

Mutatie-uitwisseling op basis van standaarden

Voor veel toepassingen van geo-informatie is het van groot belang dat de gebruikte gegevens actueel zijn. In dit artikel wordt een overzicht gegeven van de mogelijkheden voor het actualiseren van gegevens vanuit een centrale bron, gegeven de huidige stand van de techniek. De aandacht gaat speciaal uit naar op standaarden gebaseerde methoden en niet naar specifieke oplossingen. Het onderzoek waarop dit artikel is gebaseerd, is uitgevoerd in het kader van RGI-project 116, Verkenning van innovaties bij geo-standaarden voor NGII.

Om te werken met actuele informatie zijn er in de basis twee mogelijkheden: de gegevens worden voor iedere bevraging bij de bronhouder opgehaald (data bij de bron) of er wordt gewerkt met een lokale kopie van de gegevens binnen de eigen organisatie.

Het eerste principe is vanuit het oogpunt van de kwaliteit van gegevens het beste. Omdat de gegevens direct bij de bron worden geraadpleegd, kan er in principe geen discussie zijn over de actualiteit en juistheid van de gegevens. Dat wil niet zeggen dat de gegevens foutloos zijn, maar de bronhouder heeft kennelijk op dat moment geen betere gegevens beschikbaar. Deze opzet stelt een aantal eisen aan de informatie-infrastructuur binnen en tussen de betrokken organisaties: de gegevens bij de bronhouder moeten altijd toegankelijk zijn, de bronhouder moet voldoende capaciteit hebben om alle bevestigingen van alle afnemers af te kunnen handelen en de afnemer moet altijd verbinding kunnen hebben met de bronhouder. Met de huidige stand van techniek kan het al lastig zijn om voor grote en populaire datasets aan al de hiervoor genoemde eisen te voldoen. In het verleden was het probleem bij gebrek aan snelle computernetwerken nog groter. Daarom zijn systemen opgezet waarbij de gebruikers lokaal een kopie hebben van de gegevens en periodiek nieuwe gegevens krijgen van de leverancier.

Lokale kopieën actueel houden

Voor het bijhouden van lokale kopieën moeten gegevens worden uitgewisseld tussen leveranciers en afnemers. Deze gegevens werden in het verleden bijvoorbeeld aangeleverd op magnetische tapes en later ook op cd's of dvd's. Met het beschikbaar komen van snelle netwerkverbindingen kan de levering van fysieke media vervangen zijn door leveringen via bijvoorbeeld e-mail of FTP. De verwerking van de gegevens is dan echter vaak gelijk gebleven.

Voor het regelmatig bijwerken van de gegevens zijn in de basis twee opties: bij iedere levering wordt een volledige nieuwe set gegevens geleverd, of alleen de wijzigingen worden uitgewisseld. Het uitwisselen van volledige sets heeft als belangrijke voordelen dat het heel eenvoudig is om alle oude gegevens te vervangen door alle nieuwe en dat het duidelijk is wat de actualiteit van de data



Maarten Vermeij,
onderzoeker,
Onderzoeks-
instituut OTB,
Technische
Universiteit Delft

is. Een nadeel is dat als de hoeveelheid nieuwe gegevens relatief klein is ten opzichte van de volledige dataset, er erg veel gegevens worden uitgewisseld die al bij de afnemer bekend zijn. Een oplossing hiervoor is om gebruik te maken van incrementele updates. Dit houdt in dat alleen de wijzigingen in de gegevens aan de afnemer worden doorgegeven. Gegevens die niet veranderen, worden niet opnieuw verstuurd, waardoor de hoeveelheid te versturen gegevens klein kan blijven. Hier kleven ook weer een aantal voor- en nadelen aan. Het belangrijkste voordeel van het uitwisselen van alleen de wijzigingen is dat de omvang van de uitgewisselde gegevens kleiner is. Hierdoor kan de uitwisseling sneller plaatsvinden en is ook de verwerking van de wijzigingen mogelijk sneller. Een ander voordeel is dat door het specifiek aanleveren van wijzigingen, de afnemer expliciet op de hoogte wordt gebracht van wijzigingen en indien nodig daar direct aandacht aan kan besteden in interne werkprocessen. Nadelen van deze methode zijn dat het bijhouden en correct verwerken van mutaties complex kan zijn en dat de afnemer strikt alle mutaties op volgorde moet verwerken om een goede gegevensset te hebben.

