

Information Supply Chain Management is de volgende stap na EIM

Integreer twee werelden

Rob Arntz en Ad Stam

Informatiewerkers binnen organisaties leven (grotendeels) nog steeds in twee werelden. Enerzijds is er de wereld van de (gestructureerde) harde cijfers en anderzijds die van de 'fuzzy' tekst, video en audio; ongestructureerde informatie. Dit artikel is een pleidooi voor het geven van een veel bredere inhoud aan het begrip EIM.

Over het algemeen wordt onder Enterprise Information Management (EIM) het integreren van gestructureerde en ongestructureerde informatie binnen één applicatie-omgeving verstaan. Op Wikipedia is de volgende definitie van Enterprise Information Management te vinden: "Enterprise Information Management is de benaming voor het vakgebied dat de kennisgebieden Business Intelligence en Enterprise Content Management combineert."

Enterprise Information Management (EIM) gaat nog een stap verder in het organiseren van informatie dan Business Intelligence (Datawarehousing) en Enterprise Content Management al doen. Deze vakgebieden organiseren respectievelijk gestructureerde informatie (data) en ongestructureerde informatie (content), en zijn er op gericht verzuilde informatiebronnen te vervangen door enkelvoudige bronnen, waardoor informatie beter beschikbaar komt. Echter, doordat Business Intelligence en Enterprise Content Management zelf afzonderlijke oplossingen bieden, bestaan er nog altijd twee informatie-silo's, waardoor er nog steeds beperkingen zijn in het beschikbaar stellen van bedrijfsinformatie.

Twee werelden

Informatiewerkers binnen organisaties leven (grotendeels) nog steeds in twee werelden. Enerzijds is er de wereld van de harde cijfers en anderzijds die van de 'fuzzy' tekst, video en audio. Combinatie van deze twee is nog steeds een spaarzaam en ook moeizaam toegepast fenomeen. In de dagelijkse praktijk zijn wij nog maar een zeer beperkt aantal toepassingen tegengekomen. Eén daarvan was een applicatie waarin verkoopcijfers van verschillende cosmetische producten geanalyseerd konden worden en binnen dezelfde applicatie was het dan ook mogelijk reclamespots van een bepaald artikel te bekijken. Dus product y in land x verkoopt in deze maand opeens veel meer dan vorige

maand, en wat is dan de tv commercial nu en die van vorige maand?

Een ander voorbeeld is het verzamelen van allerlei bronnen van verschillende formaten (zowel in file- als database-vorm). Hierbij worden naast de data ook de geïntegreerde metadata gecentraliseerd opgeslagen. Hierdoor zijn de centrale metadata te raadplegen en via deze centrale metadata kunnen vervolgens de onderliggende data (ongeacht het formaat!) opgevraagd worden, en door standaard transformaties toe te passen op basis van de metadata kunnen ze in het gewenste formaat uitgeleverd worden. Zo kan een *fixed width flat file* ingelezen worden, vervolgens worden opgeslagen als tabel (of *comma separated* of in ieder andere gewenst open formaat) en vervolgens uitgeleverd in weer een ander formaat, bijvoorbeeld XML. In feite zijn de opgeslagen bestanden weliswaar ongestructureerd, maar de eraan gekoppelde metadata zorgen ervoor dat de ongestructureerde data zich kunnen gedragen als gestructureerde data. Ook applicaties – vaak aangeboden via een platform – waarin annotaties op of koppeling van documenten met (formele) rapportages kunnen plaatsvinden, soms samen met e-mail notificatie of een vorm van workflow, komen wel voor. Maar zoals Wikipedia ook meldt: een gezamenlijk portal maakt nog geen geïntegreerd gebruik mogelijk. En overigens is het maar de vraag of dat nu wel altijd helemaal zo nodig of nuttig is.

Het nieuwe EIM

Het nieuwe EIM houdt zich – overigens om misverstanden te voorkomen zowel in het domein van de gestructureerde als van de ongestructureerde data – bezig met een veel breder activiteitenscala. Kijken we naar een organisatie dan zien we – op een heel hoog abstractieniveau – twee processen. Allereerst is daar de zogenaamde primaire waardeketen, waarin de operationele processen (ondersteund door een aantal secun-

daire) van een organisatie zich afspelen. Deze primaire waardeketen is een 'horizontaal' proces beginnend met productdefinitie en -ontwerp en eindigend met serviceverlening na het afleveren van het product. De 'cash to cash cycle' is een financieel begrip, waarmee aangegeven wordt hoe lang het kapitaal van een organisatie 'vastzit' in de dagelijkse operatie. Hoe korter deze periode hoe efficiënter en effectiever het proces wordt uitgevoerd.

