

**DE TOEGANKELIJKHEID VAN GEGEVENS UIT
PUBLIEKE EN SEMI-PUBLIEKE ADMINISTRATIES VOOR
WETENSCHAPPELIJK ONDERZOEK**

UPDATE SEPTEMBER 2009

Aanleiding en vraagstelling

De minister van OCW zal in het najaar van 2009 een nota betreffende wetenschappelijke informatievoorziening aanbieden aan de Tweede Kamer. Met de opkomst van informatisering en digitalisering worden er in administraties van overheidsinstellingen en semi-publieke organisaties op systematische wijze steeds meer gegevens gegenereerd die voor wetenschappelijk onderzoek van belang kunnen zijn. Het betreft veelal gegevens die in juridische zin wel voor wetenschappelijk onderzoek toegankelijk zijn, maar in de praktijk niet of slechts met veel inspanning beschikbaar komen voor wetenschappelijk onderzoek. Vaak moet er door wetenschappelijke onderzoekers extra worden betaald voor gegevens die al met publieke middelen tot stand zijn gekomen, zoals de gemeentelijke basisadministratie persoonsgegevens (GBA). Bij andere gegevensbestanden zoals de belastingadministratie zijn er wettelijke beperkingen.

Het ministerie van OCW is van plan in overleg te treden met de ministeries van Justitie en Binnenlandse Zaken om de wettelijke, organisatorische en financiële barrières te verminderen en daarmee de beschikbaarheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek te verbeteren. In dit licht heeft de OCW-werkgroep die de nota betreffende wetenschappelijke informatievoorziening voorbereidt, de Sociaal-Wetenschappelijke Raad van de KNAW bij brief van 13 januari 2006 verzocht meer inzicht te verschaffen in de behoefte van wetenschappelijke onderzoekers aan verbeterde toegang tot publieke en semi-publieke administraties. In mei 2006 bood de president van de KNAW het advies 'De toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek' aan aan de minister van OCW. In dit advies gaat de SWR in op de maatschappelijke baten van een betere toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek aan de hand van een aantal voorbeelden. In de zomer van 2009 bleek dat de nog niet afgeronde nota over de wetenschappelijke informatievoorziening uit 2006 voor eind 2009 door het ministerie van OCW zal worden afgerond. In dit licht is met de directie OWB (Onderzoek en Wetenschapsbeleid) van het ministerie van OCW afgesproken dat de SWR de notitie "De toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek" actualiseert.

Preambulante opmerking

Vooralsnog beperkt de vraagstelling van de OCW-werkgroep die de nota betreffende wetenschapsinformatie voorbereidt zich tot de administraties van publieke en semi-publieke organisaties. De SWR geeft de OCW-werkgroep in overweging ook de administraties van private organisaties en publieke toezichthouders in zijn nota te betrekken. Administratieve gegevens van alle organisaties, dus ook private organisaties en publieke toezichthouders, die onder de Nederlandse wet vallen, zijn toegankelijk voor wetenschappelijk onderzoek. Artikel 9, lid 3 van de Wet Bescherming Persoonsgegevens (WBP) luidt immers: "Verdere verwerking van de gegevens voor historische, statistische of wetenschappelijke doeleinden, wordt niet als onverenigbaar beschouwd, indien de verantwoordelijke de nodige voorzieningen heeft getroffen ten einde te verzekeren dat de verdere verwerking uitsluitend geschiedt ten behoeve van deze specifieke doeleinden". In de WBP wordt geen onderscheid gemaakt tussen typen organisaties die registratiehouder zijn.

De belangrijke discussie over de daadwerkelijke toegankelijkheid van al deze administratieve gegevens voor wetenschappelijke doeleinden heeft nog lang niet geleid tot bevredigende resultaten voor wetenschappelijke onderzoekers. Deze discussie is belangrijk, omdat wetenschappelijke onderzoekers in toenemende mate ook gegevens willen verwerken uit administraties van private organisaties en publieke toezichthouders.

Een voorbeeld

In alle nutssectoren en in dienstensectoren van algemeen belang (gezondheidszorg, wonen, onderwijs) is eenzelfde beweging zichtbaar van overheid naar markt. Gedurende de transitie is er een speciale rol weggelegd voor een toezichthouder van overheidswege. Zeker gesteld zou moeten worden dat publieke toezichthouders zoals

de Nederlands Mededingingsautoriteit (NMa), de Onafhankelijke Post en Telecommunicatie Autoriteit (OPTA), de Autoriteit Financiële Markten (AFM) en het College tarieven gezondheidszorg / de Zorgautoriteit i.o. (CTG/ZAio) voldoende informatie over de markt verzamelen waarop hun toezicht betrekking heeft en deze informatie in geschikte vorm voor wetenschappelijk onderzoek beter toegankelijk maken.