Er zijn ook tussenvormen denkbaar. Zo is het ook mogelijk het interessegebied op te splitsen in kleine delen en de afnemer een complete beschrijving te geven van de delen waarbinnen wijzigingen hebben plaatsgevonden. Dit geeft een relatief eenvoudige verwerking bij de afnemer en houdt toch de hoeveelheid gegevens die verstuurd moeten worden beperkt.

Eerder werd aangegeven dat vanuit kwaliteitsoogpunt, data bij de bron het beste wordt geacht. Een opmerkelijk gevolg van het data-bij-de-bron-principe is dat het in feite inhoudt dat voor iedere bevraging een complete dataset van de bronhouder naar de afnemer wordt verstuurd. In veel gevallen zal de hoeveelheid gegevens echter wel beperkt zijn omdat het alleen de gegevens betreft waar de afnemer direct belangstelling voor heeft. Echter, als de afnemer voor zijn werk een grote hoeveelheid data nodig heeft, bijvoorbeeld voor een analyse over een groot gebied, kan dit een zware belasting zijn voor

*Het uitwisselen van grote hoeveelheden gegevens was lastig zonder goede netwerkverbindingen.
(Afbeelding van www.skierpage.com.)*



zowel de afnemer als de aanbieder. Dit is een reden waarom te verwachten valt dat voorlopig het bijhouden van lokale kopieën van datasets in de praktijk nog een veel gebruikte situatie zal zijn en dat het daarom nuttig is om te kijken naar goede methoden om deze kopieën actueel te houden.

Mutaties kunnen zich altijd voordoen en het is dan ook van belang dat afnemers regelmatig nieuwe gegevens krijgen. Wat deze regelmaat is, kan sterk verschillen. Voor verschillende geo-informatieproducten bestaan abonnementsvormen waarbij bijvoorbeeld iedere maand nieuwe gegevens geleverd worden door de bronhouder, zoals kadastrale gegevens. Een product als TOP-10Vector werd in het verleden slechts eens in de vier jaar ververst. In het algemeen zullen kleinschalige gegevens minder snel veranderen dan grootschalige. Vooral bij gegevens die afkomstig zijn van sensoren is een hoge mutatiefrequentie mogelijk omdat daarbij geen menselijke handelingen nodig zijn. Geo-informatievoorbeelden hiervan zijn de beelden van de populaire buienradar-website (www.buienradar.nl) of de actuele verkeerssituatie (www.anwb.nl).

De manier waarop de levering van mutaties het beste gebeurt, hangt af van de frequentie en de omvang van de gegevens en ook van de mogelijkheden die beschikbare technieken geven. Binnen het onderzoek is vooral gekeken naar het uitwisselen van gegevens tussen systemen van verschillende professionele organisaties. Deze zullen in de praktijk netwerken hebben die vaak verbonden zijn op een snelle en betrouwbare manier.

Een podcast voor geo-informatie

Steeds meer radio-uitzendingen bieden tegenwoordig ook een podcast aan via een website. Met behulp van speciale software kunnen luisteraars automatisch nieuwe uitzendingen downloaden - meestal in de vorm van een mp3-bestand - en beluisteren op het moment dat het hen uitkomt. De techniek die gebruikt wordt voor podcasts is RSS. Hierbij is op de website van de aanbieder een XML-bestand beschikbaar dat een overzicht geeft van de laatste nieuwe bestanden die aangeboden worden. Dit bestand heet een RSS-feed [RSS Board, 2007]. De software van de bezoeker leest dit bestand, kan automatisch zien welke



bestanden hij nog niet eerder heeft gedownload en haalt deze nieuwe bestanden op. Door regelmatig de RSS-feed opnieuw op te halen en te doorzoeken op nieuwe bestanden, blijft de afnemer op de hoogte van nieuwe gegevens bij de aanbieder.

