

Datamanagement: beter omgaan met bezit (2)

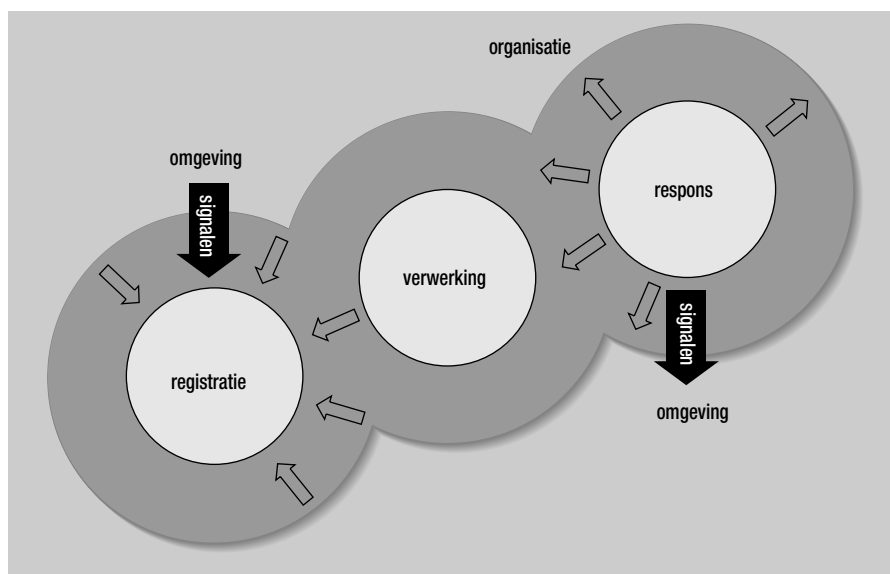
Registratieproces moet snel en nauwkeurig

Daan van Beek

Het proces van registratie van data vereist een snelle en nauwkeurige afhandeling van de stroom signalen die op een organisatie afkomt. De signalen komen uit de omgeving of de organisatie zelf. Daan van Beek bespreekt de drie doelstellingen binnen dit basisproces, die zich concentreren op doorlooptijd, nauwkeurigheid en feedback.

Data worden niet voor niets vastgelegd. In de eerste plaats bestaat er een wet die organisaties verplicht gegevens te administreren. Ten tweede beoogt een organisatie de data in te zetten in het bedrijfsproces en om te zetten tot informatie. Het bedrijfsproces levert dan niet alleen concrete goederen en diensten op. Deze gaan vergezeld van informatie voor klanten, leveranciers of andere partijen. Tegenwoordig is die informatie van onderscheidende waarde om de concurrentie het hoofd te bieden. Ten slotte vindt registratie (soms) louter en alleen plaats om er informatie van te maken.

Het registratieproces heeft als doel signalen uit de omgeving of de organisatie



FIGUUR 1: DE GENERIEKE BASISPROCESSEN BINNEN ORGANISATIES.

zelf op te vangen, om uiteindelijk hierop adequaat te antwoorden. In figuur 1 zijn de generieke basisprocessen in een organisatie weergegeven, waaronder het registratieproces. Het genereren van een adequate respons hangt onder meer af van hoe succesvol datamanagement het registratieproces ondersteunt.

