

Continue verbeteringsmethodiek geeft betere beheersing

Enterprise Information Management (2)

Mike Ferguson en Jan Henderyckx

In het vorige artikel van de serie over Enterprise Information Management (EIM) is dieper ingegaan op de risico's die een onderneming loopt als ze niet beschikt over een gestructureerde aanpak van informatie management. In dit vervolg wordt het belang van data governance en datakwaliteit voor de EIM-oplossing belicht.

Een organisatie is steeds in beweging, waardoor governance en kwaliteitsbewaking een continu proces zijn, meer dan een einddoel. Via continue verbeteringsmethodiek kan een organisatie op een efficiënte manier evolueren naar een betere beheersing van de EIM-problematiek.

Enterprise Data Governance Strategie

Hoewel specifieke ad hoc strategieën op het gebied van datamanagement en datakwaliteit zeker ook resultaat kunnen opleveren, dringt zich toch een meer grootschalige, ondernemingsbrede aanpak op. En hoewel IT-professionals zich langzaam bewust worden van het belang van een dergelijke ondernemingsbrede visie, blijken de beleidslijnen op het vlak van data ownership en enterprise datamanagement in de praktijk vaak erg onsamenhangend. Evenzeer doorslaggevend is het uitwerken van een continue strategie en verbeteringsmethodiek op het vlak van enterprise data governance.

Een doelmatige data governance veronderstelt meer bepaald:

- het vaststellen van bedrijfsbrede datastandaarden aan de hand van een *Shared Business Vocabulary* voor databenoeming, datadefinities en data integriteit regels;

- het definiëren van een ondernemingsbreed datamodel aan de hand van Shared Business Vocabulary definities;
- het stelselmatig aanwenden van Shared Business Vocabulary databenoeming en datadefinities op alle niveaus in de onderneming;
- het uitwerken van gemeenschappelijke beleidslijnen, patronen en processen met het oog op enterprise datamanagement en voor de ontwikkeling van data-integratie;
- het opzetten van een ondernemingsbreed datakwaliteitsprogramma;
- het implementeren van Master Data Management voor alle core business activiteiten van de onderneming.

Datastandaarden

Om datastandaarden te kunnen bewerkstelligen, moet eerst en vooral een Shared Business Vocabulary (SBV) voorhanden zijn. Alle master- en transactiedata die op de verschillende niveaus in de onderneming worden gebruikt, dienen te worden aangeduid en gedefinieerd en moeten zo leiden tot een gemeenschappelijke set van ondernemingsbrede standaard datadefinities.

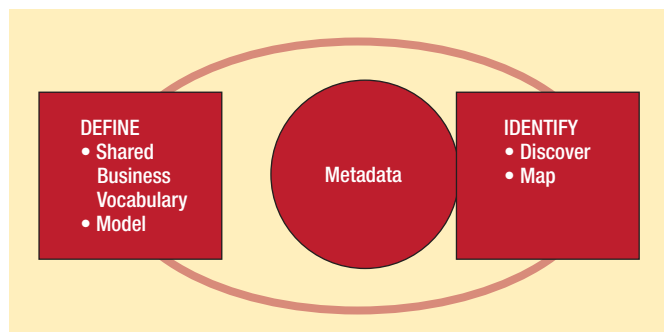
Ongelijksoortige datadefinities moeten vervolgens naar deze gemeenschappelijke definities worden omgezet. In een eerste fase zullen alle anomalieën enkel worden in kaart gebracht.

Heel concreet houdt het opzetten van een SBV dus in dat stelselmatig almaar meer ondernemingsbrede *gemeenschappelijke* definities bepaald worden voor: datadefinities, data integriteit regels, referentiedata (namelijk code set-waarden), mappings en transformaties voor alle masterdata, transactiedata. De SBV vormt als het ware de hoeksteen voor een bedrijfsbrede datastandaard en is van doorslaggevend belang voor makkelijk te begrijpen data door de gehele onderneming heen, voor masterdata management en voor data-integratie.

Dergelijke gemeenschappelijke datadefinities kunnen worden gebruikt in:

- Datamodellen om aldus consistentie te verkrijgen door meerdere modellen heen;
- Data-integratie-tools (ETL en EII) voor target-systemen;
- Applicatie-integratietechnologieën (message brokers en ESB's);
- Zakelijke rapporterings-tools;
- XML mark-up tags.

