

Ontwikkelingen voor metadata concepten

Prof.ir. Henri J.G.L. Aalders
Technische Universiteit Delft, OTB
h.j.g.l.aalders@geo.tudelft.nl
Katholieke Universiteit Leuven, Faculteit Toegepaste Wetenschappen

Introductie

Oorspronkelijk werd geografische informatie geproduceerd en gebruikt in een geo-omgeving met deskundigen die zich hadden gespecialiseerd in de geografie, kartografie, geodesie, fotogrammetrie, remote sensing, zeebodemkartering, hydrografie, geologie, bodemkunde, ruimtelijke planning, architectuur, civiele techniek, milieukunde, etc. Hun werk resulteerde in de kartografische representatie van de werkelijkheid eerst in kaartvorm en later als een digitaal computer model. Combinaties van verschillende ruimtelijke datasets in verschillende schalen, projecties en coördinaatsystemen, inhoud en verschijningsvorm creëert een gezamenlijk computermodel van de aarde als geheel of in delen. Dit omvat de aggregatie van digitale geografische, topografische en thematische gegevens (zoals ortho-fotokaarten, stadskarten, lucht- en ruimtelijke fotobeelden, (3D-)terrein modellen met gravimetrische, fotogrammetrische en kartografische informatie) geïntegreerd en geordend met een gemeenschappelijk concept zoals een geodetische basis, inhoud, ontwerp en referentie informatie en opgeslagen in digitale vorm.

Om geografische informatie geschikt te maken voor gebruikers, pasten deze deskundigen maatregelen toe met een conceptuele en methodologische basis, wettelijke regelingen en normen voor formattering en gegevens overdracht. Ook zijn implementaties op het gebied van hard- en software ontwikkeld en technologieën om de gegevens over te dragen naar gebruikers (Moellerling, 1991), (Moellerling, 1994), (Moellerling 1997), voor de opslag, analyse en verwerking en kartografische presentatie van gegevens.

In het dagelijkse leven zijn we gewend om de labels van gekochte producten te lezen om verschillende soorten informatie te verkrijgen, zoals het land van herkomst, het toegepaste productieproces, de laatste geldigheidsdatum, garantiebepalingen, toegestane gebruik, enz. Deze informatie is gebonden aan wettelijke bepalingen. Voor geografische informatie bestaan dergelijke bepalingen niet. Maar nu we in een digitale geo-informatie infrastructuur werken, die deze informatie beschikbaar stelt aan lekengebruikers, is universele informatie nodig over de locatie, de evaluatie, het verkrijgen en het toepassen van de gegevens.

Beschrijvingen van geografische datasets bestaan nu al enige tijd. Meestal zijn deze beschrijvingen ontstaan in verschillende omgevingen, in verschillende landen en voor verschillende doeleinden, waardoor ze vaak niet met elkaar zijn te vergelijken en weinig overeenkomst vertonen en niet uitwisselbaar zijn; tenminste in een internationale omgeving. Dit is de aanleiding geweest om een internationale norm te ontwikkelen voor ruimtelijke gegevenbeschrijvingen: ISO19115:2003, Geographic Information – Metadata.

Doelstelling van de ontwikkeling van ISO19115: 2003, Geographic Information - Metadata was o.m.:

- ondersteunen van het gebruik van ruimtelijke informatie;
- samenwerken met de informatie technologie bij de ontwikkeling van de norm;
- wereldwijde toepassing van ruimtelijke gegevens bevorderen over de landsgrenzen heen en in een meertalige omgeving;
- baseren op de fundamentele van nationale en regionale ontwikkelingen en ervaringen door een diepgaande analyse uit te voeren en implementaties te testen.

ISO19115:2003 Metadata definieert een set metadata elementen en karakteristieken, evenals het schema om metadata volledig te documenteren. De norm kan op alle ruimtelijke gegevens worden toegepast: datasetseries, datasets, individuele ruimtelijke objecten, en attributen. De norm definieert ook een minimum set metadata, maar laat ook optionele metadata elementen toe om een meer uitgebreidere dataset te kunnen beschrijven.

In de IT wereld bestaan nog twee andere prominente metadata initiatieven:

- Dublin Core Metadata Initiative (DCMI, Dublin Ohio, U.S.A, metadata element set (zie <http://www.xml.com/pub/2000/10/25/dublincore/>) is bedoeld om zoekopdrachten naar verschillende typen onderwerpen uit te voeren en te verkrijgen (ofwel de basis: in elkaar grijpende bouwstenen). Het kan gezien worden als een voorbeeld metadata zoektaal dat de gebruiker laat navigeren door van elkaar afwijkende onderwerpen, in verschillende talen en culturen. Het wordt veel toegepast in gouvernementele omgevingen, musea, bibliotheken, milieuwetenschappen, drukkerijen, handel, gezondheidszorg, financiële sector, enz., waaruit het succes wel blijkt.
- Institute for Electrical and Electronic Engineers (IEEE) Learning Object Metadata (LOM) Working Group, zie <http://grouper.ieee.org/p1484/doc/wg12/LOM-WD3.htm> voor technische toepassingen.

