

# Modelleren van ruimtelijke gegevens, een kort overzicht

Ir. S. de Hoop

Gastmedewerker LUW, Vakgroep Landmeetkunde  
Hesselink van Suchtelenweg 6, Postbus 339  
6700 AH Wageningen  
Telefoon 08370-82643, telefax 08370-84643  
e-mail dehoop@axp13.iend.wau.nl

agro.informatica 8(3) / juni 1995

## Referaat

Dit artikel geeft een beknopt overzicht van de concepten van enkele general-purpose gegevensmodellen (met uitbreidingen) en geowetenschappelijke gegevensmodellen. Uit het overzicht blijkt dat het object gecentreerde of object-georiënteerde paradigma ten grondslag aan de modelleringsmethoden voor geowetenschappelijke toepassingen. Redenen hiervoor zijn dat deze toepassingen een grote verscheidenheid aan gegevenstypes en bijbehorende functies, vereisen. Maar ook omdat GIS een general-purpose systeem is.

Trefwoorden: database ontwerp, ruimtelijke databases, database management systemen, geografische informatiesystemen

## Inleiding

Ruimtelijke gegevens zijn gegevens met een geometrische component. Vakgebieden waarin ruimtelijke gegevens voorkomen zijn bijvoorbeeld chemie, robotica en geowetenschappen. De geowetenschappen omvatten onder andere geografie, geologie, bodemkunde, remote sensing, kartografie en landschapsarchitectuur.

Binnen de verschillende vakgebieden van de geowetenschappen kan men overeenkomstige bewerkingen op ruimtelijke gegevens onderscheiden, zoals inwinnen, opslaan, bevragen, transformeren en representeren van ruimtelijke gegevens. De gereedschappen benodigd voor het uitvoeren van deze bewerkingen, vormen tezamen een (general-purpose) geografisch informatiesysteem (GIS) (Burrough, 1986).

In de literatuur worden verschillende definities van een GIS gegeven. Volgens Bill en Fritsch (1991) bestaat een GIS uit vier componenten: *Een GIS is een computerondersteund systeem, dat uit hardware, software, gegevens en de toepassingen bestaat.* Met toepassingen wordt hier het gebruik van het informatiesysteem bedoeld. Burrough (1986) daarentegen onderscheidt drie GIS componenten, te weten *computer hardware, applicatie software modules en een geschikte organisatorische context.* Deze laatste component verwijst naar het eerdergenoemde gebruik van het informatiesysteem.

Deze definities hebben hun equivalenten in de database technologie. Een database management systeem (DBMS) is een general-purpose software systeem dat het proces van definiëren, construeren en manipuleren van databases voor verschillende toepassingen vergemakkelijkt. De software en de database tezamen worden database systeem genoemd (Elmasri en Navathe, 1989). Analoog hieraan kan de software component van een GIS beschouwd worden als een (general-purpose) database (management) systeem voor ruimtelijke databases. Centraal in een DBMS en GIS staat het gegevensmodel. Het gegevensmodel verschaft een conceptuele representatie van de gegevens.

Parallel aan bovengenoemde ontwikkelingen heeft in het database ontwerp scenario voor geowetenschappelijke en overige geavanceerde toepassingen, een verschuiving plaatsgevonden van een overwegend applicatie programma gecentreerde wijze van ontwerpen naar een object gecentreerde wijze. Deze trend in database modellering was al ingezet voordat de zogenoemde object-georiënteerde gegevensmodellen in zwang raakten (Engels, 1992/93).

Het gevolg van deze verschuiving had consequenties voor het ontwerp van database applicatie programma's. De bewerkingen op een database zijn niet specifiek voor een bepaald applicatie programma, maar meer voor de gemodelleerde objecten en gegevensstructuren. De bewerkingen zouden dus moeten worden ontworpen teza-

men met de object types en de gegevenstypes (Engels, 1992/93).

Samenvattend, een gegevensmodel zou concepten moeten verschaffen om zowel statische en structurele eigenschappen als dynamische eigenschappen van objecten te beschrijven.

Als een aanloop naar de bespreking van enkele gegevensmodellen in de secties 3 en 4, zal in sectie 2 kort enige aspecten relevant in het database ontwerp voor geowetenschappelijke toepassingen, besproken worden. Sectie 3 bespreekt enkele bekende general-purpose gegevensmodellen en hun toepassing voor geowetenschappelijke problemen. Sectie 4 beschrijft enkele geowetenschappelijke gegevensmodellen. Conclusies en een vooruitblik naar de toekomst worden gegeven in sectie 5.

