

# Van data-overdaad naar just-enough en just-in-time

Onderzoek: datahamsteren en de volwassenheid  
van data-engineering binnen grote organisaties  
mei 2023



# Inhoudsopgave

Voorwoord

Over het onderzoek

Thema I: onbenut datapotentieel voor realtime besluitvorming

Thema II: kloof tussen business en IT(-partner)

Thema III: organisaties gaan steeds bewuster om met data

Conclusies

Een volgend niveau van datavolwassenheid: aanbevelingen

Opdracht aan onszelf, oproep aan de sector

Over AMIS Conclusion

## Voorwoord



In 2020 publiceerde AMIS Conclusion een onderzoek naar de status van IT in Nederland. De conclusie was duidelijk: IT-systemen zijn te ingewikkeld geworden.

Zo ingewikkeld zelfs dat het IT-landschap vaak niet meer de ondersteuning levert die men verwacht.

Het advies luidde: kies voor een aanpak met de juiste strategie en langetermijnvisie, houd de focus op je doelen en streef naar eenvoud en slagkracht.

Begin 2023 heeft AMIS Conclusion dit onderzoek nogmaals laten uitvoeren, maar dan gericht op de bruikbaarheid van de datavoorziening van grote organisaties. Daar komt een vergelijkbaar beeld uit naar voren: data wordt steeds ingewikkelder. Zo ingewikkeld zelfs dat organisaties worstelen met het vinden van de juiste data en het vertalen naar informatie om beslissingen op te baseren.

Eén van de complexiteiten is de overweldigende hoeveelheid data die bedrijven verzamelen. Meer data betekent immers ook meer moeite om daar waarde uit te halen, meer noodzaak tot goede data-engineering en volwassen dataplatformen, meer behoefte aan dataspecialisten en meer kosten. En dit frustreert de digitale slagkracht van organisaties die een datagedreven strategie nastreven.

Dit onderzoek, dat is uitgevoerd door Markteffect, brengt de problemen in kaart van digital leaders bij grote organisaties op het vlak van zowel data-overdaad, dataverwerking als dataplatformen. Uit het onderzoek blijkt onder meer dat:

- Slechts 39 procent van de respondenten oordeelt dat de bijdrage van de data aan de bedrijfsdoelstellingen groter is dan de kosten.
- 12 procent stelt dat de verzamelde data in het geheel niet van waarde is voor de organisatie.
- 42 procent niet altijd tot de benodigde inzichten komt door de immense hoeveelheid data die ze hebben.
- 40 procent niet (precies) weet wat de kwaliteit en herkomst van de data is.

De drie-eenheid van realtime data – beschikbaarheid, betrouwbaarheid en actualiteit – is essentieel om waarde uit data te halen – maar blijkt in de praktijk een grote bottleneck.

Centraal in het onderzoek stond ook de kloof tussen business en IT. De resultaten laten zien dat deze twee werelden te weinig van elkaar begrijpen én dat zowel eigen IT-afdelingen als externe IT-partners nog een hoop te winnen hebben wat betreft het effectief genereren van waarde uit de beschikbare dataplatformen, integraties en IoT-oplossingen.

Goed, je leest er alles over in dit onderzoeksrapport. Niet alleen bespreken en duiden we de onderzoeksresultaten. Afsluitend geven we ook waardevolle inzichten en concrete handvatten om het datapotentieel beter te benutten. We hopen dat organisaties daardoor meer controle krijgen over hun data, volwassen met data omgaan, waardevollere data-toepassingen bouwen én de volgende stap zetten naar machine learning en AI-gebaseerde modellen.

Tot slot roepen we de IT-sector op – en manen we onszelf tot – het effectiever, professioneler en beheersbaarder maken van de datalandschappen van klanten. Zodat we met zo'n allen door de bomen het databos weer kunnen zien en daar meer waarde uit kunnen halen.

Ik wens je veel leesplezier!

André van Dalen  
Algemeen directeur AMIS Conclusion



# Over het onderzoek

## Het probleem

Organisaties kampen tegenwoordig met een enorme stortvloed aan data, die met de dag groter wordt. AMIS Conclusion ziet bij veel klanten dat ze neigen tot datahamsteren: grote hoeveelheden digitale data verzamelen en bewaren, zonder dat (precies) duidelijk is wat het doel daarvan is.

De datastortvloed wordt met de dag groter doordat er meer en meer databronnen, datasoorten en data-verwerkende systemen komen. Met name video- en audiobestanden, externe data (zoals gegevens over het weer en de koersen van grondstoffen) en ruwe IoT-data hebben het datahamsteren in een stroomversnelling gebracht. Daar bovenop komt dat organisaties hun meetfrequentie verhogen en steeds meer behoefte hebben aan actuele of zelfs realtime data.

Waar organisaties hopen dat meer data leidt tot meer en betere inzichten, is het effect vaak juist averechts. Het datateveel zorgt er onder meer voor dat ze flexibiliteit en slagkracht missen, worstelen met het genereren van waardevolle inzichten en kampen met compliancy-risico's. Last but not least brengt het (toren)hoge kosten voor het opschonen, opslaan en beheren van de data met zich mee.