Het datamanagement in het registratie-

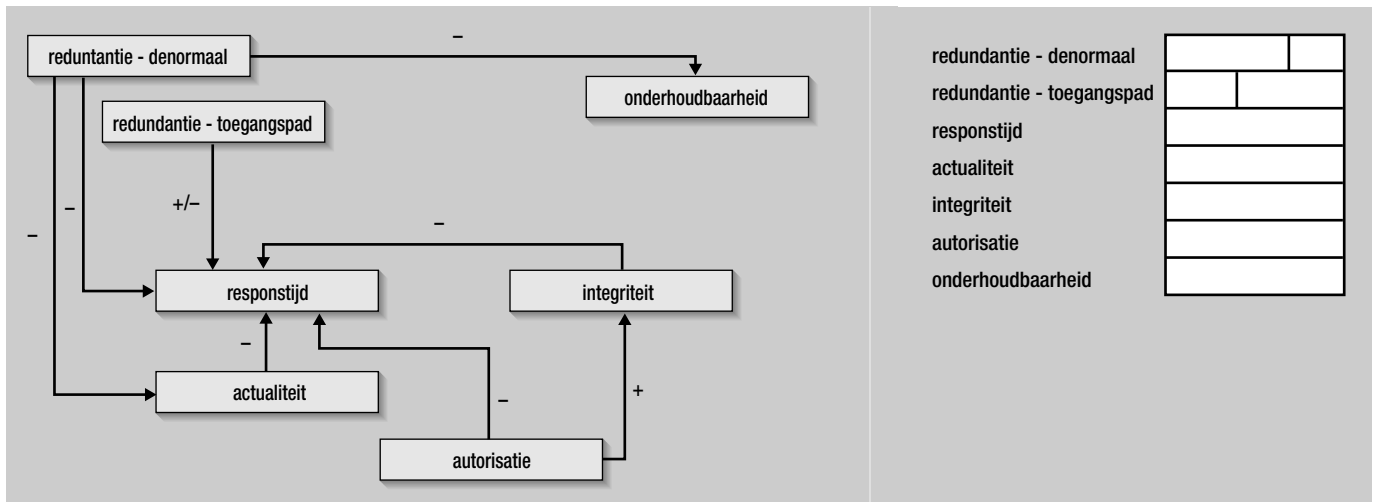
proces verschilt niet wezenlijk maar wel qua invulling van andere generieke basisprocessen. In tegenstelling tot de andere basisprocessen is redundantie tijdens dit proces eigenlijk niet gewenst. De doelstellingen van het registratieproces bepalen de wijze waarop het datamanagement in dit proces vorm krijgt. Deze doelstellingen zijn:

- het realiseren van een zo kort mogelijke doorlooptijd per gebeurtenis;
- een nauwkeurige, complete en onduidelzinnige registratie van de signalen;
- het geven van directe feedback of de gebeurtenis kan leiden tot het gewenste resultaat.

De eerste twee doelstellingen wisselen elkaar af qua belang. Bij stamgegevens met een lage mutatiegraad is nauwkeurige

Basisprocessen voor datamanagement

De drie basisprocessen registratie, verwerking en respons nopen tot een uitgekende methode voor datamanagement. In een nieuwe serie artikelen neemt Daan van Beek deze aspecten voor ieder basisproces afzonderlijk onder de loep. Na een inleidend artikel, in het vorige nummer van DB/M, komt op deze pagina's het registratieproces uitvoerig aan de orde.



FIGUUR 2: DE OORZAAK-EN-GEVOLG-RELATIES VAN DATAMANAGEMENTASPECTEN BINNEN HET REGISTRATIEPROCES.

registratie belangrijker dan snelle doorlooptijd. Stamgegevens veranderen immers niet zo vaak. Bovendien heeft een fout in een stamgegeven veel grotere consequenties dan een fout in een transactiegegevens.

Toch is validatie van de stamgegevens zelf niet zonder meer belangrijker dan die van transactiedata. Met name bij geautomatiseerde elektronische aanlevering van deze laatste kan een lek in de validatiestrategie leiden tot nare gevolgen, waaronder omzetverlies, bijvoorbeeld wanneer verminkte productnummers het systeem binnendringen zonder dat dit direct opvalt.

De laatste doelstelling wordt in dit artikel buiten beschouwing gelaten, omdat deze lastig is te verbinden met datamanagementspecten.

Een voorbeeld illustreert deze doelstellingen. Een bedrijf wil goederen afnemen en belt onze organisatie (de gebeurtenis). Een telefoniste neemt de bestelling op en registreert de naam van het bedrijf en de gewenste goederen. Tijdens dit gesprek ziet de telefoniste dat het bedrijf nog niet als klant geregistreerd staat. Naast deze gegevens zal zij ook de ordergegevens opnemen (de signalen). Nadat alle signalen volgens de regels zijn geregistreerd, geeft de telefoniste aan dat de goederen binnen twee dagen worden geleverd (de feedback). De telefoniste is tien minuten bezig geweest met het opnemen van de bestelling.

