

Definieer de informatie

Enterprise Information Management (3)

Mike Ferguson en Jan Henderyckx

In deel twee van deze serie over Enterprise Information Management (EIM), hebben we u laten kennismaken met de basiscomponenten van het EIM-platform en introduceerden we de Continuous Improvement Methodology. Hoe dan ook; informatie kan niet lukraak worden aangeboden. Van belang is te weten wat de informatie precies inhoudt en wat de relatie ervan is tot andere data.

Elke poging om EIM afdoend in te voeren, dient dan ook uit te gaan van een correcte definiëring van de relatie tussen de bron-gegevens en het Enterprise Data Model. Een correcte definiëring van deze relatie en de bepaling van het datamodel stoelt altijd op de eerste twee stappen van de Continuous Improvement Methodology, de CIM. In dit artikel geven we u dan ook graag meer inzicht in de eerste twee fases van de CIM; de Define-stap en de Identify-stap, zie afbeelding 1.

Weinig bedrijven hebben een informatiearchitectuur die naadloos aansluit bij de behoeften van de informatiegebruikers. Voortdurend evolueren is de boodschap. Als gevolg van tal van externe factoren, zoals de introductie van nieuwe pakketten, overnames en fusies, nieuwe versies van services etcetera, is informatiearchitectuur bovendien absoluut geen statisch

Enterprise Information Management (3)

In een serie artikelen introduceren Mike Ferguson en Jan Henderyckx Enterprise Information Management (EIM) en onderzoeken zij wat er gedaan moet worden om EIM te implementeren, teneinde controle over de data in de onderneming te krijgen. Het is hun bedoeling om in de komende maanden diepgaand te onderzoeken welke technologieën, methodieken en best practices er omtrent EIM voorhanden zijn. Op deze manier willen ze tot een mechanisme komen dat zowel gestructureerde als ongestructureerde data aankan en zo een bijdrage levert aan de verbetering van de algehele business performance.

gegevens, maar veeleer een cyclisch proces dat tot doel heeft de architectuur te verbeteren en het niveau van data governance te verhogen.

De eerste twee stappen in deze cyclus hebben tot doel het informatielandschap in kaart te brengen, en dit informatiemodel op te slaan in de vorm van metadata. Een informatiemodel bouwen gaat verder dan de aanmaak van een metadata repository die louter kan worden geraadpleegd. Het scenario waarin bedrijven aanstuurden op een bedrijfsbreed informatiemodel is inmiddels volledig achterhaald. Dergelijke projecten barstten veelal uit hun voegen en, erger, toonden een groot gebrek aan inzetbaarheid van de gegevens. Een datamodel zonder verdere toepasbaarheid of louter als documentatie heeft weinig reden van bestaan, zodat dergelijke projecten gedoemd zijn tot mislukken.

Aanbieders van modelleer- tools bieden steeds vaker ook de datamodellen aan

Het wezenlijke verschil met een bedrijfsbreed datamodel is de *actionability* van de oplossing. De informatie vervat in het model biedt meer dan alleen maar raadpleegbaarheid, maar maakt echt deel uit van de EIM-platform runtime-oplossing zodra de mapping met de fysieke objecten een feit is.

Metadata

Hoe belangrijk ook, de Define- en Identify-stappen zijn niet meer dan de onderliggende basis voor het opbouwen van de metadata. Metadata behelzen de Shared Business Vocabulary, de Data Modellen waaronder het Operationele Datamodel en het Multidimensionele Datamodel, de Governance regels en de Data Quality regels vallen. Al deze metadata vormen een bedrijfsplatform waarop alle data-actoren, zoals de data steward, de business modeler, de informatie-architect en de BI-specialist met elkaar in contact treden. Hoe hoger de volwassenheidsgraad van de onderneming, hoe meer gedefinieerd het datamodel en hoe completer de SBV. Dit houdt in dat alle data-elementen behoorlijk zijn gedocumenteerd en dat de data steward het gebruik en de betekenis van elk element goed in kaart heeft.

