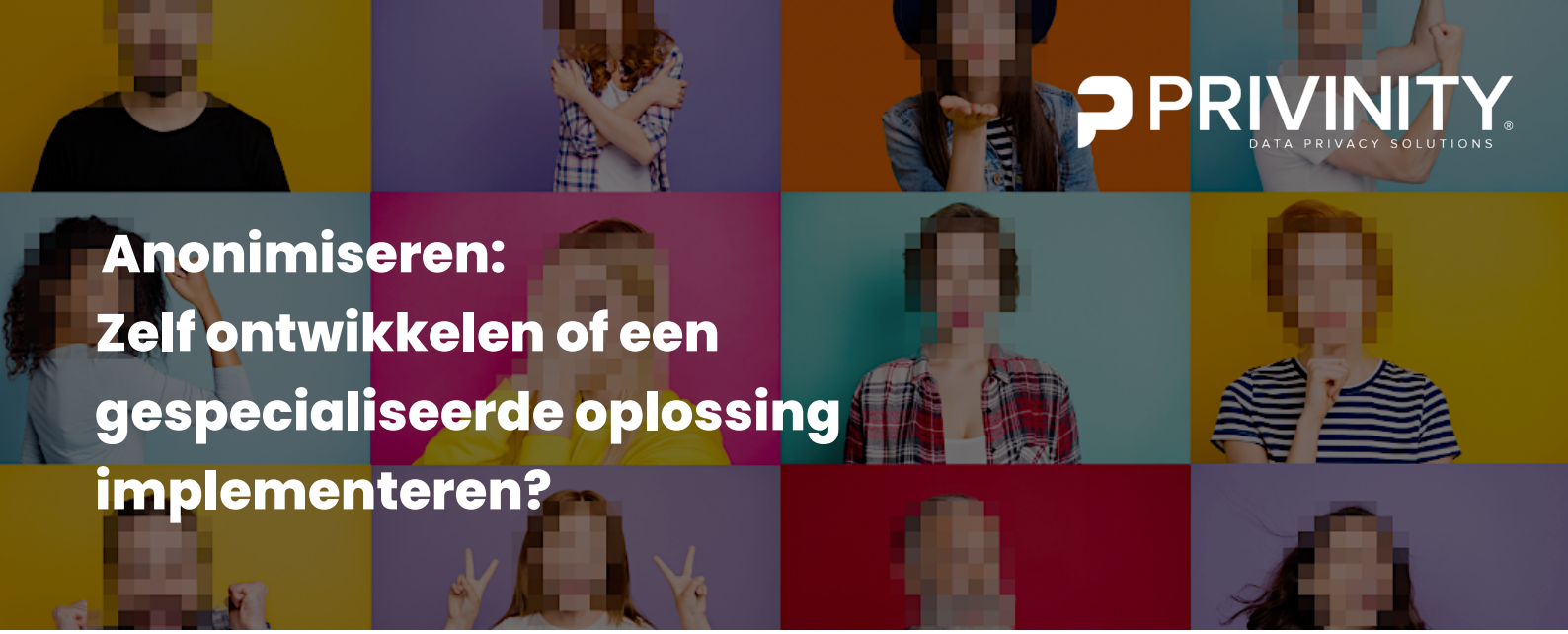

Anonimiseren: Zelf ontwikkelen of een gespecialiseerde oplossing implementeren?

Op het moment dat je als organisatie privacygevoelige data wil gaan anonimiseren sta je voor een belangrijke keuze: zelf een oplossing ontwikkelen of een gespecialiseerde oplossing implementeren. Wat de beste keuze voor jouw organisatie is hangt sterk af van de capaciteit, kennis en ervaring die je binnen je organisatie hebt. Er is – helaas - geen “one size, fits all” oplossing op het gebied van anonimiseren.

Om je te ondersteunen bij het maken van de keuze tussen zelf ontwikkelen of een gespecialiseerde oplossing implementeren hebben we dit whitepaper opgesteld. We staan in dit whitepaper stil bij de verschillende factoren die impact hebben op de keuze en zetten de voor- en nadelen tussen beide opties op een rijtje.

