

Deel 1: Business glossary als effectief communicatiemiddel en fundament voor datakwaliteit

Organisaties automatiseren besluitvorming, implementeren AI-oplossingen en moeten aantoonbaar compliant zijn met regelgeving zoals de EU AI Act. In dat aspect is datakwaliteit een randvoorwaarde voor betrouwbaarheid en compliance. Gartner schat dat slechte datakwaliteit organisaties gemiddeld 12,9 miljoen dollar per jaar kost¹. De 1x10x100-regel maakt de dynamiek concreet: een probleem oplossen bij de bron kost één eenheid. Hetzelfde probleem oplossen nadat het door systemen is verspreid kost tien. Oplossen nadat het de besluitvorming heeft bereikt kost honderd².

Maar die kosten zijn niet allemaal even zichtbaar. Haug et al. (2011) laten zien dat organisaties de kosten van datakwaliteitsfouten structureel onderschatten, terwijl ze de kosten van onderhoud juist overschatten [3]. Het gevolg is dat investeringen in kwaliteitsverbetering te lang uitblijven omdat de pijn niet als dataprobleem wordt herkend. In de praktijk gaat het met datakwaliteit eerder mis op betekenis dan de techniek. Sales rapporteert bijvoorbeeld een ander nummer aan klanten dan marketing en finance. De discussies gaan vervolgens niet over waardecreatie of capabilities, maar over de cijfers en wat een ‘klant’ nou inhoudt.

- Telt een prospect mee?
- Telt een inactieve klant mee?
- Telt een dochteronderneming apart?
- Wanneer in de sales fase is iets een klant?

Hier hebben we het dan nog niet eens over data, maar over de definities die leven – of nog niet leven – binnen een organisatie. Zolang begrippen niet eenduidig zijn vastgelegd, zullen rapportages blijven conflicteren. En zolang rapportages conflicteren, verdwijnt vertrouwen. Zonder vertrouwen geen goede besluitvorming.

Wat is datakwaliteit écht en hoe beïnvloedt het decision making?

Bij datakwaliteit denken we vaak gelijk aan volledigheid, juistheid, tijdigheid. Die zijn belangrijk, maar datakwaliteit is ook de mate waarin het:

- Eenduidig is gedefinieerd
- Herleidbaar is naar bron en eigenaar
- Geschikt is voor het beoogde besluit

Als de definitie van “klant” verschilt per afdeling, kan de technische onderkant goed zijn en alsnog onbruikbaar voor besluitvorming. De betekenis volgens de organisatie is dus het startpunt omdat het de basis vormt voor sturing.

¹ [Dataversity](#)

² [Dataversity](#)

Onderstaande matrix maakt het onderscheid concreet. Op operationeel niveau zijn directe kosten zichtbaar: foutieve leveringen, herstelwerk, foutmeldingen. Minder zichtbaar zijn de verborgen operationele kosten zoals lange doorlooptijden, medewerkers die tijd verliezen aan afstemming die niet had hoeven plaatsvinden. Op strategisch niveau geldt hetzelfde patroon: investeringsverlies en compliance-boetes zijn achteraf meetbaar, maar markt- en reputatierisico's blijven onzichtbaar tot het te laat is. Beslissingen over prijsbeleid, klantsegmentatie of productontwikkeling die op inconsistente data zijn gebaseerd, manifesteren zich pas maanden later als omzet- of marktpositieverlies.

Verborgen kosten	<p>Operationele inefficiëntie</p> <p><i>Lange doorlooptijden, lage medewerkerstevredenheid</i></p>	<p>Markt- en reputatierisico</p> <p><i>Focus op verkeerde klantsegmenten, wankel fundament</i></p>
Directe kosten	<p>Correctiekosten</p> <p><i>Foutieve leveringen, foutmeldingen</i></p>	<p>Investeringsverlies</p> <p><i>Verspilde investeringen, compliance boete</i></p>
	Impact van slechte data kwaliteit op operationeel niveau	Impact van slechte data kwaliteit op strategisch niveau