Shared Business Vocabulary en datamodel

Business processen moeten ondersteund worden door kwalitatief goede informatie; informatie die bovendien vaak door allerlei business processen wordt uitgewisseld. Deze informatie kan volatiel zijn, maar blijkt in de praktijk toch vaak persistent opgeslagen in een databank. Waar voorheen deze gegevens vaak schuilgingen achter business processen, dienen deze momenteel steeds vaker rechtstreeks aanspreekbaar te zijn. Dit uit zich in nieuwe architecturale trends als Master Data Management en Information as a Service. Rechtstreekse aanspreekbaarheid was tot voor kort alleen in een Business Intelligence omgeving terug te vinden. Daar bleek al snel de behoefte aan eenduidige definitie van de data-elementen te bestaan. Het delen van informatie is hoe dan ook alleen mogelijk als er een eenduidige definitie is van alle data-elementen.

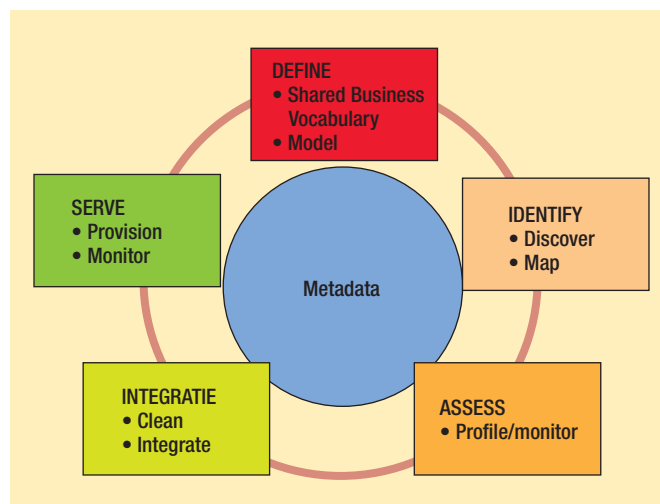
De Shared Business Vocabulary biedt een definiëring van alle data-elementen, maar het ontbreekt daarbij nog aan een beschrijving van de onderliggende verbanden; die worden beschreven in het Datamodel. Dit datamodel mag wel teruggrijpen naar elementen die zijn vastgelegd in de SBV. Van alle metadata fungeert vooral het datamodel als interface tussen de verschillende actoren. Het datamodel is dan ook even cruciaal voor de informatiearchitectuur als de klassen en methodes dit zijn voor de applicatiearchitectuur. Aangezien een datamodel vaak erg complex uitvalt, wordt in de praktijk nogal eens toevlucht genomen tot het opdelen in deelgebieden. In afbeelding 2 is te zien hoe de afdelingen Sales, Distribution, Manufacturing en Marketing elk respectievelijke data-elementen hebben met de bijhorende verwerkingsmethodes. De som van al deze deelgebieden geeft een beschrijving van het product in zijn geheel.

Define

Met de Define-fase leggen we op een eenduidige manier de beschrijvingen van de data-elementen en hun onderlinge verbanden vast. Deze fase bepaalt dus in hoge mate de kwaliteit van de uiteindelijke oplossing. Het is immers zowat de enige fase waarin de interactie met de informatiegebruikers wordt vastgelegd. De andere fasen van de CIM hebben veeleer tot doel de snelheid en de kwaliteit van de informatie te verhogen.

Doel van de Define-fase is om een conceptueel model uit te tekenen

De Define-fase is een gradueel proces, dat niet noodzakelijkerwijs als één groot project dient te worden aangepakt, maar stapsgewijs kan worden uitgevoerd. Verwijzend naar de deelgebieden in het vorige voorbeeld wordt het beste een voldoende kritische massa per deelgebied behandeld. Deze aanpak bevordert in hoge mate de actionability, omdat telkens één deelgebied de volledige cyclus kan doorlopen.



Afbeelding 1: Continuous Improvement Methodology.

