

Expertgroep

product- informatie- management

Productdatakwaliteit in Nederland



Gastheren



Voorzitter



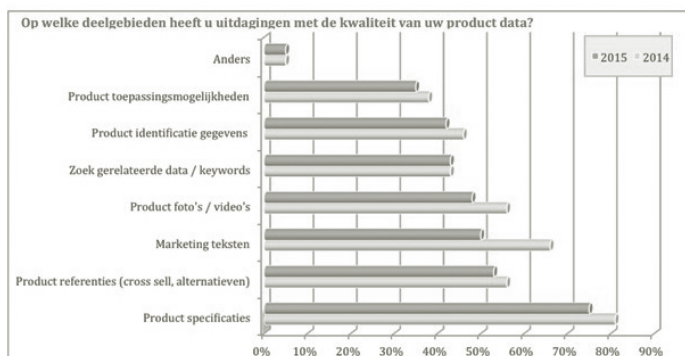
Productdatakwaliteit in Nederland

Het onderwerp *data* is niet sexy en bereikt zelden de agenda van bestuurders. Niet alleen omdat het een complex en enigszins technologisch onderwerp is, maar vooral ook omdat de waarde van data zwaar onderschat wordt. Ontwikkelingen in de markt, zoals de transparantie van internet die klanten in staat stelt producten te vergelijken, en wetgeving maken dat data en datakwaliteit een toenemende belangstelling ontvangen. Het is niet langer de vraag of het een moetje of een weloverwogen keuze is.

1. Datakwaliteit . . . geen ontkomen meer aan

De e-commerce- of e-businessmarkt, groeit jaar na jaar met dubbele cijfers, terwijl het belang van de traditionele verkoop- en marketingkanalen afneemt. Waar de traditionele kanalen zoals winkels en catalogi een beperkte hoeveelheid productinformatie bieden, kennen digitale kanalen zoals websites en e-mailcampagnes nauwelijks technische beperkingen in de hoeveelheid informatie die aangeboden kan worden. In de strijd om de klant gaat het niet alleen over de hoeveelheid informatie, maar vooral ook om de juiste informatie in de juiste context. De aangeboden productinformatie dient van een zo hoog mogelijke kwaliteit te zijn om je als bedrijf te onderscheiden en het vertrouwen van klanten te winnen of te behouden.

Het is echter niet alleen de klant die behoefte heeft aan complete en correcte productdata. Internationale normen en wet- en regelgeving, zoals energielabels, dwingen bedrijven hun productdata-huishouding goed op orde te hebben.



Op welke deelgebieden heeft u uitdagingen met de kwaliteiten van uw productdata?¹

¹ QHUBA PIM ENQUÊTES 2014 EN 2015

In dit hoofdstuk duiden we datakwaliteit in de context van productinformatiemanagement (hierna: PIM) en geven we de lezer een framework om de kwaliteit van productdata te beschouwen en te verbeteren. De eerste stap in dit model is het opstellen van een visie en strategie over datakwaliteit; de ‘waarom’-vraag wordt hier beantwoord.

De doelgroep van dit hoofdstuk zijn mensen met enige voorkennis van PIM. Mocht deze voorkennis niet aanwezig zijn, wordt geadviseerd om het hoofdstuk PIM in het boek “ShoppingTomorrow: Bent u klaar voor de consument van 2020?” te lezen.

2. Visie en Strategie

De missie en visie met betrekking tot datakwaliteit moeten aansluiting vinden bij de missie en visie van de organisatie zelf. De definitie van datakwaliteit is immers niet zonder reden ‘fit for purpose’.

Met ‘passende productinformatie’ bedoelen we:

- voldoen aan de behoefte van leverende processen en gerelateerde partners;
- voldoen aan de behoefte van afnemende processen en gerelateerde klanten;
- voldoen aan de behoefte van analyse om te kunnen meten waar een organisatie staat op het gebied van de bedrijfsdoelen en gekoppelde prognoses.



Echter, wat ‘passende productinformatie’ voor het ene bedrijf is, is onvoldoende voor een ander bedrijf. Om die reden is het dan ook schier ondoenlijk om een generieke visie, missie en strategie op te stellen voor datakwaliteit van productdata. De volgende generieke visie en strategie dient dan ook verbijzonderd te worden naar de eigen bedrijfssituatie.

