

Verwachtingsvol kijkt OVAM, Openbare Afvalstoffenmaatschappij

voor het Vlaamse gewest, al enige tijd naar GML als

uitwisselingsstandaard: een gratis, volsterkt niet-

leveranciersgebonden formaat mét datavalideringsmogelijkheden.

Maar is het een reële optie? Daarom greep ik de volgende etappe in

de Vi MATRIX Estafette aan om eens te praten met Wilko Quak, GML-

expert en onderzoeker aan de Technische Universiteit Delft.

'Zoek alle objecten binnen dit rechthoekje' zal worden vertaald in 'Zoek alle tekstobjecten waarvan het veld X-coördinaat een tekst heeft groter dan...'. Dat is heel omslachtig", legt Quak uit.

De conversie is erop gericht dat van het tekstachtige formaat weer een binair formaat wordt gemaakt, zodat er gemakkelijk kan worden. Dat is volgens Quak

'GML wordt binnen vijf jaar hét mondiale GIS-uitwisselingsformaat'

Drs. Wilko Quak, als onderzoeker toegevoegd aan de sectie GIS-technologie van het onderzoeksinstituut OTB van de TU Delft, zal metertijd promoveren op spatial databases en performance: hoe krijg je ruimtelijke gegevens zo snel mogelijk in en uit databases? Geo-analyses zullen immers steeds meer met behulp van 'gewone' databases worden gedaan in plaats van in een GIS. Niet dat dat morgen al te gebeuren staat. "Er valt nog wel wat te verbeteren aan de modellering van ruimtelijke gegevens in deze producten."

Toch willen steeds meer organisaties hun geodata niet meer in een leveranciersafhankelijk GIS-formaat opslaan. "Het opslaggedeelte wordt al afgesnoept van de GIS-verkopers", weet Quak. De meeste databases die geschikt zijn voor opslag van ruimtelijke gegevens hebben een uitbreiding op de standaard query language volgens de standaard specificaties van het OpenGIS Consortium. Voor de uitwisseling van ruimtelijke gegevens zullen de databaseleveranciers en de GIS-leveranciers óók naar OGC-specificaties moeten kijken en wel naar die voor Geography Markup Language (GML). Quak: "Het ziet ernaar uit dat GML dé uitwisselingsstandaard wordt; we zitten er als onderzoeksgroep bovenop."

GML is de uitbreiding voor ruimtelijke

data op XML, de inmiddels breed geaccepteerde uitwisselingsstandaard voor allerlei soorten gegevens via internet; veel tools zijn gratis via het internet te bekomen. In GML kun je eigenschappen kwijt van punten, lijnen en vlakken (zie kader). Bijna niemand verdient eraan, er is niet veel op tegen en het werkt uitstekend – mijn liefje wat wil je nog meer?

Gebruik

Het gebruik van GML begint nu echt op gang te komen, observeert men in Delft. Vooral in het Verenigd Koninkrijk gaat het hard. De Ordnance Survey (het Britse nationale geografisch instituut, inclusief kadaster) heeft van de ene op de andere dag besloten alles in GML uit te leveren, dus dat heeft een hele keten van volgers op gang gebracht. In Nederland overdenkt de Topografische Dienst hoe hun gegevens er in GML uit zouden moeten zien, wellicht het Belgische NGI ook.

