

# Administratieve bestanden of enquêtes?

De sterktes en zwaktes op een rijtje

*Het STILE-project, een Europese project dat wordt gecoördineerd door het HIVA, zit in haar laatste fase.<sup>1</sup> De belangrijkste doelstelling is het verbeteren van arbeidsmarktindicatoren voor de informatiesamenleving, de eEconomy. STILE staat dan ook voor Statistics and Indicators on the Labour Market in the eEconomy. De opzet is niet zozeer het bedenken van nieuwe statistieken en indicatoren, dan wel het verbeteren van bestaande instrumenten voor het monitoren van de arbeidsmarkt. Het Steunpunt WAV is betrokken bij Work Package 6 binnen het STILE-project, Mobility in the eEconomy.<sup>2</sup> Het doel van dit Work Package is tweeledig. Enerzijds wil het via het analyseren van data van de Community Labour Force Survey – CLFS (waarvan de Enquête naar de Arbeidskrachten de Belgische variant is) en van het Belgische Datawarehouse Arbeidsmarkt, een zicht krijgen op de determinanten van mobiliteit (instroom, uitstroom en jobmobiliteit) van werknemers in de ICT-sector. Anderzijds wil het via deze analyses de sterktes en zwaktes van beide bronnen voor dergelijk gedetailleerd onderzoek blootleggen. In dit artikel ligt de focus op het tweede aspect.*

## De Community Labour Force Survey

Tabel 1 toont een SWOT-analyse (Strengths, Weaknesses, Opportunities and Threats-analyse) met betrekking tot de CLFS, die de voor- en nadelen van de enquête aan het licht brengt. Een survey zoals de CLFS heeft vele voordelen. In eerste instantie wordt de enquête gecoördineerd door Eurostat zodat het mogelijk is om internationaal vergelijkbare data te analyseren. Alle Europese landen zijn immers verplicht om de door Eurostat gevraagde gegevens te leveren. Hier moeten we echter onmid-

dellijk een nuance aan toevoegen. Eurostat geeft namelijk enkel richtlijnen voor het opstellen van de vragenlijst. In praktijk leidt dat ertoe dat de vragen niet in alle landen op dezelfde manier worden geformuleerd, hetgeen kan leiden tot verschillende resultaten.

Een ander belangrijk voordeel van de CLFS is dat deze enquête reeds een lange geschiedenis kent en bijgevolg is het in vele landen een van de omvangrijkste en belangrijkste bronnen voor arbeidsmarktonderzoek. Daardoor is er veel ervaring bij gebruikers van deze data. Bovendien kunnen door de jaarlijkse bevraging lange, relatief coherente, tijdreeksen worden opgesteld. Een enquête laat ook toe om meer kwalitatieve informatie te verkrijgen, bijvoorbeeld over het doel van een opleiding of de reden waarom iemand zijn job heeft verlaten. Dergelijke kwalitatieve informatie is onontbeerlijk voor verklarend arbeidsmarktonderzoek en is niet aanwezig in administratieve bestanden zoals het Datawarehouse. Daarnaast is de bevraging gericht op iedereen, ongeacht hun arbeidsmarktsituatie. Zowel kinderen, werkenden als gepensioneerden behoren tot de populatie. En tot slot is de informatie relatief up-to-date. Na ongeveer een half jaar zijn de belangrijkste data beschikbaar voor het publiek.

Het belangrijkste nadeel is dat een enquête gebaseerd is op een steekproef zodat alle resultaten binnen een zeker betrouwbaarheidsinterval moeten

worden geïnterpreteerd. Dat is erg nadelig wanneer er een gedetailleerde opdeling en kruising van variabelen is gewenst, zoals bijvoorbeeld de NACE-code op twee of meer digits. Het overschakelen in alle landen naar een continue bevraging in plaats van enkel een bevraging in het tweede kwartaal, en dus de overschakeling naar jaargemiddelden, zal er in de toekomst wel toe leiden dat de betrouwbaarheid van de data zal toenemen. Indien we een fenomeen als jobmobiliteit in sectoren willen onderzoeken, zullen we evenwel blijven stuiten op de grenzen van de betrouwbaarheid van de data.

