



Larry English over 'information quality':

# Denk in waardeketens

Robbert Hoeffnagel

**Natuurlijk vindt iemand die van 'information quality' zijn werk heeft gemaakt dat er te weinig aandacht wordt besteed aan de kwaliteit van informatie. Maar Larry English, directeur van Information Impact International en een bekend spreker op dit gebied, onderbouwt zijn visie met goede argumenten. "Informatie van slechte kwaliteit zorgt voor enorme verspilling – van tijd, geld, systemen, van klanten, noem maar op. Denk daarom vooral in waardeketens."**

"Voegt een bepaalde handeling of bewerking waarde toe? Of eigenlijk alleen maar kosten?" Larry English begint zo langzamerhand een bekende persoonlijkheid op het gebied van informatie-kwaliteit te worden. Zijn voordrachten worden druk bezocht, zijn artikelen goed gelezen. Vorig jaar juli behandelde hij in 'Plain English about information quality' – zijn column in het maandblad 'DMReview' – een reactie die hij van een lezer had ontvangen. Het leverde een mooi voorbeeld op van de impact die slechte informatiekwaliteit kan hebben. In deze column liet hij een Engelse ambtenaar aan het woord die zich beklaagde over een mislukt ERP-project bij het Britse ministerie van Defensie. Op tal van gebieden ging het fout. Zo was het ERP-pakket ontwikkeld voor de nucleaire industrie, maar werd het zonder al te veel aanpassingen in een defensie-omgeving ingevoerd. Het gevolg? De eindgebruikers werden met termen en kreten geconfronteerd die voor hen vreemd waren. "Een typisch voorbeeld van het ontbreken van informatiestandaarden", zegt English in een toelichting. "Begrippen en afkortingen die in een applicatie worden gebruikt, moeten absoluut intuïtief zijn voor de gebruiker."

## Ramp

Er ging meer fout. Bijvoorbeeld met de gebruikersinterface. Vensters waarin de eindgebruikers geacht werden gegevens in te voeren, waren soms veel groter dan de hoeveelheid data vereiste. Dit leidde tot misverstanden, omdat gebruikers het gevoel hadden dat zij meer gegevens moesten invoeren. Waarom waren anders die vensters zo groot? Om de verwarring nog groter te maken, telden sommige vensters twee regels. Op ongeveer driekwart van de lengte van de eerste regel brak de applicatie woorden echter af en plaatste vervolgens de rest van de invoer op de tweede regel. Sterker nog, in verschillende schermen bleek bij het invoeren van een en hetzelfde gegeven tot het afbreken van de eerste regel op soms zeer verschillende posities plaats te vinden.

"Dit is een mooi voorbeeld van inconsistentie tussen de lengte van een veld in de database en de lengte van een veld in de applicatie. Zou iets dergelijks in een auto optreden, dan zou dit ongetwijfeld leiden tot een actie waarbij alle getroffen modellen door de fabrikant worden teruggehaald. Het garandeert naar mijn mening namelijk dat er geen enkele grip meer is op de kwaliteit van de informatie. Ik gebruik daarvoor wel eens de kreet 'nonquality information'." Het werd bij het ERP-project allemaal nog erger, omdat hetzelfde gegeven (de 'full item description' van een artikel) in sommige schermen in slechts één erg lange regel diende te worden ingevoerd. "De kans was groot, zo schreef de Engelse ambtenaar wiens klacht ik in mijn column behandelde, dat de eindgebruiker hierdoor niet eens in de gaten heeft dat er data ontbreken", meent English. "Temeer omdat het verschil tussen artikelen – bijvoorbeeld de maat van soldatenlaarzen – pas in de laatste posities worden aangegeven. Dit ERP-systeem was vanuit het oogpunt van informatiekwaliteit een ramp. Uiteindelijk zag ook het Britse ministerie dit in en het project werd geschrapt. De kosten? Honderdveertig miljoen pond, naar ik begrijp."

English: "De kwaliteit van informatie is helaas bij veel managers geen onderwerp waar zij zich veel mee bezig houden. Dat is een groot probleem, want slechte informatiekwaliteit leidt tot enorme verspilling. Van geld, van tijd, van resources en het kan een bedrijf bovendien klanten kosten."

## Overproductie

Over informatiekwaliteit wordt veel gepraat, maar wat is het nu eigenlijk? English verstaat onder 'kwaliteit' het continu voldoen aan de verwachtingen van klanten. Gaat het om 'informatiekwaliteit' dan breidt hij die doelgroep uit met kenniswerkers binnen de eigen organisatie. Informatiekwaliteit kan bovendien gemeten worden. De vraag is alleen: wat meten we dan precies?