## Het onderzoek

Om het groeiende data-overdaadprobleem dat we bij klanten signaleren in kaart te brengen en te kunnen duiden, hebben we onafhankelijk onderzoeksbureau Markteffect gevraagd om grotere organisaties over deze thematiek te ondervragen.

De centrale vragen in het onderzoek zijn:

- Kunnen organisaties de data die ze verzamelen optimaal benutten?
- Hoeveel van de verzamelde data zetten ze succesvol in?
- Weten ze eigenlijk überhaupt nog wel wat ze allemaal verzamelen en waarom ze data nodig hebben?
- Overtreffen de baten de kosten?
- Hoe is de samenwerking tussen IT en business? Begrijpen deze twee schakels elkaar wel?
- Hoe is het gesteld met de effectieve en beheersbare verwerking van al deze data?

## Het doel

Dit onderzoeksrapport geeft niet alleen inzicht in de datastaat van organisaties, maar reikt ook handvatten aan om het datapotentieel beter te benutten. Daarmee hoopt AMIS Conclusion organisaties op weg te helpen om (meer) waarde te halen uit de inzet van (realtime) data. En ze te ondersteunen bij het bereiken van een volgend niveau van datavolwassenheid, dat blijvend geborgd is in zowel goede data-engineering practices als een efficiënt dataplatform.

Daarnaast willen we graag – zoals we uiteenzetten in onze aanbevelingen – bewustzijn creëren binnen de IT-sector. Met z'n allen hebben we, om het dramatisch te formuleren, een datamonster van Frankenstein gecreëerd. Laten we dus ook met z'n allen toewerken naar een toekomst van lean data, waarin we op agile wijze snel, trefzeker en met precies de juiste hoeveelheid data impactvolle data-oplossingen bouwen.

## De respondenten

Het onderzoek is gehouden onder 150 digital leaders en digital change makers, die op managementniveau meebeslissen in (digitale) transformatietrajecten bij datagedreven organisaties met minimaal 250 fte.

Aantal FTE		Functieniveau	
250 tot 500	32%	Management	67%
500 to 1.000	19%	Senior-management	29%
Meer dan 1.000	49%	Lid van de directie / Raad van Bestuur	5%

## Geslacht



## Leeftijd





# Thema I: onbenut datapotentieel voor realtime besluitvorming

## Geringe beschikbaarheid, effectiviteit en efficiëntie

De respondenten geven aan dat gemiddeld zo’n 62 procent van de beschikbare data wordt gebruikt voor besluitvorming door de organisatie. In andere woorden: ruim een derde van de data ligt ergens te verstoffen – of erger: strooit zelfs zand in de raderen van de datamachine. Met alle nadelen en risico’s van dien. Zo bezien verbaast het niet dat de onder-vraagden harde noten kraken over de effectiviteit van de data: maar iets meer dan de helft van hen (56 procent) is van mening dat alle databronnen effectief gebruikt worden. Met de efficiëntie is het nog erger gesteld: 72 procent stelt dat de doelmatigheid van databronnen beter kan.

*“Ruim een derde van de data ligt ergens te verstoffen – of erger: strooit zelfs zand in de raderen van de datamachine.”*

## Data(verzameling) die z’n doel mist

Het geringe gebruik en de tegenvallende effectiviteit en efficiëntie lijken vooral voort te komen uit het feit dat een duidelijk doel voor deze data ontbreekt. Twee van de drie professionals (68 procent) vindt namelijk dat het verzamelen, verwerken en beschikbaar stellen doelgerichter en sneller kan.

## Meestal wel, soms niet van waarde

De respondenten geloven over het algemeen wél in de waarde van de data: 88 procent meent name-lijk dat de beschikbare gegevens een positief effect hebben op de bedrijfsvoering. Laten we het eens

omkeren: frappant genoeg vindt één op de acht respondenten (12 procent) dus dat de organisatie niets of niet genoeg aan de verzamelde data heeft.

## Beperkte bijdrage aan bedrijfsdoelstellingen

Eén van de opvallendste uitkomsten van dit onder-zoek? Een magere 39 procent van de respondenten stelt dat de bijdrage van de data aan de bedrijfsdoel-stellingen groter is dan de kosten die het verzamelen, opschonen, opslaan, verversen, beheren en beschik-baar houden van data met zich meebrengen. 40 procent zegt dat kosten en baten tegen elkaar weg zijn te strepen. En een verbazingwekkende vijfde van de digitale leiders – oftewel 21 procent – zegt zelfs dat de kosten qua manuren en investeringen groter zijn dan de bijdrage aan de bedrijfsdoelstellingen.