Hoewel deze techniek populair is voor het aanbieden van geluidsbestanden, is dat zeker niet de enige mogelijkheid. Op een soortgelijke wijze zouden ook aanbieders van geo-informatie nieuwe gegevens beschikbaar kunnen stellen. Zo zouden bijvoorbeeld bestanden met mutaties opgenomen kunnen worden in een RSS-feed. Op basis van deze techniek kan vrij eenvoudig een systeem van uitwisseling van mutaties worden opgezet. De eisen aan de afnemer zijn niet groot, aangezien deze niets anders hoeft te doen dan bestanden ophalen bij de aanbieder. Er bestaat inmiddels ook een specificatie voor het opnemen van geografische informatie in een RSS-feed: GeoRSS [Open Geospatial Consortium Inc., 2006]. Deze specificatie beschrijft hoe de elementen in een RSS-feed van geografische informatie, zoals bijvoorbeeld een locatie, kunnen worden voorzien.

Het is echter de vraag of dit systeem als een serieuze oplossing in een professionele omgeving ingezet kan worden. Er zitten namelijk geen voorzieningen in die de betrouwbaarheid van de uitwisseling waarborgen.

Niet halen, maar brengen: WFS-T

De hiervoor geschetste situatie met RSS gaat er vanuit dat de afnemer bij de aanbieder gaat kijken of er nog

nieuws is. Dit verhaal is ook om te keren: de aanbieder kan actief de afnemer benaderen om veranderingen door te geven. Binnen de verzameling van OpenGIS-specificaties is de Transactional Web Features Service (WFS-T) er een die het mogelijk maakt om wijzigingen door te geven [Open Geospatial Consortium Inc., 2005]. Een client-toepassing kan hierbij wijzigingen doorgeven aan een server, die de wijzigingen dan moet doorvoeren in het achterliggende systeem, bijvoorbeeld een database. Dit systeem lijkt in eerste instantie opgezet om relatief eenvoudige kleinschalige wijzigingen met bijvoorbeeld een desktop- of webtoepassing door te sturen naar een centraal



Een podcast voor geo-informatie, de laatste informatie eenvoudig beschikbaar voor de eindgebruiker.

systeem. Hetzelfde WFS-T-protocol kan in principe echter ook gebruikt worden om op grote schaal mutaties tussen organisaties uit te wisselen. De partij die de wijzigingen wil ontvangen moet dan de server draaien, in ons geval is dat de afnemer van geo-informatie. De figuur op de volgende pagina geeft schematisch aan hoe vanuit een centrale server verschillende afnemers kunnen worden bediend via WFS-T.

Bij het Britse Ordnance Survey is samen met IONIC Software een test uitgevoerd waarbij het principe van het automatische uitwisselen van wijzigingen met WFS-T is getest [IONIC Software, 2006]. Hierbij wordt de afnemer eerst voorzien van een startbestand, bijvoorbeeld in GML, zodat de uitgangssituatie bekend is. Deze gegevens worden niet via WFS-T maar via bijvoorbeeld FTP aangeleverd. Alle volgende wijzigingen worden wel via WFS-T doorgestuurd. Hiervoor moet de afnemende partij de URL van zijn WFS-T doorgeven aan de leverancier en verder het gebied waarin hij is geïnteresseerd. Ook moet hij aangeven hoe vaak hij nieuwe wijzigingen aangeleverd wil

krijgen. De aanbieder moet vervolgens bij iedere wijziging kijken voor welke afnemers deze in het interessegebied liggen en of het volgens de afgesproken bijwerkfrequentie al weer tijd is voor een nieuwe levering. Als het tijd is voor een levering, dan maakt de aanbieder als client contact met de WFS-T-server van de afnemer en geeft de wijzigingen door. Met iedere afnemer moet hiervoor apart verbinding gemaakt worden, wat bij grote aantallen afnemers wel wat organisatie vereist.

Vanuit het Open Geospatial Consortium wordt ook gekeken naar de mogelijkheden om met WFS-T mutaties uit te wisselen tussen organisaties. Onder de naam Geo-Synchronization Service is een systeem beschreven waarbij de gegevens van verschillende afnemers actueel worden gehouden met een combinatie van WFS-T en ATOM, een tegenhanger van RSS [Open Geospatial Consortium Inc., 2008]. Ook in dit systeem draaien de afnemers een WFS-T-server en worden die periodiek benaderd met transacties die door de synchronization service worden verstuurd.