"Er spelen natuurlijk nogal wat aspecten mee. Het gaat om de vraag of de data accuraat zijn, maar evenzo goed of de data valide zijn. Zijn ze compleet? Zijn ze op tijd beschikbaar? Het zijn allemaal punten die we mee moeten nemen. Data kunnen immers wel accuraat zijn, maar desondanks niet valide."

Een slechte informatie-kwaliteit leidt tot verspilling, meent English. Hij onderscheidt negen soorten of typen van verspilling (zie het kader 'Negen soorten van verspilling'). Hij maakt hierbij regelmatig een vergelijking met de industrie. "De eerste soort verspilling is overproductie. Hoeveel systemen heeft een bedrijf in principe nodig om basisinformatie over een klant vast te leggen? Precies: één. In werkelijkheid gebruiken we er echter veel meer. Dat komt omdat bedrijven verticaal georganiseerd zijn. Het verzekeringsproduct 'levensverzekering' heeft klantgegevens nodig, dus bouwen we applicaties en databases om die vast te leggen. Hetzelfde geldt voor het product 'autoverzekering': aparte applicaties, aparte databases. Dat is verspilling. We weten als bedrijf al veel van die klant, maar kiezen er toch voor die gegevens opnieuw vast te leggen. Maar ook binnen één applicatie en één database komt verspilling voor. Bijvoorbeeld: vanwege tijdsdruk kijkt een medewerker van een call center niet of een klant reeds eerder een melding of verzoek heeft gedaan, maar maakt simpelweg een nieuwe melding aan. Dezelfde persoon komt hierdoor meerdere keren in een en dezelfde database voor. Ook het feit dat gebruikers vaak niet voldoende hebben aan de functionaliteit van een applicatie en dus zelf een extra administratie gaan bijhouden, compleet met een eigen en vaak officieel niet bestaande databases is een vorm van verspilling."

De kosten kunnen hierdoor enorm oplopen, meent English. "We bouwen en onderhouden meer applicaties dan we feitelijk nodig hebben, terwijl veel van die systemen ook nog eens van interfaces moeten worden voorzien om gegevens met andere applicaties te kunnen uitwisselen. Maar ook moet tijd en energie worden gestoken in het vergelijken, consolideren en samenvoegen van data uit al die verschillende onderliggende databases."

## Ook binnen één applicatie en één database komt verspilling voor

Laten we bovendien niet vergeten dat als er in drie of vier databases gegevens over een en dezelfde klant vastliggen, de kans op het missen van commerciële kansen natuurlijk vrij groot is. De kans is immers groot dat de verkoper hierdoor niet beschikt over alle relevante gegevens."

### Voorraadbeheersing

"Het is in feite een vorm van voorraadbeheersing en dat is niet de sterkste kant van de ICT-wereld. In de industrie vertegenwoordigt



Larry English: "We bouwen en onderhouden meer applicaties dan we feitelijk nodig hebben"

ieder artikel dat op voorraad wordt gehouden niet zozeer een bepaalde zakelijke waarde, maar wordt vooral als een kostenpost gezien. Het is daarom uiterst opmerkelijk dat digitale informatie het enige niet-verbruiksartikel binnen een organisatie is. Ontwikkel liever één centrale database met alle gegevens van klanten, zodat iedere medewerker deze gegevens kan gebruiken. Ik kan eigenlijk maar één reden bedenken waarom een gegeven op twee plaatsen mag worden vastgelegd en dat is wanneer dat gegeven wordt gekopieerd naar een datawarehouse om te worden gebruikt voor strategische of tactische analyses."

Een grote organisatie legt ieder gegeven gemiddeld op tien verschillende plaatsen vast, nog afgezien van de datawarehouse-omgevingen. Dit wil dus zeggen dat maar liefst negentig procent van de vastgelegde data overbodig is en in feite een gevolg is van slecht voorraadbeheer. "In de praktijk hoor ik vaak argumenten als 'we leggen data meerdere keren vast in verband met database performance' of 'we delen niet dezelfde klanten'. Vanuit zakelijk oogpunt zijn dit echter zelden valide argumenten." Meestal gaat het simpelweg om een klassieke fout, meent hij: gegevens worden beheerd op het niveau van afdelingen en niet op bedrijfsniveau. Veel organisaties worden aangestuurd als een industrieel bedrijf en niet op basis van waardeketens. Wie wat aan de kwaliteit van zijn informatie wil doen, zal data op bedrijfsniveau moeten

beheren. Vervolgens kan technologie worden toegepast om de operationele vraag naar het gebruik van deze gegevens mogelijk te maken. De gebruikte technologie mag de datamodeltering echter op geen enkele manier verstoren.

