

## WHITEPAPER

# Een datagedreven organisatie kan niet zonder Modern Data Platform

### Axians

Eemsgolaan 15  
9727 DW Groningen

Esp 120  
5633 AA Eindhoven

[info.nl@axians.com](mailto:info.nl@axians.com)

Vergroot de  
grip op data  
en analyse-  
resultaten









# Een datagedreven organisatie kan niet zonder Modern Data Platform

Vrijwel iedere organisatie heeft ergens in het strategisch plan opgenomen dat zij beter gebruik wil maken van data. Data moet niet langer primair gebruikt worden voor managementinformatie, maar moet ook medewerkers op de werkvloer helpen betere beslissingen te nemen. Dit stelt organisaties voor een probleem, want ze zijn niet in staat om verschillende soorten data op één platform te combineren en in samenhang te analyseren. Daarom heeft Axians het Modern Data Platform ontwikkeld, een platform dat geschikt is voor opslag en analyse van alle typen data uit alle mogelijke bronnen, en dat bovendien in de governance van deze data voorziet. Hoe ziet dit Modern Data Platform eruit? En welke voordelen biedt dit in de praktijk? Lees er alles over in dit whitepaper.



## INHOUDSOPGAVE

<b>Inleiding</b>	<b>3</b>
Hoofdstuk 1	
<b>Betere beslissingen op basis van data analytics</b>	<b>4</b>
Hoofdstuk 2	
<b>Waar ligt momenteel de bottleneck?</b>	<b>6</b>
Hoofdstuk 3	
<b>Wat is een Modern Data Platform?</b>	<b>7</b>
Hoofdstuk 4	
<b>Onderdelen Modern Data Platform</b>	<b>8</b>
Hoofdstuk 5	
<b>Eén platform voor al je informatie vraagstukken</b>	<b>10</b>

# Betere beslissingen op basis van data analytics

“Nieuwe inzichten ontstaan voornamelijk door data uit verschillende bronnen te combineren en in samenhang te analyseren.”

Steeds meer organisaties zien in dat ze waarde kunnen halen uit vooral het combineren van data. Door geautomatiseerd op zoek te gaan naar patronen en afwijkingen daarin doen ze nieuwe inzichten op. Een ziekenhuis kan bijvoorbeeld op zoek gaan naar voorspellers die aangeven dat patiënten met bepaalde kenmerken minder goed op een bepaalde behandeling zullen reageren. Een school kan bijvoorbeeld op basis van verzuim, studieresultaten en iemands activiteit in de online leeromgeving voorspellen welke leerlingen voortijdig met de studie zullen stoppen. Door deze leerlingen tijdig de juiste begeleiding te bieden kan worden voorkomen dat ze afhaken. Zo kan iedere sector wel manieren vinden om data in te zetten in het alledaagse werk. Data worden dus al lang niet meer alleen gebruikt op strategisch, maar ook op tactisch en operationeel niveau.

## BETERE PERSONEELSPANNING DOOR COMBINATIE VAN DATA

Het onderwijs, de zorg, de techniek, er zijn nogal wat sectoren die te maken hebben met een ontzettend krappe arbeidsmarkt en daardoor schaarste aan personeel. Zou het niet handig zijn als de HR-afdeling een dashboard heeft dat hen helpt bij de personeelsplanning? Bijvoorbeeld een dashboard dat vertelt hoeveel FTE in welke expertisegebieden je de komende maand op welke dagen kunt inzetten. Daarvoor combineer je verschillende getallen: het aantal medewerkers en FTE's per afdeling (wie werken er in deeltijd en op welke dagen?), de verdeling van de FTE's over verschillende expertisegebieden, de overwerkuren en de limietdatum waarop overwerkuren vervallen als ze niet worden opgemaakt, geplande vakanties, het verzuim per afdeling en expertisegebied en de laatst gemeten ervaren werkdruk. Nu worden al deze gegevens in verschillende systemen bijgehouden. Door deze data bij elkaar te brengen en in samenhang te analyseren kunnen betere prognoses worden gemaakt van het aantal beschikbare uren per dag per afdeling en/of expertisegebied voor de komende maanden, rekening houdende met het risico dat medewerkers langdurig ziek worden of een andere baan zoeken.

