

L. Heres GEOSPERANTO

De lange weg naar een gemeenschappelijke taal voor
geografische
informatie

De roep om standaardisatie van geografische informatie wordt de laatste jaren steeds luider. Binnen het Europa zonder grenzen zal het zwaartepunt daarbij verschuiven van nationale naar Europese standaards. Dit artikel beschrijft de huidige stand van zaken.

Waarom standaardiseren we eigenlijk ?

Denkt u zich eens in: u bent aannemer en heeft een opdracht gekregen om een groot kantoorgebouw in Rotterdam te bouwen. Voor dit gebouw bent u op zoek naar geschikte aluminium raamkozijnen. In een brochure van een bedrijf uit Stuttgart vindt u de kozijnen van uw gading. U bestelt er 500 stuks, die vier weken later worden geleverd. Bij een Gents bedrijf, gespecialiseerd in zonwerend glas, heeft u inmiddels 500 ramen op maat besteld. Deze arriveren een dag later en worden door uw personeel in de sponningen van de kozijnen uit Stuttgart gezet. Vanzelfsprekend ? Vandaag de dag wel, maar dit had u driehonderd jaar geleden niet moeten proberen. En ik bedoel daarmee niet te zeggen dat er toen nog geen aluminium ramen bestonden of dat een transport Stuttgart - Rotterdam veel langer dan vier weken zou hebben geduurd. Ik bedoel dat u een veel fundamenteeler probleem zou hebben gehad: de ramen zouden waarschijnlijk niet gepast hebben in de kozijnen, omdat het bedrijf uit Stuttgart, dat uit Gent en uw eigen aannemersbedrijf met verschillende lengtematen zouden hebben gewerkt; maten waarvan u bovendien niet zou hebben geweten hoe die zich tot elkaar verhielden. Waar wij niet dagelijks bij stil staan is dat de lengtematen die wij anno 1992 gebruiken, alle met zeer hoge nauwkeurigheid op elkaar zijn afgestemd, waardoor de timmerman uit Reykjavik als hij een plank van 1.863 m afzaagt, inderdaad op dezelfde lengte uitkomt als een timmerman uit Buenos Aires. Toch zal een aannemer van drie eeuwen geleden het gemis van een nauwkeurige lengtestandaard waarschijnlijk niet erg gevoeld hebben. Want hij bestelde geen kozijnen uit Stuttgart en ramen uit Gent. Hij liet deze maken op de bouwplaats

zelf, en daarbij kon hij zijn eigen maten gebruiken. Gestandaardiseerde lengtematen werden pas belangrijk bij de aanleg van grote infrastructurele werken (kanalen, verdedigingswerken), waarbij bedrijven uit verschillende regio's betrokken waren. Ze werden ook van belang bij de opkomst van industriële produktiewijzen waarbij toeleveranciers een belangrijke rol gingen spelen.

Op het gebied van de produktie van geografische informatie bevonden wij ons tot voor kort in een situatie die in veel opzichten overeen kwam met de 16e eeuwse aannemerswereld: iedere organisatie vervaardigde haar eigen bestanden volgens eigen methode en gebruikte daarbij materiaal uit eigen werkplaats. Daarbij hoefde nauwelijks te worden gelet op hoe anderen te werk gingen. Deze werkwijze zal echter veranderen. In de toekomst zal het heel normaal zijn om grote hoeveelheden informatie van diverse leveranciers te betrekken, deze te bewerken, aan te vullen en weer door te verkopen aan verschillende afnemers. Daarom wordt het steeds belangrijker dat leveranciers en afnemers zoveel mogelijk op basis van dezelfde normen werken, kortom dezelfde maten hanteren.

De hoofdreden om te standaardiseren is van financiële aard. Dat kan aanschouwelijk gemaakt worden met het volgende voorbeeld: Zes organisaties, ieder met hun eigen GIS van verschillende makelij, willen elkaar bestanden toesturen. Laten we er even vanuit gaan dat er geen technische belemmeringen zijn, dat wil zeggen dat het in principe mogelijk is om alle gegevens uit systeem X zo te vertalen dat ze op systeem Y kunnen worden afgebeeld. Eén mogelijkheid is om directe onderlinge vertaalprogramma's te schrijven. Hiervoor zijn in totaal 30 vertaalprogramma's nodig (1). Zie Figuur 1. Een tweede mogelijkheid is om een "tussentaal" af te spreken en om programma's te schrijven die ieders systeemtaal in deze tussentaal omzetten of omgekeerd. Nu zijn er nog maar 12 verschillende programma's nodig: een flinke verbetering (2). Zie Figuur 2. Een nog rigoureuzere oplossing is om alle 6 bestaande systemen naar de schroothoop te brengen en ze te vervangen door systemen van identieke makelij. Er zijn dan in het geheel geen vertaalprogramma's meer nodig. Omdat deze laatste oplossing in ieder geval op middellange termijn geen haalbaar

alternatief is, besluiten de organisaties om voor de tweede oplossing, de tussentaal, te kiezen. Naast financiële zijn er ook technische redenen om standaardisatie na te streven. Vaak kan een bepaald aspect uit bestand X moeilijk of met onvoldoende betrouwbaarheid op bestand Y worden afgebeeld, hoe slim het vertaalprogramma ook is. Ook wanneer we de gegevens uit twee verschillende bestanden willen combineren, blijkt standaardisatie een puur technische noodzaak.

Wat valt er eigenlijk te standaardiseren aan geografische informatie ?

