

Beknopt overzicht van functionaliteit NDFF / EcoGRID in september 2007

Emiel van Loon, Floris Sluiter, Willem Bouten, Tom Visser, Lourens Veen

Universiteit van Amsterdam, Institute of Biodiversity and Ecosystem Dynamics. Nieuwe Achtergracht 166, 1018 WV, Amsterdam (vanloon@uva.nl)

De Gegevensautoriteit natuur in wording is al enige tijd in samenwerking met de Universiteit van Amsterdam en de stichting VeldOnderzoek Flora en Fauna (VOFF) bezig met de voorbereiding van het tot stand komen, en het verder uitbreiden van de Nationale Databank Flora en Fauna (NDFF).

De basis van de Nationale Databank Flora en Fauna is EcoGRID, het EcoGRID project richt zich op het centraal beschikbaar maken van alle natuurinformatie en omgevingsdata en zo bij te dragen aan de ontwikkeling van 'tools' voor relevante analyses en modellering. EcoGRID maakt deel uit van het Virtual Laboratory for e-Science (VL-e) initiatief en heeft als wetenschappelijk doel bij te dragen aan begripsvorming over het functioneren van ons ecosysteem. Een volwassen Nationale Database Flora en Fauna past binnen de doelstellingen van de Gegevensautoriteit natuur, om met het beschikbaar maken van natuurgegevens bij te dragen aan een afgewogen ruimtelijke inrichting en aan het informeren van goed natuurbeheer en natuurbeleid.

De belangrijkste ontwerpfilosofie achter NDFF/EcoGRID is dat het systeem zeer lang mee kan. Hiervoor zijn software-componenten gekozen die al goed uitontwikkeld zijn, heel goed schaalbaar zijn, en beschikbaar zijn in het publieke domein. De database software waarvan NDFF/EcoGRID gebruik maakt is PostgreSQL/PostGIS. Daarnaast is gekozen voor een structuur die uit gaat van verandering: het is gemakkelijk nieuwe databases aan te sluiten of bestaande databases uit te breiden. PostgreSQL/PostGIS is een ruimtelijke database waarmee het mogelijk is om rechtstreeks in de database selecties te maken op gridcellen (bijv. km-hokken), buffers (zone van 100 m links en rechts van een weg, of iets in de straal van 10 km rond een stad), of polygonen (een provincie of een perceel). Er kunnen dus erg veel vragen met een ruimte/tijdscomponent beantwoord worden door het database systeem zelf zonder dat aparte GIS-software nodig is.

NDFF wordt op dit moment getest en uitgebreid door een ontwikkelteam. Er worden nieuwe datasets toegevoegd en in een aantal geselecteerde pilot projecten gewerkt aan aansluiting en ontsluiting van gegevens van verschillende organisaties.

NDFF datamodel

Er is voor de Particuliere Gegevensverzamelende Organisaties (PGO's) één generiek datamodel ontwikkeld en al de bestaande databases zijn naar dit model geconverteerd. In dit model worden observaties altijd gekoppeld aan een ruimtelijk object en een tijdstip / tijdsinterval. De mogelijke ruimtelijke objecten zijn punten, lijnen of polygonen. Behalve de observatie zelf wordt ook de betrouwbaarheid, de gebruikte meettechniek, de observator en de eigenaar opgeslagen in de database (naast nog een groot aantal andere relevante zaken). De belangrijkste eigenschap van het datamodel is dat het wijzigingen toelaat: een tabel kan bijvoorbeeld eenvoudig toegevoegd worden, en de taxonomie kan worden gewijzigd bij voortschrijdend inzicht. Ook kan er extra informatie opgeslagen

worden, als het overkoepelende datamodel daarin niet zou kunnen voorzien. Het datamodel zorgt ervoor dat de database (en vooral de in- en uitvoer) gewoon blijft functioneren met deze wijzigingen. Het datamodel is op dit moment gedocumenteerd in de database zelf, en in een datamodel beschrijving. Daarnaast is er cursusmateriaal in ontwikkeling.

Interactie met NDFF/EcoGRID

Voor gebruikers verloopt de interactie met NDFF/EcoGRID, zowel voor het opvragen als het aanleveren van gegevens, via een gebruikersvriendelijke web-interface die in alle gangbare browsers werkt. Deze interface is nog niet gereed. De interactie bestaat kortweg uit: 1) het aanleveren van gegevens; 2) het nakijken (valideren) van aangeleverde gegevens, 3) het bekijken/opzoeken van gegevens. In de toekomst zijn meerdere portals te voorzien die gericht zijn ontwikkeld voor verschillende type eindgebruikers. NDFF bestaat uit een verzameling projecten (iedere PGO heeft meerdere projecten onder zijn hoede). Aan ieder project is één eigenaar (of verzameling eigenaren) gekoppeld en één concreet datamodel (dat past binnen het algemene NDFF datamodel). De gegevensinvoer verloopt op het niveau van een project, de gegevensuitvoer kan via de virtuele database op elk gewenst niveau plaats vinden: per project, per PGO of voor de gehele NDFF.