Goed geïmplementeerd datamanagement draagt bij aan het realiseren van deze doelstellingen. Het datamanagement ondersteunt echter iedere doelstelling op zijn eigen manier, en dit leidt tot dilemma's. Voor een doorlooptijd moet de transactie zo snel mogelijk worden opgeslagen. Maar om de gegevens ook nauwkeurig en ondubbelzinnig op te slaan, moeten ze eerst integer worden bevonden. En dat laatste kost nu eenmaal doorlooptijd. Deze dilemma's, in de vorm van relaties van oorzaak en gevolg, worden in figuur 2 weergegeven. Responsstijd en integriteit zijn respectievelijk verbonden met de eerste en de tweede doelstelling en staan daarom centraal.

DOORLOOPTIJD

In grote databases een snelle responsstijd krijgen, is lastig, maar niet onmogelijk. Verschillende aspecten van datamanagement helpen daarbij. Allereerst is een genormaliseerd datamodel van wezenlijk belang. Hierdoor wordt redundantie vermeden. De gegevens worden op één plek in de organisatie én -op één plek- in de database geregistreerd en vastgelegd. Het aantal met de gebeurtenis verbonden deeltransacties neemt in zijn totaliteit af en dat beperkt de hoeveelheid bits die de database wegschrijft. Zo wordt de doorlooptijd verkort. Verwijdering van overtoolligheid in een datamodel heeft dus een aanwijsbare

versnelling tot gevolg. Ten tweede zijn logische toegangspaden (indexen) nodig om een snelle responsstijd te garanderen. Hoewel hierdoor de redundantie weer toeneemt, vergen toegangspaden in de meeste databases geen extra applicatielogica voor het onderhoud daarop. Dit wordt immers door het dbms zelf geregeld.

Toegangspaden

Toegangspaden ter ondersteuning van het registratieproces zijn alleen nodig om gegevens te bekijken, te wijzigen of te verwijderen. Een toevoeging wordt immers gewoonweg achteraan gezet. Hoe meer toegangspaden en hoe meer attributen daarin zijn opgenomen, hoe trager het dbms een transactie verwerkt. Iedere actie van toevoegen/verwijderen leidt onvermijdelijk tot het bijwerken van de toegangspaden op schijf. Zijn er honderden tot duizenden transacties per seconde actief -binnen een registratieproces geen zeldzaamheid- dan dient het aantal toegangspaden beperkt te worden. Afgezien van geavanceerde zoekmogelijkheden volstaat indexatie op de primaire sleutel en bijbehorende omschrijving voor het registratieproces, tenzij bepaalde managementinformatie nodig is om de gebeurtenis af te ronden. Denk dan bijvoorbeeld aan het berekenen van de kredietwaardigheid, waarbij voor die klant de waarde van alle nog niet betaalde orders moet worden gesommeerd. Snelle beschikbaarheid van

deze informatie vereist een extra toegangspad op in elk geval klantnummer.

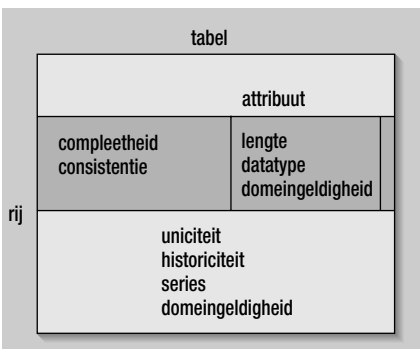
Managementinformatie buiten productie-omgeving

Dit raakt op zijn beurt een ander aspect van redundantie: managementinformatie buiten de productie-omgeving, zoals een datawarehouse. Juist die overtuiging, in de vorm van een kopie van de database, ontlast het productiesysteem; het aantal toegangspaden kan tot een minimum beperkt blijven. Vaak moeten enorme hoeveelheden rijen worden geselecteerd om de informatie te berekenen.