## Database ontwerp voor geowetenschappelijke toepassingen

Ruimtelijke objecten in de werkelijkheid worden in databases gedefinieerd en beschreven door ruimtelijke gegevens. Ruimtelijke gegevens bestaan uit thematische, geometrische en temporele gegevens. Ook worden vaak grafische gegevens onderscheiden. Ruimtelijke gegevens kunnen beschreven worden in vectorformaat en in rasterformaat. In dit artikel wordt uitgegaan van de vectorstructurering van ruimtelijke gegevens.

Geometrische gegevens omvatten vier aspecten: positie (coördinaten), vorm, grootte en topologie informatie. Positie, vorm en grootte vormen tezamen de metrische gegevens. Topologie beschrijft de onderlinge samenhang van ruimtelijke objecten. Deze samenhang kan uit de metrische gegevens van de gerepresenteerde objecten afgeleid worden of eventueel door elementaire topologische verwijzingen bij de gerepresenteerde objecten vastgelegd worden. De verscheidenheid aan ruimtelijke relaties tussen objecten kan ondergebracht worden in drie verschillende categorieën (Egenhofer en Franzosa, 1991): topologische relaties welke invariant zijn onder topologische transformaties; metrische rela-

ties in termen van afstanden en richtingen; en relaties die de gedeeltelijke of volledige ordening van ruimtelijke objecten betreffen zoals beschreven door de voorzetsels voor, achter, boven en onder.

De ontwikkeling van een database doorloopt een aantal fasen, welke gelijk zijn aan de fasen van systeemontwikkeling. De eerste fase omvat een grondige analyse van de vereisten van eindgebruikers van het systeem. De tweede fase betreft de eigenlijke ontwerpfase. De vereisten van de eindgebruikers worden vertaald naar een conceptueel schema, daarbij gebruikmakend van een hoog-niveau gegevensmodel. De term hoog-niveau wil zeggen dat de concepten van het gegevensmodel een hoog abstractieniveau beschrijven. De derde fase betekent veelal een vertaling/afbeelding (mapping) van het conceptueel schema naar een logisch schema, daarbij gebruikmakend van een (conceptueel) logisch gegevensmodel. Mogelijk zijn in deze fase niet alle implementatie vraagstukken meer afgeschermd. De vierde fase betreft de daadwerkelijke implementatie met aandacht voor onder andere een efficiënte opslag en bevraging van ruimtelijke gegevens.

De object gecentreerde wijze van gegevensmodellering is vergelijkbaar met de object-georiënteerde wijze van gegevensmodellering. De object-georiënteerde benadering kan in principe toegepast worden in elke fase (Worboys, 1994). In dit artikel wordt echter het gegevensmodellering beperkt tot het conceptueel en logisch gegevensmodellering.

## General-purpose gegevensmodellen

In de voorgaande sectie zijn enige eigenschappen van ruimtelijke objecten in de werkelijkheid, beschreven. Vereenvoudiging is inherent aan gegevensmodellering. Dit is ook het geval op het geometrisch vlak. Derhalve kan aan de eigenschappen toegevoegd worden dat een ruimtelijk object meerdere geometrische representaties in de database kan hebben.

In deze sectie zullen enkele bekende general-purpose gegevensmodellen kort beschreven worden. Kort zal worden aangegeven of het gegevensmodel in kwestie geschikt is voor geowetenschappelijke toepassingen.

## Het entity-relationship model en uitbreidingen

Het entity-relationship model (ER-model) is een hoog-niveau conceptueel gegevensmodel. De concepten van het ER-model zijn: entiteit; relatie en attribuut waarde (Chen, 1976).

De semantische uitdrukingskracht van het ER-model is vaak bekritiseerd. Dit wordt veroorzaakt doordat relaties verscheidene, semantisch verschillende relaties veronderstellen zoals onderdeel-van en is-een relaties. Een aantal abstractie concepten zijn voorgesteld om het ER-model te verbeteren (zie voor referenties Engels, 1992/93). Deze concepten vormen ook de fundamentele basis van object-georiënteerd modelleren. Maar zoals hier zal blijken, zijn een aantal van deze concepten deels al aanwezig in het ER-model.