Kortom, zo kan beter worden ingeschat of er voldoende mensen beschikbaar zijn voor de werkzaamheden die in de komende periode moeten worden gedaan.



In dit voorbeeld gaat het alleen om data die traditioneel beschikbaar is in het datawarehouse: Financiën, HR en Productie. Maar organisaties maken ook meer en meer gebruik van nieuwe typen data, zoals data van allerlei met het internet verbonden sensoren (IoT), ongestructureerde data (zoals een sentimentanalyse op social media) en metadata (bijvoorbeeld time stamps van processen).



### **BETERE INZICHTEN DOOR ANALYSE VAN NIEUWE TYPEN DATA**

We zien allemaal dat de hoeveelheid data enorm snel groeit omdat we steeds meer zaken vastleggen. Zo vragen bijvoorbeeld ziekenhuizen in het kader van de Patient Reported Outcome Measures (PROMs) patiënten op verschillende momenten tijdens en na het behandeltraject naar hun kwaliteit van leven. Juist de data in de vrije velden willen zij graag meenemen in hun analyse. Ook het analyseren van tijddata is populair. Ziekenhuizen kunnen op basis van de vastleggingstijdstippen in het Elektronisch Patiënten Dossier (EPD) zien of er onnodige vertraging is in een zorgpad. Scholen kunnen op basis van de time stamps in de online leeromgeving onder meer zien hoe vaak, hoe laat en hoe lang leerlingen inloggen om te studeren. Dit soort data zijn vaak een belangrijke voorspeller van vroegtijdige uitval.

Kortom, gegevens die tot voor kort ofwel niet werden verzameld ofwel niet eenvoudig uit een systeem konden worden gehaald, worden nu juist steeds vaker interessant gevonden. De kunst is om die data zo vast te leggen en te ontsluiten dat je er ook iets mee kunt.

# Waar ligt momenteel de bottleneck?

Het combineren en analyseren van gegevens uit verschillende bronnen is voor de meeste organisaties een ingewikkelde opgave. Gegevens zijn immers vaak in een grote hoeveelheid systemen opgeslagen. Daarnaast stelt het verschillende gebruik van data organisaties voor een aantal uitdagingen.

- ▶ Voor de productie van klassieke management-informatie volstaat doorgaans de data in het centrale datawarehouse. Wanneer informatievraagstukken echter meer operationeel worden, is data op een veel gedetailleerder niveau nodig, dat niet voorhanden is in het datawarehouse. In dergelijke situaties worden data vaak rechtstreeks uit één of meerdere bronsystemen onttrokken. In commerciële organisaties is dat bijvoorbeeld het CRM-systeem, in een ziekenhuis het Elektronisch Patiënten Dossier en in een onderwijsinstelling de Elektronische Leeromgeving.
- ▶ Operationele informatie moet ad hoc en snel opgeleverd kunnen worden op het moment dat de vraag speelt. Deze aanpak vraagt veel van een centrale BI- of analytics-afdeling. De werkvloer wil zelf snel met data aan de slag omdat vragen direct beantwoord moeten worden en er nu een beslissing moet worden genomen. Men wil niet wachten tot een zwaarbelaste BI-afdeling eindelijk tijd heeft. De vraag naar Self Service BI neemt toe. Dit stelt niet alleen eisen aan data ontsluiting, maar ook aan Data Governance.
- ▶ Informatie moet onder andere betrouwbaar en herleidbaar zijn en goed geautoriseerd. Wie heeft toegang tot welke gegevens en wie heeft gebruik gemaakt van dat recht? Voor een datawarehouse is de governance vaak redelijk goed geregeld. Maar voor data die daar buitenom onttrokken wordt aan bronsystemen is dit veelal niet het geval. Wie wanneer met welke data wat heeft gedaan is vaak onduidelijk. In het kader van de AVG is dit simpelweg niet toegestaan. We gaan verderop in deze whitepaper dieper in op het thema data governance.
- ▶ Het detailniveau waarop data kwalitatief hoogwaardig opgeslagen moeten worden, neemt sterk toe. Een datawarehouse wordt primair ontwikkeld om managementinformatie te genereren. Of het ziekteverzuim nu 5,7 of 5,8 procent is maakt voor het beeld niet zoveel uit. Maar waar data operationeel gebruikt wordt, is absolute juistheid van informatie veel belangrijker. Voor een afdelingsmanager is het voor zijn of haar planning wél heel relevant of iemand wel of niet werkt op een bepaalde dag.
- ▶ Om deze redenen zijn analyses van data uit verschillende bronnen op dit moment nog complex en tijdrovend. Daardoor is het aantal analyses dat je kunt uitvoeren beperkt. Alleen voor de heel belangrijke strategische vraagstukken zet je de BI-afdeling aan het werk. Terwijl, we benadrukken het nog maar eens, het juist de wens is om inzichten uit data ook te gebruiken bij alledaagse vraagstukken. Helaas kan aan deze wens momenteel alleen worden voldaan als de data in één en hetzelfde systeem is vastgelegd, en zelfs dat is soms lastig.