Visie: Productinformatie van hoge kwaliteit is een steeds belangrijker wordende asset en in toenemende mate bepalend voor het onderscheidende vermogen van zowel B2C- als B2B-bedrijven om te kunnen voldoen aan de groeiende informatiebehoefte van klanten en aan wetgeving op nationaal en internationaal niveau.

Strategie: Groei op een holistische wijze en op basis van een projectmatige aanpak naar een steeds hogere mate van volwassenheid wat betreft de kwaliteit van productdata, hierbij gebruikmakend van het PIM DQ Framework.

Definitie datakwaliteit

Er zijn vele definities van datakwaliteit, waarbij die van J.M. Juran een van de meest gebruikte is: “Data is van hoge kwaliteit indien ze geschikt zijn voor het beoogde gebruik in de activiteiten, besluitvorming en planning.”²

² JURAN, J.M., JURAN'S QUALITY CONTROL HANDBOOK. 4 ED, ED. F.M. GRINA. 1988, NEW YORK: MC-GRAW-HILL, INC.

Het verbeteren van datakwaliteit dient zich niet eenzijdig te richten op de data (-waarden) zelf maar vraagt om een holistische aanpak. Datakwaliteitsproblemen komen immers voort uit: het niet goed definiëren van processen, gebrek aan controles in de processen, onvoldoende ondersteunende systemen, fouten in het datamodel of de data-architectuur, een organisatie die onvoldoende aansluit op de PIM/MDM-problematiek en/of een gebrekkige governance.

3. Het datamodel

Met een datamodel (of gegevensmodel) wordt beschreven hoe de gegevens in een informatiesysteem gestructureerd zijn. In algemene zin zijn er vandaag de dag twee typen: gestructureerde en ongestructureerde datamodellen (bijvoorbeeld een Oracle RDBMS versus een Hadoop-database).

Bij de huidige PIM-systemen is er altijd een gestructureerd deel dat soms wordt aangevuld met een ongestructureerd deel (bijvoorbeeld voor datastreams uit sociale media). Het datamodel is het startpunt waar datakwaliteitseisen kunnen worden afgedwongen, bijvoorbeeld door voor bepaalde velden (attributen) regels te definiëren waaraan de in te voeren gegevens dienen te voldoen. Bij de opzet van het datamodel is het zogenaamde meta-datamodel van groot belang; in het meta-datamodel wordt in feite het ontwerp van het datamodel beschreven. Hierin staan alle specificaties en definities van de gegevenselementen (datavelden en attributen), inclusief hun relaties.

3.1 Datamodel en validaties

Vanuit het perspectief van het datamodel zijn er een aantal aandachtspunten die direct bij de data-invoer in hoge mate de kwaliteit van de data bepalen. Hieronder vallen volledigheid, consistentie, correctheid en de relevantie van gegevens. Aan de hand van een datum-attribuut bespreken we de verschillende aspecten van datakwaliteit.

Volledigheid van de data kan worden afgedwongen door het verplicht stellen van de betreffende attributen. Als we willen dat het veld met de datum steeds gevuld is, moeten we er ervoor zorgen dat dit attribuut verplicht is door dit in de opzet van de database als zodanig te definiëren.

Consistentie van data. Indien informatie niet op een gelijke wijze wordt opgeslagen in attributen is vergelijking en bewerking van deze attributen niet goed mogelijk. Een datum kan op veel verschillende manieren aangeduid worden, bijvoorbeeld:

- Zondag 3 januari 2016
- 03-01-2016
- 03/01/16

In alle voorbeelden wordt een datum opgeslagen in het attribuut. Aan volledigheid wordt voldaan maar niet aan consistentie. Door op het attribuut een invoerregel voor het gewenste formaat in te stellen wordt consistentie afgedwongen.

3.2 Meervoudige relaties en regels

Naast de geldigheid van een attribuut moet dus ook de geldigheid van een record, een gedeelte daarvan, of een combinatie met andere records gewaarborgd worden. Dit gebeurt met zogenaamde *data rules*. In data rules wordt vastgelegd welke afhankelijkheden er tussen attributen en of records bestaan. Een voorbeeld zijn de postcodes die per land verschillen. Een data rule kan dus zijn dat de postcode het formaat "9999" moet hebben indien de landcode "BE" is. Data rules zijn een krachtig hulpmiddel maar zijn ook complexer dan enkelvoudige regels.

