

Van wezenlijk belang voor de data-huishouding

Masterdata en metadata

Karien Verhagen

Masterdata en metadata staan volop in de belangstelling. Ze worden vaak in één adem genoemd. Toch zijn ze heel verschillend. Eigenlijk hebben ze weinig anders gemeen dan hun vitaal belang voor de data-huishouding. Er wordt veel geschreven en gezegd over masterdata management en metadata management. Dit artikel zet ze naast elkaar om ze uit elkaar te halen.

Als definitie voor masterdata voel ik het meest voor de uitleg van Malcolm Chisholm¹. Masterdata zijn in zijn uitleg basisgegevens die op meer plekken binnen en buiten het bedrijf worden gebruikt. Het is belangrijk dat die gegevens slechts éénmaal worden geregistreerd en op één plek onderhouden. Dat sluit niet uit dat die gegevens vele malen worden gebruikt. Denkt U bijvoorbeeld eens aan het belang van één klantenbestand voor gefuseerde banken en verzekeringsmaatschappijen. Zo weet de ING nog steeds niet dat ik een hypotheek heb bij de Postbank. De twee beconcurreren elkaar vrolijk met aantrekkelijke offertes. Dat is niet alleen lastig voor de bedrijfsvoering maar zeker ook voor de management-informatie. Zolang er niet een leidend en betrouwbaar klantenbestand is kan het datawarehouse niets zinnigs zeggen over het aantal nieuwe klanten of de gemiddelde omzet per klant.

Metadata zijn als het ware de katalysatoren

Chisholm brengt nuances aan tussen referentietabellen, masterdata en codetabellen. Referentietabellen worden in zijn uitleg gebruikt om masterdata te categoriseren zoals landcodes, maar ook valuta's of klanttypes. Een code kan een masterdata-element identificeren volgens bepaalde conventies. Zo zouden in afbeelding 1 de luchthavens gecodeerd kunnen zijn volgens de afspraken van de ICAO of de IATA.

Records uit de Item Identificatie Conventie tabel zijn:

Item Identificatie Conventie Code	Item Identificatie Conventie Naam
ICAO	International Civil Aviation Organization
IATA	International Air Transport Association

De masterdata tabel bevat de gegevens van Amsterdam Airport Schiphol:

Item Gegeneerde Sleutel	Item naam	Item Attribuut1	Item Attribuut2
234	Amsterdam Airport Schiphol	Netherlands	Amsterdam

Luchthavens kunnen ook gecategoriseerd zijn naar bijvoorbeeld de lengte van de landingsbaan. Zo'n categorieaanduiding zou Chisholm plaatsen in een referentietabel. Het masterdata item kan volgens verschillende conventies gecodeerd worden:

Item Gegeneerde Sleutel	Item Identificatie Conventie Code	Item Identificatie
234	ICAO	EHAM
234	IATA	AMS

Op deze manier wordt de naamgevingsconventie losgekoppeld van het 'echte' masterdata item. Het probleem dat een code niet uniek zou zijn of dat een lijst van luchthavens noch volgens de ene noch volgens de andere codering volledig zou zijn, is daarmee impliciet opgelost.

Dat Chisholm slechts één van de definities geeft, blijkt bijvoorbeeld uit een white paper van Knightsbridge. Volgens deze definitie zijn masterdata "the controlling set of data values for reference data, which is a category of data that defines the business context of transactions. Reference data becomes master data when business constituents across the organization agree on its definition".²

Volgens de beschrijving op Internet van Dataflux, eigendom van SAS, zijn masterdata objects "core business objects that are used

in the different applications across the organization".³ In het Centennium BI-woordenboek ontbreekt een beschrijving van masterdata.