Baten voor het wetenschappelijk onderzoek van een betere toegankelijkheid van gegevens uit publieke en semi-publieke administraties

Het beter toegankelijk maken van gegevens uit publieke en semi-publieke administraties heeft belangrijke baten voor het wetenschappelijk onderzoek zelf, niet in de laatste plaats omdat de kwaliteit van deze administraties in Nederland groot is. Indertijd is besloten de Volkstelling af te schaffen. Een overweging daarvoor was dat de kwaliteit van de administraties zodanig is dat de volkstellingsgegevens daaruit geconstrueerd kunnen worden. Inmiddels is de virtuele volkstelling een feit.

Een betere toegankelijkheid van gegevens uit dergelijke administraties maakt het wetenschappelijke onderzoekers mogelijk:

- De gemeentelijke basisadministratie persoonsgegevens (GBA) te gebruiken als steekproefkader.
- Betrouwbaarder populatieschattingen te maken.
- De selectiebias te verminderen en te corrigeren.
- Analyses van panels, vooral wat betreft toe- en uitreding, te verbeteren.
- Analyses van non-response te verbeteren.
- Analyses van zeldzame gebeurtenissen in gegevens uit survey-onderzoek te verbeteren.

Het Burger Service Nummer (BSN) zal bovendien het koppelen van administraties onderling en met gegevens uit survey-onderzoek vergemakkelijken. In februari 2008 heeft de president van de KNAW de ministers van VWS, OCW en BZK verzocht het bij Algemene Maatregel van Bestuur (AMvB) mogelijk te maken dat organisaties voor wetenschappelijk onderzoek gebruik mogen maken van het Burger Service Nummer. De Raad voor Gezondheidsonderzoek (RGO) heeft in zijn advies "Van gegevens verzekerd. Kennis over de volksgezondheid in Nederland nu en in de toekomst" het standpunt van de KNAW onderschreven. Tot nu toe is er echter nog geen AMvB tot stand gekomen.

Een voorbeeld

Voor het verklaren van pensioneringsbeslissingen willen economen graag precies weten wat de keuzemogelijkheden zijn. Zij moeten dan haast wel over de gegevens van de pensioenfondsen beschikken. Maar zij willen ook andere factoren (gezondheid, gezinssamenstelling, opleiding, etc.) laten meewegen, dus koppeling van steekproefgegevens aan administratieve gegevens is ook hier de aangewezen weg. Het Centraal Planbureau (CPB) en het Network for Studies on Pensions, Aging and Retirement (Netspar) hebben hierover een pilotstudie afgerond met als gevolg dat er meerdere remote access terminals in gebruik zijn genomen.

De eerste deelconclusie is dat de betere toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek met name de kwaliteit van het empirische wetenschappelijke onderzoek ten goede komt.

Baten voor bestuur en beleid van betere toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek

Naast baten voor wetenschappelijk onderzoek zijn er ook baten voor bestuur en beleid. De toegenomen complexiteit van de samenleving stelt nieuwe en hoge eisen aan bestuurders ten aanzien van het monitoren van maatschappelijke processen, de vroegtijdige signalering van

belangwekkende maatschappelijke veranderingen en de raming van consequenties van beleidsmaatregelen voor verschillende groepen in de samenleving. Inzicht in processen, goede indicatoren en een toets op maatregelen die hun werking hebben bewezen (evidence-based interventions), bieden bestuurders een houvast in een snel veranderende samenleving. Wetenschappelijk onderzoek kan de gegevens die op systematische wijze in administraties van publieke en semi-publieke organisaties worden gegenereerd, vertalen in informatie en kennis over die samenleving.

Het beter toegankelijk maken van gegevens uit dergelijke administraties voor wetenschappelijk onderzoek heeft een aantal baten voor de overheid en de Nederlandse samenleving. De baten zijn onder meer:

1. Beter inzicht in de Nederlandse samenleving

Door het beter toegankelijker maken van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek zullen onderzoekers betere beschrijvingen en analyses van de Nederlandse samenleving kunnen maken. Nu doet zich de merkwaardige paradox voor dat Nederlandse wetenschappelijke onderzoekers analyses doen op basis van buitenlandse gegevens omdat deze vaak eerder en gemakkelijker beschikbaar komen voor het wetenschappelijk onderzoek.