In tabel 2 wordt de jobmobiliteitsgraad op jaarbasis weergegeven voor de Europese landen (het aandeel loontrekkenden dat op een jaar tijd van job verandert). Voor een aantal landen is deze indicator niet beschikbaar, omdat de retrospectieve vraag waarmee mobiliteit wordt gemeten (zie verder) daar niet wordt opgenomen in de CLFS. Voor de andere landen is het wel mogelijk deze jobmobiliteitsgraad te berekenen, bijvoorbeeld voor mannen en vrouwen. Maar wanneer we deze indicator berekenen voor de ICT-sector dan zien we dat heel wat landen met onbetrouwbare data worden geconfronteerd.

Daarnaast kunnen in enquêtes ook steekproeffouten of andere fouten optreden, zoals het verkeerd begrip

pen van een vraag of een antwoord en het verkeerd coderen. Een ander belangrijk probleem is het gebruik van proxy-respondenten. Dit zijn leden van hetzelfde huishouden die antwoorden in plaats van een afwezig lid, een praktijk die wordt toegelaten om de non-respons te verlagen. In België wordt in ongeveer drie op tien van de gevallen geantwoord door een proxy-respondent. Er wordt verondersteld dat proxy-respondenten goed op de hoogte zijn van de meest relevante eigenschappen en activiteiten van de eigenlijke (afwezige) respondent, zoals bijvoorbeeld de arbeidsmarktsituatie (werkend, werkloos of niet-beroepsactief). De informatie over meer specifieke onderwerpen zoals het aantal gewerkte uren of de sector van tewerkstelling is echter minder accuraat (Statistics Canada, 2003).

De CLFS heeft verder een aantal nadelen die specifiek betrekking hebben op het meten van de ICT-sector en het meten van mobiliteit, en dus a fortiori op het meten van mobiliteit in de ICT-sector. Om de ICT-sector af te bakenen werd in STILE Work Package 6 vertrokken van de definitie die de Organisatie voor Economische Samenwerking en Ontwikkeling (OESO) vooropstelt (OECD, 2000). Om deze afbakening te kunnen maken, zijn echter data nodig tot op drie en soms vier digits van de NACE-codering. Het probleem is echter dat de CLFS enkel

**Tabel 1.**

SWOT-analyse van de CLFS met betrekking tot het meten van mobiliteit.

Sterktes	Zwaktes
<ul style="list-style-type: none"> <li>- Internationaal vergelijkbare data</li> <li>- Ondersteund door wetgeving (richtlijnen voor vragen)</li> <li>- Geschiedenis van organisatie/ervaring</li> <li>- De grootste databron in veel landen</li> <li>- Kwalitatieve informatie beschikbaar</li> <li>- Gericht tot iedereen</li> <li>- Relatief up-to-date</li> </ul>	<ul style="list-style-type: none"> <li>- Steekproef → Betrouwbaarheidsinterval → beperkte betrouwbaarheid voor kleine populaties</li> <li>- Steekproeffouten</li> <li>- Andere fouten: verkeerd begrijpen van vraag/antwoord; miscoderen</li> <li>- Proxy-respondenten</li> <li>- Geheugeneffecten en non-respons</li> <li>- NACE-code niet gedetailleerd genoeg om exacte definitie van (ICT-)sector af te bakenen</li> <li>- NACE-code gebaseerd op beschrijving van de respondent</li> </ul>
<ul style="list-style-type: none"> <li>- LFS wordt steeds meer gebruikt voor allerhande onderzoek: het is de grootste databron in veel landen</li> <li>- Nadruk op kwalitatieve informatie</li> <li>- Overschakeling naar continue bevraging → betrouwbaarheid van de data stijgt</li> </ul>	<ul style="list-style-type: none"> <li>- Steeds meer informatie wordt geregistreerd in administratieve bestanden</li> </ul>
Kansen	Bedreigingen

Bron: Camire, Steunpunt WAV.

gegevens heeft tot op het 2-digit-niveau en voor een minderheid van landen tot op drie digits. Daardoor moesten er beslissingen worden genomen over de inclusie en exclusie van bepaalde sectoren (afgebakend op twee digits) op basis van het voorkomen van ICT in een van de subsectoren. Algemeen is het probleem van de CLFS dus een te beperkt detail van de NACE-codering waardoor het vaak moeilijk is om zuivere afbakeningen te maken van sectoren. Hetzelfde geldt voor gegevens over beroepen op basis van de ISCO-code.

