

Datamanagement mariene projecten

OpenEarth ontsluiting van data en beheerscenario's



Datamanagement mariene projecten

OpenEarth ontsluiting van data en beheerscenario's

dr. F. Baart
dr.ir. G.J. de Boer
dr.ir. C. den Heijer
drs. ing G. Hendriksen
ing. J.F. Keppel
drs. G.H. van der Kolff
M. Pronk BSc
dr. W. Stolte

1208604-000

Titel
Datamanagement mariene projecten

Opdrachtgever RWS Water Verkeer en Leefomgeving	Project 1208604-000	Pagina's 91
--	-------------------------------	-----------------------



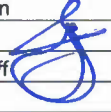
Trefwoorden
OpenEarth, mariene monitoring, hergebruik data, datamanagement, RWS-IV infrastructuur

Samenvatting
Dit project betreft een vervolg op de Pilot datamanagement mariene projecten, waarin aangetoond werd dat OpenEarth tools geschikt zijn om mariene data binnen de RWS informatievoorziening (IV) infrastructuur duurzaam voor hergebruik beschikbaar te stellen. In het huidige project is een significante hoeveelheid data van afgesloten en nog lopende mariene projecten op gestandaardiseerde wijze opgeslagen en ontsloten.

Door vertraging in de ontwikkeling van de RWS-distributielaag kunnen de gegevens niet via een centrale ingang bevroegd worden. Wel is aangetoond dat de data op de verschillende locaties waar OpenEarth stacks geïmplementeerd zijn, met behulp van een eenvoudige viewer ontsloten kunnen worden. Het gebruik van (internationale) standaards heeft aan dit succes bijgedragen.

Met de uitvoering van het project is een belangrijke stap gezet in het duurzaam voor hergebruik beschikbaar maken van mariene projectdata. Om een goed gefundeerd besluit te kunnen nemen waar en hoe mariene projectdata binnen de RWS IV-infrastructuur in de toekomst ontsloten zullen gaan worden, biedt dit rapport een afwegingskader voor RWS, gebaseerd op de ervaringen van het huidige project, en worden aanbevelingen gedaan.

Referenties
RWS-WVL zaaknummer 31085608; 23-09-2013

Versie	Datum	Auteur	Paraaf	Review	Paraaf	Goedkeuring	Paraaf
	dec. 2014	dr. F. Baart		drs. A.J. Nolte	AN	drs. S. Tatman	
		dr.ir. G.J. de Boer					
		dr.ir. C. den Heijer					
		drs. ing G. Hendriksen					
		ing. J.F. Keppel					
		drs. G.H. van der Kolff					
		M. Pronk BSc.					
		dr. W. Stolte					

Status
definitief

Inhoud

1	Inleiding	1
1.1	Voorgeschiedenis	1
1.2	Doel huidige project	1
1.3	Open data en ontsluiting in Nederland	2
1.4	Leeswijzer	5
2	OpenEarth en aansluiting op RWS-standaarden	7
2.1	Deltares OpenEarth software stack	7
2.1.1	Repository	9
2.1.2	Database	9
2.1.3	Catalogus	10
2.1.4	Authenticatie	10
2.2	Aquo-standaard	11
2.3	Koppeling RWS-IV architectuur: datadistributielaag	12
2.4	Aansluiting bij dataportalen	16
3	Mariene projectdata	19
3.1	PMR-natuurcompensatie Voordelta	19
3.2	Zandmotor	22
3.3	Shortlist Wind op Zee	25
3.4	Noordzeewind; Offshore Windpark Egmond aan Zee (OWEZ)	27
3.5	MEP-duinen	28
3.6	Overige datasets	31
3.6.1	EcoLIMS	31
3.6.2	RijMaMo	32
3.6.3	PoC Lodingen	32
3.6.4	BWN	32
3.6.5	KPP B&O Kustlijnzorg	33
3.6.6	MCI Old data	34
4	Implementatie op vier locaties	35
4.1	Deltares	35
4.1.1	Technische implementatie stack PMR-NCV	35
4.1.2	Organisatorische invulling PMR-NCV	36
4.1.3	Autorisatie en online beschikbaarheid data tbv DL	36
4.1.4	Technische implementatie tweede stack tbv overige mariene projecten	37
4.1.5	Organisatorische invulling Mariene Projecten	38
4.1.6	Autorisatie en online beschikbaarheid Mariene Projecten	38
4.2	IMARES	39
4.2.1	Technische implementatie stack	39
4.2.2	Organisatorische invulling	39
4.2.3	Autorisatie en online beschikbaarheid data tbv DL	40
4.3	TU Delft / 3TU.Datcentrum	40
4.3.1	Technische implementatie stack	40

4.3.2	Organisatorische invulling	41
4.3.3	Autorisatie en online beschikbaarheid data tbv DL	42
4.4	RWS-CIV	43
4.4.1	Technische implementatie stack	43
4.4.2	Organisatorische invulling	43
4.4.3	Autorisatie en online beschikbaarheid data tbv DL	44
5	Afwegingskader beheersscenario's	45
5.1	Enkelvoudige versus meervoudige bronhouder	45
5.2	Waarde, onderhoudskosten en adaptieve opwerking	45
5.3	Enkelvoudige versus meervoudige dataset	47
5.4	Versies van data, releases van data, foutmeldingen, bevroren basisdata	47
5.5	Open data vs. data onder embargo: authenticatie	48
5.6	Data expertise, helpdesk	50
5.7	Vinden van data en opvragen van data zonder DL: voordeel van standaarden	50
5.8	Hosting van een OpenEarth stack met mariene data	52
6	Conclusies en aanbevelingen	57
7	Literatuur	60
	Bijlage(n)	
A	Gedetailleerd overzicht opwerking data per project	A-1
A.1	PMR-NCV	A-1
A.2	Zandmotor	A-2
A.3	Wind op zee	A-2
A.4	MEP NSW	A-3
A.5	MEP bestaande duinen	A-4
A.6	MEP Nieuwe duinen	A-5
A.7	ECOLIMS-BENTHOS	A-6
A.8	KPP B&O Kustlijnzorg (ecologie)	A-7
B	Uitwisselmodel Mariene Data	B-8
C	Aquo als standaard voor mariene data	C-10
C.1	Ervaringen met betrekking tot het werken met Aquo.	C-10
C.2	Observaties van terrestrische soorten (vegetatie)	C-11
C.3	Observaties van soorten uit het mariene domein (Fauna)	C-12
D	Mapping SeaDataNet - Aquo	D-31

1 Inleiding

1.1 Voorgeschiedenis

Het ministerie van Infrastructuur en Milieu heeft de ambitie om vanaf 1 januari 2015 in principe al haar data open toegankelijk te maken. Deze ambitie past in een internationale trend van het inrichten van infrastructuren voor open toegang tot gestandaardiseerde data. Binnen het ministerie heeft Rijkswaterstaat (RWS) het initiatief genomen om data duurzaam op te slaan en voor hergebruik geschikt te maken.

In 2012 heeft Rijkswaterstaat aan Deltares gevraagd om in een pilotstudie te inventariseren welke korte en langere termijn oplossingen mogelijk zijn voor duurzame opslag en hergebruik van mariene projectgegevens. De pilot moest niet alleen input leveren voor de technisch inhoudelijke kant van opslag en distributie van data, maar ook suggesties aanreiken voor de formele en organisatorische kant.

De pilot toonde aan dat OpenEarth technieken succesvol kunnen worden toegepast in combinatie met de RWS data architectuur. Gebruikers krijgen toegang tot de laatste versie van data, terwijl de data bij verschillende bronhouders kan blijven. Het pilotsysteem maakt gebruik van open source componenten en omvat de procedures die de verbinding met de Watermanagement distributielaag (DL) van Rijkswaterstaat verzorgen. Hiermee worden op projectbasis verzamelde gegevens op gestandaardiseerde wijze ontsloten. Tegelijkertijd verschaft de pilot inzicht in nationale en internationale ontwikkelingen, op het gebied van datamanagement.

Begin 2013 heeft RWS in haar Portfolio-overleg besloten de pilot een vervolg te geven, op basis van de Businesscase "Hosting scenario's mariene projectdata". Data van mariene projectmonitoring worden momenteel alleen gedurende de looptijd van het project opgeslagen en beheerd. Vindbaarheid voor andere gebruikers tijdens de looptijd is beperkt en wordt na afloop van het project nog slechter. Het moeizame zoeken naar relevante data of het opnieuw inwinnen van data is een kostenpost voor RWS en andere opdrachtverlenende partijen (initiatiefnemers, EZ, kennisinstituten), die vermeden kan worden door geavanceerd datamanagement.

1.2 Doel huidige project

Het huidige project richt zich op een werkend instrument voor Rijkswaterstaat, dat ervoor zorgt dat een significante hoeveelheid mariene projectdata beheerd en herbruikbaar, en voor andere projecten toepasbaar wordt. De data die binnen dit project ontsloten worden, betreffen enerzijds data van afgesloten projecten, waarvoor destijds geen speciale maatregelen getroffen zijn om de data te borgen. Deze data lopen het risico lopen verloren te raken, omdat ze achteraf onvindbaar blijken, of omdat niet duidelijk is wat de data precies voorstellen. Binnen het project worden ook data van lopende projecten ontsloten: met enkele aanpassingen kunnen deze doorgaans (beter) geschikt gemaakt worden voor hergebruik. Voor nieuwe projecten is een handreiking opgesteld (Handreiking datamanagement Versie 1.0: Datamanagement (mariene) projectmetingen in Rijkswateren; Hendriksen, G., Van der Kolff, G.H., Van Splunder, I., Borst, K. en J. Staeb, 2014) die projectleiders richtlijnen verschaft om reeds bij aanvang van het project de juiste maatregelen te treffen om de te verzamelen data duurzaam te bewaren en te ontsluiten. De ontwikkelde bouwstenen om dit mogelijk te maken, moeten aansluiten bij de RWS-Informatievoorziening (IV-) architectuur zodat een duurzame datamanagement oplossing gecreëerd wordt. De distributielaag heeft hierin een

kerntaak: deze draagt er namelijk zorg voor dat de projectdata centraal via webservices ontsloten worden.

Korte termijn doelstelling van het project is het vergroten van de vindbaarheid, bereikbaarheid en de borging van data van de grote mariene monitoringprojecten: Project Mainport Rotterdam (PMR: Maasvlakte 2), Zandmotor en Wind op zee. Het eindproduct is een via een selectietool benaderbare, gedistribueerde database, die de gegevens ontsluit. Alhoewel het huidige project zich geheel toespitst op mariene projectdata, is de voorgestelde OpenEarth oplossing generiek en dus ook geschikt voor andere domeinen (bijvoorbeeld zoet water of “droge” data). Hetzelfde geldt voor de aard van de data: het huidige project richt zich op puntdata, maar de gekozen oplossing is met slechts enkele aanpassingen direct toepasbaar voor het bewaren en ontsluiten van rasterdata. De OpenEarth oplossing biedt Rijkswaterstaat niet alleen intern voordeel, maar biedt ook andere gebruikers (nationaal en internationaal) de gelegenheid om van de verzamelde en ontsloten gegevens te profiteren. Door de OpenEarth oplossing op verschillende locaties te implementeren, biedt het project Rijkswaterstaat ook inzicht in opties voor het uiteindelijke beheer van de data. Een afwegingskader biedt handvatten om hiervoor een onderbouwd besluit te nemen.

1.3 Open data en ontsluiting in Nederland

Er zijn veel definities van open data. Wikipedia omschrijft open data als een term die wordt gebruikt om vrij beschikbare informatie te duiden. De voorwaarden waaronder deze informatie beschikbaar is, wordt beschreven in licenties en gebruiksvoorwaarden. Bij open data wordt er naar gestreefd om de beperkingen voor hergebruik zoveel mogelijk terug te dringen. De toegevoegde waarde van open data wordt in toenemende mate erkend.

Om partijen te stimuleren om hun data in een zo herbruikbaar mogelijk formaat beschikbaar te stellen, heeft Tim Berners-Lee (grondlegger van het World Wide Web-WWW) een vijfsterrenmodel voorgesteld (Tabel 1.1). Hoe meer sterren, hoe beter de data toegankelijk en herbruikbaar zijn. Linked open data staat in Nederland nog in de kinderschoenen. Voor eenvoudige toegang is één ster minimaal noodzakelijk; de ambitie voor het ontsluiten van mariene projectdata is drie à vier sterren.