3.3 Hiërarchieën en classificaties

Producten worden om meerdere redenen in groepen ingedeeld; bijvoorbeeld voor category-management of voor publicatie en vergelijkingsdoeleinden op een website. Gebruikelijk is dat er minimaal twee structuren worden gehanteerd: een interne en een externe. Bij de opzet van het datamodel is de interne het belangrijkste omdat dit een zuivere hiërarchie dient te zijn waarin ieder product maar één keer voorkomt. De presentatie- of publicatiestructuur aan de andere kant stelt eisen aan de te gebruiken attributen. Bij een classificatie van de zuivere producthiërarchie is het gebruikelijk om met een viertal niveaus te werken waarbij attribuut 'overerving' van hoog naar laag in de boomstructuur plaatsvindt.

3.4 Datamodel en standaarden

Het gebruik van een standaard is een middel om sneller waarde toe te voegen in de samenwerkingsketen. Men spreekt als het ware dezelfde taal bij het hanteren van een standaard. Blijf je gebruik maken van een eigen datadefinitie dan heb je steeds een tolk (converter) nodig.

De attributen voor indentificatie spelen een belangrijke rol bij de koppeling tussen producten en de daarbij behorende productinformatie. Wanneer de productinformatie van externe partijen afkomstig is, moet het identificatienummer ook uniek zijn in de informatiesystemen waar de gegevens worden opgeslagen. Dit kan op twee manieren:

1. Zelf een nieuw uniek nummer toewijzen en een vertaling maken naar het externe nummer van de partij die dit nummer heeft toegekend
2. Of je zorgt dat je gebruikmaakt van wereldwijd unieke nummers zoals GS1-artikelcodes (GTIN). Deze nummers kunnen worden gebruikt door alle partijen in de keten zonder dat er vertaalslagen nodig zijn.

3.5 Datapools

Datapools worden ingezet om de distributie van artikelgegevens tussen grote groepen gebruikers te regelen. Leveranciers sturen hun productgegevens naar de datapool en geven aan welke partijen toegang mogen hebben tot deze data. Afnemers halen realtime of periodiek alle nieuwe data, waartoe zij toegang hebben, op uit de datapool. Op deze manier kunnen beide partijen hun databeheer vereenvoudigen en makkelijk nieuwe handelspartners toevoegen. Datapools maken vaak gebruik van standaarden. Kernpunten zijn:

- Zorg dat de beschrijving van het datamodel (de metadata) goed wordt opgezet en bijgehouden
- De keuze voor een classificatie en hiërarchie heeft grote invloed op de inrichting van het datamodel en de te stellen datakwaliteitseisen
- Standaarden zijn van groot belang voor de datakwaliteit in de keten. Het probleem is wel dat er zoveel standaarden zijn (ter vergelijking met gesproken talen: een Esperanto is er helaas niet)
- Voor identificatie is de trend wereldwijd naar één standaard genaamd GTIN (Global Trade Identification Number).

4. Processen en controls

Gezien vanuit het ketenperspectief, zijn er meerdere momenten waarop het proces van product creatie begint. De fabrikant ontwikkelt het product (bijvoorbeeld in het PLM-systeem), maar ook de tussenhandel en retailer moeten dit product bij opname in hun assortiment creëren en verrijken in hun PIM-omgeving. Bij de beschrijving van hoe datakwaliteit in de productcreatie-processen terugkomt, gaan we hier voor een generiek inzicht alleen in op de PIM-omgeving. Hierbij koppelen we per procesdeel de benodigde controles voor de datakwaliteit.

4.1 Procesdeel productcreatie

1. Communicatie vooraf

Om het proces goed te laten starten met de juiste kwaliteit, is het nodig eerst alle benodigde datavooraarden te communiceren en af te stemmen met betrokken partijen (intern én extern), zodat de data voldoet aan de behoefte van de gebruiker(s). Het is wenselijk om zoveel mogelijk de kwaliteit al bij de invoer te controleren (bij voorkeur automatisch):

- a. Wat: welke data is nodig, in welk veldformaat en volgens welke definities en afspraken? (zie paragraaf 3.5 Datamodel)
- b. Wie: rekening houden met meerdere doelgroepen, bijvoorbeeld interne afdelingen (marketing), uitgaande portalen (webshop) of externe partijen (klanten, retailers).
- c. Wanneer: hoe vaak en op welke momenten is de data nodig?