# Wat is een Modern Data Platform?

“Het Modern Data Platform ontsluit alle beschikbare data uit een veelheid aan bronnen op een gestructureerde en gecontroleerde manier.”

Om het hoofd te bieden aan bovenstaande uitdagingen heeft Axians het Modern Data Platform ontwikkeld. Het platform ontsluit alle beschikbare data uit een veelheid aan bronnen - interne en externe - op een gestructureerde en gecontroleerde manier. Het biedt daar bovenop verschillende tools waarmee medewerkers zelf analyses kunnen maken. Hoe die tools eruit zien en van welke onderliggende datasets ze gebruikmaken, hangt helemaal af van de gebruiker, diens rechten en het type data dat hij nodig heeft voor zijn analyse. Een data scientist heeft iets anders nodig dan een medewerker met een relatief simpele analysevraag. Daarnaast biedt het platform alle functionaliteit die nodig is om de governance op data te regelen.

## WAT MOET JE REGELEN OM DATA GOVERNANCE TE WAARBORGEN?

Wie verschillende soorten data in samenhang wil analyseren, zal een aantal zaken goed moeten regelen:

- ▶ **Datakwaliteit:** in hoeverre is de data volledig, actueel en betrouwbaar?
- ▶ **Definitie:** welke schaal is gebruikt om een bepaalde waarde te meten? Welke definitie gebruiken we als we praten over ziekteverzuim? Meten we een lengte in centimeters of inches? Noteren we een datum op de Nederlandse of Amerikaanse manier en betekent 1-6-2020 daarom 1 juni of 6 januari?
- ▶ **Herleidbaarheid:** wat is het bronsysteem? Hoe en in welke context is de data tot stand gekomen?
- ▶ **Beschikbaarheid:** wie hebben allemaal toegang tot welke typen data? Via welke devices kunnen zij bij de data?
- ▶ **Beveiliging:** wie heeft toegang tot welke data? Op welke manieren is de data afgeschermd voor onbevoegden (enkele authenticatie, dubbele authenticatie)?
- ▶ **Privacy:** is traceerbaar wie welke data heeft vastgelegd? Weten we waar, door wie en voor welke doeleinden de data is gebruikt? Kunnen we garanderen dat data die niet herleidbaar mag zijn naar een persoon ook nooit herleidbaar is?

