

H.J.G.L. Aalders, J.D. Bulens, L.A.E. Vullings, P.A.L.M. Janssen,

P.J.M. van Oosterom, M. Reuvers, C.W. Quak en H.T.J.A. Uitermark

Een standaard voor semantiek in een Geo-Informatie Infrastructuur

Standaarden en GII

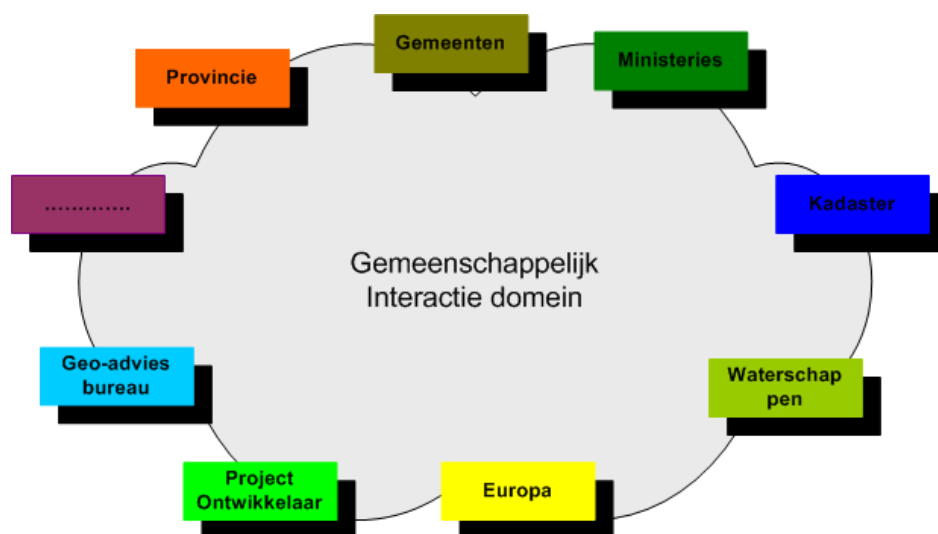
Binnen een Geo-Informatie Infrastructuur (GII) gaat het in essentie om Geo-informatie en mensen en hoe deze twee met elkaar in een goede communicatie worden gebracht. Om dit te realiseren is het van belang dat belangrijke aspecten van organisatie, dataleidend, technologie en standaarden aanwezig zijn.

Standaarden zijn onder te verdelen in standaarden voor technologie en semantiek. Met standaarden voor technologie wordt bedoeld dat communicatie en interactie tussen verschillende software systemen door middel van gezamenlijke interfaces mogelijk wordt. De OpenGIS standaarden zijn hier h t voorbeeld van. Met standaarden voor

semantiek wordt het mogelijk om informatie eenduidig te begrijpen en te gebruiken die door verschillende organisaties op verschillende tijden en voor verschillende doelen zijn gemaakt. Kortweg: standaarden voor technologie zorgen ervoor dat software systemen elkaar begrijpen en standaarden voor semantiek zorgen ervoor dat de mensen achter deze software systemen de informatie met haar betekenis kunnen begrijpen die gepresenteerd wordt.

Wat is semantiek?

