

How To's voor dataopwerking

KPP 2017 Mariene projecten

Gerrit Hendriksen, Willem Stolte, Dave de Koning, Joan Staeb

1230040-002

Titel

How to's voor dataopwerking

Opdrachtgever**Project****Kenmerk****Pagina's**

1

Trefwoorden

Type hier de trefwoorden

Samenvatting

Type hier de samenvatting

Referenties

Type hier de referenties

Versie	Datum	Auteur	Paraaf	Review	Paraaf	Goedkeuring	Paraaf

Status

concept

Dit document is een concept en uitsluitend bedoeld voor discussiedoeleinden. Aan de inhoud van dit rapport kunnen noch door de opdrachtgever, noch door derden rechten worden ontleend.

Inhoud

1 Inleiding	1
1.1 Afbakening	1
1.2 Leeswijzer	1
2 Werkwijze vectordata	1
2.1 Standaard voor vectordata - IM Metingen	1
2.2 Bestandsformaten voor vectordata	2
2.3 Workflow voor vectordata van ruwe data naar IM metingen	8
2.4 Workflow van IM metingen naar database	9
3 Werkwijze rasters	11
3.1 Standaarden voor raster files	11
3.2 Rasterdata – fileformaten	11
3.3 Rasterdata – van ruwe data naar netCDF	11
3.4 Rasterdata – van ruwe data naar tiff plus metadata xml	12
3.5 Issues met rasters	13
3.6 Use case 1 netCDF met meerdere rasters	13

1 Inleiding

Dit document bevat een beknopte omschrijving van de werkwijze om te komen van 'ruwe' data tot aan OGC webservices welke onderdeel kunnen vormen van het IHM data portaal. In dit document worden diverse afkortingen en vaktermen gebruikt. Er wordt verondersteld dat deze bekend zijn.

1.1 Afbakening

Het gaat hier alleen over NL data. Desalniettemin is de werkwijze gelijk aan die van data elders, namelijk van ruwe data naar opgewerkte data volgens een bepaalde standaard. De gebruikte standaard in dit document is, tenzij anders vermeld, AQUO. AQUO is de standaard voor watergerelateerde gegevens in Nederland. Wanneer nodig kan hiervan afgeweken worden (bijv voor niet-watergerelateerde gegevens).

1.2 Leeswijzer

Stapsgewijs wordt in de komende paragrafen aangegeven data gestandaardiseerd wordt. Hierbij dient onderscheid gemaakt te worden tussen twee typen gegevens, vector (tijdreeksen) en raster (tijdreeksen). De werkwijze om te komen tot een OGC webservice is nogal verschillend tussen vector en raster.

2 Werkwijze vectordata

Er wordt hier verondersteld dat bekend is wat een vector is. Zo niet check http://wiki.gis.com/wiki/index.php/Vector_data_model.

Allereerst wordt ingegaan op het uitwisselmodel IM metingen. Dit vormt de basis voor de workflow die ervoor zorgt dat de gegevens in een bepaald vastgesteld datamodel kunnen worden opgeladen.

2.1 Standaard voor vectordata - IM Metingen

AQUO kent een datamodel voor uitwisseling van gegevens, IMWA. Deze data kunnen ook als "platte" tabellen worden uitgewisseld volgens het Informatiemodel metingen (IM_metingen). In dit document wordt verwezen naar kolommen (hierna velden genoemd) uit het IM metingen model die vereist zijn voor data die opgeleverd worden naar RWS.

Veel aangeleverde ruwe data is nu nog niet in het IM metingen format gebracht. De werkwijze om een willekeurige dataset naar IM metingen te brengen is kan verschillen per dataset, en persoon. Het is wel belangrijk dat dit goed gedocumenteerd wordt, waarbij de documentatie naast de data op de repository bewaard blijven.