- Entiteit, entiteit type, classification, instantiation. Een entiteit representeert een reeel, kunstmatig of abstract fenomeen in de werkelijkheid, welke uniek identificeerbaar is onder alle entiteiten. Entiteiten worden onderverdeeld in verschillende entiteiten verzamelingen, de entiteit types. De tweeledige concepten van classification en instantiation komen overeen met de concepten van entiteit types (object types) en individuele entiteiten (object instances).
- Relatie, relatie type, associatie. Associatie wordt toegepast om verzamelingen van entiteiten van bepaalde entiteit types te vormen. Relaties zijn de gebruikelijke vorm van associatie in ER modelleringen. Een relatie onder entiteit types associeert een of meerdere entiteiten van deze types. Bijvoorbeeld de rivier de Rijn stroomt door de provincie Gelderland. Relaties worden ingedeeld in relatie types.

- **Attribuut waarde, attribuut, domein.** Entiteiten en relaties kunnen gedefinieerd en beschreven worden door attributen. Aan attributen worden waarden toegekend. De waarden die aan een attribuut toegekend kunnen worden wordt het domein van het attribuut genoemd. Een attribuut kan een atomaire, enkelvoudige waarde hebben, maar ook een verzameling waarden (meer-voudig attribuut).
- **Identificatie.** Elke afzonderlijke entiteit is uniek identificeerbaar, dit onafhankelijk van zijn attribuutwaarden. Deze identificatie is voornamelijk een systeem concept, maar het appelleert aan een natuurlijk idee (Worboys, 1994). In het ER-model wordt een entiteit geïdentificeerd door een of meerdere attributen, de primaire sleutel genaamd.
- **Aggregatie.** De eigenschap-van relaties tussen attributen en entiteit types (of relatie types) is een vorm van aggregatie. Een relatie type representeert naast associatie ook aggregatie. Daarnaast kan een aggregaat een samengesteld entiteit representeren, dat uit verschillende entiteiten is samengesteld. Aggregaten representeren in het laatste geval onderdeel-van relaties.
- **Generalisatie/specialisatie.** Het abstraheren van eigenschappen van verschillende entiteit types tot een gegeneraliseerd entiteit type, wordt generalisatie genoemd. Het omgekeerde proces, het onderverdelen van een entiteit type in subtypes wordt specialisatie genoemd. Generalisatie en specialisatie representeren is-een relaties.

In de voorgaande secties is aangegeven dat naast de statische en structurele eigenschappen van ruimtelijke objecten, ook de dynamische eigenschappen beschreven dienen te worden. In het onderhavige geval betekent dat in principe het definiëren van abstracte data types. Voorbeelden van abstracte data types zijn bijvoorbeeld geometrische data types, zoals punt, lijn en polygoon met geschikte operatoren. De geometrische gegevens van een ruimtelijk object worden in dit geval als attribu(u)t(en) beschouwd. Een ruimtelijk object kan in dit

geval meerdere ruimtelijke representaties hebben.

### Het relationeel model en uitbreidingen

Het huidige paradigma voor general-purpose informatie beheer is het relationeel model. Het relationele model representeert de gegevens in een database als een verzameling relaties (tabellen). Een rij, record of tuple in de tabel representeert een entiteit of een individuele relatie tussen entiteiten. Een kolom representeert een attribuut van het entiteit type. Aan een attribuut kunnen slechts enkelvoudige waarden toegankelijk worden. Een entiteit wordt geïdentificeerd door een primaire sleutel. Relaties tussen entiteiten worden vastgelegd door zogenoemde vreemde sleutels, dat wil zeggen een attribuut van een bepaald entiteit type wiens waarde verwijst naar de waarde van de primaire sleutel van een ander entiteit type.

Het relationele model is ongeschikt voor het beheer van ruimtelijke gegevens. Dit wordt nog versterkt door de restrictie van enkelvoudige attribuutwaarden. De positie van een punt kan alleen worden geïmplementeerd met de standaard gegevenstypes van het systeem, bijvoorbeeld integers. Polylijnen en polygonen kunnen enkel gerepresenteerd worden door een herhaling van rijen in een tabel.

De afwezigheid van geometrische gegevenstypes veroorzaakt tevens dat er geen geometrische operatoren zijn. Dit heeft tot gevolg dat een complexe combinatie van beschikbare relationele operatoren gekozen moet worden voor ruimtelijke bevragingen.

Het is zelfs denkbaar dat ruimtelijke bevragingen alleen opgelost kunnen worden door middel van het vastleggen van de gegevensstructuur in applicatie programma's.

Derhalve zijn er al vele voorstellen gedaan om het relationele model uit te breiden met opties waardoor complexere problemen aangepakt kunnen worden. Voorbeelden daarvan zijn bijvoorbeeld Postgres (Stone-

braker en Rowe, 1986) en zijn commerciële opvolger Illustra, Ask OpenIngres/Intel-  
ligent Database (Maw, 1993), en de voorgestelde SQL3 uitbreidingen (Bundock, 1991).