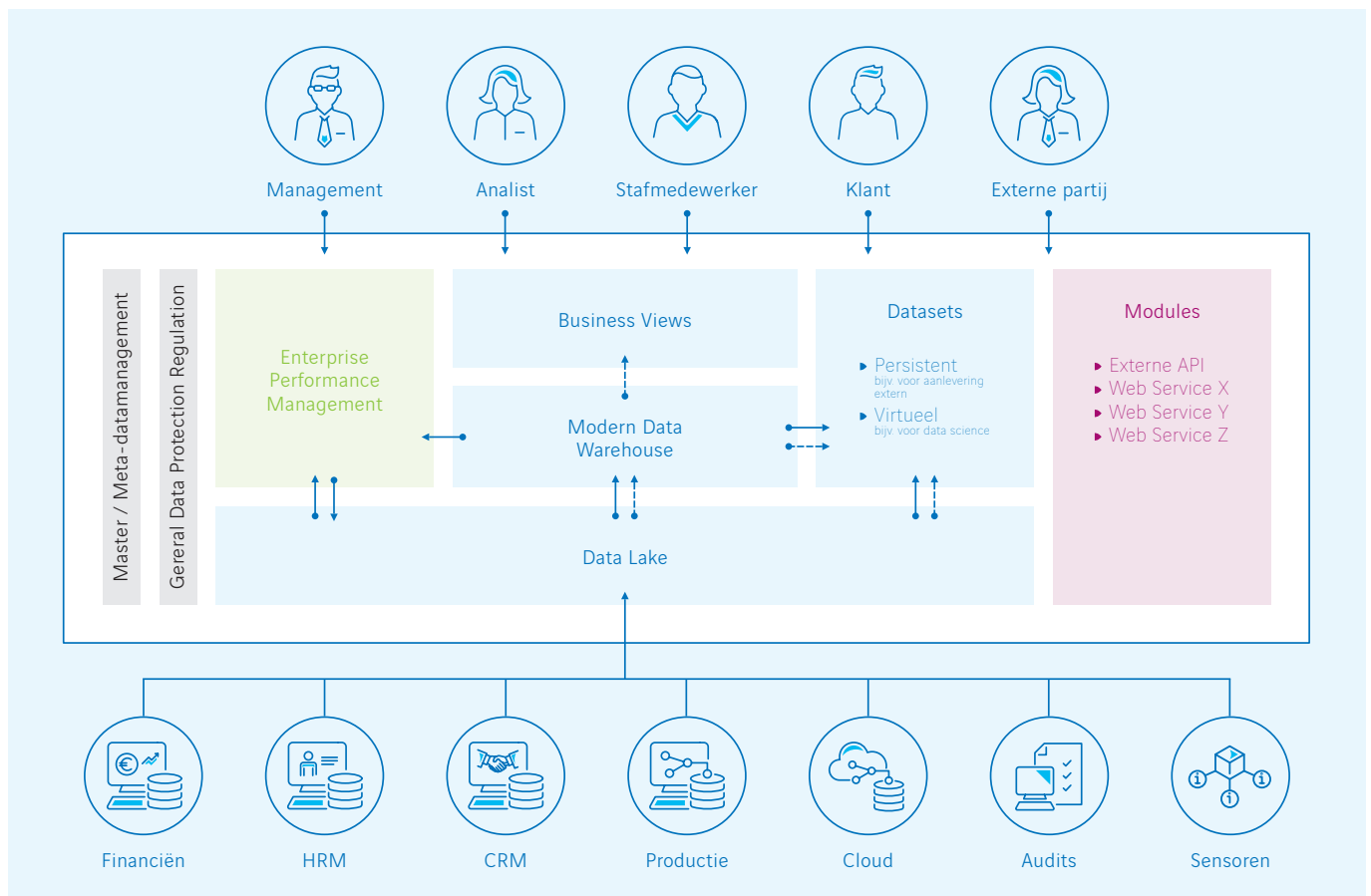
## GEBRUIK VAN MACHINE LEARNING VERGROOT BELANG VAN GOVERNANCE

Nu steeds meer organisaties gebruik gaan maken van Machine Learning en het daardoor voor gebruikers steeds moeilijker wordt om te begrijpen hoe een algoritme tot zijn advies of conclusie komt, wordt data governance nog belangrijker dan het al was. ‘Garbage in’ is immers ‘garbage out’. Als straks niet meer helemaal herleidbaar is hoe het algoritme tot een aanbeveling of conclusie komt, moet je heel zeker weten dat de onderliggende data goed is. Omdat verschillende bronsystemen vaak verschillende definities gebruiken, de datakwaliteit niet altijd even betrouwbaar is en ook de context lang niet altijd goed wordt vastgelegd, is data niet altijd zomaar te combineren.