De meest gebruikte manier om mobiliteit te meten via de CLFS is het gebruik van antwoorden op re-

trospectieve vragen. Zo wordt bij de respondenten gepeild naar de arbeidsmarktsituatie één jaar voor de bevraging. Deze vragen zijn zo verwoord dat het resultaat makkelijk vergelijkbaar is met het moment van bevraging. Een eerste probleem bij deze retrospectieve bevraging is dat ze niet in alle landen is opgenomen in de enquête. Dat betekent dat slechts voor een beperkt aantal landen vergelijkbare data beschikbaar zijn over mobiliteit. Een tweede probleem van retrospectieve bevraging is het bestaan van geheugen- of herinneringseffecten die zowel een random als een systematisch karakter kunnen hebben. Verschillende factoren hebben immers een invloed op herinneringsfouten: onder

**Tabel 2.**

Jobmobiliteitsgraad op jaarbasis (EU-25; tweede kwartaal 2002).

(%)	ICT sector		Totale economie	
	Vrouw	Man	Vrouw	Man
België	8,8u	8,6	6,6	7,0
Tsjechië	5,9	8,1	6,4	6,2
Denemarken	:u	12,1	10,4	12,4
Duitsland	:	:	:	:
Estland	:u	:u	7,5	11,1
Griekenland	:u	:u	5,0	4,9
Spanje	5,2	7,5	9,3	8,9
Frankrijk	6,4	8,5	7,4	8,2
Ierland	:	:	:	:
Italië	:	:	:	:
Cyprus	:u	:u	7,5	9,4
Letland	:u	:u	8,9	10,8
Litouwen	:u	:u	7,0	12,0
Luxemburg	:	:	:	:
Hongarije	6,7	3,4u	4,6	5,6
Malta	:	:	:	:
Nederland	:	:	:	:
Oostenrijk	:	:	3,6	5,3
Polen	:	:	5,6	7,8
Portugal	:u	:u	4,9	4,8
Slovenië	:u	:u	2,5	3,9
Slowakije	8,7u	7,4	10,3	9,8
Finland	:u	:u	4,7	4,8
Zweden	10,2	11,2	11,4	11,5
Ver. Koninkrijk	:u	:u	13,0	13,1

Uitzondering op het referentiejaar 2002: Duitsland en Luxemburg = 2001  
 : = niet beschikbaar; :u = niet betrouwbaar; u = data met omzichtigheid gebruiken.

**Bron:** EU LFS, data tweede kwartaal (Bewerking Camire).

andere de lengte van de retrospectieve periode, de sociale wenselijkheid en de aard van de te herinneren gebeurtenis. Zo blijkt bijvoorbeeld dat periodes van werkloosheid minder accuraat worden gerapporteerd dan periodes van werk en ook korte periodes (van om het even welke aard) worden minder correct herinnerd en gerapporteerd (Paull, 2002). Dit probleem van de retrospectieve bevraging leidt volgens onderzoek van het Centraal Bureau voor de Statistiek in Nederland voornamelijk tot een onderschatting van mobiliteit (Allaart & van Ours, 2001). Bovendien wordt dit geheugeneffect vermoedelijk nog versterkt bij het gebruik van proxy-respondenten.

## Het Datawarehouse

Een aantal van de nadelen van de CLFS kunnen opgevangen worden door het gebruik van administratieve databanken zoals het Belgische Datawarehouse Arbeidsmarkt.<sup>3</sup> Tabel 3 laat de SWOT-analyse zien die werd uitgevoerd met betrekking tot het Datawarehouse.