Nadeel is dat soms vanuit het productiesysteem een aanroep moet plaatsvinden die de gevraagde informatie oplepelt uit de managementinformatie-omgeving. De complexiteit van de IT-infrastructuur neemt hierdoor toe, als dit aanroepen al mogelijk is. In ieder geval zal de responstijd voor andere gebruikers niet nadelig worden beïnvloed. In sommige gevallen moet de managementinformatie-omgeving realtime worden bijgewerkt, als de actualiteit dit vereist. Hierop voortbordurend zal de managementinformatie-omgeving bij veelvuldige aanroepen vanuit het productiesysteem aan dezelfde ICT-eisen moeten voldoen als de productie-omgeving.

Autorisaties

Autorisaties, waarbij delen van de database worden afgeschermd, maken het vraagstuk nog complexer. Medewerkers mogen bijvoorbeeld alleen gegevens raadplegen en muteren van hun werkmaatschappij. Toegangspaden dienen nu voorzien te zijn van een extra attribuut werkmaatschappij.



FIGUUR 3: NIVEAUS VAN DE FACETTEN VAN INTEGRITEIT.

Dit leidt echter tot een dilemma. Uit redundantie-oogpunt is ooit bepaald dat dit gegeven slechts wordt vastgelegd bij een client. Hoewel moderne dbms'en toegangspaden over meerdere gekoppelde tabellen ondersteunen, bijvoorbeeld door middel van indexed views, is vaak voor het gemak een extra attribuut opgenomen in de transactietabel. Beveiliging en responstijd hebben hier prioriteit boven het beperken van redundantie. Met autorisaties wordt voorkomen dat onbevoegde personen de geregistreerde signalen verminken of weggooien. De integriteit van de data is ook een kwestie van een goede beveiliging.

NAUWKEURIG EN COMPLEET

Op integere data kun je bouwen. Ze zijn betrouwbaar en congruent. Integere gegevens zorgen naast andere eigenschappen voor goede toepassingen. Het datamanagement dat het registratieproces in organisaties ondersteunt zal gegevens moeten valideren voordat ze worden opgeslagen in de database. Het validatieproces houdt rekening met diverse facetten van integriteit, zoals lengte, datatype, uniciteit, historie, domeingeldigheid, compleetheid en consistentie. Al deze facetten beperken de mogelijkheden van het informatiesysteem. Strikte afdwinging van integriteitsregels beoogt kwalitatief hoogstaande data, maar vermindert ook de flexibiliteit die een eindgebruiker ervaart.

Figuur 3 geeft de niveaus van de facetten van integriteit weer. Over het algemeen geldt dat: hoe breder het facet, hoe meer processing plaatsvindt. Controle van lengten en datatypes is relatief eenvoudig, maar controle op compleetheid of uniciteit vereist mechanismen die de computer zwaarder belasten, omdat meer attributen of rijen bekeken worden. Het aanhouden van deze 'volgorde' bij de controle van de data voorkomt onnodige processing.

Uniciteit

De uniciteit zorgt ervoor dat waarden in de primaire sleutelvelden niet in meerdere rijen kunnen voorkomen. Een meer geavanceerde vorm hiervan bewaakt dat

klanten met de dezelfde naam of postcode en huisnummer niet worden opgeslagen. Vaak gebeurt dit toch ongemerkt, en de daarmee gepaard gaande vervuiling tast het efficiënt en doelgericht functioneren van een organisatie aan. Het aantal klanten dat men denkt te hebben komt niet overeen met de werkelijkheid. De gemiddelde klantwaarde neemt relatief gezien af; trendbreuken treden op. Kortom, de betrouwbaarheid van managementinformatie staat ter discussie.

Soms kiest men er zelfs bewust voor klanten dubbel op te voeren, omdat het productiesysteem niet de gewenste functionaliteiten biedt die eindgebruikers of klanten wensen. Dit is een van de dubblures om aan te pakken bij de invoering van een klantrelatiesysteem of een datawarehouse. Het is vaak onhaalbaar deze dubblure in het productiesysteem zelf op te lossen vanwege de complexiteit van de conversie van alle aan de klant gekoppelde gegevens.

