

# Eens gegeven blijft gegeven?

J. Bade en T.W.G. van Bergen\*

**D**e markt voor databestanden is ondoorzichtig en werkt slecht. Het wetenschappelijk onderzoek wordt daardoor geschaad. Omdat dit onderzoek belangrijke positieve externe effecten kan hebben, is het van belang na te gaan hoe de werking van de markt kan worden verbeterd en de beschikbaarheid van data kan worden vergroot.

In 1992 heeft, in opdracht van de Stichting ECO-ZOEK, een onderzoek plaatsgevonden naar de data-voorziening voor economisch onderzoek<sup>1</sup>. Aanleiding vormden de klachten van economische onderzoekers over de slechte beschikbaarheid, de hoge prijs en de ontoereikende kwaliteit van databestanden die voor onderzoek noodzakelijk zijn. De belangrijkste conclusie uit het onderzoek onder 130 economen in Nederland is dat empirisch economisch onderzoek voornamelijk 'data driven' is: de beschikbaarheid van data bepaalt het onderzoek. Dit betekent dat kansen worden gemist.

Omdat wetenschappelijk onderzoek belangrijke externe effecten kan hebben in de vorm van maatschappelijk nuttige onderzoeksresultaten, is het van belang om na te gaan waarom en door wie databestanden worden gegenereerd en in hoeverre ze beschikbaar zijn. In dit artikel wordt een analyse gegeven van de markt voor databestanden. Na een algemene beschrijving van de markt mondt een analyse van de marktonvolkomenheden en de nadelige gevolgen daarvan uit in een aantal concrete aanbevelingen.

## De vraag naar databestanden

De vraag naar databestanden is in te delen naar het doel waarvoor de bestanden primair worden gebruikt. Er zijn drie categorieën:

- vraag naar databestanden voor administratieve doeleinden en openbaar bestuur;
- vraag naar databestanden voor commerciële doeleinden;
- vraag naar databestanden voor wetenschappelijke doeleinden.

Voor categorie a is de overheid de grootste vrager (en aanbieder). De instantie die registreert is vaak ook de voornaamste en eerste gebruiker van het bestand. Bij deze categorie kan gedacht worden aan de (registratie)bestanden van het Centraal Bureau voor de Statistiek (CBS), bevolkingsregisters, bestanden van de belastingdienst, enz. De recente discussies over koppeling van de bestanden van sociale diensten en kentekenregistratie (om fraude met

bijstandsuitkeringen op te sporen) hebben aangetoond dat het voor de ene overheidsinstantie vaak niet mogelijk is het bestand van een andere overheidsinstantie te krijgen vanwege privacybescherming.

Categorie b bestaat uit bij voorbeeld de vraag naar specifieke adressenbestanden, marktonderzoeksgegevens of opiniepeilingen, waarbij de vrager de gegevens direct gebruikt voor marketingdoeleinden of als input voor commercieel onderzoek.

Categorie c betreft databestanden die gebruikt worden als input voor wetenschappelijk onderzoek. Op dit soort bestanden was ons onderzoek met name gericht. Deze bestanden kunnen door onderzoekers zelf zijn samengesteld, al dan niet op basis van een enquête. De data kunnen ook worden gekocht of verkregen van een overheidsinstelling, een commercieel bureau, een databank of van een archief.

Een bestand dat primair voor het ene doel is gegenereerd, kan daarnaast natuurlijk voor een ander doel worden gebruikt. Een wetenschappelijke instelling kan bij voorbeeld data verzamelen voor derde-geldstroomonderzoek en deze data ook gebruiken voor wetenschappelijk onderzoek (of het derde-geldstroomonderzoek zelf is wetenschappelijk). In dat geval zal het volgens bovenstaande indeling om een commercieel bestand gaan, dat secundair gebruikt is voor wetenschappelijk onderzoek. Het omgekeerde kan ook voorkomen.

## Het aanbod van databestanden

Overheidsinstellingen, financiële instellingen, markt-onderzoeksbureaus en wetenschappelijke instellingen zijn de belangrijkste binnenlandse aanbieders van databestanden. Vanuit het buitenland worden

\* J. Bade werkt als onderzoeker bij het Economisch en Sociaal Instituut van de Vrije Universiteit. T.W.G. van Bergen is via BOFEB werkzaam bij het Ministerie van Financiën.

