



## Nederlandse data zijn te duur

**Auteur(s):**

Borghans, L.

*Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA), Universiteit Maastricht. Met dank aan onder andere Ron Dekker, Joachim Frick en Dean Lillard voor hun informatie en commentaar. [l.borghans@roa.unimaas.nl](mailto:l.borghans@roa.unimaas.nl)***Verschenen in:**

ESB, 88e jaargang, nr. 4397, pagina 132, 21 maart 2003

**Rubriek:****Trefwoord(en):**

onderzoek

*In plaats van forse bedragen te vragen voor datasets, kan de overheid het gebruik van data beter subsidiëren. Dat bevordert onderzoek naar de Nederlandse economie.*

Data voor sociaal-wetenschappelijk onderzoek zijn in Nederland moeilijk toegankelijk en duur. [tabel 1](#) geeft als voorbeeld een vergelijking tussen vier landen voor twee databronnen die voor arbeidseconomen erg belangrijk zijn. De kolommen betreffen de Labor Force Survey en het Household Panel. De Labor Force Survey is een enquête met een omvangrijke steekproef, maar een relatief korte vragenlijst, die als doel heeft algemene ontwikkelingen op de arbeidsmarkt in kaart te brengen. De Nederlandse Labor Force Survey is de Enquête Beroepsbevolking (ebb). Een huishoudenspanel is een longitudinale enquête met een kleinere steekproef waarbij echter meer informatie wordt verzameld dan in de enquête beroepsbevolking, zodat sociaal-economische ontwikkelingen in een bredere samenhang kunnen worden bestudeerd. De Nederlandse versie hiervan is het Sociaal Economisch Panel<sup>1</sup>.

**Tabel 1. Prijs en beschikbaarheid van diverse arbeidsmarktdata op 1 februari 2003**

Land	labor force survey	household panel
Nederland	ebb 1987-2000 alles: € 19.299 2000: € 2275	sep 1984-1999 alles: € 19.299 1999: € 2275
Duitsland	Mikrozensus wordt niet beschikbaar gesteld	gsoep 1982-2001 € 100 vragenlijsten in het Engels beschikbaar
Verenigd Koninkrijk	qlfs 1975-augustus 2002 gratis	bhps 1991-2001 gratis
Verenigde Staten	cps 1962-december 2002 gratis	psid 1968-2001 gratis

### Recente ontwikkelingen

In alle vier de landen, Nederland, Duitsland, het Verenigd Koninkrijk en de Verenigde Staten, heeft de laatste jaren een ontwikkeling plaatsgevonden waardoor de toegang tot microbestanden voor onderzoekers is verbeterd. Deze ontwikkeling wordt veroorzaakt door innovaties in computertechnologie waardoor via cd-roms en het internet gemakkelijker data kan worden gedistribueerd, maar ook door een rationalisering van het onderzoeksproces, waardoor er meer aandacht is voor de vraag of de onderzoeker ook wel aan geschikte data kan komen.

#### *Beperkingen versus openheid*

In Nederland is enkele jaren geleden door nwo het Wetenschappelijk Statistisch Agentschap (wsa) opgericht, dat tot doel heeft belangrijke databestanden toegankelijk te maken voor onderzoekers. Hoewel het wsa een belangrijke rol heeft gespeeld bij het vergroten van de beschikbaarheid van data, blijkt uit de tabel dat Nederland ongunstig afsteekt bij de andere landen. In de Verenigde Staten is openheid en toegankelijkheid al sinds lange tijd zeer gebruikelijk. Wie onderzoek wil doen met de Amerikaanse dataset cps kan op [www.nber.org](http://www.nber.org) alle data van 1962 tot enkele maanden geleden zonder problemen en met uitgebreide documentatie downloaden. In het Verenigd Koninkrijk regelt het Data Archive op zeer efficiënte wijze de verspreiding via internet en wordt de gebruiker alleen gevraagd zich te registreren. In Duitsland, waar het van oudsher met name voor niet-Duitsers moeilijk was om data te krijgen, is het beleid ook sterk aan het veranderen. Een cd-rom met zo'n twintig jaar van het bestand gsoep kan men inmiddels zonder veel moeite voor honderd euro productie- en verzendkosten krijgen. Van alle vragenlijsten is een Engelse vertaling beschikbaar. Alleen de Duitse Labor Force Survey is voor onderzoekers buiten Duitsland nog steeds moeilijk te krijgen.

## Nederland

In Nederland kan men de twee databestanden inmiddels ook via het genoemde wsa bestellen, maar in vergelijking met de andere landen zijn deze bestanden erg duur en vaak niet erg actueel. De documentatie is veel minder uitgebreid en aan de uniformiteit van bestanden over de verschillende jaren is nauwelijks aandacht besteed. De vergoeding die wordt gevraagd voor het gebruik van deze databestanden is waarschijnlijk een poging om de kosten van dataverzameling gedeeltelijk te verhalen op de gebruiker. In mijn ogen is een dergelijk tariefstelsel voor onderzoeksbestanden maatschappelijk gezien inefficiënt. Het zou beter zijn om de data gratis ter beschikking te stellen aan onderzoekers, terwijl er zelfs voor subsidie op gebruik iets te zeggen valt. Er zijn drie argumenten waarom de data op zijn minst gratis zouden moeten zijn.