Voor meer informatie over aquo wordt verwezen naar → <http://www.aquo.nl/>

Voor meer informatie over IM metingen wordt verwezen naar → <http://www.aquo.nl/over-aquo/aquo-onderdelen/aquo-modellen/imwa/>

Een aantal velden in IM metingen kunnen niet vrij worden ingevuld. Hiervoor moeten waarden gebruikworden uit de AQUO "domeintabellen". Dit is aangegeven in Tabel 1. De

domeintabellen zijn interactief te raadplegen via de AQUO domeinserver (AQUO DS). (<http://www.aquo.nl/tools/aquo-ds-webservice/>)

Tabel 1 IM Metingen geeft een volledig overzicht van de velden die gebruikt kunnen worden.

2.2 Bestandsformaten voor vectordata

Vectordata worden aangeleverd in verschillende bestandsformaten, zoals MS-Excel, CSV, MS-Access, Shape files.

Voor opname in de database zullen deze getransformeerd moeten worden naar een eenduidig formaat. IM metingen schrijft voor:

- CSV bestand(en) voor meetpunt en meetwaarden in het geval van puntdata
- SHP voor meetobject en CSV voor meetwaarden in het geval van lijnen of polygonen

Voor nieuwe projecten is het de bedoeling dat data-leveranciers alleen IM metingen conforme datasets aanbieden (tenzij anders is afgesproken).

Tabel 1 lijst met velden volgens IM Metingen

Veldnaam	Voorbeeld	RWS	opmerking / toelichting
		vulling van het veld is: v=verplicht, c=conditoneel, o=optioneel	
Meetpunt.identificatie	NOORDWK10	v	Id nummer van meetobject (of locatie of trek) waar "monster" en "meting" naar verwijzen
Metingomschrijving	Noordwijk 10 km uit de kust	o	naam van meetobject, locatie of trek bij Noordwijk-10
GeometriePunt.x	gg.ggggg	c	de x en y in etrs89. (wgs84 mag ook als benadering gebruikt worden, omdat de afwijking op de Noordzee max 80 cm is) Bij een trek of een raai kan het middelpunt aangegeven worden of een polygoon. Indien polygonen gewenst zijn, kan dat met een string worden aangegeven in het veld "geometrie",
GeometriePunt.y	gg.ggggg	c	
Geometrie		c	(vlak, lijn, raster, polygoon) als de meetlocatie geen punt is
Referentiehorizontaal.code	EPSG4258 of EPSG:4258	v	EPSG4258 (=etrs89) Op dit moment accepteert aquo geen : maar dat is internationaal wel gebruikelijk. Hiervoor is een change verzoek ingediend bij AQUO in dec door Joan Staeb
Monster.identificatie	Imaresnr_2014_43256421	v	unieke identificatie monster (max 36 tekens)
Compartiment.code	BS	v	~28 domeinwaarden zoals: oppervlaktewater, organisme (biota) etc
Orgaan.code	FI	c	indien organisme dan verplicht ~32 domeinwaarden zoals: spierweefsel, dielijk weefsel etc
Organisme.naam	Macoma balthica	c	indien organisme dan verplicht ~duizenden

Begin diepte_m	-1	c	domeinwaarden zoals Abra, Abra alba, etc. monsternemingsdiepte t.o.v. referentievlak (-1 betekent 1m onder referentievlak; +1 betekent 1 m boven referentievlak)
Eind diepte_m	-1	c	verplicht als een begin diepte is opgegeven.
Referentievlak.code	WATSGL	c	~28 domeinwaarden zoals: tov mean sea level, tov NAP, tov bodem, tov WATSGL
Monsterbewerkingsmethode.code	ISO5667-3	o	~23 domeinwaarden zoals: NEN-EN-ISO 5667-3, etc
Bemonsteringsmethode.code	rws huismethode 231	v	~53 domeinwaarden zoals: NENs en RWSVs
Bemonsteringsmethode.codespace	www.rws.nl	c	Indien hierboven een vrije waarde is ingevuld (dus geen domeinwaarde) dan moet de namespace worden ingevuld. Bijv www.rws.nl of www.imares.nl
Monstercriterium.code		o	De domeintabel bestaat uit twee subtabellen van de domeintabel Hoedanigheid: Biologische Kenmerk en Korrelgroottefractie. om een nadere classificatie en/of beperking van het monster te geven. Veel beperkingen kunnen al via de opdeling van het monsterobject verder worden uitgewerkt, maar in sommige gevallen zoals bij biologische metingen is het noodzakelijk om aan te geven dat er enkel een specifiek onderdeel bemeten is. Bijv: alleen bepaalde vislengtes, of van bodemmonster wordt alleen het fijne deel geanalyseerd.
Veldapparaat.omschrijving	Boxcorer	v	De domeintabel Veldapparaat bestaat uit de subtabellen 'Meetapparaat', 'Bemonsteringsapparaat' en 'Plaatsbepalingsapparaat'. Veelal zal het gaan om een bemonsteringsapparaat. Er zijn ~67 domeinwaarden zoals: bodemschaaf, boxcorer, etc.

