*Dr. ir. H.J.W. van Heerde**

Vervagende persoonsgegevens: een levensloopregeling voor privacygevoelige gegevens

246

Trefwoorden:
beperken bewaartermijnen, geleidelijk vervagen, meetnetwerken

Het is in de praktijk moeilijk, zo niet onmogelijk, te voorkomen dat privacygevoelige gegevens op straat komen te liggen. Een manier om de nadelige gevolgen van privacyschending te beperken, is het tijdig verwijderen van gegevens. Het gaat daarbij niet alleen om het verwijderen van direct identificeerbare persoonsgegevens, maar ook om indirect herleidbare, privacygevoelige persoonsgegevens. Voor het verwijderen van persoonsgegevens is weinig aandacht. Zelden bevat een IT-opdracht specifieke vereisten voor tijdige verwijdering. Opdrachtgevers zien het nut van tijdig weggooien niet in en soms worden andere wettelijke bewaartermijnen (bijvoorbeeld in verband met belastingwetgeving) als excuus opgevoerd om alles te bewaren.

In dit artikel wordt ingegaan op de verschillende mogelijkheden om persoonsgegevens te bewaren, bijvoorbeeld in minder privacygevoelige varianten met behulp van bepaalde technieken. Geadviseerd wordt een levensloopregeling voor persoonsgegevens op te stellen waarin aangeven wordt hoe en wanneer gegevens 'vervaagd' moeten worden. Nieuwe types databasesystemen kunnen dergelijke levensloopregelingen autonoom en automatisch uitvoeren, zodat de oorspronkelijke individuen na verloop van tijd niet meer te achterhalen zijn.

1 Inleiding

Bij gegevens die beschermd moeten worden tegen privacyinbreuk, denkt men vaak aan persoonlijke, direct identificeerbare, gegevens die op internet verzameld en achtergelaten worden. Bedrijven als Google, Facebook en Apple worden steeds vaker in verband gebracht met aantasting van privacy. Ook gaat er steeds meer belangstelling uit naar grootschalige projecten als de OV-chipkaart, het elektronisch patiëntendossier en rekeningrijden. Bij deze diensten en projecten speelt privacy een direct aanwijsbare rol. Het lastige bij dit soort toepassingen is om het concrete, direct kwantificeerbare belang van dataverzameling en -opslag af te wegen tegen het wettelijk vastgelegde maar moeilijk te definiëren privacybelang van de gebruikers. Drie problemen spelen hierbij een rol. Ten eerste, omdat de kosten (in termen van het risico op privacyschending) zo onduidelijk zijn, wordt het privacybelang ten onrechte vaak niet afgewogen bij de beslissing welke direct of indirect herleidbare persoonsgegevens verzameld en opgeslagen mogen worden, en voor hoe lang. Ten tweede wordt veelal onterecht vertrouwd op het idee dat persoonsgegevens volledig zijn te beschermen tegen ongeautoriseerd gebruik door het toepassen van goede beveiligingstechnieken. Ten derde is het doel waarvoor persoonsgegevens worden verzameld vaak onduidelijk of op zijn minst niet goed afgebakend, waardoor bedrijven denken dat het maar beter is om gegevens zo lang mogelijk op te slaan, in de hoop er in de toekomst nog wat aan te kunnen hebben.

2 Sensoriseringstrend als privacyprobleem

Het probleem wordt groter als we verder kijken dan de genoemde, vaak aangehaalde voorbeelden van privacyschendende instellingen en projecten. In Nederland verschijnen steeds meer sensoren. Het meten van waterkwaliteit, verkeersintensiteit, luchtkwaliteit, et cetera, neemt een steeds belangrijkere rol in om de verschillende fysische systemen te kunnen doorgronden. Meetgegevens kunnen worden gebruikt om modellen te kalibreren, zodat betere voorstellingen kunnen worden gegeven, bijvoorbeeld van waar en wanneer sprake kan zijn van wateroverlast. Daarnaast kunnen de meetgegevens worden gebruikt om beleid te toetsen en nieuw beleid te maken. De meetnetwerken die hiervoor worden gebruikt, worden steeds fijnmaziger. Om een voorbeeld te geven: om in een gescheiden rioolstelsel te kunnen controleren of afvalwater illegaal gemengd wordt met hemelwater, wordt de temperatuur van het rioolwater gemeten.[1] Dit gebeurt op huisaansluitingsniveau, wat als privacygevoelig bijeffect heeft dat in theorie bepaald zou kunnen worden hoeveel leden een huishouden telt en welk leefpatroon die volgen, zodat het niet moeilijk is vast te stellen wanneer een gezin op vakantie is.