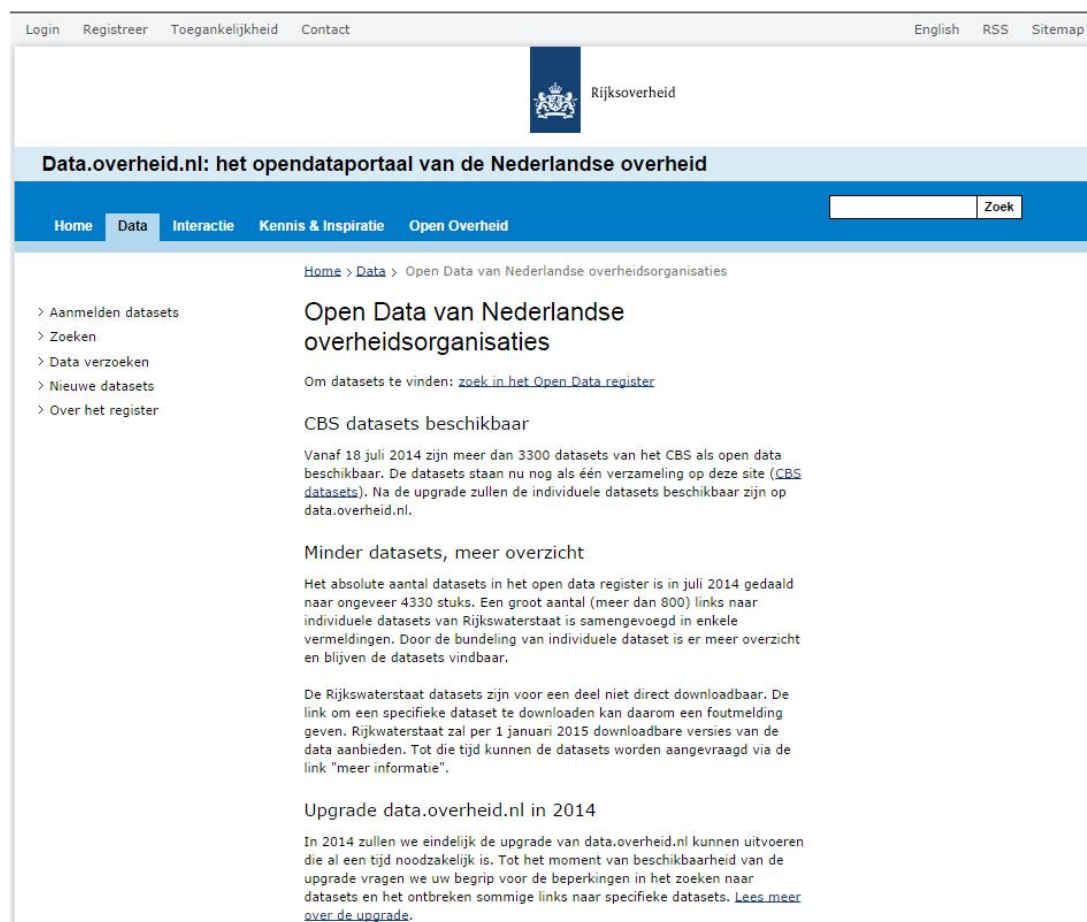
Tabel 1.1 Vijfsterrenmodel van Berners-Lee

*	De informatie is beschikbaar op het internet, in welk formaat dan ook.
**	De informatie is online beschikbaar in een gestructureerd formaat, dat geschikt is voor automatisch hergebruik (zoals Excel in plaats van een plaatje van een tabel).
***	De informatie is online beschikbaar in een open bestandsformaat (zoals CSV in plaats van Excel).
****	Al het bovenstaande, en bovendien wordt gebruikgemaakt van de open standaarden Resource Description Framework (RDF) en SPARQL , zodat anderen makkelijk naar de dataobjecten kunnen verwijzen.
*****	Al het bovenstaande, en bovendien wordt er naar data van anderen verwezen voor meer context van de data (Linked open data).

Ook in de Nederlandse politiek krijgt open data aandacht. Op 30 mei 2011 stuurde minister Donner (BZK) zijn visie op het hergebruik van overheidsinformatie en open data naar de Tweede Kamer. Naast zijn visie schetste hij de activiteiten die reeds zijn ondernomen ter bevordering van het (her)gebruiken van overheidsinformatie en de belemmeringen die er nog bestaan.

Een jaar later stuurde zijn opvolger het rapport 'De waarde van open data - Keuzes en effecten van open-datastrategieën voor publieke organisaties' naar de Tweede Kamer. Dit rapport noemt de volgende voordelen van open-datastrategieën voor publieke organisaties:

- *Financiële besparing*
Besparingen kunnen ontstaan als andere instrumenten niet meer ingezet hoeven te worden of als er andere financiële voordelen ontstaan. Doordat een individuele publieke organisatie met een open data aanpak informatie-uitwisselingen anders organiseert, kunnen ook andere publieke instellingen daarvan de (financiële) vruchten plukken.
- *Verhoging datakwaliteit*
De datakwaliteit kan omhoog gaan doordat de organisatie gedwongen wordt om de eigen datakwaliteit te verhogen alvorens data naar buiten te brengen, of doordat er feedbackmechanismen worden ingericht.
- *Versterking van de legitimiteit*
De legitimiteit van de organisatie wordt versterkt doordat de informatie die de organisatie in beheer heeft vaker wordt gebruikt. Het effect van haar bestaan heeft vaker impact op de samenleving.
- *Vergroting van de externe oriëntatie van de organisatie*
Door voor een open data strategie te kiezen, kan een organisatie gedwongen worden om naar buiten te kijken en meer samen te werken met ketenpartners.
- *Standaardisatievoordelen*
Wanneer meer data volgens (open) standaarden ontsloten worden, kunnen andere overheidspartijen deze data ook eenvoudiger hergebruiken. De interoperabiliteit van de overheid wordt hierdoor vergroot.



Figuur 1.1 Open overheidsdata wordt online ontsloten door Data.overheid.nl

Data.overheid.nl

In Nederland wordt open overheidsdata ontsloten via Data.overheid.nl (Figuur 1.1). Data.overheid.nl signaleert als gevolg van open data positieve effecten in de organisatie, zowel op financieel als organisatorisch gebied. Door open data wordt niet alleen economische waarde gecreëerd, maar wordt ook de openheid en transparantie van overheidsorganisaties vergroot. De Rijksdienst voor het Cultureel Erfgoed (Beeldbank en Bibliotheek) zag een enorme verbetering in de kwaliteit van hun data, door de feedback van het publiek, die ze na de openstelling ontvingen. Bij de wettenbank droeg het openstellen van data bij tot standaardisatie van de datastructuur. Zo werd het vinden van officiële documenten eenvoudiger en werd de samenwerking tussen onderdelen van de overheid bevorderd.

Nationaal Georegister

Het Nationaal Georegister (nationalegeoregister.nl) is dé centrale voorziening voor het ontsluiten van geo-informatie in Nederland (Figuur 1.2). Door een zoekterm in te voeren wordt een overzicht aangeboden van beschikbare datasets en services. Deze datasets zijn in veel gevallen direct te downloaden, de services zijn te bekijken of in eigen toepassingen te gebruiken.

Figuur 1.2 Het Nationaal Georegister ontsluit geo-informatie in Nederland

RWS-IV infrastructuur

Met het huidige project beoogt Rijkswaterstaat invulling te geven aan haar ambitie om haar data open toegankelijk te maken en voor hergebruik te ontsluiten. Op basis van de ervaring met het gebruik van de Deltares OpenEarth software stack zal Rijkswaterstaat een besluit nemen over de inrichting van haar IV-infrastructuur.

1.4 Leeswijzer

De oplossing die Deltares voorstelt voor opslag en ontsluiting van mariene projectdata is gebaseerd op de OpenEarth workflow. In het volgende hoofdstuk wordt uiteengezet uit welke functionele elementen de OpenEarth software stack bestaat. Tevens wordt aangegeven hoe de OpenEarth stack aansluit op de RWS-standaarden, zowel wat betreft semantiek (Aquo-standaard) als wat betreft syntax (RWS-IV infrastructuur; distributielaag). Hier worden ook de consequenties beschreven van de vertraging in het beschikbaar komen van de DL.

Hoofdstuk 3 beschrijft de mariene projecten, waarvan de data via de OpenEarth stack opgeslagen en ontsloten worden. Rode draad in deze projecten is het bestuderen van de ecologische en morfologische effecten van menselijke ingrepen in het mariene milieu. Het betreft zowel reeds afgesloten projecten als lopende projecten, die op hetzelfde thema

voortborduren. Voor toekomstige projecten is een handreiking datamanagement opgesteld. De handreiking adviseert RWS-projectleiders hoe zij met een beperkte inspanning vóóraf, optimaal kunnen profiteren van de voorgestelde wijze van databeheer. Met de handreiking wordt beoogd te voorkomen dat pas na afloop van projecten nagedacht wordt hoe de veelal bonte verzameling van gegevens voor hergebruik ontsloten kan worden, terwijl de projectbudgetten op dat moment al vaak opgesoupeerd zijn. Hoofdstuk 4 laat zien dat de OpenEarth stack op verschillende manieren op verschillende locaties technisch geïmplementeerd kan worden, terwijl het geheel als één virtuele database blijft werken. Welke organisatorische consequenties de verschillende implementaties met zich meebrengen, wordt in Hoofdstuk 5 nader uitgewerkt. Daarmee verschaft dit hoofdstuk RWS handvatten om te besluiten over de Businesscase “Hosting scenario’s mariene projectdata”.

De mariene projectdata worden online ontsloten via:

- <http://marineprojects.openearth.eu/geonetwork>
- <http://pmr-geoserver.deltares.nl/geonetwork/srv/en/main.home>
- <https://zandmotordata.nl/>

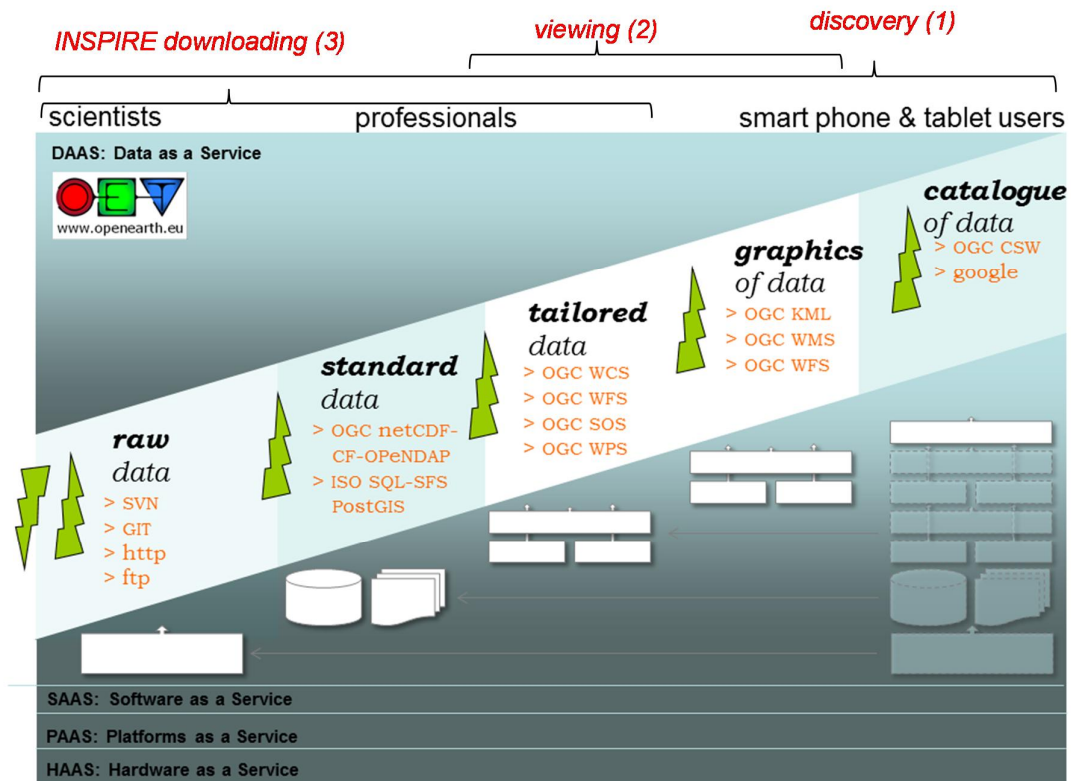
De mate van opwerking (van ruw, naar gestandaardiseerd tot in de catalogus ontsloten), wordt tenslotte in Bijlage A samengevat. Bijlage B beschrijft het uitwisselmodel dat voor standaardisatie nodig is. In Bijlage C worden de ervaringen van het gebruik en eventueel gewenste aanvullingen van de Aquo-standaard beschreven terwijl Bijlage D ten slotte de mapping beschrijft van de SeaDataNet standaard op de Aquo-standaard.

2 OpenEarth en aansluiting op RWS-standaarden

2.1 Deltares OpenEarth software stack

Als platform voor gegevensuitwisseling is gekozen voor de Deltares OpenEarth software stack. De OpenEarth filosofie gaat ervan uit dat gegevens alleen op een zinnige manier herbruikbaar zijn, als zij beschreven worden in open standaarden, bruikbaar zijn voor alle systemen, toegankelijk op verschillende niveaus, van ruwe data tot afgeleide beelden, schaalbaar en eenvoudig vindbaar zijn, door middel van een goede catalogus.

