KWANTIFICATIE VAN HERLEIDBAARHEID

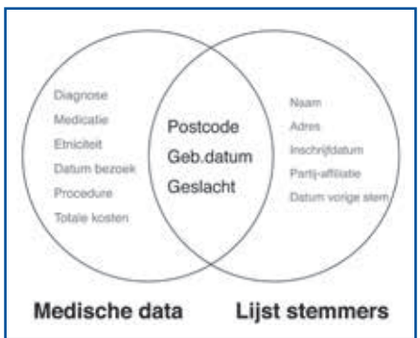
Matthijs R. Koot is als ethisch hacker werkzaam bij Madison Gurkha. Hij is op 27 juni 2012 aan de Universiteit van Amsterdam gepromoveerd op het proefschrift "Measuring and Predicting Anonymity". Dit artikel beschrijft enkele resultaten van dit onderzoek en verscheen eerder in het tijdschrift Privacy & Compliance. Matthijs is inzake dit artikel te bereiken op koot@uva.nl.

Elke ontwerpbeslissing die resulteert in betere bescherming van privacy kan worden aangeduid als Privacy by Design. Daarbij kan het gaan om het ontwerp van IT-systemen, maar ook om de opzet van onderzoek en enquêtes waarbij gegevens over personen worden verwerkt. Privacy by Design kan bijvoorbeeld zijn gebaseerd op het principe van gegevensminimalisatie: verwerk uitsluitend gegevens die noodzakelijk zijn voor het beoogde doel, en geen andere gegevens. De Wet bescherming persoonsgegevens (Wbp) maakt dit principe expliciet.

Het is ook mogelijk om verwerking van 'persoonsgegevens' in juridische betekenis uit te sluiten. De Wbp stelt dat elk gegeven betreffende een geïdentificeerde of identificeerbare natuurlijke persoon een persoonsgegeven is. Welnu, gegeven een database met gegevens over personen: welke aanpassingen zijn nodig zodat er niet langer sprake is van 'geïdentificeerd' of 'identificeerbaar'? Het verwijderen van persoonsgebonden nummers, namen en adressen uit een bestand leidt waarschijnlijk al snel tot opheffing van 'geïdentificeerd'. Maar wanneer is ook niet langer sprake van 'identificeerbaar'? Eén gedachte is dat gegevens niet langer identificeerbaar worden geacht wanneer deze alleen met *onevenredige inspanning* zijn te herleiden tot personen. Wanneer sprake is van *onevenredige* inspanning, dan is afhankelijk van de aard (medisch, financieel, justitieel, etc.) de misbruikwaarde van de privacygevoelige informatie lager dan de kosten. In een wereld van toenemende gegevensverwerking waarbij gegevens op steeds nieuwe manieren toegankelijk worden, moet worden aangenomen dat wat vandaag onevenredige inspanning kost, in de nabije toekomst mogelijk slechts *evenredige inspanning* gaat kosten. Kortom, dat het voor kwaadwillenden mogelijk rendabel wordt gedeïdentificeerde gegevens tot natuurlijke personen te herleiden. Het fenomeen datalekken en gerelateerde berichtgeving in de media mag worden opgevat als indicatie dat als gevolg van miserabele informatiebeveiliging, de mogelijkheden voor kwaadwillenden om gegevens te kunnen herleiden via gegevensdiefstal toenemen. In zekere zin is de uitspraak "attacks only get worse" van bekende IT-beveiliger Bruce Schneier hier ook van toepassing.

Sweeney en *k*-anonimiteit

Eind jaren '90 heeft de Amerikaanse onderzoekster Latanya Sweeney voor een gedeelte van de populatie van Massachusetts uitgezocht hoe herleidbaar die is via demografische informatie. Sweeney bleek records in een publiek beschikbare geanonimiseerde medische dataset te kunnen herleiden tot natuurlijke personen door de data op bepaalde demografische gegevens te koppelen met een niet-anonieme dataset van stemgerechtigden die zich hadden ingeschreven om te kunnen stemmen (zie figuur 1). De geanonimiseerde gegevens raakten hierdoor 'gedeanonimiseerd'. Gezien de triviale herleidbaarheid van de medische data was in feite in beginsel al geen sprake van anonimiteit, en mag dus eigenlijk niet van 'anonimiseren' worden gesproken. Nauwkeuriger is te spreken van 'de-identificeren' en 'heridentificeren'. Een combinatie van kolommen die tot heridentificatie kan leiden wordt 'Quasi-IDentifier' (*QID*) genoemd.