### *Gratis data*

Het eerste argument is dat data een niet-rivaliserend goed zijn<sup>2</sup>. De marginale kosten om een extra kopie van de data te verstrekken zijn erg laag en dus is er vanuit economisch perspectief geen goede reden om meer dan die productiekosten in rekening te brengen. Hoogstens wordt een herverdeling van geld bewerkstelligd tussen instanties (cbs en universiteiten) die allemaal door de overheid worden bekostigd<sup>3</sup>. Het tweede, meer principiële argument, dat met name in de Verenigde Staten van groot belang wordt geacht, is het democratisch principe. Als de overheid haar beleid baseert op gegevens uit enquêtes, dan is het redelijk dat de burgers in staat worden gesteld op basis van deze gegevens hun eigen analyses uit te voeren, zodat ze de argumenten die ten grondslag liggen aan het beleid met gelijke middelen ter discussie kunnen stellen. Zo ontstond er in de vs veel commotie omdat data van het star-experiment, waarbij het effect van klassenverkleining op scholen werd onderzocht, niet publiekelijk beschikbaar werden gesteld. Voor Nederland geldt dat als iemand op basis van gegevens uit het Sociaal Economisch Panel in esb kanttekeningen bij het wao-beleid van de regering wil maken, hij eerst €4.550 moet neertellen om over twee jaren van dit panel te beschikken. Ook voor de wetenschappelijke discussie is openheid en replicerbaarheid van groot belang.

### *Nederland in beeld*

Het derde argument is dat het principe dat de gebruiker betaalt weinig oog heeft voor de prikkels van onderzoekers en de belangen van de Nederlandse samenleving. Voor beleidsmakers en de Nederlandse samenleving in haar geheel is het gunstig als veel onderzoek wordt gedaan naar actuele ontwikkelingen in Nederland. De praktijk van het onderzoek is echter dat het voor de onderzoeker niet zo veel uitmaakt of hij een artikel schrijft over Nederland of een ander land. Ook als recente data niet voorhanden zijn, kan een nieuwe theorie ook wel getoetst worden op basis van oudere data. Door data over Nederland duur te maken en pas laat beschikbaar te stellen, stimuleert de overheid dus dat Nederlandse onderzoekers hun aandacht verleggen naar andere landen en vervlogen tijden. De onderzoeker die zich wel wil richten op de Nederlandse situatie (vaak omdat zijn onderzoek direct betaald wordt door een ministerie) gaat bekibbelen op de uitgaven en gebruikt in plaats van de hele reeks aan beschikbare jaren een of twee jaren. De kwaliteit van het onderzoek loopt dus terug, terwijl slechts een herverdeling van geld tussen verschillende overheidsinstanties tot stand komt.

### **Analyse publicaties**

Om na te gaan of een goede beschikbaarheid van onderzoeksdata inderdaad onderzoek in binnen- en buitenland stimuleert, heb ik de ontwikkeling van het aantal publicaties op basis van het Duitse databestand gsoep geanalyseerd. Alle publicaties in het publicatiebestand EconLit die in de samenvatting het gebruik van gsoep als databron vermelden, zijn geteld. Hoewel de gebruikte data niet altijd in de samenvatting te vinden zijn, geeft deze indicator een goed beeld van de ontwikkelingen als de kans om genoemd te worden niet verandert. Gsoep is in 1991 (vrijwel) gratis toegankelijk geworden voor onderzoekers in binnen- en buitenland. Sinds 1993 wordt deze mogelijkheid nadrukkelijk in de publiciteit gebracht. Dit beleid heeft zijn uitwerking niet gemist. Met een vertraging van enkele jaren is het aantal getelde publicaties op basis van gsoep gestegen van een niveau van vier à vijf naar rond de vijftien per jaar. Terwijl voor 1996 75 procent van de publicaties door Duitsers werd geschreven, is sinds 1996 nog slechts 43 procent van de publicaties van Duitse origine, werd 34 procent geschreven door Amerikanen en 23 procent door onderzoekers uit andere landen.

### *Beleidswijziging gewenst*

Juist met een voortvarender beleid ten aanzien van het beschikbaar stellen van data kan dus worden gestimuleerd dat onderzoekers zich in hun onderzoek op Nederland richten. Nederland zou wellicht zelfs gratis profijt kunnen hebben van de onderzoeksinspanningen van buitenlandse onderzoekers, net zoals Nederlandse onderzoekers nu sociaal-economische ontwikkelingen in de vs, Engeland of Duitsland analyseren. Ook valt te verwachten dat Nederland vaker opgenomen zal worden in internationaal vergelijkende studies. Eigenlijk zou een subsidie dus op zijn plaats zijn. Dit kan natuurlijk niet door aan iedereen die de data van internet haalt geld te geven, omdat dit ook een kostbare stimulans voor kwalitatief laagwaardig onderzoek zou betekenen. Echter door te investeren in een toegankelijke website, met goede documentatie en hulpprogramma's voor het gebruiken van de data, met Engelse vertalingen van de data en documentatie, kan onderzoek naar de Nederlandse samenleving op een eenvoudige en effectieve manier worden gestimuleerd.

---

<sup>1</sup> Overigens bestaat er in Nederland ook het vergelijkbare OSA-aanbodpanel. Volgens opgave van het WSA zijn hiervan voor de periode 1985-1998 acht bestanden beschikbaar voor € 453,78 per stuk (samen dus € 3630,24).

<sup>2</sup> L.L.G. Soete en B. ter Weel gaan in ICT and access to research data: an economic review (Ministerie van OC&W, Zoetermeer, december 2002) in op de vraag welke onderzoeksdata als publiek goed moeten worden beschouwd.

<sup>3</sup> Ook voor vrijwel al het contractonderzoek dat wordt uitgevoerd door universiteiten en onderzoeksbureau's is de overheid opdrachtgever, waardoor ook hier de overheid uiteindelijk haar eigen rekening krijgt gepresenteerd.