Operationele fouten zijn direct herleidbaar naar gebreken in het dataverzamelingsproces; afgebakend, zichtbaar en herstelbaar³. Strategische fouten werken anders: een verkeerde definitie voedt een beslissing, die de basis vormt voor de volgende. Tegen de tijd dat de fout zichtbaar wordt, is hij al meerdere lagen diep. Hoe verder rechts in de matrix, hoe groter het ripple effect en hoe hoger de uiteindelijke rekening. Het doel is dan ook niet perfecte datakwaliteit, maar het punt waarop onderhoud en foutkosten samen het laagst zijn.⁴

³ *Int. J. Environ. Res. Public Health* 2014, 11, 5170-5207; doi:10.3390/ijerph110505170
International Journal of
Environmental Research and
Public Health
ISSN 1660-4601 www.mdpi.com/journal/ijerph

⁴ doi:10.3926/jiem.2011.v4n2.p168-193 JIEM, 2011 – 4(2): 168-193 – Online ISSN: 2013-0953
Print ISSN: 2013-8423
The costs of poor data quality 168
A. Haug; F. Zachariassen; D. van Liempd

De verborgen oorzaak: semantische versnippering

In veel organisaties groeien definities organisch. Afdelingen optimaliseren voor hun eigen doelstellingen en werkwijzen. Hierdoor worden datamodellen ook projectmatig opgezet.

Het resultaat hiervan is:

- Begrippen krijgen meerdere betekenissen.
- Datamodellen zijn niet altijd reproduceerbaar.
- Discussies gaan over cijfers in plaats van over keuzes

De business glossary als fundament

Een business glossary lost dit op door definities expliciet, herleidbaar en bestuurbaar te maken.

Een business glossary:

- Legt bedrijfsbegrippen expliciet vast
- Verbindt deze begrippen aan onderliggende (kritieke) data-elementen
- Benoemt eigenaarschap over de definitie en het data-element
- Maakt definities toetsbaar en reproduceerbaar

Een bruikbare entry bevat minimaal:

Attribuut	Toelichting
Term	De naam van het begrip zoals gebruikt in de organisatie
Definitie	Eenduidige omschrijving, goedgekeurd door de eigenaar
Eigenaar	De rol of persoon verantwoordelijk voor de definitie
Status	Conceptueel, in review, of goedgekeurd
Gerelateerde begrippen	Synoniemen, verwante termen, uitsluitingen

Het is geen document ter archivering maar bedoeld als een governance-instrument. Het is ook niet statisch. In combinatie met een enterprise datamodel ontstaat een semantisch fundament: Het woordenboek definieert betekenis en het datamodel structureert de samenhang.

Waar te beginnen: de 10-30 begrippen die pijn doen

Niet alles tegelijk. De eerste stap is het identificeren van de begrippen die op dit moment de meeste conflicten veroorzaken in rapportages en besluitvorming. Dat zijn doorgaans de begrippen die terugkeren in overleggen tussen afdelingen of verschillend worden gebruikt in dashboards die naast elkaar bestaan.

Wat verandert er operationeel

Veel organisaties hebben inmiddels data stewards en data-eigenaren benoemd, maar zonder gedeelde definities missen deze rollen een gemeenschappelijk referentiekader. Een steward kan geen kwaliteit borgen als onduidelijk is wat een begrip exact betekent. Een eigenaar kan geen verantwoordelijkheid nemen voor een KPI die per afdeling anders wordt geïnterpreteerd.

Zodra kernbegrippen eenduidig zijn vastgelegd, verandert de dagelijkse praktijk concreet. Een analist die voorheen tijd verloor aan definitie-afstemming voordat een rapport gebouwd kon worden, heeft nu een gevalideerd startpunt. Een data steward die een kwaliteitsmelding ontvangt, toetst direct aan een vastgelegde definitie in plaats van de discussie opnieuw te voeren. Een nieuwe collega begrijpt binnen een dag wat de organisatie verstaat onder haar kernbegrippen.