Figuur 1: Heridentificatie via overlappende kolommen

Als oplossing voor het probleem van heridentificatie bedacht ze '*k*-anonimiteit'. In dat model wordt vereist dat een dataset vóór publicatie een toets doorstaat: elke combinatie van waarden in de QID-kolommen *moet* ten minste *k* keren in de dataset voorkomen. Is dat niet het geval, dan dient de data via 'generalisatie' en 'onderdrukking' te worden aangepast, totdat de toets alsnog wordt doorstaan. Bij generalisatie wordt een gegeven opgehoogd: een geboortedatum wordt bijvoorbeeld vervangen door een leeftijd. Bij onderdrukking wordt een gedeelte van de data in een kolom weggelaten: bij een postcode kan bijvoorbeeld een letter worden weggelaten. Als de toets dan nog niet wordt doorstaan, kunnen beide letters worden weggelaten. Als de toets dan nog niet wordt doorstaan, kan weer worden gekeken naar generalisatie: de postcode vervangen door een plaats- of gemeentenaam, bijvoorbeeld.

Nota bene: Het model van *k*-anonimiteit is niet perfect. Dit heeft geleid tot diverse uitbreidingen op *k*-anonimiteit, elk met eigen voordelen en beperkingen. Een bespreking daarvan valt buiten het blikveld van dit artikel. Het is noodzakelijk dat bij het in praktijk nastreven van onherleidbaarheid rekening wordt gehouden met die verfijningen.

Heridentificeerbaarheid

(Her)Identificeerbaarheid kan worden uitgedrukt in termen van anonimiteit. Een definitie van anonimiteit die in het onderzoeksgebied aan belangstelling wint, luidt [1]:

"Anonimiteit is de onlinkbaarheid van een subject en een voorwerp van belang, gezien vanuit het perspectief van een aanvaller."

In de context van dit artikel vertegenwoordigt 'subject' een persoon waarover gegevens in een database staan geregistreerd; het 'voorwerp van belang' vertegenwoordigt een record over die persoon in de database; en 'aanvaller' vertegenwoordigt een model van een kwaadwillende.

De heilige drie-eenheid bij analyse van anonimiteit is dus:

- subject
- voorwerp van belang
- aanvaller

In deze conceptie heeft een claim over anonimiteit dan ook uitsluitend betekenis als aan al deze begrippen inhoud is gegeven. Een claim als "u bent anoniem" of "deze gegevens zijn anoniem" is dus betekenisloos. Een claim over anonimiteit is gebonden aan één subject, één voorwerp van belang en één kwaadwillende. Strikt genomen is aan elk record in een gedeïdentificeerde database een aparte claim over anonimiteit verbonden. De kwaadwillende wordt gemodelleerd in termen van de gegevens die tot zijn/haar beschikking staan en waarmee gedeïdentificeerde gegevens kunnen worden getracht te worden herleid tot natuurlijke personen.

> Niet identificeerbaar houdt in dat gegevens alleen met *onevenredige inspanning* zijn te herleiden tot personen

Persoonsgegevens zijn onontbeerlijk voor bepaalde terreinen van beleidsonderzoek. Een voorbeeld daarvan is het beleidsonderzoek naar patiëntenzorg in ziekenhuizen. Medio 1960 is de Landelijke Medische Registratie (LMR) opgericht [3]: een centraal archief waar ziekenhuizen kopieën van hun administratie van ziekenhuisontslagen aan verstrekken, bedoeld voor spiegelen van ziekenhuizen onderling, maar ook ten behoeve van verbetering van zorg. Ziekenhuizen kunnen vrijwillig meedoen aan die registratie, en vrijwel alle Nederlandse academische en reguliere ziekenhuizen doen er tegenwoordig aan mee. Ze sturen dan jaarlijks een kopie van hun administratie in. In de resulterende database representeert elke rij één ziekenhuisontslag. Er zijn kolommen voor NAW-gegevens en kolommen voor informatie over de behandelingen, inclusief medische diagnoses (ICD-10 codes). Het CBS beheert een kopie van de LMR-database, en onderzoekers kunnen bij CBS onder bepaalde voorwaarden toegang krijgen tot (gedeeltes van) die data. In de nabije toekomst worden de LMR en een vergelijkbare registratie over ambulante zorg geïntegreerd tot het de Landelijke Basisregistratie Ziekenhuiszorg (LBZ).