1. Voor een volledig verslag van het onderzoek, zie J. Bade en T.W.G. van Bergen, *Datafaciliteiten voor economen. Van file tot snel verkeer*, VU Uitgeverij, Amsterdam, 1992.

	Vrije markt	Hoge drempels	Monopolie
<b>Registratie</b>	verkeerstellingen	import- en exportgegevens	belastinggegevens
<b>Commercie</b>	marktonderzoeksgegevens	telepanel data	bevolkingsregister
<b>Wetenschap</b>	enquêtes voor NWO of VF onderzoek	Sociaal Economisch Panel	CBS Bedrijven-enquête

**Figuur 1.**  
**Voorbeelden**  
**van typen**  
**data-**  
**bestanden**

data aangeboden door overheden, internationale instellingen (OESO, EG, VN) en particuliere databanken.

Afhankelijk van de hoogte van toetredingsdrempels zijn drie deelmarkten te onderscheiden, namelijk:

- een vrije markt, zonder hoge toetredingsdrempels<sup>2</sup>;
- een markt met hoge toetredingsdrempels in de vorm van organisatiekosten en/of goodwill;
- een monopolistische markt, ten gevolge van institutionele drempels.

Bij de eerste deelmarkt is sprake van volledige mededinging en gaat het vooral om data die niet continu worden verzameld en om informatie die op basis van vrijwilligheid kan worden gekregen of gekocht. Het grote aantal enquêtebureaus en de enquêtes die onderzoekers houden, bevestigen dat toetreding tot deze markt, waarop voornamelijk microbestanden worden verhandeld, betrekkelijk eenvoudig is. De beschikbaarheid van deze data kent dan ook geen drempels.

Anders ligt het bij data die doorlopend of periodiek verzameld worden (de tweede deelmarkt). Voorbeelden zijn: handelsdata, voorraadgegevens of paneldata. Voor het verzamelen van dergelijke gegevens is, behalve medewerking van de respondenten, een grote en/of goede organisatie nodig. Hier bestaan dus wel toetredingsdrempels, maar is mededinging nog steeds mogelijk. Ook voor deze data geldt nog dat ze vrij gemakkelijk te krijgen zijn.

Ten slotte zijn er data die om institutionele redenen slechts door één enkele instantie kunnen worden verzameld (de derde deelmarkt). Tijdreeksen van de diverse geldhoeveelheden bij voorbeeld. Maar ook data die alleen verkregen kunnen worden indien de respondenten met wettelijke middelen worden gedwongen informatie te verschaffen. In Nederland hebben het CBS, de Nederlandsche Bank en een aantal andere overheidsinstellingen (zoals het GAK, de arbeidsbureaus, onderwijsinstellingen, belastingdiensten, politie) bijzondere bevoegdheden met betrekking tot de registratie en/of het opvragen van informatie van individuen en bedrijven<sup>3</sup>. Deze speciale bevoegdheden leiden ertoe dat deze instellingen een monopoliepositie voor bepaalde gegevens hebben. Dit heeft implicaties voor de kwaliteit van deze data en de mogelijkheid deze informatie te verhandelen. Met name op grond van de Wet Persoonsregistratie (WPR) worden deze data nauwelijks beschikbaar gesteld.

De rest van dit artikel heeft met name betrekking op de bestanden die primair of secundair door de economische wetenschap worden gebruikt. Uit ons

onderzoek blijkt dat twee derde van de economen die met databestanden werken hun wetenschappelijk onderzoek als 'data driven' kenmerken. Dat betekent dat de onderzoeker zich op de hoogte stelt van de aanwezigheid van bepaalde data alvorens een onderzoeksvoorstel te formuleren. De overigen maakten vaak de kanttekening dat ze bij het schrijven van een onderzoeksvoorstel 'enigszins' rekening hielden met de data die nodig zouden zijn. De belangrijkste reden hiervoor is dat het verzamelen van data een kostbare en tijdrovende zaak is. Data worden zelden speciaal voor wetenschappelijk onderzoek verzameld (primair gebruik). Voor het gebruik van databestanden voor zuiver wetenschappelijk onderzoek wordt ook nauwelijks geld uitgegeven. Wel voor commercieel onderzoek, maar dan is het wetenschappelijk gebruik secundair. De conclusie lijkt gerechtvaardigd, dat bij datagebruik voor wetenschappelijk onderzoek het aanbod de vraag grotendeels bepaalt.