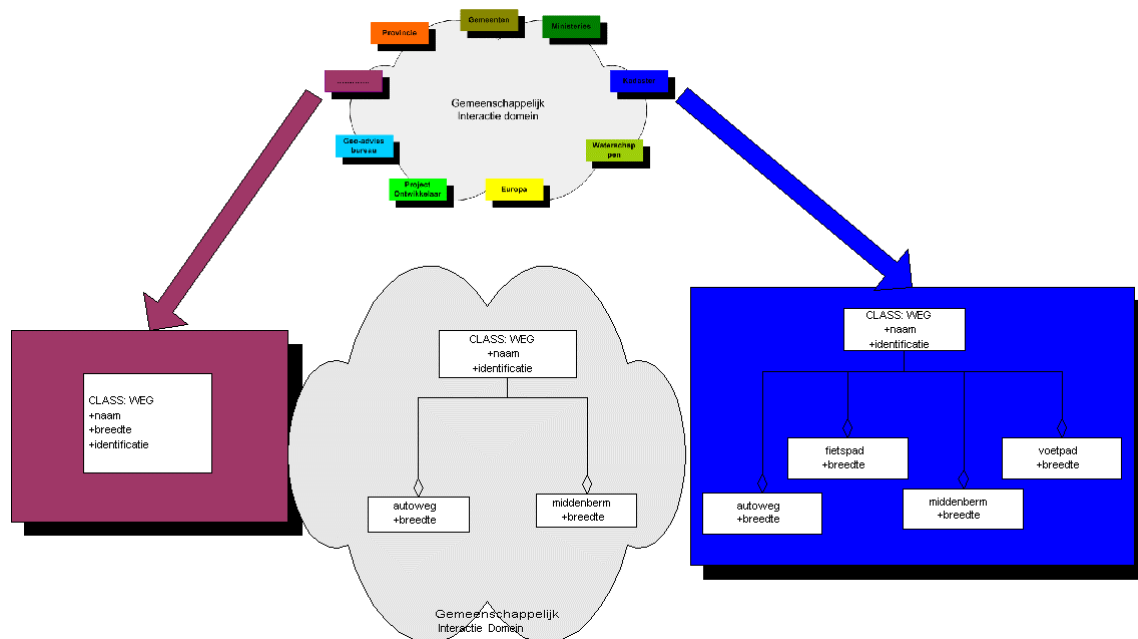
Door organisaties wordt steeds meer informatie uitgewisseld met een ruimtelijke component, Geo-informatie. Het betreft een grote verscheidenheid aan organisaties en Geo-informatie.



Figuur 1 Gemeenschappelijk Interactiedomein

Om duidelijk te maken wat de rol van semantiek is kunnen we als voorbeeld het informatie-element *weg* nemen. Een weg wordt gedefinieerd met een aantal attributen als naam, breedte en identificatie. Verder kan gedefinieerd worden dat de naam een tekst string bevat, de breedte is uitgedrukt in meters en de identificatie een nummer waarmee een weg uniek is te identificeren. Echter er wordt nergens aangegeven wat er wordt bedoeld met een weg. Is

een weg gedefinieerd inclusief fiets- en voetpaden? Wat wordt bedoeld met de breedte van de weg als niet duidelijk is welke onderdelen tot een weg worden gerekend? Deze laatste problematiek rekenen we tot de semantiek van een informatie element, wat is de gemeenschappelijke veronderstelling hoe een weg wordt gedefinieerd.



Figuur 2 Model van het gemeenschappelijk interactiedomein en het bijbehorend informatiemodel.

Deze laatste problematiek is in bovenstaande figuur weergegeven waarin twee organisaties een interne definitie hebben wat zij onder een weg verstaan. Op het moment dat beide organisaties besluiten om informatie over de breedte van de weg met elkaar uit te wisselen is een gemeenschappelijke notie vereist wat er bedoeld wordt met de breedte. Als voorbeeld is in de figuur een model geschetst van het gemeenschappelijk interactie domein (resultaat van harmonisatie) en het een bijbehorend informatie model. Dit model bevat de gemeenschappelijk uit te wisselen informatie, en door de relaties wordt er semantiek aan toegevoegd.

Behoeftte aan Semantiek

Wordt op dit moment nog de meeste Geo-informatie vrij direct gebruikt door de mens, in de toekomst zal ook een groot deel van de informatie (eerst) door machines gebruikt worden alvorens met de mens te communiceren. Is de mens (soms) nog in staat om de verschillende begrippen juist te interpreteren door impliciet gebruik te maken van enige context informatie (welk domein betreft het, welke leveranciers). Voor de machine is het noodzakelijk deze kennis expliciet te maken. Onze netwerk (informatie) maatschappij met internet als belangrijk communicatiemedium zorgt ervoor dat zelfverklarende informatie noodzakelijk is. Een flink deel van de formele structurele kennis rondom de begrippen (objecten die gemodelleerd worden) zit in de relatie die een object heeft met andere objectsoorten.