Vanuit semiotisch gezichtspunt (zie Intermezzo I), zijn er vier hoofdgebieden te onderscheiden die zich voor standaardisatie lenen: begrippen, conceptuele modellen, gegevensstructuren en syntax.

Bij begrippen, in de informatie-analyse ook vaak "entiteiten" genoemd, moet men denken aan zaken zoals "dorp", "bos", "huis", maar ook aan wiskundige begrippen als "coördinaat", "punt", "lijn" en daarnaast aan concepten als "getal", "cijfer" en "letter".

Een conceptueel model geeft aan hoe de verschillende entiteiten zich tot elkaar verhouden. Belangrijke relaties die in een dergelijk model worden gedefiniëerd zijn (met voorbeelden):

- classificatierelaties: het begrip "dorp" en het begrip "stad" beide zijn van het het begrip "nederzetting".
- aggregatierelaties: de entiteit "dorp" is opgebouwd uit de entiteiten "huis", "weg" en "park".
- associatierelaties: een "bocht" is een associatie tussen twee "weg-elementen"
- attributrelaties: een bijzonder soort associatie. Voorbeeld: het begrip "dorp" kan worden geassocieerd aan het begrip "inwoneraantal"
- representatierelaties: een bijzonder soort associatie tussen geografische begrippen enerzijds en wiskundige begrippen anderzijds. In een representatierelatie kan bijvoorbeeld worden bepaald dat het begrip "huis" door een "punt" wordt voorgesteld.

Zie [Molenaar, 1992] voor een meer gedetailleerde behandeling van deze relatietypen.

Een gegevensstructuur definieert een hiërarchische verzameling symbolen, bedoeld om de begrippen uit het gegevensmodel weer te geven.

De syntax van een bestand definieert een aantal "hulp"symbolen waarmee de gegevensstructuur daadwerkelijk opgebouwd kan worden.

Aan de hand van bovenstaande typologie zijn de huidige (geografische) informatie standaards als volgt te rangschikken:

- Internationale byte standaarden
- Nationale en multinationale standaarden voor gegevensstructuren
- Internationale syntaxstandaard
- Europees Conceptueel Gegevensmodel
- Nationale en multinationale standaards voor geografische begrippen

Internationale byte standaarden

Een zeer verbreide conventie is om bits (3) te groeperen in groepjes van acht

of een veelvoud daarvan, de zogenaamde byte. Hoewel er diverse uitzonderingen

bestaan, kunnen we de byte toch met een gerust hart een de facto wereldstandaard noemen, zeker voor het soort toepassingen waarover we in dit

artikel spreken.

Een enkele byte kan 256 verschillende waardes hebben. Daarom leent de byte zich

goed voor de weergave van een verzameling letters, cijfers en leestekens, in

vaktaal meestal als "karakters" aangeduid. Een veel gebruikte byte-code

standaard is ISO 646, die ook vaak ASCII-code wordt genoemd omdat het een van

oorsprong Amerikaanse standaard is die, met enige wijzigingen, later tot ISO-

standaard is gepromoveerd (4). ISO 646 is een zogenaamde 7-bits code, wat

inhoudt dat van de 8 beschikbare bits er slechts zeven worden gebruikt om

karakters aan te duiden. De 8e bit heeft een contrôle functie (pariteitsbit).

Dit betekent dat er met ISO 646 slechts 128 karakter codes kunnen worden

gedefinieerd in plaats van 256. Hiervan zijn er 52 bestemd voor de letters van

het Latijnse alfabet (26 hoofdletters, 26 kleine letters), 10 voor de cijfers 0

t/m 9 en 22 voor leestekens en andere veelgebruikte tekstsymbolen (&, %, *, \$, >

etc.). Vier en dertig codes zijn toegewezen aan "onzichtbare" karakters, dat

wil zeggen tekens zonder een grafisch equivalent. Deze tekens hebben meestal

een syntax-functie, wat inhoudt dat ze gebruikt worden om het begin en einde

van bepaalde structuren aan te geven. Een veel gebruikt teken is het "Carriage

Return" (5) karakter (binair code: 00001101). Dit teken wordt door de meeste

tekstverwerkers wordt gebruikt om het einde van een alinea aan te geven (harde

return). Tenslotte zijn er 10 codes gereserveerd voor variabele, taalafhankelijke letters. Met 01111110 wordt in de Duitse versie de

"ü"

weergegeven, in de Franse de "è" en in de Noorse de "å". Intrigerend is dat ook begrippen als "niets", "spatie" en het cijfer "0" hun eigen codes hebben. Ook het niets moet blijkbaar door iets worden voorgesteld. Zie Figuur 3 voor een overzicht van de codes.

Omdat ISO 646 voor vele toepassingen te beperkt is (bijvoorbeeld voor een gemengd Duits - Franse tekst, of voor een tekst met Griekse symbolen), zijn er later ISO sets gedefiniëerd waarbij de contrôlebit is opgeëfferd, zodat er 256 verschillende karakters kunnen worden onderscheiden [ISO 2022, 4873, 6429, 6937/2 en de 8859 serie].