Van het veldboekje tot via internet uitwisselbare, (inter-)nationaal gestandaardiseerde en gecatalogiseerde, digitaal te verwerken gegevens, moet een serie stappen doorlopen worden. Deze stappen moet traceerbaar zijn, zodat niet alleen de opgewerkte producten, maar ook de basisgegevens waaruit de opgewerkte producten afgeleid zijn, terug te vinden zijn.



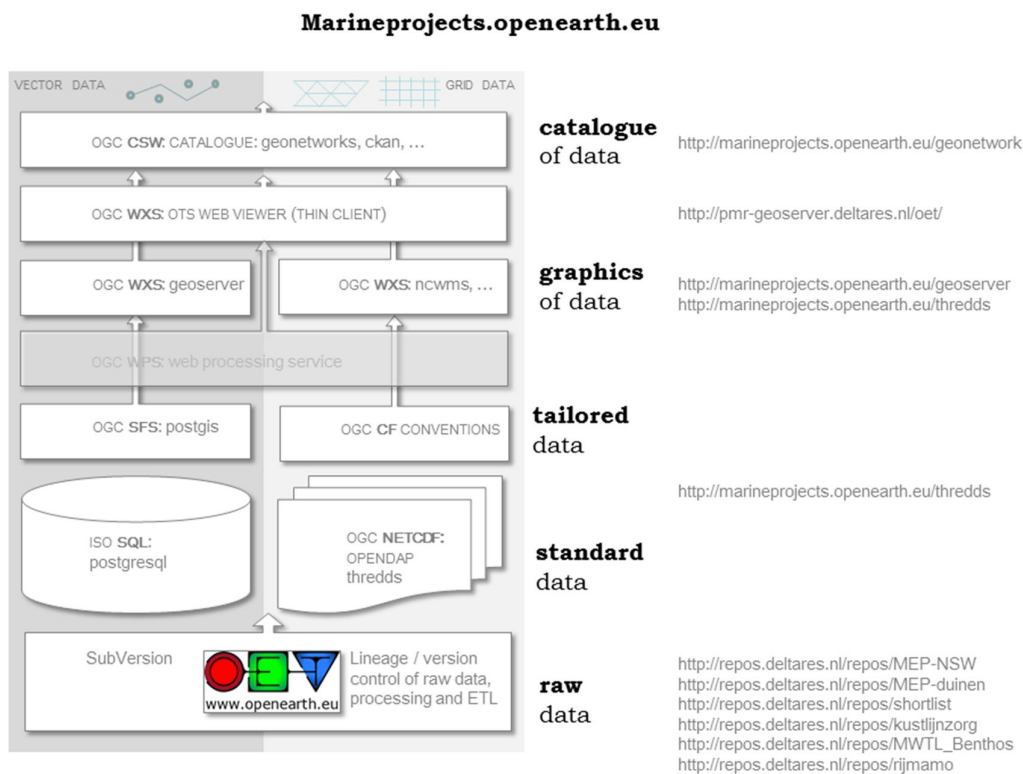
Figuur 2.1 Opwerking en hergebruik van data volgens OpenEarth

Voor de opwerking van data is een bepaalde inspanning nodig (de verticale as in Figuur 2.1). Wetenschappers, die zelf de data verzamelen, zullen zelf vaak ook de opwerking doen. Zij weten precies wat de gegevens voorstellen en kunnen die ook gebruiken. Het is echter niet gezegd dat hun (buitenlandse?) collega's daar dan ook mee overweg kunnen. Wat was bijvoorbeeld het gebruikte referentievlak of coördinatenstelsel? En zijn de Nederlandse soortnamen, ook in het buitenland bekend? Kortom, om hergebruik door derden mogelijk te maken, is er informatie over de verzamelde gegevens nodig, de metadata. Aangezien metadata voor eigen gebruik door de wetenschappers doorgaans niet nodig is -zij weten toch

wat de data voorstellen- komt het in de praktijk vaak voor dat metadata achteraf moet worden toegevoegd of erger, niet meer traceerbaar is, waardoor collega's later slechts kunnen gissen naar de werkelijke betekenis van de gegevens (kilometers of miles, graden Celsius of Fahrenheit etc.). Dit geldt niet alleen voor de inhoud van de data, maar ook van de wijze waarop ze aangeboden worden. Zijn de gegevens opgeslagen in een voor de computer leesbaar formaat, dat eenvoudig verdere verwerking toelaat, of moeten de gegevens bij wijze van spreken worden overgetypt? Hoe lager de benodigde inspanning voor hergebruik, hoe groter de kans dat hergebruik zal plaatsvinden.

De (her-)gebruiker benadert de gegevens in Figuur 2.1 vanaf rechtsboven: vanaf een centrale ingang zal hij zich afvragen of er überhaupt gegevens beschikbaar zijn (discovery) en zo ja hoe die er uit zien en waar ze opgenomen zijn (viewing), om tenslotte te besluiten of hij ze zelf wil gebruiken (downloading).

OpenEarth koppelt het aanbod van gegevens aan de vraag naar gegevens. Om dat mogelijk te maken moeten vraag en aanbod dezelfde taal spreken. Daarom schrijft OpenEarth voor hoe gegevens bewaard dienen te worden, zodanig dat zowel de inhoud (semantiek) als de wijze van aanbieden (syntax) volgens internationale normen zijn gestandaardiseerd. Dit vergt weliswaar bij de opslag van ingewonnen gegevens een extra investering, maar die kan worden terugverdiend doordat het voor de hergebruiker mogelijk wordt te profiteren van de gegevens die zijn voorganger reeds verzameld heeft. Zonder deze extra investering, is hergebruik achteraf niet (goed) mogelijk en zal er opnieuw gemeten moeten worden, met extra kosten (en tijdverlies) als gevolg. Hoe meer hergebruikers, des te groter het rendement van de extra investering (eenmalig inwinnen, meervoudig gebruik). Ervaringen zoals in Hoofdstuk 1 aangehaald, bevestigen dit beeld.



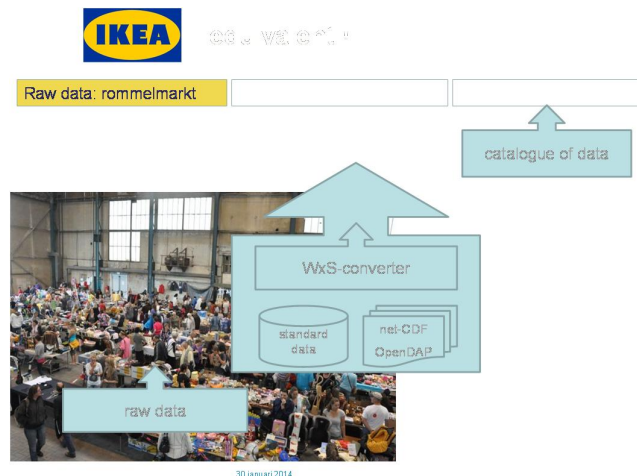
Figuur 2.2 OpenEarth Software Stack Deltares

De Deltares OpenEarth stack is een door Deltares beheerd en onderhouden samenstel van opensource software producten (Figuur 2.2). In z'n algemeenheid bestaat de OpenEarth stack uit een repository, waar de ruwste vorm van data (in versiebeheer) opgeslagen wordt. Vanuit de repository kunnen de opgewerkte data overgezet worden naar een geordende database met geo-informatie functionaliteit, om vervolgens via Geoserver in een catalogus ontsloten te worden. Merk op dat in de figuur twee parallelle sporen, met veel overeenkomsten, te onderscheiden zijn. In het huidige project ligt de focus op het linker spoor. Dit betreft de OpenEarth oplossing voor puntdata. De OpenEarth oplossing voor rasterdata is in het rechterspoor weergegeven en is (en wordt) in andere projecten vormgegeven.

Afhankelijk van de gekozen implementatievorm kunnen delen van de stack door verschillende partijen beheerd worden. Indien gekozen wordt voor beheer door RWS, zullen die componenten deel moeten uitmaken van de RWS bouwstenen catalogus. Zo niet, stelt de RWS bouwstenen catalogus wel eisen aan de interfaces tussen de door RWS en niet door RWS beheerde delen. In de volgende paragrafen worden de functies van de diverse onderdelen van de OpenEarth stack nader besproken. Die zullen aan de hand van een markt analogie geïllustreerd worden.

2.1.1 Repository

De repository is de plaats waar ruwe data opgeslagen worden. Volgens het vijfsterrenmodel van Berners-Lee is dit de (belangrijke) eerste stap. Zorg dat de data, in welke vorm ook, online bereikbaar zijn. De gebruiker moet dan zelf maar uitvinden waar de data te vinden zijn en wat zij voorstellen. Je zou het kunnen vergelijken met een rommelmarkt (Figuur 2.3): er is van alles te krijgen, maar het ligt rijp en groen door elkaar: je moet zelf rondneuzen of er iets van je gading bij is. De rommelmarkt zegt niets over kwaliteit; tussen de rommel kan je ook plotseling op een originele schets van Rembrandt stuiten!



Figuur 2.3 De repository is net een rommelmarkt

2.1.2 Database

Het wordt een stuk eenvoudiger winkelen, als alle producten netjes verpakt en geordend zijn. Dat betekent voor de leverancier wel dat hij zijn spulletjes stuk voor stuk moet verpakken en sorteren, zodat de klant weet wat hij koopt en waar het in het schap staat, maar dat heeft de leverancier er graag voor over. De klanten staan in de rij! IKEA heeft dat goed begrepen: de "Billy" is wereldwijd een van de best verkochte kasten! De stap van repository naar database en Geoserver is een vergelijkbare stap met die van rommelmarkt naar IKEA (Figuur 2.4).



Figuur 2.4 Eenvoudiger winkelen dan op de rommelmarkt

2.1.3 Catalogus

Om de vergelijking nog een stap verder door te voeren: met de IKEA-catalogus in de hand of de IKEA-app op je telefoon/tablet, hoef je de deur niet eens meer uit om je spulletjes in huis te halen (Figuur 2.5). Dat is nog eens comfortabel! Stel nu ook eens dat je vanaf je eigen laptop eerst alle data op je gemakje kunt bekijken en uitproberen in plaats van partij voor partij te benaderen of ze alstublieft een CD-rommetje of USB-stick met de felbegeerde data willen toesturen! Welke data zal het meest gebruikt worden?



Figuur 2.5 Gebruikersgemak ten voeten uit: gebruik de catalogus

2.1.4 Authenticatie

Tenslotte speelt natuurlijk de vraag of eenieder geautoriseerd is om data zomaar op te halen (Figuur 2.6). De eigenaar / bronhouder kan daar, indien gewenst, voorwaarden en beperkingen aan stellen. Bijvoorbeeld een jaar embargo om eigen onderzoekers de ruimte te geven om eerst zelf te kunnen publiceren. Technisch is dit mogelijk door een authenticatiefunctie op te nemen. In het geval van OpenEarth en open data is het streven om data in een zo vroeg mogelijk stadium al naar buiten te brengen. Daar gaat wel een afweging aan vooraf: wie zijn veldboekje online zet, riskeert dat er talloze vragen voor verduidelijking komen, die voor de onderzoeker zelf vanzelfsprekend leken te zijn. Daarom is het handig om enige vorm van kwaliteitscontrole vooraf te verzorgen. Wie echter wacht tot hij volstrekt zeker



Figuur 2.6 Authenticatiefunctie om gebruikersrechten te bepalen

is dat er geen enkele onduidelijkheid of fout meer in zijn gegevens zit, onthoudt de hergebruiker al die tijd toegang tot zijn gegevens en maakt het daarmee voor derden ook niet mogelijk om (eventueel vanuit een ander perspectief) op onvolkomenheden in de data te wijzen. Kwaliteitsverbetering is dus ook mogelijk door feedback van derden te incorporeren. Het moment van naar buiten treden met de data wordt wel een release genoemd. Tussentijdse verbeteringen en aanvullingen op de release worden (in de repository) bijgehouden in zogenaamd versiebeheer (vergelijk Track-changes in een Word-document). Afhankelijk van het aantal en/of het gewicht van aanpassingen kan periodiek besloten worden of er een nieuwe release (Accept Track-changes) wordt uitgegeven. In de repository blijven de track changes ook dan nog bewaard. In Hoofdstuk 5 zal hier in meer detail op worden ingegaan.

2.2 Aquo-standaard

In het bovenstaande zijn de technische randvoorwaarden voor de uitwisseling van gegevens besproken (de syntax). Maar het feit dat gegevens technisch uitgewisseld kunnen worden, biedt nog geen meerwaarde, zolang er onenigheid of onduidelijkheid is over wat de gegevens voorstellen. Daarom is het bij het uitwisselen of hergebruik van gegevens essentieel dat de betekenis van deze gegevens eenduidig is vastgelegd (de semantiek). Zo bepaalt de syntax de technische uitwisselingsstandaard; de semantiek definieert de inhoud.