### Bijzondere kenmerken van databestanden

#### *Niet-rivaliserend goed*

Allereerst is een databestand een niet-rivaliserend goed: het kan meerdere keren en door verschillende gebruikers worden gebruikt. Dit leidt ertoe dat iedere koper onmiddellijk een potentiële concurrent van de producent wordt. Het is dus van belang of het al dan niet exclusieve recht op wederverkoop bij de prijs is inbegrepen. Tevens leidt de non-rivaliteit tot extra transactiekosten, omdat er kosten verbonden zijn aan uitsluiting (beveiliging tegen illegaal kopiëren, juridische kosten, enz.). Naarmate men meer zekerheid over de uitsluiting wil hebben, zullen de kosten toenemen<sup>4</sup>. De standaardoplossing voor de verschaffing van niet-rivaliserende goederen, namelijk collectieve voorziening, werkt niet bij databestanden omdat algemene beschikbaarheid juist niet gewenst is. De enige mogelijkheid is dus om de kosten van uitsluiting zoveel mogelijk te drukken. Dit kan door het straf-

2. Bij toetredingsdrempels moet gedacht worden aan beperkingen in de mogelijkheid om data te verzamelen en/of te verhandelen. Als het gaat om micro- en mesobestanden hangt die mogelijkheid onder meer samen met de bereidheid van subjecten om informatie te verschaffen en de praktische mogelijkheid de subjecten te ondervragen. Bij macrobestanden vormt met name de organisatie die vereist is om gegevens te verzamelen een toetredingsdrempel.

3. Dat deze gegevens door de overheid worden verzameld, komt niet voort uit een van de standaardmotieven voor overheidsvoorziening, te weten de hoge kosten van uitsluiting (dijkaanleg), het 'merit good'-motief (scholing) of het externe-effecten motief (riolering). De reden dat de overheid deze informatie verzamelt is dat de overheid zelf de grootste gebruiker van de verzamelde gegevens is en dat de informatieplicht overheidscontrole op de zorgvuldigheid vereist. Een vergelijking met het drukken van geld maakt duidelijk dat privatisering in principe wel mogelijk is.

4. Dit zou een argument voor het CBS en DNB kunnen zijn om vertrouwelijke informatie niet aan onderzoekers af te staan zonder toestemming van de informant. De kosten van uitsluiting en de kosten van verkeerd gebruik van de informatie (onder andere verlies van vertrouwen) zouden de baten van het onderzoek (inclusief externe effecten) kunnen overtreffen.

baar stellen van illegaal gebruik (net als bij het auteursrecht) en door het voorkomen daarvan, bij voorbeeld door het opzetten van gebruikersgroepen.

#### *Specifieke produktiefunctie*

Ten tweede wordt de produktie van een databestand gekenmerkt door zeer hoge vaste kosten en zeer geringe marginale kosten. Dit geeft de producent een natuurlijk monopolie. Als de data zijn verzameld, geschoond, gedocumenteerd en eventueel geanonimiseerd, zijn bijna alle kosten gemaakt. De kosten van een extra floppy zijn verwaarloosbaar.

### **Marktimperfecties**

#### *Externe effecten*

Onderzoek kent niet alleen directe baten voor degenen die het onderzoek uitvoert, maar dikwijls is er sprake van een extern effect. Bestanden zijn inputs voor wetenschappelijk onderzoek. Voor de onderzoeker zijn er substitutiemogelijkheden, want er kan ook onderzoek gedaan worden zonder data of met andere data. Indien een onderzoeker zuiver wetenschappelijk onderzoek uitvoert met bepaalde data zijn de wetenschappelijke artikelen de output. Als het onderzoek beleidsrelevant blijkt te zijn is die beleidsrelevantie een extern effect.