Nationale en multinationale standaarden voor gegevensstructuren
Met bytes kan men "velden" opbouwen en met velden "records". Hiermee betreft men het terrein dat vaak met de term "record-formaten" wordt aangeduid. In combinatie met een bepaald soort bestandsgebruik (het leveren van informatie aan anderen) leidt dit tot lelijke termen als "uitwisselingsformaat" (6). In feite is een uitwisselingsformaat de gegevensstructuur die behoort bij de "tussentaal" oplossing zoals geschetst in figuur 2. Op dit terrein bestaan nog geen ISO of CEN normen (7). Wel bestaan er inmiddels een groot aantal nationale en multinationale standaarden [Moellering, 1991]. Eén van de oudste van de huidige generatie is de Britse NTF (8), [Haywood, 1987; Sowton, 1991]. De NTF is een moderne standaard in die zin dat hij topologische relaties tussen wiskundige elementen als lijnsegment, knooppunt en vlak op een expliciete manier kan weergeven. De NTF 1.0 is inmiddels opgevolgd door de NTF 1.1 en die wordt binnenkort opgevolgd door de BS 7567 (9). Deze zal gebruik maken van het Europese Conceptuele Gegevensmodel en de internationale syntaxstandaard ISO 8211. In de volgende paragrafen ga ik daar nader op in. Van iets recentere datum is de Amerikaanse SDTS (10), ontwikkeld door het National Committee for Digital Cartographic Data Standards [Rossmeyssl & Rugg, 1991]. De eerste officiële versie werd gepubliceerd in januari 1988. De SDTS is duidelijk wetenschappelijker van opzet dan de NTF en heeft een strengere theoretische basis. De documentatie is daardoor voor de gemiddelde leek minder toegankelijk, dit in tegenstelling tot de NTF, waarvan de documentatie

eenvoudig is gehouden.
Een andere standaard die niet onvermeld mag blijven is DIGEST (11).
Deze
standaard is in eerste instantie bedoeld voor het versturen van
topografische
bestanden van en naar NATO partners. DIGEST is in sterke mate
beïnvloed door de
Amerikaanse SDTS, maar is minder open, dat wil zeggen: maakt een
keuze uit één
van de alternatieven die de SDTS biedt.
EDIGÉO (12) is de Franse nationale standaard. Qua gegevensstructuur
is EDIGÉO
identiek aan DIGEST. Het verschil is dat de documentatie Franstalig
is en dat
de bijbehorende Feature en Attributen Catalogus gericht is op civiele
toepassingen in plaats van op militaire [Salgé, 1991].
In Duitsland wordt momenteel nog EDBS gebruikt (13). EDBS is evenals
het
Nederlandse SUF (14) een generatie ouder dan de andere in dit
artikel
besproken standaards. Bij deze old-timers ontbreekt de mogelijkheid
om
topologische relaties expliciet aan te geven. Ze zijn uitstekend voor
het
versturen van "plot"bestanden, of Digitale Kartografische Modellen
zoals van
der Schans ze noemt [v.d. Schans, 1992], maar minder geschikt voor
het
verzenden van informatie uit GIS bestanden. In Duitsland wordt
momenteel
trouwens hard gewerkt om EDBS te vervangen door zijn modernere
opvolger
[Brüggemann, 1991].
In Duitsland wordt verder veel gestandaardiseerd door middel van
ATKIS (15).
Maar let op, ATKIS is geen uitwisselingsformaat maar een
systeemstandaard. Dit
heeft te maken met het feit dat alle deelstaten in Duitsland over een
eigen
topografische dienst beschikken. Deze diensten wilden verder gaan dan
de
"tussentaal" oplossing en kozen voor de meest rigoureuze oplossing:
identieke
systemen.
Tenslotte is er nog GDF (16). GDF definieert een gegevensstructuur
die
gehanteerd wordt door een consortium van zeven bedrijven die
produkten
ontwikkelen ter verbetering van verkeer en vervoer en die voor deze
produkten
digitale wegenbestanden nodig hebben [Heres, 1991]. De
gegevensstructuur van
GDF is gebaseerd op die van NTF. Een aantal voor verkeersinformatie
overbodige
zaken zijn weggelaten (bijvoorbeeld: records voor rastergegevens en
records
voor de vastlegging van de positie van de tekst in het kaartbeeld) en
een
aantal zaken zijn toegevoegd, zoals records om associaties tussen
objecten te
kunnen weergeven [Heres & Wood, 1992].

Internationale syntax standaard

Men hoort in discussie over uitwisselingsstandaarden vaak de term "ISO 8211" vallen. ISO 8211 definieert echter geen volledige gegevensstructuur. Het is een syntaxstandaard die de hulpstructuren definiëert waarmee men de hogere orde structuren (records) uit de lagere orde structuren (velden en subvelden) kan opbouwen. ISO 8211 is een theoretisch zeer streng onderbouwde specificatie die grote nadruk legt op de "zelf-beschrijving" van bestanden en daarmee een grote flexibiliteit toestaat. Een flexibiliteit die echter niet door iedereen op prijs wordt gesteld en vaak als gecompliceerdheid wordt ervaren. Hierdoor bestaat er een hoge drempel om ISO 8211 ook daadwerkelijk te gebruiken. DIGEST en SDTS hebben al vanaf het begin gekozen voor ISO 8211 als onderliggende syntaxstandaard. EDIGÉO biedt de gebruikers de keuze: hetzij ISO 8211, hetzij een andere (naamloze) syntax. De NTF, die aanvankelijk zijn eigen syntax definieerde zal als BS 7567 ook gebruik maken van ISO 8211, naast de bestaande syntax. Ook het GDF consortium heeft voor deze strategie gekozen.