Monsternemingsdatum	2014-06-01	v	datum in jjjj-mm-dd (denk aan de verplichte nullen!!)
Monsternemingstijd	09:44	o	tijd als hh:mm of hh:mm:ss indien seconden relevant zijn
Tijd_UTCOffset	+1	c	Als monsternametijd is ingevuld en niet in MET zijn opgenomen (UTC / GMT + 1) dan hier de afwijking ten opzichte van UTC opnemen (wintertijd = MET, zomertijd = UTC + 2)
Typering.code		c	~57 domeinwaarden, subdeel van domeintabel parameter , je hebt of een typering, of een grootheid
Grootheid.code	AANTPOPVTE	c	je hebt of een typering, of een grootheid
Parameter.code		c	~3000 domeinwaarden (bij chemische stoffen en objecten verplicht)
Parameter.omschrijving	Macoma balthica	c	~3000 domeinwaarden, bij chemische stoffen is de code bepalend en is dit alleen een optionele toelichting: bij biotaxon is dit veld verplicht
Eenheid.code	mg/m2	v	~167 domeinwaarden
Hoedanigheid.code	adg	o	~415 domeinwaarden verdeeld in referentievlakken, fracties, equivalenten, referentiebases, biologische kenmerken en combinaties hiervan (bijvoorbeeld: asvrij drooggewicht: adg)
Waardebepalingsmethode.code		o	~50 domeinwaarden verdeeld in percentielen en statistiek
Waardebepalingsmethode.code	Imaresmethode 2014-781	v	~600 domeinwaarden, vooral NENs en ISOs. Huismethodes dienen opgenomen te worden als "other: Naam van de huismethode incl nr en versie"
Waardebepalingsmethode.codespace	www.imares.nl	c	Indien hierboven een vrije waarde is ingevuld (dus geen domeinwaarde) dan moet de namespace worden ingevuld. Bijv www.rws.nl

Waardebepalingstechniek.code		o	of www.imares.nl ~150 domeinwaarden, laboratoriumprincipes, geven globaal idee van de techniek. Van belang voor mapping naar internationale standaarden
Begindatum	2014-06-01	v	Datum waarover de waarneming geldig is. In de praktijk meestal de monsternemingsdatum
Begintijd	09:44	o	
Einddatum		o	Vooraf nodig bij tijdsproportionele bemonsteringen.
Eindtijd		o	
Tijd_UTCOffset	+1	c	Als begintijd is ingevuld en niet in MET zijn opgenomen (UTC / GMT + 1) dan hier de afwijking ten opzichte van UTC opnemen (wintertijd = MET, zomertijd = UTC + 2)
Resultaatdatum	2014-06-01	v	Datum waarop de meting is gedaan (dit veld is nieuwe ISO O&M verplichting). Tijdelijke oplossing Indien dit niet bekend is wegens ontbreken in de aanleverende database dan laatste datum uit bovenstaande velden overnemen
Limietsymbool		c	"<" indien waarneming beneden de rapportagegrens is, ">" indien waarneming boven het bereik is.
Numeriekewaarde	24.2	c	De waarneming!!, Let op alleen significante decimalen meegeven
Alfanumeriekewaarde		c	indien geen numerieke waarde is gegeven
Kwaliteitsoordeel.code	0	v	~13 domeinwaarden. Uit mappingwerkgroep is gebleken dat alleen waarde 0 en 3 betekenis hebben.
Accreditatiestatus.omschrijving	ISO17025	o	ISO17025;geen;ander
Meetonzekerheid.bias.waarde	+10	o	
Meetonzekerheid.reproduceerbaarheid.waarde	30	o	2 sigma waarde

Meetonzekerheid.eenheid.code	%	o	%, ug/L, mg/L, pH etc
Meetonzekerheid.referentie.waarde	10	c	indien hierboven % dan hier referentiewaarde opgeven
Meetonzekerheid.referentie.eenheid	n	c	indien hierboven waarde ingevuld, dan hier de eenheid: ug/L, pH etc

Deze velden vormen het startpunt voor de conversie van ruwe data naar een format wat kan worden ingeladen in een database.