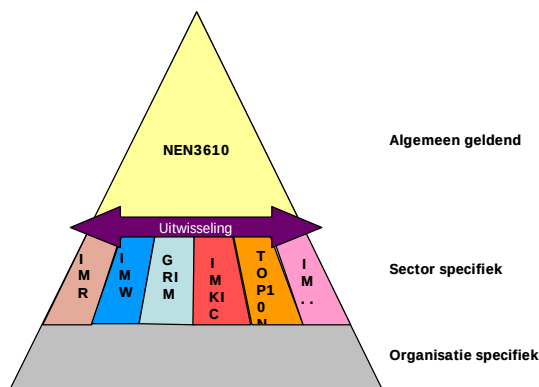
Afstemming domeintabellen: harmoniseren

Binnen een domein is het al niet eenvoudig om overeenstemming te krijgen over de begrippen en de definities hiervan, laat staan tussen de verschillende domeinen. In Nederland kennen we verschillende domeinstandaarden (IMRO/ruimtelijke ordening, IMWA/water, IMKICH/cultuurhistorie, GRIM/groen, kabels/leidingen, topografie, eigendom, bodem, etc.), waarbij vaak al vele partijen binnen een domein na overleg op een begrippenstandaard zijn uitgekomen. Veel van de meer recente domeinmodellen zijn al in UML klasse diagrammen beschreven (wat dus al een eerste stap richting het formaliseren is). Overigens is het in Internationale context ook belangrijk om dergelijke domeinmodellen op elkaar af te stemmen (bv. INSPIRE). Dit kan in sommige gevallen tot nog grotere discussies leiden dan welke plaatsvinden op nationaal niveau (omdat daar b.v. wet- en regelgeving tenminste voor partijen gelijk is). Wel speelt in de internationale context ook het gebruik van meerdere talen en zal dus voor elk begrip een vertaling gegeven moeten worden. Nog moeilijker wordt het om tot overeenstemming te komen over begrippen die in verschillende domeinen voorkomen: soms worden dezelfde woorden (labels) gebruikt voor iets wat geheel of gedeeltelijk een andere betekenis heeft. Bijvoorbeeld het begrip BANK, heeft in de

meubelindustrie een geheel andere betekenis dan een bank in de financiële wereld. En soms worden identieke begrippen compleet verschillend benoemd. Bijvoorbeeld salaris en loon zijn verschillende benamingen voor hetzelfde begrip. Het zal duidelijk zijn dat in onze netwerk (informatie) maatschappij dit in toenemende mate een probleem zal worden. Hoe moeilijk het ook zal zijn, er zal toch getracht moeten worden om tot harmonisatie tussen deze domeinen te komen.

Hoe pakken we dit op?

Sinds midden jaren negentig is er een algemene standaard die voorziet in de definitie van verschillende objectgroepen en kenmerken (entiteiten en attributen) in het veld van Geo-informatie. Dit is de NEN3610 standaard, het terreinmodel Vastgoed, en is bedoeld voor het uitwisselen van informatie op het gebied van aan het aardoppervlak gerelateerde objecten. NEN3610 definieert objectgroepen en kenmerken op een generieke manier en is hiermee afgestemd op het detailniveau dat aansluit bij de uitwisseling van informatie tussen sectoren. Detailinformatie die binnen een sector wordt vastgelegd en uitgewisseld wordt uitgewerkt in specifieke sectorstandaarden. Deze sectorstandaarden zijn in dat geval een toepassing van NEN3610 binnen een sector.



Figuur 3 Toepassing NEN3610 binnen een sector

Geconstateerd wordt dat de beperkte uitwerking van semantiek en structuur in NEN3610 een beperking is voor een brede toepassing van het

model. Deze beperkte uitwerking geeft een grote mate van vrijheid en maakt dat voor elke toepassing (sectoraal) aparte regels opgesteld moeten worden. Zoals eerder geschetst is door het toenemende gebruik van het elektronisch uitwisselen van informatie een toenemende vraag om een gemeenschappelijk kader waarin deze informatie gedeeld kan worden. Dit was de reden om NEN3610 als standaard te verbeteren.