In de Define-fase moeten twee zaken worden uitgetekend, namelijk de Shared Business Vocabulary (SBV) en het datamodel. De SBV is inmiddels al uitvoerig aan bod gekomen. Een datamodel kan worden voorgesteld als een UML-model, een Entity Relationship Model of als een multidimensioneel model. Het is zinvol de verschillende modellen tegelijk aan te houden, omdat elk model een specifiek doelpubliek heeft. In ideale omstandigheden zijn de drie verschillende modellen niet onsamenhangend, maar wel aan elkaar gelinkt.

Doel van de Define-fase is om een conceptueel model uit te tekenen. Daarbovenop heeft de Database Administrator nog behoefte aan andere parameters die niet in dit conceptuele model zijn vervat; we spreken dan over het fysieke model. Essentieel is om een link te voorzien tussen conceptueel model en fysiek model. Zo kan de DBA teruggrijpen naar de definities en omschrijvingen van het conceptuele model en hoeft hij niet steeds van voor af aan te beginnen. De latere Identify-fase heeft precies tot doel deze link vast te stellen.

Een datamodel kan op verschillende manieren worden gerealiseerd: via aankoop, door zelf te bouwen of door handig gebruik te maken van één van de beschikbare industriemodellen. De keuze om het datamodel zelf te bouwen laat verschillende werkwijzen toe. Neem nu de fusie van het getrouwheidsprogramma van KLM en Air France. In een eerste werkwijze, die uitgaat van de Define-fase, gaan de business specialisten van beide fusiepartners in gezamenlijk overleg het nieuwe ideale model voor een getrouwheidsprogramma uittekenen. In de daaropvolgende Identify-fase wordt nagegaan hoe dit ideale model zich verhoudt tot beide bestaande systemen. Toegevoegde waarde van deze werkwijze is dat het model toekomstgericht is, daarbij niet gehinderd door historische factoren. In de praktijk blijkt evenwel dat de oude systemen vaak nog een tijdje worden aangehouden, zodat moet worden voorzien in mapping naar het nieuwe systeem, uiteraard met verlies aan flexibiliteit.

In de regel wordt dan ook vaker voor een omgekeerde methodologie gekozen, één waarbij wordt uitgegaan van de Identify-fase. In het voorbeeld van het getrouwheidsprogramma worden zodoende de modellen van KLM en Air France naast elkaar gelegd. De vergelijking leert welk model het best beantwoordt aan de nieuwe behoeften. In de daaropvolgende Define-fase wordt het model dan verder aangepast tot het gewenste resultaat ontstaat. Hoewel een zelfgebouwde en dus per definitie op maat gesneden oplossing erg doelmatig blijkt, is het een tijdrovend proces.

Een datamodel louter als documentatie heeft weinig reden van bestaan

En aangezien tijd cruciaal is wordt steeds meer gebruik gemaakt van kant en klare oplossingen, vrij op de markt te koop – een algemene trend overigens in de software-markt. Allerlei aanbieder spelen gretig in op deze trend. Aanbieder van modelleer-tools bieden steeds vaker niet alleen de software aan, maar ook de datamodellen. Andere leveranciers laten zich kennen als specialisten in bepaalde deelgebieden, bijvoorbeeld per product, per klant, en verkopen niet alleen het datamodel, maar ook

de bijhorende bedrijfsprocessen. Het gaat hier om echte Master Data Management oplossingen.

Als alle data op één fysieke locatie zouden zijn opgeslagen, zou het werk er nu grotendeels opzitten. In de bedrijfsrealiteit zijn data echter opgeslagen in tal van platformen en in niet noodzakelijk compatibele vormen. Bij bedrijven die nog een hele weg hebben af te leggen op de volwassenheidsschaal is er soms zelfs geen enkel inzicht in welke gegevens zich precies waar bevinden. De hieropvolgende Identify-fase voorziet dan ook in de brug tussen de ideale conceptuele wereld van de Define-fase en de meer chaotische bedrijfsrealiteit.