Een voordeel van methodes, gebaseerd op WFS-T, is dat het een geaccepteerd protocol is waarvan verschillende implementaties beschikbaar zijn in zowel commerciële als open source-software. Hierdoor vereist het van de afnemers relatief weinig inspanning om op deze wijze mutaties te kunnen verwerken en zijn ze niet gebonden aan een software-leverancier.

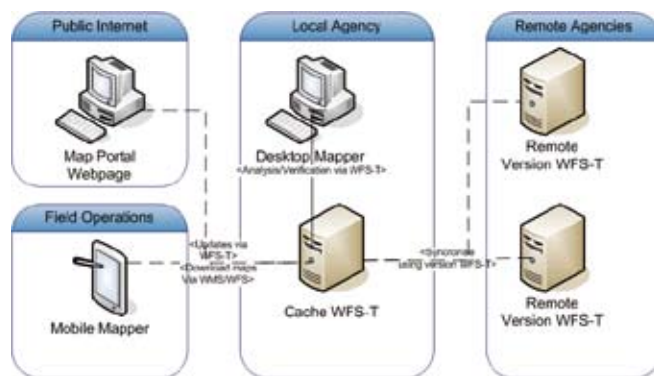
Bij het testen van het doorgeven van grotere aantallen mutaties aan een WFS-T-server zagen we voor de open source-toepassing Geoserver dat er problemen optraden. De interne manier van verwerken van transacties in Geoserver blijkt niet ingericht te zijn op grote aantallen mutaties in een enkele transactie. Hoewel de specificaties van het protocol het toepassen voor het uitwisselen van omvangrijke transacties niet in de weg staan, kunnen implementaties problemen geven omdat deze manier van werken niet voorzien werd.

Een aandachtspunt van een andere orde is dat hier gebruik gemaakt wordt van een geo-oplossing, terwijl het uitwisselen van wijzigingen helemaal geen specifiek geo-probleem is. De kans is daarom groot dat er naast deze oplossing voor andere gegevensverzamelingen andere oplossingen zijn en komen die meer generiek bedoeld zijn. Binnen de Nederlandse overheid bestaat al zo'n formaat: het Standaard Uitwisseling Formaat (StUF).

StUF 3.0

Het Standaard Uitwisseling Formaat is ontwikkeld om de uitwisseling van gegevens tussen overheidsorganisaties te standaardiseren en te voorkomen dat voor iedere gegevenssoort weer een nieuw systeem wordt ontwikkeld [ICTU, 2007]. Sinds versie 2.04 maakt StUF gebruik van XML als notatietaal voor berichten. Onder andere de communicatie met de landelijke voorziening van de Basisadministratie Adressen en Gebouwen maakt gebruik van StUF en codeert in de berichten ook geometrische gegevens. De geometrische attributen worden in GML beschreven. Ook binnen de gemeente Amsterdam worden geometrische gegevens via StUF-berichten verspreid. Daar is een systeem opgezet waarbij afnemende partijen zich kunnen abonneren op mutaties van gegevens. Dit lijkt op het hiervoor beschreven systeem van Ordnance Survey, echter nu met het belangrijke verschil dat dit systeem veel breder toepasbaar is binnen en tussen organisaties. Het vereist namelijk niet de inzet van services die heel specifiek voor het geo-domein zijn ontwikkeld.

Op basis van ervaringen uit de praktijk ligt nu een voorstel



Schema van koppelingen via WFS-T protocol (Afbeelding met toestemming van C. Shorter van <http://techblog.terrapages.com/2007/05/federated-geo-synchronization-standards.html>).

klaar voor StUF-versie 3.0. Hier zijn onder andere verbeteringen aangebracht waardoor StUF nu beter is in te zetten voor het synchroniseren van gegevens tussen verschillende partijen. Ook voldoet het nu aan de uitgangspunten voor een service-georiënteerde architectuur.