Functie

Over de exacte scheiding tussen referentiedata, codetabellen en masterdata is men het blijkbaar nog niet eens. Over de *functie* van masterdata in relatie tot datawarehousing is meer duidelijkheid. Masterdata zijn voor het datawarehouse belangrijke elementen omdat het volgens elke definitie systeemoverschrijdende basisgegevens zijn die in het datawarehouse, al of niet slowly changing, de dimensies vormen. Het datawarehouse moet er ook vaak meer versies van bijhouden voor een juist historisch inzicht. Toch vormen ze geen exclusief onderdeel van de BI-infrastructuur. Systemen die masterdata onderhouden zijn *operationele* systemen. Je moet klanten kunnen opvoeren, afvoeren, wijzigen. Ondanks het feit dat ze voor het datawarehouse van levensbelang zijn ligt de verantwoordelijkheid voor het onderhoud dus *niet* bij datawarehouse-beheer. Het zijn bronsystemen.

Toch worden datawarehouse en masterdata vaak in een adem genoemd. Dat is verklaarbaar. Voor goed masterdata management is kennis van data-integratie, cleansing, datakwaliteit, gegevensaudits en brongebruik belangrijk. Die kennis zit vaak bij het datawarehouse team. Het voert echter te ver om de verantwoordelijkheid voor goed masterdata management bij het DWH-team te leggen. Doel van masterdata management is eenmalige vastlegging, duidelijke verantwoordelijkheden, eenduidige definities over systeemgrenzen heen. Er moet een eigenaar worden gezocht van elke basisregistratie. De afgifte van de masterdata moet bij voorkeur via SLA's worden geregeld. Nog mooier is het natuurlijk als ze als een SOA-dienst beschikbaar worden gesteld.

En er is nog een tweede link met datawarehousing. Masterdata management is een technisch lastig karwei maar de *zachte* kant is net zo belangrijk. Gegeveneigendom, geringe betrouwbaarheid, lokale heiligdommen, gebrekkige communicatie, privacy-aspecten en verwachtingsmanagement: ook bij MDM liggen ze op de loer en kost het pareren van de risico's net zoveel aandacht als de techniek.

Bij BI dient de business naast ICT aan het stuur te zitten. Ambitieniveau, 'readiness', bereidheid tot samenwerking; u kunt zo langzaam de bekende randvoorwaarden voor BI wel opsommen. Die gelden voor MDM net zo. MDM dient te leiden tot 'data governance', maar als de tijd rijp is en als de business toe is aan data governance zal de business daarvoor waarschijnlijk een andere term hanteren.

Masterdata management en datawarehousing hebben een verschillend doel. Masterdata management beoogt de integratie van administraties van basiselementen, waaraan door vele bedrijfsonderdelen wordt gerefereerd. Die integratie bespaart administratieve kosten en ondersteunt de één-loket gedachte.

Denkt u daarbij bijvoorbeeld aan de overheid die deze gedachte veel en hardop uit. Eenmalige vastlegging verhoogt bovendien de betrouwbaarheid en de kwaliteit van de gegevens en maakt transacties of feiten uit verschillende systemen vergelijkbaar. Dat gebeurt in de Business Intelligence omgeving. Een BI-omgeving moet betrouwbare informatie leveren om de bedrijfshuishouding en de uitvoerende processen te verbeteren. Goed masterdata management is dus een voorwaarde voor het correct functioneren van een BI-omgeving.