Door data (bijvoorbeeld in XML-formaat) aan te reiken gebruikmakende van XML tags gebaseerd op SBV databenoeming,



Afbeelding 1: Stappen op weg naar een Shared Business Vocabulary.

kunnen de gebruikers erg snel het systeem van de databenoeming doorgronden. Als de data dan ook nog eens op precies dezelfde manier beschikbaar worden gesteld voor gebruik door toepassingen in alle niveaus van de onderneming, is er sprake van een consistent, ondubbelzinnig beleid van datamanagement bij de voorstelling van de gegevens. De SBV vormt dus absoluut de hoeksteen in een efficiënt systeem van informatie management. Aan de hand van deze set van gemeenschappelijke metadata kunnen naderhand datamodellen, XML-schema's, cubes, business verslagen en dergelijke worden gecreëerd, met een hogere graad aan consistentie. Als dit allemaal is uitgewerkt, is het makkelijker om ongelijksoortige datadefinities en datarelaties in bestaande systemen te gaan vaststellen en deze om te zetten naar de SBV van de onderneming (zie de Identify/Identificeer stap in afbeelding 1).

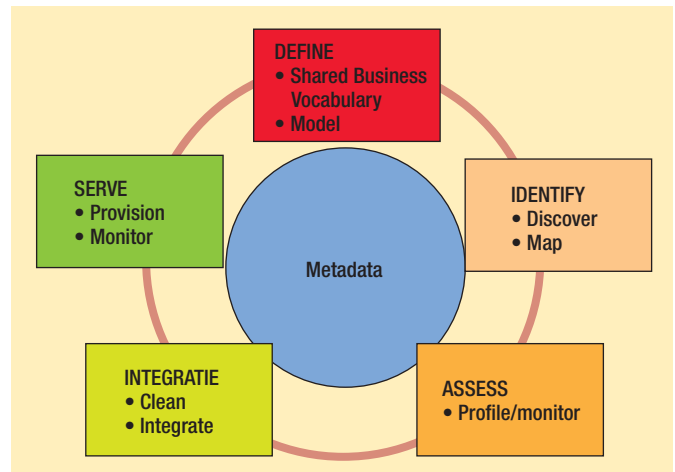
Toegang vragen tot lage latentie data kan via een query interface of services interface

Op die manier definiëren we de relaties tussen ongelijksoortige definities in één enkele, ondernemingsbrede shared business vocabulary: een belangrijke mijlpaal in het verwerven van controle over de bedrijfsdata.

Continue verbeteringsmethodiek

Veel organisaties die zich in de beginfase van de EIM maturiteitsmatrix bevinden, hebben het moeilijk om voldoende inhoud te creëren om nuttig aan de slag te gaan. De Shared Business Vocabulary is nog geen feit of bevindt zich in een pril stadium (de Define/Definieer stap in afbeelding 1) en de datamodellen maken niet altijd gebruik van deze definities. Hoe dan ook blijft het opzetten van een Shared Business Vocabulary met voldoende kritische massa een absolute vereiste. Als een voldoende basis is opgebouwd, kan dit worden uitgebreid in een steeds hoger tempo.

Het is vaak belangrijk om bij het opzetten van de SBV te beginnen bij de aanduiding van de core master data entities en attributen. Alle nieuwe datamodellen kunnen vervolgens worden gecreëerd aan de hand van deze gemeenschappelijke data-definities om zo geleidelijk aan te evolueren naar een gemeenschappelijke data vocabulary op alle niveaus. Datamodellen creëren vanuit SBV-definities maakt deel uit van de Define stap in afbeelding 1. In een volgende fase worden de gelijksoortige maar anders benoemde data in de bestaande systemen aangegeleid en worden deze omgezet naar Shared Business Vocabulary definities. Deze relaties en definities worden opgeslagen in een gemeenschappelijk metadata-platform, waardoor de zonet gedefinieerde relaties bruikbaar worden in de verdere stappen.



Afbeelding 2: Continuous Improvement Methodology.