Het wordt nog lastiger als naast eigen data ook data van externe partijen wordt toegevoegd. Veel ziekenhuizen staan nu bijvoorbeeld voor de uitdaging om data die patiënten zelf met een smart watch genereren te gebruiken in hun analyses. Hoe weet je of de data uit deze externe bron betrouwbaar is?

Minstens zo belangrijk is het garanderen van de privacy. Je kunt bijvoorbeeld ten behoeve van research patiëntdata anonimiseren. Maar als geanonimiseerde data uit verschillende bronnen bij elkaar komt, bestaat de kans dat die anonieme data toch te herleiden is tot een specifieke patiënt. Bijvoorbeeld omdat iemand een zeldzame bloedgroep heeft, meedoet aan een wetenschappelijk onderzoek en in een cohort maar één patiënt met die bloedgroep voorkomt.

# Onderdelen Modern Data Platform



## Data lake

De basis van het platform is een data lake dat wordt gevuld met data uit een veelheid van bronnen. Het data lake bevat een kopie van de brondata. Per bron wordt bepaald hoe vaak de data in het data lake moet worden ververs. Voor sommige analyses is het immers niet belangrijk dat altijd de laatste stand van zaken wordt gebruikt – denk bijvoorbeeld aan een behandeladvies op basis van de uitkomstdata bij vergelijkbare patiënten – terwijl het voor andere vraagstukken cruciaal is – denk bijvoorbeeld aan de beddenscapaciteit.

Het data lake houdt de historie van data bij. Je kunt altijd teruggaan naar een bepaalde kopie. Als in het bronsysteem data wordt overschreven, is er nog altijd

een historische file beschikbaar in het data lake.

Niet alle data die je genereert hoeft te worden opgeslagen in het data lake. Met sensoren gegenereerde real-time data die alleen wordt gebruikt voor monitoring, hoeft niet bewaard te blijven. Deze data wordt ‘at the edge’ geanalyseerd en alleen analyseresultaten die aanleiding gaven voor een bepaalde actie worden opgenomen in het data lake. Denk bijvoorbeeld aan sensordata van een auto: alleen als de sensor een signaal genereert wordt de data opgeslagen in het data lake voor eventuele verdere analyse. Het platform beschikt over tooling die automatisch filtert welke data wel en niet moet worden gekopieerd naar het data lake.



“Door de keus voor de cloud en het bijbehorende pay-per-use model wordt het betaalbaar om complexe analyses uit te voeren op basis van hele grote datasets.”

### Enterprise Performance Management

Dit onderdeel bevat alle financieel gedreven data, specifiek gericht op het management. Het betreft data die traditioneel vaak in een klassiek datawarehouse werd opgeslagen.

### Modern datawarehouse

Dit onderdeel bevat alle andere data die vaak wordt gebruikt in analyses, denk bijvoorbeeld aan data over klanttevredenheid of medewerkerstevredenheid. Immers, driekwart van de in het data lake beschikbare data wordt nooit of slechts heel zelden gebruikt voor analysedoeleinden. Het is zonde om er tijd en energie in te steken om deze data te prepareren. In het modern datawarehouse worden alleen data geprepareerd waarvan je vrijwel zeker bent dat die vaker gebruikt gaan worden in analyses.

### Business views

Business views zijn voorgeconfigureerde ‘vertaallagen’ die de gebruiker in staat stelt om zijn eigen analyse bij elkaar te slepen, te knippen en te plakken.

### Datasets

Datasets zijn op de gebruiker afgestemde sets van data uit verschillende bronsystemen. De gebruiker bepaalt hoe diep de data moet gaan. Een data scientist heeft voor complexe analyses immers vaak veel diepgaandere informatie nodig dan een gemiddelde businessgebruiker. De datasets kunnen putten uit zowel het modern datawarehouse als uit het data lake. Ze kunnen virtueel zijn (dat wil zeggen: tijdelijk) of persistent (ze blijven gedurende langere tijd in gebruik).