De glossary werkt alleen als de rollen die het beheren en handhaven ook zijn ingericht. Hoe dat in de praktijk werkt, van het signaleren en oplossen van kwaliteitsissues en de concrete uitwerking van die rollen — is het onderwerp van deel 2.

Deel 2: Processen en tools om datakwaliteit in de praktijk aan te pakken

Datakwaliteit is een fundamentele voorwaarde voor betrouwbare analyses en effectieve besluitvorming. In deel 1 legden we het fundament: het belang van datakwaliteit, opstellen van eenduidige definities en business rules en wat de gevolgen zijn van slechte datakwaliteit. Alleen wie bewaakt of data daar ook aan voldoet? Hoe weet je welke issues er zijn, hoe zorg je dat je als organisatie weet wat er speelt en de controle behoudt? Hoe zorg je dat issues structureel gestandaardiseerd worden opgelost? Datakwaliteit is een continu doorlopend proces dat niet enkel de technische mensen raakt, maar van uitvoering tot back-office gedragen dient te worden.

Datakwaliteitsmonitoring

Sommige data-elementen zijn crucialer voor de bedrijfsvoering dan anderen. Wanneer er issues plaatsvinden is het belangrijk om te prioriteren op de kritieke data-elementen.

Kritieke data-elementen zijn data die essentieel zijn voor processen, bedrijfsvoering of rapportage. Voor deze elementen worden datakwaliteitsregels opgesteld. Wanneer deze gegevens niet op orde zijn of ontbreken kan dit leiden tot misverstanden, vertraging in processen en uiteindelijk minder tevreden klanten.

Deze kritieke gegevens kunnen op verschillende manieren gemonitord worden, afhankelijk van de context:

Volledigheid

Mate waarin velden zijn ingevuld

Is er een e-mailadres ingevuld?

Nauwkeurigheid

Mate waarin data overeenkomen met de werkelijkheid

Juiste geboortedatum

Consistentie

Uniformiteit tussen systemen

Klant staat zowel in het CRM- als facturatiesysteem aangemeld als inactieve klant

Validiteit

Naleven van formats en definities

Alle datum-elementen zijn in DD-MM-YYYY format

Tijdigheid

Actualiteit van data

Heeft de pipeline de data van vandaag ingeladen in het dashboard?

Uniekheid

Voorkomen van duplicaten

Klant heeft 4 accounts gemaakt

Praktijkmechanismen voor datakwaliteitsbeheer

De organisatie, systemen en data zijn continu in beweging. Nieuwe producten en/of diensten, versie releases, wijzigingen in processen, nieuw gewenste inzichten vanuit rapportages/BI, aanpassingen in de functionaliteit van applicaties, veranderde wet- en regelgeving. Allerlei redenen voor technische wijzigingen waar potentiële datakwaliteitsissues uit voort kunnen komen. Hoe zorg je ervoor dat je weet wat er allemaal speelt, en hoe behoud je controle?

Dashboard

Het monitoren van kritieke data-elementen gebeurt vaak via een dashboard waarin KPI's zijn geformuleerd op basis van de bovenstaande 6 kwaliteitsdimensies. Afhankelijk van het systeemlandschap en de proces- en gegevensstroom kunnen de meetpunten en doeleinden van monitoring variëren. Zo kan de focus bijvoorbeeld liggen op de consistentie van een gegeven tussen verschillende applicaties; *“Is de startdatum van iedere klant in beide systemen hetzelfde?”*, of de compleet- en validiteit in een datawarehouse; *“Hebben alle klanten een geldige startdatum in het verleden in DD-MM-YYYY format?”*

Data Governance Tooling

Organisaties met een hoge volwassenheid op data governance gebied beschikken vaak over tooling, zoals Colibra, waarin uniforme definties, eigenaarschap, data lineage en business rules centraal belegd zijn. In combinatie met monitoring en een issue loket zorgt dit dat je niet enkel reactief (handelen na waarnemingen in dashboard of ingeschoten tickets in het loket), maar juist proactief kan sturen op datakwaliteit en governance.

