

XML content management

Toepassing Oracle XML database en Xopus

In dit artikel komen XML content management oplossingen aan bod. Naast het beschrijven van deze oplossingen in algemene zin worden er twee business cases behandeld: onderhouden en publiceren van (XML) documenten bij Schuitema en de toepassing van XML content management binnen Oracle Portal bij Zorg en Zekerheid. In latere artikelen volgen andere praktijkvoorbeelden van het gebruik van XML op het Oracle platform.

Eerder verscheen in dit blad een reeks van vier artikelen van Erwin Groenendal en Marc Vahsen over XML-ondersteuning in de Oracle database. Daarin werd veel technologie besproken: 'SQL/XML: Extensies voor het genereren van XML data', 'XMLType: Native XML datatype in de database', 'Oracle XDB Repository: File system in de database' en 'DBMS_XMLDOM: XML Document Object Model API'. Na de technologische en theoretische aspecten, is het nu tijd voor praktijkvoorbeelden van oplossingen.

Data centric XML versus document centric XML

XML-gegevens kunnen in twee categorieën ingedeeld worden: data centric XML voor machine consumption en document centric XML voor human consumption.

Data centric XML: machine consumption

In deze categorie vallen bijvoorbeeld objecten uit (administratieve) applicaties: gegevens van een werknemer, een verzekeringspolis, dienstregelingen, et cetera. Dergelijke XML-data worden in de regel gegenereerd vanuit een systeem – er is niet iemand die de gegevens met een text editor of een XML-ontwikkelt tool zoals XMLSPY intypt – en wordt vrijwel altijd weer door een systeem verwerkt: machine consumption.

De toepassing van data centric XML is dan ook vooral te vinden in business-to-business (B2B) en application-to-application (A2A) integratieoplossingen, waarbij elektronische gegevensuitwisseling plaatsvindt tussen bedrijven en applicaties. Data centric XML wordt daarnaast ook veel gebruikt als tussenformaat voor het

presenteren van data in HTML, SVG, PDF et cetera. Hierbij worden relationele data omgezet in data centric XML en vervolgens met XSLT getransformeerd naar één van de genoemde presentatieformaten.

Document centric XML: human consumption

Voorbeelden van document centric XML zijn nieuwsberichten, tekst op webpagina's en webcontent in het algemeen. Maar ook whitepapers, dit artikel (zie www.cumquat.nl om dit artikel in XML te bekijken), polisvoorwaarden en documenten in brede zin. In tegenstelling tot data centric XML wordt document centric XML in de regel juist wel door een persoon (auteur) geschreven en is de informatie bedoeld om door een mens gelezen te worden: human consumption. Document centric XML wordt toegepast bij medianeutrale opslag van content,

De toepassing van data centric XML is vooral te vinden in B2B en A2A-integratie-oplossingen

waarbij de content op verschillende manieren en via verschillende kanalen gepubliceerd kan worden. Deze toepassing wordt XML content management genoemd en hiervan worden in dit artikel praktijkvoorbeelden gegeven.

Er zou ook nog een derde categorie van XML documenten onderkend kunnen worden, waarin op XML gebaseerde programmeertalen vallen: XHTML, SVG, XSLT, XSL-FO, XML Schema, et cetera.

Voordelen XML content management

De twee belangrijkste voordelen van XML content management zijn het scheiden van data en presentatie, en verrijking.