Een slimme en pragmatische oplossing is deze dubbele 'klanten' in het productiesysteem aan elkaar te knopen, zodat een hoofdklant ontstaat. Door bij het ontwerp en het vullen van het klantrelatiesysteem of het datawarehouse hiermee rekening te houden, ontstaat toch een integraal beeld van de klant.

Domeingeldigheid

De domeingeldigheid let erop dat gegevens waarnaar verwezen wordt, zogenaamde vreemde sleutels, bekende en dus geldige waarden zijn. Bij het opvoeren van een order van een klant wordt eerst gekeken of het ingevoerde klantnummer voorkomt in het klantenbestand. Zo niet, dan wordt de order niet opgeslagen in de database en zal de gebruiker eerst de klant moeten opvoeren.

Testen of het domein geldig is, volstaat niet altijd. Er kan namelijk een weliswaar bekende waarde worden ingegeven, maar dat hoeft niet de juiste te zijn. Dit is alleen te vermijden door ook te testen of alle ingevoerde gegevens van dat moment met elkaar in overeenstemming zijn én met de rest van de database. Het facet consistentie let daar op.

Consistentie

Een voorbeeld verduidelijkt het facet consistentie. Iemand die niet in ploegendienst werkt, mag ook niet toebehoren aan een ploeg. Dit raakt een ander punt, namelijk de relatie tussen de integriteit van bedrijfsdata en het ontwerp van de gebruikersinterface. Vaak houdt men hiermee rekening door het veld ploeg uit te schakelen zolang het veld ploegendienst niet aan staat. Zo voert de gebruiker geen overbodige gegevens in en wordt mede de integriteit van de data bewaakt.

Er schuilt een addertje onder het gras: als de gebruiker eerst ploegendienst aanzet en een ploeg invult en daarna ploegendienst weer uitschakelt. De twee oplossingen liggen voor de hand: valideer de data door validatieregels op te nemen in de database of zorg dat het veld ploeg wordt leeggemaakt zodra ploegendienst wordt uitgezet. Beide oplossingen zijn legitiem, hoewel de eerste het fraaist is; tijdens de invoer zijn wellicht niet alle data voorhanden om alle facetten van de integriteit te waarborgen. Sommige facetten vereisen namelijk gegevens uit andere tabellen. Dit genereert extra netwerkverkeer door de intensievere communicatie tussen gebruikersinterface en database als de interface draait op een andere machine dan de database.

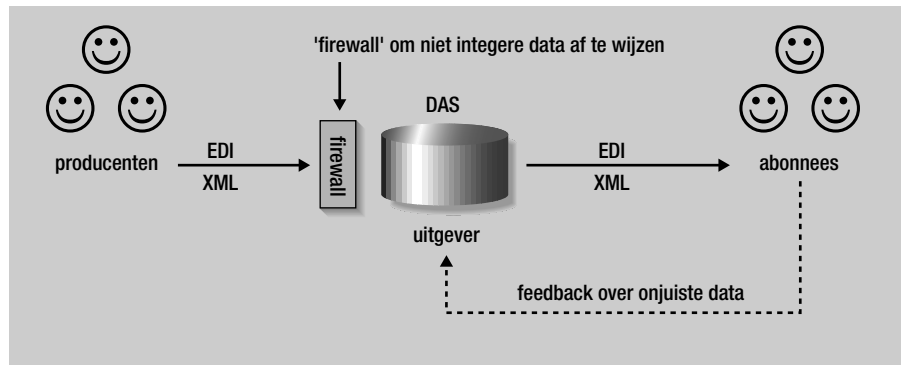
Validatieregels horen thuis in de database. Dat biedt zeker grote voordelen als gegevens vanuit meerdere toepassingen worden gemuteerd.

Compleetheid

De compleetheid test of alle met een gebeurtenis samenhangende verplichte gegevens een waarde bevatten. In principe is de compleetheid te testen op attribuutniveau, behalve waar het invullen van bepaalde velden afhankelijk is van de ingevoerde waarde van een ander veld. In dat geval worden meerdere attributen in ogenschouw genomen, of zelfs de hele rij.