Daarnaast is er sprake van een 'verticale' planning and control cyclus, zie afbeelding 1. In deze cyclus wordt de organisatie aangestuurd; te beginnen op strategisch niveau; 'wat is het doel van de organisatie en zijn we nog met de goede dingen bezig?' Vervolgens op tactisch niveau waarin als het ware een voortdurende bewaking en optimalisatie van de totale waardeketen plaatsvindt. Op het operationele niveau wordt over individuele functies en (sets van) transacties binnen een proces gerapporteerd en geanalyseerd. Deze planning & control cyclus is het traditionele domein van Business Intelligence.

Deze beide ketens bestaan uit een groot aantal elkaar opvolgende functies. Met de buitenwereld worden goederen of diensten 'geruild' tegen geld. De voortgang en resultaten van dit ruilproces wordt door middel van het overdragen van informatie tussen de elkaar opvolgende functies bewerkstelligd en bewaakt. Het reguleren van al deze informatieoverdracht is in feite te vergelijken met het inrichten en reguleren van een logistiek proces ofwel het inrichten en managen van de 'Information Supply Chain' (ISC). Het proces zou dan aangeduid kunnen worden met Information Supply Chain Management (ISCM). En daarbij komt veel meer kijken dan het plaatsen van een portal bovenop twee afzonderlijke silo's met gestructureerde, respectievelijk ongestructureerde gegevens.

Componenten

De volgende componenten zijn nodig om een efficiënt en effectief verplaatsen, bewerken, bewaren en beheren van informatie in de ISC te kunnen faciliteren.

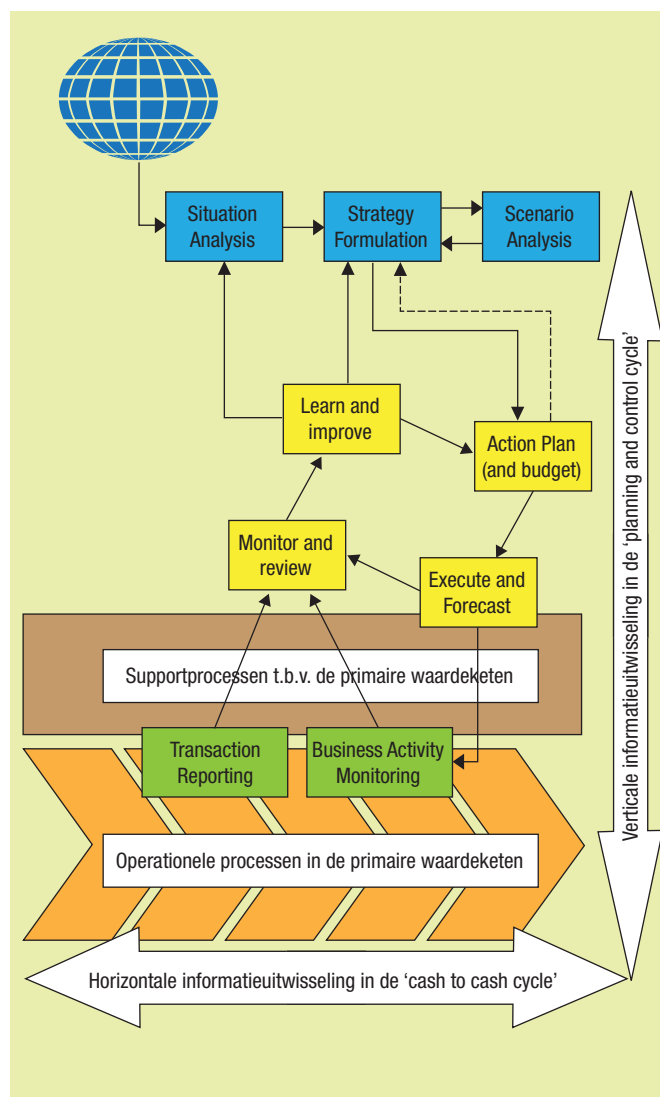
- Verplaatsen: EAI (Enterprise Application Integration); ETL (Extraction, Transformation, Load).
- Bewerken: ETL (Extraction, Transformation, Load); DQM (Data Quality Management).
- Bewaren: Datawarehousing; ECM (Enterprise Content Management).
- Beheren: BRM (Business Rules Management); DQM (Data Quality Management)
- ECM (Enterprise Content Management); MDM (Master Data Management).