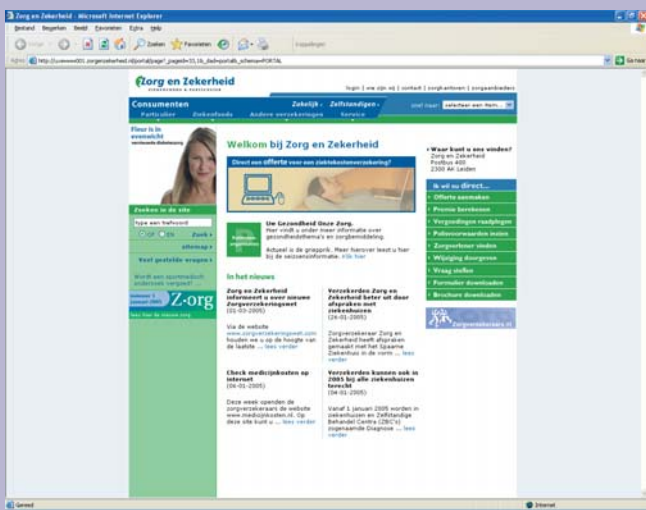
Businesscase: Zorg en Zekerheid

Zorg en Zekerheid is een zorgverzekeraar voor Ziekenfonds en particulier verzekerden. Zorg en Zekerheid vindt dat zorg voor iedereen toegankelijk en betaalbaar moet blijven en werkt daarom zonder winstoogmerk. Bij de dienstverlening staan service en kwaliteit voorop. Zorg en Zekerheid is bewust een regionale zorgverzekeraar, zodat gunstige afspraken met zorgverleners kunnen worden gemaakt. Voor de

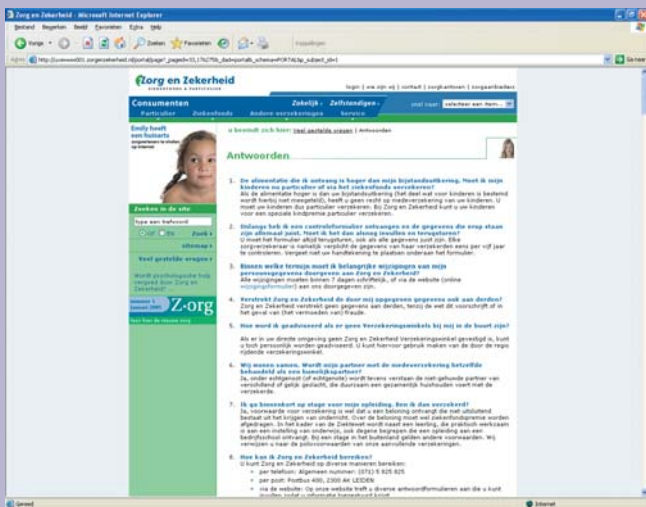
zorgkantoren Amstelland en Meerlanden en Zuid-Holland Noord is Zorg en Zekerheid de concessiehouder. De verzekeraar was voor haar Internetsite op zoek naar een portaalraamwerk waarbinnen selfservice functionaliteit gerealiseerd kan worden en content gemakkelijk onderhoudbaar is door (niet-technische) gebruikers.

Oplossing

De keuze viel op Oracle Portal met als aanvulling de 'Basismodules', een product van Cumquat voor effectief (XML-) content management in Oracle Portal. Met de News Engine en FAQ-modules worden nieuwsberichten en veelgestelde vragen door redacteurs beheerd en gepubliceerd op het portaal. Hierbij is deze content niet gebonden aan een bepaalde pagina, maar worden items (nieuwsberichten en FAQ's) aangemaakt en opgeslagen (als XMLType) in een eigen XML-opslaglaag (gerealiseerd met Oracle XDB) en vervolgens via standaard meegeleverde portlets gepubliceerd. Van nieuwsberichten worden op de hoofdpagina automatisch samenvattingen geplaatst met een "lees verder" link naar het volledige nieuwsbericht. Nieuwsberichten worden na verloop van tijd automatisch naar het nieuwsarchief verplaatst. Vanuit nieuwsberichten, veelgestelde vragen en page content (zie afbeeldingen 1 t/m 4) kan gebruik worden gemaakt van



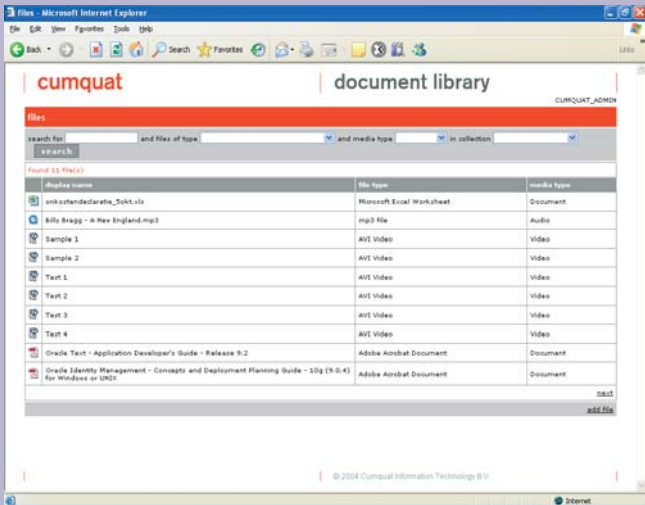
Afbeelding 1. Nieuwsberichten op de site van Zorg en Zekerheid