Binnen de watersector is er in Nederland een semantische standaard ontwikkeld die zorgt voor een uniforme taal: de Aquo-standaard. Het beheer van de Aquo-standaard ligt bij het Informatiehuis Water ([IHW](#)). De Aquo-standaard is ontstaan uit 'best practices' in de watersector. De standaard is bedoeld voor iedereen die te maken heeft met het vastleggen en gebruiken van gegevens in de watersector. Zowel op zee als binnendijs, in beekdalen en polders, bij afvalwater en grondwater, voor ecologie en waterkwantiteit. Door actieve bijdragen van de Aquo-gebruikers blijft de standaard actueel (zie voor ervaringen Bijlage C). Ook wet- en regelgeving bepalen de inhoud van de Aquo-standaard.

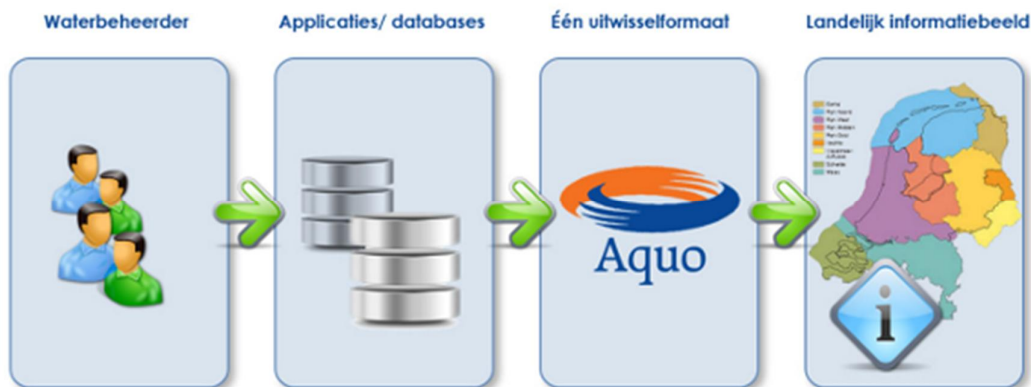
Het ideaalbeeld, gezien vanuit de Aquo-standaard, is dat waterbeherend Nederland gegevens altijd uitwisselt conform de Aquo-standaard. Wat wordt daarmee bereikt?

- Er wordt één taal gehanteerd, zowel in de beleidsstukken als in de digitale wereld. Het Aquo-lex biedt een 'standaardtaal' die als doel heeft om spraakverwarring tussen mensen en/of informatiesystemen te voorkomen. Fouten die bij gegevensuitwisseling

optreden, ontstaan namelijk vaak door verschil in betekenis. Aquo-lex, het 'waterwoordenboek', bevat circa 7500 definities voor termen; in zogenaamde domeintabellen worden de keuzemogelijkheden aangegeven.

- De waterbeheerder is nauwelijks nog tijd kwijt aan het interpreteren of overzetten van informatie van de ene naar de andere database/applicatie/formaat
- De gegevens zijn betrouwbaar, bruikbaar en beschikbaar
- Waterinformatie is over de verschillende domeinen goed te koppelen en landelijke informatiebeelden zijn eenvoudig te maken (zie Figuur 2.7).

Door met één standaard te werken kunnen gegevens van waterbeheerders eenvoudig worden verzameld en beschikbaar worden gesteld aan derden. Daarmee draagt deze manier van werken bij aan nationale beleidskaders zoals Open Data.



Figuur 2.7 Vereenvoudiging informatiestromen door toepassing Aquo-standaard (bron: www.aquo.nl)

Rijkswaterstaat heeft gekozen voor het adopteren van de Aquo-standaard als semantische standaard. Deze keuze sluit tevens aan bij de keuze van het Informatiehuis Mariene (IHM), dat uniformering van mariene data nastreeft. Zij heeft daartoe onder meer het Protocol Mariene Data opgesteld ([referentie](#)).

Het is niet vanzelfsprekend dat een Nederlandse standaard ook wereldwijd als standaard aangenomen wordt. Voor mariene data bestaan ook diverse internationale standaarden. Vanuit het OpenEarth perspectief is het niet zozeer van belang welke standaard er gekozen wordt, als er maar een standaard gekozen wordt. De standaard met het grootste draagvlak geniet daarbij wel de voorkeur. Waar niet gekozen wordt voor een internationale standaard, is het vaak mogelijk een "mapping" te maken van de lokale standaard naar de internationale standaard, waardoor uitwisseling van gegevens toch mogelijk wordt. De keuze voor Aquo in dit project is vooral ingegeven door de vraag van RWS (en IHM) voor het gebruik ervan. Vanuit Aquo is er een mapping naar SeaDataNet, zodat internationale ontsluiting van gegevens die op de Aquo-standaard gebaseerd zijn, relatief eenvoudig is (zie Bijlage D).

2.3 Koppeling RWS-IV architectuur: datadistributielaag

RWS hanteert voor de implementatie van haar softwaresystemen een RWS-IV architectuur. Alle door RWS beheerde data (lange termijnmonitoring en projecten) worden in principe via de Watermanagement distributielaag (ook wel datadistributielaag of distributielaag, DL genoemd) ontsloten, waarbij de aansluiting moet voldoen aan de distributielaag standaarden. De toegepaste software moet voldoen aan de RWS bouwstenen catalogus. Deze catalogus biedt oplossingen aan op verschillende niveaus: data-laag, applicatielaag, presentatielaag en eventueel proceslaag, in internationaal jargon aangeduid als HaaS, PaaS, SaaS en DaaS

(voor resp. Hardware as a Service, Platform as a Service, Software- en Data as a Service). Ook als een softwarelaag door externen beheerd wordt, dient deze, samen met de onderliggende niveaus, aan de hostingeisen uit de bouwstenen catalogus te voldoen.

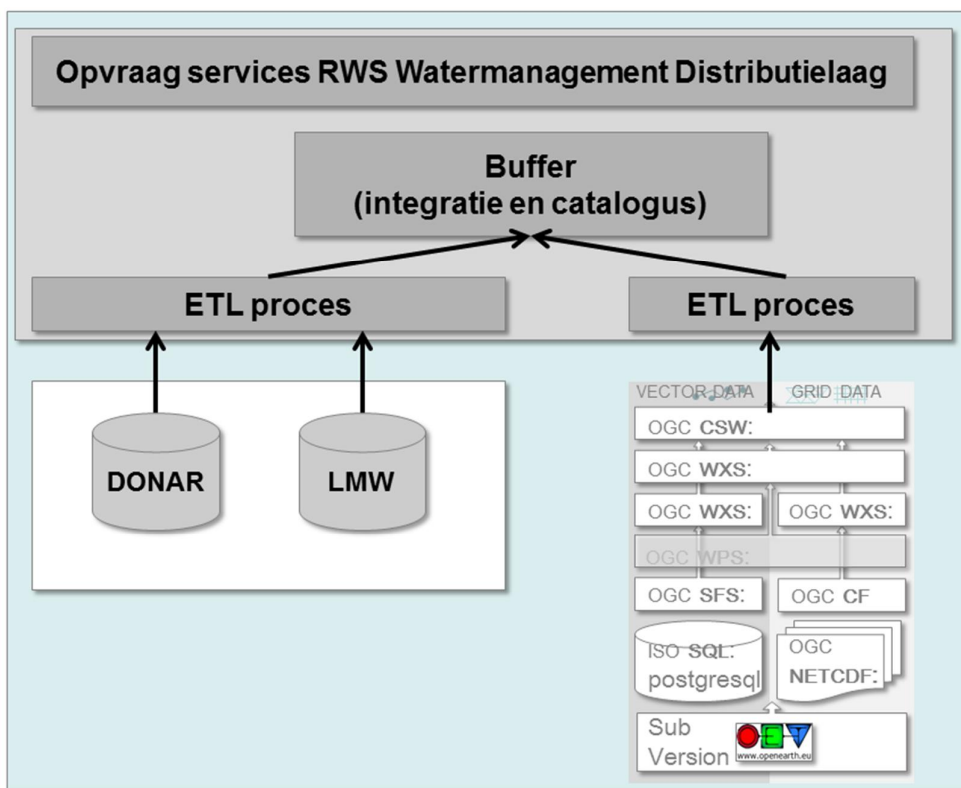
De kaders en randvoorwaarden waaraan de OpenEarth oplossing voor de opslag en distributie van mariene data moet voldoen, zijn beschreven in een zogenaamde Project Start Architectuur (PSA). In deze PSA (Dirksen & Van Splunder, in concept, 2014) beveelt RWS aan dat de OpenEarth software stack door RWS zal worden voorgeschreven bij de aanbesteding van biologische monitoring projecten voor de opslag en distributie van de verzamelde gegevens. Op deze manier wordt de opslag en distributie van de monitoring gegevens geüniformeerd.

Het geheel zal functioneren via de Watermanagement distributielaag van RWS en, indien operationeel, via het waterportaal als ingang. De oplossing moet zowel technisch als organisatorisch voorzien in de mogelijkheid om te draaien:

- geheel binnen de RWS firewall (binnen overheids-raamcontract centrale dataopslag RWS) met koppeling naar de DL;
- geheel buiten de RWS firewall met koppeling naar de DL;
- deels erbinnen en deels erbuiten, waarbij alle delen een koppeling hebben naar de DL.

Om te voorzien in de klantbehoefte zijn in deze PSA de volgende ontwerpbeslissingen genomen:

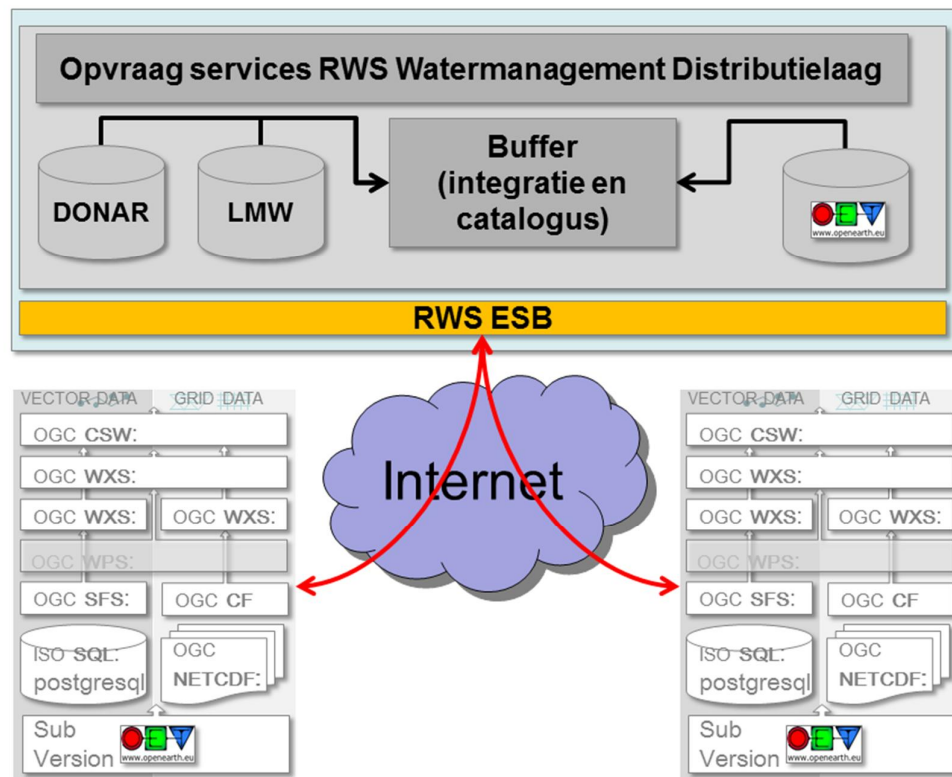
- De data-integratie tussen partners wordt gerealiseerd door bij alle partners dezelfde OpenEarth-architectuur te gebruiken
- De integratie binnen RWS van monitoring en projectdata wordt conform de Domeinarchitectuur vorm gegeven binnen de Distributielaag in overeenstemming met de daarbij gehanteerde standaarden.



Figuur 2.8 Architectuur DL (naar: Flip Dirksen, Ingeborg van Splunder, 2014, PSA Mariene Data) met aansluiting naar Landelijk Meetnet Water (LMW), DONAR en een OpenEarth stack binnen RWS

In Figuur 2.8 is de architectuur van de 2.0 versie van de distributielaag gegeven met daarin de benodigde aanpassingen voor het ontsluiten van de OpenEarth software stack. De in de OpenEarth stack gedefinieerde webservices moeten voldoen aan de meest recente OGC standaard voor WFS (Web Feature Service) en CSW (Catalogue Service for the Web). Als blijkt dat deze webservices niet voldoende performance kunnen bieden, kan als alternatief gebruik gemaakt worden van JSON-services zoals nu plaatsvindt tussen het Waterportaal en de distributielaag voor zowel het bijwerken van de catalogus als het ophalen van waarnemingen.