Tijd nu om over te gaan tot de beoordelingsfase (zie de Assess/Beoordeel stap in afbeelding 2), namelijk hoe is het gesteld met de kwaliteit van de gelijksoortige databronnen. De bestaande data worden bestudeerd om te bepalen of de waarden wel degelijk zijn wat ze zouden moeten zijn. Deze stap verzuimen en de informatie zonder meer afleveren is niet zonder gevolgen, aangezien verondersteld wordt dat de inhoud van de data in lijn is met de datadefinities en de integriteitbeperkingen van de Shared Business Vocabulary en het datamodel. Data waarvan de kwaliteit te wensen overlaat, moeten worden gezuiverd vooraleer over te gaan tot datatransformatie en -integratie (de Integreer/Integreer stap in afbeelding 2). De actuele informatie wordt op die manier afgestemd op de nieuw gemodelleerde versie. Nu pas zijn de data klaar voor bedrijfsgebruik in allerlei toepassingen, processen en portals.

Informatieservices die draaien op een EIM-platform dienen uiteindelijk het volledige spectrum te bestrijken van information delivery requirements die business processen nodig kunnen hebben. Dit betekent dat de informatiediensten zowel data-

Enterprise Information Management (2)

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.

integratie, datazuivering en zelfs toegang tot metadata moeten kunnen bieden en dit in diverse formaten (zie de Serve/Bied aan stap in afbeelding 2), zoals XML, HTML, Office Excel bestandsformaat, e-mail bijlage, of rechtstreeks geleverd in een database. Alle bovenvermelde stappen kunnen aan elkaar worden geschakeld in een continue verbeteringsmethodiek, zoals te zien in afbeelding 2.

Twee factoren

Met het oog op het beschikbaar stellen van de data (in de Serve/Bied aan stap) hebben twee factoren een belangrijke invloed op de keuze van beste EIM-oplossing, met name de informatie-latentie en het vereiste integratieniveau.

Datalatentie is een makkelijk te begrijpen concept: hoe actueel moeten de data zijn voor het bedrijfsproces dat de data gebruikt.

De validering van alle data gebeurt in de meeste ondernemingen nog te fragmentarisch

Een verslag met de top-tien van best verkopende boeken op een website hoeft niet per se de allerlaatste informatie te bevatten. Een bedrijfsprocedure daarentegen die de rekeningstand controleert vooraleer een bedrag over te maken, heeft er alle belang bij toegang te krijgen tot real-time data, de zogenaamd zero latentie (nul vertraging) data. In andere gevallen is dan weer de stabiliteit van de data geprefereerd boven een lage latentie.

Een mooi voorbeeld hiervan is een subject oriented datamart, gebruikt door bedrijfsanalisten, om een subset van data te kunnen begrijpen. Een proces dat een dergelijke hoge mate van datastabiliteit vereist, kan beter werken met een (omgevormde) kopie van de operationele gegevens. Processen die evenwel een lage latentie of zelfs real-time toegang tot data vereisen hebben behoefte aan informatiediensten die de gevraagde gegevens rechtstreeks halen uit het operationele systeem.

Toegang vragen tot lage latentie data kan via een query interface of services interface. Een directe query interface is een voorbeeld van een hechte koppeling. Er wordt toegang verkregen tot de informatiebron via een XQUERY of SQL query. Als een dergelijke hechte koppeling niet toegelaten is of als het schema van de onderliggende informatieverstrekker te complex uitvalt, is een lossere koppeling te verkiezen in de vorm van een ESB-achtige benadering. Deze benadering wordt soms *Information as a Service* genoemd. Op die manier wordt toegang verkregen tot de laag latente data en kunnen de gevraagde data in XML-formaat worden afgeleverd.

We stellen vast dat de drie belangrijke opties, te weten ESB, EII

en ETL, elk een deel van het spectrum bestrijken in functie van de latentie en integratieniveaus (zie afbeelding 3). Mits met een combinatie van de drie opties kan worden ingespeeld op elke mogelijke informatiebehoefte.