WEBDATABASES

Klanten worden steeds meer eigenaar van hun eigen data. Dat daarmee wordt



FIGUUR 4: HET CONCEPT VAN DATA ALIGNMENT SERVICES.

bespaard op arbeidskosten is evident. Door het fenomeen ketenintegratie ontstaan min of meer 'virtuele' ondernemingen, geleid door externe partijen. Dit stelt nog hogere eisen aan de ICT-omgeving, in het bijzonder op het gebied van datamanagement. Acceptable responstijden, autorisatie en actuele en valide data zijn hierbij essentieel. Validatie van gegevens speelt e-business-systemen parten, ondanks het gewicht ervan. Uit onderzoek komt naar voren dat de kwaliteit van via het Web ingevoerde gegevens in veel gevallen niet bijster hoog is. Dat mensen met name hun persoonlijke gegevens nog niet goed durven toevertrouwen aan het www is daar mede debet aan. Maar er is nog een andere reden.

De voertaal van het Web, html, kent nog niet de rijkdom aan interfacecontrols die nodig is om de invoer van de gebruiker in goede banen te leiden die eigen is aan

Windows-gebaseerde systemen. Allerlei essentiële vormen van interacties en feedback worden gemist, waardoor het systeem de invoer van onjuiste of inconsistente data toestaat.

Validatie van de data kan vaak niet worden uitgevoerd door de database. Immers, deze staat vaak op grote afstand van de client, waardoor snelle feedback over de kwaliteit van de ingebrachte data niet mogelijk is. Mid-office-databases, vaak ingezet ter verbetering van web-responstijden, bevatten vaak weer niet de benodigde validatieregels. Het succes van e-business hangt mede af van de juistheid en integriteit van de gegevens.

DATA ALIGNMENT SERVICES

Binnen diverse bedrijfskolommen, zoals food en farma, komen *data alignment ser-*

Redundantie

Een kopie van de database, of elementen daaruit - voor veel ICT-professionals is het vloeken in de kerk. En dat is niet voor niets. Het up-to-date houden van deze kopieën vereist immers de nodige aandacht en vaak ook het verrichten van handmatige handelingen of een complexe technologische infrastructuur die de kopie (realtime) synchroniseert met de bron. Behalve in het geval van grote e-business-databases, klantrelatiesystemen en datawarehouses verdient het aanbeveling een atomair gegeven slechts op één plek vast te leggen, zowel in de database zelf als binnen het ondernemingsnetwerk. De genoemde uitzonderingen lijken te verdwijnen door voortschrijdende technologieverbeteringen op het gebied van systeempartitioning, clustering, capaciteit en performance. Maar met name managementinformatietransacties zijn zo anders van aard dan operationele transacties dat we denken dat de toename van het aantal data (ieder jaar een verdubbeling) gelijke tred zal houden met de toegevoegde waarde van de technologische verbeteringen. Dit impliceert dat de managementinformatie- en de productie-omgeving nooit kunnen worden samengevoegd, de toekomstvisie van enkele deskundigen ten spijt. Het vraagstuk van redundantie blijft voorlopig nog op de agenda.

vices (DAS) op voor productgegevens. Zo'n service bestaat uit een gevulde database met een datamodel waarin alle aangesloten leveranciers op uniforme wijze de gegevens aanleveren. Andere partijen in de logistieke keten, bijvoorbeeld groothandels en klanten, abonneren zich op de DAS. Deze services sluiten aan bij de eerste twee doelstellingen in het registratieproces: het verkorten van de doorlooptijd én het zo goed mogelijk vastleggen van data. Productgegevens worden pas in de DAS opgenomen nadat een groot aantal controles zijn doorlopen. Daarnaast worden steekproefgewijs metingen verricht aan de producten zelf om de door de leverancier ingevoerde afmetingen te controleren.

Een DAS helpt organisaties met minder mensen de kwaliteit van productdata te verhogen. Zij voorkomt daarmee foutieve orders, lege schappen en ontevreden klanten. Logistieke processen in de keten, die steeds meer elektronisch geschakeld zijn, zullen door een DAS vlotter en met minder fouten verlopen.