Steeds meer informatie wordt geregistreerd en ter beschikking gesteld in databanken. Door het exhaustieve karakter is het mogelijk om analyses te doen op populatiebestanden zodat de typische fouten die voortvloeien uit het gebruik van een steekproef worden vermeden en het niet nodig is om de resultaten te interpreteren binnen een betrouwbaarheidsinterval. Dat maakt dergelijke bronnen bijzonder geschikt voor onderzoek van kleine populaties en voor onderzoek waar een verregaand detail van bepaalde variabelen vereist is. Tabel 4 illustreert de mogelijkheid om op basis van het Datawarehouse specifieke indicatoren te berekenen voor kleine populaties, bijvoorbeeld NACE-sectoren op drie digits.

Ook het probleem van vertekening door geheugeneffecten bij het meten van mobiliteit is niet meer aan de orde, evenmin als de effecten door het gebruik van proxy-respondenten. Bovendien zijn de mogelijkheden voor mobiliteitsonderzoek niet beperkt tot collectieve mobiliteit, maar kunnen we ook loopbaananalyses uitvoeren gezien in het Datawarehouse gegevens beschikbaar zijn over de

**Tabel 3.**

SWOT-analyse van het Datawarehouse met betrekking tot het meten van mobiliteit.

Sterktes	Zwaktes
<ul style="list-style-type: none"> <li>- Exhaustief → gedetailleerde opsplitsing van variabelen; geen steekproeffouten; geen proxy-respondenten; geen geheugeneffecten</li> <li>- Mogelijkheid om combinatie van activiteiten te meten</li> <li>- Kwartaalgegevens op individueel niveau: loopbaananalyses mogelijk</li> <li>- Relatief up-to-date</li> </ul>	<ul style="list-style-type: none"> <li>- Niet internationaal vergelijkbaar</li> <li>- Dekking rond 90% van de inwoners</li> <li>- Geen kwalitatieve data – data worden verzameld vanuit administratief doel, niet voor onderzoek</li> <li>- Vertragingen/fouten in registratie beïnvloeden de kwaliteit van de data</li> <li>- Jobmobiliteit = verandering van onderneming (geen informatie op niveau van de lokale vestiging)</li> <li>- Administratieve wijzigingen van het identificatienummer van ondernemingen (door overname, afsplitsing, ...) leiden tot meten van 'schijnmobiliteit'</li> <li>- Geen informatie over onderwijsniveau</li> </ul>
<ul style="list-style-type: none"> <li>- Toenemend belang van administratieve data</li> <li>- 'Schijnmobiliteit' zal worden opgelost door het gebruik van een uniek identificatienummer voor lokale vestigingen</li> <li>- Gebruik van dit uniek identificatienummer zal ook data beschikbaar maken op het niveau van de lokale vestiging</li> <li>- Dekking zal toenemen door meer bestanden te integreren in het Datawarehouse</li> </ul>	<ul style="list-style-type: none"> <li>- Verandering in publieke opinie/wetgeving over privacy kan het gebruik van administratieve bestanden beïnvloeden</li> </ul>
Kansen	Bedreigingen

**Bron:** Camire, Steunpunt WAV.

socio-economische positie van ieder individu op het einde van elk kwartaal.

