Testen van informatiesystemen en het gebruik van (geanonimiseerde) persoonsgegevens

Ir. Hilbrand Kikkers, Bert Nienhuis B.ICT en drs. Erik Rutkens RE

Ten behoeve van het testen van de juiste werking van informatiesystemen worden gegevens vaak in hun geheel vanuit de database van een bestaand informatiesysteem naar de database van het te testen systeem gekopieerd. Hierbij worden ook bestaande persoonsgegevens gebruikt. De Wet bescherming persoonsgegevens (Wbp) maakt onderscheid tussen 'normale' persoonsgegevens en bijzondere persoonsgegevens. Onder de bijzondere persoonsgegevens vallen alle gegevens die informatie kunnen geven over godsdienst, ras, politieke gezindheid, gezondheid, lidmaatschap van een vakvereniging of strafrechtelijke verleden van een persoon. Deze wet geeft aan dat, in het geval van bijzondere persoonsgegevens, bij voorkeur fictieve persoonsgegevens dienen te worden gebruikt. Om te voldoen aan de Wet bescherming persoonsgegevens hebben organisaties een aantal mogelijkheden. Verreweg de meest eenvoudige (en goedkoopste) is het anonimiseren van bestaande persoonsgegevens. Dit artikel gaat in op de risico's van het testen van de juiste werking van informatiesystemen met behulp van bestaande persoonsgegevens en legt uit hoe organisaties ten behoeve van testdoeleinden bestaande persoonsgegevens eenvoudig kunnen anonimiseren.

Inleiding

Consumenten en organisaties kunnen bankzaken tegenwoordig via internet regelen. Zorgverzekeraars bieden mogelijkheden om declaraties online in te dienen. Paspoorten en rijbewijzen kunnen bij gemeenten digitaal worden aangevraagd. De hiervoor genoemde diensten worden mogelijk gemaakt door informatiesystemen[1] die via internet worden ontsloten dan wel aangeboden. Voordat een dienst, zoals internetbankieren, daadwerkelijk in gebruik kan worden genomen gaat hier een uitgebreid systeemontwikkelingstraject aan vooraf, grofweg: het ontwerpen, ontwikkelen en testen van het informatiesysteem.

Voor het ontwikkelen en testen van informatiesystemen zijn testgegevens nodig. In de praktijk worden hiervoor, om praktische redenen, vaak bestaande (tot natuurlijke per-

1 Onder een informatiesysteem verstaan we in dit artikel een set aan elkaar gerelateerde bedrijfsmiddelen die informatie verzamelen (zoeken), verwerken, opslaan en verspreiden ter ondersteuning van besluitvorming, coördinatie en controle binnen een organisatie.

Ir. H. Kikkers
is senior consultant bij ITCG. Vanuit zijn functie adviseert hij organisaties bij de inrichting van Test Data Management. Tevens leidt hij projecten op het gebied van data-integratie (datawarehousing en dataconversies).

hi.kikkers@itcg.nl

B. Nienhuis B.ICT
is consultant bij ITCG. Vanuit zijn functie voert hij informatieanalyses uit en adviseert organisaties bij het genereren van in te zetten testdata en anonimisatiesoftware.

be.nienhuis@itcg.nl

E.P. Rutkens RE
is senior manager bij KPMG IT Advisory, specialisatie onderwijs. Daarnaast heeft hij een deeltijdaanstelling als universitair docent bij de afdeling Bestuurlijke Informatievoorziening bij de vakgroep Accountancy binnen de faculteit Bedrijfskunde van de Rijksuniversiteit Groningen.

rutkens.erik@kpmg.nl

sonen herleidbare) gegevens gebruikt. Er wordt, ten behoeve van één of meer testomgevingen, simpelweg een kopie van een 'productiedatabase' met bestaande gegevens beschikbaar gesteld. Het kopiëren van een bestaande databaseomgeving is relatief eenvoudig te realiseren en te onderhouden. Daarnaast bevat een kopie van een productiedatabase alle relevante gegevens en is zij derhalve representatief voor alle bestaande situaties. Deze methode kent echter twee belangrijke nadelen:

- Het gebruik van een veelheid aan kopieomgevingen kan leiden tot aanzienlijke kosten voor opslag, beheer, lange testdoorlooptijden, testinzet, et cetera.
- De Wet bescherming persoonsgegevens (Wbp) verbiedt weliswaar het testen met bestaande persoonsgegevens niet expliciteit, maar het College bescherming persoonsgegevens (Cbp) is helder voor wat betreft het testen met persoonsgegevens. In het document 'Achtergrondstudies en verkenningen Nr. 23' stelt het dat voor het testen van informatiesystemen met bijzondere persoonsgegevens uitsluitend gegevens van fictieve personen mogen worden gebruikt.