Overheidsservicebus

De overheid is een grote producent en afnemer van geo-informatie. Om te zorgen voor samenhang in de bouw en inrichting van de informatie-infrastructuur van de overheid is de Nederlandse Overheid Referentie Architectuur (NORA) ontwikkeld. Hierin staan inrichtingsprincipes die overheidsorganen kunnen toepassen bij de projecten. Voor de uitwisseling van gegevens tussen verschillende overheidsinstellingen is binnen NORA de OverheidsServiceBus (OSB) neergezet. 'De OverheidsServiceBus verzorgt de logistieke kant van het berichtenverkeer en omvat standaarden om berichten juist te adresseren en veilig en betrouwbaar te kunnen verzenden.' Binnen de OSB wordt onderscheid gemaakt tussen twee soorten berichtenverkeer: bevestigingen en meldingen. Bij een 'bevestiging' wordt direct een antwoord verwacht op een vraag. Bij 'meldingen' kan de reactie veel langer op zich laten wachten, bijvoorbeeld omdat de ontvangende partij eerst een aantal handelingen moet verrichten met de melding, voordat er een reactie teruggestuurd kan worden. Het is hierbij van belang dat de verzendende partij wel zeker weet dat zijn bericht aankomt bij de ontvanger. Wanneer er precies een reactie terugkomt is van minder belang.

Gezien de verschillende eisen die deze twee soorten berichtenverkeer met zich meebrengen, zijn ook twee verschillende protocollen vastgesteld voor de realisatie. Voor bevestigingen wordt in beginsel gebruik gemaakt van SOAP en voor meldingen van ebXML.

Door de belangrijke rol van de Overheids-ServiceBus binnen de informatieuitwisseling van de Nederlandse overheid, is het van belang om ook voor de uitwisseling van mutaties van geo-informatie hierbinnen te kunnen werken.

Als we dit systeem bekijken, dan valt het uitwisselen van mutaties van geo-informatie op twee manieren te bekijken. Als mutaties als bevraging worden gezien, dan neemt de afnemende partij het initiatief tot het opvragen van mutaties en doet zich een bevraging voor volgens het systeem van de OSB. Qua protocol kan dan dus SOAP worden gebruikt. Ligt het initiatief tot het doorgeven van mutaties echter bij de leverancier van de data, dan doet zich een melding voor en is ebXML de geschikte methode. Welke methode ook gekozen wordt, er is binnen de OSB geen plaats voor de tot nu toe meestal bij WFS gebruikte manier van web-services. Er wordt echter wel gewerkt aan SOAP-bindings voor WFS, waardoor hier wellicht meer mogelijkheden ontstaan. We zitten dan echter nog steeds met een aparte methode voor het synchroniseren van geo-informatie ten opzichte van andere, administratieve informatie.

Uitwisseling op basis van StUF staat hier neutraler in en biedt met bijvoorbeeld de inbedding van GML wel de mogelijkheid tot het uitwisselen van geometrische informatie.

Conclusies

Op basis van de huidige stand van de techniek zijn meerdere oplossingen mogelijk voor het uitwisselen van mutaties van geo-informatie. Oplossingen binnen het geo-domein kunnen worden gerealiseerd op basis van WFS-T, mits de gekozen software kan omgaan met grote transacties. Voor de integratie met andere gegevens lijkt een inpassing binnen algemenere oplossingen beter omdat het achterliggende probleem van het actueel houden van lokale kopieën van gegevens niet specifiek is voor het geo-domein. ■

Literatuur

- ICTU, programma EGEM, *StUF 03.00: Kandidaat Aanbeveling*, 2007.
- IONIC Software, *The Ordnance Survey Change Only Data Distribution using WFS-T*, 2006.
- Open Geospatial Consortium Inc., *Web Feature Service Implementation Specification Version: 1.1.0*, 2005.
- Open Geospatial Consortium Inc., *An Introduction to GeoRSS: A Standards Based Approach for Geo-enabling RSS feeds.*, 2006.
- Open Geospatial Consortium Inc., *OGC® Geo-Synchronization Service*, 2008.
- RSS Board, *RSS 2.0 Specification*, 2007.