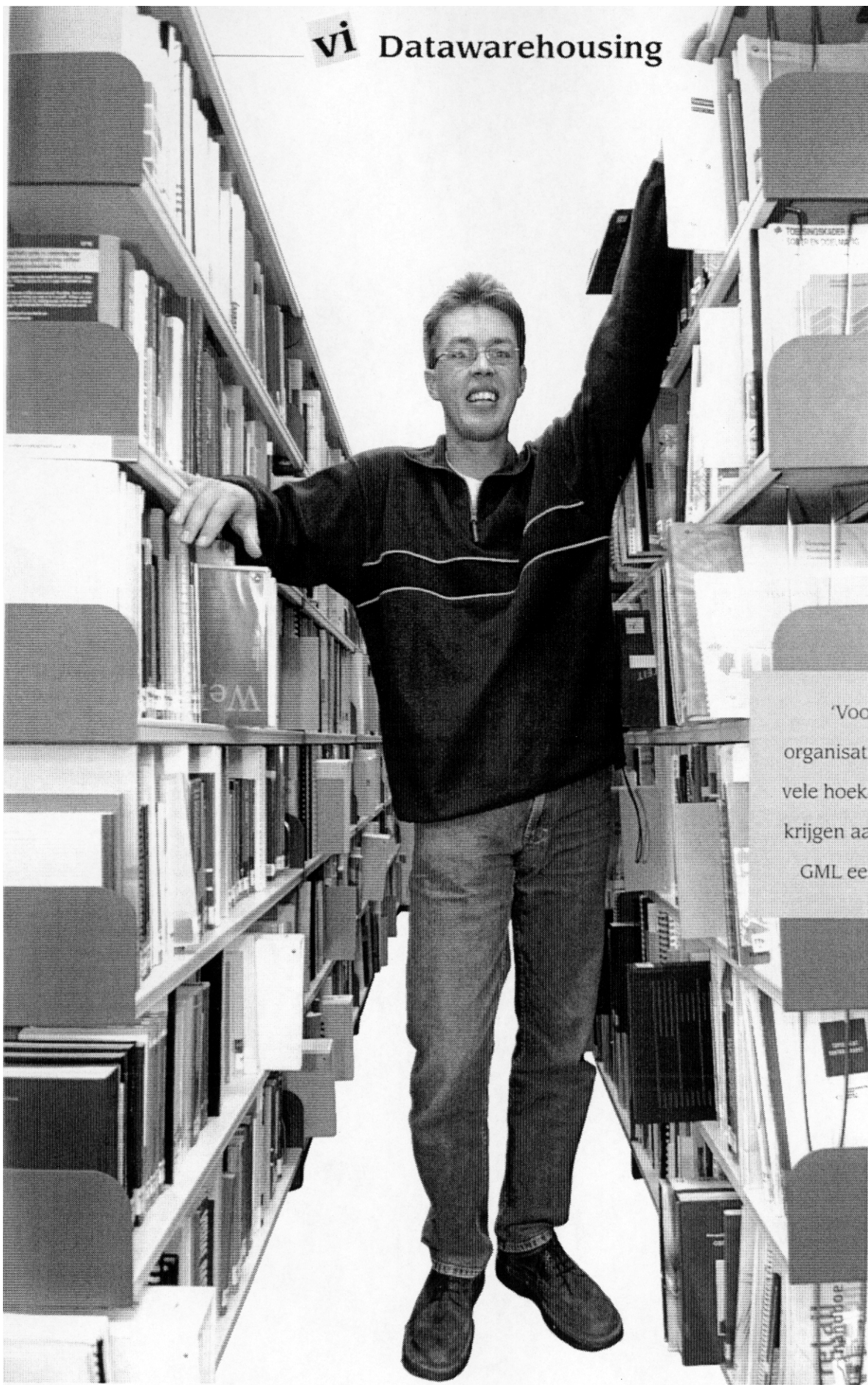
Met GML verstuurde en ontvangen data worden geconverteerd naar het bestandsformaat van de database waarin ze worden opgeslagen. GIS-analyses of -bewerkingen direct uitvoeren op een GML-document is namelijk niet handig. "Snelle verwerking van grote bestanden gaat helemaal mis als je dat in GML zou doen. In GML wordt een bestand 20 keer zo groot en 'tekstachtig'. De opdracht

geen probleem: "Een als XML- of GML-object ontvangen bestand, wordt met een paar geautomatiseerde handelingen opgeslagen als een polygoonobjectbestand. Dat neemt niet weg, dat nog lang niet alle GIS-pakketten dat kunnen."

De sectie GIS-technologie heeft samen met de KvAG al twee jaar achtereenvolgens een zogenaamde GML-estafette gehouden. Op een plenaire bijeenkomst moet een GML-document met een standaard pakket worden ingelezen, op een bepaalde manier gewijzigd, weggeschreven en dan doorgegeven aan de volgende GIS-leverancier. "Het viel ons de laatste keer erg mee. Er waren veel grote leveranciers en ze slaagden er allemaal in om het GML-bestand te lezen, de een wat makkelijker dan de ander."

