1.

Overzicht van de nomenclatuurposities, 18-64 jaar (België, 2de kwartaal 1998).

n	%	Code + toelichting van de socio-economische positie
5905276	100,0	<i>Totaal</i>
3704142	62,7	1. Werkend
2990375	50,6	1.1. Werkend in loondienst
2873600	48,7	1.1.1. Werkend in één job in loondienst
116775	2,0	1.1.2. Werkend in meerdere jobs in loondienst
547846	9,3	1.2. Werkend als zelfstandige
513550	8,7	1.2.1. In hoofdberoep
20884	0,4	1.2.2. In bijberoep
13412	0,2	1.2.3. Werkend als zelfstandige na pensioenleeftijd
43141	0,7	1.3. Werkend als helper bij een werkgever met zelfstandigenstatuut
42295	0,7	1.3.1. Werkend als helper in hoofdberoep
354	0,0	1.3.2. Werkend als helper in bijberoep
492	0,0	1.3.3. Werkend als helper na pensioenleeftijd
122780	2,1	1.4. Werkend in loondienst en in zelfstandigenstatuut
111615	1,9	1.4.1. Voornaamste job wordt uitgeoefend in loondienst
11165	0,2	1.4.2. Voornaamste job wordt uitgeoefend in zelfstandigenstatuut
374613	6,3	2. Werkzoekend met tussenkomst van de RVA
250493	4,2	2.0.1. Werkzoekend na voltijdse tewerkstelling
101447	1,7	2.0.2. Werkzoekend na studies, gerechtigd op wachtuitkering
22554	0,4	2.0.3. Werkzoekend na een vrijwillig deeltijdse job
119	0,0	2.0.4. Werkzoekend na studies, gerechtigd op overbruggingsuitkering
0	0,0	2.0.5. Werkzoekend zonder uitkering (nieuw statuut)
276794	4,7	3. Niet beroepsactief (met tussenkomst van de RVA)
117438	2,0	3.0.1. Voltijds bruggepensioneerd
17163	0,3	3.0.2. Voltijdse loopbaanonderbreking
142193	2,4	3.0.3. Vrijstelling van inschrijving als werkzoekende
1549727	26,2	4. Onbekend
1191	0,0	4.0.1. Geschorste werkzoekende
1548536	26,2	4.0.2. Andere

Bron: KSZ-DWH Arbeidsmarktgegevens (Bewerking Steunpunt WAV).

Tabel 2 toont dat een werkzaamheidsgraad op basis van het Datawarehouse tussen 1,6 en 2,5 procentpunten lager ligt dan een werkzaamheidsgraad op basis van de Steunpunt WAV-raming. Een deel van dit verschil is terug te brengen tot aanwijsbare factoren. Deze oefening brengt de cijfermatige weerslag hiervan in kaart.

Tabel 2.

Vergelijking van de werkzaamheidsgraad (18-64 jaar), Datawarehouse versus Steunpunt WAV-raming (België, 2de kwartaal 1998).

	DWH (%)	Steunpunt WAV (%)	Verskil	
			Voor afstemming (ppn.)	Na afstemming (ppn.)
Vlaanderen	62,4	64,8	+2,5	+0,4
Wallonië	53,8	56,2	+2,4	-0,2
Brussel	49,3	50,9	+1,6	+0,7
België	58,4	60,8	+2,4	+0,2

Bron: Steunpunt WAV, KSZ-DWH Arbeidsmarktgegevens (Bewerking Steunpunt WAV).

Een eerste factor, die aanzienlijk doorweegt in het verschil tussen beide bronnen, betreft de zogenaamde *dubbeltellingen*. Personen die meerdere jobs uitoefenen, worden ook meerdere malen aangegeven aan de socialezekerheidsinstellingen van ons land. Op basis van de administratieve bronnen kunnen enkel de dubbeltellingen binnen het RSZ-LATG-bestand worden uitgezuiverd. Op die manier wordt bijvoorbeeld een zelfstandige in hoofdberoep, met daarnaast een kleine job in loondienst, als 2 personen geteld. Het Datawarehouse sluit alle dubbeltellingen uit, wat leidt tot een lagere werkzaamheidsgraad.⁷