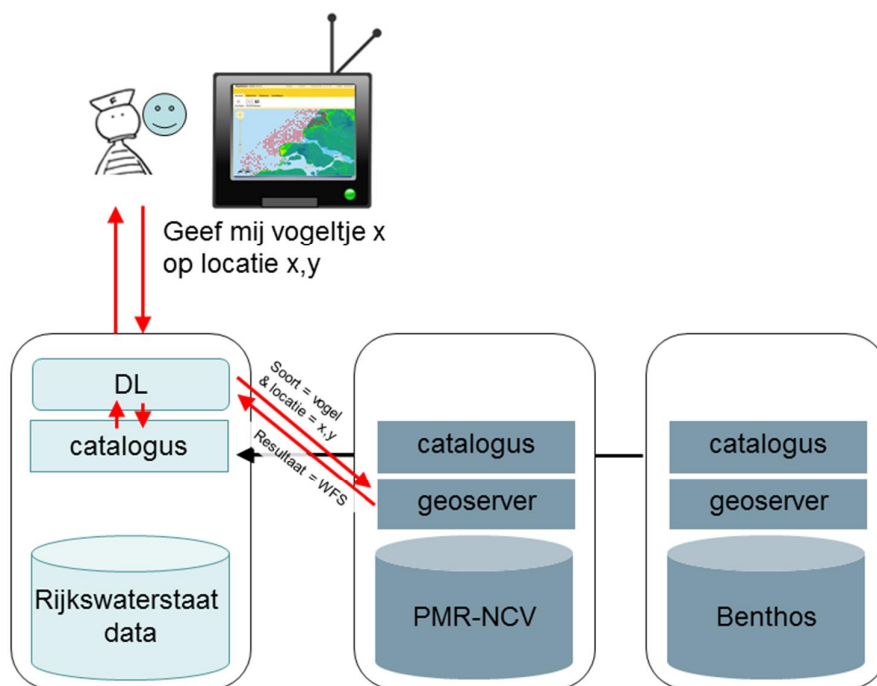
Via de distributielaag is het mogelijk om via webservices informatie op te vragen over locaties, waarnemingen en metadata, binnen de RWS firewall. Voor koppeling van bronnen op basis van de beschreven webservices die zich buiten de RWS firewall bevinden, dienen extra maatregelen getroffen te worden. De architectuur voor een DL met externe bronnen is weergegeven in Figuur 2.9.



Figuur 2.9 Architectuur DL met externe bronnen (naar: Flip Dirksen, Ingeborg van Splunder, 2014, PSA Mariene Data)

De webservices die de distributielaag gebruikt om een bron te ontsluiten buiten de RWS firewall, lopen via de RWS-Enterprise Service Bus, ESB. De OpenEarth stack voorziet alle databronnen van een set OGC-services die een interface vormen met de, conform de door de DL gehanteerde standaarden. Deze webservices worden vanuit de distributielaag aangeroepen zodra er een verzoek is ontvangen voor data uit deze bronnen. Ook worden zij periodiek opgeroepen om de DL-catalogus up-to-date te houden. RWS kan vervolgens kiezen voor een tijdelijke opslag (caching) in de DL ten behoeve van integratie van de data uit verschillende bronnen of voor een meer permanente opslag (vanwege performance en/of beschikbaarheid).

De DL maakt geen kopie van de data van alle databronnen, maar neemt van de aangesloten catalogi alleen de metadata over. Een zoekopdracht binnen de DL kan als resultaat hebben dat een dataset van bijvoorbeeld PMR-NCV wordt opgevraagd. Figuur 2.10 maakt duidelijk hoe een dergelijke bevraging in zijn werk gaat. Bij de specifieke zoekopdracht naar een vogelsoort op locatie x, y, leidt de DL de gebruiker naar de services die door de (externe) database van PMR-NCV worden vrijgegeven. Zodra de DL van de interne catalogus te horen krijgt dat de gevraagde data niet binnen de firewall beschikbaar zijn, maar daarbuiten (immers de interne catalogus houdt bij wat er in de catalogi van de externe databronnen beschikbaar is), bevraagt hij de aangesloten databronnen buiten de firewall. De OpenEarth stacks leveren vervolgens de opgevraagde data via OGC webservices.



Figuur 2.10 OpenEarth ontsluit verschillende (externe) databases via een aanspreekpunt: de DL

Een complicerende factor is dat gedurende de uitvoering van het project, de ontwikkeling van de distributielaag vertraging heeft opgelopen. Hierdoor is de mogelijkheid om de uitwisseling tussen de OpenEarth stack en de DL te testen, komen te vervallen. Ook de centrale bereikbaarheid van data via één ingang, de DL, kon binnen het tijdsbestek van het project niet worden gerealiseerd. Om niettemin de werking van de OpenEarth stack vast te kunnen stellen, heeft Deltares een tijdelijke viewer aangeboden en ingezet?, waarmee het toch mogelijk werd de opgewerkte en gestandaardiseerde data uit de externe databronnen elk afzonderlijk te visualiseren.

2.4 Aansluiting bij dataportalen

Het feit dat OpenEarth op open (inter-) nationale standaarden is gebaseerd, maakt de stack toch bruikbaar, ook al is de distributielaag op dit moment niet beschikbaar. De volgende voorbeelden illustreren dat.

Internationaal

Internationaal wordt de OpenEarth stack toegepast in het EMODnet-project. EMODnet is het 'European Marine Observation and Data network' wat een lange termijn initiatief is voor mariene data van het Directoraat Generaal voor Maritieme aangelegenheden en Visserij (DG MARE) van de Europese Commissie. Met EMODnet wordt de mariene kennisstrategie 2020 onderschreven.

Voor het portal EMODnet Chemistry heeft Deltares een taak in de dynamische ontsluiting van gegevens.

Voor EMODnet Biology is overeengekomen om de T₀ observaties van biota van de PMR-NCV data in het zogenaamde OBIS formaat uit te wisselen via CSW (Catalogus) en WFS.

Nationaal

Het Nationaal Georegister, en het opendata portaal van de Nederlandse overheid zijn in Hoofdstuk 1 al even aangestipt. Zij hosten beide een web-based catalogus voor web services, waarbij NGR metadata doorlevert aan het opendata portaal. Beide catalogi bevatten data van Nederlandse overheidspartijen, die andere partijen integraal kunnen overnemen, en samenvoegen met behulp van de OGC CSW standaard voor het uitwisselen van catalogi. OpenEarth stacks zouden zo aan deze catalogi doorgegeven kunnen worden.

In een eerder uitgevoerd project (DANK: Digitale Atlas Natuurlijk Kapitaal) heeft Deltares laten zien dat de OpenEarth workflow voldoet aan de eisen die het Nationaal Georegister stelt. Via OGC standaarden worden datasets beschikbaar gesteld aan DANK. Deze worden vervolgens een thematische dataset binnen het Nationaal Georegister. Een van de voordelen van deze werkwijze is dat data vanuit diverse bronnen afkomstig kunnen zijn waarbij het Nationaal Georegister fungeert als zoekmachine.

Ook het Informatiehuis Marien is bezig met een portaal voor het ontsluiten van mariene data. Het IHM portaal maakt ook gebruik van de OpenEarth services. Deltares adviseert IHM m.b.t. een viewer, die de data via WFS van de Deltares-server moet visualiseren.

3 Mariene projectdata

Voor het huidige project heeft RWS zich tot doel gesteld om de (lange termijn) bereikbaarheid van de data van een aantal mariene projecten, zoals PMR-NCV, Zandmotor en Wind op Zee, te borgen. Deze projecten hebben met elkaar gemeen dat ze data inwinnen waarmee de effecten van menselijke ingrepen in het mariene systeem geanalyseerd kunnen worden. De data van de verschillende projecten zullen worden opgenomen en ontsloten in een uniforme, via een selectietool benaderbare, gedistribueerde database. Voor reeds afgesloten projecten moet de uniformering noodgedwongen na afloop van het project, in een extra slag, plaatsvinden; voor nieuwe projecten kunnen van tevoren afspraken gemaakt worden, waardoor uniformering letterlijk van meet af aan kan plaatsvinden (zie Tabel 3.1). Hiervoor is de handreiking datamanagement opgesteld. Het resultaat zal zijn dat de data zowel tijdens de uitvoering van het project, als direct na afloop van het project eenvoudig toegankelijk zijn.

Tabel 3.1 Datamanagement voor afgesloten, lopende en nieuwe marine projecten

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
PMR-NCV															
Zandmotor															
Wind op Zee Shortlist															
MEP NearShoreWind															
MEP Bestaande Duinen															
MEP Nieuwe Duinen (Kappittelduinen)															
Benthos/Vis															
RijMaMo															
PoC Lodingen															
Ecolims-benthos															
BWN															
KPP B&O Kustlijnzorge															
MCI Old data															

In de volgende paragrafen volgt een korte beschrijving van de betreffende mariene projecten. Daarbij wordt ingegaan op de doelstellingen van de projecten en de gegevens die daarvoor verzameld zijn of nog verzameld moeten worden en de wijze waarop de gegevens geborgd worden.

De (opgewerkte) data zijn te vinden op:

<http://marineprojects.openearth.eu/geonetworkhttp://pmr-geoserver.deltares.nl/geonetwork/srv/en/main.home>

<https://zandmotordata.nl/>

<http://marineprojects.openearth.eu/geonetwork>

Zoals eerder opgemerkt, vindt de ontsluiting niet plaats via de DL. De lokale ontsluiting via de bovengenoemde servers laat evenwel het potentieel zien van ontsluiting via de OpenEarth stacks. Op laatstgenoemde server zijn meer data ontsloten dan strikt genomen voor dit project gevraagd. Daarvoor is gekozen, omdat het opnemen van de (ruwe) data in de repository (één ster volgens Berners-Lee) relatief eenvoudig is en weinig kosten met zich meebrengt. Daarmee zijn de data toch vindbaar en toegankelijk, zij het in onbewerkte vorm. Wanneer de gelegenheid zich voordoet (daarover meer in Hoofdstuk 5) kan alsnog opwerking naar hogere niveaus plaatsvinden.

3.1 PMR-natuurcompensatie Voordelta

Het project PMR-natuurcompensatie Voordelta (PMR-NCV) richt zich op de monitoring van de effecten van de maatregelen die getroffen zijn om de gevolgen van de aanleg van de Tweede Maasvlakte in het Natura 2000-gebied Voordelta te compenseren. Door de aanleg

van de haven gaat 2.455 hectare van habitatype H1110 (ondiepe zeeën met zandbanken die permanent onder water staan) verloren en worden er negatieve effecten verwacht op de foerageerfuncties voor de grote stern, visdief en de zwarte zee-eend. Ter compensatie is er binnen de Voordelta een bodembeschermingsgebied ingesteld met daarbinnen een aantal rustgebieden. In het bodembeschermingsgebied worden beperkingen opgelegd aan vormen van visserij die de zeebodem beroeren, waardoor de bodemdierenbiomassa toe zal kunnen nemen. In de rustgebieden geldt een verbod op, of beperking van, menselijke activiteiten om het verlies in foerageerfunctie als gevolg van de aanleg van Maasvlakte 2 te compenseren zodat de beschermde vogelsoorten optimaal kunnen profiteren van het voedselaanbod.

Het effect van de maatregelen wordt bepaald door de ontwikkelingen in het bodembeschermingsgebied en de rustgebieden te vergelijken met ontwikkelingen in referentiegebieden, waar geen compensatiemaatregelen getroffen zijn. De uitgangssituatie - vóór de aanleg van Maasvlakte 2- is in de periode 2004-2007 vastgelegd door het uitvoeren van een breed scala aan nulmetingen. Het monitoringprogramma richt zich niet alleen op de primaire parameters die direct voortvloeien uit de compensatieopgave, zoals de bodemdierenbiomassa of aantallen vogeldagen, maar bestaat uit een veel breder meetprogramma. Het programma is zo ingericht dat niet alleen een antwoord gegeven moet kunnen worden op de vraag of de compensatieopgave is behaald, maar daarnaast ook veel informatie beschikbaar komt over het functioneren van de Voordelta in ecologische zin en over de rol van menselijke activiteiten in dat gebied. Deze kennis kan gebruikt worden om zonnig de in 2008 genomen compensatiemaatregelen beter te onderbouwen dan wel bij te stellen. PMR-NCV kan profiteren van de biotische en abiotische gegevens die via de andere mariene projecten over een groter gebied verzameld worden. De eerste fase van monitoring is in 2013 afgesloten. Thans worden voorbereidingen getroffen voor de uitvoering van de tweede fase van monitoring voor de periode tot en met 2019.