- d. Hoe: via welk medium (denk aan AS2, e-mail, datapool) en in welke vorm (XML, Excel) wordt de data uitgeleverd?
2. Contractafspraken met leveranciers
Om te garanderen dat afspraken over uitwisseling van data, en daarmee de kwaliteit, worden nagekomen, raden we aan dit onderdeel te maken van het contract, bijvoorbeeld van de Service Level Agreement die geldt voor de leveranciers van de producten. Belangrijk hierbij is dat de kwaliteit van data ook gemeten wordt om de SLA te kunnen toetsen (zie paragraaf 6. Metrics)
 3. Ontvangst input voor PIM
 - a. Bij ontvangst zorgen validaties voor de controle of de data daadwerkelijk voldoende is om in het PIM-systeem (c.q. datapool) te brengen. Afhankelijk van het volume van de ontvangen data, vinden er na ontvangst de volgende validatieprocessen plaats:
 - b. Technische controle: voldoet de data aan de regels van 'Wat'?
 - c. Visuele controle: inhoud is niet altijd geautomatiseerd te controleren, dit is een snelle menselijke check op de data
 - d. Feedback naar toeleverancier van de data voor ontbrekende of onjuiste info. Dit zorgt ervoor dat de kwaliteit in de keten verbetert, zodat alle partijen dezelfde data hebben
 4. Verwerking in PIM
Nadat de data bij ontvangst akkoord is bevonden, start het PIM-verwerkingsproces. Het is aan te raden dit zoveel mogelijk, maar niet volledig, te automatiseren waarbij ondersteuning via een workflow-systeem tegenwoordig gangbaar is.

4.2 Procesdeel productverrijking

Zodra de basis van het artikel in PIM is aangemaakt, is er een productidentificatie (artikelnummer) en kan de aanvulling van de productdata voor de verschillende toepassingen starten. Afhankelijk van de PIM-inrichting, kan elke afdeling zijn benodigde data (laten) aanvullen. Ook is er vaak nog aanvulling nodig vanuit de leverancier, fabrikant en/of datapool. Opnieuw speelt de inrichting van de workflow een belangrijke rol.

Via een workflow is het mogelijk om per stap een controle in te bouwen. Voorbeelden van controles zijn:

1. Fysieke controle
Hierbij wordt het product fysiek met het oog gecontroleerd in vergelijking met de data in het PIM-systeem. Dit zijn controles die niet geautomatiseerd kunnen worden.
 - Klopt het beeld met de productomschrijving?
 - Zijn de afmetingen juist?

2. Logische controles, ook wel: technische controles

Dit zijn automatische controles die in het PIM-systeem of een aanvullend data-qualitysysteem in te richten zijn:

- Zijn alle velden gevuld?
- Zijn de juiste velden gevuld voor dit specifieke product (uitsluitende kenmerken)?
- Zijn de juiste datatypes gebruikt (attribuut lengte, format, enz.)?

3. Relationale controles

Hiermee controleer je een groep van artikelen, ofwel een reeks van vergelijkbare producten. Maar dit is ook toepasbaar voor hetzelfde product afgenomen van verschillende partijen. Dit kunnen geautomatiseerde validaties zijn, maar ook visuele of handmatige.

4.3 Procesdeel productpublicatie

De voorwaarden voor publicatie maken eigenlijk onderdeel uit van het procesdeel 'verrijking', maar is hier verder uitgeschreven. Zodra het product en alle bijbehorende data kwalitatief goed en compleet is, wordt het product gepubliceerd voor alle klanten. Maar wat is compleet? Welke aanvulling is nodig, waaraan dient het product te voldoen? En de definitie die nodig is om te bepalen wanneer een artikel compleet is en dus gepubliceerd kan worden, is afhankelijk van het publicatiekanaal. Per kanaal kan er een andere behoefte en dus kwaliteitsniveau nodig zijn. Voor de publicatie, en dus ook het verrijgingsproces, is het dus noodzakelijk om te weten welke publicatiekanalen er zijn en welke informatie en kwaliteit daarvoor gewenst is.

De eerste stap bij het zorgen voor continue kwaliteit is het inregelen van processen en procedures voor het bijwerken van de productdata. Hierbij dient rekening gehouden te worden met de volgende zaken:

- Wie is verantwoordelijk voor het productdatabeheer?
- Versiemanagement
- Welke data mag wel/niet wijzigen?
- Waardoor ontstaan mutaties, waar kan een wijziging ingevoerd worden?