Niettegenstaande het feit dat zowel de Steunpunt WAV-raming als het Datawarehouse op administratieve bestanden zijn gebaseerd, is de werkende populatie in beide instrumenten niet op volledig dezelfde wijze afgebakend. Zo houdt de Steunpunt WAV-raming in haar bepaling van het aantal werkzame personen rekening met de *uitgaande grensarbeid*, op basis van de bestanden van het RIZIV. Deze bestanden zijn echter niet opgenomen in de

Datawarehouse-koppeling, zodat uitgaande grensarbeid in het Datawarehouse ontbreekt. De *niet-vergoede helpers* worden niet aangegeven bij de RSVZ en zijn om die reden niet gekend op basis van administratieve bronnen. De Steunpunt WAV-raming lost dit op door deze (overwegend) meewerkende echtgenotes bij te ramen op basis van NIS-EAK; het Datawarehouse ontbeert echter deze informatie. Verder telt de Steunpunt WAV-raming de *zelfstandigen en helpers in bijberoep* niet mee in de veronderstelling dat deze personen al in de telling zijn opgenomen op basis van hun hoofdberoep. Er zijn echter ook zelfstandigen en helpers die weliswaar het statuut 'bijberoep' hebben, maar verder geen andere betrekking uitoefenen. Waar de Steunpunt WAV-raming deze personen ten onrechte 'wegcorrigeert', neemt het Datawarehouse hen gewoon mee op in de telling. Ook *PWA's* worden in beide bronnen op verschillende wijze geteld. De Steunpunt WAV-raming ziet hen als uitkeringsgerechtigde werklozen, het Datawarehouse klasseert deze groep als personen met een job in loondienst.

Wanneer rekening wordt gehouden met bovenstaande verschillen tussen de Steunpunt WAV-raming enerzijds en de Datawarehouse-populatie anderzijds, krimpt het verschil in werkzaamheidsgraad tot 0,2 à 0,7 procentpunt. Conclusie is dan ook dat beide instrumenten versterkt uit de oefening komen.

DWH versus NIS-EAK

Een tweede analyse is qua opzet volledig gelijklopend als de eerste, alleen wordt als vergelijkingsmateriaal naar de NIS-Enquête naar de Arbeidskrachten (EAK) gegrepen. De aanzet tot deze oefening werd gegeven op de studiedienst van het federaal Ministerie voor Tewerkstelling en Arbeid, waarna het Steunpunt WAV naging hoeveel van de afwijking kan verklaard worden door verschillen in definitie of populatie.

Tabel 3 vergelijkt de werkende populatie volgens de twee gegevensbronnen. In het tweede kwartaal 1998 zijn er volgens NIS-EAK 3 842 000 personen aan de slag in een betaalde job. Het Datawarehouse telt zo'n 120 000 werkende personen minder en komt op 3 720 000 personen.

Tabel 3.

Vergelijking van het aantal werkende personen (15-64 jaar), Datawarehouse versus NIS-EAK (België, 2de kwartaal 1998).

	DWH (n)	NIS-EAK (1998) (n)
werkende personen	3 720 000	3 842 000
uitgaande grensarbeiders	-	+
niet-vergoede helpers	-	+
voltijds loopbaanonderbrekers	-	+/-
werkstudenten	+	-

Bron: KSZ-DWH Arbeidsmarktgegevens, NIS-EAK (Bewerking Steunpunt WAV).

De tabel biedt ook informatie over de verschillen tussen beide gegevensbronnen. *Uitgaande grensarbeiders* bijvoorbeeld worden in EAK meegeteld bij de werkende populatie, terwijl het Datawarehouse geen informatie heeft over deze personen (81 196 personen in 1998). Er zijn ook de *niet-vergoede helpers*, die door EAK gekend zijn maar in het Datawarehouse ontbreken (75 637 personen in 1998). Een andere groep zijn de *voltijds loopbaanonderbrekers*, die in het Datawarehouse steeds tot de niet-beroepsactieve bevolking worden gerekend. In EAK werd in 1998 de vraag naar voltijdse loopbaanonderbreking nog niet expliciet gesteld en werd de respondent afhankelijk van hoe hij/zij zichzelf zag ingedeeld bij de werkenden of de bij niet-beroepsactieven. Het is mogelijk dat een aantal van deze loopbaanonderbrekers zichzelf nog steeds als loontrekkende beschreef. Ten slotte zijn er de *werkstudenten*, die er naast hun studies ook een baan(tje) op nahouden. In de mate dat deze studenten socialezekerheidsbijdragen betalen worden ze opgenomen in de werkende populatie van het Datawarehouse. De EAK daarentegen ziet deze personen als student en dus per definitie niet werkend.⁸

