



Het hoe en wat van gegevenskwaliteit

Een whitepaper van The Lifecycle Company

In dit whitepaper lichten we het onderwerp gegevenskwaliteit nader toe, evenals de opties rond het meten, bewaken en verbeteren van de kwaliteit van bedrijfsgegevens. Tevens adresseren we de verantwoordelijkheden op het gebied van gegevenskwaliteit. Hiermee bieden we de basis voor besluitvorming rond

Achtergrond

Goede informatievoorziening is voor veel organisaties een belangrijke levensader. Een zeer relevant aspect van de informatievoorziening is de informatie zelf: de data, de content, de gegevens. Logischerwijs verdient dus ook het beheer van deze informatie veel aandacht.

Om deze reden is er binnen het BISO-model een procesgebied beschreven dat Beheer Bedrijfsinformatie heet. Hierbinnen zou men zes aandachtsgebieden kunnen onderscheiden:

- Gegevenskwaliteit
- Managementinformatie en query's
- Parameterbeheer
- Autorisatiebeheer
- Bewaking bedrijfsinformatiemodel
- Afwijkende mutaties.

Dit whitepaper gaat over het eerste onderwerp: gegevenskwaliteit.

Het belang van gegevens

Gegevens vormen voor de meeste organisatie de input en de basis van de basisprocessen, ze maken de interne strategievorming, sturing, beleidsimplementatie en verantwoording mogelijk, maar ze vormen ook de basis voor de informatie-uitwisseling met onze omgeving.

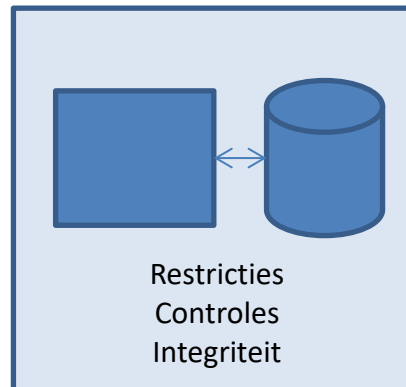
Omdat gegevens belangrijk zijn is de kwaliteit van veel van deze gegevens ook van essentieel belang.

Het bewaken van gegevenskwaliteit

Het bewaken en controleren van gegevenskwaliteit vindt in eerste instantie plaats in de geautomatiseerde systemen. Deze systemen voeren geautomatiseerde invoercontroles uit en dwingen bepaalde formats af. Dit geldt voor grootschalige gegevensoverdracht (bijvoorbeeld het binnen halen van basisadministraties en andere externe bestanden) maar ook voor bediening van de applicatie door eindgebruikers waar gegevens worden ingevoerd. Het is een taak van IT (de interne IT-organisatie en de externe leveranciers) om deze geautomatiseerde kwaliteitsmaatregelen te realiseren en onderhouden en het is de taak van informatiemanagement om software hierop te selecteren en de leveranciers op dit aspect aan te sturen. Zie figuur 1.



IT

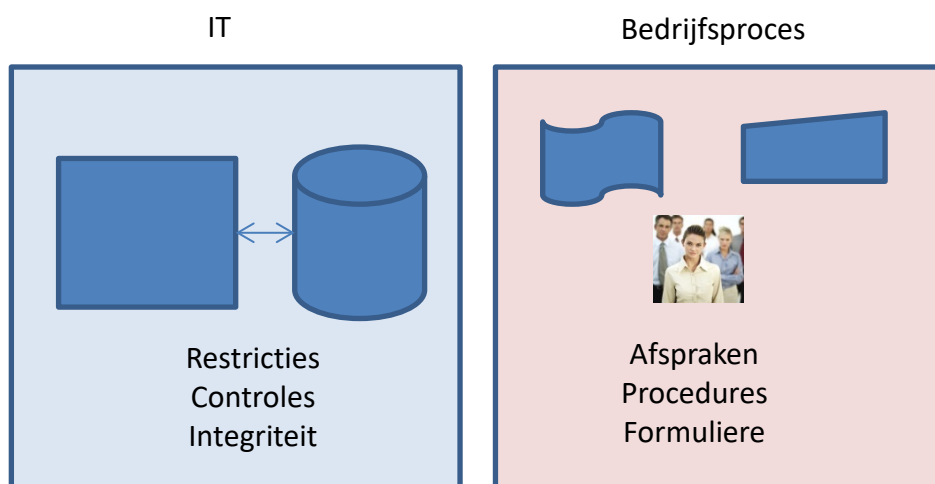


Figuur 1: IT en gegevenskwaliteit

Anderzijds kan men niet alles geautomatiseerd afdwingen, want dat zou onwerkbaarheid tot gevolg hebben. Dit betekent dat de medewerkers in de bedrijfsprocessen toch ook veel vrijheid (en dus verantwoordelijkheid) hebben bij het invoeren en muteren van gegevens in de geautomatiseerde systemen. Daarmee komt een heel groot stuk van de verantwoordelijkheid voor de kwaliteit van de bedrijfsgegevens bij de eindgebruiker te liggen: de medewerker en diens manager(s). Het is ten slotte hun bedrijfsproces en het zijn hun bedrijfsgegevens.