> Bij het in praktijk nastreven van onherleidbaarheid moet rekening wordt gehouden met verfijningen op het model van *k*-anonimiteit

Wat nu als een beleidsonderzoeker geen arts is, d.w.z., niet wettelijk gebonden is aan het medische beroepsgeheim? In dat geval is zal de data anoniem moeten worden gemaakt alvorens de beleidsonderzoeker er toegang toe krijgt. Het is evident dat daartoe, zoals eerder genoemd, persoonsgebonden nummers, namen en huisnummers uit de data moeten worden verwijderd. Afhankelijk van de (beleids)onderzoeksvragen zal bepaalde demografische informatie echter beschikbaar moeten blijven, zoals enige indicatie van geslacht, geboortedatum en postcodegebied.

Hoe weten we of de overblijvende demografische informatie niet leidt tot heridentificeerbaarheid? Een mogelijk antwoord hierop is: inventariseer het aantal mogelijke combinaties en tel per combinatie het aantal personen dat dezelfde informatie deelt. Deze personen zitten dan in dezelfde 'anonimiteitsgroep'. In geval van geslacht, geboortedatum en postcode is die telling in Nederland triviaal uitvoerbaar: die gegevens staan in de GBAs. En met dank aan de driejaarlijkse GBA-audit, waarbij steekproefsgewijs de juistheid van GBA-gegevens wordt gecontroleerd, neemt de kwaliteit van GBA-gegevens in principe toe.

Nederland

Recent is de empirische studie van Sweeney in Nederland herhaald met GBA-gegevens [2]. Van 16 gemeenten (zie Tabel 1) is de lijst van geboortedatum, geslacht en postcode van alle

Gemeente	Aantal inwoners
Amsterdam	766656
Rotterdam	591046
Den Haag	487582
Utrecht	305845
Nijmegen	161882
Enschede	156761
Arnhem	147091
Overbetuwe	45548
Geldermalsen	26097
Diemen	24679
Reimerswaal	21457
Enkhuizen	18158
Simpelveld	11019
Millingen a/d Rijn	5915
Terschelling	4751

Tabel 1: onderzoekspopulatie

inwoners verzameld. In totaal omvatte de onderzoekspopulatie 2.7 miljoen burgers.

Van diverse (deel)combinaties van deze GBA-kolommen is geteld hoe vaak dezelfde waarden voor kwamen; dit leverde inzicht in de groottes van anonimiteitsgroepen. Tabel 2a geeft per (deel) combinaties (*Quasi-identifier)* weer:

1. Het aantal verschillende waarden dat in de data voorkomt en een eigen anonimiteitsgroep vormt (*# of sets*)
2. De omvang van de kleinste (anonimiteits)groep (*Min.*)
3. De groepsgrootte waar de 25% kleinste groepen onder vallen (*1st Qu.*; eerste kwartiel)
4. De groepsgrootte waar de 50% kleinste groepen onder vallen (*Median*; tweede kwartiel)
5. De groepsgrootte waar de 75% kleinste groepen onder vallen (*3rd Qu.*; derde kwartiel)
6. De gemiddelde groepsgrootte (*Mean*)
7. De omvang van de grootste groep (*Max.*)

De (deel)combinaties bestaan uit:

- geboortedatum (*DoB*, bijvoorbeeld '1 januari 1970')
- geboortemaand (*MoB*, bijvoorbeeld 'januari')
- geboortejaar (*YoB*, bijvoorbeeld '1970')
- geslacht (*gender*, in de onderzoeksdata uitsluitend 'M' of 'V')
- viercijferige postcode (*PC4,* bijvoorbeeld '1098')
- volledige postcode (*PC6*, bijvoorbeeld '1098 XH')
- plaatsnaam (*town,* bijvoorbeeld 'Amsterdam')
- gemeentenaam (*municipality*, bijvoorbeeld 'Overbetuwe')

Met behulp van kwartielen kunnen uitspraken worden gedaan als: "75% van de anonimiteitsgroepen bestaat uit X of minder personen".