Datakwaliteitsissue loket

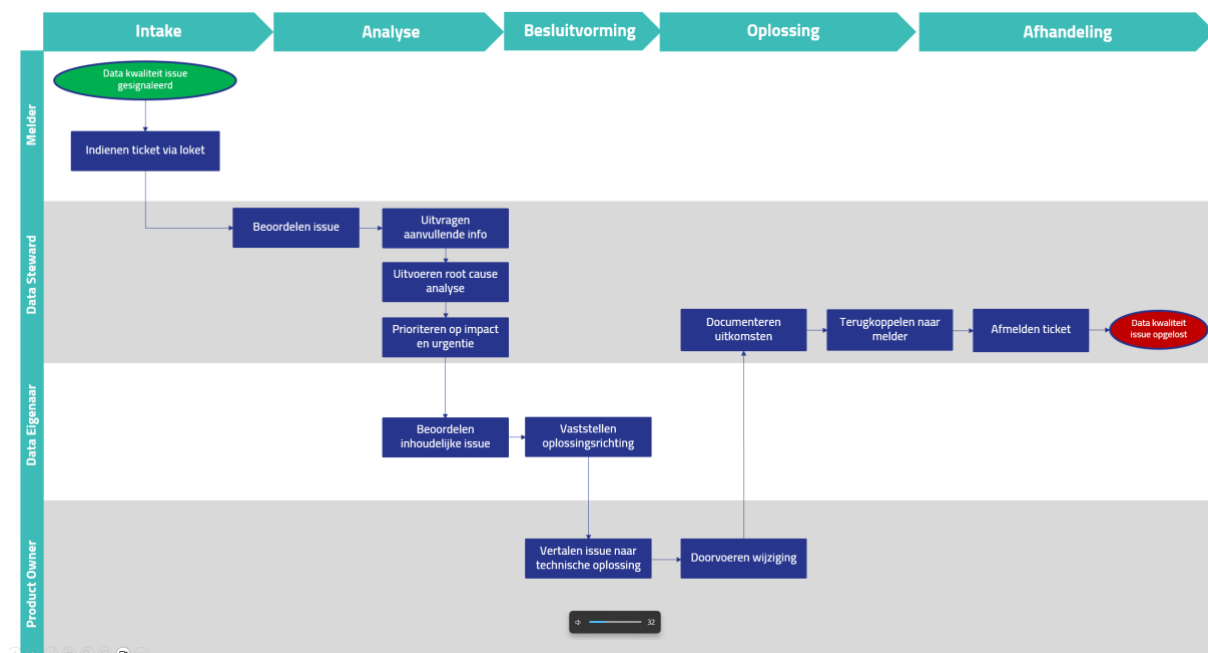
Een datakwaliteitsissue loket fungeert als centraal meldpunt voor datakwaliteitsproblemen en zorgt voor een uniforme afhandeling van issues. Door datakwaliteitsissues centraal te registreren ontstaat inzicht in structurele

dataproblemen waardoor er gerichtere verbeteringen in systemen, processen of datadefinities kunnen worden doorgevoerd.

Zonder een centraal loket ontbreekt vaak overzicht over de datakwaliteitsproblemen die spelen binnen de organisatie. Issues worden verspreid via e-mail, chats, of los opgepakt waardoor het onduidelijk is welke problemen al bekend zijn, welke prioriteit zij hebben en of ze daadwerkelijk worden opgelost. Het loket vervult daarom meerdere belangrijke functies:

- Centraal meldpunt
- Overzicht en transparantie
- Voorkomen van ad-hoc oplossingen
- Borging van opvolging
- Datagedreven kwaliteitsbeheer; meer inzicht in wie, wat, waar, hoe, hoevaak

Wanneer afwijkingen in datakwaliteit worden gesignaleerd is het belangrijk dat deze volgens een **vast afhandelingsproces** worden aangepakt zodat iedereen weet **wie** verantwoordelijk is voor **wat** op **welk** moment. Na registratie in het loket wordt een issue beoordeeld op impact en urgentie. De data steward speelt hierin een centrale rol door de issues te analyseren en prioriteren. Op basis hiervan worden de juiste stakeholders betrokken. De data-eigenaar is verantwoordelijk voor de inhoudelijke besluitvorming en prioritering, terwijl de product owner of functioneel ontwerper zorgt voor de implementatie van de structurele verbetering in systemen en/of processen. Terugkoppeling naar de melder en borging van de oplossing zijn essentieel om vertrouwen in het proces te waarborgen en herhaling te voorkomen.