De problematiek van de kosten voor opslag is op te lossen door een subset uit de productiedatabase in plaats van de gehele database beschikbaar te stellen. Het maken van een consistente en representatieve subset is voor organisaties vaak lastig te realiseren, maar hiervoor zijn (technische) oplossingen beschikbaar. Dit artikel gaat verder in op de problematiek van het testen van informatiesystemen met persoonsgegevens.

Allereerst wordt ingegaan op de eisen die vanuit de Wbp worden gesteld aan het testen met persoonsgegevens. Vervolgens wordt toegelicht op welke wijze een consistente en representatieve set fictieve persoonsgegevens kan worden gegenereerd door gebruik te maken van anonimisering. Het artikel wordt afgesloten met een conclusie.

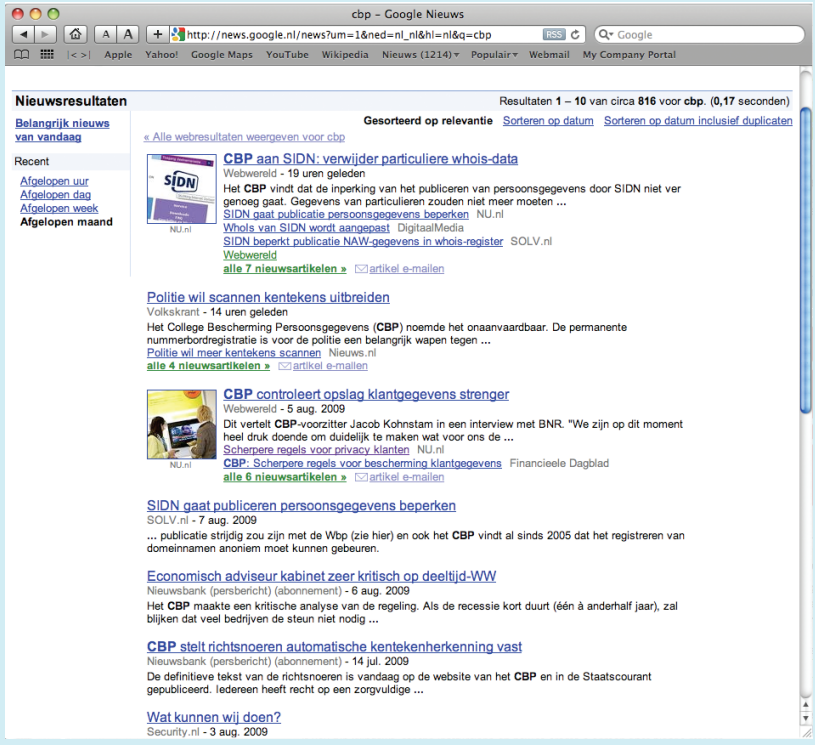
Wbp en testen met persoonsgegevens

Wet bescherming persoonsgegevens

Sinds 1 september 2001 is de Wet bescherming persoonsgegevens (Wbp) van kracht. De Wbp verving de Wet persoonsregistraties (Wpr) en is gebaseerd op de privacyrichtlijn van de EU. De wet verschilt, onder invloed van de Europese richtlijn, op belangrijke punten van de Wpr. Het grootste verschil met de Wpr is wel dat het bereik van de Wbp aanzienlijk is verruimd. Zo is het criterium van toepasselijkheid van privacyregelgeving niet meer de aanleg of het houden van een persoonsregistratie als zodanig, maar elke (afzonderlijke) verwerking van persoonsgegevens van één of meer geregistreerden. De termen van de Wbp, met name de begrippen persoonsgegevens en gegevensverwerking, zijn ruimer gedefinieerd. Een persoonsgegeven heeft betrekking op alle informatie over een geïdentificeerde of (direct of indirect) identificeerbare natuurlijke persoon. Elke theoretische mogelijkheid om een gegeven te herleiden tot een persoon geeft dit gegeven een karakter van een persoonsgegeven.

Natuurlijke personen waarvan gegevens worden verwerkt zijn bijvoorbeeld identificeerbaar als hun NAW-gegevens bekend zijn. In beginsel zijn alle gegevens die over identificeerbare personen worden verwerkt persoonsgegevens, denk aan:

- naam, adres, postcode, woonplaats, telefoon- en faxnummers, e-mailadressen;
- leeftijd, opleiding, werkervaring, gezondheid;



Figuur 1. **Bescherming persoonsgegevens is een actueel thema.**

- strafrechtelijke antecedenten;
- schulden, vorderingen, kenmerken/kentekens van eigendommen en videobeelden.

Onder verwerking verstaat de Wbp: elke operatie of geheel van operaties die al dan niet worden uitgevoerd met behulp van

geautomatiseerde procedures die worden toegepast op persoonsgegevens, zoals verzameling, vastlegging, bewaring, ordening, aanpassing, raadplegen, gebruik, verspreiding of elke andere vorm van gebruik of beschikbaarstelling van persoonsgegevens. Let wel, de Wbp is dus van toepassing op alle geautomatiseerde verwerkingen van persoonsgegevens, alsmede op alle niet-geautomatiseerde verwerkingen waarbij de persoonsgegevens zijn opgenomen in een gestructureerd geheel, zoals kaartenbakken.