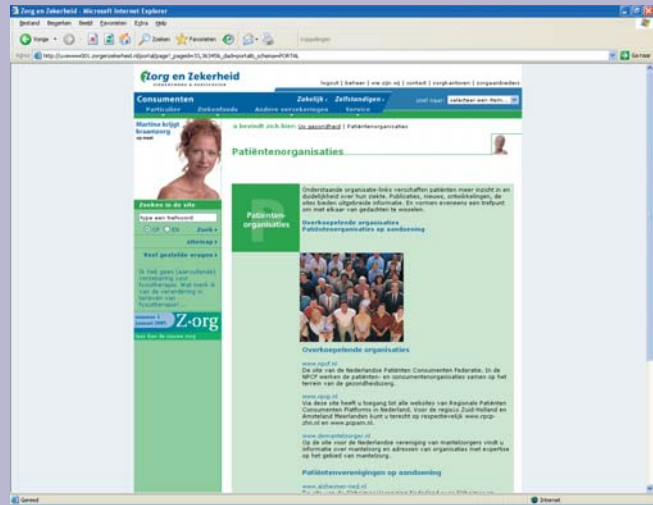
Afbeelding 2. Veelgestelde vragen



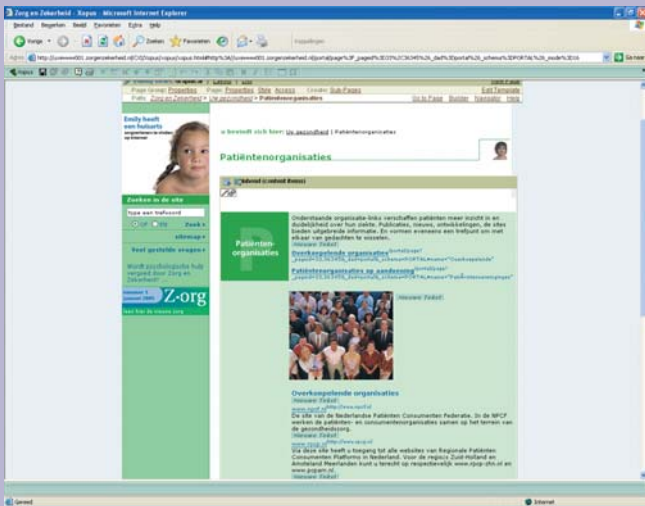
Afbeelding 3. Image library



Afbeelding 4. Document library



Afbeelding 6. Dezelfde pagina in 'view' mode



Afbeelding 5. Pagina in 'edit' mode waarin de content met Xopus op een WYSIWYG manier gewijzigd kan worden

afbeeldingen uit de Image Library en kunnen downloadlinks opgenomen worden naar documenten uit de Document Library. Contentmanagers kunnen in deze modules afbeeldingen en documenten uploaden en beheren.

Voor het editen van nieuwsberichten en FAQ's wordt gebruik gemaakt van Xopus. Deze software wordt ook gebruikt in de XML Items module voor het in-place WYSIWYG editen van page content, zoals te zien is in afbeelding 5 en 6, bij het editen van een speciale themapagina.

Over de content op het portaal kan integraal gezocht worden. De zoekresultaten worden hierbij gegroepeerd getoond

(zoals gevonden nieuwsberichten, veelgestelde vragen en pagina's). Vervolgens kan op een groep ingezoomd worden en de zoekopdracht worden verfijnd. Voor autorisatie en authenticatie wordt via de Autorisatie module gebruik gemaakt van Oracle Internet Directory (OID) en Single Sign-On.