Figuur 1 – Voorbeeld van het proces van een datakwaliteit issue loket

Verantwoordelijkheden in Data Governance

Het doel van data governance is om de controle te behouden over je systemen, gegevens en de kwaliteit van gegevens. Data moet betrouwbaar, bruikbaar, veilig en goed beheerd zijn voor effectieve processen en besluitvorming. Datakwaliteit is echter geen eenmalige actie, maar een doorlopend proces van Plannen, Uitvoeren, Controleren en Verbeteren. Een ingerichte PDCA-cyclus belegt duidelijke verantwoordelijkheden in een formeel doorlopend proces.

Het aanwijzen van een data-eigenaar is een stap in de goede richting, maar nog niet het eindproduct voor een waterdichte data governance. De data-eigenaar (**PLAN**) is in hoofdlijnen degene die de kaders bepaalt voor de data; vaststellen van kwaliteitseisen, definities, en het vaststellen van beleid en regels. De Product Owner (**DO**) zorgt dat het systeem goed functioneert en dat verbeteringen worden doorgevoerd. De Data Steward (**CHECK**) bewaakt en monitort de datakwaliteit, signaleert datakwaliteitsissues en beheert de issues die ingeschoten worden via het datakwaliteit loket.

De uiteindelijke besluiten, aanpassingen en verbeteracties (**ACT**) worden door het gehele team opgepakt. Op basis van de bevindingen, monitoring en geregistreerde datakwaliteitsissues worden oplossingen bepaald waarbij wordt gekeken naar de onderliggende oorzaak van het probleem zodat niet alleen het symptoom wordt opgelost maar het proces structureel wordt verbeterd. De data-eigenaar is hierin de eindverantwoordelijke voor het nemen van een besluit en bepaalt de prioritering van verbeteracties. De Product Owner vertaalt deze verbeteringen door naar aanpassingen in systemen en datastromen, waarbij de Data Steward een ondersteunde rol heeft in dit proces door de doorgevoerde verbeteracties te monitoren, datakwaliteitsproblemen te analyseren en eventuele wijzigingen door te voeren in het business glossary.



Figuur 2 – PDCA cyclus voor datakwaliteit

Wanneer rollen zijn belegd, een loket functioneert en kwaliteitsissues opgevolgd worden is nu de vraag: wie bepaalt wat er met terugkerende patronen gebeurt? Wie beslist als een issue de grenzen van een domein overstijgt? Er is dus een passende structuur nodig over hoe besluitvorming wordt uitgevoerd en hoe informatie de juiste mensen bereikt. Dit is het onderwerp voor in het volgende en laatste artikel.

Deel 3: Van instrumenten naar een werkend governance systeem

In de eerste twee delen stonden we stil bij enkele data governance instrumenten en mechanismen: het business glossary als fundament voor gedeelde betekenis, data governance rollen en het datakwaliteitsloket als centraal mechanisme voor het signaleren en oplossen van issues. Echter verbinden de bouwstenen zichzelf niet.

Onderzoek van laat zien dat tot 75% van de governance-initiatieven mislukt doordat eigenaarschap onduidelijk blijft⁵. Niet omdat de instrumenten slecht zijn, maar omdat er geen structuur is die bepaalt hoe beslissingen worden genomen, hoe informatie doorstroomt en wie verantwoordelijk is als het fout gaat. Dat is wat een data governance operating model doet.