Volgens de Wbp mogen persoonsgegevens alleen voor welbepaalde uitdrukkelijk omschreven en gerechtvaardigde doeleinden worden verkregen. De 'voor de gegevensverwerking verantwoordelijke' moet een duidelijk doel omschrijven waarvoor de gegevensverwerking nodig is.

De Wbp bepaalt in het algemeen dat persoonsgegevens in overeenstemming met de wet en op behoorlijke en zorgvuldige wijze moeten worden verwerkt. Gedoeld wordt op het principe van doelbinding, waarborging van rechten van geregistreerden en rechtmatigheid van de gegevensverwerking (c.q. de gronden waarop gegevens mogen worden verwerkt). Vervolgens wordt een aantal gronden opgesomd op basis waarvan persoonsgegevens mogen worden verwerkt, namelijk slechts indien:

1. de betrokkene (van wie gegevens worden verwerkt) daarvoor ondubbelzinnige toestemming heeft verleend, of
2. de verwerking noodzakelijk is voor de uitvoering van de overeenkomst waarbij de betrokkene partij is, of
3. verwerking noodzakelijk is om een wettelijke plicht waaraan de verantwoordelijke is onderworpen na te komen, of
4. de gegevensverwerking noodzakelijk is ter bestrijding van ernstig gevaar voor de gezondheid van de betrokkene, of
5. de gegevensverwerking noodzakelijk is voor de vervulling van een publieke taak door een bestuursorgaan, of
6. de verwerking noodzakelijk is voor de behartiging van het gerechtvaardigde belang van de voor de gegevensverwerking verantwoordelijke of van een derde aan wie de gegevens worden verstrekt, tenzij het belang van de betrokkene (het recht op privacy) prevaleert.

Ook de Wbp kent naast de bovenstaande regels voor het gebruik van gegevens, verplichtingen voor de verantwoordelijke, bijvoorbeeld een meldingsplicht. Geautomatiseerde gegevensverwerkingen dienen doorgaans te worden gemeld bij het Cbp. Belangrijk is, in het licht van dit artikel, dat de Wbp de 'verantwoordelijke' een beveiligingsplicht oplegt. De verantwoordelijke moet passende technische en organisatorische maatregelen nemen om de persoonsgegevens te beschermen tegen een onwettige verwerking.

Testen met persoonsgegevens

De Wbp bepaalt dat persoonsgegevens niet mogen worden gebruikt voor andere doeleinden dan waarvoor deze gegevens zijn verstrekt. De wet verbiedt het testen met persoonsgegevens van informatiesystemen niet expliciet. Aan de andere kant: het testen van informatiesystemen met persoonsgegevens is over het algemeen geen doel waarvoor persoonsgegevens worden verwerkt. Het is niet altijd te voorkomen dat persoonsgegevens voor andere doeleinden worden gebruikt zoals het testen van de juiste werking van een informatiesysteem met tot natuurlijke personen herleidbare gegevens.

Het Cbp is echter helder voor wat betreft het testen met persoonsgegevens. In het document 'Achtergrondstudies en Verkenningen Nr. 23' stelt het college dat voor het testen van informatiesystemen met bijzondere persoonsgegevens[2] uitsluitend gegevens van fictieve personen mogen worden gebruikt. Wanneer afgeweken wordt van het gebruik van fictieve persoonsgegevens, dient invulling te worden gegeven aan de algemene eisen, zoals aldaar beschreven in hoofdstuk 2, en bovendien aan de aanvullende richtlijnen, zoals beschreven in het document. Het is van het grootste belang dat ervoor gezorgd wordt dat het testen geen nadelige invloed kan hebben op de beschikbaarheid,

2 Het gaat om persoonsgegevens in risicoklasse II en hoger.

Kader 1. Eisen aan testen met persoonsgegevens.
Samengevat zijn de belangrijkste eisen voor het testen van informatiesystemen met persoonsgegevens:

- In beginsel mogen bijzondere persoonsgegevens niet voor testdoeleinden worden gebruikt.
- Wijzigingen in het systeem moeten voordat deze wijzigingen in productie worden genomen, worden getest.
- Indien mogelijk wordt bij het testen gebruikgemaakt van geanonimiseerde (fictieve) gegevens.
- In beginsel mag niet worden getest in de productieomgeving.
- De toegang tot testgegevens moet worden beperkt.
- Voor het verlenen van toegang gelden dezelfde autorisatieprocedures als in de productieomgeving.
- Toegang tot testgegevens dient te kunnen worden herleid tot natuurlijke personen.
- De gegevens van de test- en productieomgeving zijn (logisch) van elkaar gescheiden.
- In het geval testactiviteiten zijn uitbesteed aan een derde partij dienen contractuele afspraken over het gebruik van productiegegevens voor testdoeleinden te worden gemaakt (onder meer over verantwoordelijkheid en geheimhouding).

integriteit en vertrouwelijkheid van productiesystemen en/of productiegegevens.