Een tweede stap is het borgen van kwaliteit, oftewel het wederom toepassen van controles op de datakwaliteit. Zodra de productinformatie opgenomen is in de database is het mogelijk om validaties toe te passen die gericht zijn op de totale consistentie binnen de database. Dit is tevens onderdeel van een goede inrichting voor productinformatiemanagement, waarbij zowel informatie-specifieke als overkoepelende controles uitgevoerd worden. Dit kun je inrichten door dezelfde controles ook hier toe te passen, zoals beschreven in de paragraaf 4.2 Procesdeel productverrijking.

4.4 Procesdeel uitfaseren product

Voor elk product komt er een moment waarop deze niet meer geproduceerd en/of verkocht wordt (afhankelijk van de positie in de keten). Wanneer het zover is, zijn er geen validaties meer nodig op de inhoudelijke data, maar nog wel controles op de juiste afronding van de levenscyclus in het bedrijfsproces.

Het uitfaseren van producten heeft gevolgen voor de klant. Het is nog steeds van groot belang de klant van de juiste informatie, en daarmee kwaliteit, te voorzien. Denk bijvoorbeeld aan:

- online informeren klant over product status;
- prijsaanpassingen;
- advies geven over vervangende producten en opvolgers.

Intern zijn natuurlijk ook nog zaken van belang, zoals:

- afbouwen van de voorraad;
- (wettelijke) bewaartermijnen voor (specifieke) productdata;
- opname van nieuwe (vervangende) artikelen: nu start de cyclus opnieuw met het proces productcreatie.

Het belangrijkste aandachtspunt binnen de PIM-omgeving in het uitfaseren en opnemen van producten is dat de benodigde doorlooptijd gekoppeld is aan het moment van verwijderen van het oude product. Oftewel: stem de levenscycli van oude en nieuwe producten goed op elkaar af en vergeet niet de tijd die nodig is om de informatie gereed te maken voor publicatie.

5. Organisatie en governance

De mate van succes van een bedrijf wordt steeds afhankelijker van de kwaliteit van data en productdata in het bijzonder. Het is dan ook van groot belang om een duidelijke strategie voor datakwaliteit op te zetten en dit binnen de hele organisatie uit te dragen. Het beleggen van deze rollen en verantwoordelijkheden wordt ook wel 'data governance' genoemd, vaak een onderdeel van productinformatiemanagement.

Om te beginnen is het noodzakelijk te bepalen wat het gewenste kwaliteitsniveau van productdata is. Dit is in paragraaf 2. Visie en strategie omschreven als de 'fit-for-use'-toepassing en kan dus variëren per bedrijf of bedrijfsonderdeel.

Hoe beheer en bewaak je de processen om de 'fit-for-use'-kwaliteitsniveaus te borgen? Hiervoor moeten de diverse rollen die relevant zijn voor datakwaliteit ingevuld (organisatie) en bewaakt (governance) worden. Deze rollen zijn te koppelen aan de verschillende niveaus binnen de organisatie:

1. Strategisch
2. Tactisch
3. Operationeel.

Het behalen van de gewenste kwaliteit is hierin dus afhankelijk van de invulling per niveau. De relatie tussen strategie, tactische uitvoering en operationele uitvoering moet helder en volledig doorgevoerd zijn.

5.1 Strategisch niveau

Het strategisch niveau is het meest abstract van de drie niveaus. Dit deel is ingebed op directieniveau. Om het draagvlak voor iedereen duidelijk te maken, is het van groot belang om de eindverantwoordelijkheid binnen het managementteam neer te leggen. Op dit niveau gaat het over de lange termijn en de uitwerking zit daarom dicht tegen het organisatiebeleid aan. De volgende aspecten spelen hierbij een rol:

- Er is vastgelegd hoe goede productdata bijdraagt aan de doelstelling van het bedrijf
- De doelen zijn bekend in de gehele organisatie en de middelen voor het realiseren van de gewenste kwaliteit van productdata zijn beschikbaar
- Er is sturing om rollen en verantwoordelijkheden op tactisch niveau te bepalen en de kwaliteitseisen zijn vastgelegd.