Realisatie

Bij de realisatie van Internetsite voor Zorg en Zekerheid is de volgende standaardaanpak gevolgd:

- ontwikkelen UI templates (voor banners en dergelijke; de 'buitenkant' van de pagina's), op basis van een bestaande vormgeving.
- ontwikkelen menu (en andere navigatie-componenten).
- configuratie van basismodules (opvoeren van redacties, channels, subjects, item types, et cetera).
- maken XML Schema definities en XSLT stylesheets ("item templates") voor de XML-content (nieuwsberichten, FAQ's, page content, et cetera).
- ontwikkelen selfservice functionaliteit en dynamische content, dit wordt gerealiseerd door Java/J2EE-webapplicaties te ontwikkelen die geportletized worden voor integratie met Oracle Portal, maar ook 'los' gebruikt kunnen worden.
- inrichten portaal: aanmaken pagina's, plaatsen menu en andere navigatie-componenten plaatsen portlets, et cetera.

In een tweedaagse workshop is vervolgens aan de niet-technische gebruikers uitgelegd hoe content management gedaan moet worden en een start gemaakt met het 'vullen' van de pagina's van het portaal. De gebruikers hebben na de workshop de site verder afgemaakt en kunnen deze nu eenvoudig zelf onderhouden.

Businesscase: Schuitema

Het beursgenoteerde Schuitema is de supermarktorganisatie achter C1000. Met meer dan 450 supermarkten is C1000 de tweede supermarktformule van Nederland. Het gros van de C1000 winkels wordt gerund door zelfstandige ondernemers. Sinds de overname van A&P in 2001 beschikt Schuitema bovendien over ruim tachtig eigen filialen. Schuitema biedt de C1000 ondernemers en bedrijfsleiders ondersteuning in de vorm van producten en diensten, van de levering van producten tot en met promoties, van loonadministratie tot en met de financiering van een winkelverbouwing, van personeelsbeleid tot en met winkelautomatisering.

Functionele eisen

Schuitema was op zoek naar een oplossing voor het effectief beheren en publiceren van een groot aantal uitgebreide handboeken (voor zaken als marketing, personeel en ICT) op Intranet (800 kantoormedewerkers) en Extranet (1.700 winkeliers van C1000). Een aantal specifieke functionele eisen waren:

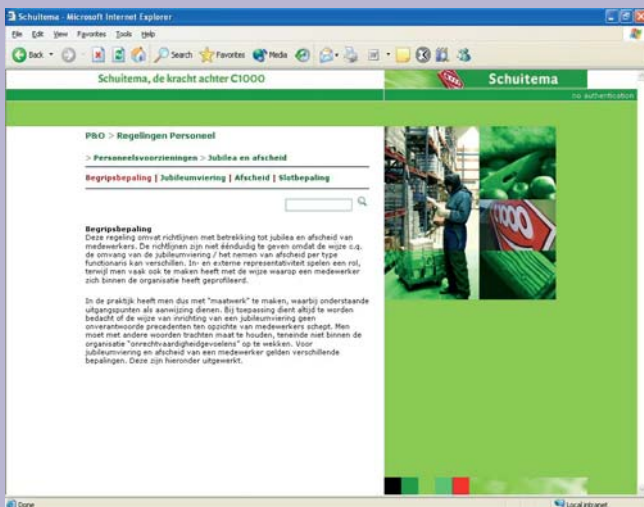
- integratie met Oracle Portal.
- kunnen uitprinten van handboeken via PDF.
- autorisatie: verschillende rollen (schrijven, goedkeuren, publiceren, et cetera), documenten en delen van documenten niet voor iedereen zichtbaar.
- versiebeheer.

- hetzelfde document in verschillende contexten – Intranet en Extranet – kunnen tonen.
- voorkomen van 'broken links' bij verwijzingen tussen handboeken.