Bij het voorbeeld van rioleringsmetingen is onduidelijk hoe groot het gevaar voor privacyschending nu eigenlijk is. Vrijwel iedereen zal van mening zijn dat het kunnen

* **Harold van Heerde is adviseur informatiemanagement bij Witteveen+Bos en recentelijk gepromoveerd aan de Universiteit Twente op het gebied van privacy-aware datamanagement. Dit artikel is gebaseerd op onderzoek uitgevoerd aan de Universiteit Twente in samenwerking met INRIA Roquencourt, Frankrijk.**

1 O.A.C. Hoes, R.P.S. Schilperoort, W.M.J. Luxemburg, F.H.L.R. Clemens & N.C. van de Giesen, 'Locating illicit connections in storm water sewers using fiber-optic distributed temperature sensing', *Water Research*, Volume 43, Issue 20, December 2009.

ontdekken van illegale lozingen zwaarder weegt dan het gevaar dat een rioolonderzoeker ontdekt dat een gezin uit drie personen bestaat, en dat een gezinslid op zeker moment naar het toilet is gegaan. Maar de afweging blijft vaak volledig uit. Gegevens worden vaak voor onbepaalde tijd opgeslagen zonder dat enig potentieel privacygevaar wordt onderkend, ervan uitgaand dat de gegevens in vertrouwde handen zijn en niet gestolen zullen worden, en in de hoop dat de historische gegevens in de toekomst nog meer waarde op zullen leveren. Als deze grote hoeveelheden gegevens om wat voor reden dan ook op een later tijdstip toch openbaar gemaakt worden, kan dat tot gevolg hebben dat alsnog inbreuk wordt gemaakt op de persoonlijke levenssfeer van mensen. Niet alleen de rioolonderzoeker krijgt dan de voor hem nietszeggende gegevens onder ogen; vrienden én vijanden kunnen wellicht wél interessante informatie herleiden uit deze gegevens. Zo zal het dievengilde buitengewoon geïnteresseerd zijn in nauwkeurige voorspellingen van het afwezigheidpatroon van een gezin, of kan een alimentatieplichtige zijn of haar voordeel halen uit het feit dat het huishouden van zijn of haar ex niet uit één, maar uit twee personen blijkt te bestaan.

Er is een tweede belangrijke factor die vaak wordt onderschat, en dat is de mogelijkheid tot het combineren van verschillende gegevensbronnen. Een meetproject wordt niet vaak afzonderlijk uitgevoerd. Vaak voert een gemeente meerdere, al dan niet op zichzelf staande projecten uit, of begeleidt een adviesbureau meerdere projecten waardoor het indirect toegang krijgt tot verschillende gegevensbronnen. Uitgaande van en vertrouwend op de integriteit van het gemiddelde adviesbureau kan dit op zichzelf geen kwaad; een garantie dat gecombineerde gegevensbronnen nooit op straat komen te liggen kan echter nimmer gegeven worden. Niet alleen juristen en politici laten onbeveiligde USB-sticks of laptops achter in een trein; ook ingenieurs kunnen dergelijke fouten maken.

Als de huidige 'sensoriseringstrend' doorzet, wordt de hoeveelheid indirect herleidbare persoonsgegevens, zoals de resultaten van temperatuurmetingen in riolen, alsmaar groter. De hoeveelheid wordt niet alleen groter, de veelzeggendheid van de gegevens zal ook toenemen. Zo wordt in de Verenigde Staten reeds geëxperimenteerd met zogeheten labs-on-a-chip waarmee individuele stoffen in het rioolwater gedetecteerd kunnen worden. Hiermee is het niet alleen mogelijk om drugs op te sporen, het is ook mogelijk vast te stellen dat een vrouw zwanger is, reeds voordat de vrouw in kwestie zélf weet dat ze zwanger is.