*“Meer dan de helft van de respondenten oordeelt dat de kosten voor het verwerken en beschikbaar stellen van data qua manuren en investeringen groter zijn dan de bijdrage aan de bedrijfsdoelstellingen.”*

## AMIS CONCLUSION SIGNALEERT...

*...dat zowel tijd als investeringen gemakkelijk onbe-heersbaar worden. Dat is met name zo wanneer een geautomatiseerde aanpak via een dataplatform ontbreekt. Denk maar aan de grote hoeveelheid hand-matig werk die kruipt in het verversen en opschonen van de data. De gemiddelde data-analist is pak ’m beet zo’n 80 procent van z’n tijd kwijt met het ophalen en geschikt maken van de data, voordat hij aan zijn*

*echte werk kan toekomen. Boven-dien: wordt data niet automatisch ververst, dan beschikt men niet over actuele data – wat de besluitvorming negatief beïnvloedt.*

*Los van de arbeidsinvesteringen lopen ook de kosten voor bijvoorbeeld opslag, licenties en cloud-services voor het verversen van de data snel op naarmate de dataverwerking niet op vakkundige wijze plaatsvindt.*

*De grootste uitdaging? Dat is in ieder geval níét om eenmalig een dataset te verzamelen. Nee, het is juist het herhaalbaar en betrouwbaar beschikbaar stellen van actuele data. Deze data-actualiteit zorgt wel voor onderscheidend vermogen. Vergelijk het met verkeersinformatie: up-to-date data laat je snel je doel bereiken, maar als het maar één keer per week wordt bijgewerkt verliest het zijn waarde.*

## Data-overdaad schaadt

Bijna twee van de drie organisaties (63 procent) verza-melt bewust méér data dan zij momenteel gebruiken, aangezien de gegevens mogelijk in de toekomst nog van waarde zijn. Dit is niet geheel zonder risico: zo stelt 42 procent van de respondenten dat zij niet altijd tot de benodigde inzichten komen door de grote hoeveelheid data die ze hebben.

## AMIS CONCLUSION ZIET...

*...in de praktijk dat data die wordt bewaard voor de toekomst over het algemeen niet van voldoende detailniveau en de juiste meetfrequentie is. Dat komt dan omdat er op het moment van verzamelen nog geen doel voor was. Deze data hindert echter wel het dataverwerkingsproces én legt beslag op de resources van het dataplatform.*

*Ook geeft deze extra data onnodig veel complexiteit bij het identificeren en categoriseren. De hoeveelheid (met name IoT-telemetrie)data kan bovendien de totale verwerking ernstig vertragen. Daarom adviseert AMIS Conclusion om, zoals we bij de aanbevelingen*

*uitleggen, data uitsluitend doelgericht te verzamelen en bij de verwerking de principes van software- en data-engineering toe te passen.*

## Beschikbaarheid en kwaliteit van data

27 procent van de respondent zegt dat men niet altijd toegang heeft tot de verzamelde data. 31 procent meent dat data niet snel genoeg voorhanden is op de momenten waarop het nodig is. En slechts 60 procent geeft aan dat het gebruikers volledig duide-lijk is wat de herkomst en kwaliteit van data is.

## AMIS CONCLUSION VINDT...

*...het ernstig dat de respondenten aangeven dat de data-actualiteit te wensen overlaat én dat de herkomst en kwaliteit van de data in het gros van de gevallen niet duidelijk is. Steeds meer beslissingen worden immers op data gebaseerd. Zowel de herkomst, de traceerbaarheid als het meetmoment zijn cruciale fundamenteen voor datagedreven werken. Datas-cientists en machine learning-specialisten moeten te allen tijde kunnen vertrouwen op de juistheid van de dataproducten waarmee ze werken. AMIS Conclusion zorgt dat deze hoogkwalitatieve data er altijd is – zoals water uit de kraan – zodat zij hun kostbare tijd kunnen besteden aan het genereren van toegevoegde waarde.*





## Thema II: kloof tussen business en IT(-partner)

### Begrijpen en begrepen worden

79 procent van de beslissers begrijpt wat de eigen IT-afdeling doet op het gebied van dataverwerking, terwijl dat voor IT-partners iets lager ligt (67 procent). 70 procent voelt zich begrepen door de eigen IT-afdeling, terwijl het aantal respondenten dat zich begrepen voelt door de IT-partner op 74 procent ligt. Ook al zijn het slechts 4 procentpunten, markant is het verschil hoe dan ook.

### Dataverwerking en bijdrage aan doelstellingen

Nog zo'n opvallend feit: waar 74 procent van de digitale leiders het gevoel heeft dat de IT-partner voldoende kennis heeft van de doelstellingen van de organisatie, ligt dit voor de eigen IT-afdeling met 70 procent iets lager. Daar staat tegenover dat net iets meer respondenten aangeven dat de eigen IT-afdeling op datagebied een waardevolle bijdrage levert aan het behalen van de bedrijfsdoelen: 77 procent, tegenover 75 procent voor IT-partners.

### Meer begrip, meer waarde uit data

De respondenten geloven dat het beter begrijpen van de bedrijfsdoelstellingen leidt tot meer data-succes. 46 procent stelt namelijk dat meer begrip van

de doelstellingen leidt tot effectiever datagebruik bij de IT-afdeling, terwijl dit voor IT-partners 60 procent is.

### AMIS CONCLUSION VERMOEDT...

*...dat er meerdere redenen zijn waarom respondenten*  
1) zich beter begrepen voelen door een IT-partner en  
2) stellen dat IT-partners de bedrijfsdoelstellingen beter begrijpen.