Borging van de data in een online repository

Het PMR-NCV project heeft aan de wieg gestaan van de huidige serie overkoepelende projecten op het gebied van datamanagement voor mariene projecten. In PMR-NCV is het initiatief genomen om de data die door de diverse partijen werden ingezameld centraal op te slaan in een repository en via de (toen in beginstadium van ontwikkeling verkerende) OpenEarth software stack te ontsluiten. Gaandeweg zijn oplossingen gezocht voor knelpunten en heeft overleg met RWS plaatsgevonden om de aansluiting van de stack met de RWS-IV- infrastructuur te garanderen.

Opwerking van de data tot Aquo conforme data

PMR-NCV heeft voor de datalevering “in een voor RWS acceptabel formaat” al in de startfase gekozen voor het hanteren van de volgende internationale standaarden:

- WORMS (World Of Marine Species, <http://marinespecies.org>), waar het de soorten lijst betreft
- CF standaard (<http://cfconventions.org>) voor fysische parameters
- HILUCS (Hierarchical Land Use Classification System) voor gebruiksparemeters

Als uitkomst van de pilot en in overeenstemming met het vervolgtraject heeft op verzoek van RWS een vertaling plaatsgevonden naar de Aquo-standaard. Voor PMR NCV betekende dit een mapping van de internationale standaarden naar Aquo. Een andere mogelijkheid zou bestaan uit het totaal opnieuw opbouwen van de database. Daarvoor is niet gekozen. De opbouw van het gebruikte datamodel komt tegemoet aan de minimum afspraken die met IHW, RWS, Deltares, IMARES en Ecosys zijn overeengekomen (zie Bijlage B).

Tabel 3.2 *Opwerkingspercentage van data uit PMR-NCV*

Beschikbaarheid	Percentage	Tim Berners-Lee classificatie
% opgeslagen in Repository	85	*/**
% in Database	80	***
% OGC	80	****
% Aquo conform	80	****
Catalogus	Ja	****

Tabel 3.2 geeft aan welke percentages van de data beschikbaar zijn in welk format. Het traject van opwerking start onderin bij de SubVersion repository. Gezien de structuur van het project is ervoor gekozen om de observaties die gedaan zijn door de partijen in de PostgreSQL/PostGIS database te laten plaatsnemen. Dat is in onderling overleg gedaan onder meer door het geven van workshops.

De data van het laatste meetjaar 2014 zijn nog niet geheel door de partijen verwerkt, vandaar dat deze gegevens ook niet allemaal (=100%) beschikbaar zijn. Dat er meer data in de repository staat dan in de overige onderdelen (de catalogus even buiten beschouwing gelaten), heeft de volgende oorzaken:

- Data kunnen niet verwerkt worden (Bijvoorbeeld: INBO data van vogels betreft observaties zonder exacte locatiegegevens)
- Data zijn onlangs ter beschikking gesteld
- Data bevatten gegevens over de fysieke gesteldheid van de observatie, waaraan geen prioriteit is gegeven

De schematische weergave van het opwerkingspercentage van de PMR-NCV data is een gemiddelde over alle percelen. Binnen de percelen (perceel vogels, vissen, benthos, gebruik en abiotiek) zijn grote verschillen met betrekking tot bovenstaande. Vooral het perceel vis ligt erg achter op data verwerking, controle en analyse ten opzichte van abiotiek en vogels (beiden 100% opgewerkt op alle terreinen).

In de tabel is tevens aangegeven over welke sterrenstatus de data beschikken. Vijfsterren in termen van Tim Berners-Lee worden niet gehaald. De data wordt op dit moment niet verder gelinkt naar andere data.

Verwachte aanpassing in volgende release

Zoals in paragraaf 2.1.4 aangegeven, kunnen er aan het vrijgeven van data voorwaarden gesteld worden. Als aan die voorwaarden is voldaan, kan een nieuwe versie van de data uitgegeven worden. Dit heet een release. Op dat moment kunnen eventuele fouten of onvolkomenheden in de dataset hersteld worden. De verwachting is dat er, door het intensievere gebruik van open data, meer feedback vanuit de gebruikers over de kwaliteit van de data zal plaatsvinden, waardoor actualisatie en aanpassing van de dataset nodig zal blijken. Om hieraan tegemoet te komen, kan een planning opgesteld worden voor de volgende release. Dat kan zijn op basis van het aantal correcties dat aangemeld is, maar kan ook op periodieke basis. Dit is aan de beheerder om te besluiten. De volgende release van de PMR-NCV data zou kunnen plaatsvinden als alle gegevens zijn verwerkt en dus 100% van de data in de repository beschikbaar is.

Wat betreft de opwerking van data is het de verwachting dat 100% van de observaties die bijdragen aan de beantwoording van de PMR-NCV vraagstelling beschikbaar zullen zijn. Er zijn echter beschrijvende gegevens van de survey, waarvoor het de vraag is of die opgenomen moeten worden. Dat zijn gegevens als:

- Dieetonderzoek kuikens (geen locatie-gegevens, alleen gebiedsnamen)
- CTD data van vis en benthos survey's
- Weersomstandigheden van gebruiksonderzoek en vogeltellingen

Deze gegevens vormen naar schatting nog geen 5 procent van de totale hoeveelheid gegevens. De opwerking van deze beschrijvende gegevens is een intensief karwei, waarvan het de vraag is of dat een goede besteding is.

Als alle PMR-NCV data verzameld zijn, zal blijken in hoeverre deze een plek hebben kunnen vinden in de Aquo-domeintabellen, of dat deze moeten worden uitgebreid. Vooral wat betreft de biota zal, als een juiste vorm van wijzigingsvoorstellen is gevonden, Aquo naar verwachting een groter aantal soorten bevatten dan nu het geval is. In Bijlage C is verder uitgewerkt welke soorten binnen PMR-NCV zijn gevonden en gematcht zijn met de WoRMS lijst. Deze lijst is vergeleken met de biotaxon lijst (TWN-Lijst) om een Aquo conforme WFS service te kunnen opstellen. Het verschil tussen de lijsten mondt uit in een wijzigingsverzoek dat wordt voorgelegd aan het IHW.

3.2 Zandmotor

De Zandmotor is een kunstmatig schiereiland, opgespoten tussen Ter Heijde en Kijkduin. Het innovatieve aspect van de Zandmotor is dat er gebruik gemaakt wordt van de krachten van de natuur voor de bescherming tegen de zee. Door wind, golven en stroming verspreidt het opgespoten zand zich de komende twintig jaar langs de kust tussen Hoek van Holland en Scheveningen. Het vormt daar nieuw strand en duin, dat het achterland tegen zeespiegelstijging beschermt en extra ruimte voor natuur en recreatie biedt.

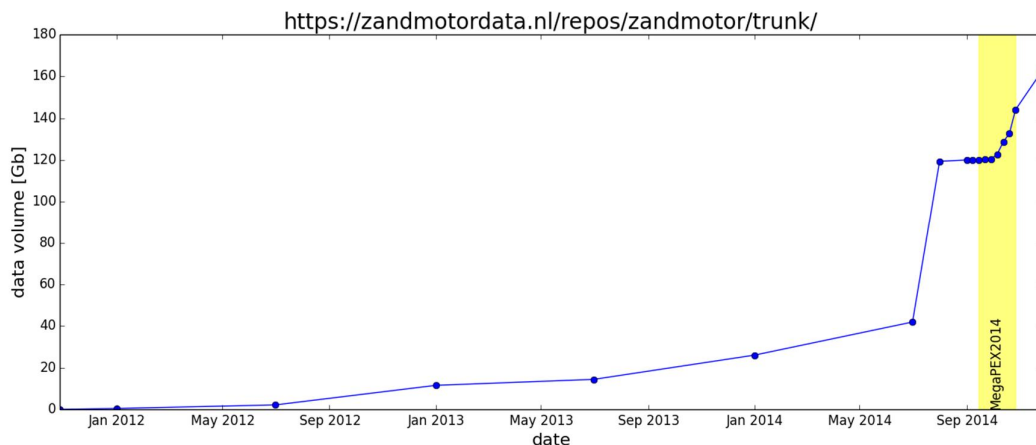
Tussen maart 2011 en oktober 2011 legden Rijkswaterstaat en de provincie Zuid Holland het schiereiland in de vorm van een haak aan. Sleephopperzuigers zogen het zand tien kilometer ver uit de kust op en brachten het vervolgens naar de juiste plek. Het schiereiland stak bij aanleg één kilometer ver in zee. De aanleg van de Zandmotor is voltooid in november 2011.

Onderzoek moet uitwijzen of de Zandmotor zich inderdaad ontwikkelt zoals modellen voorspeld hebben en meerwaarde heeft voor recreatie en natuur. Rijkswaterstaat, Ecoshape, verschillende universiteiten en kennisinstellingen zijn bezig om data over morfologie, hydrologie, ecologie, bestuurlijke aspecten en recreatief gebruik te verzamelen en te analyseren. Twee wetenschappelijke onderzoeksprogramma's richten zich geheel (STW onderzoeksprogramma NatureCoast) of gedeeltelijk (NEMO) op de Zandmotor. Metingen vinden onder meer plaats in de duinen, op de Zandmotor zelf en rondom de Zandmotor. Vanuit NatureCoast zijn al diverse presentaties en publicaties gerealiseerd. In totaal zijn er in

het onderzoeksprogramma 12 promovendi, 3 postdocs en diverse afstudeerders actief. Het onderzoek naar de ontwikkeling van de Zandmotor loopt door tot 2021. In 2016 worden de eerste conclusies getrokken over dit pilotproject.

Borging van de data in een online repository

De ruwe zandmotor data wordt opgeslagen in een repository die sinds eind oktober 2011 in de lucht is. De ruwe data is vooral in 2014 sterk toegenomen, o.a. tijdens de gezamenlijke meetcampagne (MegaPEX 2014) die in september-oktober heeft plaatsgevonden.



Figuur 3.1 Ontwikkeling van ruwe zandmotor datavolume in de tijd

Opwerking van de data tot Aquo conforme data

Het OpenEarth DataLab t.b.v. de zandmotor data bevat zowel een THREDDS server om netCDF bestanden aan te bieden als een PostgreSQL database voor vector data. Sinds het beschikbaar komen van de infrastructuur om netCDF data op te werken (zomer 2014), zijn er acht datasets die als zodanig beschikbaar zijn gemaakt. De infrastructuur om data in de database te laden, middels een script, is inmiddels (eind 2014) beschikbaar maar bevat nog wat kinderziektes. Zodra de laatste problemen opgelost zijn, kunnen de eerste drie datasets, waarvoor scripts gemaakt zijn, in de database geladen worden. Er zijn op dit moment dus nog geen puntdata in de database geladen.

Conform de andere overzichten is zo goed als mogelijk aangegeven welke percentages kunnen worden toegekend. 100 % betekent hier dus dat alle gegevens die tot nu toe binnen zijn voor het specifieke onderwerp verwerkt zijn tot aan OGC en Aquo conforme datasets. Hierbij moet dan wel de kanttekening worden geplaatst dat er nog geen specificaties zijn gedefinieerd voor Aquo conforme rasters (WMS, WCS).

