

Scandinavische toestanden in Nederland: toegang tot registerbestanden

Decennialang hebben Nederlandse onderzoekers met afgunst gekeken naar de mogelijkheden die er in de Scandinavische landen zijn om analyses te verrichten op basis van registergegevens. Inmiddels hebben we die mogelijkheden ook in Nederland, een ontwikkeling die zondermeer is toe te juichen.

In de Wet Bescherming Persoonsgegevens (WBP) is vastgelegd dat gegevens over individuen die voor een ander doel zijn verzameld en waarvoor geen beroepsgeheim geldt, in principe beschikbaar zijn voor wetenschappelijk onderzoek. Uiteraard gelden hier de regels van anonimiteit en geheimhouding en mogen de gegevens uitsluitend worden gebruikt voor wetenschappelijk onderzoek. In december 2005 zijn de implicaties van de WBP voor wetenschappelijke onderzoekers door de VSNU uitgewerkt in een gedragscode.

Wat betekent een en ander in de praktijk? Een concrete verandering is dat de toegang tot de Gemeentelijke Basisadministratie Persoonsgegevens (GBA) is geregeld. Voorheen was toegang tot de GBA uitsluitend mogelijk ten behoeve van de uitvoering van wetgeving, hetgeen inhield dat instellingen als het RIVM, SCP en RPB toegang hadden tot de GBA, maar universiteiten en KNAW-instituten niet. Deze ongelijkheid is er niet meer. Wetenschappelijke instellingen kunnen nu, op basis van een goedgekeurd onderzoeksplan, 'geautoriseerd afnemer' worden van de GBA. Verschillende typen gegevensverstrekkingen zijn mogelijk. Ten eerste is er een selectieverstrekking: men verkrijgt gegevens over alle personen die voldoen aan specifieke voorwaarden (bijvoorbeeld alle gescheiden vrouwen in de leeftijd 55 tot 89 jaar). Ten tweede is er een *ad hoc*-gegevensverstrekking: over specifieke steekproefleden wordt aanvullende informatie uit de GBA gevraagd (bijvoorbeeld of betrokkene nog op een bepaald adres woont). Ten derde zijn er spontane mutaties: hierbij controleert de GBA automatisch of bij steekproefleden veranderingen zijn opgetreden en geeft deze door (bijvoorbeeld een overlijden of verhuizing buiten de gemeente). De gegevens die een afnemer verstrekt mag krijgen, worden vastgelegd in een autorisatieregel.

Het is prachtig dat wetenschappelijk onderzoekers toegang kunnen krijgen tot de GBA, maar hiervoor moeten wel hobbels worden genomen. De procedure om geautoriseerd afnemer te worden kost tijd: ongeveer een jaar. De verstrekkingen zijn niet kosteloos en in het geval van een aansluiting op Gemnet (het elektronische netwerk van de publieke sector), zijn er kosten voor hard- en software.

Een geïntegreerd bestand voor alle gemeenten is er nog niet. Ter illustratie van de problemen die men kan tegenkomen, noem ik de selectieverstrekking aan een onderzoeker van het Utrechts Instituut voor Linguïstiek. Betrokkene zal de voornaam ontvangen van alle personen met de Nederlandse nationaliteit die in 1995 nog levend waren of daarna zijn geboren. Van de ouders worden de voornaam, de geboortedatum, het geboorteland en de postcode geleverd.

Van de kinderen is ook de voornaam bekend. Via koppelingen is in een groot aantal gevallen de voornaam van de grootouders en van ooms en tantes te traceren. Deze onderzoeker zal een complete genealogie van Nederlanders kunnen construeren. Hij krijgt de gegevens aangeleverd op 460 cd-roms (voor iedere gemeente één) en heeft een apart bureau ingehuurd om de data, die niet zomaar in te lezen zijn, te verwerken tot een formaat die met bestaande statistische pakketten te gebruiken is. Het bureau heeft ook tot taak de 460 gemeenten ertoe aan te zetten de gevraagde gegevens te leveren. Onderzoekers moeten er iets voor over hebben om gegevens aan de GBA te ontfemen.