Dit betekent dus ook dat de medewerker zich aan bepaalde richtlijnen en (vaak ongeschreven) normen en standaarden dient te houden die niet persé geautomatiseerd worden afgedwongen, op basis van eigen verantwoordelijkheidsgevoel en op basis van organisatie-richtlijnen en afspraken.

Zo kan men dus vanuit twee richtingen de gegevenskwaliteit bewaken en borgen: vanuit IT en vanuit de bedrijfsprocessen (lijn en medewerkers). Zie figuur 2.



Figuur 2: Actoren m.b.t. gegevenskwaliteit



De praktijk

Binnen het bedrijfsproces is niets menselijk ons vreemd. Soms kennen we de afspraken niet, soms vergeten we de afspraken en richtlijnen even, soms is er sprake van haast en tijdsdruk en soms hebben we er gewoon ook even geen zin in.

In een document van een willekeurige organisatie is te lezen:

- *De medewerkers houden zich te weinig aan de generieke afspraken en definities. Vaak is men daar niet van op de hoogte.*
- *Er is nog te weinig kennis, kunde, ervaring, motivatie en soms tijd om volledig en nauwkeurig gegevens in te voeren.*

De consequentie is dat veel gegevensverzamelingen gaandeweg inboeten aan kwaliteit: gegevens vervuilen, raken incompleet worden en/of gedateerd en anderszins. We zouden het niet willen, maar het is een 'fact of life'.

Op zich is hier wel mee te leven, mits de organisatie redelijk actueel inzicht heeft in wat dan precies de gegevenskwaliteit is en waar het niet voldoet aan de eisen, zodat daar rekening mee gehouden kan worden en zodat tijdig kan worden ingegrepen op plaatsen waar de norm niet meer gehaald wordt, bijvoorbeeld door middel van organisatorische maatregelen of door herstelacties.

En dit laatste is helaas niet het geval. Grosso modo weten de organisatie niet wat de huidige kwaliteit van onze bedrijfsgegevens is en we wordt er ook niet over de volle breedte of heel gericht periodiek onderzoek naar gedaan. En daar loopt men risico's. En het vervelende is dat deze risico's steeds groter worden, aangezien de druk en de eisen uit de omgeving ook steeds groter worden. Er worden namelijk steeds meer gegevens uitgewisseld en de vraag om direct goede gegevens uit te wisselen wordt steeds korter van tevoren gesteld.

Enkele voorbeelden van waar in praktijk problemen optreden, waar het mis is gegaan, kan gaan of zou kunnen gaan:

- Bij het koppelen van systemen blijkt vrij regelmatig dat bestanden vervuild zijn. Snelle koppeling lukt niet en analyse en herstel vragen veel tijd en leiden tot druk en problemen in het bedrijfsproces.
- Wat gebeurt er als uw organisatie in het nieuws is en direct gegevens op moet leveren waar de pers gaten in schiet?
- Wat gebeurt er als de directie snel managementinformatie nodig heeft die een week op zich laat wachten?
- Wat gebeurt er als er vergunningverlening (of juist niet), handhaving, subsidieverstrekking of welk bedrijfsproces dan ook, plaatsvindt op basis van slechte gegevens?



Een concreet voorbeeld: een tekst uit een strategisch analyserapport van een organisatie

....

Het samenstellen van een bruikbare dataset voor analyse heeft veel moeite gekost. Dat heeft veel te maken met het systeem waarin de data zijn geregistreerd, maar ook met de opzet van de registratie. In feite ging het hier niet om inspecties, maar om meldingen in de vorm van signalen. Vanuit pragmatische overwegingen is er voor gekozen ze als inspecties vast te leggen

.....

Dit heeft wel geleid tot allerlei inconsistenties in de dataset. Zo zijn nogal eens de melder en de locatie waarover gemeld werd door elkaar gehaald.

.....

