

Hergebruik van content cruciaal

# Content management en XML: een veelbelovend stel

Paul Baan en Marc Speyer

**I**n de voor Internet zo kenmerkende dadendrang zijn veel sites al snel tegen de befaamde webmaster-bottleneck aangelopen. De site is verworpen tot een onbeheersbare hoeveelheid informatie, waarvoor geen enkele garantie met betrekking tot correcte links, consistente lay-out of actualiteit van de informatie valt af te geven. Content moet daarom 'gemanaged' worden. En gelukkig zijn er allerlei marktpartijen die deze behoefte tijdig hebben ingezien en nu in staat zijn zogeheten (web)content management-pakketten aan te bieden. Maar deze pakketten, zo blijkt, schieten vaak tekort in de ondersteuning van XML. Hergebruik van content is dus niet altijd mogelijk.

Nu is slecht contentbeheer helemaal niet zo'n recent ervaren probleem. Eind jaren zestig werden reeds de eerste stappen gezet om een oplossing te vinden voor het optimaal beheren van allerlei soorten informatie. Toen was het onder andere het

*Kernprobleem blijft de inhoud te scheiden van de opmaak*

Amerikaanse ministerie van Defensie dat inzag dat bijvoorbeeld specificaties en documentatie rondom gepantserde tanks en duikboten wel eens langer konden meegaan dan een versie van een tekstver-

werker en misschien wel langer dan het computersysteem waarop deze tekstwerker draait. De vliegtuigindustrie kreeg ook al vroegtijdig te maken met content management. Zeker als daarbij in acht wordt genomen dat een vliegtuig conform de wensen van de luchtvaartmaatschappij wordt opgeleverd. Beheren van de tientallen jaren te gebruiken handboeken en

documentatie is zonder content management eenvoudigweg niet mogelijk. De grootste vliegtuigbouwer levert al jarenlang zijn gebruikershandboeken en documentatie op cd-rom en in diverse formaten. De klant bepaalt zelf welk formaat het meest geschikt is voor bijvoorbeeld een geprinte versie. Het kennisgebied van contentbeheer kent een voorgeschiedenis



van zo'n drie decennia, waarin al vele standaards tot volle wasdom zijn gekomen. Het is dan ook zeer opvallend te constateren dat de meeste partijen die nu het hardst hun content management-software aanprijzen, relatief zeer jeugdig zijn, en claimen een oplossing te hebben gevonden voor een recent probleem. Geregeld komt het voor dat dezelfde partijen zich weinig tot niets aantrekken van de vele beschikbare standaards. Eerlijkheidshalve moet worden opgemerkt dat de oudere vaak ISO-standaards (SGML en aanverwante standaards) minder goed zijn voor nieuwe media.

### WAAR GAAT HET MIS?

De karakteristieken van nieuwe media en de mediaspecifieke functies mogen dan anders zijn, het kernprobleem blijft aanwezig. Inhoud scheiden van opmaak blijft nodig. Opmaak is mediumspecifiek, inhoud kan worden hergebruikt. Het is deze inhoud die vandaag de dag juist als een asset moet worden gezien. De investeringen in een asset zijn sneller terug te verdienen als we deze vaker kunnen inzetten of zelfs kunnen verkopen. Concreet betekent dit dat content waar mogelijk mediumneutraal moet worden opgeslagen. Eveneens noodzakelijk is het verrijken van content. Iedereen heeft het tegenwoordig over personalisatie. Het kunnen vinden van de juiste content is weliswaar de eerste stap naar personalisatie, maar als de content eenmaal is gevonden, is presentatie (mediumtype, wijze van presenteren) de volgende stap.

***XML bleek een verrassend eenvoudig te implementeren standaard met een sterke uitdrukkingvaardigheid***

Het moge duidelijk zijn dat met opge maakte content geen volledige personalisatie kan worden bereikt. Vandaag de dag moeten we zelfs de content zodanig structureren, dat de inhoud zich kan aanpassen

## Wie heeft gelijk?

Als we wat uitgebreider stilstaan bij het begrip content management, blijkt dat dit voor tweërlei uitleg vatbaar is, namelijk het beheren van:

- alle assets voor een bepaalde toepassing;
- assets op basis van de inhoud.