2.3 Workflow voor vectordata van ruwe data naar IM metingen

Vector gegevens worden opgewerkt naar het IM metingen model zoals dat hierboven is opgenomen. Zie **Fout! Verwijzingsbron niet gevonden..** In het algemeen zal gelden dat een nieuw bestand gemaakt wordt waarbij de opmaak zodanig is dat deze overeenkomt met de in de lijst genoemde velden. De ETL verwacht dat alle IM_metingen velden aanwezig zijn. Het is niet zo dat alle velden ingevuld dienen te worden.

Bij het invullen van de velden dient te worden gelet op:

- Is van de meting bekend of informatie over de verplichte velden bekend is en of deze matchen met die in de Aquo standaard
- Of de waarden van de meting (en de karakteristieken zoals eenheid etc.) (indien vereist) overeenkomen met de Aquo domein tabellen.

De Aquo domein tabellen zijn te bevragen met de zogenaamde aquo domein tabellen service, deze is online te bevragen via <http://domeintabellen-idsw.rws.nl/>.

Valkuilen en instinkers

Grootheid – alleen wanneer een numeriekewaarde is toegekend

Typering – allen wanneer een alfanumeriekewaarde is toegekend

Op elke regel moet of een grootheid of een typering worden ingevuld. Niet beiden

Op elke regel moet of een numeriekewaarde of een alfanumeriekewaarde worden ingevuld.

Organisme.naam – alleen wanneer in of aan een dier/plant iets gemeten wordt, bijv de lengte, maaginhoud, littekens, etc.

Biotaxon.naam of **parameter.omschrijving** – wanneer aantallen of dichtheden van een dier of plant gemeten zijn.

Denk aan: **Nulwaarden** (NULL, 0, Zero, NA,) NIET weglaten. Deze zijn soms noodzakelijk om een dataset goed te kunnen interpreteren. Vraag bij twijfel na bij een expert.

Het is van belang goed op de hoogte te zijn van de betekenis van verplichte velden zoals hoedanigheid, grootheid, typering, parameter, biotaxon en organisme.naam. Naast de betekenis moet ook duidelijk zijn wanneer deze velden wel of niet kunnen worden gebruikt. In sommige gevallen zoals grootheid en typering, mag er maar maximaal één van de twee kolommen gebruikt worden. Een uitvoerige beschrijving van de standaard is hier te vinden (<http://www.aquo.nl/documents/2015/04/rapport-imwa-metingen-2015.pdf>)

Speciale aandacht verdienen de geometrieën lijn en polygon. De doeldatabase is een zogenaamde [PostgreSQL/PostGIS](#) database. Hierin is het mogelijk om naast een punt (de gemakkelijkste geometrie) een lijn of polygoon op te nemen en alle varianten daarop. Omdat het niet mogelijk is om de geometry binair in een csv op te slaan is afgesproken om het WKT (Well Known Textpad) te gebruiken. Lees meer hierover in het volgende wiki document https://en.wikipedia.org/wiki/Well-known_text.

Voor niet ingewijden is het een leuke puzzel om uit een bepaald type GIS bestand de WKT notatie van de geometrie te halen. Omdat dit voor ieder type GIS bestand anders is wordt hier

verder niet op ingegaan hoe dat te doen. Met de meeste GIS pakketten en/of scripting talen is het mogelijk deze notatie op te vragen en te bewaren. Hier geldt zeker dat je favoriete zoekmachine je grootste vriend is.