Verplaatsen.

Transport is een belangrijk onderdeel van iedere logistieke keten. In een ISC vindt transport van informatie plaats tussen verschillende applicaties. (Transport binnen verschillende functies van één applicatie wordt niet beschouwd als functie van

de ISC). Er zijn twee vormen van transport. In de eerste plaats transport van individuele transacties¹ (een bestelling vanuit het order entry systeem naar de planningsapplicatie van de werkvoorbereiding), in de tweede plaats het transport van groepen van transacties (alle call detail records van gisteren vanuit het billing-systeem naar een datawarehouse).

De eerste vorm van transport vindt plaats in de 'cash to cash cycle', de tweede in de 'planning & control cycle'. Individuele transacties worden vaak verplaatst met EAI-tools en in de cash to cash cycle, batches vaak met ETL-tools en in de planning & control cycle. Maar omdat in toenemende mate het onderscheid tussen de tools aan het verdwijnen is kan het ook heel goed 'andersom'. Een verschil tussen de twee transportvormen is nog wel dat bewerkingen in de cash to cash cycle niet vaak voorkomen en als ze voorkomen dat ze toch vaak redelijk eenvoudig zijn. In de planning & control cycle vinden veel vaker bewerkingen plaats en deze bewerkingen kunnen ook (zeer) complex zijn (vandaar ook de T van Transformation in ETL).



Afbeelding 1: De informatie-uitwisseling in een organisatie.

Bewerken.

De bewerkingfunctie (transformatie) van ETL-tools is hierboven al behandeld. Voorbeelden van bewerkingen zijn het integreren van gegevens uit verschillende bronnen, het toevoegen van berekende gegevens of het 'manipuleren' van gegevens om gegevens 'te laten passen' op het gegevensmodel van de doel-applicatie, of het verrijken en schonen van data aan de hand van bedrijfsregels).

Het is overigens natuurlijk ook mogelijk dit soort bewerkingen uit te voeren met 'gewone' hulpmiddelen. Zo wordt in Oracle-omgevingen nog vaak gebruik gemaakt van PL/SQL als gereedschap om het ETL-proces mee uit te voeren. Aangezien databases zelf steeds krachtiger worden, zien we deze verschuiving vaker. Een nieuwe 'trend' is zelfs EL-T: waarbij EL het daadwerkelijke transport tussen de logistieke ontkoppelpunten laat plaatsvinden door een 'loader' en daarna de transformatie in de database laat plaatsvinden. Als we de logistieke analogie hanteren dan werden vroeger in het distributiecentrum de pallets erwtensoep afkomstig van de fabriek omgezet in dozen erwtensoep voor de winkeliers. Nu zegt de winkelier: doe mij maar gewoon de pallets, ik zet ze direct in mijn schap. Naast E-T-L en ET-L komt ook nog T-EL voor, waar de leverende database al de transformaties doet alvorens de boel wordt omgezet.

Naast E-T-L en ET-L komt ook nog T-EL voor

Hulpmiddelen op het gebied van data quality management zijn in staat om gegevensbestanden (bijvoorbeeld bestanden met heel veel naam/adres/woonplaatsgegevens) te analyseren, te controleren op geldigheid van adressen en op het voorkomen van dubbele gegevens.

Afhankelijk van de geconstateerde fout en de instelling van de tool zullen fouten hersteld of gemeld worden (eventueel voorzien van een advies hoe te handelen). Het uitvoeren van een data-kwaliteitsanalyse kan deel uitmaken van een ETL-proces, maar kan ook heel goed een zelfstandige functie zijn op een bepaald bestand, zonder dat er sprake is van transport naar een andere omgeving.

Bewaren.