Afhankelijk van de definitie kom je tot verschillende systemen. Zo vallen webcontent management-systemen meestal onder de eerste categorie. Eigenlijk zijn dit soort systemen niet meer dan document management volgens een werkstroom, maar dan geoptimaliseerd voor elektronisch leveren van documenten. Deze systemen hebben hun bestaansrecht, maar lossen niet het feitelijke contentbeheerprobleem op, namelijk het eenvoudig kunnen hergebruiken van de inhoud. Hergebruik is moeizaam doordat inhoud en opmaak met elkaar zijn verweven. De eindbestemming is vanaf het begin reeds bekend: bijvoorbeeld een website.

Content management-systemen die in de tweede categorie vallen, kijken als het ware naar de inhoud van de content en nemen op basis van deze inhoud beslissingen. Zo kan de werkstroom, afhankelijk van het wel of niet aanwezig zijn van afbeeldingen of verwijzingen, een andere route nemen. Een wijziging van een foto als gevolg van een nieuwe verpakking voor een product kan zo automatisch een werkstroom in gang zetten, niet onbelangrijk voor het consistent en up-to-date houden van de content.

aan de gebruiker, zoals het voorbeeld van een recept waarin de hoeveelheid ingrediënten, bereidingswijze en bereidingstijd afhankelijk is van het aantal in te geven personen.

### MEDIUMNEUTRAAL

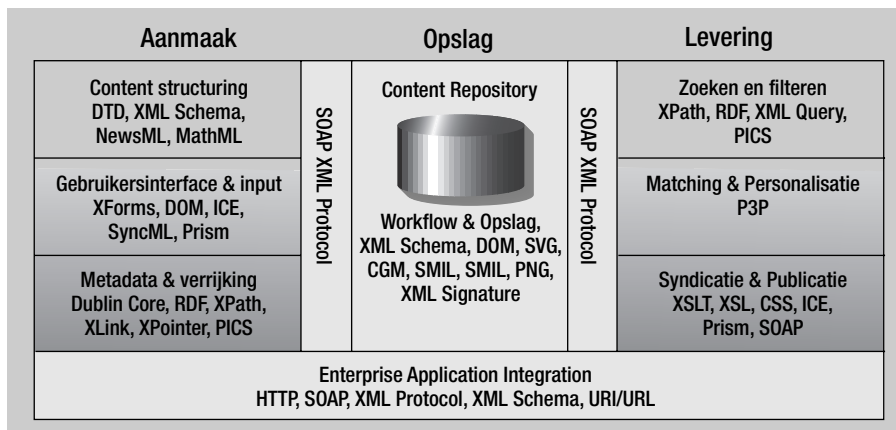
Als we vandaag de dag spreken over mediumneutraal, hebben we het al snel over XML. XML is afgeleid van de ISO-standaard SGML, waarbij de ontwerpers minder gebruikte -maar moeilijk te implementeren- onderdelen hebben geschrapt en faciliteiten ten behoeve van nieuwe (elektronische) media hebben toegevoegd. Het resultaat bleek een verrassend eenvoudig te implementeren standaard met een sterke uitdrukkingvaardigheid. Vele leveranciers, waaronder Microsoft en Sun, adopteerden al snel de XML-standaard. De XML Working Draft dateert van eind 1996. Ruim een jaar later gaf het World Wide Web Consortium (W3C) de XML-standaard de proposed recommendation-status. Vanaf dat moment ging het hard met de toepassingen gebaseerd op XML.

Aanverwante standaards, zowel W3C als niet-W3C, ontwikkelden zich in hoog tempo. Vandaag de dag zijn er meer dan een dozijn XML-standaards van W3C, waarvan het merendeel nog niet de genoemde status heeft bereikt. Het aantal XML-standaards buiten de W3C voor een specifiek toepassingsdomein is nog veel

***Voor systemen die daadwerkelijk op basis van inhoud van content werken, is XML al snel een vereiste***

groter (zoals WML, ICE, SOAP, ebXML, Prism, NewsML enzovoort).

Terwijl SGML voornamelijk werd toegepast voor het structureren van content voor publicatiedoeleinden, is het toepassingsgebied van XML nu al vele malen groter. Een belangrijke toepassing van XML is het uitwisselen van informatie tussen systemen en applicaties. Deze niet zichtbare toepassing van XML speelt nu al een belangrijke rol in middleware. Dankzij de beschikbaarheid van diverse XML-



**FIGUUR 1: XML CONTENT MANAGEMENT FRAMEWORK (GEÏNTEGREERDE TOEPASSINGEN ZIJN DONKER GEKLEURD).**

standaards en het brede toepassingsgebied kunnen -of beter gezegd: moeten- content-beheersystemen zich baseren op XML-standaards. Hierdoor is men verzekerd van werkelijk open systemen en optimale uitwisseling van content. Zoals reeds eerder betoogd, moet content waar mogelijk worden hergebruikt. Een content management-systeem op basis van XML-standaards zal dit zeker bevorderen.