## Reparatie

English stoot zich aan het feit dat nogal wat mensen – ook database-deskundigen – denken dat 'data cleansing' hetzelfde is als 'information quality management'. "Dat is natuurlijk onzin. Data cleansing is een kostenpost die het gevolg is van het feit dat informatie een slechte kwaliteit heeft. We hebben het over het herstellen van fouten doordat een gegeven niet correct is vastgelegd of later verkeerd is bewerkt." Dat is heel iets anders dan het beheren van de kwaliteit van informatie. Bedrijven die geld investeren in software voor data cleansing volgen in feite een verkeerde aanpak. De kwaliteit van informatie is het gevolg van het proces waarmee gegevens worden vastgelegd. Verbeter dit proces en de behoefte aan data cleansing neemt af.

## Applicatieontwikkelaars zijn in veel gevallen de helft van hun tijd kwijt met het onderhouden van interfaces

Er ontstaan grote problemen als eindgebruikers twijfels hebben over de kwaliteit van de vastgelegde data. Dan gaan zij in veel gevallen een eigen administratie aanleggen, met alle gevolgen van dien. "Het is opmerkelijk om te zien dat zowel ICT-managers als business managers vaak dezelfde fout maken: zij verwijten die gebruikers dat zij een eigen gegevensverzameling hebben aangelegd. Dat is natuurlijk de wereld omdraaien. Richt die energie liever op het verbeteren van het proces waarmee de gegevens langs de officiële weg worden vastgelegd. Kijk dus ook naar de vraag waarom zij besloten hebben een eigen administratie aan te leggen."

Over de relatie tussen information quality management en data cleansing zegt English dat het doel van een weldoordachte vorm van beheer van informatie kwaliteit het voorkomen is van situaties waarin het opschonen van gegevensbestanden noodzakelijk wordt. Data cleansing maakt dus per definitie geen deel uit van een information quality management-beleid, stelt hij vast. "In de praktijk moeten we soms echter toch aan data cleansing doen. Naar mijn mening zal er dan echter altijd sprake van een dubbelslag dienen te zijn: een eenmalige opschoonactie in een specifieke database die is gekoppeld aan een verbetering van het proces waarmee deze data worden vastgelegd. Wie dat niet doet, veroordeelt zichzelf tot het keer op keer uitvoeren van data cleansing. En dat kost veel geld."

## Beweging

Als op een fabrieksvloer de opstelling van machines, materialen en mensen niet optimaal is geregeld, gaat vaak veel tijd verloren omdat mensen teveel moeten lopen om hun werk te kunnen doen. Hetzelfde geldt voor assemblage-lijnen die te lang zijn. "Tets dergelijks gebeurt ook wanneer gebruikers teveel gegevens moeten invullen. Het klassieke voorbeeld is natuurlijk: waarom een gebruiker of een klant een blanco formulier voorschotelen als je ook een reeds gedeeltelijk ingevuld formulier kunt presenteren? Dat scheelt niet alleen tijd, maar levert ook een betere kwaliteit op." Er kunnen bovendien tal van hulpmiddelen worden toegepast. Leg postcode en huisadres vast en achterhaal daarmee het adres. Gebruik dropdown menu's om bijvoorbeeld lijsten van branches en dergelijke te presenteren en laat de gebruiker waar mogelijk keuzes maken door in checkboxes vinkjes te plaatsen. Creëer bovendien een schermhiërarchie die zodanig in elkaar zit dat een gebruiker met minimale navigatie bij de juiste schermen terecht komt. Gebruik een pull-mechanisme zodat een gebruiker een lijst kan opvragen en invullen, waarna deze bij voorkeur lokaal kan worden geprint. "Nog teveel ICT-afdelingen drukken dit soort lijsten centraal af en versturen deze vervolgens per interne post."

## Verwerking

Waarom ontwikkelen veel ICT-afdelingen toch steeds nieuwe applicaties om bepaalde gegevens vast te leggen, terwijl zij allang over toepassingen voor hetzelfde doel beschikken? De reden die English vaak hoort, is dat op die manier applicaties sneller beschikbaar kunnen worden gesteld. Pure verspilling, meent hij, en wel om meer dan één reden. Soms worden min of meer kopieën van informatiesystemen gemaakt om – bijvoorbeeld – verschillende klanten te bedienen. Niet ideaal, meent English, maar op zich nog geen onoverkomelijk probleem zolang de twee



**GURU4Pro**  
**Dr. Harm van der Lek**  
'Meta Driven  
Data Warehouses'

Woensdag 19 april  
in Amstelveen

**GURU  
4PRO**

Geïnteresseerd in deze of andere Guru4Pro's?  
Kijk op [www.logicacmg.nl/guru4pro](http://www.logicacmg.nl/guru4pro).