Geschiedenis van metadata

Voor velen zijn de concepten gegevensbeschrijving, metadata, legenda en marginalia hetzelfde: allen beschrijven ze de eigenschappen en karakteristieken van datasets.

Vroeg in de jaartelling pasten de Griekse geografen al beschrijvingen van kaartobjecten toe. Veel middeleeuwse geografen maakte prachtige tekeningen om hun visie van de wereld in kaarten weer te geven: veelal met religieuze motieven en altijd met de auteursnaam. Vanaf het midden van de 18^e eeuw topografische, zee- en navigatiekaarten lieten beschrijvingen en verklaringen zien om een betere kaartinterpretatie mogelijk te maken; dit werd kaartmarginalia genoemd, omdat de informatie vaak in de marge van de kaart werd vermeld. Marginalia vermeldde o.m. verkenningsdata, begrenzingen, coördinaatsystemen, schaalinformatie, kwaliteit, uitgever of eigenaar, enz.

ISO/TC 211, de auteur van de ISO19115:2003, Metadata norm, omschrijft metadata met: "Gegevens over de inhoud, kwaliteit, conditie en andere karakteristieke gegeven van een dataset".

De ICA Commissie voor Standaarden voor Ruimtelijke gegevens gebruikt de omschrijving: "informatie over een ruimtelijke dataset voor een beter begrip van ruimtelijke gegevens".

Volgens deze definitie bevat metadata: templates voor overdrachtformaat, tellingen van data items in de overdracht, conceptuele gegevensmodellen, catalogi, referentiesystemen, kwaliteitsrapportage en de logische beschrijving van de metafile.

De term **metadata** * – of ook gebruikt in andere vormen, zoals *meta data* (McIntosh, 1968), *meta-data* (Homer, 1978), (Ziegler, 1978), (Weber, 1978), (Schelling, 1978) en *metadata* (DLC, 1979) – dook op in de Computer wetenschappen/Informatie literatuur in 1968. Sindsdien wordt de term gebruikt in alle drie vormen: metadata werd populair in Europa in het begin van de zeventiger jaren en werd later in Amerika gebruikt in militaire documenten als één woord: meta-data.

Theoretische aspecten voor metadata

Volgens de concepten uit de verzamelingentheorie, kan een verzameling geografische gegevens verschillende andere homogene verzamelingen bevatten met een of meerdere exemplaren van objecten, attributen en/of relaties. Daarmee kan men onderscheid maken in verschillende hiërarchische niveaus van verzamelingen. Bovendien blijkt het voor de toepassing van dergelijke verzamelingen noodzakelijk om ook de beschrijving van de inhoud van iedere verzameling te geven. Naast een gegevensbeschrijving van de verzameling bevat de metadata ook een beschrijving van de beheerorganisatie, en het gegevensonderhoud, van zowel de verzameling van geografische gegevens, als van de metadata van de verzameling. Complete metadata beschrijvingen van de structuur, inhoud en nauwkeurigheid van alle gegevensverzamelingen is een belangrijke eis voor het ontwerpen van een gegevensbestand om bescherming en misbruik van de verzameling te voorkomen.

Een metadatamodel omvat de definitie van verplichte informatie, mogelijke metadata en conditionele metadata (metadata die verplicht is onder bepaalde omstandigheden). Bovendien moet de minimale metadata gegevensset worden weergegeven (meestal een opsomming van de verplichte metadata).

Metadata kunnen worden opgeslagen vanuit verschillende principes, bijvoorbeeld vanuit:

- de managementstrategie;
- financieel oogpunt;
- institutionele of technische achtergrond.

De beste tijd om metadata te definiëren en te verzamelen is tijdens het verzamelen van de gegevens zelf, omdat afwachten tot na de gegevens invoer zal resulteren in een slechtere beschrijving van de te verzamelen gegevens of tot een kostenverhoging omdat later aanvullende informatie moet worden verzameld. Ook kan het zijn, dat achteraf bepaalde details worden vergeten.

Het toepassen van gegevens uit verschillende bronnen, en mogelijk ook het toepassen van processen op gegevens op verschillende plaatsen door een GRID-computing benadering met gedistribueerde verwerking maakt nieuwe en innovatieve applicaties mogelijk. Dit zou kunnen uitmonden in een Virtueel GIS (V-GIS omgeving), waarin de Digitale Wereld, opgeslagen in bestanden, naadloos kan worden bestudeerd, gebruikmakend van het gehele, gecombineerde spectrum van karakteristieken, die opgeslagen zijn in de afzonderlijke bestanden en door processen die op verschillende computers beschikbaar zijn. Dit concept is onderwerp van discussie ge-

* De Metadata Company registreerde de term 'Metadata' in 1986. De Metadata Company verkoopt goederen en diensten gekoppeld aan metadata. Omdat de FGDC geen goederen of diensten verkoopt, vindt ze dat ze de registratie niet overtreedt zoals bepaald door het Department of the Interior, Office van de Solicitor (zie FGDC Official statement op http://www.fgdc.gov/metadata/meta_trademark.htm). Omdat de registratie goederen en diensten betreft, hebben verschillende bedrijven die metadata producten vervaardigen de neiging om de term metadata te vermijden en neigen ertoe de term metadata te gebruiken of zelf andere termen zoals gegevenseigendom of gegevenscatalogisering.

weest tijdens de G8 conferentie in 2001 met de Verenigde Naties, verschillende staatshoofden, non-profit organisaties en internationale, particuliere instellingen.