Het gebruik van productiegegevens voor testdoeleinden brengt een aantal risico's met zich mee. Een belangrijk risico is dat door het gebruik van productiegegevens in testomgevingen onbevoegde personen toegang hebben tot gegevens die zij uit hoofde van hun functie niet nodig hebben. Denk aan salarisgegevens of gegevens over ziekteverzuim. Een andere belangrijk risico is bijvoorbeeld dat de testgegevens, in het geval de ontwikkeling van een informatiesysteem is uitbesteed, op straat komen te liggen. Onzorgvuldig gebruik of misbruik van deze gegevens kan een organisatie schaden. Daarnaast stelt het Cbp zoals gezegd dat voor het testen van informatiesystemen met bepaalde persoonsgegevens uitsluitend gegevens van fictieve personen mogen worden gebruikt. Als een organisatie deze regel overtreedt kan het Cbp een boete als sanctie opleggen.

Desondanks gebruiken veel organisaties veelal bestaande (tot natuurlijke personen herleidbare) gegevens voor testdoeleinden. De reden hiervoor is vaak een economische: er doen zich situaties voor waarin er simpelweg geen tijd is, of de kosten niet opwegen tegen de baten om fictieve gegevens te gebruiken. Door het anonimiseren van persoonsgegevens kan dit probleem worden opgelost.

Anonimiseren en filteren van persoonsgegevens

Anonimiseren van persoonsgegevens

Het anonimiseren van persoonsgegevens houdt in dat (persoons)gegevens zodanig kunnen worden gemanipuleerd, dat deze niet meer te herleiden zijn tot de oorspronkelijke (natuurlijke) persoon. We geven hier enkele voorbeelden van.

Verwisselen van gegevens

Bestaande gegevens worden verwisseld, zodat een oorspronkelijk identificerend gegeven niet meer herleidbaar is naar dezelfde natuurlijke persoon. Een nadeel van deze methode kan zijn dat de verwisselde gegevens herkenbaar zijn als bestaande gegevens.

Voorbeeld – **verwisselen van naamsgegevens**

Productie		Test	
Jan	Jansen	Jan	***de Boer***
Frans	de Boer	Frans	***Jongsma***
Kees	Jongsma	Kees	***Jansen***

Maskeren van gegevens

Bestaande gegevens worden (deels) onleesbaar gemaakt. Dit kan bijvoorbeeld door karakters te vervangen door een standaardwaarde (bijvoorbeeld een 'x'). Een groot voordeel van deze methode is dat – bij voldoende maskering – de gegevens niet te herleiden zijn tot de oorspronkelijke persoon. Een mogelijk nadeel van deze methode is dat software hiermee niet overweg kan. Denk hierbij aan banknummers, waarbij maskering zal leiden tot invalide bankrekeningnummers (een dergelijk nummer moet onder meer voldoen aan de zogenaamde 11-proef).

Voorbeeld – **maskeren van e-mailadressen**

Productie	Test
k.jansen@hotmail.nl	***k.jansen@xxxxxxx.xx***
deboer@gmail.com	***deboer@xxxxx.xxx***
jjongsma@provider.nl	***jjongsma@xxxxxxxx.xx***

Retoucheren van gegevens

Bestaande gegevens worden enigszins aangepast. Een numerieke waarde wordt bijvoorbeeld aangepast door deze maximaal tien procent te laten afwijken van de oorspronkelijke waarde. Een belangrijk voordeel is dat trends in de gegevens nagenoeg identiek blijven, maar de individuele gevallen niet exact gelijk zijn aan het oorspronkelijke gegeven.

Voorbeeld – **retoucheren van maandsalaris**

Productie		Test	
Jansen	€ 3500	Jansen	***€ 3780***
de Boer	€ 4500	de Boer	***€ 4050***
Jongsma	€ 5000	Jongsma	***€ 5100***

In situaties waarbij berekeningen plaatsvinden op basis van een datum, is het in dit geval goed denkbaar om een willekeurige datum te genereren, die valt binnen dezelfde maand. Dit zou toegepast kunnen worden bij berekeningen, die afhankelijk zijn van de geboortedatum van een persoon. Een nadeel van deze methode kan zijn dat de gegevens in grote mate gelijkenis vertonen met de oorspronkelijke gegevens.

Voorbeeld – **retoucheren van datum**

Productie	Test
21-04-1967	***01-04-1967***
12-09-1943	***01-09-1943***

Genereren van gegevens

Gegevens kunnen op basis van regels gegenereerd worden. Zo kan bijvoorbeeld een bankrekeningnummer worden gegenereerd dat voldoet aan de 11-proef.