Een voorbeeld

Nederlandse wetenschappelijke onderzoekers waren de eersten om het effect van overgewicht op levensverwachting en het verwachte aantal jaren met hart- en vaatziekten aan te tonen (zwaarlijvigheid vermindert de levensverwachting met zeven jaar). Voor het onderzoek werd gebruik gemaakt van longitudinale gegevens uit de VS (Framingham Heart Study; surveydata, geen administratieve gegevens). De resultaten van het onderzoek verschenen in 2003 in *Annals of Internal Medicine*, een vooraanstaand Amerikaans tijdschrift. Het artikel is volgens Web of Science meer dan 260 keer geciteerd in vooral Amerikaanse tijdschriften. De lagere levensverwachting van mensen met overgewicht vormde een aanleiding voor een gezamenlijk position statement door de American Society for Nutrition en de Obesity Society (2005).

2. De mogelijkheid om tijdig voor het beleid relevante ontwikkelingen te signaleren

Door het beter toegankelijker maken van gegevens uit deze administraties voor wetenschappelijk onderzoek kunnen onderzoekers actuelere beschrijvingen van de Nederlandse samenleving geven.

Een voorbeeld

Decennia lang had Nederland één van de hoogste levensverwachtingen in de wereld. Dat veranderde in de jaren 80. Terwijl in andere landen de levensverwachting verder toenam, stagneerde de sterftedaling in Nederland. Aanvankelijk werd de verklaring gezocht in traditionele risicofactoren zoals roken. Sedert 2002 stijgt de levensverwachting weer, vooral als gevolg van een sterftedaling op hoge leeftijd. Verandering in leefstijl kan die omslag niet verklaren. Daarom wordt thans de verklaring gezocht in de gezondheidszorg. Een recente analyse van administratieve gegevens toont dat de omslag in 2002 samenvalt met een toename van ziekenhuisopnamen na een verruiming van het budget voor de gezondheidszorg. De beschikbaarheid van administratieve data is essentieel om tijdig voor het beleid relevante ontwikkelingen te kunnen signaleren.

3. De mogelijkheid structurele ontwikkelingen in de Nederlandse samenleving te beschrijven

In publieke en semi-publieke administraties komen doorgaans gegevens voor die over een langere tijdsperiode zijn of worden verzameld. Door het beter toegankelijker maken van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek kunnen onderzoekers lange termijntrends in de Nederlandse samenleving beter beschrijven.

Een voorbeeld.

Op macroniveau zijn de Nederlandse R&D-uitgaven en het aantal bedrijven met R&D-activiteiten in Nederland vrij constant in de tijd. Wetenschappelijk onderzoek wijst echter uit dat er op bedrijfsniveau sprake is van een zeer grote dynamiek. Voor de meeste bedrijven variëren de uitgaven voor R&D zeer sterk van jaar tot jaar. Zo heeft meer dan eenderde van de bedrijven met R&D-uitgaven het jaar erop geen R&D-uitgaven. Deze ook voor beleidsmakers interessante bevindingen van wetenschappelijke onderzoekers worden over het hoofd gezien indien de Nederlandse R&D-uitgaven en het aantal bedrijven met R&D-activiteiten in Nederland alleen op macroniveau zouden worden gezien.

4. De mogelijkheid gegevens uit publieke en semi-publieke administraties verder te verbeteren

Wetenschappelijke onderzoekers zorgen er vaak voor dat de kwaliteit van de gegevens uit publieke en semi-publieke administraties verder verbetert, omdat zij uitvoeriger analyses kunnen uitvoeren dan degenen die de administraties beheren. Laatstgenoemden beperken zich veelal tot het presenteren van eenvoudige tabellen en kijken niet hoe fenomenen betreffende personen of organisaties die zij administreren, zich ontwikkelen in de tijd.

Wetenschappelijke onderzoekers zijn in staat de kwaliteit van de gegevens te verbeteren omdat zij als geen ander panelanalyses beheersen en daardoor eenvoudig controles in de tijd op micro-niveau kunnen uitvoeren. Ook kunnen zij verbanden leggen met informatie uit andere administraties en survey-onderzoek.

Een voorbeeld

Europa streeft naar een gecoördineerd en samenhangend migratiebeleid. De migratiestatistieken van Eurostat zijn echter maar zeer beperkt bruikbaar. In samenwerking met Eurostat en nationale overheden ontwikkelen consortia van onderzoekers uit verschillende landen van Europa methoden om migratiestatistieken te verbeteren. Onderzoekers maken gebruik van geavanceerde statistische theorie en technieken, en combineren gegevens uit verschillende databestanden. Via simulaties identificeren zij de zwakste plekken in het statistisch systeem en onderzoeken zij mogelijke oplossingen.

