

Een gevecht tegen windmolens?

De bescherming van de persoonlijke levenssfeer van individuele medeburgers tegen eventueel onoirbare wensen en doelstellingen van het sociaal-wetenschappelijk onderzoek sluit direct aan bij het grondwettelijk recht op eerbiediging van de persoonlijke levenssfeer. Het is in principe een goede zaak dat de Wet Persoonsregistraties (WPR) – van kracht sedert 1 juli 1990 – tracht dit grondrecht te waarborgen. In dit verband wordt in de Memorie van Toelichting op de wet aangegeven wanneer de persoonsgegevens wel of niet tot individuele personen herleidbaar zijn: 'Indien (echter) een onevenredige hoeveelheid tijd, geld en mankracht nodig zou zijn om een dergelijke herleiding mogelijk te maken, dan zal er voor de toepassing van de wettelijke regeling van moeten worden uitgegaan dat deze gegevens geen persoonsgegevens in de zin van dit artikel zijn' (MvT, p.35).

Het is ook in het belang van het sociaal-wetenschappelijk onderzoek dat deze grondrechten onverkort worden nageleefd. Geïnterviewden, geënquêteerden en proefpersonen moeten er blindelings op kunnen vertrouwen dat hun individuele gegevens (persoonsgegevens) 'veilig' zijn in de handen van de onderzoekers. De bedoelde sociaal-wetenschappelijke onderzoekers zullen de resultaten van hun onderzoek uitsluitend naar buiten brengen op geaggregeerd niveau. Strikt genomen is er dus geen sprake van een wanverhouding tussen doelstellingen en de wijze van gebruik van individuele databestanden door onderzoekers enerzijds en de individuele rechten op bescherming van de persoonlijke levenssfeer anderzijds. Geanonimiseerde microdatabestanden, dat wil zeggen ontdaan van persoonsgegevens zoals naam, adres, geboortedatum, en met een verminderd informatiegehalte van enkele andere gevoelige gegevens zoals bijvoorbeeld beroep, zouden vrij algemeen beschikbaar moeten zijn voor sociaal-wetenschappelijke onderzoekers. Zeker voor die onderzoekers, die werken vanuit de gedragscode zoals vastgelegd door de Vereniging van Onderzoek Instituten (1991).

Het CBS als beheerder van een groot aantal van deze microdatabestanden, wenst daarnaast óók ingedekt te zijn tegen *onverlaten*, die expliciet op zoek zijn naar de gegevens van individueel herkenbare, bijvoorbeeld publiek bekende personen. Bijgevolg streeft men dan naar volledige anonimiteit; gegevens uit de CBS publieksbestanden zijn nooit herleidbaar tot een enkel individu. Bij het beschikbaar stellen van gegevens volgt men dan een beleidslijn die terughoudender is dan de WPR voorschrijft. Deze publieksbestanden zijn vervolgens veel minder interessant en veel minder geschikt voor gebruik in wetenschappelijk onderzoek; de informatie is teveel ingedikt, de details van belangrijke variabelen zijn aanzienlijk verminderd/verminkt. In dit verband stelt Vuijsje (1992) terecht: 'Het ontzien van deze gevoeligheid (voor privacy-schending) is een maatschappelijk goed, maar het is van belang te constateren dat daar ook een prijs tegenover staat. Waar de individuele privacy wordt afgeschermd, wordt het verzamelen van statistische gegevens moeilijker en duurder. Bovendien geldt dat de 'kosten' van privacybeschermende maatregelen toenemen, en de 'meeropbrengst' afneemt, naarmate die maatregelen perfectionistischer zijn'. Het komt er met andere woorden op aan de juiste afweging te maken tussen enerzijds de baten en anderzijds de kosten van bescherming van de persoonlijke levenssfeer. Van verschillende zijden zoekt men daarom ook naar een nadere bepaling van deze kosten en baten. Zo neemt de roep om een meer praktische operationalisering van de garantie op anonimiteit toe en wordt wel gestreefd naar 'feitelijke anonimiteit', waarmee een situatie van niet volledige garantie van anonimiteit wordt bedoeld. Echter re-identificatie van individuele personen mag binnen dit streven niet gerealiseerd worden anders dan met inzet van een onevenredige hoeveelheid tijd en kosten van de zijde van de 'brekers' van de anonimiteit.

Onlangs werd in Duitsland een project afgerond, waarin het Statistische Bundesamt, ZUMA en de universiteit van Mannheim met de bekende Duitse Gründlichkeit samenwerkten onder leiding van Walter Müller teneinde empirisch experimenteel vast te stellen tegen welke tijds- en financiële inzet bepaalde microdatasets konden worden gedeanonimiseerd. Zowel de koppeling van een feitelijk geanonimiseerd databestand aan het bijbehorende identificatiebestand van persoonsnamen, als andere vormen van gericht zoekgedrag (het identificeren van achtduizend personen vermeld in een 'Gelehrtenkalender' tegen de gegevens van een micro-census en Fischzug-doelen) werden nagestreefd. De gehanteerde methoden van onderzoek omvatten vergelijkingstechnieken (sorteer- en selectieroutines) en re-identificatietechnieken gebaseerd op discriminant-analyses.

Van de achtduizend personen uit de 'Gelehrtenkalender' werden er volgens de onderzoekers met behulp van sorteer- en selectieroutines mogelijk veertien

in een micro-databestand geïdentificeerd. Bij controle door een vertrouwenspersoon bleek echter dat deze conclusie slechts voor vier personen terecht zo getrokken was. Gelet op de omvang van het identificatiefile staat dit gegeven gelijk aan een kans op correcte toedeling van 0,0005.

In geval gebruik werd gemaakt van discriminant-analyses, ter vergelijking van een databestand en een identificatiefile met maar liefst 35 overlappende variabelen, werden ogenschijnlijk 35 personen geïdentificeerd. Bij controle door de vertrouwenspersoon bleken alle 35 cases op een foutieve koppeling gebaseerd te zijn.

De conclusie dat feitelijke anonimiteit 'werkt', ligt daarmee voor de hand. Zolang de gezochte persoon niet gekenmerkt wordt door een zeer bijzondere combinatie van zeldzame karakteristieken, worden in de poging tot koppeling vooral veel statistische 'dubbelgangers' gevonden. Het uitsélectioneren van deze sets van dubbelgangers vereist vervolgens een groter aantal overlappende variabelen in beide files. Met elke toevoeging van een extra overlappende variabele neemt echter de kans op data-incompatibiliteit toe en daarmee de waarschijnlijkheid dat foute koppelingen zullen resulteren. Aangezien databestanden, die uitgezet worden volgens het principe van feitelijke anonimiteit de facto niet te kraken blijken te zijn, is het te vuur en te zwaard verdedigen van volledige anonimiteit een onnodige, te restrictieve en voor het sociaal-wetenschappelijk onderzoek veel te kostbare onderneming.

Wanneer mag in Nederland een herafweging van de kosten en baten van de huidige praktijk van bescherming van de persoonlijke levenssfeer verwacht worden?

Jenny de Jong-Gierveld

Literatuur

Walter Müller, Uwe Blien, Peter Knoche, Heike Wirth u.a., *Die faktische Anonymität von Mikrodaten*. Wiesbaden: Statistisches Bundesamt, 1991.

Herman Vuijsje, *Mens, erger je niet; privacybescherming en wetenschappelijk onderzoek*. Ministerie van Onderwijs en Wetenschappen, 1992.