Contouren van NEN3610

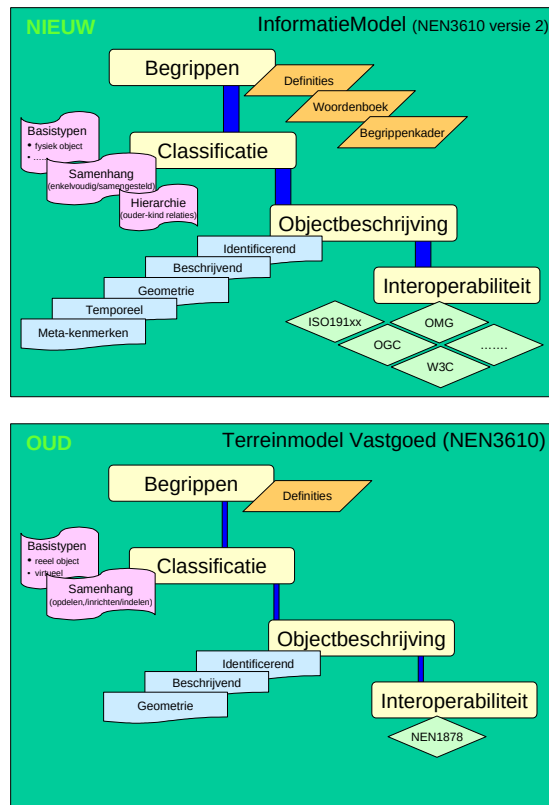
versie 2

Voor het Informatiemodel gaat het om objecten die in het terrein binnen zekere tijdsgrenzen onveranderlijk wat locatie en geometrie betreft, aanwezig zijn. We spreken over objecten, of iets specifieker ruimtelijke objecten waaruit het terrein bestaat. Binnen alle sectoren die onder het domein van het Geo-Informatiemodel vallen moeten deze ruimtelijke objecten hetzelfde gedefinieerd zijn. Voor het informatiemodel willen we uitsluitend het begrip Geo-object gebruiken en daarbij de definitie vaststellen die in het werkveld eenduidig gehanteerd kan worden. De objectgerichtheid is al bepaald in de huidige NEN3610 en we willen deze lijn nog strakker doorzetten in versie 2. Dit heeft uiteraard consequenties omdat we daarmee ook aangeven dat we alleen om kunnen gaan met discrete objecten. Continue objecten, in GIS-termen vaak in een rasterrepresentatie gebruikt, kunnen strikt genomen niet worden gemodelleerd. Vanuit de praktijk zal het gewenst zijn hier wel een oplossing voor te bieden.

Het informatiemodel bestaat uit een aantal samenhangende onderdelen die gezamenlijk de beschrijving van het model vormen. Deze onderdelen zijn:

- Begrippen
- Classificatie
- Objectbeschrijvingen
- Interoperabiliteit

In figuur 4 wordt d.m.v. een globaal schema weergegeven hoe deze onderdelen in respectievelijk NEN3610 en NEN3610 versie 2 zijn uitgewerkt.



Figuur 4 Schematische weergave van de onderdelen van het Geo-Informatiemodel (NEN3610 versie 2) in vergelijking met het oude Terreinmodel Vastgoed (NEN3610).

De verschillen uiteten zich in een completere en uitgebreidere beschrijving voor het informatiemodel als we dit vergelijken met het huidige model. Bovendien wordt de samenhang tussen de verschillende onderdelen verder uitgewerkt (de verbindingsstrookjes zijn breder weergegeven). Daarnaast is voor de interoperabiliteit aangegeven dat er aansluiting is met bestaande standaarden die een relatie hebben met het informatiemodel.