Hiërarchie in standaardisatie

In sommige landen wordt de toepassing van normen en standaarden bij wet geregeld (bijvoorbeeld Portugal en Rusland), terwijl in andere landen het gebruik ervan wordt overgelaten aan het particuliere initiatief (bijvoorbeeld Nederland en Zweden). Landen waar de toepassing bij wet wordt geregeld maken onderscheid tussen:

- internationale normen, meestal ontwikkeld door de ISO, terwijl regionale normen worden ontwikkeld door een regionale internationale organisatie zoals Comité Européen de Normalisation (CEN). Veel nationale regeringen en norminstituten nemen deze normen over in hun nationale normen bibliotheek.
- nationale normen aangenomen door de (Staats-)Normcommissie, een normalisatiewet of een nationaal decreet (Koninklijk Besluit);
- domeinstandaarden die worden gebruikt door verschillende (overheids-)organisaties of een groep organisaties in hetzelfde werkgebied;
- toepassingsstandaarden, toegepast binnen een specifiek bedrijf.

Veel normen staan niet op zichzelf maar maken onderdeel uit van een serie normen die ieder hun eigen aspect van de normalisatie in een bepaald toepassingsveld beschrijven (ook wel een normen complex of serie normen genoemd, zoals de ISO normen 191xx voor Geografische informatie).

Omdat normen onder andere ook de objectdefinities bevatten (in de geografische informatie ook wel ontologie genoemd) en het moeilijk is om voor alle voorkomende toepassingen de definities te harmoniseren (of identiek te maken), wordt vaak uitgegaan van een hiërarchie in de objectdefinities, waarbij bij een lager niveau in de hiërarchie de genormaliseerde definitie uit een hoger niveau wordt geïmplementeerd en eventueel nader gepreciseerd. Daarbij wordt voorgaande indeling aangehouden. Overdracht van gegevens kan dan ook alleen maar plaatsvinden binnen een bepaald niveau of hoger.

Normen kunnen verschillende aspecten voorschrijven voor de levering van gegevens, ontologie, kwaliteit, conceptuele modellen, enz. Normen worden gebaseerd op gegeneraliseerde resultaten van wetenschappelijk, technisch en toegepast onderzoek en als resultaat van de noodzaak in de praktijk. Bovendien is de intentie dat er publiek voordeel mee gedaan kan worden. Normen voor metadata vormen daar geen uitzondering op.

Metadata modulariteit en flexibiliteit

Een ander concept bij metadata normen is de **modulariteit** (Duval, 2002): metadata modulariteit is een fundamenteel principe dat ontwerpers toestaat nieuwe samenstellingen te creëren, gebaseerd op bestaande metadata schema's: in een modulair systeem kunnen metadata elementen uit verschillende gebieden worden samengevoegd zowel syntactisch als semantisch. Daarmee kunnen metadata gegevens hergebruikt worden in grotere verbanden zelfs als de semantiek niet overduidelijk te combineren is, maar er wel een gemeenschappelijk syntactische fundament onder ligt.

Metadata modulariteit kan er ook voor zorgen in een specifieke noodzaak binnen een bepaalde toepassing wordt voorzien. Zo zal bijvoorbeeld de *identificatie* bij elke dataset voorkomen, terwijl de *graad van bewolking* alleen bij luchtopnames en remote sensing beelden wordt weergegeven.

Bovendien verlangt metadata modulariteit voldoende verfijning om meer specifieke informatie kenbaar te maken; bijvoorbeeld *auteur*, *illustrator*, *componist of beeldend kunstenaar* zijn een verbijzondering van een meer algemene term *ontwerper*.

Toepassing van meerdere talen en culturen wordt ook steeds belangrijker nu Internet zich aandient voor de distributie van (ruimtelijke) gegevens: Internet zal aan betekenis inboeten indien we niet in staat zijn om gebruikers in hun moedertaal te bedienen. Hier komen twee tegengestelde begrippen als **internationalisering** en **lokalisering** naar voren. Zoeken naar gegevens via Internet over de gehele wereld is het best gediend bij een internationalisering (met gebruik van gemeenschappelijke afspraken, taal en karaktersets), terwijl de behoefte van een specifieke gemeenschap beter is gediend met lokale afspraken.

Een oplossing kan zijn om in de internationale metadata architectuur automatische vertalingen te laten plaatsvinden in verschillende omgevingen, culturen en talen. DCMI (Dublin Core Metadata Initiative) onderhoudt hiertoe een lijst met vertaalde documenten en de European Workshop on Learning technologies doet hetzelfde voor de LOM specificaties.

Tenslotte vormen namespaces een integraal onderdeel van de Web infrastructuur. Namespaces worden gebruikt in bevestigingen via het Internet als onderdeel van een modulaire metadata-systeem. Zo kan bijvoorbeeld in een bevestiging via het Web de URL de prefix bevatten *dc* om een Dublin Core metadata gestructureerde elementset te zoeken terwijl *lom* een zoekopdracht van metadata volgens de IEEE-LOM structuur uitvoert.