*Allereerst zal een goede IT-partner vragen naar het eindresultaat. Wat wil de business bereiken en welke dataproducten horen daarbij? Ook hebben IT-partners vaak meer drive om impact te maken. Ze kijken vooral heel kritisch: wat moet er nou écht gebeuren? Eigen IT-afdelingen hebben juist sneller de neiging te doen wat de business vraagt en vinden het lastiger om 'nee' te zeggen. Plus: omdat externe partijen meer kosten, kijken organisaties beter naar wat ze vragen. En dat leidt tot strakker afgebakende opdrachten.*

## Thema III: organisaties gaan steeds bewuster om met data

### Data-engineering als speerpunt

Uit het onderzoek blijkt dat data-engineering en data-governance bij veel organisaties op de agenda staan. 49 procent geeft aan dat de eigen IT-afdeling verantwoordelijk is voor het databeleid, 29 procent heeft hiervoor een externe partner en 23 procent zegt zowel een eigen afdeling als een externe partner te hebben.

### Goede plek voor eigenaarschap van data

81 procent van de grotere organisaties zegt dat het eigenaarschap van de data en het databeleid zijn geborgd op board- en/of directieniveau. Respondenten geven aan dat er beleid is opgesteld ten aanzien van:

- Toegang tot data (70 procent)
- De bewaartermijn van data (63 procent)

Toepassingen voor verschillende databronnen (55 procent).

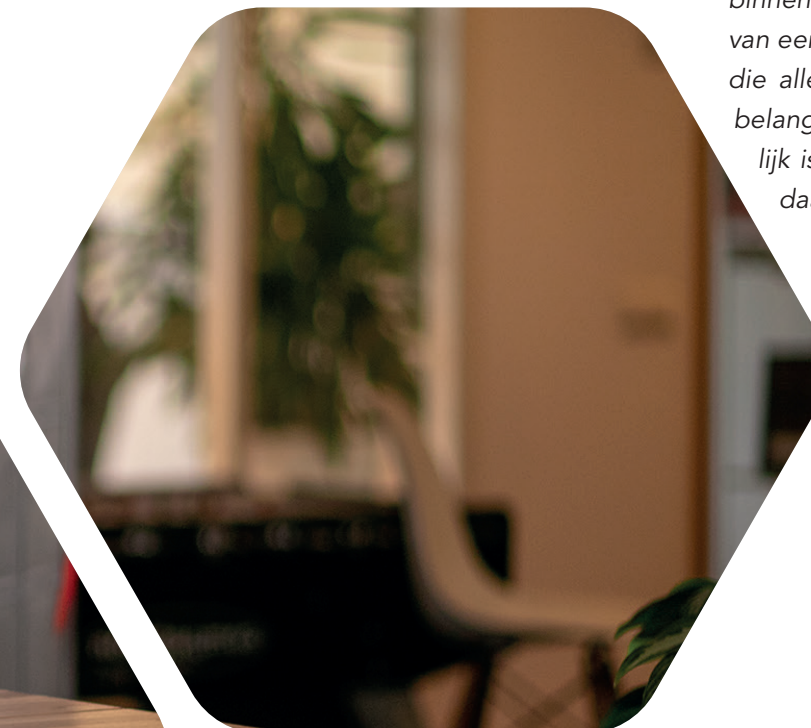
### AMIS CONCLUSION UIT...

*...zorgen over het niveau van data-engineering van organisaties. Veel organisaties zeggen weliswaar datagedreven werken als speerpunt te hebben, maar in de praktijk is er vaak een hoop mis. Ze kampen bijvoorbeeld met een gebrek aan duidelijk beleid, een kloof tussen beschikbare data en de bedrijfsdoelen, inadequate datakwaliteit, beperkte beschikbaarheid van data en/of uitdagingen op het gebied van security en compliancy.*

*De achterliggende oorzaak is vaak onvoldoende focus op een volwassen dataplatform en serieuze data-engineering practices. De datacompetentie is bij veel organisaties gegroeid vanuit het datawarehouse. Daarbij geldt een statische dataverwerkingsaanpak met een lage verversfrequentie. Datagedreven werken vergt juist een engineering-aanpak die vergelijkbaar is met die van software-engineering: professioneel, lage fouttolerantie, hoge frequentie, goede governance en monitoring, continu testen en volledig geautomatiseerd.*

*Dit vereist een andere aanpak, volledige ondersteuning van het senior management en investeringen in zowel goede dataproducten als nieuwe competenties binnen het datateam. Belangrijk daarbij is het hebben van een datasteward, oftewel iemand uit de business die alles rondom de data begrijpt. Ook is het van belang dat iemand vanuit de business verantwoordelijk is voor de governance, aangezien de business daadwerkelijk iets met de data doet.*

*Organisaties die voldoen aan deze vereisten, creëren een beheersbaar datalandschap met overzichtelijke, flexibele, consistente, betrouwbare en vooral effectieve dataproducten.*





# Conclusies

Veel organisaties denken dat ze het goed voor elkaar hebben omdat ze in de cloud zitten of beschikken over een datalake. Wat in ieder geval duidelijk wordt uit deze onderzoekresultaten: ze hebben geen moeite met het verzamelen van data. Nog steeds lijken organisaties te denken dat meer data automatisch leidt tot meer inzichten. En daar gaat het wringen. Van het tijdperk van big data lijken we nu namelijk een fase van big, fat data te zijn ingegaan – waarbij het teveel aan data juist een ongezonde ballast wordt.