Het 'nadeel' van deze centrale vorm

van registratie is een grote afhankelijkheid ten opzichte van het functioneren van een DAS. De gehele bedrijfskolom ondervindt wellicht ernstige hinder als deze niet goed functioneert. Omzetverliezen strekken zich nu uit over de totale keten. Daarom zijn goede afspraken tussen uitgever en abonnee nodig.

UITDAGINGEN

Datamanagement ter ondersteuning van het registratieproces kenmerkt zich door twee grote uitdagingen: het garanderen van een zo kort mogelijke doorlooptijd en het bewaken van de integriteit van de gegevens. Redundantie binnen een database moet zoveel mogelijk worden vermeden, behoudens elementaire toegangspaden. Redundantie in de vorm van een kopie van de database kan de responstijd voor bijvoorbeeld webdatabases juist minimaliseren.

Met name (internationale) consumentenwebsites, waarbij de vraag onvoorspel-

baar is, zijn gebaat bij een eigen database en server, mits deze goed worden ontworpen en meer dan gewone aandacht uitgaat naar automatische synchronisatie en validatie van gegevens.

Zowel webdatabases als Data Alignment Services minimaliseren de doorlooptijd per gebeurtenis en kunnen, mits voorafgegaan door een uitgekiend beleid, de organisatie op een hoger plan brengen. ●

Literatuur

Meer informatie over Data Alignment Services is te vinden op www.eandas.nl

Daan van Beek MSc (daanvanbeek@hetnet.nl) is Manager Datamanagement Services bij een groothandel in geneesmiddelen en farmaceutische producten.

U P D A T E

INFORMATION BUILDERS PRESENTEERT PLATFORM VOOR REALTIME 'INFORMATISERING'

Information Builders heeft op de jaarlijkse gebruikersconferentie "Summit 2002" in de Amerikaanse stad Baltimore een platform voor *information delivery* ten doop gehouden. Dit platform, onder de WebFocus-noemer van deze leverancier, biedt een enkelvoudige infrastructuur voor het opvragen en analyseren van informatie, voor bedrijfsbrede toegang tot en distributie van rapportages en voor het realiseren van bedrijfsportals.

WebFocus Information Delivery Platform stelt organisaties in staat essentiële, tijdkritische informatie snel over de 'extended enterprise' te verspreiden: onder medewerkers, business partners, leveranciers, klanten en prospects.

"Business intelligence ontwikkelt zich van een zeer gespecialiseerd werkgebied voor analisten tot de praktijk van alledag voor de

doorsnee zakenman. Bedrijven beginnen BIsystemen te gebruiken voor het afleveren van informatie buiten hun firewalls", zegt David Folger, vice president Web & Collaboration Strategies bij Meta Group. Volgens hem biedt Information Builders "een heel flexibel platform voor het aanboren van een groot aantal databronnen, inclusief legacy systemen, en het verspreiden van realtime informatie. Gebruikers stellen de parameters in voor de specifieke informatie die zij willen."

Information Builders kondigde op de Summit verder aan de WebFocus-lijn uit te breiden met ondersteuning voor webservices. Alle WebFocus-applicaties zullen in de loop van 2002 geschikt gemaakt worden voor webservices en via een reeks nieuwe releases beschikbaar komen. De leverancier beschouwt webservices als de snel opkomende standaard voor *cross-environment computing* via Internet.

Op dit moment is het aantal via web-services uitwisselbare soorten gegevens beperkt; Information Builders zegt daar met WebFocus

meer dan 85 uiteenlopende bronsoorten aan toe te voegen.

Webservices-toepassingen zijn ook beschikbaar om bedrijven te helpen de multiplatform integratieproblematiek op te lossen. iWay Software, dochterbedrijf van Information Builders, biedt Business Services, een webservices-raamwerk om meer dan 200 verschillende proprietary informatiesystemen te benaderen.

iWay kondigde op de Summit de Enterprise Integration Suite 2002 aan. De nieuwe release brengt technologieën samen waarmee applicatie-integratie op een grafische manier gerepresenteerd kan worden.

Verder zijn twee tools uitgebracht die het makkelijker moeten maken webservices te creëren vanuit bestaande applicaties en legacy systemen.

Voor meer informatie: www.information-builders.com/new/index.html ●