Dit is een argument voor overheidsregulering; de overheid zou het gebruik van bepaalde data (bij voorbeeld Nederlandse in plaats van buitenlandse) moeten stimuleren, door te zorgen dat ze zo goedkoop mogelijk beschikbaar worden gesteld. In de praktijk valt het tegenovergestelde waar te nemen. Bij de microbestanden van het CBS gaat het om gegevens die het CBS zelf gebruikt. Die data worden enigszins bewerkt en vervolgens wordt het gebruiksrecht van die databestanden tegen een, in vergelijking met het buitenland hoge, prijs aan onderzoekers verkocht. Het welvaartsverlies dat het resultaat is van het niet hanteren van marginale kostprijzen, wordt dus nog versterkt door het verlies aan positieve externe effecten.

Het is in dit verband belangrijk om op te merken dat in Nederland een zeer groot deel van het economisch onderzoek door de overheid wordt betaald. Alleen als onderzoek niet in opdracht van de overheid wordt gedaan of als bestanden van een niet-publieke instelling worden betrokken hebben we te maken met echte transacties. Voor de overige transacties geldt dat het eigenlijk om intra-overheidsleveranties gaat. De overheid is dus bezig met het rondpompen van geld. De prijsstelling is echter wel van invloed op de mate waarin (bepaalde) data worden verzameld en gebruikt voor onderzoek en dus op de externe effecten daarvan. Onderzoekers onderling blijken daarentegen over het algemeen wel bereid elkaar, onder bepaalde voorwaarden zoals wederkerigheid, bronvermelding enz., data te leveren tegen marginale (te verwaarlozen) kosten<sup>5</sup>.

#### *Gebrekkige prijsvorming en negatieve prikkels*

Dikwijls is de producent van een bestand tevens gebruiker (vrager) en bereid de volledige kostprijs te betalen<sup>6</sup>. Uit een oogpunt van efficiënte allocatie is

het dan optimaal indien dit bestand tegen marginale kosten wordt aangeboden voor secundair gebruik. De praktijk wijkt hiervan op twee punten af. Ten eerste worden vaak meer dan marginale kosten in rekening gebracht. Ten tweede wordt een bestand veelal niet aangeboden. Dit omdat de prikkel ontbreekt om een bestand te leveren. Hierdoor ontstaat de kans dat data dubbel verzameld worden. Instellingen zijn vaak zo georganiseerd dat de baten van de verkoop van een bestand niet ten goede komen aan de afdeling die het werk doet. Indien er extra werk gedaan moet worden, zoals wanneer bestanden slim gekoppeld of geanonimiseerd moeten worden, is er zelfs sprake van een negatieve prikkel. Een gevolg hiervan is ook dat, als er een bestand geleverd wordt, er veel klachten over het produkt zijn<sup>7</sup>.

De WPR en het begrip 'herleidbaarheid' spelen eveneens een grote rol<sup>8</sup>. Er gaat een negatieve prikkel van uit. De angst om problemen te krijgen op grond van de WPR leidt ertoe dat overheidsinstanties erg terughoudend zijn met het aanbieden van gegevens. Het ontbreken van jurisprudentie ten aanzien van de WPR en gebrek aan kennis van de juridische mogelijkheden om misbruik van gegevens tegen te gaan zijn daar gedeeltelijk debet aan. Verder worden anonimiseringsmethoden niet of in onvoldoende mate toegepast<sup>9</sup> en is er verwarring over het eigendomsrecht van informatie<sup>10</sup>.

#### *Ondoorzichtige markt en onzekerheid*

Als externe afnemers nodig zijn om het verzamelen van de data kostendekkend te maken, is het de vraag of de data wel beschikbaar komen. De prikkel tot

5. Zie J. Bade en T.W.G. van Bergen, op.cit., 1992, blz. 46.

6. Over de prijs die de producent als gebruiker voor het bestand moet betalen is een uitgebreide discussie gevoerd, vooral tussen onderzoekers en het CBS. Zie bij voorbeeld J. Hartog en H. Oosterbeek, Een beschouwing over de prijsstelling van het CBS, *Openbare Uitgaven*, 1990, nr. 22, blz. 5-10 en W. Begeer, De prijs van informatie, *Openbare Uitgaven*, 1990, nr. 22, blz. 131-138. In het kort komt het erop neer dat de onderzoekers vinden dat de overheid als sociale planner voor een 'first best'-allocatie moet zorgen en het CBS dus zodanig moet subsidiëren dat het de marginale kostprijs kan berekenen. Het CBS wijst op de hem opgedrongen budgettering en op extra vaste kosten die gemaakt moeten worden. Als de reserveringsprijs van de producent niet bekend is heeft een dergelijke discussie weinig zin.