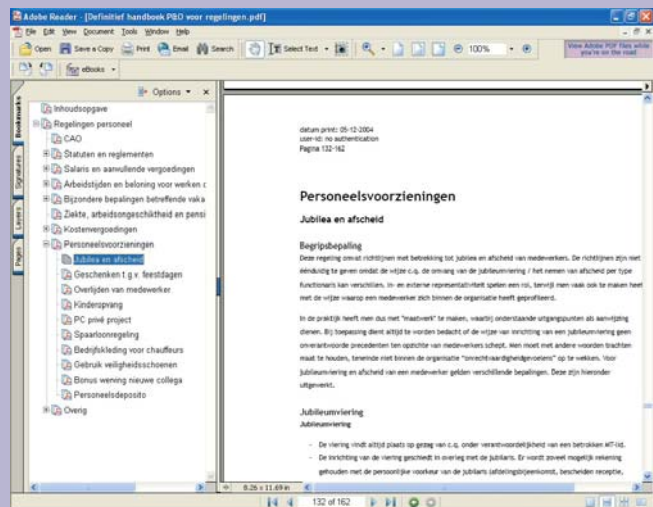
Oplossing

Gekozen is voor de 'Document Editor', een product van Cumquat voor het beheren en publiceren van (XML-)documenten. De oplossing is gebaseerd op Oracle XDB en bestaat uit een J2EE-webapplicatie en een XML-opslaglaag. De integratie met Oracle Portal is gerealiseerd met standaard meegeleverde portlets voor het browsen van een document en het tonen van zoekresultaten.

Een handboek wordt automatisch in hoofdstukken en paragrafen online gepresenteerd, en is gemakkelijk doorbladerbaar met de bijhorende gegenereerde navigatiestructuur. Uiteraard is de online presentatie het resultaat van het toepassen van een XSL-stylesheet op de XML-content. Hierdoor is de lay-out consistent en gemakkelijk aan te passen. Van het XML-document kan met behulp van een transformatie naar XSL Formatting Objects (XSL-FO) via Apache's Formatting Object Processor (FOP) een PDF-document gegenereerd worden.



Afbeelding 7. Deel van document online



Afbeelding 8. Hetzelfde deel in PDF

De content management applicatie maakt gebruik van Oracle Internet Directory (OID) en Single Sign-On (SSO) voor autorisatie en authenticatie. Op documenten en delen van documenten kunnen privileges (rollen) toegekend worden. Op deze manier kunnen delen van handboeken afgeschermd worden en kan er een werkverdeling en workflow gerealiseerd worden. Contentmanagers kunnen gelijktijdig aan een document werken. Versiebeheer stelt de contentmanager in staat om een nieuwe versie van een document of hoofdstuk te maken en hieraan te werken terwijl de huidige gepubliceerde versie voor bezoekers zichtbaar blijft. Dezelfde documenten worden in meerdere contexten (delen van het portaal) getoond. Bij het volgen van links voor verwijzingen naar andere documenten blijft de bezoeker in de juiste context. Bij de verwijzingen tussen handboeken komen géén broken links voor. Verwijzingen worden alleen als link getoond als de gebruiker privileges heeft om het betreffende deel van het andere handboek te mogen zien en als het andere handboek gepubliceerd is. Bij het maken van een nieuwe versie van een handboek worden verwijzingen automatisch verplaatst.

Nieuwsbrieven

Bij Schuitema zijn ook de 'Basismodules' van Cumquat geïmplementeerd (zie kader: 'Businesscase: Zorg en Zekerheid'). Met deze modules worden onder andere de nieuwsberichten



Afbeelding 9. Nieuwsberichten in PDF als nieuwsbrief

voor alle corporate channels, vestigingen en regio's verzorgd. Omdat niet iedere medewerker een werkplek met toegang tot het Intranet heeft, kunnen nieuwsberichten periodiek automatisch in een PDF-nieuwsbrief gezet worden. Deze wordt vervolgens geprint en in de postvakjes gelegd of uitgedeeld. Afbeelding 9 is een mooi voorbeeld van het publiceren van dezelfde content op verschillende manieren (in verschillende verschijningsvormen).