Tabel 3.2 Overzicht zandmotor datasets en status Zandmotor

#	path	date	raw size	OGC	% Aquo	Catalogus
1	aeolian/megapex	28-11-14	914.3 MB	netCDF		Nee
2	ecology/benthos/coast	01-07-14	1.1 MB			Nee
3	ecology/fishes	01-07-14	372.5 KB			Nee
4	meteohydro/Golfkarakteristieken	21-07-14	366.7 MB			Nee
5	meteohydro/Golftransformatie	26-11-14	124.0 MB			Nee
6	meteohydro/STRAINS	21-07-14	2.3 GB			Nee
7	meteohydro/hydrodynamics/adcp	29-07-14	1.7 GB	netCDF		Nee
8	meteohydro/hydrodynamics/drifters	24-11-14	304.8 MB	netCDF		Nee
9	meteohydro/hydrodynamics/drone	20-10-14	1.2 GB			Nee
10	meteohydro/hydrodynamics/tiltmeters/ihe2014	21-07-14	44.4 MB			Nee
11	meteohydro/hydrodynamics/tiltmeters/wltrial	21-07-14	6.1 MB			Nee
12	meteohydro/meteo	09-10-14	10.2 KB			Nee
13	meteohydro/wind	21-07-14	5.9 MB	netCDF		Nee
14	meteohydro/xband	29-11-14	13.8 GB			Nee
15	morphology/ARGUS/cBathy	24-02-14	1.9 GB			Nee
16	morphology/ARGUS	10-06-14	1.9 GB			Nee
17	morphology/CombiDelflandseKust	10-06-14	264.5 MB			Nee
18	morphology/JARKUS	10-06-14	11.8 MB			Nee
19	morphology/JETSKI	29-11-14	398.3 MB	netCDF PostgreSQL	100	Nee
20	morphology/LIDAR	10-06-14	49.4 MB			Nee
21	morphology/Multibeam	10-06-14	451.9 MB			Nee
22	morphology/NeMo	02-12-14	157.5 MB	netCDF		Nee
23	morphology/construction	11-10-14	5.3 GB	PostgreSQL	100	Nee
24	morphology/megapex/topography	01-12-14	19.2 MB	netCDF/ PostgreSQL	100	Nee
25	morphology/sediment_composition/Data_T0_Oct 2010_Imares	29-04-14	4.2 MB			Nee
26	morphology/sediment_composition/Data_T2_Aug 2012_Imares	29-04-14	3.8 MB			Nee
27	morphology/sediment_composition/Data_T3_Feb 2013_NEMO	30-04-14	48.4 MB			Nee
28	morphology/sediment_composition/Data_T4_Oct 2013_Imares	28-04-14	4.6 MB			Nee
29	morphology/sediment_composition/Data_T5_Feb 2014_NEMO	28-04-14	1.0 MB			Nee
30	morphology/sediment_composition/MacroPhotography_NEMO	06-01-14	6.8 GB			Nee
31	morphology/sediment_composition	05-05-14	6.9 GB	netCDF	100	Nee
32	naturedunes/01_sanddynamics	25-11-14	15.6 GB			Nee
33	naturedunes/02_dynamic_geomorphology	22-10-14	6.7 MB			Nee
34	naturedunes/03_altitudechanges	23-10-14	9.3 GB			Nee

35	naturedunes/04_sand_spray	21-10-14	146.5 MB			Nee
36	naturedunes/05_saltspray	21-10-14	32.7 MB			Nee
37	naturedunes/06_windenergy	21-10-14	186.3 MB			Nee
38	naturedunes/07_water_levels	21-10-14	1.5 GB			Nee
39	naturedunes/08_vegetation_habitats_enroachment	23-10-14	514.4 MB			Nee
40	naturedunes/09_vegetation_habitats_rewetting	23-10-14	20.7 MB			Nee
41	naturedunes/10_vegetation_sand_deposition	21-10-14	453.3 MB			Nee
42	naturedunes/11_higher_plants	21-10-14	2.2 MB			Nee
43	naturedunes/14_breeding_birds	22-10-14	13.8 MB			Nee
44	naturedunes/15_embryonal_dunes	21-10-14	897.0 KB			Nee

Verwachte aanpassing in volgende release

Zowel zandmotor monitoring als de onderzoeksprojecten (NatureCoast en NEMO), waar data verzameld wordt, lopen nog een aantal jaren door. Het aantal datasets en de hoeveelheid data per dataset zal dus nog verder groeien. De scripts die er al zijn, zijn zo goed mogelijk voorbereid om nieuw binnengekomen data zonder veel moeite op te werken.

3.3 Shortlist Wind op Zee

De Shortlist Ecologische Monitoring Windmolenparken op Zee (2010-2011), kortweg Shortlist, komt voort uit het "Masterplan Ecologische Monitoring Windmolenparken op Zee", een gezamenlijk initiatief van de ministeries van IenM en EZ. Doel van het Shortlist-onderzoeksprogramma is het verwerven van aanvullende kennis rond de effecten van windmolenparken op zee. Deze kennis is onder meer nodig voor de vergunningverlening. Onderzoeksinstituten TNO en IMARES voeren de Shortlist-onderzoeken uit, in samenwerking met onder meer SeaMarco en Bureau Waardenburg. Het Shortlist project is afgesloten.

Het onderzoek bestaat uit zeven deelonderzoeken

- 1) Vislarven survey
- 2) Vislarven ingreepeffect onderzoek
- 3) Zeevogeltellingen (vliegtuig en schip)
- 4) Bruinvistellingen (vliegtuig en schip)
- 5) Kleine Mantelmeeuw verspreidingsonderzoek
- 6) TTS-onderzoek bij zeezoogdieren
- 7) Geluidsonderzoek

Vooraf voor de onderdelen 1, 3, 4 en 5 zijn grote hoeveelheden data verzameld.

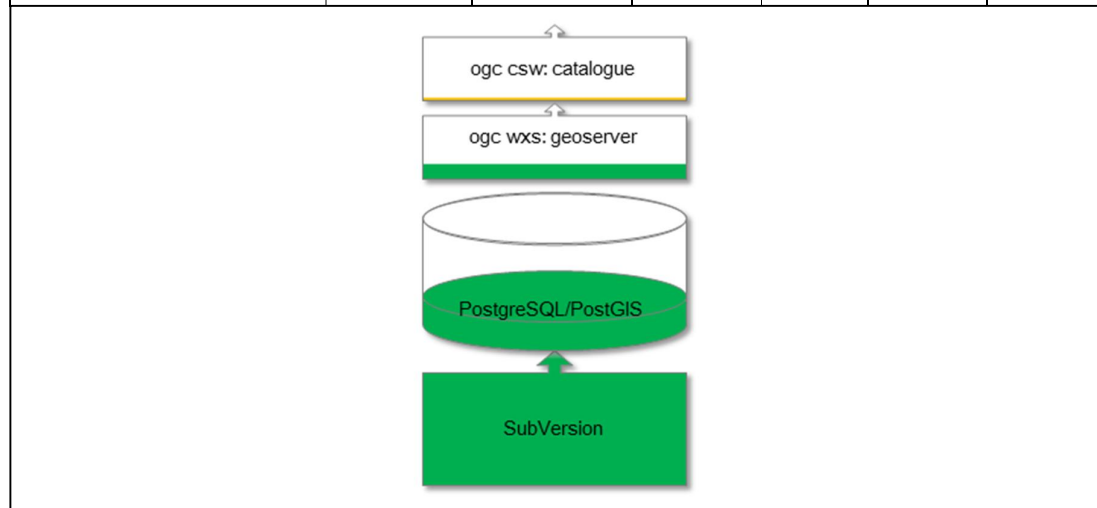
Borging van de data in een online repository

De data van het Shortlist Wind op Zee monitoring project zijn na afloop van het project niet direct online beschikbaar gesteld. Wel is door Deltares, na afloop van het monitoringproject, een advies uitgebracht over de oplevering van de data (referentie Stolte et al., 2011). Begin 2014 was er nog geen beslissing genomen over de opslag van deze gegevens, en de data waren dan ook niet online beschikbaar, of in een database (zoals DONAR) verwerkt. Ook bestond er geen volledige duidelijkheid of opgewerkte en/of ruwe data bij het afsluiten van het project opgeleverd waren aan RWS, in ieder geval waren deze begin 2014 niet bij RWS beschikbaar. Juni 2014 is de dataleveranciers (onderzoekers van IMARES en BuWa) gevraagd om de data op een repository te zetten. Aangezien dit ten tijde van de oplevering van het project niet gebeurd was, heeft RWS extra budget beschikbaar moeten stellen aan de onderzoekers ter compensatie voor de tijd die nodig is voor het opnieuw uitzoeken, controleren, en transporteren van de data naar een repository ten behoeve van het

datamanagement project. Oktober 2014 waren alle data beschikbaar gemaakt op de Deltares repository, maar nog zonder metadata. Over het begrip metadata blijken veel misverstanden te bestaan. Vaak wordt er wel een beschrijving van de data (velden) meegeleverd. Uiteindelijk is door Deltares een template voor metadata aangemaakt op projectniveau, waaraan door de dataleveranciers de dataset specifieke metadata (contactgegevens, keywords) zijn toegevoegd. Hiervoor is in eerste instantie de INSPIRE metadata editor gebruikt. Informatie over de inhoud van de data (bounding box, temporele dekking etc.) wordt later door Deltares aan de metadata toegevoegd op basis van de data. November 2014, dus ruim drie jaar na afloop van het project zijn alsnog voor alle datasets metadata beschikbaar gekomen. Deze metadata zijn echter nog niet door de dataleveranciers gecontroleerd, en dus nog niet definitief.

Tabel 3.3 Opwerkingspercentage van data uit het Shortlist Wind op Zee project

thema	% opgeslagen in Repository	% in Database	% OGC WFS/WMS	% Aquo conform	catalogue	Tim Berners-Lee *
Kleine Mantelmeeuw	100	100	100	0	Ja	***
Vislarven verspreiding	100	0	0	0	Ja	*
Vliegtuigtellingen zeevogels en bruinvissen					Ja	
Ruwe data					Ja	
Opgewerkte data	100	100	100	100	Ja	***
Schiptellingen zeevogels	100	0	0	0	Ja	*
Schiptellingen bruinvissen	100	0	0	0	Ja	*



Opwerking van de data tot Aquo conforme data

Op dit moment zijn nog niet alle data opgewerkt tot Aquo-standaard. Voor opgewerkte data (aantal per km² voor de belangrijkste soorten) van vogels en bruinvissen geteld vanuit het vliegtuig is dit wel gedaan. Dit zijn relatief simpele data, die maar een kleine inspanning nodig hebben om ze Aquo-conform on line uit te leveren. Echter, er is veel informatie verloren gegaan over bijvoorbeeld hoeveel individuen geteld zijn, meteorologische condities, en informatie over de uitvoerende persoon. Het voordeel is dus dat de verwerking van de data tot Aquo-standaard service zeer snel kan gaan, en dat er direct bruikbare data beschikbaar

zijn (verspreiding in aantallen per km²). Een afweging zal moeten worden gemaakt welk type van verwerking voor RWS het beste resultaat geeft.

De GPS gegevens van de getagde Kleine Mantelmeeuwen zijn ook in een tabel in de database gezet, en via WxS ontsloten. Hierbij is nog geen Aquo-standaard toegepast.

Verwachte aanpassing in volgende release

De reden dat andere data nog niet verwerkt zijn, is de late aanlevering van de data op de repository. Er is nu veel ervaring opgedaan met opwerking van soortgelijke data, en de verwachting is dat het verwerken van de rest van de data niet veel moeilijkheden gaat opleveren.

3.4 Noordzeewind; Offshore Windpark Egmond aan Zee (OWEZ)

Het Offshore Windpark Egmond aan Zee (OWEZ) is het eerste windpark dat in Nederland op zee werd gebouwd. De windturbines op zee staan bloot aan zware omstandigheden: zout, zeewater, golven, stromingen en harde wind. Om in de toekomst windparken op zee efficiënter en goedkoper te kunnen maken is meer kennis nodig over de invloed van die omstandigheden op windturbines. Ook over eventuele effecten op de natuur en het milieu zijn voor de Nederlandse situatie nog onvoldoende gegevens beschikbaar.

Aan het OWEZ is een uitgebreid Monitoring- en Evaluatieprogramma (MEP-NSW) gekoppeld dat wordt uitgevoerd in samenwerking met diverse onderzoeksinstituten. Het MEP-NSW Ecologie moet de leemten in kennis en ervaring opvullen op het gebied van natuur, milieu en gebruiksfuncties. Het totale onderzoeksprogramma is gestart tijdens de bouw van het windmolenpark in 2006 en loopt tot 2012.

Borging van de data in een online repository

In opdracht van Rijkswaterstaat heeft Deltares eind 2012, begin 2013, dus aan het eind van het MEP-NSW project, de verzameling van alle biologische meetdata in een online repository gecoördineerd (intern project nr. 1207392, Deltares; ca. 8 k€). Dit proces is begeleid door middel van templates voor metadata, richtlijnen voor data formats, en een ontmoeting met de dataleveranciers in de vorm van een workshop (IJmuiden). Hierna is nog regelmatig contact onderhouden met dataleveranciers om de data ook daadwerkelijk binnen te krijgen.

Voor zover nu is vast te stellen zijn alle meetdata vastgelegd op een online repository bij Deltares en daardoor in ruwe vorm beschikbaar.

Bij opwerking in het huidige project zijn wat onvolkomenheden aangetroffen in de data. Deze betreffen o.a. niet-passende CTD data (ander jaar dan de metingen), en het ontbreken van details in de metadata. Op dit moment is er contact met de dataleveranciers om deze onvolkomenheden te herstellen.

Opwerking van de data tot Aquo conforme data

De monitoringsdata verzameld in het kader van MEP-NSW zijn zeer divers, en conformeren niet volledig aan een bekende (inter)nationale semantische datastandaard als Aquo, CF, of BODC.

Tevens verschillen de dataformats voor identieke data binnen de dataset van elkaar wat betreft veldnamen, aantal velden, parameters etc. Dit maakt alleen al het beschrijven van deze datasets gecompliceerd, en stelt hoge eisen aan de opwerking naar een gestructureerde en gestandaardiseerde omgeving.

De verwerking van de data in de OpenEarth omgeving is goed gevorderd (Tabel 3.4). Voor verschillende datatypen zijn WxS services beschikbaar die voldoen aan het Aquo-uitwisselmodel, en zijn opgenomen in de catalogus (GeoNetwork).