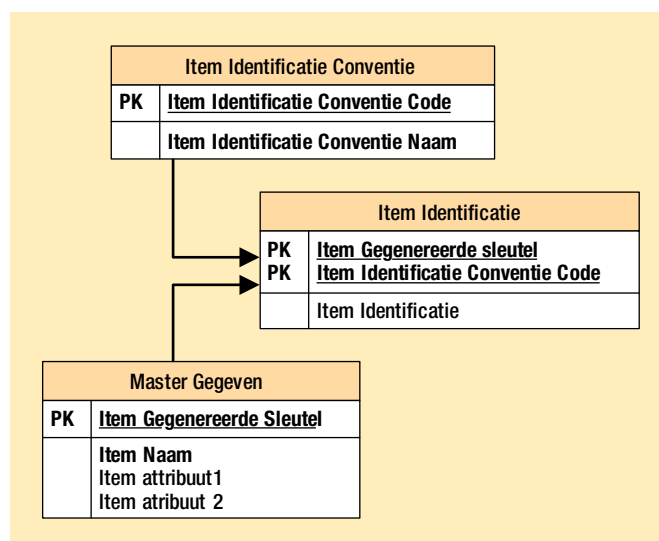
En dan nu ... Metadata

Datawarehousing heeft tot doel de bedrijfshuishouding te beheersen en waar nodig en waar mogelijk te verbeteren. Metadata hebben betrekking op de data-huishouding zelf. Metadata tonen bijvoorbeeld aan dat de load vanuit de bron tot en met de Information Delivery laag technisch goed en logisch betrouwbaar en controleerbaar is. Metadata tonen, volgen en beheren *het gebruik* van de beschikbare data en informatie. Als metadata bijdragen aan de verbetering van het bedrijfsproces (en dat doen metadata natuurlijk wel) gebeurt dat indirect door de data-huishouding te verbeteren. De term 'metadata' is overigens niet voorbehouden aan datawarehousing. Ook bij document management of workflow management worden metadata geproduceerd die bijvoorbeeld het documentvergaar- en distributieproces beschrijven en besturen.

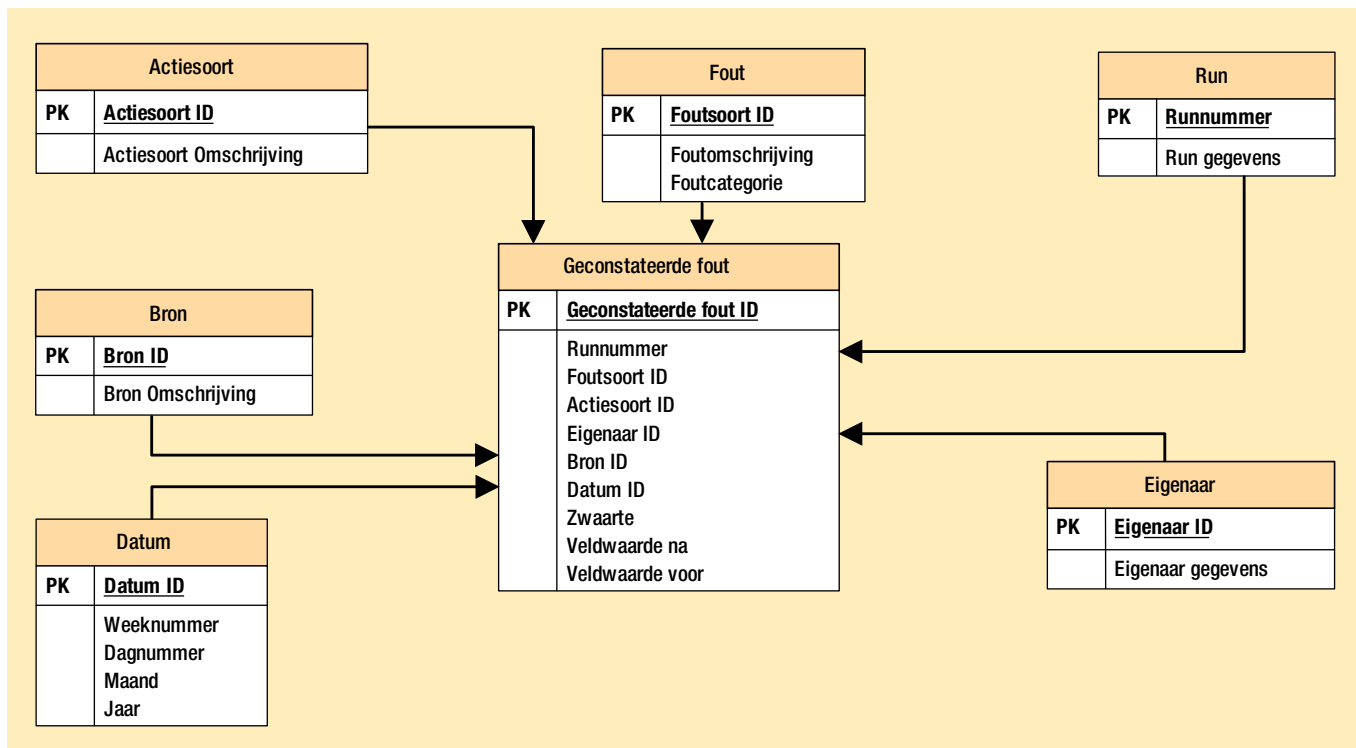
In de beschrijving van Rochade (spreek uit: Rosjeet), marktleider op het gebied van metadata management, vullen de metadata uit het datawarehouse slechts een deel van de metadata repository. Rochade noemt zich een oplossing voor Enterprise Metadatamanagement.