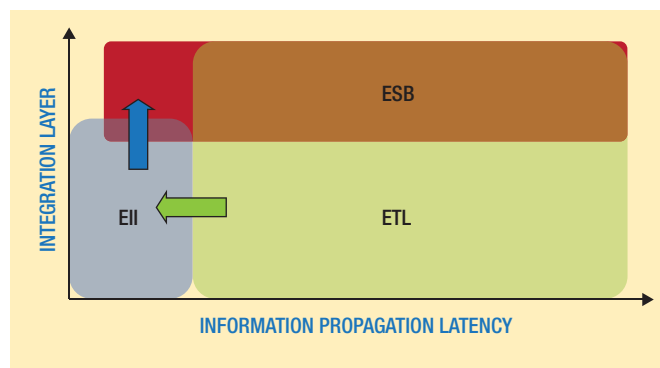
Merk op hoe ETL steeds meer evolueert in de richting van lage latentie. Met de toenemende vraag van de bedrijven naar geïntegreerde lage latentie data in hun datawarehouses en datamarts, stijgt ook het doelgerichte gebruik van lage latentie ETL-tools. De ETL-functie wordt daarbij meestal aangesproken als een service. Ook EII kan steeds meer abstraheren door de output van webservices te gebruiken als informatiebron. Op die manier is een hoger abstractieniveau mogelijk dan met louter rechtstreekse ondervragingen via SQL of XQUERY. De EII-technologie is ook toegankelijk via een webservices interface om gefedereerde query's die zijn opgeslagen op een EII server ter beschikking te stellen op de Enterprise Service Bus.

EIM Data Governance Patronen

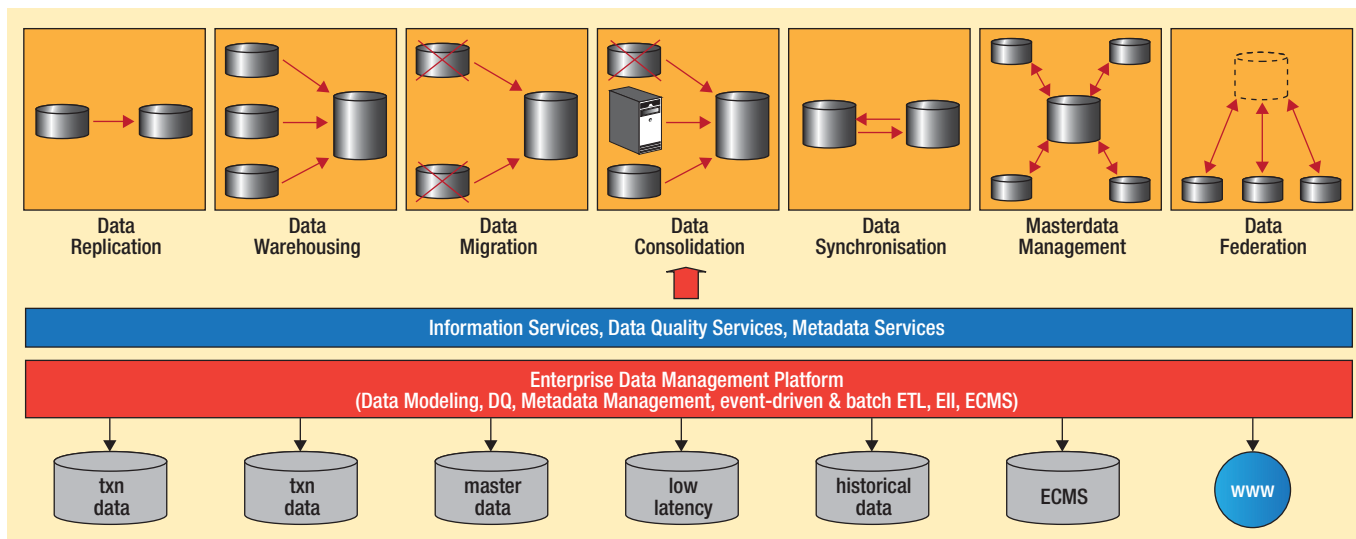
Een enterprise data governance strategie is sterk gebaat met vuistregels voor een correcte benadering van datamanagement. Deze kunnen we best uitdrukken als een reeks van toepasbare patronen.

Afbeelding 4 toont een aantal vaak voorkomende data-management-taken die de meest bedrijven dienen uit te voeren. Zo onder meer: Datareplicatie; Datawarehousing; Datamigratie; Dataconsolidatie; Data Synchronisatie; Masterdata Management; Datafederatie.