Het omzetten van de gegevens naar IM metingen format kan op verschillende manieren gedaan worden:

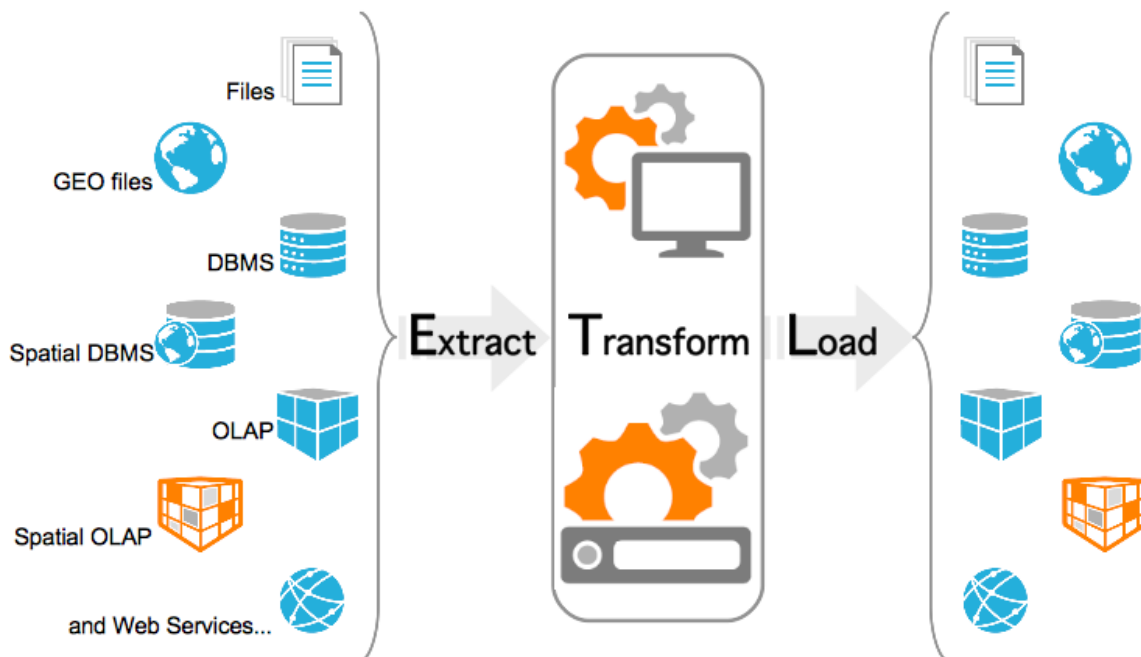
1. Handmatig – bijv in een tekst editor of Excel. Bij gebruik van Excel moet goed opgelet worden dat er geen Excel specifieke formaten (zoals datum) gebruikt worden, of, als deze gebruikt worden, dat ze overeenkomen met het gewenste formaat. Een nadeel van deze methode is het gebrek aan reproduceerbaarheid.
2. Scripting - Feitelijk is scripten één van de methoden om omzetten van ruwe data naar basis data reproduceerbaar te maken, de keuzes worden immers in code vastgelegd. Het maakt hierbij niet uit welke scriptingtaal wordt gebruikt.
3. ETL met Spoon - Is het nu zo dat er veel bestanden zijn die moeten worden omgezet dan kan een ETL van dienst zijn. Net als voorgaande paragraaf geldt hier dat dit maatwerk is en een recept niet echt te geven is.

Bij zowel scripting als ETL wordt verwacht dat de code wordt geplaatst op de repository waar ook de data vandaan komt.

Tijdreeksen zijn data waarbij een of meerdere locaties zijn gedefinieerd, met voor elke locatie een reeks metingen in een apart bestand of apart deel van een bestand is opgenomen. De omzetting van een tijdreeks welke op deze manier beschikbaar wordt gesteld naar IM_metingen formaat resulteert in een grote mate van herhaling (namelijk van de locaties). Dat lijkt inefficiënt, maar wordt tijdens de ETL in de database weer verder genormaliseerd. Kortom, het datamodel is erop gericht redundantie te voorkomen.

2.4 Workflow van IM metingen naar database

De workflow van losse bestanden naar entries in de database wordt ETL (Extract Transform Load) genoemd, om aan te geven dat het hier om data verkeer tussen 2 opslag systemen gaat.