In Tabel 2a zien we bijvoorbeeld bij quasi-identifier *PC6* (dus: alleen de volledige postcode) dat de *Median*-waarde 35 is, dat

Quasi-identifier:	# of sets	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PC4	388	2	3,278	7,090	7,188	10,300	22,330
PC6	66,883	1	24	35	41	50	1,322
PC4+DoB	2,267,700	1	1	1	1	1	42
PC6+DoB	2,759,422	1	1	1	1	1	5
PC4+gender	776	1	1,652	3,536	3,594	5,151	11,730
PC6+gender	133,012	1	11	18	21	25	954
gender+YoB	221	1	5,219	14,570	12,550	19,740	25,580
gender+YoB+MoB	2,699	1	397	1,177	1,028	1,594	2,326
gender+YoB+MoB+PC4	635,679	1	2	3	4	6	40
gender+YoB+MoB+municipality	34,790	1	6	18	80	96	733
gender+DoB	71,318	1	21	40	39	54	571
gender+DoB+PC4	2,488,828	1	1	1	1	1	22
gender+DoB+PC6	2,766,475	1	1	1	1	1	4
town+gender	134	1	222	1116	20,700	3259	347,100
town+YoB	5,642	1	6	29	492	101	14,270
town+YoB+MoB	49,207	1	2	5	56	20	1,262
town+DoB	463,134	1	1	2	6	7	419
town+YoB+gender	10,492	1	4	17	264	60	7,515
town+YoB+MoB+gender	83,172	1	1	3	33	14	695
town+DoB+gender	697,875	1	1	2	4	5	226

Tabel 2a: per quasi-identifier: aantal groepen en verdeling van groepsgroottes

wil zeggen: 50% van de anonimiteitsgroepen bestaat uit 35 of minder personen. De helft van de 66,883 (*# of sets*) verschillende postcodes is dus zodanig identificerend dat gedeïdentificeerde data dat die postcodes bevat herleidbaar is tot 35 of minder natuurlijke personen. Ook zien we dat de grootste anonimiteitsgroep aanzienlijk groter is: 1322 personen (een postcode in Amsterdam). En de kleinste anonimiteitsgroep is aanzienlijk kleiner: 1 persoon. Voorbeeld van een correct geformuleerde claim over anonimiteit: de anonimiteit van personen (subjecten) ten aanzien van records in een database (voorwerpen van belang) die quasi-identifier *PC6* bevat tegenover iemand die beschikt over een (voldoende volledige en nauwkeurige) database met postcodes en persoonsnamen (aanvaller) loopt uiteen van 1-anonimiteit tot 1322-anonimiteit.

De tabel toont van enkele quasi-identifiers een *Min.*-, *1st Qu.*-, *Median*- en *3rd Qu.*-waarde van 1. Dat betekent dat 75% van de anonimiteitsgroepen omvang 1 heeft; anders gezegd, dat 75% van de waarden (*# of sets*) voor die quasi-identifier een natuurlijke persoon ondubbelzinnig identificeert.

Tabel 2b toont voor dezelfde quasi-identifiers het volgende:

1. Het aantal personen in groep van grootte 1 (*k =1,* dus ondubbelzinnig identificeerbaar)

Quasi-identifier:	$k = 1$	$k \leq 5$	$k \leq 10$	$k \leq 50$	$k \leq 100$
PC4	0	9	19	345	996
PC6	429	6,109	25,103	1,459,939	2,354,255
PC4+DoB	1,861,081	2,754,465	2,765,932	2,774,476	-
PC6+DoB	2,744,653	2,774,476	-	-	-
PC4+gender	4	27	103	889	2,555
PC6+gender	1,854	31,262	184,803	2,342,242	2,629,017
gender+YoB	5	14	53	250	516
gender+YoB+MoB	55	356	712	4,478	9,674
gender+YoB+MoB+PC4	137,035	279,100	2,196,950	2,774,476	-
gender+YoB+MoB+municipality	2,186	22,565	59,597	244,152	619,671
gender+DoB	2,014	14,506	40,322	1,392,622	2,725,472
gender+DoB+PC4	2,240,461	2,765,067	2,772,205	2,774,476	-
gender+DoB+PC6	2,758,578	2,774,476	-	-	-
town+gender	4	4	28	372	896
town+YoB	499	3,172	7,225	50,985	103,145
town+YoB+MoB	10,083	61,073	112,850	287,173	394,844
town+DoB	185,042	596,769	1,045,559	2,730,668	2,750,700
town+YoB+gender	1,153	7,195	16,333	102,018	150,135
town+YoB+MoB+gender	22,260	109,126	170,351	398,601	826,744
town+DoB+gender	288,409	1,029,601	1,813,559	2,750,669	2,764,050

Tabel 2b: per quasi-identifier: aantal personen in groepsgrootte k

2. Het aantal personen in groep van grootte 1 t/m 5 ($k \leq 5$)
3. Het aantal personen in groep van grootte 1 t/m 10 ($k \leq 10$)
4. Het aantal personen in groep van grootte 1 t/m 50 ($k \leq 50$)
5. Het aantal personen in groep van grootte 1 t/m 100 ($k \leq 100$)

Nota bene: de maat $k=1$, ondubbelzinnige identificeerbaarheid, is niet de enige maat die relevant is bij analyse van anonimiteit. Afhankelijk van aanwezigheid van andere informatie kan een anonimiteitsgroep verder worden gereduceerd. Een anonimiteitsgroep kan daarom ook het best zo groot mogelijk worden gemaakt. Een minimale waarde van k kan, afhankelijk van de aan- en afwezigheid van andere informatie, een bruikbare norm zijn. Om die reden zijn grotere groepsgroottes meegenomen in tabel 2b.

Amsterdam versus Terschelling

Het volgende illustreert hoe dezelfde quasi-identifier bij alle burgers in de onderzoekspopulatie ongeveer even identificerend is. Daarna zullen we een voorbeeld geven waarbij de mate van identificerendheid juist sterk verschilt.

Dit is een voorbeeld van een quasi-identifier in de Landelijke Medische Registratie:

QID_A = *geslacht + geboortejaar + geboortemaand + 4-cijferige postcode* (in tabel 2a/2b: gender + YoB + MoB + PC4)

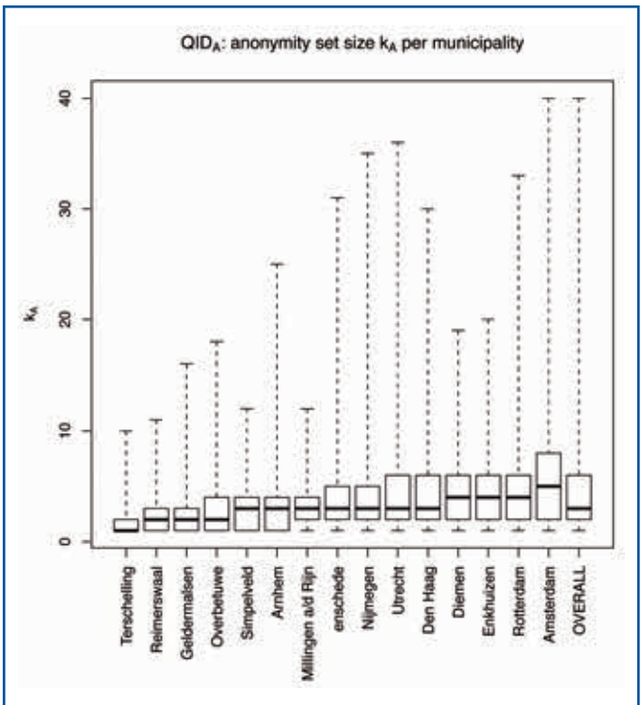
> In deze conceptie heeft een claim over anonimiteit alleen betekenis als inhoud is gegeven aan subject, belang en aanvaller

Het tellen van de groottes van anonimiteitsgroepen en het aantal groepen per grootte levert de uitkomsten op die zichtbaar zijn in Figuur 2a.