Illustratie: de arbeidsmarkt op wijkniveau

Ter illustratie van de nieuwe mogelijkheden die het Datawarehouse biedt, wordt ook een oefening voorgesteld op basis van arbeidsmarktgegevens op

het kleinste geografische niveau, de statistische sector. Via de geanonimiseerde koppeling met het rijksregister beschikt het Datawarehouse over de NIS-statistische sectorcode. Op die manier is een zeer diepgaande geografische uitsplitsing, tot op buurtniveau, van de gegevens mogelijk. Zo kunnen een aantal arbeidsmarktindicatoren, zoals bijvoorbeeld de werkzaamheidsgraad, op buurtniveau worden berekend. In het Jaarboek over de Arbeidsmarkt in Vlaanderen, te verschijnen medio december 2002, wordt deze analyse verder uitgediept. Hieronder vindt u kort de voornaamste bevindingen.

Aan de hand van een eenvoudige spreidingsmaat wordt nagegaan in welke mate werkenden en werkzoekenden gaan 'samenwonen' in dezelfde wijken. Is het zo dat bepaalde wijken een zeer hoge werkzaamheidsgraad hebben en andere net een laag aandeel werkenden? Vind je hoge concentraties werklozen of beroepsinactieven in de straten die eerder als 'achteruitgesteld' geboekstaafd staan? De analyse werd gedaan voor de 13 Vlaamse centrumsteden en het Brussels Hoofdstedelijk Gewest, en beslaat het tweede kwartaal van 1999. Zowel op het vlak van werkzaamheidsgraad als op het vlak van werkloosheidsgraad blijken de buurten binnen het Brussels Gewest verder uiteen te liggen dan in de beschouwde Vlaamse steden. Ook in Antwerpen en in Genk is het verschil tussen de werkzaamheidsgraden en de werkloosheidsgraden van de verschillende buurten groter dan in de andere steden.

In een tweede luik werd de arbeidsdeelname in de buurten op kaartjes gevisualiseerd. Grofweg zien we vooral in de grotere steden een concentrische beweging: de buurten met de laagste werkzaamheidsgraad en de hoogste werkloosheidsgraad zijn vooral terug te vinden in de stadskern, de buurten waar de arbeidsdeelname gunstiger is liggen eerder in de stadtrand.

Concreet: wat is er intussen beschikbaar?

De meest recente beschikbare gegevens die al in een basistoepassing werden gegoten betreffen het derde kwartaal van 1999. Van sommige deelnemende instellingen zijn ook al latere kwartalen in

het Datawarehouse ingevoerd. Naast de beschikbare basistoepassingen is er voor onderzoekers met een specifieke onderzoeksvraag ook de mogelijkheid om een gerichte gegevensaanvraag aan de KSZ te richten. Hiervoor moet de creatie van de basistoepassingen niet afgewacht worden. Onderstaand schema geeft per deelnemende instelling weer hoe ver het proces van gegevensinvoer in het Datawarehouse nu staat. De eerste 8 kolommen geven per instelling de kwartalen aan waarvoor de administratieve bestanden in het Datawarehouse zijn ingevoerd, de kolom 'socio-economische posi-

tie' geeft weer voor welke kwartalen de populatie is ingedeeld volgens de nomenclatuur van het Datawarehouse en de laatste kolom geeft aan voor welke kwartalen de zogenaamde 'afgeleide variabelen', variabelen op basis van de variabelen van de deelnemende instellingen, reeds zijn gecreëerd.

*Caroline Vermandere
Steunpunt WAV*

Tabel 4.

Overzicht ingeladen data in het Datawarehouse.