### Master Data Management & GDPR

Met Master Data Management regel je alles wat nodig is voor de governance op het gebruik van data, denk aan logging, definities, authenticaties, controle op datakwaliteit. Door dit goed te regelen, voldoe je automatisch ook aan de GDPR/AVG.

### Modules

Dit zijn tools waarmee je specifieke taken kunt doen. Voorbeelden van modules zijn:

- ▶ Dataverrijking. Denk bijvoorbeeld aan het toevoegen van een geografische locatie aan adresdata, zodat je op een kaart kunt mappen waar je medewerkers wonen of hoe groot je verzorgingsgebied is.
- ▶ Dataclassificatie, waarmee je op basis van tekst in vrije tekstvelden automatisch een classificatie toekent.
- ▶ Datakwaliteit. Er is een module waarmee je automatisch data kunt opschonen of ontbrekende data op basis van externe databronnen kunt aanvullen. En er is een module die kan automatisch kan herkennen of een veld op de juiste manier is ingevuld.
- ▶ Data-uitwisseling. Deze module heeft als taak om data uit te wisselen met externe organisaties, zodat er geen aparte interface ontworpen hoeft te worden tussen de bronsystemen en het systeem van de externe samenwerkingspartner.

Kortom, je kunt het zo gek niet bedenken of er is in het Modern Data Platform wel een module beschikbaar om extra acties uit te voeren op de data.

### CLOUD ALS BASIS

Het Modern Data Platform heeft de cloud als basis. Daar bovenop draaien verschillende databases, aangevuld met programmatuur die ervoor zorgt dat alles als één geheel samenwerkt en die extra functionaliteit biedt. Door de keus voor de cloud en het bijbehorende pay-per-use model wordt het betaalbaar om complexe analyses uit te voeren op basis van hele grote datasets. Bovendien is de cloud schaalbaar, zowel opwaarts als neerwaarts. Heb je extra capaciteit nodig voor een grote berekening? Geen probleem, je schaaft deze tijdelijk bij. En je stopt met betalen zodra de berekening klaar is en je de capaciteit niet meer nodig hebt. Uiteraard maken we uit oogpunt van wet- en regelgeving alleen gebruik van de Europese cloud.

# Eén platform voor al je informatievraagstukken

Het combineren van data uit verschillende bronnen voor andere dan managementtoepassingen is voor de meeste organisaties op dit moment ontzettend complex. Natuurlijk, ze werken met analytics software, maar dit betreft in negen van de tien gevallen alleen financiële, productie- en HR-data. Daarnaast maken ze gebruik van analysetools binnen applicaties. Dan ontbreekt de mogelijkheid om die data te combineren met data uit andere systemen. Voor de gecombineerde berekeningen is Excel nog altijd erg populair, ondanks het nadeel van de beperkte governance mogelijkheden. De kans dat er een foutje in zo'n Excelsheet sluipt is immers groot, zeker als het analyseresultaat uit de ene Excelsheet als input wordt gebruikt voor een nieuwe berekening in een andere Excelsheet.

## VERSCHILLENDE TYPEN GEBRUIKERS, ÉÉN PLATFORM

Aan deze problemen maakt het Modern Data Platform een eind. Het Modern Data Platform bedient alle typen gebruikers – management, medewerkers, data scientists – vanuit één platform. Een platform dat geïntegreerde tooling bevat om de data governance te regelen en zo te allen tijde 'in control' te blijven op zowel de data als de uitkomsten van de analyses die je doet op basis van die data. Op die manier maak je data analytics beheersbaar, en voldoe je en passant ook aan de geldende wet- en regelgeving zoals de GDPR/AVG.

## MAAK ANALYTICS BEHEERSBAAR EN BEHAPBAAR

Het Modern Data Platform neemt de bottleneck weg die nu nog ligt bij de BI-afdeling. Zij kunnen momenteel de toenemende vraag naar allerhande data-analyses vaak al niet aan, terwijl we nog maar aan het beginpunt staan van een intensiever gebruik van data. Ontwikkelingen als Internet of Things, Artificial Intelligence en Machine Learning maken toekomstige analysevragen veel data-intensiever en complexer. Daarom is het cruciaal om het beheersbaar en behapbaar te houden.