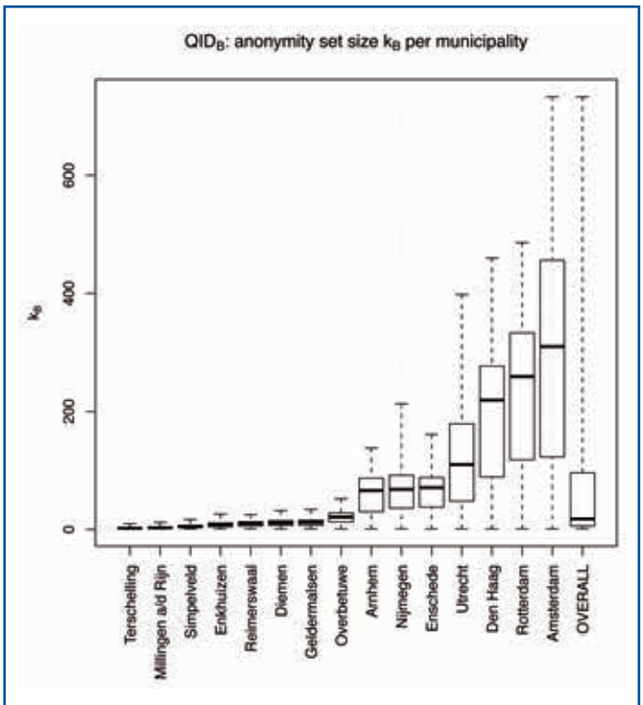
In Figuur 2a is per gemeente weergegeven wat de omvang is van de kleinste en grootste anonimiteitsgroep (de uitlopers van de stippellijn); en in de verticale rechthoeken het eerste, tweede (dikgedrukt horizontaal streepje) en derde kwartiel. We zien dat de anonimiteit van inwoners van Terschelling en inwoners van Amsterdam min of meer gelijk is: allen zitten in groepen van $k\_a < 10$, dat wil zeggen: indien een kwaadwillende beschikt over een bepaald gedeelte van de LMR en daarnaast beschikt over (toegang tot) de volledige GBA-administratie van deze gemeenten, dan kan deze de LMR-records herleiden tot 10 of minder personen.

> Inventariseer het aantal mogelijke combinaties en tel per combinatie het aantal personen dat dezelfde informatie deelt

Ter illustratie is bij de studie ook een andere database bekeken: de Bijstands Fraude Statistiek (BFS) [4]. Die database bevat informatie over onderzoek van gemeenten naar bijstandsfraude. Elk record komt overeen met één afgerond onderzoek. Er is een kolom die aangeeft of er wel of geen sprake bleek te zijn van fraude, en een andere kolom die in het eerste geval het bedrag bevat dat gemoeid is met de fraude. Uit de BFS-database is de volgende quasi-identifier gekozen:



Figuur 2a: grootte kA van anonimiteitsgroepen bij QIDA (LMR)



Figuur 2b: grootte kB van anonimiteitsgroepen bij QIDB (BFS)

QID_B = *geslacht + geboortejaar + geboortemaand + gemeentenaam*
(in tabel 2a/2b: gender + YoB + MoB + municipality)

De resultaten van die telling staan in Figuur 2b. In die figuur wordt duidelijk dat de groottes van anonimiteitsgroepen per gemeente wél (aanzienlijk) kunnen verschillen: inwoners van Terschelling zijn aanzienlijk makkelijker te herleiden dan inwoners van Amsterdam. Op basis van deze verschillen zou er bij het ontwerp van bijvoorbeeld een enquête voor kunnen worden gekozen om bij onderzoek onder inwoners van Terschelling omwille van privacybescherming bepaalde informatie *niet* te vragen die in het algemeen wél 'privacyveilig' zou kunnen worden gevraagd van inwoners van Amsterdam. Dit soort grafieken heeft ook gebruikswaarde voor het individu: die kan er beter geïnformeerde keuzes mee maken over het wel/niet prijsgeven van bepaalde informatie.