5. Vermindering van de administratieve lastendruk bij het bedrijfsleven

Wetenschappelijke onderzoekers verzamelen soms rechtstreeks gegevens over bedrijven en personen bij het bedrijfsleven. Zij kunnen, indien gegevens uit publieke en semi-publieke administraties beter toegankelijker zijn gemaakt voor wetenschappelijk onderzoek, afzien van deze manier van gegevensverzameling, wat de administratieve lastendruk bij het bedrijfsleven zal verminderen.

De tweede deelconclusie is dat betere, actuelere en trendmatiger kennis van de Nederlandse samenleving die bovendien efficiënt is gegenereerd, in belangrijke mate kan bijdragen aan het werk van beleidsmakers en besluitvormers in de publieke sector. Ook zij hebben dus belang bij een betere toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek.

Behoeftte van wetenschappelijke onderzoekers aan gegevens uit publieke en semi-publieke administraties

In het algemeen hebben wetenschappelijke onderzoekers die empirisch onderzoek verrichten, behoefte aan gegevens over personen en organisaties (bedrijven e.d.). Een indicatie van de omvang van deze behoefte zou kunnen worden ontleend aan het aantal actieve gewone hoogleraren in de gedrags- en maatschappijwetenschappen. In september 2009 waren er circa 2000 van deze hoogleraren. Hoewel niet alle leerstoelen zijn gericht op empirisch onderzoek, gaat het toch om zeer grote aantallen wetenschappelijke onderzoekers die behoefte hebben aan gegevens over personen en organisaties, want het is niet ongebruikelijk dat een hoogleraar leiding geeft aan een onderzoeksgroep van tien personen of meer.

Het Centraal Bureau voor de Statistiek (CBS) is vanouds een belangrijke toeleverancier voor wetenschappelijke onderzoekers wat betreft de gegevens over personen en organisaties. Het gaat dan onder meer om gegevens die afkomstig zijn uit publieke en semi-publieke administraties. Momenteel hebben wetenschappelijke onderzoekers via het CBS toegang tot een rijke verzameling aan administratieve gegevens zoals de gemeentelijke basisadministratie persoonsgegevens (GBA) en gekoppelde gegevenssets zoals het Sociaal Statistisch Bestand (SSB) en het Gezondheidsstatistisch Bestand. Ook is het CBS bijvoorbeeld bezig met gegevens van het ABP en PGGM over pensioenafspraken, die uiteindelijk gekoppeld zouden moeten worden aan andere gegevens en geanonimiseerd ter beschikking komen van wetenschappelijke onderzoekers. Het CBS publiceert geaggregeerde resultaten in zijn elektronische output-database (StatLine). Om de privacy te beschermen heeft het CBS de toegang tot de gegevens op microniveau aan strenge regels gebonden. Wetenschappelijke onderzoekers die toegang verkrijgen tot deze gegevens kunnen:

- Door hen gespecificeerde tabellen opvragen.
- Gegevensbestanden on site op het CBS analyseren.
- Gegevensbestanden analyseren via remote execution. Dit gaat als volgt in zijn werk. De wetenschappelijke onderzoeker stuurt vanaf de eigen werkplek programmatuur op waarmee het CBS de gegevens analyseert, het CBS controleert de uitkomsten op onthullingsrisico's en zendt deze na goedkeuring terug aan de wetenschappelijke onderzoeker.
- Gegevensbestanden analyseren met remote access. Hierbij biedt het CBS gemachtigde onderzoekers toegang tot microdata zonder dat zij naar het CBS hoeven te komen. Het CBS installeert bij de onderzoeker een beveiligd werkstation (uitgerust met biometrische authenticatie) waarmee op afstand analyses kunnen worden uitgevoerd. De microdatabestanden blijven op een daarvoor ingerichte server bij het CBS en kunnen door de gebruiker niet worden gedownload.
- Beveiligde microbestanden analyseren, die worden aangeboden op cd-rom via DANS. Hierbij zijn de gegevens verder geanonimiseerd en zijn niet alle originele gegevens beschikbaar. Voor alle microdata die ter beschikking worden gesteld, geldt dat ze zijn ontdaan van directe identificatoren, zoals naam en adres. Met uitzondering van de databestanden die via cd-rom worden aangeboden geldt dat de resultaten van de analyses door het CBS worden beoordeeld op risico's voor statistische onthulling. Het CBS brengt kosten in rekening voor het gebruiksklaar maken van de benodigde data en voor het gebruik van de faciliteiten.