Nominale grondslag

Iedereen neemt zijn omgeving waar en creëert daar een model van volgens eigen inzichten. Ruimtelijke bestanden streven een objectieve digitale voorstelling van de werkelijkheid na door IT-oplossingen toe te passen, maar ook ruimtelijke computer analyses met grafische voorstellingen. Echter, iedere beschrijving van de werkelijkheid is altijd een abstractie van de objecten die waarneembaar zijn in de werkelijkheid en die abstractie is altijd subjectief: het is een van de vele voorstellingen die we ervan kunnen maken. Bovendien is het niet een exacte kopie van de werkelijkheid: zaken worden benaderd en gedeeltelijk voorgesteld, gesimplificeerd of soms zelf niet voorgesteld. Om er zeker van te zijn dat een gegevensbestand niet misbruikt wordt door verkeerde interpretatie, aannames of beperkingen moet elk bestand volledig worden beschreven voor wat de inhoud betreft: de metadataset.

De nominale grondslag wordt in het Engels: 'UoD, Universe of Discourse' genoemd. In de Verenigde Staten van Amerika, wordt meestal de term 'abstract view of the universe' gebruikt; dat geeft echter niet het selectieproces (zie hieronder) weer en dat is nu juist een integraal onderdeel van de nominale grondslag definitie. ISO/TR 9007, 1987 (ISO, 1946) definieert de UoD als "alle objecten of fenomenen van interesse voor een bepaalde toepassing, die de werkelijkheid in een abstracte, karakteristieke vorm beschrijven door de thematische, geometrische en temporele semantiek". Daarvoor wordt een **ontologie** gebruikt, waaronder in de geografische informatie wordt verstaan: *een definitie, classificatie en structuur van objecten, als een limitatieve opsomming van eenduidige uitgekozen - beschouwd vanuit een specifieke toepassing - concepten* (Uitermark, 2001). Deze definitie volgt de doctrine van de Griekse filosoof Parmenides (500 - 450 BC) en de Indiase Nyāya-Vaiśeṣika filosofie school op het gebied van ontologie. Op deze wijze bevat de nominale grondslag een limitatieve set karakteristieken van geselecteerde objecten uit de werkelijkheid met hun beschrijvingen. Objecten in de werkelijkheid worden in een gegevensbestand gerepresenteerd en daardoor bestaat er een één-op-één relatie tussen de bestand objecten en de werkelijkheid. Het proces om de objecten uit de werkelijkheid in een bestand te representeren bestaat uit drie stappen:

1. selectie van objecten uit de werkelijkheid die in aanmerking komen voor een bestandrepresentatie. Door deze selectie, uitgevoerd door de potentiële gebruiker, wordt het bestand direct geënt op de gebruikers wensen en waarschijnlijk allen voor zijn – beperkte – doeleinde toepasbaar. Dat geeft een productbeeld, zowel emotioneel als cognitief.
2. abstractie van die objecten, dat resulteert in de inhoud van het te vormen bestand;
3. meting van de objecten en opname in het bestand volgens een gespecificeerde werkwijze.

Toepassing van metadata

Geen enkele metadataset kan alle functionele eisen voor alle toepassingen behappen, terwijl met de ingebruikname van het Web de toegang tot datasets steeds eenvoudiger wordt. Daarom wordt het steeds belangrijker om modellen te ontwikkelen voor bestandbeschrijvingen en bijbehorende zoekstrategieën, door het ontwikkelen van:

- *Conceptuele modellen* die de nominale grondslag definiëren, nodig voor de representatie van de syntax en semantiek van metadata, als vervanging van de tekstuele beschrijvingen, lijsten met objecten en attributen, die momenteel nog zo vaak in metadata worden aangetroffen en gemakkelijk tot onjuiste interpretatie leiden. Een nominale grondslag bevat daarom object(-typen) met hun attribuut(-typen), en de relatie(-typen) tussen die objecten met hun attribuut(-typen) en de betreffende domeinen voor alle attributen evenals het gedrag van de objecten in een objectgeoriënteerde model.
- *Syntax en Semantiek*. Syntax gaat over de vorm, semantiek over de betekenis. Er moeten over beiden afspraken worden gemaakt om metadata te kunnen overdragen. Syntax definieert de wijze waarop de gegevens worden overgedragen, terwijl semantiek definieert wat er wordt overgedragen. Er bestaat nog een derde component in de interoperabiliteit, dat hierboven uitstijgt: de inhoudelijke woordenschat (lexicon). Die kan zijn in natuurlijke taal, zoals in het Nederlands of Engels, maar ook in specifieke formele taal die de betekenissen van beschrijvingen in de overdracht verder beperkt en daarmee de kwaliteit van de overdracht verhoogt.
- *Toepassingsprofielen*, een combinatie van metadata elementen uit verschillende schema's die de principes van modulariteit en uitbreidbaarheid mogelijk maakt, maar is toegespitst op een specifieke toepassing. Dit wordt mogelijk door:
 - definitie van de *cardinaliteit*, waarbij de beperkingen elk metadata element worden aangeduid (als verplicht, facultatief of conditioneel);
 - gebruik van *waarde beperkingen* voor specifieke toepassingen;
 - *relatie- en afhankelijkheidsspecificaties* tussen verschillende object(-type)en.
- *Declaratie van namespaces* als er meerdere worden ondersteund.