Identify: discover en mapping

Binnen de Identify-fase onderscheiden we enerzijds de Discover-taak en anderzijds de Map-taak. Discover en Mapping samen moeten de fysieke objecten koppelen aan het uitgetekende datamodel en aan de definities uit de SBV. In een efficiënt EIM-platform kunnen de gegevens via deze koppelingen effectief in de nieuwe vorm ter beschikking worden gesteld. Men moet niet vergeten dat het niet de bedoeling is om alleen documentatie aan te maken, maar actionability te verzekeren door de koppelingsinformatie te laten doorstromen naar de latere Integrate-fase.

In de Discover-taak kan worden nagegaan of er al dan niet metadata voorhanden zijn. Zijn deze aanwezig, dan wordt nagegaan of ze in het vereiste platform zitten. Zo niet, dan dient



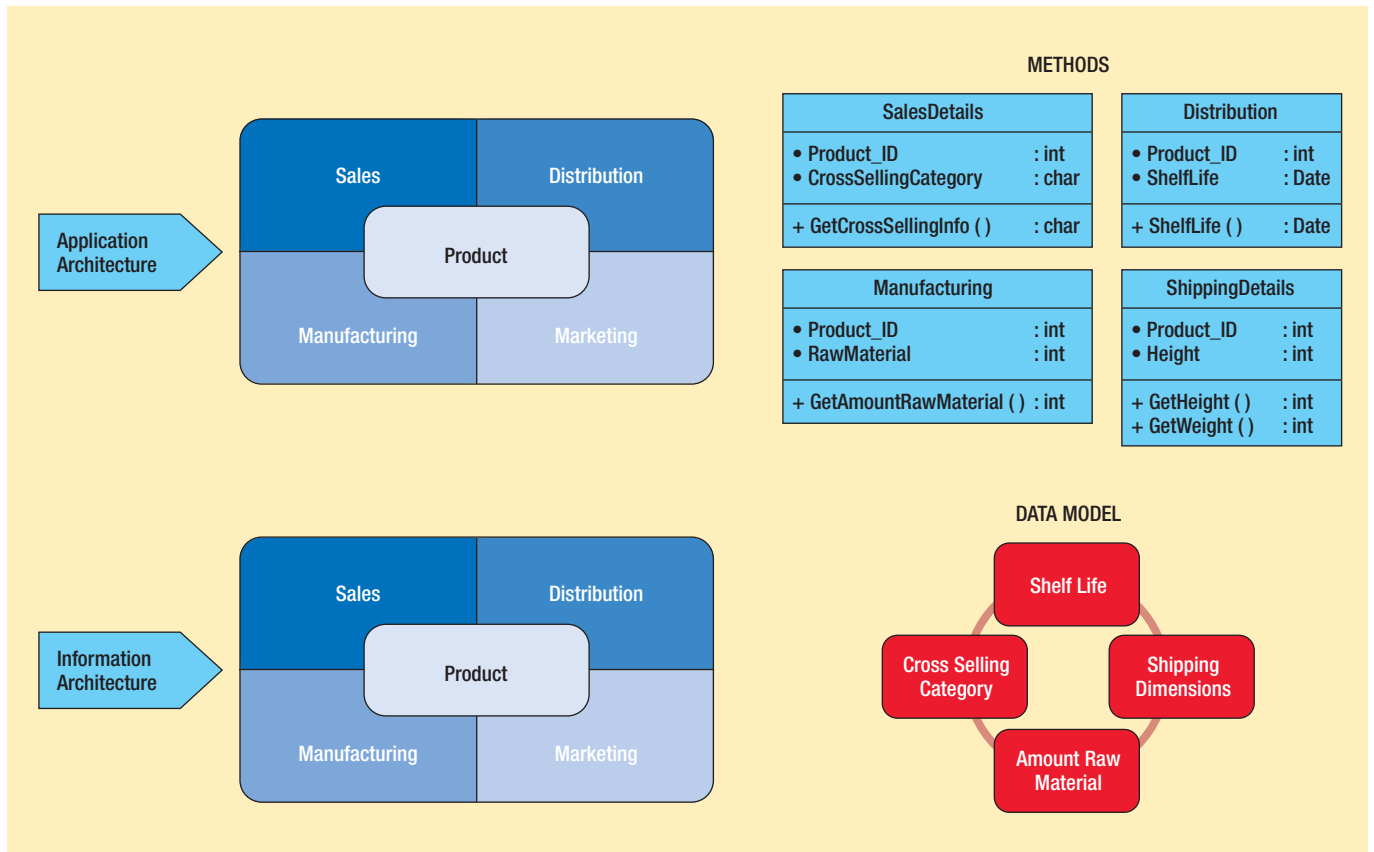
Stilstand is achteruitgang.
Hoe staat het met *jouw* carrière?

