
Datawarehouse: een nieuwe bron voor arbeidsmarktonderzoek?

Het is inmiddels al meer dan een jaar geleden dat we in dit tijdschrift – toen nog Nieuwsbrief – de komst van een datawarehouse met arbeidsmarktgegevens aankondigden.¹ Sinds enkele weken is de permanente koppeling van een aantal belangrijke administratieve databestanden met socio-economische gegevens een feit. Tijd om weer enkele artikelen aan dit Agora-project te besteden. Deze bijdrage geeft een idee van de nieuwe analysemogelijkheden voor het arbeidsmarktonderzoek, een voorstelling van de inhoud van het datawarehouse, alsook van de planning van de verdere werkzaamheden. Inzicht in enkele concrete cijfergegevens is, zo hopen we, voor een volgend OVER-WERK.

Ter herinnering: de doelstellingen

Het datawarehouse Arbeidsmarkt werd opgestart als een Agora-project van de federale Dienst voor Wetenschappelijke, Technische en Culturele aangelegenheden. De Kruispuntbank Sociale Zekerheid (KSZ) en de deelnemende socialezekerheidsinstellingen² wensten een permanente koppeling van arbeidsmarktgegevens om flexibeler, sneller en goedkoper te kunnen inspelen op vragen van onderzoekers of beleidsondersteunende organisaties. Vele vragen vereisen immers een combinatie van verschillende gegevens(bestanden).

Bovendien zou zo'n koppeling de bestaande socio-economische statistieken aanzienlijk kunnen verbeteren. Dubbeltellingen tussen verschillende administratieve bestanden worden immers uitgeschakeld zodat een exactere bepaling van de omvang van de beroepsbevolking mogelijk wordt. Anderzijds biedt de koppeling van individuele gegevens

doorheen de tijd een zicht op het dynamisch functioneren van de arbeidsmarkt: hoe verlopen de in- en uittredeprocessen op de arbeidsmarkt, en wat is hierin de rol van specifieke werkgelegenheidsmaatregelen?

De informatica-technische kant van het Datawarehouse werd door de Maatschappij voor Meconografie (MvM) waargenomen, de wetenschappelijke ondersteuning van het project werd verzorgd door de Steunpunten WAV en TEF. Zij formuleerden een eerste reeks van mogelijke statistiektoepassingen

(zie verder), gebaseerd op behoeften die leven binnen de onderzoekswereld.

Te koop in het warehouse

Het project is begin 1999 van start gegaan met een inventaris van alle socio-economische gegevens in de databestanden van de deelnemende socialezekerheidsinstellingen. Voor 1997 en 1998 zijn de geselecteerde variabelen intussen per kwartaal in een datawarehouse verzameld.³ Voor meer dan 8 miljoen personen bevat het datawarehouse geanonimiseerde informatie over de *persoonskenmerken* (geslacht, leeftijd, nationaliteit, woonplaats), de *gezinsamenstelling*, en desgevallend over de geleverde *arbeidsprestaties* (werkgever, arbeidsregime, arbeidsvolume, loon, eventuele werkgelegenheidsmaatregel, enz.), het *RVA-statuut* (uitkeringsgerechtigde volledig werkloze, tijdelijk werkloze, bruggepensioneerde, loopbaanonderbreker, enz.),

de *kinderbijslagsituatie* en eventuele inactiviteit wegens *invaliditeit*.

Deze geanonimiseerde gegevens moeten toelaten om de *socio-economische positie* voor iedere persoon te bepalen. Er wordt een onderscheid gemaakt tussen de werkenden, de werkzoekenden en de inactieven, maar vooral de diversiteit binnen deze categorieën staat centraal. Met name de personen die in meerdere administratieve bestanden voorkomen – en daarom bij vroegere tellingen van de beroepsbevolking twee keer geteld werden – worden in aparte categorieën ondergebracht. Hierdoor kunnen op een eenvoudige manier statistieken opgesteld worden over bijvoorbeeld de zelfstandigen die ook een job in loondienst uitoefenen, over de loontrekkenden die verschillende jobs in loondienst combineren, over de werkenden waarvoor de werkloosheidsverzekering tussenbeide komt (PWA, doorstromingsprogramma's, ...), enz.

Nieuwe socio-economische statistieken

Op lange termijn is het immers de bedoeling om trimesterieel op basis van het datawarehouse een aantal *statistieken* aan te maken. Elk kwartaal zal wellicht een telling gebeuren van het aantal personen naar geslacht, leeftijdsklasse en socio-economische positie. Via het gebruik van de zogenaamde statistische sectoren kan dit zelfs tot op buurtniveau. Enkele specifieke groepen – de hoger vermelde posities, maar ook de deeltijds werkenden, invaliden, enz. – kunnen het onderwerp uitmaken van meer gedetailleerde statistieken, met bijvoorbeeld een uitsplitsing naar activiteitssector, arbeidsregime of werkgelegenheidsmaatregel.

Bovendien zouden statistieken over de *mobiliteit* op de arbeidsmarkt kunnen bijgehouden worden. Dergelijke statistieken kunnen een antwoord bieden op vragen zoals: Hoeveel werknemers in het onderwijs (of een andere sector) zijn tussen 31 maart en 30 juni 1998 van werkgever veranderd? Ligt dit aantal hoger dan een jaar eerder? In welke sectoren komen deze mensen terecht? Hoeveel personen die in het eerste kwartaal van 1998 geschorst waren omwille van langdurige werkloosheid (cf. artikel 80 van de werkloosheidsreglemen-

tering) hebben voor het einde van datzelfde jaar werk gevonden? Hoe gaat het met de jongeren die in het eerste kwartaal van 1997 actief waren als zelfstandige: hoeveel hebben tegen het einde van 1998 hun zaak moeten stopzetten? Geraken deze jongeren snel weer aan de slag? Opnieuw als zelfstandige, of in loondienst? Enzovoort.