KEY TAKEAWAYS:

- **Besef het belang van het beschermen van privacygevoelige data**
- **Stel vast welke factoren impact hebben op de keuze om zelf een oplossing te ontwikkelen of een gespecialiseerde oplossing te implementeren**
- **Weeg de voor- en nadelen tussen beide opties af door de organisatorische impact te bepalen**

HET BELANG VAN DATAPRIVACY BESCHERMING

Het is een feit dat het beschermen van privacygevoelige data de laatste jaren voor een groeiend aantal organisaties een belangrijk onderwerp is geworden. Niet alleen vanwege geldende wet- en regelgeving (de Algemene Verordening Gegevensbescherming, oftewel de AVG), maar ook vanwege de impact op het imago en het vertrouwen in een organisatie wanneer privacygevoelige data ongewenst ingezien, gebruikt of zelfs gestolen wordt. Helaas zijn er in de media vele voorbeelden te vinden van momenten waarop verschillende organisaties de controle verliezen op hun privacygevoelige data, met alle gevolgen van dien.

Je zal dus als organisatie maatregelen moeten nemen om de privacygevoelige data die je verwerkt te beschermen. Dat varieert van informatiebeveiligingsmaatregelen – zoals het maken van backup's en het inrichten van authenticatie en autorisaties - tot aan daadwerkelijke privacybeschermende maatregelen, waar bijvoorbeeld het anonimiseren van privacygevoelige data onder valt.

Het uitgangspunt van de AVG is dat persoonsgegevens alleen verwerkt mogen worden voor het primaire doel waarvoor ze zijn aangeleverd. Alleen onder strikte voorwaarden mogen ze voor andere (secundaire) doelen worden ingezet.

Aan deze voorwaarden kan in de praktijk echter vaak niet worden voldaan waardoor je aangewezen bent op privacybeschermende maatregelen om deze gegevens alsnog te kunnen gebruiken[1].

Een belangrijke privacybeschermende maatregel die het mogelijk maakt om privacygevoelige data voor secundair gebruik in te zetten is **anonimiseren**.

VOORBEELD

Een ziekenhuis verzameld veel privacygevoelige gegevens van haar patiënten voor het primaire doel: het leveren van zorg. Op het moment dat het ziekenhuis besluit om dezelfde privacygevoelige gegevens te gebruiken om een nieuwe software release van haar Elektronisch Patiënten Dossier (EPD) te testen dan komt dit niet overeen met het originele doel waarvoor deze gegevens verzameld zijn (het leveren van zorg). Hierdoor zal het ziekenhuis privacybeschermende maatregelen moeten toepassen om alsnog de nieuwe release van de EPD software te mogen testen.

[1] Een andere mogelijkheid om persoonsgegevens alsnog te kunnen gebruiken voor secundair gebruik is door expliciet toestemming voor dat gebruik te vragen aan de persoon die die data aanlevert. Hoewel dit bij kleinschalig gebruik mogelijk nog te overzien is, is het bij uitgebreider gebruik van deze data nagenoeg onmogelijk om deze toestemming voor ieder gebruik van de data van de persoon te vragen.

ANONIMISEREN

In de kern zorgt het anonimiseren van privacygevoelige data ervoor dat de persoon die deze data aangeleverd heeft niet meer te identificeren is aan de hand van de aangeleverde data. Dit heeft als groot voordeel dat voldoende geanonimiseerde data niet onder de AVG valt. Oftewel, je mag als organisatie deze data voor ieder doel inzetten, zolang je ervoor zorgt dat de geanonimiseerde data niet gebruikt kan worden om een persoon binnen de data te identificeren.

Voor het ziekenhuis in het vorige voorbeeld betekent dit dat zij de verzamelde privacygevoelige data van haar patiënten zonder problemen mag inzetten voor het testen van haar EPD-software release zolang de data geanonimiseerd is.

Hoewel het anonimiseren van privacygevoelige data misschien relatief simpel lijkt – vaak horen we dat het beeld bij anonimiseren het vervangen van bestaande gegevens met “XXXXXX” is – blijkt het in de praktijk toch een stuk bewerkelijker te zijn. Want hoe zorg je er bijvoorbeeld voor dat de geanonimiseerde data nog steeds bruikbaar is voor het doel waarvoor je deze wil inzetten? Zo zal een data-analyse op een dataset waarbij alle personen volledig identiek zijn weinig inzichten opleveren.

ZELF AAN DE SLAG OF EEN GESPECIALISEERDE OPLOSSING IMPLEMENTEREN

Als je als organisatie de keuze maakt om aan de slag te gaan met anonimiseren zul je een belangrijke keuze moeten maken. Zelf een anonimiseeroplossing ontwikkelen of een bestaande – gespecialiseerde – oplossing implementeren.