Maar omdat administratieve bestanden niet zijn opgesteld met het oog op arbeidsmarktonderzoek zijn er ook een aantal nadelen verbonden aan het gebruik van dergelijke gegevens. In tegenstelling tot enquêtes is er in administratieve bestanden geen kwalitatieve informatie beschikbaar en zijn dus enkel kwantitatieve analyses mogelijk. Daarnaast volgt elk administratief bestand zijn eigen begrippenlogica. Binnen de constructie van het Datawarehouse zijn die in de mate van het mogelijke op elkaar afgestemd, maar vergelijkingen met administratieve gegevens uit het buitenland zijn problematisch. Denken we maar eenvoudig aan de manier waarop werkloosheid wordt geconcipieerd en geregistreerd in verschillende landen met verschillende wetgevingen. Een ander specifiek probleem van het Datawarehouse is de onvolledige dekkingsgraad. Niet alle inwoners van België zijn gekend in het Datawarehouse. Het probleem situeert zich voornamelijk bij niet-beroepsactieven. In de toekomst zal de dekkingsgraad wel toenemen doordat meer en meer bestanden zullen worden geïntegreerd in het Datawarehouse (bijvoorbeeld het pensioenkadaster). Een andere specifieke tekortkoming vanuit arbeidsmarktoogpunt is het ontbreken van informatie over het onderwijsniveau van de bevolking.

Verder zijn er nog een aantal knelpunten wanneer we mobiliteit willen meten. Het uitgangspunt is dat

van de populatie in het Datawarehouse ieder kwartaal een socio-economische positie gekend is (bijvoorbeeld werkend, werkloos, brugpensioen, ...). Iemand die van socio-economische positie verandert, is bijgevolg socio-economisch mobiel. Wanneer we echter mobiliteit op kwartaalbasis willen onderzoeken, worden we geconfronteerd met het probleem van 'administratieve vertragingen' doordat bepaalde veranderingen van socio-economische positie niet onmiddellijk of foutief worden geregistreerd.

Een tweede probleem duikt op bij het meten van jobmobiliteit, het veranderen van werkgever. Een job wordt in het Datawarehouse gemeten als een combinatie van een identificatienummer van een werknemer met een identificatienummer van een werkgever. Wanneer die relatie wijzigt en een werknemer gekoppeld wordt aan een andere werkgever, dan beschouwen we de werknemer als jobmobiel. Het gebeurt echter dat er wijzigingen optreden in het identificatienummer van de werkgever, bijvoorbeeld in geval van een fusie, een overname of een afsplitsing. In dat geval wijzigt ook de relatie tussen de identificatienummers van de werknemer en werkgever waardoor we jobmobiliteit meten die in feite geen jobmobiliteit is. De werknemer blijft immers bij dezelfde werkgever werken, maar het identificatienummer van die werkgever wordt gewijzigd. Tot op zeker hoogte wordt deze 'schijnmobiliteit' in het Datawarehouse gecorrigeerd. De idee is dat indien een bepaalde

**Tabel 4.**

Jobmobiliteitsgraad op jaarbasis in NACE 72 Computer en aanverwante activiteiten (België; tweede kwartaal 1999-2000).

NACE	Aantal loontrekkenden (n)	Totaal (%)	Man (%)	Vrouw (%)
NACE 72.1 Computeradviesbureaus	11 008	17,0	17,7	15,0
NACE 72.2 Realisatie van programma's en gebruiksklare systemen	12 600	15,0	15,1	14,5
NACE 72.3 Gegevensverwerking	1 815	10,6	11,7	8,2
NACE 72.4 Databanken	818	12,7	11,6	14,3
NACE 72.5 Onderhoud, reparatie van computers en kantoormachines	1 432	12,6	13,3	9,8
NACE 72.6 Overige activiteiten i.v.m. computers	81	29,6	32,7	23,1
Totale Economie	2 778 338	7,2	7,4	6,8

Opmerking: enkel de personen die gedurende de hele beschouwde periode loontrekkend waren zijn opgenomen in deze analyse.

**Bron:** Datawarehouse Arbeidsmarkt bij de KSZ (Bewerking Steunpunt WAV).

transitie van werkgever x naar werkgever y wordt gemaakt door twintig werknemers of meer, er sprake is van een collectief fenomeen dat in realiteit geen echte jobmobiliteit is.