## Verdrinken in de datazee

Het onderzoek suggereert dat veel organisaties bijna verdrinken in de grote datazee – of beter: hun eigen datalake. Verzamelen ze wel de juiste data? Of juist te véél, gezien de kosten-batenafweging? Kunnen ze de juiste data op het juiste moment vinden, gekeken naar bijvoorbeeld meetmoment en -frequentie? Hebben ze middelen om snel waarde uit deze data te halen? Zijn ze met het dataplatform in staat om realtime datadiensten aan te bieden? En beschikken ze over genoeg flexibiliteit om snel mee te bewegen met de markt en nieuwe producten te ontwikkelen? Veel respondenten blijken te worstelen met deze datathema's.

De volgende feiten ondersteunen de constatering dat grote organisaties data momenteel eerder zien als uitdaging-met-potentie dan als concrete business-enabler:

- Slechts 39 procent oordeelt dat de bijdrage van de data aan de bedrijfsdoelstellingen groter is dan de kosten.
- 12 procent stelt dat de verzamelde data níét van waarde is voor de organisatie.
- Grote datazees zorgen voor grote problemen: 42 procent van de respondenten zegt niet altijd tot de benodigde inzichten te komen door de immense hoeveelheid data die ze hebben.
- 40 procent van de respondenten weet niet (precies) wat de herkomst en kwaliteit van de data is.

## Onbegrip tussen business en IT(-partner)

Doelen, strategieën en processen die worden afgestemd tussen business- en IT-afdelingen: het blijft een knelpunt binnen de IT – en een heikel thema op datavlak. Aan het wederzijds begrip tussen beide werelden valt, getuige dit onderzoek, nog een hoop te verbeteren.

Of het nu komt door gebrekkige communicatie, verschillende doelen, culturele verschillen, onvoldoende data-engineering-competenties, een kortetermijnfocus of de razendsnelle technologische veranderingen: de kloof tussen business en IT lijkt maar niet minder te worden. In de volgende paragraaf doen we aanbevelingen om de brug te slaan tussen deze twee werelden.

Het valt op dat de eigen IT-afdelingen op datavlak niet veel beter – en in sommige gevallen zelfs slechter – scoren dan externe IT-partners. Immers:

- 74 procent van de respondenten voelt zich begrepen door de IT-partner, tegenover 70 procent door de eigen IT-afdeling.
- Eveneens 74 procent van de digitale leiders heeft het gevoel dat de IT-partner de bedrijfsdoelstellingen van de organisatie begrijpt, terwijl dit voor de eigen IT-afdeling 'slechts' 70 procent is

Typische uitkomsten, je zou immers verwachten dat de eigen IT'ers meer domeinkennis hebben. Welbeschouwd is het niet onlogisch. Externe IT-partners hebben immers meer drive om écht impact te maken en stellen zich kritischer op. Interne IT-afdelingen hebben, kort door de bocht gesteld, vaak meer een 'u vraagt, wij draaien'-mentaliteit. Daar komt bovenop dat externe IT-partners meer kosten, waardoor opdrachtgevers scherper zijn in hun vraagstelling. En dat resulteert in beter afgebakende vragen.

## Data-engineering: wel prioriteit, weinig resultaat

Organisaties stellen dat data-engineering en data-governance hoog op de agenda staan. 81 procent van de respondenten geeft bijvoorbeeld te kennen dat het professioneel omgaan met data is geborgd op board- en/of directieniveau. Blijkbaar worstelen grote organisaties om vorm te geven aan effectieve

data-engineering en dataplatforms, getuige de vele uitdagingen waar ze volgens dit onderzoek mee kampen.

## Lange weg van business reporting naar voorspellende dataproducten

Het onderzoek laat zien dat veel grotere organisaties data hoofdzakelijk inzetten voor business reporting, waarbij datawarehouses nog de nadruk hebben. Een stuk minder organisaties gebruiken data om naar het hier en nu te kijken: kunnen we bijsturen en ingrijpen? Wij zien dat data-engineering en dataplatforms het fundament zijn voor dit vermogen. Goed bruikbare, betrouwbare, actuele en kwalitatief hoogwaardige dataproducten zijn hiervoor essentieel.

Pas als deze stap is gezet zijn organisaties in staat om voorspellende modellen te ontwikkelen. Daar valt dus nog een wereld te winnen, vooral onder de motorkap op het gebied van data-engineering.





# Een volgend niveau van datavolwassenheid: aanbevelingen

Te veel data(bronnen), datalakes die overstromen, gebrekkige datakwaliteit, inzichten die daardoor worden gemist, geen flexibiliteit en slagkracht, maar wél torenhoge kosten: volgens de beloftes zou data het nieuwe goud zijn, maar voor veel organisaties betekent het eerder zand in de raderen.

Voor goud in plaats van zand moet je als organisatie een steeds hoger niveau van datavolwassenheid bereiken: beginnend bij ad hoc en dan via respectievelijk reactief, proactief en volledig beheerd met een focus op datagedreven werken naar het summum: een cultuur waarin data de basis is voor continu verbeteren en innoveren. En dat vergt een professionele aanpak voor het ontsluiten, integreren, opschonen en beschikbaar houden van de dataproducten. Met name via professionele data-engineering en goed schaalbare dataplatforms kunnen organisaties zich onderscheiden.