Figuur 1 Een algemeen overzicht van ETL procedures

Afgaande op bovenstaande figuur zijn er aan de extract kant diverse mogelijkheden. Op dit moment is er een workflow beschikbaar welke IM metingen conforme bestanden leest, (Extract), omzet naar een fysiek datamodel (we gebruiken het Informatiehuis Water datamodel) (Transfrom) en in een PostGres/PostGIS database laadt (Load).

De huidige workflow is gebaseerd op ETL software Pentaho Data Integration (Spoon) voorheen ook wel als Kettle aangeduid (download PDI van <http://sourceforge.net/projects/pentaho/files/Data%20Integration/7.1/pdi-ce-7.1.0.0-12.zip/download>). Deze software is zonder kosten beschikbaar. Op dit moment wordt het voornamelijk gebruikt om SQL statements aan te roepen, data uit csv-bestanden te lezen en logging/foutmeldingen op te slaan in een log bestand per data import. De software wordt gebruikt voor alle stappen: Inladen ruwe data, verwerken dan de data en het correct plaatsen in de database. In het geval dat de brondata al verwerkt als platte tabel in een database zit is het ook mogelijk om via een programma als pgAdmin de sql statements in de ETL procedure afzonderlijk aan te roepen.

De database moet van te voren goed ingericht zijn. Hiervoor is een SQL statement beschikbaar welke een database volgens het IHW datamodel inricht. Door deze sql statement te draaien komt een lege database tot de beschikking. Vervolgens is het mogelijk de domeintabellen van Aquo te vullen met de laatste versie van de domeintabellen op de server. Deze kunnen handmatig als csv/xml gedownload worden of direct met een ETL vanuit de server. In sommige gevallen komen de domeintabellen op de server niet overeen met de namen/invulling van de tabellen in het datamodel. Hierdoor moet er het een en ander verschoven en opgesplitst worden. Ook moeten in sommige gevallen waarden worden toegevoegd.

De ETL procedure wordt bewaard op de Deltares repository (<https://svn.oss.deltares.nl/repos/openearthtools/trunk/etl/aquo>)

3 Werkwijze rasters

Er wordt hier verondersteld dat bekend is wat een raster is. Zo niet check http://wiki.gis.com/wiki/index.php/Raster_data_model.

GIS technisch gezien zit er niet veel verschil tussen een raster en een vector met betrekking tot opslag en inhoud. Een raster bevat (op het netCDF formaat na) altijd maar 1 parameter. Groot nadeel van rasters (op het netCDF format na) is dat er nauwelijks tot geen metadata beschikbaar is in het bestand.

Rasters verdienen dus wat dat betreft speciale aandacht. Verwacht wordt dan ook dat er minimale meta informatie aanwezig is, in het slechtste geval dient dat uit de bij de data behorende rapportage te worden geëxtraheerd.

3.1 Standaarden voor raster files

Waar mogelijk kunnen AQUO termen gebruikt worden in de metadata van rasterfiles. Vaak zullen rasters gegevens bevatten die moeilijk in AQUO zijn onder te brengen. In dat geval kan uitgeweken worden naar de CF standaardnamen (<http://cfconventions.org/Data/cf-standard-names/46/build/cf-standard-name-table.html>)

3.2 Rasterdata – fileformaten

Er zijn twee manieren om rasterdata te verwerken:

1. Opwerken naar netCDF
2. Opwerken naar GTIFF en een metadata xml


3.3 Rasterdata – van ruwe data naar netCDF

netCDF is hierboven al een aantal maal genoemd. Het is een format wat zijn oorsprong heeft in de oceanografie en meteorologie. netCDF kan gezien worden als een database met metadata gevat in een enkele file. Of een netCDF als een raster kan worden gezien hangt af van de opzet van de data. NetCDF's kenmerken zich door een veelzijdigheid in manieren om data te bewaren. Een netCDF kan namelijk gebruikt worden om data in diverse dimensies op te slaan, van punten met een vrijwel oneindige tijdreeks tot x dimensionale rasters.