Scheiden van data en presentatie

Bij XML content management wordt de content medianeutraal opgeslagen; data en presentatie worden van elkaar gescheiden. Dit geeft de volgende mogelijkheden:

- dezelfde content kan in verschillende media gepubliceerd worden: browser, mobile devices, PDF, professioneel drukwerk et cetera
- dezelfde content kan in verschillende lay-outs gepubliceerd worden
- de publicatie heeft een consistente structuur en lay-out
- dezelfde content kan op verschillende manieren (in verschillende verschijningsvormen) gepubliceerd worden: bijvoorbeeld productinformatie op het portaal en in een brochure of nieuwsberichten op de website en in een nieuwsbrief

Verrijking

Naast het scheiden van data en presentatie is verrijking een belangrijk voordeel van XML content management. Bij het verrijken van tekst wordt betekenis (semantiek) toegevoegd. Zo kunnen namen van producten, personen of locaties gemarkeerd worden in de tekst. Dit maakt het mogelijk om semantisch in deze informatie te zoeken. Deze markeringen kunnen boven-

dien verwijzen naar andere gegevens. Bij gebruik van de Oracle XML-database is het zelfs mogelijk om een 'echte' foreign key constraint vanuit de XML-data naar een relationele tabel te definiëren. Referenties kunnen op deze manier bewaakt worden (referentiële integriteit wordt afgedwongen). Bovendien kunnen de verwijzingen gebruikt worden om aanvullende informatie te benaderen of data te vervangen. Deze mogelijkheden zijn overigens alleen te benutten bij het gebruik van XMLType met objectrelationele opslag (zie 'XMLType: Native XML datatype in de database', Optimize nr. 4, september 2004). Opslag van XML in een CLOB of BLOB geeft deze mogelijkheden niet.

Kenmerken document centric XML

Document centric XML heeft twee belangrijke kenmerken: het bevat geen opmaak en er wordt veelvuldig gebruik gemaakt van mixed content.

Geen opmaak

Omdat data en presentatie van elkaar gescheiden (moeten) worden bevat de XML geen opmaak. Dat wil zeggen, in het XML-document wordt nergens aangegeven dat tekst in een bepaald font, in bijvoorbeeld een bepaalde fontgrootte of met

een bepaalde kleur getoond moet worden. In het XML-document wordt alleen aangegeven dat een tekst een titel is, waar een paragraaf begint, dat iets een codevoorbeeld is, et cetera. Bij het publiceren van de tekst wordt dan door een XSLT stylesheet en eventueel een cascading stylesheet (CSS) bepaald in welke opmaak, welk font, welke grootte of welke kleuren de tekst getoond wordt.

Sommige tekstverwerkers bieden de mogelijkheid om een document als XML op te slaan. Het resultaat is dan weliswaar een XML-document, maar data en presentatie zijn niet van elkaar gescheiden. Het XML document bevat dan naast de content ook de opmaak; de XML bevat dan talrijke markeringen waarin fonts, fontgrootte, kleuren, marges en dergelijke staan aangegeven. Het niet opnemen van opmaak kan ver doorgevoerd worden. Zo zou je zelfs het gebruik van 'vetgedrukt' ('bold') niet kunnen toestaan, maar in plaats daarvan een markering gebruiken waarmee aangegeven wordt dat er op een bepaald woord of fragment van een zin 'nadruk' ('emphasize') gelegd moet worden. Bij de presentatie kan dan bepaald worden of dit vetgedrukt, in een bepaalde kleur, of genegeerd moet worden. Nog beter zou zelfs zijn, om de vraag te stellen waarom nadruk gelegd moet worden. Is het de naam van een product? Is het een nieuwe term (die geïntroduceerd wordt)? De betreffende woorden en fragmenten zouden dan ook zodanig gemarkeerd moeten worden zodat de tekst in feite verrijkt wordt. Terughoudendheid is hierbij wel op zijn plaats. Er zal altijd gekeken moeten worden naar de toegevoegde waarde en haalbaarheid hiervan.