5.2 Tactisch niveau

Het tweede niveau binnen de organisatie betreft de tactische inrichting van de data governance. Hieronder vallen de inrichting van de benodigde productdatabeheer processen en het toewijzen en verdelen van rollen, taken en verantwoordelijkheden. Integraal onderdeel van de beheerprocessen zijn de processen voor het inregelen en meten van de datakwaliteit. Op tactisch niveau dien je ervoor te zorgen dat er een totaaloverzicht geboden wordt van de benodigde datakwaliteit door de hele keten. Alleen dan wordt een goede basis teruggegeven aan het strategische niveau om daar waar nodig bij te kunnen sturen.

Rollen en verantwoordelijkheden

Verantwoordelijkheden voor de datakwaliteit worden bepaald door het belang van de data, de plaats in het proces van creatie of de verwerking van diezelfde data en de rol die een medewerker heeft in de uitvoer of het beheer. Het beleggen van de verantwoordelijkheid voor de kwaliteit van productdata wordt vaak gecombineerd met de reeds bestaande rol. Daarin is iemand verantwoordelijk voor de toepassing en het beheer van betrokken standaarden en processen. Van belang is dat er met de medewerkers ook doelen worden afgesproken (bijvoorbeeld vastgelegd in SLA's) die meetbaar zijn (zie paragraaf 6. Metrics) en onderdeel uitmaken van de jaarlijkse beoordeling.

Changemanagement

Het wijzigen van processen, standaarden en validaties moet op een gecontroleerde wijze plaatsvinden. Met goede procedures voor changemanagement kunnen deze aspecten op correcte wijze worden doorgevoerd. Hierbij is een overzicht op het totaal van de processen van belang, zodat het een interdisciplinair proces is.

De uitvoering van changes is vaak een samenspel tussen het management en commodity-/productexperts of technische engineers. Ook hier kunnen intelligente systemen worden ingezet om suggesties voor changes en alternatieven te doen aan deze interdisciplinaire groep.

Hoe bepaal je de compliance?

Om te bepalen of de organisatie handelt naar de eigen opgelegde normen, moet worden vastgesteld wat de compliance is. Dit wordt bepaald door het periodiek uitvoeren van audits op kennis en processen en door na te gaan of standaarden zijn vastgelegd en op een gecontroleerde manier worden beheerd. Op basis van de resultaten moet op managementniveau (directie) besloten worden welke verbeteringsmaatregelen er nodig zijn.

5.3 Operationeel niveau

Het operationele niveau gaat over de werking van de processen en procedures om de productdata te genereren. Aspecten die hierbij een rol spelen zijn:

- Datakwaliteit-validaties bij het opzetten en onderhoud van productdata
- Contractafspraken over productdatakwaliteit
- Compleetheid van productdata
- Kanaaloverstijgend gebruik van data.

6. Metrics

Nu gaan we in op de grondbeginselen van het meten van datakwaliteit. Hoe er gemeten wordt, wat er wordt vastgelegd en hoe de uitkomsten geïnterpreteerd kunnen worden.

Dataconsistentie

Binnen een organisatie worden over het algemeen meerdere databronnen gebruikt die productdata bevatten. Iedere databron kent daarbij zijn eigen versie en opmaak van de waarheid. In ERP wordt de hoogte bijvoorbeeld als "10cm" weggeschreven terwijl de webshop het als een echt getal "10" opslaat. Als de waarden in de diverse systemen gelijk zijn, is de consistentie hoog. Hoe lager de consistentie, hoe meer actie nodig is.

Data-accuraatheid

Accuraatheid zegt iets over de juistheid van de data. Is de 'witte wasbak' ook wit op de afbeelding. Waar de dataconsistentie iets zegt over de vraag of dezelfde afbeelding is gekoppeld, zegt de accuraatheid iets over de validiteit.

Datacompleteheid

De completeheid zegt iets over de vullingsgraad van de velden. Zijn bijvoorbeeld alle minimale velden ingevuld? De completeheid van een product staat altijd in context met de familie waar deze bij hoort en het kanaal waar deze in wordt gepubliceerd.

Een afbeelding mag niet ontbreken in de webshop maar is niet relevant voor het kassasysteem.

Bandbreedtes en grenzen (metrics)

De kwaliteit van productdata dient gemeten te worden. De vraag is alleen wat wordt er gemeten? Meten is nooit een doel op zich, maar slechts een hulpmiddel om de stand van zaken weer te geven. Het is belangrijk de KPI's goed vast te stellen door te kijken waar data waarde in de keten toevoegt.