Er is door medewerkers veel inspanning verricht om zo goed mogelijk de gegevens van de meldingen te registreren ondanks de beperkte capaciteit qua menskracht en de mogelijkheden van het registratiesysteem. Het is echter vrijwel onmogelijk om over een lange periode een grote hoeveelheid gegevens te verzamelen en te registreren met voldoende kwaliteit zonder periodieke terugkoppeling van de invoer naar de invoerders. Bovendien zou er periodiek een audit moeten plaatsvinden waarbij o.a. de kwaliteit van de gegevens wordt getoetst op volledigheid en betrouwbaarheid.

Het verdient dan ook aanbeveling dit soort aanpassingen van het werkproces zo snel mogelijk in te voeren.

Bovenstaand voorbeeld stamt uit het verleden, maar de situatie is, krijg ik te horen, sindsdien eerder achteruitgegaan dan verbeterd.

Een ander voorbeeld:

Dit betreft een analyse over twee maanden ten aanzien van zaakregistratie, waaruit blijkt dat er geen verbeteringen zijn, ondanks de rapportage die in september is verstrekt.

	aantal		
	sept 2013	okt 2013	verschil
Zaken afgerond zonder eindoordeel	780	1045	-265
Zaken afgerond zonder dat alles volledig was ingevuld	2500	2594	-94

De medewerkers zijn nog steeds niet op de hoogte van wat van ze verlangd wordt bij het invullen van het systeem.

Een klein deel van het niet invullen kun je op het conto van het systeem schrijven, bijv. Maar afronden met een eindoordeel kan altijd. En als dat eindoordeel er nog niet is, zou er niet afgerond mogen worden. Er staan nog Zaken uit 2008 open. Die zouden inmiddels gestopt of afgerond kunnen zijn.



Nog een voorbeeld

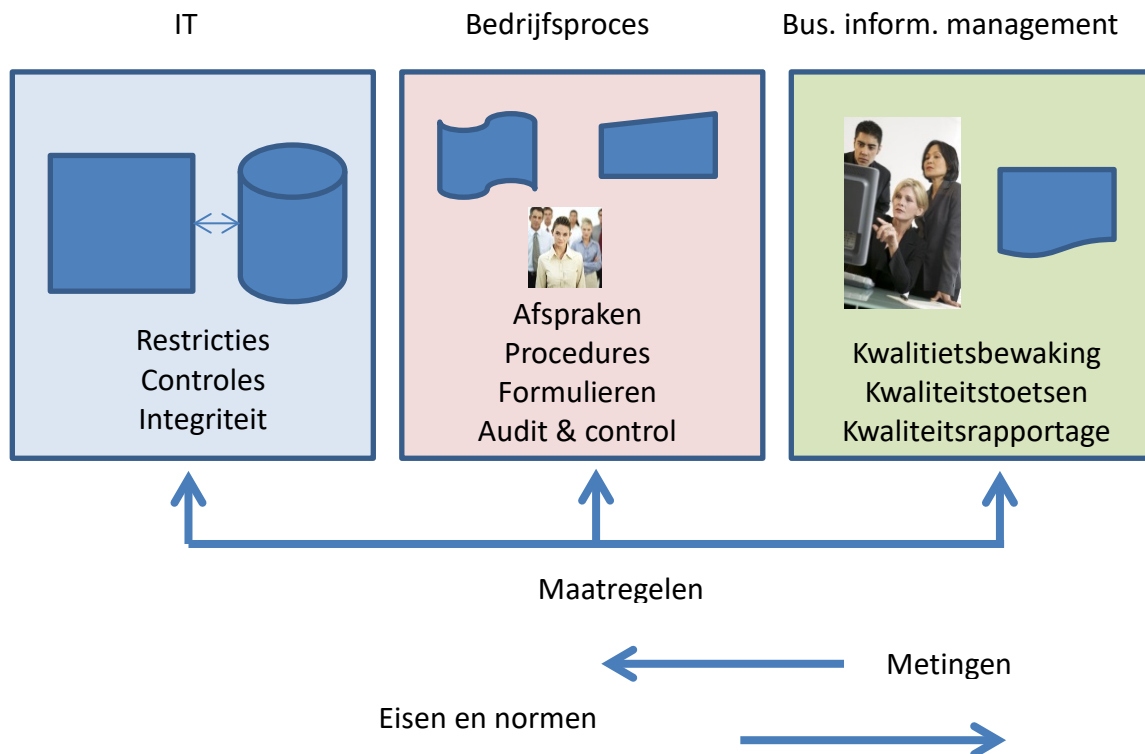
Een citaat uit een interne mail uit een andere willekeurige organisatie:

Er is helemaal geen cultuur op het gebied van datakwaliteit in deze organisatie. Dat komt mede omdat men steeds de ruimte krijgt om gegevens in rapportages op te poetsen, schaduwregistraties bijhoudt, weinig externe afname heeft van gegevens (stok achter de deur), etc.