Deze keuze wordt beïnvloed door verschillende factoren die in essentie terug te brengen zijn naar vier hoofdonderwerpen:

- **Ontwikkel capaciteit**
  Hebben we als organisatie voldoende capaciteit om een eigen anonimiseeroplossing te ontwikkelen die voldoet aan onze eisen.

- **Kennis privacybeschermende maatregelen**
  Anonimiseren is een vrij specialistische activiteit, met name op het moment dat je de geanonimiseerde data ook bruikbaar wil houden voor bijvoorbeeld test-, ontwikkel- of onderzoeksactiviteiten.

- **IT Beheer**
  Zijn we als organisatie in staat om de oplossing te installeren, onderhouden, monitoren en waar nodig bij te sturen.

- **Vertrouwen**
  Hebben we voldoende vertrouwen dat de privacygevoelige gegevens die door de oplossing geanonimiseerd worden voldoende beschermd zijn.

Ontwikkel capaciteit

De factor capaciteit is met name gericht op het moment dat je als organisatie ervoor kiest om zelf een oplossing te ontwikkelen. Als je namelijk dit pad wil inslaan dan betekent dit dat je als organisatie voldoende ontwikkel capaciteit moet kunnen aanwenden om de eigen oplossing te ontwikkelen.

Er zijn vele manieren waarop je een eigen anonimiseeroplossing kan ontwikkelen. De meest gebruikte hiervan die wij regelmatig zien is het gebruik van een script oplossing.

In het merendeel van de gevallen bevindt privacygevoelige data zich binnen één, of meerdere, database(s). De script oplossing bestaat in dit geval uit SQL[2] commando's die de gegevens die opgeslagen staan binnen de database(s) manipuleert om ze zo te anonimiseren. In deze situatie zul je dus als organisatie voldoende SQL-kennis in huis moet hebben om een dergelijk script te kunnen ontwikkelen. Daar komt bij dat naar mate je complexere anonimiseer acties wilt uitvoeren er vaak geavanceerdere SQL-kennis nodig is om dit te ontwikkelen.

Ga er dus vanuit dat er voldoende kennis aanwezig moet zijn binnen je organisatie om een dergelijke oplossing te realiseren en er ook voldoende tijd beschikbaar is om de oplossing te ontwikkelen, testen, implementeren en beheren.

Op het moment dat je als organisatie kiest voor een gespecialiseerde anonimiseeroplossing van een leverancier is dit een gebied waar je veel op kan besparen. Ten opzichte van het zelf ontwikkelen van een oplossing zorgt de implementatie van een gespecialiseerde oplossing ervoor dat je je niet bezig hoeft te houden met ontwikkeling maar direct de oplossing in gebruik kan nemen.

Hier kleeft echter ook een mogelijk nadeel aan. Een gespecialiseerde oplossing zal in de meeste gevallen een generieke anonimiseeroplossing zijn. Dat wil zeggen dat deze oplossing ontwikkeld is met een brede ondersteuning op het gebied van anonimiseermogelijkheden en ondersteunde dataformaten. Dit kan betekenen dat de mogelijkheden van de gespecialiseerde oplossing niet voldoende aansluiten bij de wens die je als organisatie hebt waardoor je mogelijk niet alle anonimiseer activiteiten binnen één oplossing kan toepassen.

Kennis privacybeschermende maatregelen

Het implementeren, inrichten en toepassen van privacybeschermende maatregelen is een specialistisch vakgebied op het snijvlak van techniek en regelgeving. Hierbij ben je continu op zoek naar de balans tussen data die voldoende beschermd is vanuit wet- en regelgevingsperspectief terwijl deze tegelijkertijd bruikbaar blijft voor het doel waarvoor je de geanonimiseerde data wilt inzetten.

[2] SQL, of Structured Query Language, is de standard programmeertaal die gebruikt wordt om gegevens binnen een relationele database (zoals Microsoft SQL Server, Oracle of MySQL) te bevragen of te manipuleren.

Bij een eigen ontwikkelde anonimiseeroplossing, als ook een gespecialiseerde oplossing, moet je dus rekening houden met op welke manier je welke data anonimiseert. Waar je bij een specialistische oplossing gebruik kan maken van vooraf ontwikkelde anonimiseer algoritmen en methode, zul je deze bij een eigen ontwikkelde oplossing zelf moeten ontwerpen, ontwikkelen en toetsen op effectiviteit.