Koppeling met andere databases

De virtuele database van EcoGRID is bijna operationeel, en wordt uitvoerig getest. Met de huidige versie kunnen de verschillende NDFF databases benaderd worden. Gebruik makend van deze structuur kan automatisch ook iedere andere database benaderd worden die gebruik maakt van het NDFF-datamodel en is geïmplementeerd in PostgreSQL/PostGIS (koppeling A). Het aansluiten van een database met een ander datamodel (maar werkend binnen PostgreSQL/PostGIS) is ook op dit moment mogelijk als deze data via een View wordt aangeboden in een formaat dat de virtuele database begrijpt (koppeling B). Het aansluiten van een database met een ander datamodel en werkend met een ander database systeem (bijv. Oracle), kan op dit moment nog niet. Webservices hiervoor zijn voorzien en in ontwikkeling.(koppeling C). De gegevens tussen de verschillende databases en virtuele database worden verstuurd via een encrypted Secure Socket Layer (SSL) verbinding. Ondanks alle technische mogelijkheden tot het koppelen van nieuwe databases (waarbij koppeling A de voorkeur verdient uit het oogpunt van ondersteuning, stabiliteit en functionaliteit), is de optie om verspreidingsgegevens van andere projecten onder te brengen in de bestaande NDFF databases (als afzonderlijke projecten) vaak de beste: 1) de bestaande datastructuur laat het toe om het eigendom van ieder record vast te leggen (ook al zitten de gegevens opgeslagen in de database waarvan de kwaliteit wordt bewaakt door een bepaalde PGO), 2) er ontstaat geen wildgroei aan verschillende datamodellen, wat het moeilijk en duur zou maken om generieke analysetechnieken te ontwikkelen, 3) er ontstaat geen wildgroei aan 'validators', 4) het risico dat er niet of onvoldoende op de kwaliteit van de binnenkomende gegevens wordt toegezien wordt verkleind (de PGO's hebben de infrastructuur en ervaring om gegevens die binnen komen op kwaliteit te controleren).

Autorisatie en eigendom

Iedereen die gebruik maakt van NDFF/EcoGRID moet eerst inloggen via de virtuele database. Via de virtuele database worden alle gebruikersgegevens van NDFF/EcoGRID gedeeld tussen alle individuele databases. Binnen EcoGRID kan een eindeloos aantal rolgebaseerde rechten worden toegeschreven aan iedere gebruiker. Zulke rechten kunnen zijn: het opzoeken/bekijken van gegevens (tot op een bepaalde schaal), het downloaden van geselecteerde gegevens, het aanleveren van gegevens voor toevoeging aan NDFF, het controleren/valideren van aangeleverde gegevens, het schrijven van nieuwe queries (etc.).

Deze rechten kunnen zelfs worden gedifferentieerd voor verschillende onderdelen van NDFF. Eén gebruiker kan bijv. voor een deel van de NDFF gegevens (bijv. 1 soortgroep, of 1 bepaald geografisch gebied) meer rechten hebben dan voor de rest van NDFF.

Het toevoegen van een nieuwe gebruiker betekent dat de gegevens van die gebruiker worden toegevoegd aan NDFF/EcoGRID, en dat de rechten van die gebruiker worden vastgelegd. Vanaf dat moment kan die gebruiker via een web-browser bij de gegevens.

Er bestaat nog geen transactie-systeem waarbij bijv. een prijs gekoppeld is aan het zien of opvragen van bepaalde gegevens. Wel kan achteraf vastgesteld worden wat iedere gebruiker aan data heeft bekeken en/of ge-download, maar er is nog geen functie gemaakt die deze informatie op een overzichtelijke manier rapporteert.

Ter info:

Alle gebruikte software valt onder een open source licentie (GPL: <http://www.gnu.org/copyleft/gpl.html>).

Het database systeem dat wordt gebruikt in NDFF/EcoGRID is PostgreSQL (<http://www.postgresql.org/>). Binnen dit systeem wordt de de PostGIS bibliotheek (<http://www.postgis.org/>) gebruikt, zodat geografische analyses mogelijk zijn.

Voor webservices wordt zoveel mogelijk gebruik gemaakt van OpenGis Web Feature Service, OpenGis Web Map Service en OpenGis Web Coverage Service.

<http://www.opengeospatial.org/standards/wfs>

<http://www.opengeospatial.org/standards/wms>

<http://www.opengeospatial.org/standards/wcs>

Apache (<http://www.apache.org/>) wordt gebruikt als webserver.

Informatie over het VL-e project is te vinden op <http://www.vl-e.nl>

De server waarop NDFF/EcoGRID wordt gehost werkt onder Linux en wordt onderhouden door SARA, het nationale rekencentrum. (<http://www.sara.nl/>). Het cluster kan (nagenoeg eindeloos) uitgebreid worden met meer CPU's en opslagcapaciteit waardoor de capaciteit toeneemt.