Europees Conceptueel Gegevensmodel

Dit model is vorig jaar tot stand gekomen in een ad hoc werkgroep die bestond uit vertegenwoordigers van de Britse Ordnance Survey, het Franse Institut Géographique National, het Duitse Institut für angewandte Geodäsie en het GDF consortium. Het model is te zien in Figuur 4. Het model definieert drie zeer algemene geografische begrippen, te weten "feature" (geografisch object), "attribute" en "relationship" en een aantal betrekkingen tussen deze begrippen. Het model geeft verder aan dat een feature tot vier verschillende categorieën kan behoren: puntvormige, lijnvormige, vlakvormige en samengestelde. Vanuit deze feature-categorieën lopen verbindingen naar drie topologische entiteiten: de "node" (knooppunt), de "edge" (lijnsegment) en de "face" (vlak). Aan de betrekkingen tussen de verschillende entiteiten zijn beperkingen opgelegd. Deze beperkingen worden gesymboliseerd door de pijltjes boven de rechthoeken waar de "rolnamen" in staan en door de dikke zwarte punten op de rand van de entiteits-cirkels. De betekenis van deze symbolen wordt besproken in Intermezzo 2.

Het Conceptueel Gegevensmodel is een zeer algemeen model dat verder uitgewerkt moet worden om het voor een bepaalde toepassing geschikt te maken. In de eerste plaats dient de entiteit "feature" te worden uitgewerkt. Zoals "dier" het overkoepelende begrip is van een hiërarchische serie begrippen zoals "gewervelden", "zoogdieren", "knaagdieren" en "veldmuizen", zo is het begrip "feature" het overkoepelende begrip van een groot aantal "feature thema's" en "feature klassen". Thema en klasse worden "taxonomische niveaus" genoemd. Ook de begrippen "attribute" en "relationship" dienen verder te worden onderverdeeld in verschillende attribuuttypes en associatietypes. Verder dient voor ieder feature type te worden aangegeven of het een enkelvoudig of samengesteld is. Van samengestelde features zal moeten worden vastgelegd uit welke andere features ze zijn opgebouwd. Verder moet worden bepaald welk feature met welk attribuut kan worden verbonden en welke features in welke associaties kunnen voorkomen. Tenslotte zal moeten worden beschreven tot welke categorie een feature behoort. De ene toepassing zal het feature "gebouw" als vlakvormig willen beschouwen, een andere toepassing ziet het liever puntvormig.

Nationale en multinationale standaards voor geografische begrippen
Alle in dit artikel besproken nationale standaards, met uitzondering van de Britse NTF, definiëren naast gegevensstructuren tevens een verzameling featuretypes en attribuuttypes. Het aantal gehanteerde taxonomische niveaus varieert sterk. De Amerikaanse SDTS kent "uit principe" geen enkele hiërarchie, dus slechts één niveau. De Duitse "Objektartenkatalog" (onderdeel van ATKIS en geënt op de NATO Feature Catalogue) kent drie verschillende niveaus: Objektbereik, Objektgruppe en Objektart. De meeste besproken standaards definiëren geen associaties en geven niet aan uit welke features een samengestelde feature opgebouwd dient zijn. Tevens ontbreken definities tot welke categorie een feature behoort. De enige standaard die zich daar wèl over uitlaat is de GDF [Heres & Wood, 1992]. De GDF geeft tevens de beperkingen aan in de betrekkingen tussen feature en de andere entiteiten. Figuur 5 laat de onderlinge betrekkingen zien tussen de features die behoren bij het thema "wegen en veerverbindingen". Zie Intermezzo 2 voor nadere uitleg van het NIAM diagram.

CEN activiteiten

Afgelopen jaar heeft het CEN, het Comité Européen de Normalisation, het leven

geschonken aan twee nieuwe Technische Commissies (TC's) die van belang zijn

voor het geografische en kartografische vakgebied. Zij luisteren respectievelijk naar de vaak verwarring stichtende codenamen TC 278 en TC 287.

TC 278, (voluit "Verkeers- en Vervoers Telematica" geheten), is midden 1991 van

start gegaan en heeft sinds kort haar werkplan gereed. Deze TC bestaat uit 12

werkgroepen waarvan Werkgroep 7, Geografische en Wegendatabases, voor kartografen het meest van belang is. Deze werkgroep heeft zich tot

doel gesteld

om de volgende standaards te ontwikkelen:

- Conceptueel Model
- Feature Catalogus
- Attributen Catalogus
- Associatie Catalogus
- Catalogus van metagegevens
- Representatie Schema
- Gegevensstructuur
- Individualiteit en Identificatie (van features)

Het Conceptueel Model zal gebaseerd zijn op het Europese Conceptueel Gevensmodel dat in de vorige paragraaf is besproken. Het onderwerp "Individualiteit en Identificatie" is een nieuw onderwerp (zie volgende

paragraaf). De andere standaards zullen gebaseerd zijn op het werk dat in het

GDF consortium is verricht.