VORMEN VAN ANONIMISEREN

Er zijn veel verschillende manieren van anonimiseren mogelijk maar globaal gezien zijn alle anonimiseermethoden op te delen in vier verschillende categoriën:

Maskeren

Maskeren is een van de meest bekende en effectieve manier van anonimiseren. Anonimiseermethoden binnen deze categorie richten zich op het onleesbaar maken van de originele – gevoelige – data door deze bijvoorbeeld te vervangen door willekeurige tekens, X'en of de bekende zwarte balken binnen documenten en afbeeldingen.

Generaliseren

Generaliseer anonimiseermethoden richten zich op het identiek maken van data waardoor unieke gegevens niet meer kunnen leiden tot een identificatie. De gedachten hierbij is dat als er geen unieke gegevens bekend zijn deze ook niet direct naar één individu kunnen leiden. Een voorbeeld hiervan is het vervangen van alle unieke e-mailadressen naar één specifiek e-mailadres.

Randomisatie

Bij het toepassen van anonimiseermethode die binnen de randomisatie categorie passen wordt data in een (bepaalde mate) willekeurige manier getransformeert. Daarbij onderkennen we nog twee subcategoriën: het toevoegen van ruis, waarbij bijvoorbeeld een geboortedatum aangepast wordt naar een willekeurige andere datum binnen een specifek interval, en permutatie waarbij we meerdere data waarden opnieuw ordenen. Een voorbeeld van permutatie is het willekeurig door elkaar "husselen" van voornamen zodat ieder persoon in de data een willekeurige andere voornaam ontvangt.

Pseudonimiseren

Hoewel pseudonimiseren strikt genomen geen anonimiseermethode is[3] , wordt deze wel vaak toegepast binnen anonimiseeroplossingen. Bij pseudonimiseren vervang je een unieke waarde voor een pseudoniem, bijvoorbeeld door de originele waarde te vervangen door een hashwaarde.

[3] Het verschil tussen anonimiseren en pseudonimiseren is dat de eerste onomkeerbaar is. Het zou dus na anonimiseren niet meer mogelijk moeten zijn om het anonieme resultaat terug te berekenen naar de originele waarde. Kan dit wel, dan spreken we over pseudonimiseren.

Bij gespecialiseerde oplossingen is zeer waarschijnlijk veel tijd en moeite gestopt in het ontwikkelen van de anonimiseermethoden zodat deze effectief en efficiënt uitgevoerd kunnen worden binnen verschillende soorten data. Op het moment dat je zelf een oplossing gaat ontwikkelen zul je zelf anonimiseer algoritmen en methoden moeten ontwikkelen waarbij in eerste instantie nog niet duidelijk is hoe effectief en efficiënt deze uitgevoerd worden.

Naast de ontwikkeling en/of toepassing van anonimiseermethoden is het ook van belang dat deze breed toepasbaar zijn op verschillende technische platformen. De verwerking van privacygevoelige gegevens vindt op zo veel plaatsen en manieren plaats dat je er bijna duizelig van zou worden. Het is dus van belang dat een anonimiseeroplossing zo veel mogelijk van deze data typen, platformen en bronnen ondersteund om een zo'n breed mogelijk dekking van privacybeschermende maatregelen te kunnen garanderen.

IT Beheer

Binnen alle IT-omgevingen dient er beheer plaats te vinden op de geïnstalleerde software en hardware binnen de omgeving. Een anonimiseeroplossing is hier geen uitzondering op. Dit betekent dat het implementeren van een anonimiseeroplossing, ongeacht of deze zelf ontwikkeld is of niet, een bepaalde mate van beheer vraagt van de IT-afdeling.

Hieronder vallen de reguliere beheersactiviteiten, zoals het installeren van nieuwe versies van de oplossing, maar ook activiteiten die gericht zijn op de inrichting en uitvoer van privacybeschermende maatregelen. Denk hierbij aan zaken zoals de tijd die het kost om de maatregelen uit te voeren of het controleren of de maatregelen correct zijn toegepast.