Een van de eerste stappen bij de ontwikkeling van een gegevensmodel is de analyse van de gebruikerswensen, gebaseerd op het toekomstige potentieel gebruik van de gegevens, resulterend in een set externe schema's voor iedere mogelijke toepassing. Het conceptuele schema voor het informatiesysteem is de combinatie van al deze externe schema's. Echter, informatiesystemen en hun toepassingen hebben een beperkte levensduur en waarschijnlijk zullen nieuwe ideeën en technologieën vereisen dat nieuwe conceptuele modellen worden ingevoerd.

Metadata elementen set

Metadata normen zijn een middel om de inhoud van informatiesystemen te beschrijven op verschillend detailniveau. Het meest vage niveau is een *samenvatting* of *overzicht* met een korte inhoudsopgave van het informatiesysteem, meestal in natuurlijke taal en ongestructureerd. Dit

geeft de leverancier van de metadata wel veel vrijheid, maar het gebruik wordt dan wel afhankelijk van de mogelijkheden van de gebruiker, omdat in de natuurlijke taal inconsistente en dubbelzinnige uitdrukkingen voorkomen, die vaagheid en onbegrip in de hand werken ondanks een goed gedefinieerde grammatica en daardoor automatische interpretatie bemoeilijken. Een lijst met object- en attribuutbenamingen inclusief de relatietypen en de bestaande hiërarchieën geeft al een veel betere automatische interpretatie; vooral als ook een conceptueel schema wordt meegeleverd. Daarom worden in de metadata elementenset ontwikkeld; men kan daarin drie typen element onderscheiden:

1. *Metadata van metadata*, die gaan over de beschrijving en definitie van metadata (let wel dit zijn niet de metadata zelf, maar gegevens om de metadata te begrijpen), zoals de gebruikte taal voor de metadata, de karakterset en het coderingsysteem (bijvoorbeeld ASCII), de referentiesystemen voor temporele, thematische en metrische attributen, het referentiekader voor de kwaliteitsdefinities en de administratieve informatie voor de metadataset (contactpunt, bijhouding enz.).
2. *Directory metadata* met de identificatie van de dataset, de afkomst informatie van de dataset, de plaats waar de dataset is te vinden en onder welke voorwaarden de dataset is te verkrijgen, adres van eigenaar, beheerder en distributeur en hun rol t.a.v de dataset, contactpunt voor informatie over de dataset, de gebiedsbeschrijving: zowel ruimtelijke als temporeel en thematisch, kwalitatieve informatie, en maatregelen voor de beveiliging van de dataset.
3. *Gegevensdefinitie* met een definitie voor de semantische, thematische geometrische en temporele eigenschappen van de objecten, conceptuele schema's voor de gegevensstructurering met objectclassificatie, verwijzingen naar externe thesauri.

Meestal worden metadata normen voorzien van een minimaal vereiste set van metadata elementen. Zo zegt de CEN ENv (Voornorm) 12657:1998, Geographic Information – Data description – Metadata dat tenminste een van de vermelde parameters van de metadata norm moet worden vermeld.

ISO19115: 2003, Geographic information – Metadata heeft een specifiek opgegeven minimale dataset die een opsomming is van de verplichte velden uit de complete metadataset.

Ontwikkelingen

De ontwikkeling van de ISO, Geographic information – Metadata norm was gebaseerd op studies van verschillende eerdere ontwikkelingen op het gebied van metadata normen, zoals:

- ANZLIC Metadata: Core Metadata Elements, 1995, toegepast in Australië en Nieuw Zeeland;
- Canadian Directory Information Describing Digital Geo-referenced Data Sets, 1994;
- CEN ENv 12657:1998, Standard for Geographic Information – Metadata, de Europese voor-norm voor metadata;
- FGDC Content Standard for Geo-spatial Metadata, 1994, de metadata norm voor de VS;
- DIGEST, 1994 voor toepassing binnen de NAVO;
- IHO S 57, 1995, ontwikkeld voor hydrografische toepassingen;
- Dublin Core Metadataset. Deze metadataset is oorspronkelijk ontwikkeld voor de biblio-theken, maar nu ook veel toepassing vindt in uitgeverij en overheden.