### Negen soorten van verspilling

Larry English verwijst in zijn betogen over informatiekwaliteit graag naar het werk van de Japanner Masaaki Imai. In zijn boek 'Gemba Kaizen' beschrijft hij een aantal soorten van 'muda' (Japans voor verspilling) die een slechte informatiekwaliteit tot gevolg heeft: overproductie; voorraad; herstel en afwijzing; beweging; verwerking; transport; wachten en tijdverlies; procesverstoring; foute beslissingen.

systemen identiek blijven. Als na verloop van tijd de eisen die aan het ene systeem worden gesteld echter gaan afwijken van die voor de andere applicatie, dan ontstaat een stevig onderhouds-probleem. Wie in deze situatie met gekochte software werkt, maakt dus simpelweg teveel licentiekosten, luidt zijn stelling. In andere gevallen worden gegevens eerst op papier ingevuld, vervolgens overgedragen aan een andere functionaris waarna alle data in een systeem wordt ingevoerd. Ook dan geldt dat er in feite twee systemen worden gebruikt voor het vastleggen van dezelfde informatie. Dat levert bovendien een grotere kans op fouten op dan wanneer de gegevens in een keer via bijvoorbeeld een laptop

worden vastgelegd. Ook een slecht ontworpen formulier kan tot problemen leiden, terwijl door de vertraagde invoer van de gegevens in een database de informatie bovendien minder actueel en up-to-date zal zijn dan mogelijk of gewenst is.

Veel verspilling ontstaat ook doordat rapporten worden gedraaid die eigenlijk niet meer nodig zijn. Het aantal rapportages dat in de praktijk door niemand meer wordt gebruikt, is vaak vrij groot. Besparingen van tachtig procent of meer zijn zeker geen uitzondering. Bovendien zijn interessante besparingen mogelijk door een reeks van standaard rapportages aan te maken die door middel van parameters kunnen worden aangepast.

Een veel voorkomende situatie is een applicatie die data uit een database haalt, deze gegevens bewerkt en vervolgens in een andere database plaatst. Data die al in een database aanwezig zijn extraheren en transformeren, voegt echter geen waarde toe maar levert alleen maar kosten op, meent English. Bovendien voegt het complexiteit toe, omdat data in meerdere databases voorkomen en consistent moeten worden gehouden, wat met name lastig zal zijn als de data in verschillende formaten worden vastgelegd. Hoewel de ene eindgebruiker vaak net even iets andere informatie over klanten of producten nodig hebben dan de andere, is er volgens English geen enkele reden om meerdere applicaties te



## CaseTalk

Fact-based information modeling made easy



Informeer nu naar onze vrijblijvende modelleersessies bij u op locatie.

CaseTalk, de workbench voor feitgebaseerd informatiemodelleren, helpt u om de complexiteit van uw IT-systemen in de vingers te krijgen. CaseTalk verzekert u van kraakheldere user requirements als fundament voor hoogwaardige IT-oplossingen.

CaseTalk werkt met uitspraken over feiten en regels die mensen in de organisatie gebruiken. Daardoor is een gebruikersorganisatie op elk moment in staat om te controleren op juistheid en volledigheid. Bovendien bereiken gebruikers sneller overeenstemming over requirements, omdat inconsistenties en dubbelzinnigheden direct aan het licht komen.

CaseTalk beschikt over diverse modules voor forward engineering. U kunt direct een database of datawarehouse creëren of uw lower case tool van een ER-model, dimensionaal model of UML-schema voorzien.

Met de reverse modeling module brengt u het verborgen informatie-model in bestaande systemen terug naar de mensen-wereld.

**CASETALK LAAT DE FEITEN VOOR ZICH SPREKEN.**



creëren om basisgegevens als bedrijfsnaam, adresgegevens en dergelijke mee vast te leggen. Er is ook geen enkele zakelijke rechtvaardiging te geven voor het op meerdere manieren coderen van dezelfde informatie. Bijvoorbeeld of een klant al of niet getrouwd is.