Een ander voorbeeld hiervan is het doorvoeren van wijzigingen in de te-anonimiseren data. In de praktijk zal het werken met geanonimiseerde data – met name tijdens de eerste paar keer anonimiseren – leiden tot wijzigingen op de manier waarop data geanonimiseerd wordt. Zo kan bijvoorbeeld een tester erachter komen dat door een specifiek attribuut op een specifieke manier te anonimiseren een test scenario niet meer mogelijk is. Hierdoor zal de bestaande anonimiseerinrichting aangepast moeten worden waardoor het test scenario weer uit te voeren is en het attribuut nog steeds beschermd blijft.

Nog een voorbeeld zijn wijzigingen in de databronnen welke geanonimiseerd worden. Dit kan bijvoorbeeld ontstaan doordat er een nieuwe software update geïnstalleerd is waardoor er nieuwe attributen toegevoegd zijn of bestaande attributen zijn gewijzigd.
Indien er door deze wijzigingen nieuwe attributen ontstaan die mogelijk privacygevoelige gegevens bevatten dan zullen deze ook toegevoegd moeten worden aan het anonimiseerproces.

De hoeveelheid tijd en effort die het kost om met dergelijke wijzigingen om te gaan is sterk afhankelijk van de hoeveelheid databronnen die geanonimiseerd worden, de hoeveelheid wijzigingen binnen deze databronnen en het gebruik van de geanonimiseerde data.

Een eigen ontwikkelde anonimiseeroplossing kan hier voordelen hebben doordat deze dicht aan kan sluiten bij het specifieke gebruik binnen je organisatie. Zo kun je tijdens de ontwikkeling van een eigen oplossing bijvoorbeeld al rekening houden met het releaseschema van je IT-afdeling of zorgen dat de eigen oplossing gebruik maakt van technologieën waar je IT-afdeling veel ervaring mee heeft.

Een gespecialiseerde oplossing kan ook veel positieve impact hebben op de beheer lasten. Zo kunnen deze oplossingen verschillende functionaliteiten bevatten die de impact van wijzigingen in je databronnen minimaliseert. Een voorbeeld hiervan is de automatische detectie van privacygevoelige gegevens welke wij binnen de Privinity oplossingen hebben ontwikkeld. Via deze functionaliteit worden wijzigingen in databronnen automatisch gedetecteerd en wordt er geanalyseerd of deze wijzigingen privacygevoelige gegevens raken. Indien dit het geval is, dan kan de anonimiseer inrichting hier automatisch op aangepast worden waardoor je het beheer op die inrichting aanzienlijk verlaagd.

Vertrouwen

Een belangrijke factor die invloed heeft op de keuze om zelf een anonimiseeroplossing te ontwikkelen of een gespecialiseerde oplossing aan te schaffen is het vertrouwen dat de data die door de oplossing geanonimiseerd is voldoende beschermd is.

Je ontwikkelt of implementeert een oplossing omdat je als organisatie de keuze hebt gemaakt dat je op een veilige manier met privacygevoelige gegevens wilt werken. Je wilt er dan ook zeker van zijn dat de geanonimiseerde resultaten voldoen aan de eisen die je eraan stelt. Daarnaast moet de oplossing deze resultaten keer op keer kunnen leveren waardoor je erop kan vertrouwen dat de oplossing op een voorspelbare manier functioneert.

Vertrouwen krijgen in een anonimiseeroplossing is iets dat tijd kost en over verschillende niveaus binnen je organisatie moet ontstaan. Zo zullen de beheerders van de oplossing vertrouwen moeten krijgen dat de oplossing stabiel is en de uitvoering van het anonimiseren geen impact heeft op andere processen. Gebruikers van de oplossing, als ook gebruikers van de geanonimiseerde data, moeten vertrouwen krijgen dat de oplossing voldoende mogelijkheden biedt om in hun gebruiksdoelen te voorzien (een tester moet natuurlijk kunnen blijven testen, ook op geanonimiseerde data). En als laatste – en zeker niet de minst belangrijkste – zullen de rollen die verantwoordelijk zijn voor het dataprivacy en securitybeleid binnen je organisatie – vaak de Functionaris Gegevensbescherming (FG) of de Chief Information Security Officer (CISO) – erop moeten vertrouwen dat de geanonimiseerde data voldoende beschermd is. Kortom er is heel wat vertrouwen te winnen binnen een organisatie.