Mixed content

In document centric XML is vrijwel altijd mixed content terug te vinden. Een element in een XML document heeft mixed content wanneer dit element tekst bevat met daarin op willekeurige plekken gemarkeerde data, veelal verrijking. Hieronder is een voorbeeld te zien.

```
<p>
  Lorem ipsum <term>dolor</term> sit amet, consectetur adipiscing
  elit.
  Vivamus suscipit magna vel lectus. Sed eget urna. Maecenas lobortis
  porttitor lacus. Curabitur nisl mauris, <product id=123>interdum</
  product>
  sit amet, lobortis ut, pulvinar a, odio. Vestibulum nulla. Sed id
  diam vel
  <term>felis</term> congue dapibus. Ut ultrices semper magna.
  Suspendisse
  purus.
</p>
```

In dit paragraafelement (<p>) zijn twee termen gemarkeerd en één product. Bij het product wordt via een attribuut verwezen naar de productidentificatie.

XML content management oplossingen

In de kaders worden twee concrete business cases besproken. Hieronder worden de onderdelen van XML content management oplossingen in algemene zin beschreven.

Database

De basis voor een XML content management oplossing is een database waarin zowel XML als relationele data opgeslagen kunnen worden. De opslag van XML spreekt voor zich, maar ook belangrijk zijn de relationele data voor de administratieve en procesmatige aspecten van de oplossing. De Oracle database beschikt vanaf 9i release 2 over de XML database (XDB) en is daarmee bij uitstek geschikt als database voor XML content management oplossingen. De mogelijkheden van Oracle XDB zijn uitgebreid beschreven in de eerder verschenen reeks van vier artikelen in dit blad.

XML editor

Verder is het noodzakelijk om over een gebruiksvriendelijke XML editor te beschikken om document centric XML te kunnen editen ('authoren'). Op een groot aantal projecten heeft Cumquat succesvol de Xopus XML editor ingezet. Xopus is ook onderdeel van een tweetal producten van Cumquat. Later in dit artikel wordt verder ingegaan op Xopus.

Applicatie

Hoewel de Oracle XML database en Xopus onmisbaar zijn in de door Cumquat gerealiseerde XML content management oplossingen is er geen oplossing zonder applicatie (zie de kaders met de business cases over *Schuitema en Zorg en Zekerheid*). Het is namelijk de applicatie die de content management functionaliteit inpast in allerlei bedrijfsprocessen, zoals het beheren van polisvoorwaarden, change management, het opstellen van offertes of het samenstellen van studiegidsen. Het gaat hier om meer dan de XML content alleen. Cumquat heeft een product ontwikkeld waarmee zowel (XML-) documenten op een directe manier beheerd en gepubliceerd kunnen worden, en waarmee ook deze functionaliteit ingepast kan worden in applicaties voor bedrijfsprocessen waarbinnen content management een rol speelt. Daarnaast biedt Cumquat XML content management aanvullingen op Oracle Portal.

Printing (on demand)

Naast online publicatie (in HTML) is het meestal ook wenselijk om content 'in print' te kunnen publiceren. Dit kan via PDF, gegenereerd met XSL Formatting Objects (XSL-FO) en Apache's Formatting Objects Processor (FOP), of door de XML-content te importeren in desktop publishing (DTP) tools zoals Adobe FrameMaker, QuarkXPress of Adobe InDesign. In de producten van Cumquat worden deze mogelijkheden standaard geboden.

Dit stelt bedrijven en organisaties in staat om zelf professioneel drukwerk voor te bereiden (brochures, handleidingen, reglementen, polisvoorwaarden, studiegidsen et cetera), zonder hiervoor afhankelijk te zijn van tussenpartijen. Via printportalen is printing on demand mogelijk. Hierbij worden via een portaal, wanneer er behoefte aan is, elektronisch (gepersonaliseerde) printbestanden aangeleverd die vervolgens door een drukker digitaal of in offset gedrukt worden en afgeleverd worden. Cumquat werkt hiervoor samen met Relate4U (www.relate4u.com) en Paro (www.paro.nl).