Deze pakketten worden wel object-relatie-  
neel genoemd. Maar wat hebben deze pakketten relevant voor GIS te bieden? Onder andere dat de gebruiker applicatie-specifieke gegevenstypes, operatoren, functies en zoekstructuren zelf kan definiëren, en overerving (zie volgende sectie) vindt plaats van zowel statische en structurele als dynamische eigenschappen van ruimtelijke objecten in een generalisatie/specialisatie hiërarchie.

### Een object-georiënteerd model

De afgelopen jaren is regelmatig aangetoond dat de object-georiënteerde benadering in hoge mate geschikt is voor geowetenschappelijke gegevensmodellering (zie voor referenties Worboys, 1994).

Men zou zich kunnen afvragen waarin de object-georiënteerde benadering zich onderscheidt van de object gecentreerde wijze van gegevensmodellering. De object gecentreerde wijze van gegevensmodellering vindt zijn ontstaan in het semantisch gegevensmodellieren. Semantische gegevensmodellen zijn over het algemeen gericht op de representatie van gegevens, terwijl object-georiënteerde programmeertalen zich vooral richten op het manipuleren van gegevens (King, 1989)

Een nadeel bij de object-georiënteerde methode is dat de betekenis van algemeen gebruikte termen (nog steeds) niet gestandaardiseerd is (Worboys, 1994). Aan de concepten, genoemd bij de bespreking van het ER-model en uitbreidingen, kunnen nog de volgende concepten toegevoegd worden:

- **Overerving.** Het concept overerving hangt samen met de begrippen generalisatie en specialisatie. Het kan leiden tot overervingshiërarchieën.
- Daarnaast zijn *encapsulation* en *polymorphisme* belangrijke object-georiënteerde begrippen, maar omdat ze voor-

al systeemconcepten representeren worden ze hier niet verder behandeld.

## Geowetenschappelijke gegevensmodellen

Ook binnen de geowetenschappen zijn gegevensmodellen ontwikkeld voor diverse doelstellingen, bijvoorbeeld support voor ruimtelijke analyses of gegevensuitwisseling. Twee gegevensmodellen zullen hier kort beschreven worden.

### De formele gegevensstructuur voor enkelvoudige vectorkaarten

Een vector kaart representeert een verzameling vector-gestructureerde terrein gegevens. Een enkelvoudige vectorkaart kan worden beschouwd als een enkele kaartlaag. De term enkelvoudig verwijst naar een aantal conventies (zie Molenaar, 1989). De formele gegevensstructuur (fds) voor enkelvoudige vectorkaarten is een topologisch gegevensmodel. Het model heeft aspecten van het ER-model en het netwerk model. In database termen is de fds een gegeneraliseerd database schema. De fds is gebaseerd op drie concepten, namelijk: terrein objecten, onderverdeeld in punt, lijn en vlakobjecten; thematische attributen van de terrein objecten en geometrische attributen van de terrein objecten.

De thematische attributen zijn gestructureerd in een hiërarchisch classificatie schema. De geometrische attributen daarentegen worden gerepresenteerd door geometrische entiteit types, namelijk Node, Arc en eventueel Face. Als gevolg hiervan kunnen in de fds verschillende semantische niveaus worden onderscheiden: een geometrisch niveau en een terrein object niveau. De basis topologische verwijzingen worden vastgelegd op het geometrisch niveau. Een Arc entity heeft een begin Node en een eind Node, evenzo heeft een Arc een Face aan zijn linkerkant en (wellicht dezelfde) een Face aan zijn rechterkant. De (topologische) relaties tussen de terrein objecten kunnen afgeleid worden uit de gegevensstructuur.

In de fds zijn de geometrische gegevens niet opgenomen als attributen van de enti-

teit types, maar op zichzelf staande entiteit types. Een volledige beschrijving van de fds wordt gegeven in (Molenaar, 1989).

In (Molenaar, 1991) wordt een membership functie gedefinieerd waarmee de meeste topologische relaties tussen terrein objecten geanalyseerd kunnen worden.

### SAIF

Het Spatial Archive and Interchange Format (SAIF) is een gegevensmodel voor gegevensuitwisseling en archivering (Kucera en Sondheim, 1992). Het model is gebaseerd op object-georiënteerde concepten. De basis concepten van SAIF zijn: class (entiteit type), object (entiteit) en overerving. Entiteit types zijn aan elkaar gerelateerd door overervingshiërarchieën.