Om tevens te voldoen aan de eis dat gegevens gevisualiseerd kunnen worden door bijvoorbeeld QGIS, Geoserver of ArcGIS is het meest voor de hand liggend dat er 3 dimensies worden gebruikt, lon, lat en parameter waarden. Dit betekent wel dat de raster gegevens naar WGS84 geconverteerd dienen te worden. Dat is het meest voor de hand liggend en genereert de minste problemen.

Het format netCDF maakt het tevens mogelijk (het is niet verplicht) een set metadata op te nemen, deze worden in de header van het bestand geplaatst waarmee het format een ideale manier is om persistente data te bewaren. Immers waarden en metadata zijn niet gescheiden.

De Climate and Forecast wereld heeft zijn eigen metadata standaard gecreëerd welke specifiek voor netCDF toegepast wordt, een netCDF is met volledige metadata zogenaamd CF compliant. Er zijn tools beschikbaar om van CF naar ISO19135 te converteren of direct naar een metadata catoalogus (geonetwork).

Om zeker te zijn van correcte netCDF's die verder goed gebruikt kunnen worden is het gebruik van de [GDAL](#) routines [gdal_translate](#) en [gdalwarp](#). Deze tools zijn zowel los installeerbaar als via QGIS te gebruiken. Indien QGIS is geïnstalleerd zijn de tools ook goed in batch mode te gebruiken. Hiervoor dient gewerkt te worden vanuit de OSGeo4W Shell (te activeren via  gevolgd door OSGeo4W shell). Dit activeert een specifieke voor deze sessie te gebruiken set aan gdal routines.

Randvoorwaarde voor het gebruiken van de GDAL routines is dat men goed inzicht heeft in het ruimtelijke referentiestelsel waarin de data staat. Voor Nederland is dat doorgaans RD ([EPSG:28992](#)). De data dient echter te worden getransformeerd naar WGS84 ([EPSG:4326](#)). Meer informatie over referentie stelsels is te vinden op:

- <http://spatialreference.org>
- <http://epsg.io/>

Een voorbeeld. Er is een ASCII bestand beschikbaar van parameter x in RD (x.asc). Dit bestand is in 3 stappen te converteren naar een netCDF in WGS84.

```
gdalwarp -s_srs EPSG:28992 -t_srs EPSG:4326 -of GTIFF -overwrite x.asc x.tiff
```

```
gdal_translate -of netCDF -overwrite x.tiff x.nc
```

De 3^{de} stap is wat lastiger en kan het beste via een scripting taal gerealiseerd worden. Via de [netCDF kickstarter](#) is voorbeeld code te genereren waarmee de header volledige kan worden ingevuld. Het is vervolgens aan de gebruiker te bepalen via welke taal dit gebeurt. Ook hier geldt, leg de code vast.

3.4 Rasterdata – van ruwe data naar tiff plus metadata xml

De keuze voor een GeoTIFF kan worden bepaald door het toekomstig gebruik, maar ook de verwachte grootte van het bestand. GeoTIFF's zijn er in verschillende smaken waaronder BigGeoTIFF's die groter mogen worden dan 2.1 Gb. Dat dit werkt kan worden aangetoond in het MI-SAFE portaal. Hierin zitten diverse BigGeoTIFFS die goed gevisualiseerd kunnen worden. GeoTIFF's maken het mogelijk interne tiling toe te passen waardoor visualisatie performance omhoog gaat. Door het gebruik van Geoserver is een geotiff bovendien weer beschikbaar te maken als een online raster waardoor 'third party' software een stukje van de data kan oppakken en voor een bepaald proces gebruiken. Let wel, dit kan bij ieder type rasters (ook netCDF).

GeoTIFF's worden bij voorkeur gebruikt als het om gemozaiekte bestanden gaat. Bestanden met een tijdsdimensie zijn niet hiervoor niet geschikt, vanwege het ontbreken van een goede tijdsdimensie (de opslag vind dan plaats in banden).

Zoals hierboven al genoemd, bestanden zijn gemakkelijk om te zetten naar een *.tif middels de gdalwarp utility.

Tiling kan worden gerealiseerd met behulp van de utility [gdaladdo](#). GTIFF zijn het default format van GDAL.