Dat is precies wat het Modern Data Platform doet met voorgeconfigureerde datasets, analysetools en governance instrumenten. Hierdoor kunnen managers en medewerkers zelf aan de slag met hun eigen analysevragen, zonder een beroep te hoeven doen op de schaarse data science capaciteit. De data scientists hoeven minder tijd te besteden aan toegang geven, vergaren en opschonen van data, maar kunnen zich richten op datgene waar ze écht goed in zijn: complexe analysevraagstukken beantwoorden.



## TRANSITIE NAAR DATAGEDREVEN ORGANISATIE

Het beschikken over een Modern Data Platform betekent overigens niet dat de organisatie ook direct datagedreven is. Processen en procedures zullen doorgaans moeten worden aangepast, maar bovenal is er een cultuurverandering nodig.

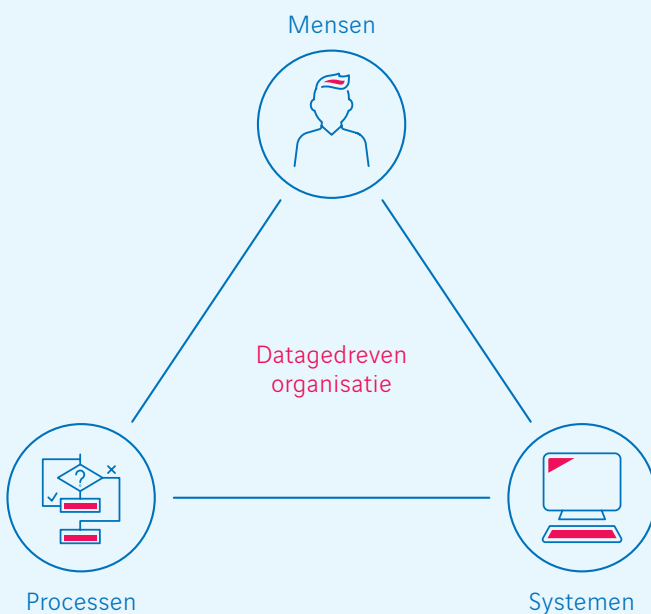
Medewerkers moeten doordrongen raken van het nut (en wellicht zelfs de noodzaak) van data. En weten hoe ze data goed kunnen vastleggen en op welke manieren zij zelf data kunnen gebruiken om inzichten te krijgen, beslissingen te nemen en snel te kunnen anticiperen op ontwikkelingen. In eerste instantie hebben medewerkers vaak het gevoel dat vrijheden worden ingeperkt.

Als medewerkers niets terug krijgen voor hun inzet om data beter vast te leggen, zal de animo om hier energie in te steken laag zijn. Hebben ze echter zelf baat bij hun inspanningen, dan is de kans dat ze meewerken en het onderwerp serieus nemen een stuk groter. Als ze merken dat data en informatieproducten sneller beschikbaar zijn, met een hogere kwaliteit en meer toegepast op het eigen werk zal het enthousiasme al snel toenemen.


### VOORDELEN VAN EEN MODERN DATA PLATFORM

- ▶ Eén platform voor al je informatievraagstukken.
- ▶ Nieuwe inzichten door data uit verschillende bronnen te combineren en in samenhang te analyseren.
- ▶ Self Service voor gebruikers; zij kunnen met hun eigen informatievragen aan de slag.
- ▶ Geïntegreerde governance, rekening houdend met wetgeving zoals de AVG.
- ▶ Voorgeconfigureerde datasets, analysetools en governance instrumenten maken de opzet en het gebruik eenvoudig.

**Wil je zelf graag stappen zetten richting een Modern Data Platform?** Wij komen graag bij je langs om hierover te praten. Je kunt ons bereiken via [info.bi.nl@axians.com](mailto:info.bi.nl@axians.com) of telefonisch via 088 597 55 00.







The best  
of ICT with  
a human  
touch

---

**axians**

Eemsgolaan 15  
9727 DW Groningen

Esp 120  
5633 AA Eindhoven

[info.nl@axians.com](mailto:info.nl@axians.com)