Voorbeeld – **genereren van banknummers**

Productie		Test	
Jansen	41.85.82.971	Jansen	***12.93.69.950***
de Boer	38.97.19.722	de Boer	***81.19.93.140***
Jongsma	13.35.96.079	Jongsma	***10.96.99.726***

Verwisselen van sleutelvelden

Een bijzondere manier van anonimiseren is het verwisselen van (onderdelen van) sleutelvelden. In bepaalde situaties komt het voor dat in databases een betekenisvolle en identificerende sleutel (bijvoorbeeld burgerservicenummer) wordt gebruikt als verwijzende sleutel naar een andere tabel. Indien deze sleutel wordt geanonimiseerd, bijvoorbeeld door die te verwisselen met andere waarden in dezelfde kolom, worden hiermee ook de koppelingen tussen gegevens uit verschillende tabellen gewijzigd. Indien deze koppeling in stand gehouden moet worden, dient hetzelfde gegeven in de tabel waarnaar verwezen wordt volgens dezelfde regels verwisseld te worden. Voordeel van deze methode is dat betekenisvolle sleutels verwisseld kunnen worden en de database consistent blijft.

Voorbeeld – **verwisselen van sleutelgegevens (BSN) over twee tabellen**

Productie (Polis)			Test (Polis)		
009732445	Jan Jansen	€ 150	***261215218***	Jan Jansen	€ 150
261215218	Kees Fransen	€ 190	***009732445***	Kees Fransen	€ 190

Productie (Polis-regel)			Test (Polis-regel)		
009732445	Basis	€ 110	***261215218***	Basis	€ 110
009732445	Aanvullend	€ 40	***261215218***	Aanvullend	€ 40
261215218	Basis	€ 140	***009732445***	Basis	€ 140
261215218	Aanvullend	€ 50	***009732445***	Aanvullend	€ 50

Filteren van persoonsgegevens

Stel dat alle salarisinformatie van eigen personeelsleden is overgenomen van de productieomgeving naar de testomgeving en dat dat de persoonsgegevens zijn versleuteld. Dan zou nog steeds achterhaald kunnen worden hoeveel de directie verdient. Of stel dat bepaalde situaties een beperkt aantal malen voorkomen, dan nog zou na anonimisering achterhaald kunnen worden welke personen dit betreft. In beide voorbeelden is het dan mogelijk om de (geanonimiseerde) persoon alsnog te identificeren. Daarom is het in dergelijke gevallen gewenst de eenvoudig herleidbare gegevens uit de testset te filteren.

Technische en organisatorische inrichting

In onderstaande sectie wordt een drietal mogelijke inrichtingen beschreven om een representatieve testset met fictieve persoonsgegevens te genereren. Voor alle inrichtingen wordt gebruikgemaakt van zowel subsetgeneratie als anonimisering.

De eerste variant betreft een enkelvoudige inrichting. Hierbij wordt slechts één productiesysteem gebruikt als basis om een consistente subset aan gegevens te extraheren en te anonimisering richting een testomgeving. De tweede variant betreft een keteninrichting, waarbij consistente subsets en anonimisering plaatsvinden 'over de keten heen'. De derde variant is een uitbreiding op variant 2, waarbij de keten-testomgeving als bron voor specifieke afgeleide testomgevingen dient.

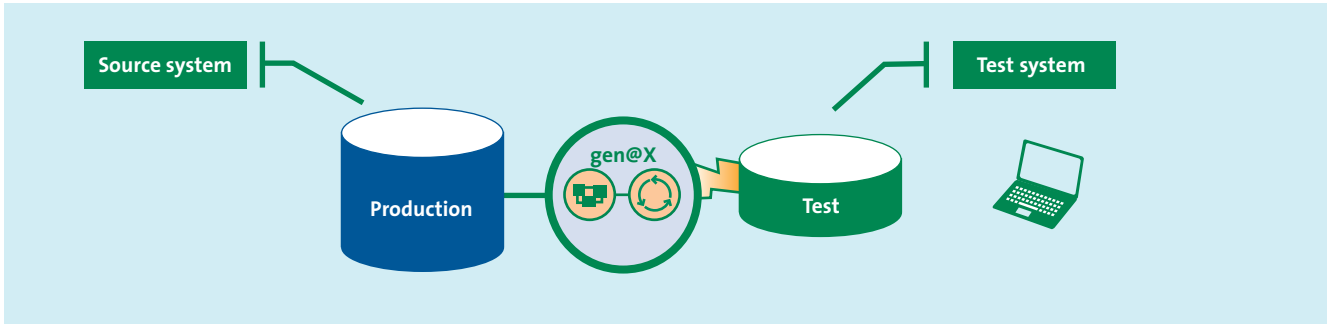
Variant 1. Enkelvoudige inrichting

In de enkelvoudige inrichting wordt op basis van een productieomgeving een representatieve en consistente subset aan gegevens geëxtraheerd. Deze subset wordt geanonimiseerd en kan worden ingezet als testomgeving. Deze testomgeving is vanwege de subsetgeneratie van beperkte omvang en kan door het toepassen van anonimiseringsregels worden gebruikt als basis voor de uitvoering van diverse tests.