Join FourPoints !!

Dé specialist in Data Warehousing & Business Intelligence

DATA WAREHOUSING | BUSINESS INTELLIGENCE | ARCHITECTUUR

Interesse? Neem contact op met Steven Tjin | Stjin@FourPoints.nl | T +31 (0)20 - 452 75 05 | www.FourPoints.nl



Afbeelding 2: Datamodel bestaat uit data-elementen en verwerkingsmethodes.

een brug te worden gelegd tussen het bestaande platform en het vereiste platform. Vaak wordt hiervoor gebruik gemaakt van OMG's Common Warehouse Metamodel (CWM), binnen de metadata-wereld algemeen aanvaard als de standaard. Daarnaast is ook Meta Integration Technology Inc. (MITI) beschikbaar met heel wat connectoren, waarop handig ingespeeld wordt door tal van bedrijven. Blijkt uit de analyse dat er geen metadata voorhanden zijn, dan schieten de mogelijkheden van de modelleertools vaak te kort en moet worden gebruik gemaakt van gespecialiseerde software, zoals DataFlux dfPower Explorer, IBM Rational Data Architect, Informatica PowerCenter and Data Explorer of Sypherlink Harvester.

Wanneer alle fysieke objecten degelijk in kaart zijn gebracht, kan de Discover-fase worden afgesloten om over te gaan tot de mapping: het koppelen van de fysieke objecten aan de elementen uit het datamodel.

Afhankelijk van de maturiteitsgraad van het bedrijf in kwestie zal de complexiteit van de mapping sterk uiteenlopen. In een omgeving waar zeer sterke naamgevingsregels gelden, kan vrij makkelijk op naam worden gemapt. Ontbreekt het aan een dergelijke naamgeving, dan moet mapping worden gerealiseerd via synoniemen of glossaria. Voor de mapping in kwestie moet dan een synoniemenlijst worden aangelegd. Er zijn diverse tools beschikbaar om de daadwerkelijke inhoud van de gegevens te controleren op correctheid van relaties. Soms, als één-op-één

mapping uitgesloten blijkt, dienen zelfs transformatieregels te worden toegepast.

Een bijkomende bron van informatie over mapping kan worden gevonden in de ETL-programma's of in de EAI-softwarestack. Deze geven vaak een goed beeld van de toe te passen regels om data-elementen aan elkaar te koppelen. Hoe mooi het softwareaanbod ook lijkt, het is toch vereist om de kwaliteit van de resultaten daadwerkelijk te toetsen, aangezien deze toch sterk afhankelijk zijn van de complexiteit van de bedrijfsgebonden realiteit.

Conclusie

Zowel de Define-fase als de Identify-fase zijn doorslaggevend voor het welslagen van elk EIM-project, aangezien ze de interface vormen tot de aangeboden gegevens. De grondslag is gelegd voor het verder bouwen aan een coherent EIM-platform.

Mike Ferguson en Jan Henderyckx

Mike Ferguson is Managing Director van Intelligent Business Strategies Limited. Als analist en consultant is hij gespecialiseerd in Business Intelligence en enterprise business integration.

Jan Henderyckx is onafhankelijk consultant, spreker en auteur op het gebied van informatie-architectuur en databases.

Het copyright van de in deze serie beschreven methode voor Enterprise Information Management ligt bij de auteurs.