Het operating model als lijm

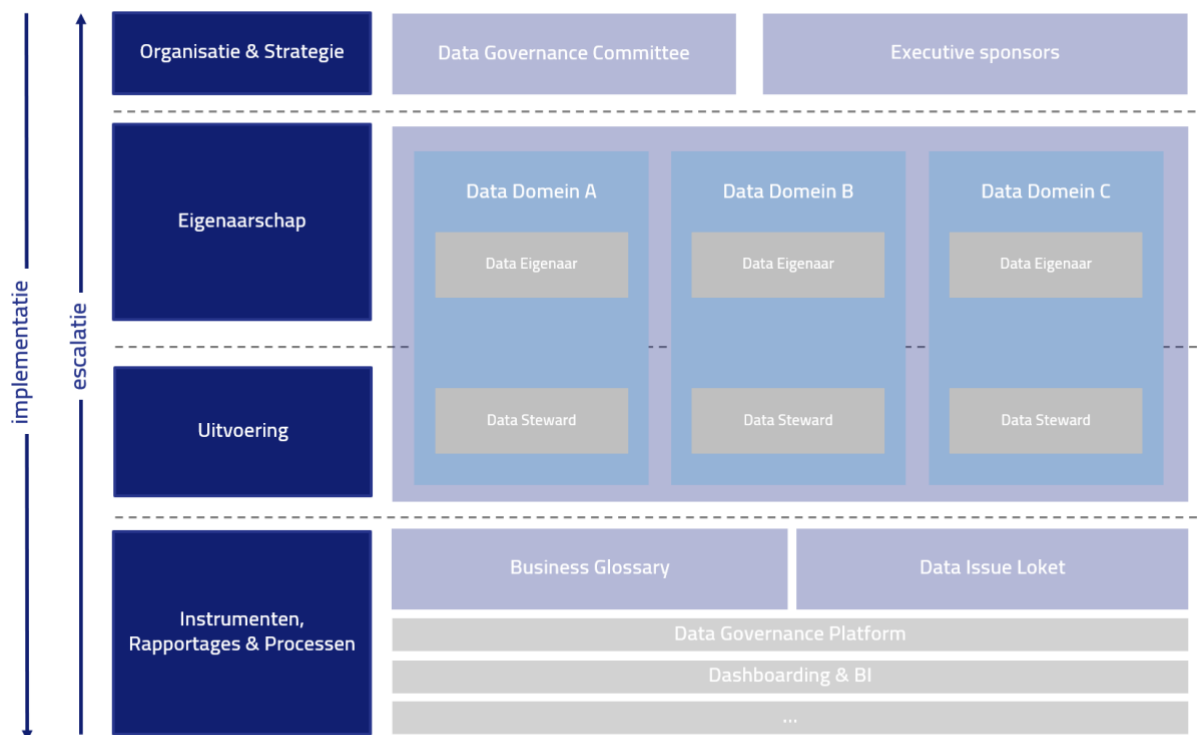
Een governance operating model is niet een extra laag van beleid of overhead. Het is een structuur die bepaalt hoe de instrumenten en rollen uit deel 1 en 2 met elkaar samenwerken. Concreet regelt het vier dingen.

1. Besluitvorming: wie beslist over definitiewijzigingen, kwaliteitseisen en verbeteracties, op welk niveau en hoe conflicten tussen domeinen worden opgelost.
2. Mandaat: wat data stewards en -eigenaren wel of niet mogen en hoe de RACI eruit ziet.
3. Informatiestromen en escalatieladders: hoe bereiken signalen uit het loket en de monitoring de juiste beslissers? Structurele rapportage zorgt ervoor dat inzicht niet alleen lokaal blijven maar ook overstijgend worden.
4. Ritme: zonder vastigheid verwatert het. We zien vaak dat een maandelijks ritme wordt aangehouden.

Er zijn diverse vormen van een operating model, centraal, decentraal of gefedereerd. De keuze daartussen heeft directe consequenties hoe middelen zoals het business glossary en de processen in de praktijk worden ingericht en gebruikt. Er bestaat geen universele oplossing. De juiste vorm hangt af van de complexiteit en volwassenheid van de organisatie, de mate van domeinautonomie en de cultuur. Wat wel geldt: het operating model moet aansluiten op de organisatiestructuur zoals die werkelijk functioneert.