7. De meest geuite klachten zijn dat de levertijd te lang is, dat de kwaliteit van de data onvoldoende is, dat databronnen verschillende definities hanteren, dat reeksen ophouden te bestaan of lacunes vertonen, dat er ongewenste aggregatie plaatsvindt, dat de voorwaarden voor datagebruik te restrictief zijn, dat het dataformat niet gestandaardiseerd is en dat de databestanden onvoldoende gedocumenteerd zijn. Zie J. Bade en T.W.G. van Bergen, op.cit., 1992, par. 4.2.2.

8. Zie W.J. Keller en J.G. Bethlehem, Disclosure protection of micro data, *CBS Select*, 1987, nr. 4, blz. 87-96; J. Bade en T.W.G. van Bergen, op.cit., 1992.

9. Voor een uitstekende beschrijving van anonimiseringsmethoden, zie P.F.A. de Guchteneire en J.G.M. Timmermans, *Wetenschappelijk gebruik van overheidsbestanden*, SWIDOC, Amsterdam, 1990, Appendix 7: Overzicht privacy-maatregelen.

verkoop van aanwezige informatie is er wel, maar een verkeerde inschatting van de vraag of een sterke aversie tegen risico kan ertoe leiden dat data niet verzameld worden. Van belang is of de markt doorzichtig is.

Bij een ondoorzichtige markt is de verwachte opbrengst van wederverkoop van belang. Of er data verzameld worden en een transactie tot stand komt en voor welke prijs hangt bij een ondoorzichtige markt af van hoe hoog de vrager en aanbieder die opbrengst schatten en wie die opbrengst krijgt. De opdrachtgever verwerft zich meestal het recht op wederverkoop en draagt het risico van tegenvallende verwachtingen. De prijs die bij wederverkoop tot stand komt hangt weer van dezelfde (bovengenoemde) factoren af. Verschillen tussen verwachtingen en realisatie leiden tot inefficiëntie. Een ondoorzichtig aanbod kan er toe leiden dat data dubbel worden verzameld of wederverkoop niet plaatsvindt, omdat de marktpartijen elkaar niet vinden.

Anders ligt het bij een doorzichtige vraag, waarbij aanbieder en vragers met hun reserveringsprijzen bekend zijn<sup>11</sup>. Nu is er geen sprake meer van onzekerheid en zal in de praktijk de vrager alleen het gebruiksrecht krijgen en het bestand dus niet kunnen doorverkopen.

## Markten voor wetenschappelijke data

Hieronder worden de imperfecties op de onderscheiden deelmarkten behandeld voor databestanden die door wetenschappers worden gebruikt. Dat zijn dus niet alleen bestanden die primair voor wetenschap zijn gegenereerd, maar ook registratiebestanden en commerciële bestanden, die secundair door wetenschappers worden gebruikt.

### *Vrije markt*

Allereerst de markt zonder toetredingsdrempels<sup>12</sup>. Omdat een databestand een niet-rivaliserend goed is, is een noodzakelijke voorwaarde voor de produktie ervan dat de vragers samen bereid zijn de kostprijs op te brengen<sup>13</sup>. Bij een betwistbare markt is dat ook de maximale prijs die de producent kan krijgen.