Systematiek

Zodra er veel attributen in een systeem aanwezig zijn, is het lastig om goed overzicht te bewaren. Het is raadzaam de juiste systematiek te kiezen en attributen te groeperen. Per groep wordt er een indicatie van de kwaliteit gegeven. Indien deze niet voldoende is, kan er specifiek op het betreffende groepje attributen ingezoomd worden.

Effectiviteit

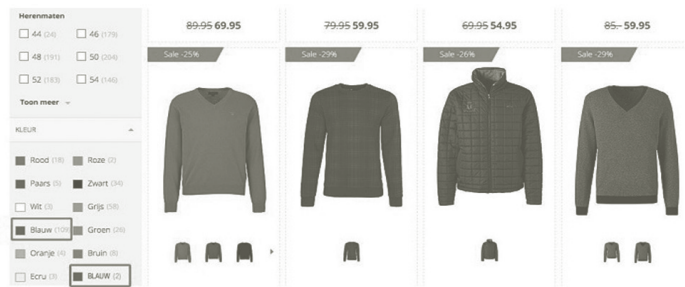
Datakwaliteit moet altijd gerelateerd zijn aan het gewenste eindresultaat, bijvoorbeeld om meer verkoop te bereiken. Iedere meting dient hier rekening mee te houden. Uiteraard is het niet voldoende om te oordelen dat de data niet deugt, enkel omdat een product niet verkocht wordt.

Data-profiling

Zoals eerder is aangegeven, moet je naast de aanwezigheid van data in een veld ook kijken naar de inhoud van het veld om een uitspraak over de kwaliteit te kunnen doen. Dit principe heet *data-profiling*. Bij data-profiling worden overeenkomende producten geanalyseerd om eventuele afwijkingen te ontdekken. Zo kunnen we voor bijvoorbeeld alle balpennen onderzoeken of het attribuut "Kleur" afwijkingen vertoont, met de volgende uitkomst:

Aantal	Waarde
10	BLAUW
30	Blauw
40	Unknown
10	Donker Blauw
10	NULL

De *Completeness Score* zou in bovenstaand voorbeeld 90% aangeven, omdat slechts 10 van de 100 waarden leeg zijn. Op het eerste gezicht dus een prima score. Werkelijke inspectie van de inhoud laat echter een heel ander beeld zien. Dit heeft gevolgen voor bijvoorbeeld de website, waar de gebruiker niet meer in staat is om alle blauwe truien te filteren, sterker nog: er komen meerdere filters blauw naar voren.



Een goede profiling-tool helpt dan ook bij de inspectie van productdata. Het geeft de gebruiker een overzichtelijk beeld van de inhoud en de spreiding van attribuut waarden. Naast de statistische waarde hoe vaak een waarde voorkomt, geeft profiling ook een beeld van:

- minimale en maximale waarden;
- patronen;
- afwijkingen.

Condities

In bovenstaand voorbeeld moet de gebruiker iedere mogelijke waarde een voor een analyseren. Daarom biedt een goede profiling-tool de mogelijkheid om logische grenzen te zetten. Hiermee bepaalt het systeem zelf wanneer attributen buiten de grenzen lopen.

Proces meetwaarden

Tot slot zijn er nog de meetwaarden die zijdelings met datakwaliteit te maken hebben en bijvoorbeeld aantonen hoe efficiënt het proces verloopt. Een bekende meetwaarde is de zogenaamde 'time to market' ofwel hoelang het duurt voordat een product verkocht kan worden op de website. Ook andere doorlooptijden zijn meestal vanuit het workflow-systeem te herleiden.

7. Systemen

Productdatakwaliteit gaat verder dan een PIM-systeem en betreft ook het optimaliseren van de processen, de organisatie en het opzetten van een datakwaliteitssysteem. Met betrekking tot datakwaliteit worden verschillende systemen ingezet, waarvan het PIM-systeem de belangrijkste is.

7.1 Kwaliteit in vier stappen

Stap 1: Data-onboarding

Producten worden niet altijd 'geboren' in een PIM-systeem. Dit vindt regelmatig plaats in een ERP- of PLM-systeem. Hierdoor moet met enige regelmaat productdata in PIM geladen worden, waarbij de eerste kwaliteitscontrole plaatsvindt. Foutieve, incomplete of incorrecte informatie wordt niet door het PIM-systeem geaccepteerd

of zorgt binnen het PIM-proces voor problemen. Een *datacleansing-tool* kan hierbij uitkomst bieden en bewaakt de 'inbound QA'.