Een file wordt in feite in drieën verstuurd. Behalve het GML-document, is er als eerste een computerleesbaar document dat meldt hoe de data in het GML-document gestructureerd zijn. Dit is de schemadefinitie. Als ESRI-software een schemadefinitie binnenkrijgt met twee geometrieën bij één object, zal het twee kaartlagen aanmaken. GML staat echter toe, dat er meer geometrieën in één tabel staan. Een weg kan zowel een lijn zijn (voor navigatiedoeleinden bijvoorbeeld) als een polygoon (voor berekening oppervlakte). In ESRI krijg je een shapefi-

vi Datawarehousing



Wilko Quak:

"Het grootste risico dat de opmars van open uitwisseling en integratie van geodata in de weg kan staan, is dat de GIS-vendors het te ingewikkeld vinden en zelf met een nieuwe GML komen."

'Vooral voor organisaties die vanuit vele hoeken bestanden krijgen aangeleverd, is GML een uitkomst'

le met de hartlijn van de wegen en één met de contour. "GIS-vendors die zelf geen eigen uitwisselingsformaat hebben staan meer open voor uitwisselingsformaten van anderen", is de indruk van Wilko Quak.

Het derde element in een GML-zending – naast GML-data en schemadefinitie – is een document waarin staat hoe het GML-document te vertalen is in kleuren op het scherm. In GML 3 zit een standaard om dat naar eigen wens te specificeren.

In een GML-document
zijn heel eenvoudig
geometrie plus
administratieve data
vast te leggen

Profielen

"GML is een belangrijke aanjager voor de integratie tussen administratieve en geografische data", bevestigt Quak de hoop van OVAM. "In een GML-document zijn heel eenvoudig geometrie plus administratieve zaken vast te leggen. Ook in een shapefile kunnen administratieve gegevens worden vastgelegd, dat is waar,

maar bij GML is het sterker geïntegreerd en kan het uitgebreider. Bovendien kunnen de data telkens gevalideerd worden, wat een belangrijke meerwaarde is van GML", vindt hij. Er is immers een schemadefinitie bijgeplaatst, waardoor de computer controleert of de data die binnenkomen voldoen aan de gewenste structuur.

"In Groot-Brittannië konden de GIS-vendors eerst alleen het GML lezen dat van de Ordnance Survey afkomstig was", vertelt Quak, "want daarvan hadden ze de schemadefinitie uitgeplozen. Maar het gaat steeds beter en verschillende GIS-leveranciers kunnen nu ook willekeurige GML's lezen."

GML is namelijk een bouwdoos. Het OGC heeft bovenop XML – met de X van extensible – de basis gelegd voor GML, maar dat moet per branche of organisatie nader worden uitgewerkt. Die uitbreiding geeft aan hoe *jouw* documenten er uit komen te zien. Voor de OVAM zouden we bijvoorbeeld voor bodemonderzoeksbureaus vastleggen dat in door hun aan-geleverde gegevens zeker één polygoon moet zitten en die en die attribuu-ten. Voldoet een bestand daar niet aan, dan blijkt dat uit de schemadefinitie en zegt

ons GIS-programma na validatie: 'Sorry, onvolledig bestand'."

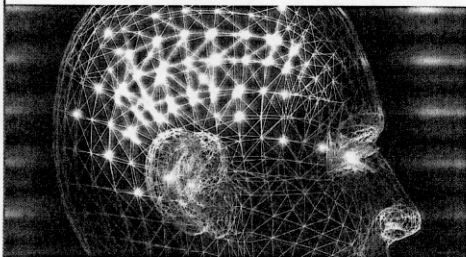
Omdat een GML-document zo'n vaste structuur heeft, is het dus ook makkelijk om externe bestanden te integreren in een database. "Vooral voor die organisaties die verschillende GIS-producten door elkaar gebruiken of vanuit vele hoeken krijgen aangeleverd, is GML als niet-vendorgebonden standaard een uitkomst", beaamt de onderzoeker. En binnen de GML-standaard zullen dus op Europees niveau allemaal profielen worden ontwikkeld, voor kadasterdata, milieu, waterkwaliteit, et cetera.