NEN3610 kennen we als het terreinmodel vastgoed. Met vastgoed worden objecten bedoeld die in het terrein aanwezig zijn en waarvan locatie en geometrie binnen zekere tijdsgrenzen onveranderlijk zijn. In de gebruikte classificatie zijn niet alleen de in het terrein waarneembare objecten opgenomen maar ook de niet zichtbaar begrensde, virtuele, objecten zoals verzorgingsgebied, kadastrale percelen, administratieve grenzen e.d. Deze objecten worden normaal gesproken niet met vastgoed geassocieerd. In het gebruik blijkt dan ook de term vastgoed voor verwarring te zorgen, zeker ook omdat de link wordt gelegd met de term zoals deze in de bouwwereld gebezigd wordt.

NEN3610 versie 2 bestaat uit een aantal samenhangende onderdelen die gezamenlijk de beschrijving van het model vormen. Deze

onderdelen zijn:

- **Begrippen**
Het eenduidige vastgelegde begrippenkader geldend binnen het domein (de thematisch afgebakende wereld)
- **Classificatie**
Het onderscheiden van objecten binnen het domein om daarmee een voor de praktijk bruikbare, maar ook noodzakelijke typering van objecten vast te leggen
- **Objectbeschrijvingen**
Het vastleggen van de algemeen geldende kenmerken van de objecten met de mogelijk daarbij behorende geldigheidsregels (bijv de inwinningsregels)
- **Interoperabiliteit**
het gebruik maken van bestaande standaarden van modellen die relaties kennen met het eigen informatiemodel. Daar waar geen standaarden bestaan moet voor het eigen model een implementatie als standaard worden bepaald.

Begrippen

Voor het informatiemodel is het gebruik van de semantiek binnen één begrippenkader van belang. De begrippen worden binnen het werkveld herkend en erkend wordt. Alle woorden die als jargon binnen het werkveld worden gebruikt en daarom hetzij als begrip, hetzij in de definitie voor een begrip gebruikt worden, worden in een woordenboek verzameld en verklaard. In de sectoren worden deze drie onderdelen dan uitgebreid met de sectorspecifieke begrippen met definities en opgenomen in de woordenboeken.

Het semantische model wordt geformaliseerd door het vast te leggen in een meer syntactisch model. Hiervoor wordt het UML klasse diagram gebruikt.

Classificatie

De waarde van een classificatie zit in het inzicht dat gegeven wordt in de systematiek van de indeling van die objecten en daarmee ook hoe de verschillende classificatiesystemen naar elkaar toe vertaald kunnen worden. Dit betekent dat in de classificatie ook semantiek zit.

Samenhang

Bij het informatiemodel gaat het om de objecten waaruit het terrein in ruime zin bestaat. Het zijn dan die geo-objecten die generiek zijn in hun eigenschappen voor alle onderliggende sectoren. Nu is een classificatie alleen zinvol als daarmee een zeker doel is gediend. Op het niveau van het informatiemodel spreken we van geo-objecten die generiek zijn, waarmee we willen uitdrukken dat

van de objecten alleen die eigenschappen vastgelegd worden die algemeen geldend zijn. Deze generieke objecten kunnen sectoraal specifiek gemaakt worden door de sectorspecifieke eigenschappen op dat niveau aan dat object toe te voegen. Door deze werkwijze blijven de objecten een zekere mate van abstractie behouden.

Hiërarchie

Voor het zinvol opnemen van hiërarchie in het classificatiestelsel moeten er in de set van geo-objecten subsets aan te wijzen zijn die een hoofd- en subniveau relatie met elkaar hebben. Een niveau wordt gekenmerkt door een gezamenlijke set van kenmerken die op dat niveau gelden. Het moet tevens zo zijn dat kenmerken die op een hoofdniveau gelden ook op alle onderliggende niveaus van toepassing zijn. In het model worden hiërarchie met relaties vastgelegd en weergegeven.

Het informatiemodel is objectgericht, dwz dat het informatie geeft over individueel te onderscheiden objecten binnen de beschreven wereld. Het object is de eenheid van informatie. Echter het onderscheiden van object kent een sterke afhankelijkheid van de schaal waarin de abstractie van de werkelijkheid plaatsvindt. Dit wordt ondervangen door in de classificatie onderscheid te maken in eenvoudige en samengestelde objecten. Samengestelde objecten bieden de mogelijkheid zowel een netwerk (bij infrastructuur), als een conglomeraat (een hiërarchische samenstelling) te classificeren.