TC 287 (Geografisch Informatie geheten) is in januari van dit jaar van start

gegaan. Haar werkplan is nog niet geheel gereed. Het is echter waarschijnlijk

dat de volgende werkgroepen zullen ontstaan:

- Organisatie en Strategie
- Conceptuele Modellen (inclusief een kwaliteitsmodel)
Data Descriptive Languages
- Mechanismen voor Gegevensuitwisseling
- Referentiesystemen voor plaatsbepaling

"De categorieën en grondvormen die we in de wereld van fenomenen menen te zien,

vinden we daar niet omdat zij iedere waarnemer in het gezicht staren; integendeel, de wereld trekt aan ons voorbij als een kaleidoscopische stroom

van indrukken die door onze geest moet worden geordend - en dit betekent:

grotendeels door de taalkundige systemen in onze geest. Lee Whorf (17)

Individualiteit en Identificatie

Dit onderdeel van de CEN TC 278 activiteiten wil ik wat nader toelichten. Het

betreft een vrij nieuw probleemveld dat vaak onderbelicht en vaak niet goed begrepen wordt. Ik heb het hier "Individualiteit en Identificatie" genoemd. In het Engels heet dit onderwerp meestal "Location Referencing", een term die volgens mij de lading niet volledig dekt. Waar gaat het eigenlijk om ? Dat kan het best aan de hand van een klein voorbeeld worden uitgelegd.

Stel dat A en B een bestand willen samenstellen van alle molens van Nederland en België. Om wat vaart achter de opbouw te zetten, besluiten zij de taken als volgt te verdelen: A verzamelt eigennaam, gemeentenaam, bouwjaar en hoogte; B legt zich toe op eigennaam, gemeentenaam, eigenaar, type en bouwkundige staat. Vervolgens voegen zij de gegevens samen, waarbij de combinatie eigennaam, gemeentenaam wordt gebruikt als unieke identificatie voor ieder individueel object (waarbij ze ervan uitgaan dat deze combinatie inderdaad altijd uniek is).

Nadat ze dit project succesvol hebben afgesloten, besluiten A en B een bestand te maken van alle parken in Nederland en België. Ze gaan weer op dezelfde manier te werk: A verzamelt eigennaam, gemeentenaam, oppervlakte en jaar van aanleg; B registreert eigennaam, gemeentenaam, eigenaar en soort beplanting. Wederom voegen ze de bestanden samen. Wat blijkt echter: een aantal parken die B heeft verzameld, ontbreken bij A en omgekeerd. Maar wat nog problematischer is: er zijn parken uit het deelbestand van A die B als twee verschillende parken heeft beschouwd en omgekeerd. Het is blijkbaar niet altijd duidelijk wat als één enkel park moet worden gezien. Hoe A en B de definities van "park" ook aanscherpen, de individualiteit van parken blijft problemen geven. A en B besluiten het daarom anders aan te pakken: het bestand van A wordt als uitgangspunt genomen en B verzamelt gegevens over de parken die door A als zodanig zijn geïdentificeerd. Met andere woorden: A stelt vast wat de individuele parken zijn en hoe ze worden aangeduid en B conformeert zich daaraan. En nu gaat het wel goed. De conclusie die we hieruit kunnen trekken is dat de "molenmethode" alleen werkt als:

- het onbetwistbaar is wat de individuele exemplaren uit een verzameling zijn.
- een identificatievoorschrift te geven is (bijvoorbeeld: eigennaam + gemeentenaam) dat voor ieder individueel exemplaar uit die verzameling een

naam oplevert die gegarandeerd uniek is binnen die verzameling. Molens voldoen aan beide eisen, maar zijn in feite uitzonderingen. Bij de meeste geografische objecten is het niet zonder meer duidelijk wat de enkele exemplaren zijn of is er geen gegarandeerd unieke identificatie te bedenken. Voor deze objecten is de "parkenmethode" de aangewezen oplossing. De parkenmethode is in feite een zeer klassieke manier die voor een aantal geografische objecten al eeuwenlang wordt gehanteerd. Neem bijvoorbeeld het "kadastraal perceel", dat zo'n centrale rol vervult in de onroerend goed registratie. Het is beslist niet zo dat kadastrale percelen "iedereen waarnemer in het gezicht staren", zoals Lee Whorf dat zo treffend uitdrukt (zie het citaat aan het begin van deze paragraaf). Een enkel kadastraal perceel is een enkel kadastraal perceel omdat de landmeter die dat perceel heeft opgemeten dat zo vond en omdat iedereen zich aan zijn keuze conformeert. Een ander goed voorbeeld zijn de straten en bijbehorende straatnamen uit de gemeentelijke stratenleggers. Ook deze objecten "staren de waarnemer niet in het gezicht", maar bestaan bij gratie van gemeentelijke straatnamcommissies die de individuele straten "bepalen" en "benoemen" en dankzij het feit dat iedereen zich daaraan conformeert. Ons hele adresseringssysteem zou in het water vallen als ieder er zijn eigen straatnaamgeving op na zou houden.

Overheid en industrie werken momenteel aan systemen die wegeninformatie, vastgelegd op een CD of geheugenkaart, combineert met verkeersinformatie die in gecodeerde vorm door de ether wordt gezonden [Heres, 1991]. Deze beide soorten informatie kunnen alleen succesvol samengevoegd worden, als die betrekking heeft op dezelfde verzameling wegelementen en als deze elementen bovendien op dezelfde manier geïdentificeerd worden. En omdat het niet altijd evident is wat een "enkel wegsegment" is, kan alleen de "parkenmethode" een oplossing bieden. Er is al een naam voor bedacht: het International Standard Road Number (ISRN). Nu nog de organisatie die de toekenning van deze nummers gaat organiseren en coördineren.

Samenvatting en conclusies

De ontwikkeling van standaards in het algemeen en internationale in het

bijzonder, is een tijdrovende bezigheid: er zijn vaak vele partijen bij betrokken, taalbarrières dienen te worden genomen en tegengestelde belangen dienen te worden overbrugd. De ontwikkeling van informatiestandaards heeft als bijkomende complicatie dat het soms zeer abstract is en snel tot misverstanden leidt.