Zoals gezegd gaat het hier bij voorbeeld om datasets die, in opdracht, verzameld worden door commerciële enquêtebureaus, of om databestanden die door het CBS verzameld worden en waarbij het CBS geen gebruik maakt van de responsplicht (bij voorbeeld de enquête over verplaatsingsgedrag). Imperfecties zijn hier ten eerste het gevolg van gebrek aan informatie (ondoorzichtigheid van de markt) waardoor data niet verzameld worden (bij voorbeeld milieudata), of dubbel worden verzameld (bij voorbeeld kiezersonderzoek). Gebrek aan prikkels spelen hier een minder grote rol. Inefficiënte prijsvorming is waarschijnlijk wel belangrijk, maar moeilijk vast te stellen vanwege gebrek aan informatie over de reserveringsprijs van de producent die tevens gebruiker is, zoals hiervoor beschreven.

### *Toetredingsdrempels*

Toetredingsdrempels leiden ertoe dat de producent meer dan de kostprijs kan krijgen. Dat komt omdat

er nu wel 'sunk costs' zijn, bestaande uit de opgebouwde goodwill en organisatie. Van betwistbaarheid is dus geen sprake meer. Als bij voorbeeld de gebruikers van paneldata (bij voorbeeld het Sociaal Economisch Panel (SEP) van het CBS) niet de namen van de respondenten hebben, kunnen ze na de eerste ronde niet zonder meer naar een andere producent stappen. De nieuwe producent zou gewoon opnieuw moeten beginnen. Bij handelsdata of andere data die continu worden verzameld, geldt dat met name de bereidheid van respondenten om mee te werken en de hoge organisatiekosten een drempel voor toetreding zijn.

Om vast te stellen of er sprake is van monopolistische prijsvorming dient in het geval van een producent die ook gebruiker is onderzocht te worden wat zijn reserveringsprijs is. In de meeste concrete gevallen geldt echter dat een overheidsinstelling de producent is en de reserveringsprijs van de producent gelijk is aan de kostprijs. Afwijken van de marginale kostprijs is dan inefficiënt. Een goed voorbeeld is het SEP.

Informatiegebrek is er op deze deelmarkt nauwelijks. Gebrek aan prikkels wel, wederom met name bij overheidsinstellingen.

### *Monopolie*

Bij deze marktform is het voor een producent mogelijk om perfecte prijsdiscriminatie toe te passen en dus het gehele consumentensurplus af te romen. In de praktijk is de prijsvorming op deze deelmarkt echter niet het grootste probleem. Bij de bestaande monopolies op grond van de wet is er altijd sprake van een contract waarbij de vrager alleen het gebruiksrecht krijgt, als er al geleverd wordt<sup>14</sup>, omdat de overheid wil bepalen aan wie wordt verkocht. De overheidsinstellingen die de wettelijke bevoegdheden hebben om data af te dwingen blijken echter

10. De ervaring met commercieel marktonderzoek wijst uit dat respondenten steeds vaker betaald worden voor hun participatie en informatie. De respondent doet dan tegen betaling geheel (of gedeeltelijk) afstand van het eigendomsrecht van zijn informatie. Over de doeleinden waarvoor de gegevens gebruikt mogen worden, kan ook onderhandeld worden. Zo ontstaat een markt, waarbij het goed 'informatie' wordt getypeerd door inhoud en verhandelbaarheid. Overheidsinstanties als het CBS en andere institutionele dataverzamelaars en registrerende instellingen dienen respondenten en informatieverschaffers te vragen waarvoor de gegevens gebruikt mogen worden en hen eventueel voor het gebruik van hun informatie te betalen. Dit laatste zou ook de kwaliteit van de informatie kunnen verbeteren.

11. De reserveringsprijs is de prijs waartegen vrager en aanbieder niet langer bereid zijn af te nemen resp. te produceren.

12. De hoge vaste kosten vormen geen toetredingsdrempel indien ze gemaakt worden nadat een leveringscontract is afgesloten, dus als afzet en prijs al gegarandeerd zijn.

13. De volledige kostprijs kan een soort algemeen aanvaardde winstcomponent bevatten, die concurrenten ook in rekening brengen.