Onderstaand diagram toont een governance structuur die dit mogelijk maakt: governance committee op strategisch niveau, data-eigenaren per domein die eigenaarschap borgen, data stewards die uitvoering geven aan de dagelijkse opvolging en instrumenten, processen en tools (processen, instrumenten en rapportages voor besluitvorming) die het geheel ondersteunen. De domeinen opereren autonoom, maar zijn verbonden via gedeelde besluitvorming en gedeelde instrumenten. Dit is kenmerkend voor een gefedereerde inrichting.

⁵ [Acceldata](#)



Voorbeeld operating model

De werkelijke waarde van een governance operating model wordt zichtbaar op het moment dat een lokaal issue structureel doorwerkt naar de bredere organisatie

Stel dat het loket in één domein structureel dezelfde melding ontvangt: klantgegevens zijn onvolledig op het moment dat een order wordt aangemaakt. In een organisatie zonder operating model wordt dit domein-voor-domein opgelost, of niet opgelost. Met een operating model wordt het patroon zichtbaar in de maandelijkse governance review. De data-eigenaar concludeert dat de oorzaak niet in de data zit maar in een definitie die per domein anders wordt geïnterpreteerd. Die bevinding gaat naar de governance committee, die bepaalt of de definitie in het business glossary wordt aangepast en of het probleem organisatiebreed speelt.

Dit is het mechanisme dat de drie delen van deze serie verbindt. In een gefedereerd model profiteert vervolgens elk domein van die definitiewijziging en niet alleen het domein waar het issue oorspronkelijk werd geregistreerd.

Bruikbaarheid en eenvoud bevorderen de inbedding en duurzaamheid van je operating model

Het inrichten van zo'n model hoeft niet complex te zijn. Volgens de DAMA zijn een aantal factoren essentieel voor succes:

- Koppel het model aan de werkelijkheid: Het moet aansluiten op hoe de organisatie nu al functioneert. Gebruik zoveel mogelijk bestaande overlegstructuren.

- Focus op hoge waarde: Start met pilots in domeinen waar de datacomplexiteit hoog is en de impact op de business groot.
- Executive sponsorship: Zonder steun van bovenaf is een model niet duurzaam. Het Committee moet een echt 'decision-making body' zijn.

Quick wins van samenwerking: De drie delen als werkend systeem

De opbrengst van een werkend systeem hoeft niet te wachten op volledige governance-volwassenheid. Zodra instrumenten zoals het business glossary, het loket en de rollen ook maar gedeeltelijk functioneren en verbonden zijn via een operating model, verandert de dagelijkse praktijk.

Het eerste dat verdwijnt is de definitiediscussie. Zodra kernbegrippen zijn vastgelegd en een eigenaar hebben, stopt het overleg over wat een cijfer betekent voordat de analyse kan beginnen. Directe tijdswinst voor analisten, stewards en business owners.

Het tweede is overzicht. Een centraal loket maakt voor het eerst zichtbaar welke datakwaliteitsproblemen er daadwerkelijk spelen, hoe vaak ze voorkomen en in welke domeinen ze zich concentreren. Dat inzicht alleen al is waardevol.

Het derde is grip. Zodra duidelijk is wie eigenaar is van een begrip of data-element, weet de rest van de organisatie wie ze moeten benaderen. Issues blijven niet langer zweven tussen afdelingen maar worden opgepakt door de juiste rol met het juiste mandaat.

Ter afsluiting van de serie

In deze serie hebben we bewust gefocust op een aantal concrete instrumenten en mechanismen die in de praktijk het verschil maken voor data governance: van gedeelde definities tot gestructureerde kwaliteitsopvolging en de structuur die beide verbindt. Het is een beginpunt om vervolgens het gehele data governance framework te verbeteren, successen in één domein de springplank te laten zijn voor de rest van de organisatie en uiteindelijk data governance als ondersteuning van de organisatiestrategie.