Opmars

Dat GML snel gemeengoed wordt, daar twijfelt Wilko Quak niet aan. "Over vijf jaar is alles standaard in GML", klinkt het stellig, gevolgd door een schuchter "maar ik ben te jong om een visionair te zijn." Hij memoreert evenwel dat een overheid min of meer verplicht is open standaarden te gebruiken. Achter GML zit niet één verkoper, dus de belangrijkste datale-ve-ranciers zullen beslissen om in dat for-maat te gaan uitleveren.

Bovendien, op het gebied waar veel ruimtelijke data worden uitgewisseld, de

(Ingezonden mededeling)



Van oudsher probeert de mens wat te krijgen op de wereld om zich heen. Waar de eerste inspanningen zich vooral richtten op het in kaart brengen van de omgeving, gaat het nu meer om het optimaliseren van de reeds aanwezige informatie en het intelligent omgaan met gegevens. Overal waar plannen worden ontwikkeld, werken worden uitgevoerd en

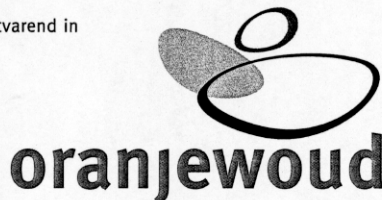
gebieden worden beheerd, neemt de vraag naar gedetailleerde en veelzijdig bruikbare informatie toe. **Oranjewoud Object & Informatie**

Dimensie geven aan informatie

voorziet voortvarend in

de groeiende behoefte aan een efficiënte inwinning en verwerking van gegevens.

Dimensie in informatievoorziening? Kijk op www.oranjewoud.nl.



Heerenveen • Deventer • Almere • Rijswijk • Capelle a/d IJssel • Goes • Oosterhout • Geleen

XML en GML

Webservices zijn gestoeld op XML, Extensible Markup Language. Om ook geografische data hierin te 'verpakken', dusdanig dat deze met webservies toegankelijk gemaakt kunnen worden, is Geography Markup Language (GML) ontwikkeld. Deze taal ontstond in 2000 op initiatief van het Open GIS Consortium. Het OGC definieert het zelf als: 'een op XML gebaseerde codering voor het transport en de opslag van geografische informatie, die zowel de ruimtelijke als niet-ruimtelijke eigenschappen van geografische objecten beslaat'.

Van origine is het bedoeld als uniform uitwisselingsformaat tussen diverse GIS-pakketten, waarmee het 'interfaceprobleem' opgelost zou moeten worden. Nog steeds is dit een belangrijk doel, maar de aansluiting op de web-servicetechnologie is een goede tweede.

In GML worden geometrieën beschreven door 'features': punten, lijnen en vlakken. Ook complexere objecten zijn te representeren, doordat features weer uit andere features kunnen bestaan. Ook niet-ruimtelijke eigenschappen als administratieve data over objecten en meta-informatie (coördinaatsystemen of datum van creatie) kunnen zonder probleem in het GML-document worden opgenomen, waardoor een sterke integratie ontstaat tussen administratieve en geografische gegevens.

Schema

Het document voldoet aan een altijd meegeleverd XML-schema, waarin de structuur van een GML-document staat: welke ruimtelijke objecten kunnen op welke manier beschreven worden? De structuur van een GML-document is grofweg hetzelfde als die van een XML-document. Met 'tags' - uitdrukkingen tussen haken ('<' en '>') - wordt een document geconstrueerd. Een browser interpreteert de code.

Iedereen kan zijn of haar eigen tags definiëren, zodat deze toepasselijk zijn in die bepaalde context (zie bijvoorbeeld figuur 1). De toegestane tags moeten wel vastgelegd worden in het XML-schema. Dit schema geeft aan welke tags gebruikt mogen worden in het XML-document en kunnen regels opleggen aan de (tekst)waarden tussen de tags. Dit schema kan gebruikt worden om een XML-document te valideren. Vaak worden schema's sectorbreed opgesteld om uitwisseling van informatie te vergemakkelijken.

Geography Markup Language (GML) werkt volgens hetzelfde principe als XML. De representatie van geografische data in een vorm van XML is echter niet eenvoudig. Daarom heeft het OGC twee GML-schema's opgesteld: feature.xsd en geometry.xsd. Hierin staat precies welke tags zijn toegestaan en hoe deze ingevuld dienen te worden. Daarnaast kunnen in een schema ook de administratieve gegevens gemodelleerd worden. Ook moeten eigen profielen in een extra definitieschema worden toegevoegd, omdat bedrijfs- of branche eigen elementen niet standaard in het feature.xsd of geometry.xsd schema aanwezig zijn.