Een laatste probleem in verband met het meten van jobmobiliteit is dat het Datawarehouse gebruik maakt van de gecentraliseerde statistiek van de RSZ, hetgeen betekent dat we spreken over ondernemingen en niet over de lokale vestigingen van die ondernemingen. Een verandering van job is dus een verandering van onderneming. Dat betekent concreet dat een verandering van job binnen de onderneming, bijvoorbeeld een verandering van lokale vestiging maar ook een verandering van functie binnen de onderneming, niet wordt gemeten. In de toekomst zal het gebruik van een uniek identificatienummer voor lokale vestigingen ook informatie beschikbaar maken op dit niveau zodat de mobiliteit tussen vestigingen kan worden becijferd. Bovendien zal ook het probleem van de 'schijnmobiliteit' hierdoor worden opgelost.

## **CLFS of Datawarehouse?**

---

Uit de analyses van Work Package 6 *Mobility in the eEconomy* van het STILE-project blijkt dat het in theorie mogelijk is om mobiliteit in de ICT-sector te becijferen op basis van de CLFS, maar de praktische uitwerking ervan geeft weinig bevredigende resultaten. Dit is voornamelijk het gevolg van de beperkte betrouwbaarheid van de gegevens door het vereiste detail van bepaalde variabelen en ook door de nadelen die verbonden zijn aan de retrospectieve wijze van bevraging. Voor gedetailleerd onderzoek is de CLFS een ongeschikt instrument. Het gebruik van administratieve gegevens zoals het Belgische Datawarehouse, geeft daarentegen wel bevredigende resultaten voor dergelijk onderzoek, maar kampt vanzelfsprekend ook met een aantal beperkingen zoals een gebrek aan internationaal vergelijkbaar cijfermateriaal en het ontbreken van meer kwalitatieve informatie.

Maarten Tielens  
Steunpunt WAV  
Alex Stimpson  
Camire – Luxemburg

## **Noten**

1. Het eindrapport *Mobility in the eEconomy* is beschikbaar op de website van het STILE-project ([www.stile.be](http://www.stile.be)). Hier kan u ook terecht voor meer informatie over andere Work Packages van STILE.
2. De andere partner in Work Package 6 is Camire, een Spaans consulting bedrijf (met een vestiging in Luxemburg) dat zich onder andere specialiseert in het verzamelen, beheren en analyseren van sociale statistieken (arbeid, onderwijs, wetenschap en technologie).
3. Het Datawarehouse Arbeidsmarkt, een project binnen het AGORA-programma, behelst een samenwerkingsverband tussen de Kruispuntbank, de RVA, de RSZ, het RSVZ, het RIZIV, de RKW, de RSZPPO en de RVP. Het beoogt de koppeling van arbeidsmarktgegevens afkomstig van deze instellingen van sociale zekerheid. Tevens kan het aangevend worden om statistieken op te stellen die niet door één enkele instelling van sociale zekerheid kunnen worden berekend. Het datawarehouse bevindt zich fysiek in de lokalen van de SmalS-MvM. Het Steunpunt Werkgelegenheid, Arbeid en Vorming (Steunpunt WAV), een inter-universitair samenwerkingsverband gehuisvest aan de KU-Leuven en het Centre de Sociologie du travail, de l'Emploi et de la Formation (Point d'appui TEF) van de ULB verlenen wetenschappelijke ondersteuning. Het beheer van het datawarehouse werd toevertrouwd aan de Kruispuntbank van de Sociale Zekerheid. Voor meer informatie over het Datawarehouse verwijzen we naar artikels in de vorige edities van *Over.Werk* of naar de website van de KSZ: [http://ksz-bcss.fgov.be/nl/statistiques/stats\\_1.htm](http://ksz-bcss.fgov.be/nl/statistiques/stats_1.htm).

## **Bibliografie**

- Allaart, P.C. & van Ours J.C. (2001). *Stromen op de Nederlandse arbeidsmarkt*. OSA.
- OECD (2000). *Measuring the ICT sector*. Paris: OECD.
- Statistics Canada (download 16/04/2003). Data Quality Statement for the Labour Force Survey, reference number 3701, [www.statcan.ca/english/sdds/document](http://www.statcan.ca/english/sdds/document)
- Paull, G. (2002). *Biases in the reporting of labour market dynamics*. London: The Institute for Fiscal Studies.