Allereerst is er het datawarehouse als ontkoppelpunt in de planning and control cycle. In het datawarehouse worden gegevens uit diverse bronsystemen opgeslagen. We hebben allemaal geleerd dat redundante opslag (en een datawarehouse is redundant) vermeden moet worden. Strikt genomen – en er zijn ook wel mensen (vooral leveranciers van bepaalde BI-tools) die dat propageren – zou dat ook kunnen. Echter er is een aantal redenen waarom een datawarehouse toch de betere oplossing is. Allereerst is het in real-time integreren van gegevens uit een – soms zeer groot – aantal bronsystemen een behoorlijke klus,

zeker als het om data sets gaat en niet om individuele transacties. Daar komt bij dat het informatiemodel van bevraging vaak een heel ander model is dan het datamodel van de bronnen, ook dat maakt het niet gemakkelijker om in real-time opvragingen te doen. Ook is vaak de historische reikwijdte van een bronsysteem te klein voor de behoeftes van BI-applicaties. Zeg nooit nooit, maar datawarehouses zullen naar onze mening waarschijnlijk nog lange tijd deel uitmaken van de ISC. Datzelfde geldt overigens waarschijnlijk ook voor datamarts waarin dan weer een bepaalde subset van gegevens (al dan niet nog weer bewerkt) uit het datawarehouse wordt bewaard (ook weer redundant) voor een specifieke gebruikersgroep.

Daar waar een datawarehouse en een datamart specifieke ontkoppelpunten hebben voor de planning & control cycle, daar zal ECM (Enterprise Content Management) zowel in de cash to cash cycle als in de planning & control cycle toegepast kunnen worden. ECM heeft in zich zelf niet een echte zelfstandige opslagfunctie. Documenten blijven staan waar ze staan en een ECM-omgeving zal daar – anders dan bij een datawarehouse – geen extra kopie van maken. Wat ECM wel doet is – op basis van de aard en inhoud van het document – informatie aan het document toevoegen, waarmee het document – langs een aantal invalshoeken – 'terugvindbaar' wordt. Deze toegevoegde informatie wordt ook opgeslagen in een metadata repository van de ECM-omgeving.

Beheren.

Het beheren van informatie is geen tool of systeem, maar een volwaardig proces. Het proces heeft tot doel de informatie over de informatie zoals deze in de cash to cash cycle en de planning & control cycle wordt gebruikt éénduidig te definiëren en vast te leggen. En zoals ieder proces begint ook dit met het organiseren ervan. Wie is eigenaar van het proces? Wie zijn verantwoordelijk voor het uitvoeren van welke stappen en op welk moment en in welke volgorde? Wie moeten worden geraadpleegd en/of geïnformeerd? Als dit is vastgelegd kan via een workflow management omgeving deze routing worden vastgelegd. Informatie over informatie heeft een drietal aspecten:

- De logische betekenis (semantiek): wat wordt met een bepaald informatie-element bedoeld; wat is de definitie en aan welke regels moet het voldoen; wie is – in het proces waar het informatie-element wordt gebruikt – de eigenaar van het informatie-element;
- De technische betekenis: aan welke technische regels moet het gebruik van het informatie-element in een applicatie voldoen;
- De plaats van gebruik: in welke applicaties wordt het informatie-element gebruikt?

Business Rules Management (BRM) en het eerder besproken DQM, ECM en MDM dekken alle vier een ander deelgebied (zij het met een grote overlap) van het beheer van informatie af. DQM is een omgeving waartegen en waarmee informatie die in

andere systemen is opgeslagen gecontroleerd kan worden op een aantal kwaliteitsaspecten. ECM biedt een beheerfunctie via het hebben en onderhouden van een repository-functie, waarmee ongestructureerde informatie benaderbaar wordt en mogelijk ook ongestructureerde informatie uit andere systemen.

Daarnaast regelt ECM ook de bewaar- en vernietigingstermijnen van ongestructureerde informatie. MDM is het proces waarin de definities van individuele informatie-elementen (masterdata) worden vastgelegd en beheerd. Bij informatie-elementen moet dan gedacht worden aan 'wat is een klant en welke klanten hebben we', maar ook aan 'wat zijn de structuren (dimensies) waarlangs informatie gerapporteerd en geanalyseerd wordt'. BRM betreft in feite een specifieke vorm van masterdata. Het gaat hier niet over individuele informatie-elementen, maar er worden beslissingsregels vastgelegd.

Het belang van ISCM

Het is duidelijk dat ISCM een cruciale rol speelt bij het op een effectieve en efficiënte wijze uitwisselen en verwerken van informatie in zowel de 'horizontale' als de 'verticale' bedrijfsprocessen. Efficiënt, omdat heel veel van de technische en organisatorische inspanning – met alle daaraan verbonden kosten – die nodig is om delen van processen via applicaties met elkaar te laten praten, nu besteed wordt aan het oplossen van de Babylonische spraakverwarring over informatie. Effectief, omdat doorlooptijden van processen vele malen sneller kunnen zijn, indien het beheren en uitwisselen van informatie een geolied proces is. Heel lang heeft men gedacht dat het implementeren van standaard applicatiepakketten de manier zou zijn om ISCM te regelen. De praktijk heeft – zeker in grotere en complexere organisaties – uitgewezen dat hiermee wellicht binnen een (deel) proces verbetering bereikt wordt, maar dat de verwarring tussen (deel)processen even groot is gebleven of misschien zelfs nog wel groter is geworden.