Basistypen voor hiërarchische indeling

Binnen de context van het terreinmodel komen we tot een indeling op twee niveaus. Op het hoogste niveau een superklasse met één objecttype het geo-object. Direct daaronder een niveau waarvoor we een aantal basistypen onderscheiden. Voor de eenvoudige objecten stellen we een indeling voor die bestaat uit het volgende supertype:

1. Geo-object
alle objecten in het terrein met ruimtelijke kenmerken met een gedurende relatief korte tijd onveranderlijke locatie en geometrie

Daarnaast zal het uit basistypen bestaan:

1. Fysiek object
alle objecten die meetbaar en waarneembaar begrensd in het terrein aanwezig zijn
2.
3.

Objectbeschrijving

Bij de objectbeschrijving gaat het om de eenduidige

beschrijving van de kenmerken van de onderscheiden objecten. Bijvoorbeeld het type wegdek van een weg. Om structuur te kunnen bieden voor uitwisseling van objecten is het essentieel dat de objectkenmerken die voor de verschillende sectoren samen gelden eenduidig gedefinieerd worden.

De volgende kenmerken kunnen onderkend worden:

- **Identificerend**
Voor management van objecten, volgen van objecten in keten relaties, wordt aangegeven hoe dit op basis van nationaal unieke objectnummers vorm gegeven kan worden.
- **Beschrijvend**
De beschrijvende attributen krijgen door relaties een duidelijke plek in het model en zijn geen losse tabellen zoals in de huidige NEN3610. De beschrijvende kenmerken worden opgenomen in de semantiek.
- **Geometrie**
Voor de uitwerking van deze kenmerken wordt gebruik gemaakt van de OpenGIS / ISO specificaties. Zowel voor de ruimtelijke vastlegging van de vorm van het object als voor de gebruikte geo-referentie. Dit betekent dat de keuze voor de geometrie-kenmerken al goeddeels vastligt. Waar deze specificaties niet voorzien in een voor het informatiemodel beschikbare uitwerking wordt deze toegevoegd
- **Temporele**
Voor bijvoorbeeld monitoringsdoeleinden zal ieder object tijd afhankelijke kenmerken meekrijgen, zodat veranderingen in de tijd traceerbaar worden. Tot op zekere hoogte kan hiermee ook de dynamiek van objecten worden vastgelegd
- **Meta-informatie**
In het informatiemodel zal een verwijzing worden gemaakt naar de metadata-standaard (ISO19115) zoals die wereldwijd gebruikt wordt. Mogelijk dat er voor de metadata-standaard zoals deze voor het informatiemodel zal worden gebruikt een uitbreiding nodig is. Anders gezegd, inhoudelijk kan het informatiemodel eigen specifieke Metadata nodig hebben die in het model op objectniveau vastgelegd worden. Met de meta-informatie kenmerken is de kwaliteit van het object beschreven en vastgelegd.

Interoperabiliteit

In informatieoverdracht is interoperabiliteit een belangrijk begrip. Aan dit begrip zitten twee kanten. De ene kant waarbij vanuit het eigen informatiemodel wordt geredeneerd wordt

meegegeven dat het model wat is vastgelegd bepalend is voor het gebruik over en weer van informatie. De andere kant is die vanuit de omgeving. Op veel terreinen zijn al standaarden vastgelegd die een deel van het informatiemodel omvatten. Om interoperabiliteit te bereiken zijn deze standaarden nodig. De OMG (Object Management Group) Object Oriëntatie standaarden, (bv. UML), OpenGIS en ISO 191** zijn voor Geo-informatie de belangrijkste. Interoperabiliteit zal in het informatiemodel verwerkt worden door zoveel mogelijk gebruik te maken van deze standaarden en ze als onderdeel van informatiemodel te beschouwen. Dit kan door middel van verwijzingen naar de juiste standaarden.

Voor informatiemodel betekent dit concreet o.a. dat het model wordt beschreven m.b.v. UML (klasse) diagrammen en dat XML/GML wordt voorgeschreven als uitwisselingsformaat.