Metadata bevraging en distributie

Er zijn vele mogelijkheden om metadata te verspreiden, waarbij Interent een belangrijk hulpmiddel is geworden. Bij het zoeken op Internet naar gegevensbronnen kan men drie niveaus onderscheiden:

1. *Metadata ontdekken.* Daarvoor is nodig een set van gemeenschappelijke termen, die de inhoud van de informatiebronnen beschrijven, een standaard grammatica om deze termen te kunnen samenvoegen in begrijpbare zinnen en een raamwerk dat gebruikers in staat stelt de metadata over te dragen. Samen vormen ze een architectuur voor de beschrijving en overdracht van metadata bronnen op het Internet. In de eenvoudigste vorm bestaat het systeem uit een centrale database met beschrijvingen van verschillende dataset, waarin iedere leverancier via speciale programmatuur zijn eigen gegeven kan bijhouden. Daarnaast bestaan er ook vormen waarin de metadata bij de leverancier blijft en via Internet toegang tot de beschrijvingen wordt verschaft.
2. *Metadata interpreteren.* Hierbij zijn de formele afspraken over de inhoud van de metadataset van belang: gebruikers moeten een indruk kunnen krijgen van de inhoud van de dataset, waar de dataset is verkrijgen en de voorwaarden waaronder de dataset beschikbaar is voorbeelden hiervan zijn:
 - MEGRIN Geographical Data Description Directory service in Europe;
 - European Spatial Metadata Infrastructure (ESMI);
 - Asia-Pacific Spatial Data Infrastructure (APSDI) zoals voorgesteld door PC GIAP (Permanent Committee for Geographic Information in Asia en de Pacific);
 - Inter-American Geo-spatial Data Network (IGDN) in Latin America and the Caribbean;
 - National Geospatial Data Clearinghouse in the U.S.A., following the Executive Order 12906 van President W. Clinton van 4 april 1994;
 - Canadian Geospatial Data Infrastructure GEONet, ontwikkeld door GeoConnections;
 - Australian Spatial Data Infrastructure Distribution Network;
 - Japanse GSI clearinghouse gateway system met multi-byte karakter code sets uit het GEO profiel, die is gebaseerd op ISO 23950.

De meeste van de huidige normen bevatten metadata naast de geografische gegevens. Bijvoorbeeld CEN ENv12657,1998 Geographic Information – Data description – Metadata en de ISO 19115, 2001 Geographic Information – Metadata geven een beschrijving van gebruikte terminologie, metadata elementen en een schema, dat, indien juist toegepast, gebruikers in staat stelt gegevens te lokaliseren, toegang te verkrijgen, te evalueren, te selecteren, aan te schaffen en toe te passen op een efficiënte wijze. Bijkomend voordeel voor de leveranciers is dat ze de organisatie en faciliteiten van de gegevens voor hen ook regelen en ook de karakteristieke eigenschappen van de gegevens opsommen om ze te catalogiseren.

CEN ENv 12657 gebruikt EXPRESS en EXPRESS-G als modelleertaal, terwijl ISO19115 UML (Universal Modeling Language) toepast; ze zijn een integraal deel van de abstracte modellen voor geografische informatie, zijn computer leesbaar, hebben een grafische voorstelling en verschaffen een beeld van de klassentypen, attributen en het type en kardinaliteit van de relaties.

3. *Data verkrijgen en toepassen.*

Gebruik van metadata

In het gebruik van metadata zijn enkele opvallende voorbeelden bekend geworden (Ahonen-Raino, 2001), (McCelland, 2002), die ons het volgende leren:

1. Meestal achten gebruikers de verstrekte gegevens juist, wat ten onrechte blijkt te zijn, omdat metadata incompleet zijn en fouten bevatten. Onverwachte toepassingen zoeken naar informatie die niet bestaat en dan denken gebruikers dat de metadata onvolledig is.
2. Gespecialiseerde woordenboeken bemoeilijken de gegevens invoer en het begrip. Het gebruik van metadata door leken leidt vaak tot een Lego™ methode van bouwstenen en kinderen zijn niet bevooroordeeld door de bedoelingen van de leveranciers: zij passen dezelfde ontwerpen toe in zowel een waterrijke omgeving als in de woestijn. Leken kunnen dezelfde filosofie toepassen binnen een geografische omgeving.
3. Het is gebruikelijk dat normen tegelijk worden ontwikkeld met de toepassingen; dat geldt ook voor metadata. Conceptuele modellen zouden deze nieuwe toepassingen moeten volgen: standaarden zijn daarom ook nooit af!

Vanuit de gebruiker gezien is het belangrijk te weten hoe goed de garantie van de producent is voor de metadata informatie; dit is ook van belang voor de producenten omdat onjuiste metadata informatie hen duur kan komen te staan. De praktijk laat een tendens zien dat klanten metadata evalueren in de volgende volgorde, bedekking, actualiteit en volledigheid en kwaliteit (Jakobsson, 2002).

De Werkgroep Kwaliteit in CERCO (Comité Européen des Responsables de la Cartographie Officielle) heeft in 1998 een questionnaire gemaakt, die werd toegezonden aan de 32 nationale karteringsorganisaties in Europa. De resultaten zijn geheel beschikbaar op de website van EuroGeographic – <http://www.eurogeographics.org> en er kunnen enkele interessante conclusies uit worden getrokken:

- Gebruikerseisen worden door interviews met de klanten vastgesteld.
- De meeste datasets hebben een product specificatie, die is gebaseerd op 'in-house' ontwikkelde methodologie omdat internationale normen nog niet beschikbaar waren (nu ISO9115), hoewel enkele datasets op internationale productspecificaties als Corine of Digest, enz. waren gebaseerd. Productspecificaties omvatten:
 - gegevenswoordenboeken;
 - inhoudspecificaties;
 - inwinningsspecificaties;
 - technische specificaties;
 - gebruikershandleidingen.