En de volgende ontwikkeling die het belang van ISCM alleen maar doet toenemen is zich aan het manifesteren. Service Oriented Architecture (SOA) is een software architectuurmodel dat enerzijds services (geautomatiseerde diensten) kent en anderzijds een mechanisme om informatie tussen services uit te wisselen. In principe zou dit het einde van de monolithische applicaties kunnen betekenen. Applicaties kunnen 'naar believen' worden opgesplitst in functies die vanuit meerdere processen aangeropen kunnen worden.

Het moge duidelijk zijn dat waar nu de problemen over betekenis, plaats en gebruik van informatie al heel groot zijn, dat die problemen werkelijk gigantisch zullen worden bij het verder opknippen van applicaties in services. ISCM is naar onze mening nu al onontbeerlijk, maar het is een *conditio sine qua non* voor het invoeren van een succesvolle SOA. Immers, SOA vereist zeer heldere definities over de functie van een service en de manier waarop een service aangesproken dient te worden.

De markt heeft dit onderkend en je ziet dan ook een 'exploderend' aantal van leveranciers die producten op de markt brengen

– soms al gedurende geruime tijd – die een deel van de functionaliteit van ISCM ondersteunen. Tegelijkertijd is er een ontwikkeling waarbij de 'big four' (IBM, Microsoft, Oracle en SAP) bijna wekelijks wel een aankondiging doen, waarin de overname van een – soms kleine – speler in dit domein gemeld wordt.

Invoering van ISCM

Het zal best nog de nodige tijd en inspanning vragen van de leveranciers om een goed gebalanceerd en geïntegreerd productaanbod te ontwikkelen. Betekent dat dan dat we daarop moeten wachten, of kunnen we daarop vooruit lopen? Om te beginnen is het natuurlijk zo dat veel organisaties al heel wat elementen van ISCM in een of andere vorm geïmplementeerd hebben.

Vervolgens is het ontwikkelen van een ISCM-beleid en het implementeren daarvan een proces dat op zichzelf al de nodige inspanning en doorlooptijd zal kosten, dus wachten op uitgekristalliseerde producten is wellicht niet zo verstandig. Beter is – misschien in 'proeftuin' omgevingen – ervaring op te doen met beperkte implementaties. Hierbij is het overigens van het allergrootste belang dat 'de business' betrokken wordt en niet zomaar betrokken, maar in een leidende rol. Net zoals SOA niet mogelijk is zonder ISCM zal ISCM niet mogelijk zijn zonder eigenaarschap van de gebruikersorganisatie.

Een manier om gebruikers beter in staat te stellen dat eigenaarschap ook werkelijk 'handen en voeten te geven' is het gebruik van het Metadata Frame². Daarbij wordt immers het belangrijkste onderdeel van de *conditio sine qua non* – het helder krijgen van de betekenis van de informatie – aangepakt op de enige wijze die *echt* door de business te valideren is. Immers binnen dit Framework wordt de semantiek ontleend aan de uiteindelijke kenner van de informatie: de echte Subject Matter Expert (domeindeskundige) en die zit in de business. En als die persoon historisch in de IT-organisatie is 'ingegroeid' (wat we meer dan eens tegenkomen) dan is ISCM meteen een mooie aanleiding om de kennis weer terug te brengen waar ze hoort: binnen de business. En kunnen de IT'ers weer de rol vervullen waarin ze behoren te zitten: het zorgen dat business wensen worden gerealiseerd.

Pas wanneer Information Supply Chain Management volledig is geïmplementeerd – zowel voor gestructureerde als ongestructureerde data – is er pas werkelijk sprake van een overkoepelend Enterprise Information Management in de ware zin van het woord.

Noten

1. Voor het gemak gaan we even voorbij aan het onderscheid tussen transacties en masterdata-elementen.
2. Zie het artikel *Het Metadata Frame in Actie* in DB/M 6 en 7, 2008.

Rob Arntz en Ad Stam

Drs. Ing. R.G. Arntz (rob.arntz@atosorigin.com) en A.A. Stam zijn beiden werkzaam als Senior Consultant bij Atos Origin/BI-CRM.