3 Gegevensopslag beperken

Ten onrechte heerst de gedachte dat gegevens afdoende beschermd kunnen worden tegen oneigenlijk gebruik ervan, en dat kan worden voorkomen dat gegevens op straat komen te liggen. De praktijk wijst uit dat er geen beveiligingsmethode bestaat die én garandeert dat gegevens niet in handen van ongeautoriseerde gebruikers komen én de bruikbaarheid van de gegevens niet zodanig aantast, dat ze nutteloos worden. Bovendien, een kluis kan nog zo kraakbestendig zijn, als de sleutel 'voor het gemak' in het slot gelaten wordt, is geen enkele kluis veilig. Onachtzaamheid en fouten zijn menselijk; privacybeleid kan helpen deze te voorkomen maar kan ze niet uitsluiten. Controles achteraf kunnen privacyinbreuken aan het licht brengen, maar dan is het kwaad reeds geschied.

Nadat ruwe gegevensverzamelingen, zoals gegevens afkomstig uit meetnetwerken, eenmaal op straat zijn komen te liggen, kunnen ze nauwkeurig geanalyseerd, verrijkt, en inzichtelijk gemaakt worden via websites, doorzoekbaar voor iedereen die interesse heeft. Als dit eenmaal is gebeurd, is er geen weg meer terug en is het kwaad geschied. Alle e-mails verzonden en ontvangen door werknemers van het bedrijf Enron, die in 2004 openbaar zijn gemaakt zijn door het Ministerie van Justitie in de Verenigde Staten naar aanleiding van een grootschalig fraudeonderzoek, zijn verzameld en doorzoekbaar gemaakt.[2] Vlak nadat America Online de bijzonder slecht geanonimiseerde zoektermgeschiedenis van haar gebruikers openbaar maakte, werden allerlei websites opgericht waarmee de gegevens eenvoudig konden worden doorzocht.[3] Onbeschermde, weliswaar min of meer vrijwillig opengestelde, profielinformatie van 100 miljoen Facebookgebruikers is verzameld en gebundeld, en verspreid over het internet zodat deze makkelijk doorzoekbaar is.[4] Weliswaar gaat dit om vrijwillig, door gebruikers zelf gepubliceerde en vrij toegankelijke gegevens, maar de gebruikers hebben geen mogelijkheid meer zelf controle uit te oefenen over de gegevens, bijvoorbeeld deze te verwijderen. Er is geen reden om aan te nemen dat het verrijken en makkelijk doorzoekbaar maken niet ook zal gebeuren op het moment dat interessante gegevens uit meetnetwerken openbaar worden gemaakt.

Om de gevolgen van mogelijke latere openbaarmaking te verzachten, is het beperken van zowel de hoeveelheid als de gedetailleerdheid van de gegevens noodzakelijk. Het beperken van de duur van de opslag van gegevens is verankerd in de Europese richtlijn en in de Wet bescherming persoonsgegevens. In verscheidene academische onderzoeken wordt getracht oplossingen te vinden om gegevens automatisch te laten verwijderen uit gegevensbanken. Een van de grootste problemen is om te bepalen hoelang gegevens bewaard mogen worden. Vanwege dat probleem wordt in de praktijk vaak een oneindige bewaarperiode gehanteerd. Om die reden moet niet alleen het doel waarvoor de gegevens bewaard worden maatgevend zijn, ook het risico op (toekomstige)

2 Zie <www.cs.cmu.edu/~enron/>.
3 Zie <http://elliottback.com/wp/aol-search-data-tools-list/>.
4 Zie <http://headlines.nos.nl/forum.php/list_messages/21092>.

privacyschending moet een rol gaan spelen. Ook al is het lastig te bepalen hoe groot dat risico is, gesteld kan wel worden dat het risico toeneemt naarmate gegevens langer worden opgeslagen en de gegevensbanken groter worden. Als we daarnaast aannemen dat gegevens minder waard worden naarmate ze ouder worden, dan kunnen we stellen dat er een optimale bewaarperiode gevonden kan worden, waarbij zowel het risico op privacyschending als het nut van het bewaren van gegevens is afgewogen.

4 Geleidelijk vervagen van gegevens

Een tekortkoming van beperkte bewaartermijnen is dat het gaat om alles of niets. Als gegevens worden verwijderd, dan worden de volledige records die privacygevoelige informatie bevatten verwijderd, inclusief gegevens die wellicht wel nuttig zijn, maar minder privacygevoelig. Dit terwijl het geen probleem hoeft te zijn om alleen die gegevens te bewaren die minder privacygevoelig zijn. Een voorbeeld. Stel een meting bevat een tijdstip, een locatie, en bijbehorende meetwaarden zoals temperatuur. Het tijdstip van de meting wordt hier als meest privacygevoelig beschouwd en moet, om risico's niet te groot te laten worden, na een maand verwijderd zijn. Het volstaat om alleen het tijdstip uit de record te verwijderen, en de rest van de informatie te behouden.