Een (GIS-)applicatie leest daarna het GML-document met het bijbehorende schema en 'weet' precies wat alle beschrijvingen betekenen. Dit houdt in, dat een dergelijke applicatie niet meer hoeft te wachten totdat een converterende interface aanwezig is. Daarnaast is het nu mogelijk om een document te valideren op basis van het definitieschema: voldoet het hier niet aan, dan is er in de creatie van het document of in het datatransport iets fout gegaan.

De weergave van het geheel is niet gestandaardiseerd, maar kan worden geregeld aan de hand van zelf te definiëren 'style sheets'. Daarnaast heeft dit het voordeel dat bestanden platform- en systeemafhankelijk zijn. Sterker nog, door de tekstvorm is het document te openen in simpele programma's als 'notepad'.

Er zijn weinig nadelen aan GML verbonden. Het is niet erg compact; door de structuur met 'tags' wordt veel informatie herhaald. Dit heeft een negatief effect op de netwerkbelasting en de benodigde rekenkracht van afzonderlijke systemen. De bestanden laten zich door middel van compressie qua grootte wel indikken, maar dit heeft een negatief effect op de performance van de oplossing. Een ander nadeel was tot voor kort dat door het tekstformaat geen rasterdata konden worden verwerkt. Deze moesten eerst gevectoriseerd worden. In versie GML 3.0 is dat probleem verholpen. (FvB)

ruimtelijke ordening, gebeurt dat in Nederland nu via een genormaliseerd formaat NEN 1878, maar die gaat op de schop. "De vernieuwde versie daarvan zal volgend jaar GML zijn. GML 3 is inmiddels zo uitgebreid, dat men er behalve geometrie ook temporele aspecten en topologie in kwijt kan. Dat biedt GIS vendors dus ook ruimte tot differentiatie: het ene pakket ondersteunt bijvoorbeeld alle definities voor 3D objecten en een ander niet."

Er zijn een paar kleine bedrijven die in het gat gesprongen zijn en als eerste met een GML-conversieprogramma zijn gekomen: van ArcView naar GML of van GML naar een willekeurige database. Vanuit de UK opereert Snowflake Software. "Je schrijft een GML-bestand en met een druk op de knop zit het in Oracle." Vanuit

```
<tdn:WegDeel fid="TOP10.402814">
  <tdn:top10_id>2400001</tdn:top10_id>
  <tdn:bronRef xlink:type="simple"
xlink:href="metadata.xml#TOP10.9000017"/>
  <tdn:object_begindatum>2001-12-
17T13:24:10+02:00</tdn:object_begindatum>
  <gml:geometryProperty>
    <gml:LineString srsName="EPSG:28992">
      <gml:coordinates>106045.396,449000 105965.228,448967.705 ...
    </gml:coordinates>
    </gml:LineString>
  </gml:geometryProperty>
  <tdn:wegtype>Autosnelweg</tdn:wegtype>
  <tdn:wegnummer>A12</tdn:wegnummer>
  <tdn:wegnummer>E25</tdn:wegnummer>
</tdn:WegDeel>
```

Figuur 1: in een stukje GML-bestand van de Topografische Dienst Nederland staan zowel administratieve gegevens (voorafgegaan door 'tdn') als geografische ('gml').

de Benelux zijn Ionic en eXQte (Safe Software) actief. Maar op termijn zal het dus in elk standaard GIS-pakket zitten en in alle belangrijke databases.

Het grootste risico dat de opmars van open uitwisseling en integratie van geo-data in de weg kan staan, is dat de GIS-vendors het te ingewikkeld vinden en zelf met een nieuwe GML komen. "Dat mondt uit in een hele serie, waardoor je door de bomen het bos niet meer ziet, ook al zijn ze onderling dan nog best te converteren", waarschuwt Quak. "Het is de taak van het OGC om er voor te zorgen dat iedereen zijn best blijft doen om méé te stappen."

ERWIN SEVENS
GIS-COÖRDINATOR OVAM