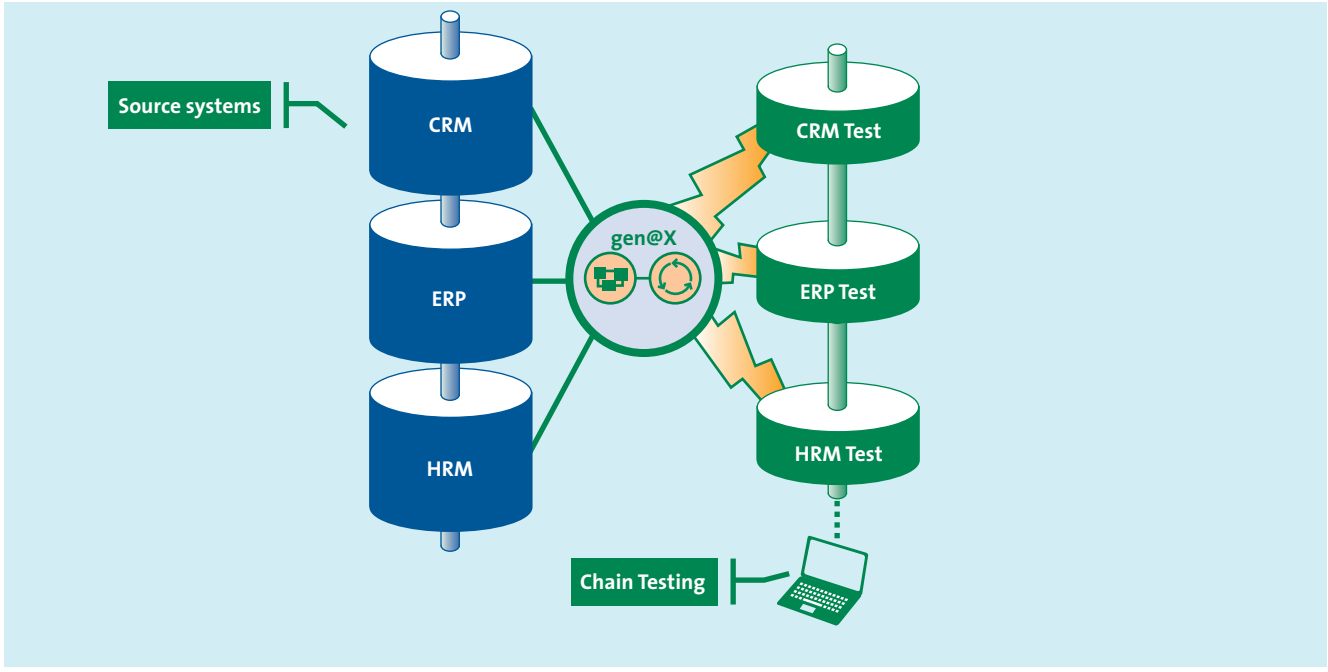
Aangezien de testers slechts kunnen beschikken over de testomgeving, zijn de productiegegevens (en dus de persoonsgegevens) afgeschermd.

Variant 2. Keten-inrichting

Indien er zich meerdere systemen in het productielandschap bevinden, kan het wenselijk zijn om een testomgeving in te richten die de ketentest kan ondersteunen. In deze variant worden representatieve en consistente subsets 'over de keten heen' geselecteerd. Indien – in onderstaand voorbeeld – bepaalde gevallen worden geselecteerd uit het CRM-systeem en overgezet naar de CRM-testomgeving, dan worden eveneens de gerelateerde gegevens van die klanten uit het ERP- en HRM-systeem geselecteerd en overgezet naar de bijbehorende testomgevingen. Ondanks dat er een selectie aan gegevens uit de bronsystemen is gehaald, is deze subset toch consistent over de systemen heen.