		RSZ	PPO	RVA	RSVZ	RSVZ – kinderbijzsl.	RKW	RIZIV	Rijks-/ Bisregister	Soc.-Eco. Pos.	Afg. variabelen
1997	1ste trimester	x	x		x		x	x	x		x
	2de trimester	x	x		x		x	x			x
	3de trimester	x	x		x		x	x			x
	4de trimester	x	x		x		x	x			x
1998	1ste trimester	x	x	x	x		x	x	x		x
	2de trimester	x	x	x	x		x	x		x	x
	3de trimester	x	x	x	x		x	x		x	x
	4de trimester	x	x	x	x		x	x		x	x
1999	1ste trimester	x	x	x	x		x	x	x	x	x
	2de trimester	x	x	x	x		x	x		x	x
	3de trimester	x	x	x	x		x	x		x	x
	4de trimester	x	x	x	x		x	x		x	x
2000	1ste trimester	x	x	x			x	x			
	2de trimester	x	x	x			x	x			
	3de trimester	x	x	x			x	x			
	4de trimester	x	x	x			x	x			

Gearceerd: niet beschikbaar

Bron: Kruispuntbank Sociale Zekerheid.

Noten

1. De Gebruikersgroep DWH Arbeidsmarktgegevens is een gemeenschappelijk initiatief van het Steunpunt WAV, het Point d'Appui TEF en de Kruispuntbank Sociale Zekerheid.
2. Het betreft hier een verdeling van de totale populatie gekend in het Datawarehouse naar geslacht, leeftijd en arbeidsmarktpositie. Deze verdeling gaat tot op het geografisch niveau 'statistische sector'. Momenteel is ze beschikbaar voor de laatste drie kwartalen van 1998 en de eerste drie kwartalen van 1999. Geïnteresseerden kunnen contact opnemen met Dhr. Chris Brijs (KSZ), chris.brijs@ksz-bcss.fgov.be
3. Via de geanonimiseerde koppeling met het Rijksregister kent het Datawarehouse voor alle personen die gekend zijn in één van de zes deelnemende socialezekerheidsinstellingen het gezinshoofd. Voor de gezinshoofden zijn tevens de gezinsleden gekend. Indien deze personen echter niet gekend zijn bij één van de zes deelnemende socialezekerheidsinstellingen, is van hen in het Datawarehouse geen informatie over de arbeidsmarktpositie gekend.
4. Stevens E. en Vermandere C., Een volgende aflevering in het datawarehouse-feuilleton... In: *OVER•WERK. Tijdschrift van het Steunpunt WAV*, nr. 1/2, jg. 12, pp. 78-86 en van der Hallen P. en Van Mechelen N., Datawarehouse: een nieuwe bron voor arbeidsmarktonderzoek? In: *OVER•WERK. Tijdschrift van het Steunpunt WAV*, nr. 1/2, jg. 11, pp. 78-80.
5. Hiervoor zouden de databanken van de regionale arbeidsbemiddelingsdiensten (VDAB, BGDA, Forem) opgenomen moeten worden in de datawarehousestructuur.
6. De methodologie voor deze raming werd uitgewerkt in samenspraak met de Studiedienst van het Federaal Ministerie voor Tewerkstelling en Arbeid en met de Dienst Studies en Statistiek van het Ministerie van het Waals Gewest.
7. Het RSZ-LATG-bestand wordt ook uitgezuiverd op dubbelstellingen. Personen die met meerdere jobs zijn aangegeven bij de RSZ worden dus ook slechts 1 maal geteld.
8. Voor een goed begrip, het gaat hier om EAK anno 1998. In 1999 werd de vragenlijst grondig herwerkt, met als gevolg dat sindsdien zowel voltijds loopbaanonderbrekers als werkstudenten via EAK gekend zijn.

De Powerpoint-presentaties met betrekking tot deze oefeningen zijn beschikbaar via www.steunpunt-wav.be, rubriek gebruikersgroepen.

Wie geïnteresseerd is in deelname aan de Gebruikersgroep rond het Datawarehouse Arbeidsmarktgegevens kan contact opnemen met het Steunpunt WAV via caroline.vermandere@wav.kuleuven.ac.be. De volgende samenkomst van de gebruikersgroep is gepland op 9 december 2002 om 14u in de lokalen van de KSZ. Op de agenda staat onder meer de bespreking van een aantal nieuwe basistoepassingen die eerstdaags beschikbaar zullen zijn.