Een andere verandering (die maar voor een deel is toe te schrijven aan veranderingen in de privacywetgeving) is dat de toegang tot registergegevens via het CBS enorm is verbeterd. Momenteel zijn zo'n 120 microbestanden (details zijn te vinden op de CBS-website) beschikbaar voor analyse. Een eerste groep betreft zogenaamde REOS-bestanden, bestanden die gekoppeld zijn aan de GBA via een record-identificatienummer. De REOS-bestanden zijn sterk sociaal-economisch georiënteerd (bijvoorbeeld verzuimstatistiek, arbeidsongeschiktheidsuitkeringen, inkomenspanelonderzoek). Een tweede groep betreft bedrijfseconomische microdata (onder andere productiestatistiek, transportgegevens, werkgelegenheidsenquête). Een derde groep is het Sociaal Statistisch Bestand (SSB). De 'kern' van het SSB bestaat uit een aantal onderling gekoppelde en opgeschoonde registers waarin demografische en sociaal-economische gegevens zijn opgenomen. Daarnaast zijn er zogenaamde SSB-satellieten, die betrekking hebben op specifieke onderwerpen, zoals de maatschappelijke positie van allochtonen en de ontwikkeling van criminaliteit in Nederland. Het SSB is nog volop in ontwikkeling. Toegang tot CBS-microbestanden wordt via een contract geregeld waarin het onderzoeksdoel is beschreven (doelbinding) en de geheimhouding en de benodigde databestanden zijn vastgelegd. De procedure neemt één à twee maanden in beslag. Te publiceren resultaten moeten door het CBS worden gecontroleerd op het mogelijke gevaar van statistische onthulling. Aan het gebruiken van CBS-microbestanden zijn kosten verbonden voor het gereedmaken en documenteren van de bestanden.

Analyse van de microbestanden is mogelijk door tegen betaling *on site* te werken bij het CBS op een beveiligde werkplek. De databestanden mogen niet mee naar buiten worden genomen. Een tweede mogelijkheid is via *remote execution*. Hierbij maakt de onderzoeker *set-ups*, op basis van een testbestand, die vervolgens naar het CBS worden gestuurd om tegen betaling uitgevoerd te worden op het 'echte' bestand. Na de succesvolle afronding van een pilot bij Netspar in Tilburg is er nu de mogelijkheid voor *remote access*: op de werkplek van de onderzoeker wordt op eigen kosten een speciale ruimte gecreëerd met een apart en beveiligd elektronisch netwerk via welke men toegang heeft tot contractueel overeengekomen databestanden. Najaar 2006 zullen op vijf instituten in Nederland *remote access*-ruimten zijn gerealiseerd.

Een praktisch obstakel dat onderzoekers tegenkomen die gebruik willen maken van gegevens die het CBS in beheer heeft, is dat toegang op korte termijn niet mogelijk is omdat de betreffende bestanden nog niet voldoende zijn gedocumenteerd. Met het SSB bijvoorbeeld is nog niet goed te werken. Beperkte menskracht is hier voor een deel debet aan, maar ook speelt mee dat het CBS op basis van zijn meerjarenprogramma andere prioriteiten heeft.

In een in mei 2006 uitgebracht advies aan de Minister van OCW heeft de Sociaal-Wetenschappelijke Raad de stand van zaken heel mooi samengevat. In juridische zin zijn gegevens uit registerbestanden toegankelijk voor wetenschappelijk onderzoek, maar in de praktijk komen deze niet of slechts met veel inspanning beschikbaar. Als onderzoekers hebben we de mogelijkheid onze wensen kenbaar te maken bij de door KNAW en NWO geïnitieerde Data Archiving and Networked Services (DANS). Ook kunnen we de Centrale Commissie voor de Statistiek (CCS), verantwoordelijk voor het meerjarenprogramma van het CBS, met verzoeken bestoken. Laten we dat vooral doen.

Pearl Dykstra