Uiteindelijk dient elke enterprise data governance strategie standaarden te ontwikkelen voor deze vaak voorkomende datamanagement-patronen. Ook het maximale hergebruik van diensten voor datazuivering en data-integratie, nodig bij replicatie, synchronisatie, consolidatie, federatie en datamigratie, staat daarbij voorop. Hiertoe dienen informatiediensten (workflows) gecreëerd te worden voor elke masterdata-entiteit en elk type transactiedata gebruikt in deze patronen. Er moeten onder meer informatiediensten worden ontwikkeld voor datazuivering, transformatie en data-integratie van productgegevens, bestelingsgegevens, en dergelijke. Wanneer deze diensten ontwikkeld zijn kunnen zij ook worden gebruikt in elk van de bovenvermelde patronen. Op die manier zijn er gemeenschappelijke



Afbeelding 3: Integratie en Latency.



Afbeelding 4: Veelgebruikte datamanagement-taken.

procedures voorhanden om gebruikersdata te migreren, te synchroniseren, te federeren en te consolideren. Waarom proberen het wiel uit te vinden?

Enterprise Data Quality

Om betrouwbare data te creëren in een data governance strategie is het zaak om een enterprise data quality 'firewall' te voorzien voor overal waar gedeelde data quality 'services' (met gemeenschappelijke gedeelde regels) worden gebruikt door toepassingen en infrastructuur. Dit is de beste manier om te voorkomen dat slechte gegevens worden ingevoerd en zich verder verspreiden door de onderneming heen. Probleem daarbij is dat in de huidige context data de onderneming op allerhande niveaus binnenkomen. Data worden daarnaast ook intern aangemaakt door programma's als deel van een bedrijfsproces. De validering van al deze data gebeurt momenteel in de meeste ondernemingen nog te fragmentarisch, door middel van erg uiteenlopende valideringsmethodes die her en der in de loop der tijd zijn ontwikkeld. Conflicten wegens verschillende valideringsmethodes met tegenstrijdige regels, 'out of date' validering of zelfs helemaal geen validering, zijn daarbij niet uit te sluiten. Uiteraard doen dergelijke tegenstrijdige procedures afbreuk aan de kwaliteitscontrole, met defecten, operationele fouten, incorrecte beslissingen en dergelijke tot gevolg. Geen beste zaak voor de algehele business performance.

Door te opteren voor een bedrijfsbrede enterprise data quality firewall kunnen al deze fragmentarische methodes onder het hoedje van één gemeenschappelijke aanpak worden gevangen, waarbij gemeenschappelijke data quality services (regels) op een standaard manier voor de bedrijfstak via webservices kunnen worden ingeroepen. De data quality regels kunnen vanuit één centrale plaats worden opgeroepen, on-demand, in batch, als deel van een bedrijfsproces of naar aanleiding van een specifiek gebeuren en op die manier data valideren en zuiveren, waarbij

het niet uitmaakt hoe de data de onderneming zijn binnengekomen. Met een data quality firewall kan het beheer van de data quality in het bedrijf worden toevertrouwd aan technologie en kan er ontwikkeld worden met een bedrijfsbrede standaard voor ogen.

Master Data Management (MDM)

Aangezien masterdata meestal in alle bedrijfsentiteiten worden gebruikt en *herhaaldelijk* worden gebruikt in tal van bedrijfsprocessen is het in data governance van het grootste belang om precies deze masterdata onder controle te krijgen. Essentieel daarbij is dat wordt ingezien dat MDM draait rond het opzetten van een ondernemingsbrede set van beleidslijnen, diensten, processen en technologieën, om de data die te maken hebben met de echte core business van het bedrijf te creëren, te onderhouden en te beheren als een *system of record* (SOR). Tot deze kernactiviteiten behoren de Klanten, Leveranciers, Medewerkers, Activa, enzovoort. Aangezien MDM zo beslissend is, komen we er graag uitgebreid op terug in een volgend artikel.

In de eerste twee artikelen van de reeks rond EIM hebben we het kader geschetst van waaruit een enterprise informatie-architectuur zou moeten bestaan. De volgende artikelen vertrekken opnieuw vanuit de continue verbeteringsmethodiek en zullen dieper ingaan op de problemen en mogelijke oplossingen die we in elke fase van de methodiek tegenkomen.

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.