14. Respondenten van een enquête daarentegen, die vaak een monopoliepositie hebben waar het hun eigen informatie betreft, zijn meestal bereid gratis en zonder voorwaarde te leveren.

ook vaak de enige gebruikers van daarmee geproduceerde databestanden. De prikkel om de bestanden te anonimiseren, te documenteren en aan te bieden ontbreekt. Dit leidt ertoe dat een schat aan gegevens gedeeltelijk ongebruikt blijft. Hierbij kan gedacht worden aan gegevens van het GAK, de belastingdienst en het Ministerie van O&W. Een bestand met karakteristieken van personen die in de WAO zitten is hierdoor bij voorbeeld niet te krijgen en onderzoek op het gebied van inkomensontwikkeling en scholing wordt onnodig bemoeilijkt.

### **Verbetering van de marktwerking**

De bestaande marktimperfecties kunnen grotendeels worden opgeheven. Ten eerste dient daartoe de markt zo doorzichtig mogelijk gemaakt te worden. Dat betekent dat vraag en aanbod geïnventariseerd en bij elkaar gebracht worden. Ten tweede dienen er voldoende prikkels te zijn voor met name overheidsinstellingen om bestanden aan te bieden voor onderzoek. Te denken valt natuurlijk aan dwang door middel van een algemene maatregel van bestuur, maar er zou ook een financiële prikkel gegeven kunnen worden. Deze dient dan wel uit de algemene middelen betaald te worden (en niet door de vragers). De te verwachten positieve externe effecten rechtvaardigen een dergelijke stimulering. Ten slotte zou de prijsvorming bewaakt moeten worden. Bij een goede inventarisatie van vraag en aanbod moet het rondpompen van geld voorkomen kunnen worden. Bij secundair gebruik zouden slechts marginale kosten (manipulaties, schoning, anonimisering, documentatie, service) en transactiekosten (beveiliging) in rekening mogen worden gebracht.

### **Aanbevelingen**

Om de gesignaleerde problemen op te lossen, hebben we een aantal aanbevelingen gedaan. De belangrijkste twee zullen we hier noemen.

De eerste aanbeveling is één centrale instantie te vormen met een drievoudige taak, te weten informatie verschaffen, prikkels geven en misbruik tegengaan.

De eerste taak bestaat uit informatie verzamelen en verschaffen over enerzijds de beschikbaarheid van bestanden en anderzijds de behoefte aan databestanden. Hiervoor bleken onderzoekers de voorkeur te geven aan een 'makelaar', die vraag en aanbod bij elkaar brengt. De makelaar kan niet volstaan met het leggen van bilaterale contacten. Hij dient de markt zo doorzichtig mogelijk te maken.

De tweede taak voor een centrale organisatie betreft het geven van prikkels. Hierbij valt te denken aan krachtig lobbyen en aan financiële prikkels, zodat aan de gecoördineerde databehoeftte wordt voldaan. Een Statistisch Agentschap blijkt een dergelijke functie nu ten opzichte van het CBS te gaan vervullen, waarbij een 'multi user'-licentie voor een speciaal geprepareerd bestand van de Enquête Beroepsbevolking wordt afgesloten<sup>15</sup>. Dit is echter slechts het begin van wat er moet gebeuren. Voorwaarde is dat degene die extra werk verricht ook de prikkel krijgt. Het moet bij voorbeeld lonen om tijd in het koppen en anonimiseren van bestanden te steken. Ook moet het lonen om informanten te vragen of hun informatie voor een bepaald doel gebruikt mag worden en eventueel moeten informanten voor die toestemming kunnen worden betaald. Op grond van het eerder genoemde externe effect lijkt het voor de overheid gerechtvaardigd hier geld in te steken.

De derde taak is het garanderen van uitsluitbaarheid van het gebruik van bestanden. Een centrale instantie zou kunnen garanderen dat bestanden alleen gebruikt worden door bij haar geregistreerde onderzoekers. Hoewel dit verre van waterdicht is, kan het helpen om de kosten die met uitsluiting (en misbruik) gepaard gaan te verkleinen.

De tweede aanbeveling is dat de overheid een prikkel moet geven, door publiekrechtelijke instellingen in principe te verplichten bestaande databestanden aan elkaar te leveren tegen marginale kosten.

### **J. Bade en T.W.G. van Bergen**

---

15. Zie L. van der Laan en J.H.G. Segers, Databestanden en economisch onderzoek, *Economienblad*, 1992, nr. 7.