## Transport

In de industrie is het verplaatsen van goederen vaak pure noodzaak, maar voegt het geen waarde toe. Het is een kostenpost, waarbij bovendien de kans op beschadiging tijdens transport nog eens tot extra uitgaven als verzekeringspremies met zich meebrengt. Hebben we het over informatiekwiteit, dan geldt in feite hetzelfde. Ook dan geldt dat er vooral kosten ontstaan en geen waarde wordt toegevoegd. Waarom is die tweede database nodig als de bewerking net zo goed door een applicatie kan gebeuren die rechtstreeks met de oorspronkelijke data werkt? 'Data movement' creëert vaak alleen maar nieuwe momenten waarop fouten kunnen ontstaan. Niet doen dus, adviseert English.

## Werkprocessen kunnen tevens problemen vertonen doordat data 'defect' zijn

Daar komt nog een tweede argument bij. Uit onderzoek blijkt dat applicatieontwikkelaars in veel gevallen de helft van hun tijd kwijt zijn met het onderhouden van interfaces die data van de ene database naar de andere overbrengen. Er is slechts één argument om data wél te transporteren en dat zakelijk gezien valide is, meent English. Dat is in die situaties waarbij de data op het punt waar deze worden vastgelegd niet rechtstreeks in een deelbare database kunnen worden geplaatst.

Het is van groot belang om ten aanzien van informatiekwiteit te denken in waardeketens en kostenketens. Zijn werk heeft English namelijk geleerd dat de meeste bedrijven meer geld uitgeven aan datatransport dat geen enkele waarde toevoegt, dan aan handelingen die wél waarde toevoegen.

## Online archief Database Magazine

Database Magazine-lezer opgelet! Artikelen over onderwerpen als Datawarehousing, SQL, ETL, Business Intelligence, Relationale databases, modellering en nog veel meer vindt u in het Online Archief van Array Publications. Vaktijdschriften als Storage Magazine, Database Magazine, IT Service Magazine, Java Magazine en ons Oracle vakblad Optimize hebben hun artikelenarchief online gezet. Met een Google-achtige zoekstructuur vindt u snel wat u zoekt op [www.dbm.nl](http://www.dbm.nl)

## Wachten, procesverstoring en suboptimalisatie

'Information float' is eveneens een probleem. Het gaat er in dit geval om dat eindgebruikers van een organisatie moeten wachten totdat een bepaald gegeven waarvan men weet dat het aanwezig is, ook daadwerkelijk beschikbaar is. Dit levert dus een wachttijd op die soms wel en soms niet productief kan worden doorgebracht. In het ene geval gaat het over seconden, maar wanneer sprake is van complexe datatransporten en werkprocessen, kan dit oplopen tot uren of zelfs dagen. Doordat gegevens niet (near) realtime beschikbaar zijn, kan de organisatie dus gemakkelijk kansen mislopen.

Werkprocessen kunnen tevens problemen vertonen doordat data 'defect' zijn. Het begrip 'defect' dient hierbij breed geïnterpreteerd te worden. Het kan bijvoorbeeld betekenen dat de datadefinitie niet voldoende duidelijk is of dat bepaalde data simpelweg ontbreken. We kennen allemaal de problemen die ontstaan als namen foutief gespeld worden, waardoor een en dezelfde klant meerdere keren in een database voor komt. Hierdoor ontstaat een foutieve lijst met klanten, met alle gevolgen van dien voor bijvoorbeeld de kosten voor direct mail of de kwaliteit van analyses.

## Conclusie

Er bestaat ten aanzien van informatiekwiteit een groot misverstand, meent English. Het probleem is namelijk niet die kwalitatief laagwaardige informatie zelf. Waar het wel om gaat zijn de kosten die hiermee gemoeid zijn. Dat zijn de kosten voor herstel maar ook de kosten van misgelopen omzet of weggelopen klanten. Bij het opstellen van de business case voor informatiekwiteit draait het dan ook om twee aspecten: de kosten van slechte kwaliteit en de uitgaven die nodig zijn om de processen waarmee data worden vastgelegd te verbeteren.

**Robbert Hoeffnagel** is freelance journalist.

The poster features a yellow header with the 'logicaCMG' logo. Below this, the text reads 'GURU4Pro Dr. Harm van der Lek "Meta Driven Data Warehouses"'. The event date and location are 'Woensdag 19 april in Amstelveen'. At the bottom, there is the 'GURU4PRO' logo and the text 'Geïnteresseerd in deze of andere Guru4Pro's? Kijk op [www.logicacmg.nl/guru4pro](http://www.logicacmg.nl/guru4pro)'.