We kunnen nog een stap verdergaan, door zogenoemde attributen (zoals tijdstip) niet in een stap volledig te verwijderen, maar geleidelijk in meerdere stappen te vervagen. Met vervagen wordt bedoeld dat de precisie waarmee de gegevens opgeslagen zijn afneemt. Zo kan een datum met tijdstip vervangen worden door de datum met alleen een uuraanduiding, vervolgens met alleen de jaar-, maand - en dagaanduiding, om vervolgens alleen het jaar waarin de meting gedaan is over te houden. Met elke vervagingsstap neemt zowel de bruikbaarheid als de privacygevoeligheid van de gegevens af. Indien echter de privacygevoeligheid sterker afneemt dan de bruikbaarheid, dan valideert dit een langere bewaarperiode dan mogelijk zou zijn als de gegevens in één stap verwijderd zouden kunnen worden.

Het concept van het vervagen van privacygevoelige gegevens is vergelijkbaar met de werking van het menselijk brein. Details van gebeurtenissen in het verleden worden langzaam vergeten, waardoor de scherpe randjes van de gebeurtenissen ervan afgaan. In de natuur komt dit vaker voor: details van voetafdrukken in het zand gaan langzaam verloren doordat weer en wind er vat op krijgen. Computers, gegevensbanken en internet in het algemeen zijn niet ontworpen om gegevens te verwijderen of te laten vervagen, maar juist om deze te bewaren. Zonder ingrijpen ontstaat er zodoende een oneindig groot geheugen, waarin details uit het verleden telkens weer te traceren zijn. Het is echter technisch mogelijk om het ontwerp van gegevensbanken aan te passen zodat de natuur nagebootst kan worden, met als bijkomend voordeel dat we zelf exact kunnen bepalen hoe en wanneer gegevens moeten vervagen.

5 Levensloopregeling van gegevens

Om gegevens te kunnen vervagen is het nodig om vast te leggen:

1. *hoe* gegevens moeten worden vervaagd; en
2. *wanneer* gegevens moeten worden vervaagd.

Deze informatie wordt gebundeld in een zogenoemde *levensloopregeling* die gekoppeld wordt aan elke record dat wordt toegevoegd aan de database. Het databasesysteem moet vervolgens autonoom in staat zijn deze levensloopregeling uit te voeren.

Zoals eerder vermeld kan een record van meetgegevens worden opgesplitst in verschillende attributen, zoals tijdstip, locatie, temperatuur, *etcetera*. Voor elk type attribuut kan een zogenoemde generalisatiehiërarchie worden vastgelegd. Een generalisatiehiërarchie schrijft voor dat bijvoorbeeld een tijdstip in uren een *generalisatie* is van een tijdstip in minuten, en dat een provincie een gegeneraliseerde vorm van stad is. Voor sommige attributen, zoals tijd, is het mogelijk om deze generalisatie automatisch af te leiden. Voor andere attributen, zoals locatie, is een *generalisatieboom* nodig, waarin voor elke mogelijke waarde vastgelegd staat wat de gegeneraliseerde waarde moet zijn. Zo zal ergens beschreven moeten staan dat Enschede, Zwolle en Deventer gegeneraliseerd moeten worden naar Overijssel, Apeldoorn en Arnhem naar Gelderland, etcetera. Vervolgens staat er ook vastgelegd dat Overijssel en Gelderland worden gegeneraliseerd naar Nederland, Vlaanderen en Wallonië naar België.

Een levensloopregeling bevat informatie over wanneer welke stap in een generalisatieboom uitgevoerd moet worden: wanneer een attribuut gegeneraliseerd moet worden van stad naar provincie en wanneer van provincie naar land. Een levensloopregeling moet daarom het logische gevolg zijn van de afweging die gemaakt moet worden tussen de privacy van gebruikers en de waarde van gegevens voor het bedrijf of instelling. Idealiter zou elke gegevensverzamelende instelling een levensloopregeling moeten opstellen en de gebruikers daarover informeren.

Het verschil met bestaand privacybeleid en bestaande overeenkomsten is, dat zolang de instelling te vertrouwen is en deze de regeling ook daadwerkelijk uitvoert, gegevens ook echt niet meer op straat kunnen komen te liggen op het moment dat de instelling *niet* meer kan voldoen aan de regeling. Ook al zou de instelling haar beleid aanpassen, ze zal niet meer in staat zijn om de privacy van haar gebruikers te schenden op basis van persoonsgegevens die in het verleden zijn verzameld.