De oplossing

Naast het inbouwen van controles en bewaking in de eautomatiseerde systemen en naast het opleiden en aanspreken van medewerkers op hun verantwoordelijkheden en het sturen vanuit de lijn, waar het belang van datakwaliteit wordt gevoeld, is er dus nog een optie: het gericht periodiek of ad hoc analyses uitvoeren, al of niet steekproefsgewijs. Af en toe een onderhoudscontrole dus. De aangewezen partij hiervoor vormt business informatiemanagement (functioneel beheer en informatiemanagement): zij vormt de schakel tussen de business en IT en kunnen de vertaling maken van wat nodig is naar wat mogelijk is en vice versa.

Zie figuur 3



Figuur 3: Business informatiemanagement als bewaker

In figuur 3 zien we dat de vertegenwoordigers uit het bedrijfsproces de eisen en normen formuleren, dat business informatiemanagement op basis daarvan de kwaliteit bewaakt, toetst en verslag doet, en dat vanuit het bedrijfsproces vervolgens maatregelen worden gevraagd aan IT (aanpassen



systemcontroles), aan het bedrijfsproces (discipline, procedures) en/of aan functioneel beheer (herstel/correctie).

Hieraan zijn echter wel enkele voorwaarden verbonden:

- de business, de gebruikersorganisatie, de lijn, moet aangeven wat de kwaliteitseisen en de normen zijn. Zie de bijlage van deze notitie voor een (niet compleet) overzicht van mogelijke kwaliteitseisen.

Vragen hierbij:

- welke criteria zijn van toepassing?
- welk kwaliteitsniveau (norm) is, per criterium, gewenst en vereist?
- voor welke gegevens(-sets)?
- wie heeft het mandaat (en de verantwoordelijkheid) om dit te bepalen?
- business informatiemanagement dient tijd, capaciteit en autorisaties te hebben om de noodzakelijke controles uit te voeren;
- het dient helder te zijn aan wie business informatiemanagement dient te rapporteren over de gegevenskwaliteit.

De controles van business informatiemanagement zullen gericht zijn op het planmatig of ad hoc bepalen van de actuele gegevenskwaliteit op basis van controlecriteria. Enkele voorbeelden zijn:

- controle of gegevens in databases voldoen aan spelregels (numeriek, decimaal etc.)
- controle of er geen wijzigingen optreden op gegevens die reeds zijn geverifieerd (bijv. betalingsopdrachten);
- controle of van ontvangen goederen de magazijnvoorraden zijn bijgewerkt;
- controle of van de beëindigde inspectietrajecten de dossiers volledig gearchiveerd zijn.

De gegevenseigenaar

In een vorige alinea wordt gesteld "de business, de gebruikersorganisatie, de lijn, moet aangeven wat de kwaliteitseisen en de normen zijn". Een probleem hierbij is dat er zelden één set van eenduidige eisen is. Doorgaans zijn er veel stakeholders die allen andere eisen stellen: de "leveranciers", gebruikers en afnemers van gegevens. Deze stakeholders kunnen intern zijn, maar ook buiten de organisatie. Denk bijvoorbeeld aan een ziekenhuis: de verpleegafdeling stelt wellicht andere eisen aan de patiëntgegevens dan de behandelend arts, dan de financiële afdeling, dan de zorgverzekeraar, dan het CBS, dan

Dat betekent dat er iemand moet zijn die deze eisen verzamelt, bundelt en "er een klap op geeft". Doorgaans noemen we dit de gegevenseigenaar. Voor elk gegeven moet dus een gegevenseigenaar worden gezocht. Hier is dan weer een taak weggelegd voor business informatiemanagement: inventariseren welke gegevens er zijn en coördineren dat er gegevenseigenaren komen.

Authentieke bron

Gegevens worden op verschillende plekken gecreëerd aangepast en vernietigd, zowel binnen de organisatie als er buiten. Het is daardoor niet altijd eenvoudig om een gegevenseigenaar aan te wijzen. Als een bepaald gegeven van buiten komt, maar het kan intern in twee plaatsen, in verschillende bedrijfsprocessen, ook gecreëerd of aangepast worden, wie is dan de eigenaar? Voor gegevens die op meerdere plaatsen worden gebruikt is de meest voor de hand liggende stap om eerst te bepalen wat de authentieke bron is. Dit kan trouwens ook belangrijk zijn als er problemen ontstaan bij het matchen van gegevens: zijn de adresgegevens van medewerkers uit het HR-systeem "de waarheid" of die uit het bestand dat we van de Belastingdienst ontvangen? Doorgaans hebben we behoefte aan "a single point of truth".