Xopus

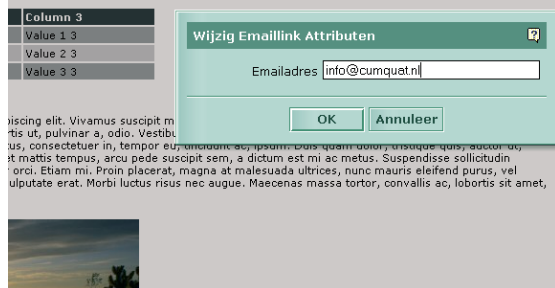
Xopus, een product van Q42 uit Den Haag (www.q42.nl), is een XML-editor die 'draait' in de browser (Internet Explorer). Het is een zogenaamde validating XML-editor, die afdwingt dat de XML voldoet aan een opgegeven XML Schema definitie. deze validatie vindt plaats tijdens het editen, in tegenstelling tot andere XML-editors die pas valideren bij het bewaren van het XML-document, waarna eventuele fouten door de gebruiker gezocht en aangepast moeten worden. Xopus is bovendien een WYSIWYG XML-editor. Editen vindt plaats in een HTML-editingview in de browser, het resultaat van het toepassen van een XSLT-stylesheet op het XML-document.

XML editing met Xopus

Hieronder is een typedefinitie te zien uit een XML Schema definitie voor een mixed content element, kenmerkend voor document centric XML.

```
<xs:complexType name="pType" mixed="true">
  <xs:choice minOccurs="0" maxOccurs="unbounded">
    <xs:element name="emailLink" type="emailLinkType"/>
    <xs:element name="downloadLink" type="downloadLinkType"/>
    <xs:element name="link" type="linkType"/>
    <xs:element name="b" type="xs:string"/>
    <xs:element name="i" type="xs:string"/>
  </xs:choice>
</xs:complexType>
```

us, consectetur in, tempor eu, tincidunt ac, ipsum. Duis quam dolor, tristique quis, auctor ut, et mattis tempus, arcu pede suscipit sem, a dictum est mi ac metus. Suspendisse sollicitudin orci. Etiam mi. Proin placerat <mailto:info@cumquat.nl>, magna at malesuada ultrices, nunc mauris gula. Proin mattis vulputate erat. Morbi luctus risus nec augue. Maecenas massa tortor, convallis

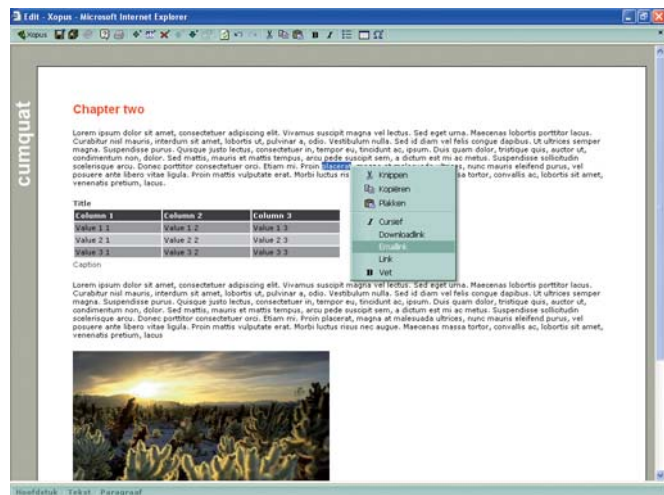


Afbeelding 10. Het markeren van een emailink in Xopus

Binnen een paragraaf kunnen e-mail-, download- en gewone links aangegeven worden en kunnen stukjes tekst bold of italic gemaakt worden. Bij het editen in Xopus worden op basis van de XML Schema definitie bepaalde opties geboden in het rechtermuisknopmenu, wanneer de gebruiker in een paragraaf tekst selecteert.

```
<p>
  Lorem ipsum dolor sit amet
  ...
  Etiam mi. Proin
  <emailLink address="info@cumquat.nl">placerat</emailLink>
  , magna at malesuada ultrices
  ...
  venenatis pretium, lacus.
</p>
```

Xopus biedt ook de mogelijkheid om een reeks van XSLT-transformaties uit te voeren (pipelined transformations) en ondersteunt XIncludes.



Afbeelding 11. Xopus toont de markeringsopties in een paragraaf op basis van XML Schema definitie

Erwin Groenendal

is senior solutions architect bij Cumquat Information Technology (www.cumquat.nl) en kan bereikt worden via erwin.groenendal@cumquat.nl.