Op welke wijze kun je als organisatie het onbenutte datapotentieel verzilveren en het niveau van je datavolwassenheid verhogen? Hoe kun je gegevens beter raffineren en zo bruikbare dataproducten en halffabricaten aanbieden? Wat is het geheim achter minder data verzamelen maar meer inzichten creëren? En hoe kun je je data dusdanig prepareren en voorbewerken dat je datascientists en machine learning-specialisten direct waarde toevoegen?

Op basis van ruim 30 jaar ervaring in de wereld van data en integratie doen we negen aanbevelingen.

## #1 Begin achteraan en stel duidelijke strategieën en doelen op

Wat vaak misgaat: organisaties duiken in hun datazee om te kijken wat er opvalt. Het is een illusie dat bergen data, wat AI en een hoop rekenkracht automatisch resulteren in waardevolle inzichten. Veel beter is om eerst duidelijke strategieën en doelen op te stellen voor zowel het verzamelen, beheren en analyseren van data. Belangrijk is dat alles in lijn is met de bedrijfsdoelen. De doelen moeten zo zijn geformuleerd dat business- en data-teams er sámen aan kunnen werken.

Zo kan ieder team zich richten op waar het goed in is. Datateams maken met data-engineering betrouwbare dataproducten. En businessteams bouwen daarop aan de hand van algoritmes beslislogica die ingrijpt op de processen van een organisatie om zo meer waarde toe te voegen.

Daarbij geldt: begin achteraan en werk achterstevoren vanuit het businessprobleem. Wat zijn, gezien vanuit de bedrijfsdoelen, de vragen die je wilt beantwoorden? Op basis van deze vragen bouw je een businesscase en definieer je welke informatie nodig is om beslissingen te nemen. Dan pas kijk je welke data hiervoor nodig is en zorg je dat dit als een betrouwbaar dataproduct aanwezig is in je dataplatform.

## #2 Start klein en werk stapje voor stapje

Het beste is om klein te beginnen en te werken in behapbare, productierijpe iteraties. Alles wat je releaset, dient een concrete businessbehoefte te vervullen én veilig zijn om te gebruiken. Voordat je toe bent aan een nieuwe stap, moet je een relatief kleine hoeveelheid data door en door analyseren.

Ga met een stofkam door je datalandschap heen en stel vast wat je allemaal wel en niet hebt. Neem alleen de data in beheer die een écht goede case heeft. Let goed op dat je focust op de juiste data, wat onder meer wordt bepaald door de bron van de data en het type oplossing. Door de meest cruciale data te identificeren, te garanderen dat die past bij het doel én daar vervolgens op te focussen, verminder je de hoeveelheid onnodige data die je moet beheren.

Belangrijk is wel dat je voor dat ene stukje data een end-to-end werkende versie van een dataplatform maakt. Door hier alle eisen van data-engineering op toe te passen maximaliseer je de kwaliteit, betrouwbaarheid en herhaalbaarheid van het proces, zonder een hele lading data met je mee te slepen. Bij uitbreiding van het dataplatform zijn deze principes van data-engineering integraal in het proces verweven. Zo blijven de kwaliteit, transparantie, governance en betrouwbaarheid behouden, terwijl het platform meeschaalt met de stijgende databehoefte.

Door zo'n iteratieve, agile aanpak is er steeds voortgang, worden inzichten op tijd geproduceerd en

kan het datateam tijdig bijsturen. Is een dataproduct operationeel, dan kun je dit met hetzelfde agile proces bijsturen zodra de bedrijfsstrategie of de markt daarom vraagt.

## #3 Teamwork makes the data dream work

Een goed datateam samenstellen en optimaal laten samenwerken is een herculeswerk. Idealiter is het team samengesteld uit een data-engineer die begrijpt wat de bron van data is en in welke vorm data binnenkomt, een data-scientist die snapt hoe waardevolle informatie uit data kan worden gehaald en een team met DevOps-specialisten die snel een data-analyse lifecycle kunnen leveren.

Afhankelijk van de grootte en het type van de organisatie moeten er één of meerdere data-stewards, data-owners en/of data-architecten worden aangesteld. Een data-steward garandeert het beheer en de kwaliteit van specifieke gegevenssets, een data-owner is verantwoordelijk voor de gegevens van een bedrijfsproces of applicatie en een data-architect ontwerpt de structuur, de integratie en het beheer van gegevens binnen een organisatie.





Het datateam moet:

- de businessdoelen begrijpen en kunnen vertalen naar databronnen
- bekend zijn met metadata- en masterdata-management
- principes en best practices op het gebied van data-engineering toepassen aan de hand van tools en methoden voor het beheersbaar opschonen en combineren van data
- voldoen aan data lineage, oftewel het nauwkeurig en stapsgewijs bijhouden en documenteren van het pad van de data door verschillende systemen en processen
- bekend zijn met datamanagement, denk aan data-governance policies, het ontwikkelen van standaarden voor datakwaliteit, het standaardiseren van dataformaten en het opstellen van data-documentatie.