Sarbanes-Oxley, Basel II, IFRS

Omdat de regels met betrekking tot compliance en de eisen aan de betrouwbaarheid steeds strenger worden, is het begrip



Afbeelding 1: De relatie tussen coderingsconventies en masterdata volgens Chisholm.



Afbeelding 2: Stermodel uit een metadatawarehouse.

metadata de laatste jaren belangrijk geworden. In extremo moet elk cijfer in een datawarehouse-rapportage *bewezen* herleidbaar en reconstrueerbaar zijn uit de bronsysteemelementen. Die samenstelling bestaat uit de business rules die in het datawarehouse-proces zijn vastgelegd en beschreven. Metadata zijn als het ware de katalysator van het BI-bedrijfsproces. Soms is het zelfs de startmotor om de grote motor op gang te krijgen. Een praktijkvoorbeeld: bij de opzet van een BI-applicatie kan het zijn dat de betrouwbaarheid te laag is om uit de inhoud al conclusies te trekken. De betrouwbaarheid van de inhoud moet eerst een flinke *boost* krijgen. In zo'n geval moet u de metadata niet verstoppen in audit-attributen in de datawarehouse-tabellen maar moet u in het vergaar-, verwerk-, en laadproces de metadata verzamelen en apart en historisch juist opslaan in tabellen. U kunt dan de metadata ad hoc bevragen en er structureel over rapporteren. Dan heeft u dus een *metadatawarehouse*.

Afbeelding 2 is een voorbeeld van een sterdiagram uit een metadatawarehouse. Met behulp van een dergelijke ster wordt het mogelijk te rapporteren op de zwaarte van de wijzigingen, het verloop in de tijd van de correcties, naar soort, naar bron, naar eigenaar. Natuurlijk zullen zulke rapportages uiteindelijk bijdragen aan de betrouwbaarheid van de inhoud door dat gericht gekeken wordt naar de oorzaken van de fouten (en de veroorzaker ervan). Een dergelijke constructie maakt de fouten en correcties meetbaar en bewijsbaar. De verantwoordelijkheid voor de inhoudelijke verbeteringen ligt niet bij het DWH team maar bij de gegevenseigenaars. De BI-rapportages zijn alleen de boodschappers.

Conclusie

Waarom masterdata en metadata in één adem genoemd worden is een raadsel. Naast elkaar getoond en beschreven hebben ze zichtbaar weinig gemeen. De enige overeenkomst is hun wezenlijk belang voor de data-huishouding. Ze dragen op een eigen maar zeer verschillende manier bij aan goede Business Intelligence. Masterdata management zorgt voor goede basisregistraties van belangrijke dimensies. Metadata management regelt, beschrijft en beheerst het datawarehouse-proces. Als de omzet van een klant wordt getoond in een datawarehouse-rapport, zorgt masterdata management ervoor dat we het over één en dezelfde klant hebben en dat de omzet van die klant bijvoorbeeld niet deels ook onder een ander klantrecord moet worden gezocht. Metadata management zorgt er naast veel andere taken voor, dat de omzet niet op twee manieren kan worden uitgelegd en dat die logische definitie aantoonbaar één-op-één loopt met de samenstelling uit de bronsystemen. Zonder masterdata en metadata dus geen goed functionerende BI-omgeving.

Noten

1. Malcolm Chisholm, *DB/M 4/2006, Eindelijk grote belangstelling voor Master Data.*
2. *Whitepaper Knightsbridge, The Seven Deadly Sins of Master Data Management.*
3. *Whitepaper DataFlux, Semantics, Metadata and Identifying Master Data.*

drs. C. Verhagen is werkzaam als senior consultant BI bij Getronics PinkRocade.