Figuur 2. **Enkelvoudige inrichting.**



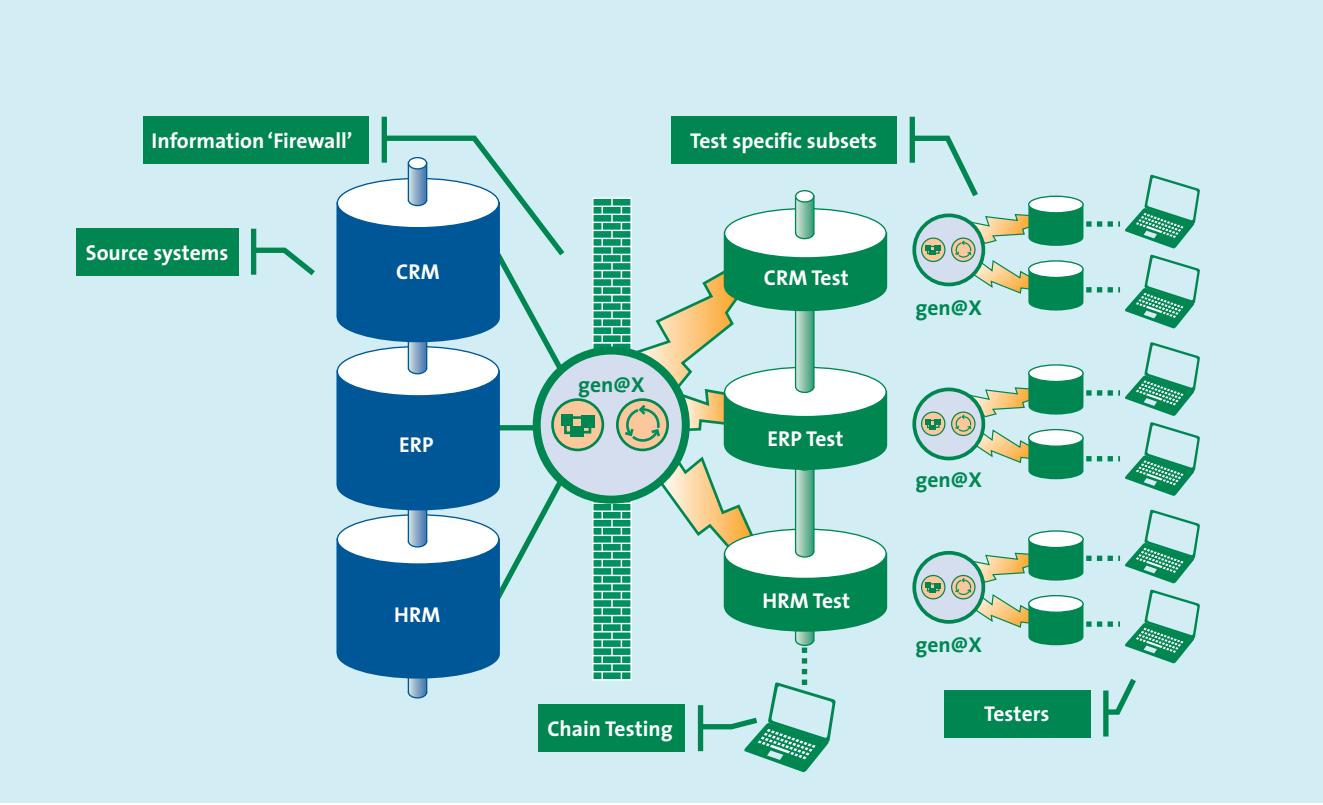
Figuur 3. **Keteninrichting.**

Daarnaast kan in deze variant anonimisering plaatsvinden 'over de keten heen'. Indien een anonimiseringsregel wordt toegepast om een bepaald gegeven uit het CRM-systeem te manipuleren en over te zetten naar de CRM-testomgeving, dan zal hetzelfde gegeven uit het ERP- en HRM-systeem op basis van exact dezelfde manipulatie worden geanonimiseerd. Hierdoor zijn de gegevens in de testomgevingen geanonimiseerd, maar toch consistent over de systemen heen. Ook in deze variant geldt dat de testers slechts kunnen beschikken over de keten-testomgeving, waarbij de productiegegevens (en dus de persoonsgegevens) zijn afgeschermd.

Variant 3. Keteninrichting met informatiebeveiliging

Deze variant is een uitbreiding op variant 2. De ingerichte keten-testomgeving kan als basis dienen voor het flexibel kunnen extraheren van specifieke testsets. Aangezien de keten-testomgeving is geanonimiseerd, kunnen testers zelf hun eigen specifieke testsets samenstellen, zonder toegang te hebben tot en inzicht te hebben in productiegegevens.

In figuur 4 is een virtuele 'Information Firewall' getekend. Hierbij is het mogelijk om tijdens het samenstellen van de keten-testsets een pseudo-identificatie uit te reiken. Aan de 'linkerkant' van de muur is de oorspronkelijke identificatie wel bekend, terwijl deze aan de 'rechterkant' niet bekend is.



Figuur 4. **Keteninrichting met informatiebeveiliging.**

Kader 2. Case landelijke zorgverzekeraar.