Bij geografische informatie heeft men bovendien met het feit te maken dat het vaak om zeer alledaagse begrippen gaat als "stad", "weg" enzovoorts. Het is soms moeilijk om uit te leggen dat het juist dit alledaagse karakter de standaardisatie van deze begrippen zo lastig maakt. Desalniettemin is er de laatste jaren vooruitgang geboekt, in het bijzonder op het gebied van gegevensstructuren. De belangrijkste standaards lijken al sterk op elkaar en groeien nog steeds naar elkaar toe. Hun totale éénwording is slechts kwestie van tijd. Met de definitie van een Europees Conceptueel Gegevensmodel is een belangrijke mijlpaal bereikt. De nadruk dient vanaf nu te liggen op de standaardisatie van de geografische begrippen zelf. Dit zal ongetwijfeld een langdurige zaak worden. De eerste aanzet is echter al gegeven en het feit dat het Europese Normalisatie Instituut (CEN) er het belang van inziet, is in ieder geval een garantie voor een gestructureerd verloop van het proces.

Van een aantal geografische objecttypen, met name die objecten die gebruikt worden als gemeenschappelijke referentie bij het samenvoegen van bestanden, moet ook de individualiteit en de identiteit worden gestandaardiseerd. Om dat te bereiken zal er een aparte organisatie in het leven moeten worden geroepen.

Epiloog

Denkt u zich nog eens in: u bent directeur van een bedrijf dat zich gespecialiseerd heeft in gegevens voor transportlogistiek. U krijgt een opdracht om een bestand te leveren voor een transportbedrijf dat actief is in het Ruhrgebied. Daarom bestelt u ISO basisbestanden bij de topografische dienst van Nordrhein Westfalen. Van een ingenieursbureau uit Düsseldorf betreft u een ISO bestand over wegbreedtes en viaducthoogtes. Bij een ander bureau koopt u een ISO bestand over de maximaal toegestane belasting van bruggen. U geeft verder een aantal eigen medewerkers opdracht om nog wat gegevens te verzamelen

in een aantal onlangs gebouwde nieuwbouwwijken. Vervolgens integreert u de data op uw werkstation, converteert ze naar ISO en stuurt ze op naar uw klant. Ondenikbaar? Nu nog wel, maar in de toekomst waarschijnlijk niet meer.

INTERMEZZO 1: Gegevensstructuren, symbolen en betekenissen

Het klinkt misschien wat vreemd, maar een gegevensbestand kan in feite worden gezien als één groot teken, één supersymbool. Dat symbool kan men beschouwen als een gestructureerde verzameling deelsymbolen, die ieder op zich ook weer zijn opgebouwd uit deelsymbolen, enzovoorts. In de informatica noemt men zo iets een geneste structuur. Dit verder uitsplitsen kan men een groot aantal malen herhalen (10 maal is geen uitzondering), totdat men op structuren stuit die nog maar slechts twee verschillende waarden kunnen aannemen (meestal aangeduid met "0" respectievelijk "1"). Deze elementaire structuur is de binaire eenheid (bit), de kleinste bouwsteen waaruit ieder digitaal bestand is opgebouwd. Op deze manier beschouwd is ieder symbool een aantal "stappen" van de bitsymbolen verwijderd. Van symbolen die zich op een gelijk aantal stappen van het bitniveau bevinden, wordt gezegd dat ze zich op één en hetzelfde symbool- of structuurniveau bevinden. Symbolen van naburige niveaus, vormen samen een gegevensstructuur. Ieder symbool, ongeacht het niveau, verwijst naar een bepaald begrip (concept, entiteit, betekenis). Bij iedere verzameling symbolen van een bepaald niveau behoort een verzameling begrippen waarnaar deze symbolen mogelijkwijs verwijzen. Bij een goed bestandsontwerp bevat deze verzameling gelijksoortige begrippen, bijvoorbeeld geografische objecten (dorp, huis) of wiskundige entiteiten (punt, lijn) of karakters (letters, cijfers). Men spreekt dan van een betekenisniveau dat bij dat symboolniveau hoort. Het is belangrijk een onderscheid te maken tussen het symbool als type, namelijk de verzameling van alle gelijksoortige exemplaren (bijvoorbeeld alle symbolen die er uitzien als {45673 23452}) en het symbool als enkel exemplaar (de {45673 23452} zoals hij op één aanwijsbare plaats in een bestand voorkomt). In een correct bestand, verwijst ieder exemplaar naar precies één begrip. Echter, twee verschillende exemplaren behorend tot hetzelfde type mogen naar

twee verschillende begrip verwijzen. Naar welk begrip een bepaald exemplaar verwijst, hangt af van de hogere orde structuren waarvan dat exemplaar deel uit maakt (hetgeen dus impliceert dat symbolen van hetzelfde type tot meerdere hogere orde symbooltypen kunnen behoren). Dit nu is de fameuze context. Andersom kan ook een bepaald begrip door meerdere verschillende symbolen worden aangeduid, zelfs binnen bestanden die tot één type behoren, al zal het aantal mogelijkheden binnen een bestandstype in zijn algemeenheid beperkt zijn. Beschouwen we echter bestanden van verschillend type, dan zien we dat het aantal mogelijkheden om een begrip weer te geven, bijna onbeperkt is. Om symbolen samen te kunnen smeden tot symbolen van hogere orde, moet op één of andere manier gebruik gemaakt worden van hulpsymbolen (die op zich uiteraard weer deel uitmaken van een bepaalde structuur). Dit specifieke symboolgebruik wordt meestal aangeduid als de syntax van een bestand. Wie opmerkt dat er opvallend veel parallellen bestaan tussen gegevensbestanden en natuurlijke talen, heeft gelijk. Beide zijn semiotische systemen en voldoen derhalve aan een aantal basiswetten. Wie in bovenstaande zinnen het woord "symbool" vervangt door "letter", "woord", "zin" of "alinea" en "bestandstype" door "taal", zal zien dat deze uitspraken nu ineens natuurlijke talen betreffen.