> Je kunt bij een enquête kiezen om bij onderzoek bepaalde informatie *niet* te vragen op Terschelling en wél in Amsterdam

Beleid

De Stichting Perinatale Registratie Nederland hanteert de volgende richtlijn ten aanzien van niet-herleidbaarheid [5]:

2.1 Basisvereisten

Bestanden zijn 'niet-herleidbaar' wanneer tegelijkertijd tenminste aan alle van de volgende vereisten is voldaan:

a) De geboortedatum van de vrouw ontbreekt: alleen de leeftijd is opgenomen.
b) De geboortedatum van het kind ontbreekt: alleen het jaar van geboorte, alsmede de weekdag waarop de geboorte plaats vond is opgenomen.
c) Van de postcode zijn alleen de eerste drie cijfers opgenomen.

> *Privacy by Design* impliceert dat ontwerpbeslissingen zijn geïnformeerd met privacyargumenten

2.2 Aanvullende vereisten

Naast bovengenoemde basisvereisten dienen voor 'niet-herleidbaarheid' tevens de volgende bewerkingen uitgevoerd te worden:

a) Exacte waarden dienen zoveel mogelijk vervangen te worden door categorieën: zo dient bij tijdstippen het exacte moment vervangen te worden door een interval waarin dit moment valt, en bij geboortegewicht het exacte gewicht in grammen door een gewichtscategorie van bv. 100 gram.
b) Voor relatief zeldzaam voorkomende waarden zal specifieke 'niet-herleidbaarheid' toegepast moeten worden.

> Het wordt niet alleen duidelijk welke gegevens in herleidbaarheid resulteren, maar ook welke gegevens *niet*

Bij het maken van dergelijk beleid zijn de kwantificaties zoals hierboven besproken van praktisch nut: het wordt niet alleen duidelijk welke combinatie van gegevens in herleidbaarheid resulteert, maar ook duidelijk welke combinatie van gegevens *niet* in herleidbaarheid resulteren. Zo kan besloten worden om bepaalde gegevens, mits deze van voldoende belang zijn voor de (beleids)onderzoeksdoelen, tóch te verwerken, waar deze anders vanwege twijfel niet zouden zijn verwerkt. (Dit is overigens niet bedoeld als pleidooi voor bovenmatige gegevensverwerking; verwerking van gegevens moet minimaal zijn en bovendien altijd noodzakelijk voor het doel van de gegevensverzameling; zeker bij mogelijk identificeerbare gegevens is dit een vereiste vanuit Wbp.)

Conclusie

Privacy by Design impliceert ontwerpbeslissingen die (gedeeltelijk) zijn geïnformeerd met privacyargumenten. Met kwantificatie van anonimiteit kan de maker van een enquête de ontwerpbeslissing om bepaalde vragen wel/niet te stellen, of om de vraagstelling om privacyredenen af te stemmen op verschillende populaties, beter beargumenteren. Een aanvullend idee is om kwantificatie na verwerking te gebruiken om bepaalde gegevens alsnog te onderdrukken of generaliseren. Idealiter zouden burgers zelfs wellicht zélf kunnen berekenen hoe identificeerbaar ze zijn op basis van door hen verstrekte gegevens.

Dankwoord

Mijn dank gaat uit naar Guido van 't Noordende (UvA) voor het reviewen van een conceptversie van dit artikel.

Verwijzingen

[1] *Andreas Pfitzmann en Marit Hansen: "A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management", documentversie 0.34, 2010 (.pdf) http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf*

[2] *Matthijs R. Koot, Guido van 't Noordende en Cees de Laat: "A Study on the Re-Identifiability of Dutch Citizens", 2010 (.pdf): http://dare.uva.nl/document/204557*

[3] *Landelijke Medische Registratie http://www.dutchhospitaldata.nl/Registraties/LMR.php*

[4] *Bijstandsfraudestatistiek http://www.cbs.nl/nl-NL/menu/themas/arbeid-sociale-zekerheid/methoden/dataverzameling/korte-onderzoeksbeschrijvingen/bijstandsfraudestatistiek.htm*

[5] *Stichting Perinatale Registratie Nederland: "Regels 'niet-herleidbaarheid'", 2011 (.docx) http://www.perinatreg.nl/uploads/76/143/Regels_niet-herleidbaarheid_versie_1_0.docx*