6 Technische mogelijkheden

Wil een nieuwe privacytechniek, zoals het geleidelijk vervagen van gegevens, in de praktijk slagen, dan mag de techniek om het mogelijk te maken niet te veel kosten. Echter, huidige gegevensbanken zijn ontworpen om grote hoeveelheden gegevens duurzaam op te kunnen slaan en te ontsluiten, niet om ze te verwijderen. Bovendien worden er juist back-ups gemaakt van de gegevens

om de gegevens niet verloren te laten gaan. Daarnaast is zelfs het verwijderen van gegevens van een harde schijf niet triviaal. Het verwijderen van een bestand betekent in de praktijk niet zelden dat het bestand slechts virtueel verwijderd is, en daardoor eenvoudig is te herstellen. Het meest eenvoudige voorbeeld is het verwijderen van een bestand in het besturingssysteem Windows: een bestand verdwijnt eerst in de prullenbak, en is pas echt weg als deze prullenbak leeggemaakt wordt. Zelfs daarna is het mogelijk voor experts om het bestand te herstellen.

Uit onderzoek is gebleken dat het wel degelijk mogelijk is om gegevensbanksystemen zodanig aan te passen dat gegevens efficiënt en onherstelbaar te vervagen en te verwijderen zijn. Mede door het gebruik van versleuteling kunnen zelfs gegevens die in back-ups terechtkomen, onherleidbaar worden gemaakt. Door gegevens te versleutelen alvorens ze in een back-up terechtkomen, kunnen deze gegevens alleen teruggehaald worden zolang de sleutel beschikbaar is. Als privacygevoelige gegevens na een bepaalde tijd verwijderd moeten worden, dan volstaat het verwijderen van de sleutel. Door meerdere sleutels te beheren per tijdsinterval, attribuut en gewenste nauwkeurigheid, kan het gehele vervagingsproces efficiënt worden geïmplementeerd. Versleuteling is echter niet voor alle componenten van gegevensbeheer een geschikte oplossing. Daarom zijn meer technieken nodig.

Een databasesysteem bestaat, naast een back-upmechanisme dat moet garanderen dat gegevens niet verloren gaan, grofweg uit twee belangrijke componenten:

1. de *opslagstructuur* waarmee gegevens opgeslagen worden op de harde schijf; en
2. *indices* waarmee gegevens sneller opgezocht kunnen worden (vergelijk met de index in een boek).

Opslagstructuren en indices zijn zodanig ontworpen dat gegevens zo snel mogelijk opgeslagen kunnen worden en zo snel mogelijk doorzocht kunnen worden. De snelheid wordt grofweg bepaald door het aantal bewegingen dat de harde schijf, waarop de gegevens opgeslagen zijn, moet maken om alle gevraagde gegevens te kunnen lezen en schrijven. Een opslagstructuur bepaalt hoe de gegevens zijn verspreid op de harde schijf. Vaak moet er, afhankelijk van het gebruik van het databasesysteem, een afweging gemaakt worden tussen snelheid van opslag en snelheid van het doorzoeken. In de context van gegevensvervaging komt er een derde ontwerpkeuze bij: gegevens moeten zo snel mogelijk aangepast en/of verwijderd kunnen worden. Bij het ontwerpen van een opslagstructuur waar gegevens snel uit verwijderd of vervaagd kunnen worden, kan gebruik worden gemaakt van het feit dat gegevens die tegelijkertijd of binnen een kort tijdsbestek toegevoegd worden, ook op hetzelfde tijdstip worden verwijderd of vervaagd. Dit pleit ervoor om gegevens altijd gesorteerd op tijd, dicht bij elkaar op te slaan op de harde schijf. Dat heeft wel weer nadelige gevolgen voor de doorzoekbaarheid van gegevens.

7 Hoe nu verder?

Op dit moment zijn de technische mogelijkheden om gegevens te verwijderen of te vervagen nog maar in beperkte mate voorhanden. De implementatie in gegevensbanksystemen zal opgepakt moeten worden door de grote leveranciers. Maar de eerste en meest belangrijke stap die gezet moet worden om het beperkt bewaren van gegevens in de praktijk te brengen, is een omslag in de manier van denken over gegevensopslag en de doelen waarvoor gegevens verzameld worden. Instellingen of bedrijven die gegevens willen verzamelen, bijvoorbeeld door het aanleggen van zowel grootschalige als fijnmazige meetnetwerken, zullen zich meer bewust moeten zijn van de risico's die de langdurige opslag van direct en indirect herleidbare persoonsgegevens meebrengt. Voor advies- en ingenieursbureaus die bedrijven en instellingen begeleiden in het opzetten van meetnetwerken is hierbij een belangrijke rol weggelegd.