Praktijkvoorbeeld

Een landelijke zorgverzekeraar maakt gebruik van een in eigen beheer ontwikkeld geïntegreerd polis- en schadesysteem. In de polis- en schadeadministratie zijn gegevens vastgelegd over verzekerden. Het betreft niet alleen NAW-gegevens maar ook bijzondere gegevens zoals medicatiegegevens. Het systeem ondergaat met grote regelmaat grote (releases) en kleine wijzigingen (updates/emergency fixes). Voor het testen van de juiste werking van het polis- en schadesysteem maakt de zorgverzekeraar zowel bij het testen van een nieuwe release als bij het testen van updates gebruik van productiegegevens waaronder bijzondere persoonsgegevens. Er zijn twee belangrijke redenen waarom gebruik wordt gemaakt van productiegegevens voor testdoeleinden. Ten eerste heeft de ontwikkelafdeling aangegeven dat een representatieve set testgegevens met fictieve persoonsgegevens niet eenvoudig is te realiseren maar vooral ook erg kostbaar is. De tweede reden heeft met tijd en snelheid te maken. In het geval van updates/emergency fixes ontbreekt de tijd om een representatieve testset te genereren. De zorgverzekeraar heeft wel een richtlijn geïmplementeerd om de risico's die samenhangen met het gebruik van productiegegevens voor testdoeleinden te beheersen. De belangrijkste uitgangspunten zijn:

Conclusie

Door de digitalisering en strikte wetgeving op het gebied van persoonsgegevens wordt het steeds moeilijker om informatiesystemen goed te testen. De vraag naar representatieve en consistente testgegevens is groot. Hiervoor productiegegevens gebruiken lijkt aantrekkelijk, maar kent een aantal knelpunten.

Het gebruik van bestaande persoonsgegevens voor het ontwikkelen en testen van informatiesystemen is niet expliciet bij wet verboden, maar is sterk af te raden. De risico's van imagoschade bij verlies of oneigenlijk gebruik is erg groot. Daarnaast brengt een kopie van productiegegevens hoge kosten met zich mee. Adequate maatregelen kunnen gezocht worden in enerzijds het extraheren van een representatieve subset van productiegegevens en anderzijds het anonimiseren van privacygevoelige informatie.

Deze maatregelen bieden grote kostenbesparingen op gebied van testhardware en doorlooptijden van tests. Tevens bieden zij een goede bescherming tegen oneigenlijk gebruik van productiegegevens, waarmee wordt voldaan aan de wet- en regelgeving op het gebied van bescherming persoonsgegevens.

Het is aan te bevelen om alle risico's te inventariseren rondom het gebruik van gegevens. Welke gegevens worden op welke plek in de organisatie vastgelegd en gebruikt en met welk doel? Stel vervolgens organisatorische en technische maatregelen vast om het oneigenlijk gebruik van productiegegevens tegen te gaan. Hierbij kan gedacht worden aan een striktere functiescheiding, het aanstellen van een functionaris voor gegevensbescherming, technische beperkingen en het beschikbaar stellen van geanonimiseerde productiegegevens voor het ontwikkelen en testen van software. Voor het anonimiseren van persoonsgegevens moet worden bepaald op welke wijze geanonimiseerd moet worden.

Door deze maatregelen kan er efficiënter worden getest, kunnen kosten worden bespaard en blijft de privacy van persoonsgegevens gewaarborgd.

Literatuur

Wet bescherming persoonsgegevens, 2001.
Blarkom, G.W. van en J.J. Borking, *Beveiliging van persoonsgegevens*, Achtergrondstudies en Verkenningen Nr. 23, 2001.
Information and Privacy Commissioner, 2001, *Privacy-enhancing technologies: The path to anonymity*, Achtergrondstudies en Verkenningen Nr. 11, 2001.
Geraadpleegde bronnen
Website College bescherming persoonsgegevens, www.cpbweb.nl

- Productiegegevens die persoonsgegevens van risicoklasse II of hoger bevatten zoals beschreven in het document 'Achtergrondstudies en Verkenningen, Nr. 23' van het College bescherming persoonsgegevens mogen onder voorwaarden worden gebruikt voor testdoeleinden.
- Productiegegevens mogen worden gebruikt voor testdoeleinden na expliciete toestemming van de systeemeigenaar. De security officer heeft hierbij een adviserende en toetsende rol. Bij het gebruik van testgegevens zijn de minimale eisen zoals beschreven in kader 1 onverkort van toepassing.
- Productiegegevens die voor testdoeleinden zijn gebruikt worden na afloop van de testactiviteiten vernietigd.

In de praktijk blijkt de implementatie en naleving van de richtlijn kostbaar. Ook ontstaan problemen met het beheer van de productie- en testgegevens. Dit betreft aan de ene kant de beheerkosten en aan de andere kant de integriteit van de productiegegevens in de testomgeving. In samenwerking met een gespecialiseerde dienstverlener is de zorgverzekeraar daarom productiegegevens gaan anonimiseren en filteren. Deze dienstverlener beschikt over software, waarmee eenvoudig en geautomatiseerd representatieve testsets kunnen worden aangemaakt. Via de software kunnen de testgegevens automatisch worden geanonimiseerd. Het samenstellen, onderhouden en gebruiken van testsets is op deze wijze voor de zorgverzekeraar veel eenvoudiger, betrouwbaarder, goedkoper en veiliger geworden.