Bij een eigen ontwikkelde oplossing begin je vaak met een achterstand op het gebied van vertrouwen. De oplossing moet nog ontwikkeld worden, of is net ontwikkeld, en moet zich nog bewijzen in de praktijk. Voor gespecialiseerde oplossingen is dit vaak al bewezen. De meeste gespecialiseerde oplossingen hebben veel referenties, of use-cases, waarin de oplossing succesvol is toegepast, vaak ook binnen de sector waarin je eigen organisatie actief is. Dit zorgt ervoor dat je bijvoorbeeld ervaringen op kan halen bij andere organisaties en zo kan bepalen of de oplossing goed binnen je eigen organisatie zou passen.

Daarnaast is het ontwikkelen van een gespecialiseerde anonimiseeroplossing vaak de “core business” van de leverancier. Dat wil zeggen dat de ontwikkelaar van de oplossing veel tijd en aandacht zal geven aan de oplossing om ervoor te zorgen dat deze zo goed mogelijk functioneert en betrouwbaar is. Op het moment dat je als organisatie zelf aan de slag gaat met het ontwikkelen van een anonimiseeroplossing dan doe je dit er in essentie als activiteit bij, naast andere activiteiten. Dit zorgt er natuurlijk voor dat de tijd en effort die je als organisatie in de eigen ontwikkelde oplossing kan stoppen waarschijnlijk aanzienlijk minder zal zijn dan wat een leverancier van een gespecialiseerde oplossing investeert.

VERGELIJKINGSOV-ERZICHT

Nu we de verschillende factoren die effect hebben op de keuze tussen het zelf ontwikkelen van een anonimiseeroplossing of een specialistische oplossing implementeren hebben beschreven, tonen we de resultaten van deze vergelijking in het overzicht op de volgende pagina.

Hoewel er verschillende gespecialiseerde oplossingen zijn die allemaal verschillen op de beschreven onderdelen hebben we een poging gedaan deze op een gemiddelde basis te scoren. Zo zullen er gespecialiseerde oplossingen zijn die veel implementatietijd in beslag nemen maar tegelijkertijd zijn er ook oplossingen die buitengewoon snel te implementeren zijn. In een dergelijke situatie hebben we dus voor de gemiddelde waarde gekozen.

Onderdeel	Zelf ontwikkelen	Gespecialiseerde oplossing
Benodigde ontwikkel kennis & capaciteit	Veel	Niet van toepassing
Beschikbare ontwikkeltijd	Veel	Niet van toepassing
Implementatietijd	Weinig	Gemiddeld
Benodigde tijd voor het testen van het anonimiseerproces	Veel	Weinig tot gemiddeld
Complexiteit implementatie	Weinig	Gemiddeld
Aansluiting technische omgeving	Veel	Gemiddeld tot veel
Ondersteuning verschillende dataformaten, typen en platformen	Weinig	Veel
Anonimiseer mogelijkheden	Weinig	Veel
Technische beheer van de oplossing	Gemiddeld	Weinig
Functioneel beheer van de oplossing	Gemiddeld	Weinig
Kwaliteit anonimiseren	Gemiddeld	Goed tot zeer goed
Integratie bestaande processen	Zeer goed	Gemiddeld tot goed
Vertrouwen in de oplossing	Weinig tot gemiddeld	Veel

CONCLUSIE

Een keuze maken tussen het zelf ontwikkelen van een anonimiseeroplossing of een gespecialiseerde oplossing implementeren is afhankelijk van verschillende factoren. Deze zullen bij iedere organisatie verschillend zijn. Het loont dus zeker de moeite om hierin een goede afweging te maken aan de hand van de situatie binnen je eigen organisatie. Dit heeft niet alleen effect op de kosten van het ontwikkelen, implementeren en beheren van een anonimiseeroplossing, maar ook op het vertrouwen en de kwaliteit van de geanonimiseerde data.

Wij hopen dat de informatie in dit whitepaper je kan ondersteunen bij het maken van een keuze die goed bij je organisatie past. Mocht je naar aanleiding hiervan nog vragen hebben, meer willen weten over dataprivacy bescherming, of ben je benieuwd hoe de Privinity oplossingen jouw organisatie kunnen ondersteunen bij het implementeren van privacybeschermende maatregelen, dan nodigen we je graag uit om onze website te bezoeken. Uiteraard is het ook mogelijk om vragen en/of opmerkingen via de mail te stellen. Onze contactgegevens zijn aan de rechterkant van deze pagina terug te vinden.

AUTEURS

Enrico van de Laar
Data Privacy Engineer @ Privinity
e.vandelaar@privinity.com

Ernst Visser
Data Protection Officer @ Privinity
e.visser@privinity.com