Al in een vroeg stadium moet bij het ontwerp van een meetnetwerk in kaart worden gebracht wat de doelen zijn waarvoor meetgegevens verzameld mogen en moeten worden en hoelang het nodig is om gegevens op te slaan om die doelen te kunnen bereiken. De vraag moet worden beantwoord *of en wanneer* oude gegevens zodanig weinig bijdragen aan het bereiken van een doel dat het risico op privacyschending niet langer in verhouding staat tot de waarde van de gegevens. Op basis daarvan kan worden bepaald hoelang en in welke vorm gegevens opgeslagen moeten worden. Daarbij moet ook worden gekeken naar in hoeverre bepaalde meetgegevens en bijbehorende attributen kunnen bijdragen aan het succesvol kunnen integreren van gegevens uit verschillende gegevensbronnen. Indien het risico op privacyschending te groot is, kan ervoor worden gekozen om bepaalde attributen (zoals tijd) te vervagen, zodat gegevensintegratie weliswaar zinvol, maar ook minder nauwkeurig en dus minder privacygevoelig wordt.

Het toepassen van gegevensvervaging en beperkte gegevensopslag hoeft niet direct te leiden tot verlies van waarde, of tot het niet kunnen halen van doelstellingen. Integendeel, juist doordat er een betere afweging gemaakt wordt tussen privacy en het oneindig lang bewaren van gegevens, is het risico op privacyschendingen kleiner, wat het draagvlak voor dergelijke netwerken vergroot. Hierdoor zal er minder weerstand zijn tegen het steeds verder 'versensoriseren' van Nederland.

8 Conclusie

In het begin van het industriële tijdperk werd op grote schaal schade toegebracht aan het milieu. De schadelijke gevolgen van het dumpen van afval in open water of het uitstoten van CO_2 waren nauwelijks bekend, waardoor het op de markt kunnen zetten van zo goedkoop mogelijke, aantrekkelijke, en welvaartvergrotende producten leidend was. Zolang er geen goede afweging gemaakt kan worden tussen de voor- en nadelen van massaproductie, simpelweg omdat de nadelen niet bekend zijn, zal er nauwelijks of geen weerstand bestaan tegen

goedkope producten. Hetzelfde geldt voor de opkomst van allerlei, op het oog gratis, diensten op het internet, maar ook voor het klakkeloos opzetten van grootschalige en fijnmazige meetnetwerken. Het feit dat privacy een moeilijk definieerbaar en kwantificeerbaar begrip is, zou niet mogen betekenen dat het concept volledig genegeerd kan worden.

In hun huidige rol wijzen ingenieursbureaus bedrijven en instellingen op de negatieve gevolgen van een bepaald productieproces op het milieu, maar er is een nieuwe rol voor hen weggelegd op het gebied van de bescherming van persoonsgegevens. Ze moeten in een vroeg stadium wijzen op de gevaren van gegevensopslag, en de noodzaak om een goede afweging te maken tussen de noodzaak van gegevens en het risico op privacyschending.

Er zijn nieuwe technieken nodig om het beginsel van beperking van de bewaartermijn van persoonsgegevens adequaat in de praktijk te kunnen toepassen. De alles-of-nietskeuze voor het volledig en in één stap verwijderen van gegevens leidt tot onnodig gegevensverlies. Geïnspireerd door natuurlijke processen kan het geleidelijk vervagen van gegevens uitkomst bieden, zodat het vergeten van gegevens niet hoeft te leiden tot inperking van de mogelijkheden die de gegevens kunnen bieden. Bedrijven en instellingen zouden daarom niet langer bang moeten zijn om afstand te nemen van de door hen soms onterecht waardevol geachte gegevens. Het is naïef om te denken dat gegevens onbeperkt lang bewaard kunnen worden, zonder dat deze gegevens uiteindelijk op straat komen te liggen. Om de impact van zo'n gebeurtenis te beperken is het tijdig verwijderen of vervagen van persoonsgegevens bittere noodzaak.