The lifecycle company

INFORMATION PROCESS IMPROVEMENT

Landelijke basisadministraties

In overheidsland wordt het streven naar authentieke bronnen vormgegeven door het definiëren van basisadministraties, die voor elke overheidsorganisatie de enige echte authentieke bron zijn. Voorbeelden zijn de BAG (Basisadministratie Adressen en Gebouwen), waarvan het Kadaster de bronhouder en verstrekker is, de NHR (Nederlands Handels Register), waarvan de Kamer van Koophandel de bronhouder en verstrekker is, en de GBA (Gemeentelijk Basis Administratie), waarvan het ministerie van BZK de registratiehouder is (de politieke eigenaar), de gemeenten de bronhouder en het Agentschap BPR de verstrekker.

Voor overheidsinstanties geldt dat zij dit soort gegevens dus altijd van de aangewezen gegevensverstrekker naar binnen moeten halen en maar één keer mogen vastleggen.

Als er eenmaal een authentieke bron is bepaald, dan kan men voor die gegevenseigenaar vaak wel makkelijk een eigenaar vinden. Voor de overige gegevens is het een kwestie van afstemmen en besluiten wie de gegevenseigenaar is en wat dat precies inhoudt.

The Lifecycle Company bestaat uit een netwerk van gelijkgezinden, allen met een ruime ervaring en expertise in advisering en inrichting van IT-regie, business informatiemanagement (functioneel beheer & informatiemanagement) en applicatiemanagement (applicatiebeheer & onderhoud).

Wij zijn het expertisecentrum voor BiSL en ASL, de control frameworks voor business informatiemanagement en applicatiemanagement.

Meer informatie?

Mail ons op info@thelifecyclecompany.nl



Bijlage: kwaliteitscriteria ten aanzien van gegevens

Enige algemeen gebruikte kwaliteitsaspecten (indicatoren) zijn:

Juistheid extern	komen de gegevens overeen met de werkelijkheid zoals de externe belanghebbenden dit vanuit hun perspectief zien
Juistheid intern	komen de gegevens overeen met de werkelijkheid zoals wij die vanuit ons perspectief zien
Consistentie intern	klopt de onderlinge samenhang van gegevens binnen de eigen organisatie
Consistentie extern	klopt de onderlinge samenhang van gegevens ook buiten de eigen organisatie
Beschikbaarheid	(Snelheid van) zijn de gegevens snel genoeg beschikbaar in het bedrijfsproces
Aanpasbaarheid	kunnen de gegevens op passende wijze gewijzigd worden.
Conform definities	worden de gegevensrepresentaties gevuld en gebruikt conform de opgestelde definities
Conform conventies	worden de gegevensrepresentaties gevuld en gebruikt conform de afspraken die gelden ten aanzien van vastlegging en gebruik van gegevens
Conform procesafspraken	worden de gegevensrepresentaties gevuld en gebruikt zoals is afgesproken voor de verschillende processen
Conformance in het algemeen	voldoen de gegevens aan wet- en regelgeving
Conformance met Governance	voldoen de gegevens aan de beleidsrichtlijnen voor gegevens van de organisatie
Controleerbaarheid	in welke mate kunnen we gegevens controleren op juistheid.
Tijdigheid	zijn de gegevens op het juiste moment beschikbaar
Volledigheid	zijn de gegevens compleet of ontbreekt er iets
Relevantie	zijn alle vastgelegde gegevens relevant of worden ook overbodige gegevens vastgelegd
Metagegevens	zijn voldoende metagegevens over de gegevens bekend.
Begrijpbaarheid	kennen we de definities van wat er staat en zijn deze definities duidelijk (staat in de meta-data)
Uniekheid	iets dat één keer in werkelijkheid is, komt ook één keer in gegevensverzameling voor
Uitwisselbaarheid (koppelbaarheid)	kunnen de gegevens gerelateerd worden aan gegevens in andere gegevensverzamelingen waarmee inhoudelijke uitwisseling plaatsvindt
Exclusiviteit	kunnen we de gegevens alleen voor onszelf houden en weten we wie er aan de gegevens gemuteerd heeft