**BEOORDELEN**

Het raamwerk in figuur 1 geeft aan waar welke XML-standaards een rol kunnen spelen. Dit raamwerk is bruikbaar om een content management-systeem, waarvan de leverancier XML-ondersteuning claimt, te beoordelen. Onder de categorie geïnte-

greerde toepassingen vallen de e-business-suites en de web- en traditionele content management-systemen - de laatste categorie wordt vaak voor uitgeefdoelein-

*De XML-familie van standaards biedt voldoende ondersteuning om een totaal content management-systeem in te richten.*

den toegepast. Momenteel maakt geen van de geïntegreerde pakketten op de markt volledig gebruik van XML-standaards. Deeloplossingen, zoals XML content repositories of XML editors, zijn meestal wel in verregaande vorm geba-

seerd op XML-standaards. In de nabije toekomst zullen fabrikanten van content management-software steeds meer standaards integreren, met name als enkele jonge standaards zich hebben bewezen.

**TOEKOMST**

Contentbeheer is niet iets van de laatste tijd, maar werd al veel langer toegepast in de defensie- en vliegtuigindustrie. Content management-systemen beheren doorgaans alleen de documenten en de werkstroom voor een bepaalde toepassing. Document management is hier meer op zijn plaats. Er zijn ook systemen die daadwerkelijk op basis van de inhoud van content werken. XML als mediumneutrale opslagvorm is dan al snel een vereiste. De XML-familie van standaards biedt voldoende ondersteuning om een totaal content management-systeem in te richten. In welke mate een softwarepakket XML ondersteunt, kan met behulp van een XML content management framework worden bepaald. De huidige pakketten schieten op dit gebied nogal eens tekort, maar de verwachting is dat dit in de toekomst snel gaat veranderen. ●

**ADVERTEERDERSINDEX**

|                           |               |                             |               |
|---------------------------|---------------|-----------------------------|---------------|
| <b>Allaire</b>            | <b>38</b>     | <b>Intersystems Benelux</b> | <b>11</b>     |
| <b>Array Publications</b> | <b>29, 50</b> | <b>ISES International</b>   | <b>7</b>      |
| <b>BEA Systems</b>        | <b>72</b>     | <b>KLPD</b>                 | <b>68</b>     |
| <b>BNT Holland</b>        | <b>14</b>     | <b>Microsoft</b>            | <b>64, 65</b> |
| <b>Borland</b>            | <b>42</b>     | <b>Microstrategy</b>        | <b>46</b>     |
| <b>CMG</b>                | <b>76</b>     | <b>Quant Systems Europe</b> | <b>30</b>     |
| <b>Cognicase</b>          | <b>62</b>     | <b>Rational</b>             | <b>2</b>      |
| <b>Compaq Computer</b>    | <b>8</b>      | <b>Roccase Public</b>       | <b>74</b>     |
| <b>Database Systems</b>   | <b>35</b>     | <b>Rogue Wave Software</b>  | <b>70</b>     |
| <b>ERD</b>                | <b>28</b>     | <b>Sitraka</b>              | <b>75</b>     |
| <b>Excelon</b>            | <b>18</b>     | <b>Software AG</b>          | <b>4</b>      |
| <b>Hyperion</b>           | <b>12</b>     | <b>Synergetics</b>          | <b>27</b>     |
| <b>IBM Nederland</b>      | <b>16, 40</b> | <b>WRQ Europe</b>           | <b>60</b>     |
| <b>Informatica</b>        | <b>66</b>     | <b>X-Hive</b>               | <b>59</b>     |

Ir. Marc Y. Speyer (marc.speyer@capgemini.nl) en

drs. Paul Baan (paul.baan@capgemini.nl) zijn respectievelijk senior consultant media en consultant media bij Cap Gemini Telecom, Media and Networks Nederland BV.

Zij houden zich sinds ruim acht jaar respectievelijk ruim drie jaar bezig met uitgeefprocessen, uitgeefsystemen en content management.