INTERMEZZO 2 Toelichting bij de NIAM diagrammen

Figuren 4 en 5 in dit artikel zijn gemaakt volgens Nijssens Informatie Analyse Methode (NIAM). In tegenstelling tot een aantal andere methodes, is deze methode goed gedocumenteerd en bestaan er geschikte boeken die stap voor stap aangeven hoe de methode moet worden toegepast. Persoonlijk kan ik het boek van Nijssen en Halpin aanraden [Nijssen & Halpin, 1989]. Het bezwaar kan wellicht zijn dat het in het Engels is geschreven. Een alternatief is het Nederlandstalige boek van Wintraecken, dat echter wat ouder is en naar mijn smaak wat te vaak in herhaling vervalt [Wintraecken, 1987].

NIAM maakt allereerst een fundamenteel onderscheid tussen "entiteiten" en "labels". Het is in feite hetzelfde als het verschil dat in dit artikel tussen "begrippen" en "symbolen" wordt gemaakt. In een NIAM diagram worden entiteiten altijd door een volle cirkel aangegeven en labels door een onderbroken cirkel.

Zie Figuur 6.

Tussen entiteiten kunnen bepaalde betrekkingen bestaan. Deze betrekkingen worden in NIAM "feittypen" genoemd. Deze feittypen worden grafisch voorgesteld door twee entiteitscirkels met elkaar te verbinden. In een feittype speelt iedere entiteit een "rol". Deze rollen worden voorgesteld door de rechthoekjes op de verbindingslijn tussen de entiteiten. De rolnamen staan hier in geschreven. Figuur 6 van links naar rechts lezend, levert dan zinnen op als:

Village met de naam X bevat House met het adres Y. Andersom kan men lezen:

House met adres Y staat in Village met naam X.

Het pijltje boven de rol "is in" geeft aan dat ieder House slechts één keer

vermeld mag worden in de relatie tussen de entiteiten Village en House, met

andere woorden: ieder House staat slechts in één Village. Boven "contains"

staat géén pijl wat inhoudt dat een Village meerdere exemplaren van House mag

bevatten (je zou er eigenlijk aan kunnen toevoegen: moet bevatten, omdat er

anders geen sprake is van een Village).

De entiteitscirkel van Village heeft een zwarte punt in de verbinding naar

House. Dit geeft aan dat deze rol "verplicht" is, met andere woorden ieder

Village heeft minstens één House. Bij House ontbreekt de zwarte punt: een House

kan bij een Village horen maar dat is geen vereiste (een alleenstaand huis

bijvoorbeeld).

Er zijn nog veel meer beperkingen die met NIAM gevisualiseerd kunnen worden,

maar het zou te ver voeren die hier allemaal te bespreken.

Geïnteresseerden

worden naar één van beide genoemde leerboeken verwezen.

Voetnoten:

(1) Voor het algemene geval van n bedrijven geldt dat er $n(n-1)$ programma's nodig zijn.

(2) Voor het algemene geval van n bedrijven geldt dat er $2n$ programma's nodig zijn.

(3) bit = binary unit, de basiseenheid uit het tweetalig stelsel

(4) ISO = International Organisation for Standardization

(5) De terminologie is ontleend aan de klassieke typemachine

(6) Deze term roep bij mij altijd beelden op van spionnenruil of van atomen

die onderling atomen uitwisselen

(7) CEN = Comité Européen de Normalisation

(8) NTF = National Transfer Format

(9) BS = British Standard

(10) SDTS = Spatial Data Transfer Standard

(11) DIGEST = Digital Geographic Information Exchange Standard

(12) EDIGÉO = Échanges de Données Informatisés dans le domaine de l'information géographique

(13) EDBS = Einheitliche Datenbankschnittstelle

- (14) SUF = Standaard Utwisselingsformaat
- (15) ATKIS = Amtliches Topographisch-Kartographisches Informationssystem
- (16) GDF = Geographic Data Files
- (17) Uit: [Crystal, 1987], vertaling van auteur

Literatuur

Brüggemann H. (1991), German Developments on Spatial Data Exchange Standards. In : Moellering H. (red.) Spatial Database Transfer Standards pp. 99-111. London: ICA/Elsevier.

Crystal D. (1987), The Cambridge Encyclopedia of Language. Cambridge: Cambridge University Press.

Haywood P. et al. (1987), The National Transfer Format, Release 1.0. Southampton: Ordnance Survey.

Heres L. (1991), Via Incognita - Stand der techniek op het gebied van Reisinformatiesystemen. Kartografisch Tijdschrift XVII.3 pp.19-25.

Heres L., de Winter N. (1990), Towards a standardized European road network database. In "Transportation 17: 301-312". The Netherlands: Kluwer Academic Publishers.

Heres L., Wood T. (1992), GDF, a Lingua Franca for Geographic Information. In: The Proceedings of the Conference on Digital Mapping and Navigation (EurNav'92). London/Düsseldorf:RIN/DGON.