Stap 2: Informatieverrijking

Enmaal langs de eerste kwaliteitscontrole van data-onboarding, vindt de verrijking van de productinformatie plaats. Er wordt rijke productdata toegevoegd en relaties met andere producten of diensten. Zo kan een wasmachine aan een garantie gekoppeld worden en een laptop aan een tas.

Stap 3: Delen van productinformatie

Hierbij wordt gekeken welke informatie er per product naar welke kanaal moet worden gedistribueerd. Een foto met een hoge resolutie werkt bijvoorbeeld niet in het mobiele kanaal, terwijl dit voor print-publicaties een must is.

Stap 4: Kwaliteit (effect) meten van je productinformatie

Tijd is geld en niet alle productdata hoeft 100% te zijn. Dit zou te kostbaar worden. Tools als Google Analytics helpen bij het prioriteren van het werk. Zo kunnen veel gezochte of bekeken producten voorrang bij de verrijking krijgen.

7.2 Workflowsysteem

Het woord 'workflow' is al op diverse plaatsen gevallen en is een systeem dat niet alleen van toepassing is voor datakwaliteit. Een workflow kan eenvoudig zijn in een kleine organisatie en meer complex als het om grotere organisaties gaat. Cruciaal is dat het workflowsysteem het gehele datakwaliteitsproces ondersteunt zodat er ook zinvolle rapportages te maken zijn over de procesefficiëntie. In sommige gevallen heeft een PIM een eigen workflow-oplossing die eventueel gekoppeld kan worden aan een ander workflow- of ticketingsysteem zodat het gehele proces wordt afgedekt.

7.3 Autorisatie

Om datakwaliteit te optimaliseren moeten de juiste mensen op het juiste moment de juiste acties kunnen uitvoeren. Zo zijn er binnen een PIM-omgeving mensen die data creëren of bewerken, andere mensen die controleren en mogelijk weer andere mensen die accorderen. Om een succesvol datakwaliteitsmodel te implementeren dienen deze verantwoordelijkheden en autorisaties te worden beschreven en waar mogelijk te worden vastgelegd in systemen, procedures en een organisatorische inrichting.

7.4 Dashboards

Organisaties die met een PIM-systeem werken, kiezen hier in de regel voor omdat ze werken met veel leveranciers, veel producten, productvarianten, assortimenten en

kanalen. Om controle op de kwaliteit van productdata te (onder)houden, maken de meeste PIM-systemen het mogelijk om dashboards op te stellen.

Veelvoorkomende zaken die gemeten worden, zijn productstatus (fase van verrijgingsproces), completeness, maar ook de snelheid waarmee een nieuw product gelanceerd wordt.

Meer lezen?

Ga naar EcommerceWiki.org om de volledige bluepaper van de expertgroep Productinformatiemanagement te downloaden.



*EcommerceWiki/
ShoppingTomorrow*

GASTEREN



Edward Haarmans
Country Manager Benelux
inRiver



Arjan Herskamp
Owner
MyDataFactory

VOORZITTERS



Wim Griffioen
Partner
Qhuba



Jos Schreurs
Partner
Qhuba

Leden expertgroep



Meindert Boorsma
Associate
Qhuba



Joost Halkes
Director Product & Content
bol.com



Frits van den Bos
ECR Manager
GS1



Andre Hageenaar
E-business Manager
Econosto



Lennart Drewes
Business Developer E-commerce
Koninklijke Jongeneel BV



Laurent Chavagne
Purchasing Director
Mag45



Daniëlle Hoekstra
Manager Product Data Management
Technische Unie



Fulco Hofmann
Project Manager Product Data Integrity
Walraven Group



Rob Intven
Artikel Data Manager
JéWéRET
Verkoopmaatschappij B.V.



Angelique Vervloet
Manager Product Informatie
Fabory



Roger Kremer
Directeur
Nigella IT BV



Luke Theissing
Client Services Manager
inRiver



Sylvia Stein
Project Manager
GS1



Onno Cleijpool
Associate
Qhuba



Peter Lummen
Manager Product Informatie
Saint Gobain



Harriet van der Winden
Director IT
Landal



Erwin Lazet
CTO
MyDataFactory



Frans Bouwmeester
Hoofd Bedrijfsbureau
DMG



Yvonne van Bruggen
Marketing and Communication
Coordinator
B&S



Norbert Donders
Associate
Qhuba