Regionale en nationale metadata ontwikkelingen

In verschillende regio's in de wereld zijn specifieke ontwikkelingen ontstaan uit een behoefte naar metadata informatie.

Europa

De CEN ENv 12657:1998, Geographic Information – Data description – Metadata is alom geaccepteerd in Europa en in veel landen wordt de norm gebruikt voor metadata beschrijving. Pan-Europese projecten zoals Geographical Data Description Directory (GDDD), Geo-Scientific Electronic Information Exchange System (GEIXS), Added Value Information Dissemination voor hydrografische Datasets (AVID), European Spatial Metadata Infrastructure (ESMI), enz. passen CEN ENv 12657:2001 toe voor de distributie van gegevens.

Ook zijn er speciale projecten geweest die zich bezighielden met de veeltalige situatie in Europa, zoals Electronic Trade for Geographic Information (GISED), Methods for Access to Data and Metadata in Europe (Madame) en CLEF (Cross-Language Evaluation Forum). Het Open

Archive Forum ondersteunt projecten en nationale initiatieven met een open archief benadering als een interoperabel raamwerk met Open Archives Initiative (OAI) metadata protocol in Europa.

Door het opstarten van het Europese project INSPIRE in het zesde Framework Programme met de noodzaak voor de ontsluiting van vele Europese bestanden is een harmonisatie tussen de ISO19115:2003 en CEN ENv 12657:1998 nodig gebleken. Daarvoor is nog een formele weg af te leggen: CEN/TC 287 in het voorjaar van 2003 weer in het leven geroepen en de eerste stappen voor dat proces zijn al ingang gezet.

Noord Amerika

Tussen Canada en de V.S. hebben informele discussies geleid tot een MoU (Memorandum of Understanding), die ertoe moet leiden dat ANSI/INCITS L1 en SCC gezamenlijk een profiel van ISO191xx normen gaat ontwikkelen. Momenteel wordt in de V.S. de FGDC Metadata Content Standard for Digital Geospatial Metadata, STD-001-1998 Version 2.0. veelvuldig gebruikt.

Azië/Pacific

Versie 1 van de ANZLIC Metadata Guidelines werd gepubliceerd in 1996 gevolgd door 'Recommended Guidelines for the Transfer of ANZLIC Metadata Core Elements'. Deze twee documenten zijn door de industrie goed ontvangen en worden veelvuldig toegepast. In het gebruikt zijn echter enkele inconsistenties en tekortkomingen gevonden, waardoor het nodig was enkele aanvullende metadata elementen te definiëren in 1999. Dit werd gezien als een interim maatregel omdat de internationale norm 'ISO 19115, Geographic Information – Metadata' op zich liet wachten.

De Aziatische regio is zeer divers en strekt zich uit van Armenië, Iran, Azerbeidjaan en Kazakstan in het Noordwesten, tot Nieuw Zeeland, Frans Polynesië, Samoa en Cook eilanden in het Zuidoosten. PC GIAP is de centrale organisatie (met 55 landen leden) en heeft een Data Node project in 2000 gestart met metadata als een centraal thema en ISO19115:2003 als basis.

Centraal end Zuid Amerika en het Caribische gebied

De regio is zeer actief in de ontwikkeling van metadata als onderdeel van verschillende SDI ontwikkelingen waarbij vier regionale organisaties een rol spelen: PCIDEA (Permanent Committee on Spatial Data Infrastructure for the Americas), PROCIG (Central American Development Project for GIS), PAIGH (Pan American Institute of Geography and History) and DIGSA (Directory of Geographic Institutes from South America, Spain and Portugal). De meeste landen in de regio passen de FGDC standaard toe, maar ook hebben ze veelal besloten om die door de ISO19115 te vervangen.

Afrika en het Midden Oosten

In deze regio gebeurt niet veel op de ontwikkeling van metadata, ofschoon in Zuid Afrika wel enkele toepassingen van ISO 19115 zijn te vinden. De meeste ontwikkelingen liggen in projecten van de United Nations Environmental Programme (UNEP) en de Regional Remote Sensing unit of the Southern African Development Community in Harare (SADC).

Conclusies

De toepassing van GIS en dus ook metadata zijn globaler en grensoverschrijdend geworden, wat is te zien in veel toepassingen op het gebied van milieu, bosbouw, mariene en culturele bescherming, rampen bestrijding, enz. De verschillende regionale ontwikkelingen hebben ertoe geleid dat een internationale norm voor metadata ISO 19115: 2001, Geographic Information – Metadata is ontwikkeld. De meeste landen nemen zich voor om norm in te voeren, hoewel in Europa en dus ook in Nederland eerst nog het formele traject voor Europese normalisatie moet worden doorlopen.

Literatuur

(Ahonen-Raino, 2001)

Description of the content of geographic datasets,

Ahonen-Raino, P., (National Land Survey, Finland) in Bjørke, J.T., Tveite, H. (eds.) Proceedings of the 8th Scandinavian Research Conference on Geographical Information Science, 25-27, June 2001, Ås, Norway

(DLC, 1979)

Technical (Unclassified) Report UG478.A88, 1979

Defense Technical Information Center, Defense Logistics Center, U.S.A.