De overheid werkt aan de zogenoemde "basisregistraties". Het kabinet heeft zes basisregistraties aangewezen die vitale gegevens bevatten van de overheid. Het ministerie van VROM is verantwoordelijk voor de totstandkoming van vier van de zes basisregistraties: Basis Gebouwen Registratie (BGR), Basis Registratie Adressen (BRA), Basisregistratie Kadaster en Basisregistratie Topografie. Het ministerie van Economische Zaken is verantwoordelijk voor het Nieuwe Handelsregister. Het ministerie van Binnenlandse Zaken en Koninkrijksrelaties is verantwoordelijk voor de gemeentelijke basisadministratie persoonsgegevens (GBA) en het gehele stelsel. Deze basisregistraties vormen zeer interessante gegevensbronnen voor wetenschappelijke onderzoekers.

Voor de Nederlandse onderzoeksgemeenschap is dus een zeer breed palet aan gegevens uit publieke en semi-publieke administraties in principe toegankelijk. In de praktijk zijn er echter aan de kant van de beheerders van de administraties aanzienlijke barrières in de toegankelijkheid van deze gegevens.

De termijn waarop data uit de basisregistraties daadwerkelijk ter beschikking komen is soms erg lang. De SWR geeft de OCW-werkgroep in overweging na te gaan of het mogelijk is een wettelijke termijn vast te stellen waarbinnen de data daadwerkelijk beschikbaar moeten zijn. Daarnaast zou de OCW-werkgroep kunnen aansturen op wetgeving die het mogelijk maakt dat bijvoorbeeld het GBA rechtstreeks opengesteld wordt voor wetenschappelijk onderzoek. Maar ook aan de kant van de wetenschappelijke onderzoekers zijn er barrières. Een belangrijke barrière is dat het hen niet altijd duidelijk is welke informatie, relevant voor hun onderzoek, deze administraties precies bevatten. Dat wetenschappelijke onderzoekers behoefte hebben aan verbeterde toegang tot publieke en semi-publieke administraties moge duidelijk zijn.

Een groot aantal gegevensbronnen is zeer boeiend voor wetenschappelijke onderzoekers. Naast de hierboven genoemde basisregistraties zijn er nog een aantal willekeurige bronnen te noemen:

- Verzekerdenadministratie werknemers (VZA), afkomstig van de uitvoeringsinstellingen werknemersverzekeringen.
- Aangiftegegevens fiscale ondernemingswinst natuurlijke personen, afkomstig van de Belastingdienst.
- Onderwijsnummer VO, afkomstig van de Informatie Beheergroep.
- BasisVoorziening Vreemdelingen, afkomstig van het ministerie van Justitie.

Het formuleren van de richting en omvang van de behoefte van wetenschappelijke onderzoekers aan gegevens uit publieke en semi-publieke administraties is een majeure uitzoekklus, waarvoor de personele en financiële middelen van de SWR te kort schieten.

Het initiatief van Data Archiving and Networked Services (DANS), om een 'landelijke data-agenda' te formuleren heeft geresulteerd in het voorstel 'Data Infrastructure for the Social Sciences' (DISS). Dit voorstel, dat door universitair onderzoekers, NWO, KNAW, CBS, SCP en DANS is ingediend, is geaccepteerd voor de Nederlandse Roadmap grootschalige onderzoeksfaciliteiten. De hoofddoelstelling van dit voorstel is het bieden van een beter inzicht in maatschappelijke veranderingen en de consequenties daarvan voor bestuur en beleid. Deze doelstelling wordt gedeeld door alle betrokken partijen, en vindt haar vorm in fundamenteel en beleidsgericht onderzoek. Het voorstel omvat drie onderdelen: surveys, methodologie en disseminatie. Belangrijke elementen uit het DISS-voorstel zijn: koppeling van surveys aan registratiedata en verbetering van de toegankelijkheid van microdata. Het voorstel wordt breed gedragen, maar heeft vooralsnog geen substantiële financiering gevonden.

Ten slotte

Gegevens uit publieke en semi-publieke administraties beperken zich veelal tot feitelijkheden over personen en organisaties. Omdat deze administraties echter geen informatie bevatten over bijvoorbeeld handelingen, houdingen of opinies, kunnen de gegevens uit deze administraties de gegevens uit survey-onderzoek slechts ten dele vervangen. In dit licht is het voor de wetenschap van groot belang dat wetenschappelijke onderzoekers survey-onderzoek blijven doen. Gegevens uit publieke en semi-publieke administraties vormen met feitelijkheden over personen en organisaties wel een belangrijke basis voor wetenschappelijk onderzoek. Het survey-onderzoek kan dan worden toegespitst op gegevens van personen en organisaties die niet uit deze administraties af te leiden zijn.