#### #4 Datamesh als web van (domein)specialisten

Datateams waarvan de leden niet over domeinkennis beschikken, hebben vaker moeite om de data waarmee ze werken te begrijpen en te interpreteren. Zonder domeinkennis is bijvoorbeeld vaak niet duidelijk wat nu precies ROT-data ('redundant, outdated en trivial') is en wat niet.

Een oplossing voor dit probleem is het concept datamesh. Binnen dit gedecentraliseerde en domeingeoriënteerde model is iedere businessunit verantwoordelijk voor zijn eigen dataproducten, waarbij ieder domein een data product owner heeft. De verschillende teams moeten kunnen beschikken over de dataproducten van de andere teams, zodat ze beter kunnen samenwerken, makkelijker inzichten kunnen delen en minder dubbel werk hoeven te verrichten.

Bij datamesh fungeert het datalake als een soort machinekamer waar de dataverwerking plaatsvindt. De eigenaars van de datasystemen (oftewel de domeinexperts) en de datascientists kunnen vervolgens achter het stuur gaan zitten om koers te zetten naar bruikbare inzichten.

Met datamesh is het mogelijk om met de producten in de machinekamer behapbare en betrouwbare dataproducten te maken, waardoor teams snel en daadkrachtig waarde uit data halen zonder dat business en IT elkaar in de weg zitten. Daarmee bevordert deze aanpak samenwerking, wendbaarheid en innovatie binnen de hele organisatie.

#### #5 Datakwaliteit: deurmat ervoor en bezem erdoor

Belangrijk om de datazee niet te laten overstromen en een zogeheten analysis paralysis te voorkomen: zorg dat er geen rotzooi door de voordeur binnenkomt. Beter is om een soort deurmat neer te leggen die de datakwaliteit garandeert. Bij binnenkomst kan bijvoorbeeld procesdata worden gevalideerd: is het waardevol, consistent, relevant en correct?

Daarnaast moet de bezem door het datalandschap worden gehaald. Het opschonen van data kan door dubbele, verouderde en irrelevante data te verwijderen. Dit verbetert niet alleen de kwaliteit van je data, het vermindert ook de hoeveelheid én geeft een boost aan de beheersbaarheid. Goed om te beseffen: organisaties die succesvol zijn in digitale transformatie, besteden meer tijd aan het beheersen en opschonen van data dan aan data-analyse en het bouwen van algoritmes.

#### #6 Beheersbaar, herhaalbaar, betrouwbaar

Heb je mooie dataproducten staan met een datakwaliteit om door een ringetje te halen? Mooi. Dan is nu de taak aan de data-engineers om alles beheersbaar, herhaalbaar en betrouwbaar te houden de komende jaren. Uit ervaring weten we: dit is een vak apart en een expliciet specialisme binnen data-engineering. Monitor daarbij ook vooral de processen voor ontsluiting en verwerking van de data. Gaat het mis, dan wil je dat immers direct weten. Door principes van software-engineering toe te passen op data-engineering krijg je volledige beheersbaarheid en traceerbaarheid binnen je dataproducten.

#### #7 Technologie is slechts een enabler

Technologie kan nooit een panacee zijn, wel een enabler. Een datalake is bijvoorbeeld geen magische doos waar je alles zomaar in kunt gooien. Als gecentraliseerde opslagplaats voor data van verschillende bronnen is het eerder een bouwblok dan een eindproduct. Toch zien we vaak in de markt dat er in plaats van een bedrijfsdoelen-first een technologie-first mindset heerst ('Wat voor merk datalake wilt u?').

Tuurlijk, technologieën zoals datalakes zijn nuttige hulpmiddelen voor organisaties die met grote hoeveelheden data werken. Heb je de – vaak stevige – uitdaging overwonnen om gegevens uit verschillende systemen en databronnen te integreren in

het datalake? Dan moet je als organisatie alsnog beschikken over de juiste analytische tools en expertise om waardevolle inzichten uit de data te halen. Daar bovenop moet je een effectief systeem voor gegevensbeheer en -validatie implementeren.

Tot slot: een datalake vereist een aanzienlijke verandering in de manier waarop je als organisatie met data omgaat. En dat vergt investeringen in het opleiden en aantrekken van gekwalificeerde mensen én een verschuiving binnen de organisatiecultuur.

#### #8 Wees voorzichtig met data op de plank

Als organisatie kan het interessant zijn om data te bewaren voor toekomstig gebruik. Met meer historische data kun je bijvoorbeeld bedrijfsprocessen verbeteren, klantinzichten verbreden, nieuwe dataproducten ontwikkelen of een steviger fundament leggen voor voorspellende analyses.

Op papier. AMIS Conclusion signaleert namelijk dat data die onze klanten op de plank hebben liggen, later vaak onbruikbaar is om alsnog in te zetten. De gegevens zijn in de regel onvolledig of ongedetailleerd, bijvoorbeeld omdat de verversfrequentie te laag was.

Ons advies luidt daarom: staat je dataproduct eenmaal, hanteer dan een strikt weggooiregime. Zit je dienst echter nog in de vormende fase, dan is het handig om wat meer data te bewaren. Zo houd je de ruimte om hypothesen te toetsen en causale relaties te onderzoeken.