(Duval, 2002)

Metadata principles and practicalities

Duval, E., (KU Leuven, Belgium), W. Hodgins, (Autodesk), S. Sutton, (University of Washington), S.L. Weibel, (DCMI), D-Lib Magazine April 2002, Vol.8, no. 4, ISBN 1082-9873.

(Homer, 1978)

A mathematical model of the flow of data in a management information system

Homer, E.D., 11th American Meeting of the Institute of Management Sciences, 1978.

(ISO, 1946)

International Organisation for Standardisation. Contact: www.iso.ch

ISO/TC 211 Geographic information/Geomatics. Contact: <http://www.isotc211.org/>

(Jakobsson, 2002)

Data Quality and Quality Management – Examples of Quality Evaluation Procedures and Quality Management in European National Mapping Agencies

Antti Jakobsson in *Spatial Data Quality* Wenzhong Shi (HK Polytechnic University), Peter F. Fischer (University of Leicester) and Michael F. Goodchild (University of California) Eds., Taylor and Francis London UK, 2002 ISBN 0-415- 25835-9

(McCelland, 2002)

Challenges for service providers when importing metadata in digital libraries

M. McCelland, D. McArthur, S. Giersch (Colleges Eduprise), G. Geisler (University of North Carolina), D-L lib Magazine April 2002, Vol.8, no. 4, ISBN 1082-9873.

(McIntosh, 1968)

Information Processing 68

S. McIntosh, D. Griffel, Proceedings of International Federation for Information Processing (IFIP) Congress, August 1968.

(Moellering, 1991)

Spatial database transfer standards: current international status

Harold Moellering, editor. Published on behalf of the International Cartographic Association by Elsevier Applied Sciences, London 1991. ISBN 185166677X.

(Moellering, 1994)

Technical characteristics for assessing standards and for the transfer of spatial data and brief international descriptions.

Harold Moellering and C. Clement, (Eds.). International Cartographic Association Standards Commission for the transfer of spatial data, Columbus, Ohio, U.S.A., 1994, 104 pp.

(Moellering, 1997)

Spatial database transfer standards 2: characteristics for assessing standards and full description of the national and international standards in the world

Harold Moellering, editor, Richard L. Hogan associate editor. Published on behalf of the International Cartographic Association by Elsevier Applied Sciences, Oxford, UK and Tarrytown, N.Y. 1997. ISBN 0 08 042433 3.

(Schelling, 1978)

The use of IBM's data dictionary

G. Schelling, Proceedings of the Conference on Data dictionary Systems, London, Nov 1978

(Weber, 1978)

Data Base theory and practice

H. Weber, A.I. Wasserman, Conference in Berlin, 13-15 Sept. 1978.

(Williamson, 2003)

Spatial Data Infrastructures – From Concept to Reality

Ian P. Williamson, Abbas Rajabifard, Mary-Ellen F. Feeney (eds) Taylor and Francis 2003.

(Uitermark, 2001)

Ontology-Based Geographic Data Set Integration

H. Uitermark, PrintPartners Ipskamp Enschede, The Netherlands ISBN 90-365-1617-X, 2001.

(Ziegler, 1978)

Distribution: A New Impetus Toward Understanding Data

K. Ziegler, (International Business Machines, Inc.), August 1978.

Organisatie	Voor on-line contact zie URL:
Dublin Core Metadata Initiative	http://www.xml.com/pub/2000/10/25/dublincore/
Institute for Electrical and Electronic Engineers, Learning Object Metadata Working Group	http://grouper.ieee.org/p1484/doc/wg12/LOM-WD3.htm
U.S.A National Spatial Data Infrastructure	http://www.fgdc.gov
FGDC Official statement, Department of the Interior, Office of the Solicitor	http://www.fgdc.gov/metadata/meta_trademark.htm
International Organisation for Standardisation	http://www.iso.ch
ISO/TC 211 Geographic information/Geomatics	http://www.isotc211.org/
EuroGeographic	http://www.eurogeographics.org

Samenvatting

Vroeger werden kaartgegevens op papier bij de gebruikers afgeleverd, die op zijn beurt in transparante overlays eigen informatie kon toevoegen of met de gegevens op de kaart analyses uitvoeren. Nu we in een digitaal tijdperk zijn aangeland worden beschrijvingen van kaartbestanden in metadatasets geleverd, die via Internet beschikbaar kunnen worden gesteld. Door de globalisering is een uniforme syntactische en semantische formulering noodzakelijk zodat gebruikers uit verschillende culturen wel dezelfde interpretatie geven aan de metadata.

Daarbij zijn verschillende ontwikkelingen van belang geweest sinds de ingebruikname van de metadata beschrijvingen, zoals regionale, nationale en toepassingsgerichte definities van meta-

data normen. Verschillende theoretische en organisatorische ontwikkelingen zijn daarbij in acht genomen.

Tot slot wordt de huidige stand van zaken m.b.t. het gebruik van metadata in verschillende regio's in de wereld toegelicht.