Een geografisch object in SAIF bestaat uit een aantal componenten, waaronder thematische attributen; geometrie welke omvat coördinaten, indexering en vormen; relaties tussen geografische objecten en metagegevens, gegevens over gegevens.

De ruimtelijke gegevens worden zowel in vector- als rasterformaat gerepresenteerd. Daarnaast beschrijft SAIF expliciet de representatie van temporele gegevens. Ook kunnen vage grenzen gerepresenteerd worden.

### Discussie

Een kort overzicht van het modelleren van ruimtelijke gegevens als in het voorgaande kan door de beperkte omvang per definitie niet volledig zijn. De beperkte omvang heeft tevens geleid tot de keuze om de gegevensmodellen niet te illustreren met voorbeelden. Het artikel beoogt dan ook slechts een indruk te geven van de trends in gegevensmodellering en de aspecten die van belang zijn bij het modelleren van ruimtelijke gegevens. Niet of nauwelijks aan de orde gekomen zijn constraints (beperkingen), dimensionele aspecten, het tijdsaspect (tijdstip of tijdsinterval van gegevensinwinning) en de nauwkeurigheid van ruimtelijke gegevens. Ook is geen aandacht besteed aan gewenste systeemvereisten en in hoeverre deze haalbaar zijn in de

huidige database management systemen, in verband met een mogelijke voorkeur voor een object-relacioneel danwel object-georiënteerd systeem.

Voor meer informatie over conceptueel gegevensmodelleren van ruimtelijke gegevens wordt verwezen naar het werk van Feuchtwanger (1993). Feuchtwanger heeft een grondige studie gemaakt van general-purpose gegevensmodellen en geografische gegevensmodellen en heeft daaruit een conceptueel geografisch semantisch database model gesynthetiseerd. Het voorgestelde model kan als een framework worden gebruikt bij de ontwikkeling van een geografische DBMS, maar representeert ook een stap in de richting van een algemene theorie voor geografische informatie beheer en analyse.

### Literatuur

- Bill R. en D. Fritsch (1991)  
Grundlagen der Geo-Informationssysteme. Band 1 Hardware, Software und Daten. Wichmann.
- Bundock M.S. (1991)  
SQL-SX: Spatially Entended SQL - Becoming a reality. Paper presented at EGIS, Brussels, April 3-5, 1991
- Burrough, P.A. (1986)  
Principles of Geographical Information Systems for Land Resources Assessment. Oxford Science Publications.
- Chen, P.P.-S. (1976)  
The Entity-Relationship Model-Toward a Unified View of Data. ACM Transactions on Database Systems, vol. 1, nr. 1, pp 9-36
- Egenhofer M.J. en R.D. Franzosa (1991)  
Point-set topological spatial relations. Int.J.Geographical Information Systems, vol.5, nr.2, pp 161-174
- Elmasri, R. en S.B. Navathe (1989)  
Fundamentals of Database Systems. Addison-Wesley.
- Engels, G., M. Gogolla, U. Hohenstein, K. Hulsmann, P. Lohr-Richter, G. Saake en H.-D. Ehrich (1992/93)  
Conceptual modelling of database applications using an extended ER model.

Data & Knowledge Engineering 9, pp  
157-204

Feuchtwanger M. (1993)

Towards a Geographic Semantic Database  
Model. Proefschrift Simon Fraser  
Universiteit, B.C.

King, R. (1989)

My Cat Is Object-Oriented.  
Object-Oriented Concepts, Databases, and  
Applications, editors W. Kim, F.H.  
Lochovsky, ACM Press.

Kucera H.A. en M. Sondheim (1992)

SAIF - Conquering space and time. GIS'92  
Symposium, Vancouver, B.C.

Maw, K.D. (1993)

Documentation of ESL Incorporated on  
Ingres Server.

Molenaar M. (1989)

Single Valued Vector Maps - A Concept in  
Geographic Information Systems.  
Geo-Informationssysteme, nr.1, pp 18-26

Molenaar, M. (1991)

Terrain Objects, Data Structures and Query  
Spaces. Geo-Informatik, editor Matthaas  
Schilcher. Siemens Nixdorf  
Informationssysteme, pp 53-70

Stonebraker, M. en L.A. Rowe (1986)

The design of Postgres. ACM SIGMOD,  
vol.15, nr.2, pp 340-355

Worboys, M.F. (1994)

Object-oriented approaches to  
geo-referenced information.  
Int.J.Geographical Information Systems,  
vol.8, nr.4, pp 385-399