Tips, let vooral ook op de no data waarden en verder op de compressie mogelijkheden. Alle opties zijn te vinden op de pagina over het [GTIF formaat](#).

4 oktober 2017

3.5 Issues met rasters

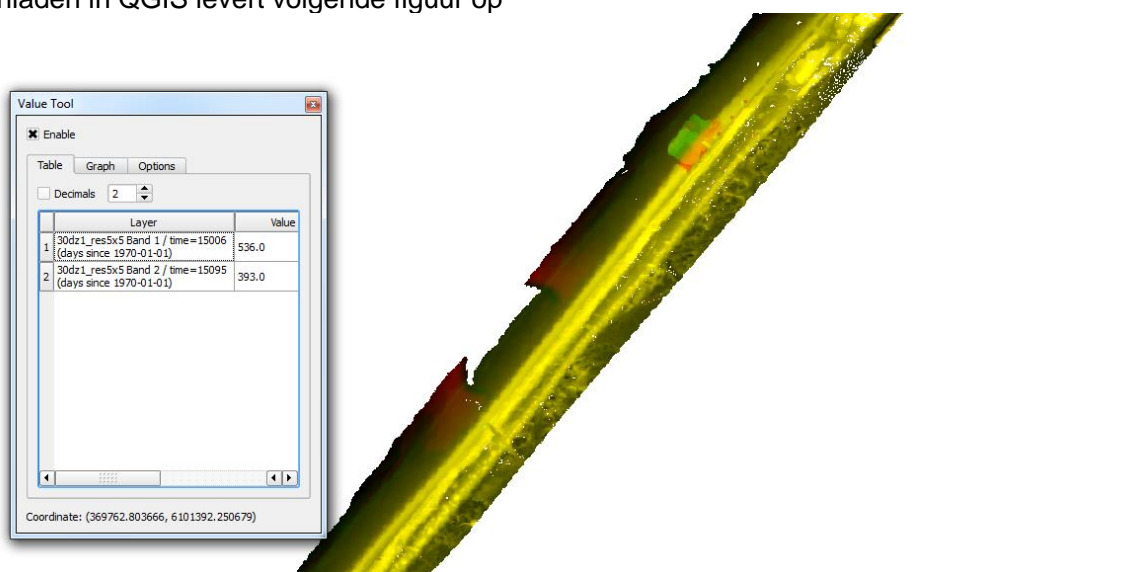
Er zijn bepaalde gevallen waarbij raster voor wat onverwachte problemen zorgen. Hieronder, en het document zal verder worden aangevuld, één of meerdere use cases waartegen aan is gelopen.

3.6 Use case 1 netCDF met meerdere rasters

Dit kan voorkomen als in een netCDF voor meer dan 1 datum waarden zijn opgeslagen.

Deze use case gaat over bestand http://opendap.tudelft.nl/thredds/catalog/data2/zandmotor/morphology/LIDAR/catalog.html?dataset=scanDatasets2/zandmotor/morphology/LIDAR/30dz1_res5x5.nc

Inladen in QGIS levert volgende figuur op



In het value tool window is te zien dat er twee lagen gegevens beschikbaar zijn. Het is zaak deze gegevens te splitsen en te voorzien van een datum. Dat zou eigenlijk voor alle bestanden een stuk makkelijker zijn.

Splitsen van netCDF over een dimensie (time in dit geval) kan met de tools die bekend zijn onder de naam NCO, a suite of netCDF operators door de University of California.

Dit is een commandline utility waarbij een specifiek onderdeel van NCO wordt aangeroepen, te weten de netCDF KitchenSink utility, ncks, op de commandline kan met behulp van de volgende code een netCDF gesplitst worden.

```
ncks -d time,1 30dz1_res5x5.nc 30dz1_res5x5 t1.nc
```

Idealiter wordt tijdstap 1 omgezet naar een leesbare datumstring. Omdat het hier om enkele tientallen bestanden gaat is een Python routine gemaakt die bovenstaande uitvoert waarbij de tijdinteger (dagen sinds 1970) wordt omgezet naar een leesbare datum welke wordt toegevoegd aan het bestand.