Voor wie meer wil

Als het Toezichtscomité van de KSZ akkoord gaat met de eerste statistiektoepassingen die door WAV en TEF zijn voorgesteld, zullen deze naar alle verwachtingen tegen het einde van mei 2001 op relatief eenvoudige aanvraag via de KSZ beschikbaar kunnen gemaakt worden. De andere gegevens in het datawarehouse kunnen uiteraard ook opgevraagd worden, via de geijkte procedure voor dataaanvragen bij de KSZ. Deze procedure zal normaliter minder tijdrovend en duur zijn dan in het verleden, gezien de koppeling van de verschillende administratieve bestanden niet meer ad hoc dient te gebeuren. Een gedocumenteerde lijst van de variabelen in het datawarehouse kan bekomen worden bij de KSZ of de Steunpunten WAV en TEF.

Geplande werken

Het datawarehouse biedt nu reeds een schat aan informatie, maar om van blijvende waarde te zijn voor het arbeidsmarktonderzoek zijn *actualisering* en verdere operationalisering onontbeerlijk. Eén van de eerstkomende werken is daarom een aanvulling van het datawarehouse met de kwartaalgegevens voor 1999, en later dit jaar ook met de gegevens voor 2000.

Een tweede taak is een grondige studie en *evaluatie van de eerste statistieken* die het datawarehouse zal opleveren. Idealiter zou dit moeten kunnen gebeuren in het kader van een Gebruikersgroep Datawarehouse Arbeidsmarkt, naar het model van de gebruikersgroepen die bestaan rond de Enquête Arbeidskrachten van het NIS en de Loon- en Arbeidstijd-Gegevens van de RSZ. Dergelijke gebruikersgroepen staan in voor de beoordeling van de volledigheid en bruikbaarheid van de gegevens en de ontwikkeling van nieuwe concepten en statistieken. Op deze manier spelen gebruikersgroepen

een cruciale rol in de *verbetering van de wetenschappelijke en beleidsmatige exploitatie* van de beschikbare gegevens.

De huidige koppeling van de socio-economische gegevens van een aantal sociale zekerheidsinstellingen is ongetwijfeld een grote stap vooruit in het arbeidsmarktonderzoek. Maar er resten uiteraard nog heel wat onderzoeksvragen waar het datawarehouse momenteel geen antwoord op kan bieden, tenzij door de *koppeling met andere gegevensbronnen*. Door de gewestelijke arbeidsbemiddelingsdiensten bij het datawarehouse te betrekken zou een vollediger beeld ontstaan van het werkzoekende segment van de beroepsbevolking. Tegelijkertijd kunnen een aantal transitieën tussen werken en werk zoeken beter in kaart gebracht worden. Ook de integratie van de gegevens van de OCMW's en het pensioenkadaster zou de mobiliteitsstatistieken aanzienlijk verbeteren, evenals de integratie van het onderwijsniveau en het beroep van de individuen op basis van de algemene socio-economische enquête 2001 (volkstelling). Om ook voor de andere jaartallen steekproefgewijs over gegevens aangaande het onderwijsniveau en het beroep te beschikken, zou beroep kunnen gedaan worden op een koppeling van de jaarlijkse Enquête naar de Arbeidskrachten van het NIS (EAK)⁴ met het datawarehouse. Een koppeling van de EAK met het datawarehouse zou kunnen bijdragen tot een vereenvoudiging van het enquêteringsopzet gezien een aantal gegevens niet meer bevraagd zou moeten worden, maar uit de administratieve bestanden geput kunnen worden.⁵ Tevens laat zo'n koppeling

toe om de bias die bestaat tussen enquêtegegevens en administratieve gegevens correct in te schatten, wat leidt tot een betere (beleids)interpretatie van de gegevens.

*Peter van der Hallen
Natascha Van Mechelen
Steunpunt WAV*

Noten

1. N. Van Mechelen & L. Van Wichelen, Een datawarehouse over de arbeidsmarkt in wording. In: Nieuwsbrief van het Steunpunt WAV, december 1999, p. 54-57.
2. De volgende sociale zekerheidsinstellingen werkten mee aan de uitbouw van het datawarehouse 'Arbeidsmarkt': de Rijksdienst voor Sociale Zekerheid (RSZ), de Rijksdienst voor Sociale Zekerheid voor Plaatselijke en Provinciale Overheden (RSZPPO), de Rijksdienst voor de Sociale Verzekering van Zelfstandigen (RSVZ), de Rijksdienst voor Arbeidsvoorziening (RVA), het Rijksinstituut voor Ziekte en Invaliditeitsverzekering (RIZIV) en de Rijksdienst voor Kinderbijslag van Werknemers (RKW).
3. Voor de RVA-gegevens waren enkel de gegevens vanaf 1998 beschikbaar.
4. Eerstkomend start een DWTC project in het kader van het programma Sociale Cohesie die de wetenschappelijke meerwaarde van zulke koppeling zal onderzoeken.
5. Enkele EAK gegevens worden momenteel aan een aantal administratieve databanken, waaronder deze van het datawarehouse, gekoppeld om na te gaan of het enquêteringsopzet kan vereenvoudigd worden.