ISO 646, Information processing - ISO 7-bit coded character set for information interchange

ISO 2022, Information processing - ISO 7-bit and 8-bit coded character sets - Code extension techniques

ISO 4873, Information processing - ISO 8-bit code for information interchange - Structure and rules for implementation

ISO 6429, Information processing - ISO 7-bit and 8-bit coded character sets - Additional control functions for character-imaging devices

ISO 6937/2, Information processing - Coded character sets for text communication - Part 2: latin alphabetic and non-alphabetic graphic characters

ISO 8211, Information processing - Specification for a data descriptive file for information interchange

ISO 8859, Information processing - 8-bit single-byte coded character sets

Moellering H. (red.) (1991), Spatial Database Transfer Standards - Current International Status. London: International Cartographic Association/Elsevier Applied Science

Molenaar M. (1992), Object-hiërarchieën en onzekerheid in GIS - of waarom is standaardisatie zo moeilijk ?. NGT Geodesia 92-5, pp. 199-205.

Nijssen G.M., Halpin T.A. (1989), Conceptual Schema and Relational Database Design - A Fact Oriented Approach. New York: Prentice Hall.

Rossmeissl H.J., Rugg R.D. (1991). The United States Spatial Data Transfer Standard. In : Moellering H. (red.), Spatial Database Transfer Standards pp. 204-217. London: ICA/Elsevier.

Salgé F. (1991), La Normalisation des Échanges d'Informations Géographiques en France. In : Moellering H. (red.), Spatial Database Transfer Standards pp. 85-99. London: ICA/Elsevier.

van der Schans R. (1992), In plaats van Kaarten; I. Digitale modellen. Kartografisch Tijdschrift XVIII.1, pp. 34-43.

Sowton M. (1991), The National Transfer Format. In : Moellering H. (red.), Spatial Database Transfer Standards pp. 191 - 204. London: ICA/Elsevier.

Tardieu H, Rochfeld A., Colletti R. (1983), La Méthode Merise. Les éditions d'organisation;

Wintraecken J.J.V.R. (1987), Informatie-analyse volgens NIAM in theorie en praktijk. Schoonhoven:Control Data - Academic Service

Summary

The development of standards, and in particular, international ones takes time, because of the many parties that are involved, the linguistic obstacles that have to be taken and the opposing interests that have to be bridged. Standardization of the representation of information has an additional complication that it is sometimes very abstract and very easily leads to misunderstandings. Geographic information has the extra difficulty that it deals with very

commonplace concepts as "towns", "roads" and so on. It is often hard to explain that it is exactly this commonplace characteristic of geographical concepts, which makes standardization so difficult. Nevertheless, progress is made, in particular in the field of data structures. The major standards currently in use are already similar in this respect and still continue to converge. With the definition of a common Conceptual Data Model an important milestone has been reached. The emphasis of the geographic standardization now can be shifted to the standardization of the geographical concepts themselves. These activities will take much time and effort. The kick-off however, has been given and the fact that the European Standardization Institute (CEN) recognizes the importance of these standards, guarantees a structured course of the discussions.

The standardization of feature instances in lieu of feature classes, will become an other important item in the near future, in particular for features as roads. However, this standardization cannot be achieved in the form of writing a document only, but has to be supported by an coordinating organisation that distributes responsibilities.

Onderschriften figuren:

Figuur 1: Zonder een gemeenschappelijke tussentaal zijn er $n(n-1)$ conversieprogramma's nodig

Figuur 2: Bij gebruik van een gemeenschappelijke tussentaal zijn er $2n$ conversieprogramma's nodig

Figuur 3: De ISO 646 7-bit karakter verzameling.

Figuur 4: Het Europese Conceptuele Gegevensmodel. Objecten behoren to een

objektklasse en object Klassen tot een objektthema. Objecten kunnen attributen hebben en kunnen deelnemen in een associatie. Objecten zijn enkelvoudig of samengesteld. Een samengesteld object is opgebouwd uit meerdere enkelvoudige objecten of uit andere samengestelde objecten. Enkelvoudige objecten zijn hetzij punten hetzij lijnen hetzij gebieden. Een punt is opgebouwd uit één knoop. Een lijn is opgebouwd uit één of meer zijden. Een gebied is opgebouwd uit één of meer

vlakken. Zie Intermezzo 2 voor een nadere toelichting van de NIAM diagrammen.

Figuur 5: Het gegevensmodel voor de objectklassen die behoren bij het thema

"Wegen en Veren". Er zijn twee generalisatieniveaus gemodelleerd. Het gedetailleerd niveau bestaat uit de objectklassen: Weg Element, Overweg, Aansluiting, Veerverbinding en Besloten Verkeersgebied. Het geaggregeerde niveau bestaat uit de objectklassen: Weg, Kruising en Veer.

Zie Intermezzo 2 voor een nadere toelichting van de NIAM diagrammen

Figuur 6: Een eenvoudig gegevensmodel dat twee objectklassen definieert: Huis en

Dorp. Dorpen worden in het bestand weergegeven door een Dorpsnaam, Huizen door een Adres.

Noot over de auteur:

L. Heres is werkzaam bij Philips Car Systems en verbonden aan het CIS-lab (laboratorium voor autonavigatie) in Eindhoven. Hij is actief betrokken bij de ontwikkeling van standaards voor digitale wegenkaarten en digitale verkeersinformatie. Hij is lid van de CEN TC 278 Werkgroep 7: Geografische Databases.