#### #9 Houd het simpel

Een conclusie uit dit onderzoek – en dus een leerpunt voor zowel eigen IT-afdelingen als externe IT-partners – is dat lang niet alle respondenten begrijpen wat de eigen afdeling of partner doet op datavlak. Een belangrijke oorzaak: techneuten maken het vaak onnodig complex. Uit het vorige onderzoek van AMIS Conclusion bleek bijvoorbeeld dat de huidige IT-oplossingen te ingewikkeld zijn geworden.

Voor datateams geldt daar bovenop dat mensen in de business vaak weinig of geen expertise op het vlak van data-analyse hebben. Ons advies? Of het nu gaat om tools, methodes of jargon: houd het simpel. Op die manier blijf je zo dicht mogelijk bij de (business) teams voor wie de inzichten bedoeld zijn.



# Opdracht aan onszelf, oproep aan de sector

AMIS Conclusion is al ruim drie decennia toonaangevend op het gebied van dataplatformen, data-engineering en realtime integraties. Dit onderzoek laat zien dat veel grote organisaties gehinderd worden in hun besluitvaardigheid en flexibiliteit door een data-overdaad en problemen op het gebied van de beheersbaarheid van datastromen. Data is vaak niet actueel, niet van de juiste kwaliteit of niet betrouwbaar.

Eén van de achterliggende oorzaken: door de focus op big data hebben IT-leveranciers klanten de afgelopen jaren (of zelfs decennia) verleid tot het verzamelen van grote bakken data en vervolgens tot het afnemen van allerlei concepten om die data te transformeren tot inzichten.

Bij AMIS Conclusion zijn we altijd kritisch, ook op onszelf – wij zijn pas tevreden als de verzamelde data ook daadwerkelijk nuttig is en wordt gebruikt. Misschien hadden wij als sector minder moeten focussen op het verzamelen van data en meer op het bruikbaar consumeren van data in de vorm van kwalitatief hoogwaardige dataproducten. En dat kan alleen door als een gatekeeper bij de voordeur eisen te stellen aan de wijze waarop de data door het IT-landschap beweegt – en vervolgens direct vast te

stellen hoe de inkomende data uiteindelijk geconsumeerd wordt.

Gezien de resultaten van dit onderzoek zijn wij voortaan scherper. Hoe kunnen we de datawaterval stoppen voor klanten om overstromende datameren en lekkages te voorkomen? Wat we in ieder geval gaan doen vanaf nu: actief aan klanten vragen welke data ze voortaan níét meer gaan bewaren.

Daarbij roepen we andere IT-leveranciers op om ook een kritischere houding in te nemen. De IT-sector heeft veel geld verdiend met onder meer het bouwen van dataproducten en het opslaan en back-uppen van data. Nu zijn we het aan onze stand verplicht om klanten te helpen om hun datalandschap bruikbaar, effectiever en beheersbaarder te maken.



# Over AMIS Conclusion

AMIS Conclusion geeft stuwkracht aan organisaties door data te laten stromen tussen sensoren, enterprise applicaties en zowel interne als externe databronnen. Onze oplossingen liggen vaak op het gebied van realtime en bijna-realtime data, IoT, data- en integratie-engineering, datagedreven toepassingen en dataplatforms. Dankzij onze oplossingen zijn organisaties in staat om direct in te grijpen in processen. Hoe geavanceerd onze oplossingen ook zijn, we houden het altijd zo simpel mogelijk.

AMIS Conclusion is, we zeiden het al, altijd kritisch. Ook richting klanten. Wij doen alleen wat nodig is. Op basis van de strategie en bedrijfsdoelstellingen bepalen we welke informatie is vereist voor operationele beslissingen op managementniveau. Pas daarna kijken we welke data daarvoor nodig is. Bij AMIS Conclusion krijg je dus niet wat je vraagt, maar wat je nodig hebt.

Domeinexpertise staat bij ons centraal. We streven er altijd naar een dataspecialist te koppelen aan een domeinexpert. Zo'n duo kan snel besluiten nemen over de waarde, het raffineren en de levensduur van de data. Daarbij kunnen we datamesh voorstellen als dit past bij het type probleem dat de organisatie heeft.

AMIS Conclusion is als een soort logistiek bedrijf. We halen data op bij een bron en bezorgen het met een bepaalde mate van prioriteit bij de persoon of afdeling die het nodig heeft. Daarbij bewaken we de hele keten: we zorgen dat het transport netjes en veilig geschiedt, houden je op de hoogte van de ETA van je pakket, zorgen dat er niets mist of kwijtraakt en zoeken naar een alternatief als er iets in de keten uitvalt. Gaat er ergens iets mis? Dan zorgen we voor herstel van de data. Eenmaal aangekomen leveren we de data af in het magazijn, waarna iemand in de winkel zorg draagt voor de display.

## Sparren over jouw data-uitdagingen?

Wil jouw organisatie de datazee indammen en ruwe data beter kunnen transformeren tot krachtig productiekapitaal? Zodat je meer slagkracht, weerbaarheid en wendbaarheid hebt én sneller en effectiever nieuwe (data)producten en -diensten kunt bouwen? Neem dan [contact](#) op met Robbrecht van Amerongen te sparren over jouw uitdagingen.







**AMIS**  

---

**CONCLUSION**