Tabel 3.4 Opwerkingspercentage van data uit het MEP-NSW project

	% in Repository	% in Database	% OGC WFS/WMS	% Aquo conform	% catalogue	Tim Berners-Lee *
Demersale vis	100	100	100	99	Ja	****
Pelagische vis	100	40	40	40	Ja	****
Gillnet	100	100	100	100	Ja	****
DIDSON	100	100	100	100	Ja	****
Vogels	100	100	0	0	Ja	***
Benthic communities	100	100	100	100	Ja	****
Bivalves	100	60	60	60	Ja	****
Macrobenthos	100	75	50	50	Ja	****
Recapture sole and cod	100	50	0	0	Ja	***
Flight patterns	100	0	0	0	Ja	*
Harbour seals	100	0	0	0	Ja	*
Harbour porpoise acoustics	100	0	0	0	Ja	*

Verwachte aanpassing in volgende release

De te verwachten inspanning voor het verwerken van de resterende datasets is lastig in te schatten, omdat alle datasets zo verschillend zijn. Ook is er bij de prioritering tot nu toe gekozen voor de op het eerste gezicht “eenvoudig” te verwerken data, en zal een volgende stap onzekerder worden.

3.5 MEP-duinen

Nieuw duingebied

Als compensatie van natuurwaarden die door aanleg van de Tweede Maasvlakte verloren zijn gegaan is bij Delfland een nieuw duingebied (Spanjaards Duin) aangelegd. De duincompensatie bestaat uit 35 hectare extra duingebied, tussen de Van Dixhoordriehoek bij Hoek van Holland en de Banken (Westland). Voor de aanleg is circa 5,4 miljoen kubieke meter zand op het strand opgespoten voor een breder strand en de natuurlijke vorming van nieuw duin. Een deel van het zand is voor het strand neergelegd als nieuwe

onderwateroever. Bij de keuze van het zand is rekening gehouden met de stuifgevoeligheid (korrelgrootte, korrelgrootteverdeling en het gehalte aan slib en grove delen).

De duincompensatie is gestart in 2008 en in 2009 opgeleverd. Nadien zal er periodiek zand worden gesuppleerd (eens per 5 jaar, gedurende circa 20 jaar). Door monitoring wordt bekeken of de duinvorming op de gewenste manier plaatsvindt. De verwachting is dat na 10 tot 20 jaar de gewenste natuurdoeltypen 'vochtige duinvallei' en 'grijze duinen' zullen zijn bereikt. In het Spanjaards Duin wordt de ontwikkeling van het stuiven, de grondwaterstand, de grondwaterkwaliteit, en de vegetatie-ontwikkeling vastgelegd. Het meetprogramma van het Spanjaards Duin is nog in volle gang.

Bestaand duingebied

In het aangrenzend duingebied worden de emissies en deposities van stikstofoxiden als nulmeting gemonitord. Doel van het meetprogramma in de bestaande duinen tussen Solleveld in het noorden en Goeree in het Deltagebied is vast te stellen wat de huidige ontwikkelingen zijn en in welke mate het toekomstig gebruik van Maasvlakte-2 de bestaande duingebieden aantast.



Figuur 3.2 Voorbeeld van vegetatielagen en opname punten MEP Bestaande Duinen

Het meetprogramma van de bestaande duinen (Solleveld, Kappitelduinen, Voorne, Goeree) is afgelopen. In Figuur 3.2 zijn van deze locaties alle overgedragen ruimtelijke bestanden

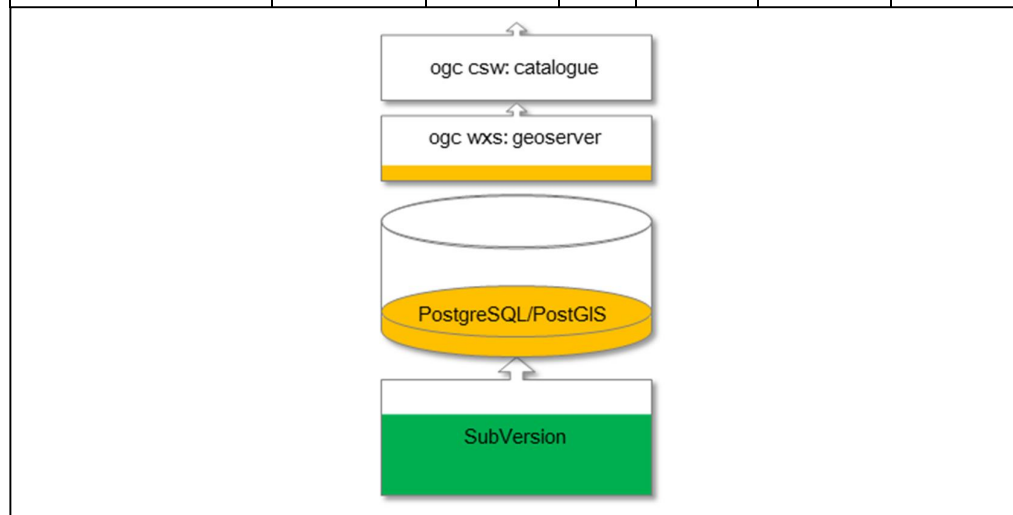
opgenomen. Over één dataset bestaat zodanige discussie dat deze niet beschikbaar wordt gesteld. Het betreft hier de depositie metingen waarvan wel de locaties zijn vrij gegeven. De volgende paragraaf geeft inzicht in de stand van zaken met betrekking tot dataopslag en vulling van de database.

Borging van de data

Tabel 3.5 geeft een schatting van de opwerkingspercentages van de gegevens van beide MEP duinen projecten.

Tabel 3.5 Opwerkingspercentage van data uit het MEP-duinen project

	% opgeslagen in Repository	% in Database	% OGC	% Aquo conform	catalogus	Tim Berners-Lee *
MEP Nieuwe Duinen	36	20	20	0	Nee	****
MEP Bestaande Duinen	95	0	0	0	Nee	*



Wat betreft het nog lopende Nieuwe duinen project, wordt de repository nog aangevuld. De laatste aanvulling dateert van 14 november 2014.

Om tot een volledige 100% vulling te komen dienen de datasets verder opgewerkt te worden. De laatste data van de Bestaande duinen zijn op 2-12-2014 opgeleverd door de data verzamelende partij. In eerste instantie werd aangegeven dat bij oplevering van de rapportage (19 september 2014) de gegevens toegeleverd zouden worden. Dat is echter om onbekende redenen niet gehaald. Waarmee het ook onmogelijk is gebleken deze gegevens te verwerken.

Opwerking van de data tot Aquo conforme data

Opwerking naar 100% Aquo conform is overigens vooralsnog niet mogelijk voor MEP Bestaande duinen omdat er discussie is over de uitkomsten van het stikstof depositie onderzoek. Tot nader order worden deze gegevens niet beschikbaar gesteld. Wat betreft uitvoering van een eventueel data herstelplan zal gewacht moeten worden totdat deze gegevens worden aangeleverd.

Verwachte aanpassing in volgende release

In de huidige opstelling van de OpenEarth stack is nog geen ervaring opgedaan met vlakdekkende gegevens, met als duidelijk voorbeeld de vegetatiegegevens. De vegetatie

gegevens zijn onder gebracht in zogenaamde habitatype kaarten, die in de vorm van polygonen (vlakken) beschikbaar zijn. Daarnaast zijn opnamen op de zogenaamde PQ's aangeleverd in TURBOVEG formaat, wat niet zonder meer te verwerken is naar een van de OGC formaten.

Buiten deze technische uitdagingen is het zo dat vegetatie associaties en/of habitatypen niet voorkomen in de Aquo-systematiek. Dat laatste is geconstateerd aan de hand van een eerste snelle scan van de gegevens die in december 2014 zijn verkregen.

Conclusie aan de hand van deze gegevens is dat er nogal wat werk gedaan moet worden om deze gegevens Aquo conform beschikbaar te maken.

3.6 Overige datasets

De uitvoering van het huidige project heeft veel belangstelling gewekt bij andere mariene projecten, waarvoor doorgaans geen datamanagement plan was opgesteld. Gaande het project is gebleken dat er een groot aantal 'overige' datasets beschikbaar is. Mede naar aanleiding van diverse demonstraties van de OpenEarth stack, zijn meerdere (mariene) datasets met opgewerkte data in diverse systemen (o.a. QuantumGIS, ArcGIS, python scripts) en een scala aan 'webviewers' (Adaguc, 'OET'-viewer, rws-mapviewer2i) geïdentificeerd of door RWS aangemeld. Het opnemen van deze datasets (of, als proef, een deel ervan) in de repository kost relatief weinig inspanning en voorkomt dat de data verloren gaat. Daarom zijn zij in Tabel 3.6 als "Overige datasets" weergegeven. Echter, in het huidige project is voor de meeste van deze datasets een pas op de plaats gemaakt. Wanneer evenwel 'eigen' financiering beschikbaar was, kon verdere opwerking plaats vinden volgens de in het huidige project toegepaste methoden. Een voorbeeld hiervan is Kustlijnzorg B&O ecologie. De gegevens van 4 jaar meten voor de kusten van Vlieland en Ameland zijn tot aan driesterren opgewerkt, maar zijn wegens restricties niet openbaar beschikbaar.

Tabel 3.6 Opwerkingspercentage van data uit overige datasets.

	% in Repository	% in Database	% OGC	% Aquo conform	In catalogus	Tim Berners-Lee sterrenstatus
Benthos/Vis						
RijMaMo	1	1	1	1	Ja	****
PoC Lodingen	10	0	0	0	Nee	*
EcoLIMS-benthos	100	100	100	0	Nee	****
BWN	100	100	0	0	Nee	***
KPP B&O Kustlijnzorg	93	0	0	0	Nee	***
MCI Old data	90	0	0	0	Nee	*

3.6.1 EcoLIMS

De EcoLIMS-Benthos data is een verzameling MWTL Benthos waarnemingen van 1992 t/m 2013 van de Noordzee en Zeeuwse Delta. Deze data zou binnen het project Digitale Delta door IBM opgewerkt en geïmporteerd worden in de EcoLIMS database van RWS. Uiteindelijk is deze vraag bij Deltares terecht gekomen en aldaar onder de paraplu van Datamanagement mariene projecten getrokken. De aangeleverde data (MsExcel, MsAccess) is in een OpenEarth stack opgewerkt, Aquo conform gemaakt en in het ontwikkelde uitwisselingsformaat gezet. Van daaruit is een mapping naar het EcoLIMS-XML-uitwisselingsformaat gemaakt. In dit proces bleek weer hoe arbeidsintensief het opwerken van data steeds weer is:

- de geleverde data was niet afdoende genormaliseerd om direct in de database van de OpenEarth stack ingelezen te kunnen worden. Het identificeren van de (unieke) meetlocaties uit de set van waarnemingen vroeg wat iteraties met de data eigenaar RWS; zo doken door de tijd heen meetlocatienamen op van (soms significant) verschillende geografische locaties.
- vooral mapping van de taxonbenamingen vroeg extra aandacht; (sub)soorten die niet in de officiële lijsten voorkomen, spellingsvarianten in de (Latijnse) naamgeving waardoor er geen match gemaakt kon worden.
- Aquo website niet altijd even gebruiksvriendelijk. Ook bleken er pointers op te staan naar verschillende taxa lijsten; een eigen Aquo-domeintabel en een taxa-lijst op de website van Ecosys die niet overeenstemden. Informatiehuis Water is tijdens de IHW netwerkdag gevraagd iets te doen aan deze issues.

Een eerste inschatting voor de opwerking van deze niet eens zo complexe dataset was 5 mensdagen. Uiteindelijk heeft het zo'n 2 maal die inspanning gekost.

3.6.2 RijMaMo

Via RWS is de vraag gesteld na te denken over datamanagement ten behoeve van online kalibratie van modellen voor het Rijn Maas Monding gebied. Als test dataset is er voor één locatie een jaar aan zoutgegevens aangeleverd. Handmatig zijn de gegevens in het datamodel van Mariene Projecten geladen, via Geoserver beschikbaar gemaakt en binnen GeoNetwork vindbaar gemaakt.

Vooralsnog is er geen verdere actie uit voortgekomen.

3.6.3 PoC Lodingen

Proof of Concept OpenEarth opslag en ontsluiting lodingen van het rivierengebied. Het PoC verzoek is ontstaan als gevolg van een demonstratie van OpenEarth opslag en visualisatie. De kustviewer aanpak zal, indien een project wordt gehonoreerd, voor een groot deel als template dienen. Er zijn verschillende extra vragen die nog niet zijn geïmplementeerd in de ¹kustviewer die wel voor de PoC van belang zijn. Hieronder puntsgewijs deze eisen, gecombineerd met een indicatie of deze al in de kustviewer zijn doorgevoerd.

Pakket van wensen:

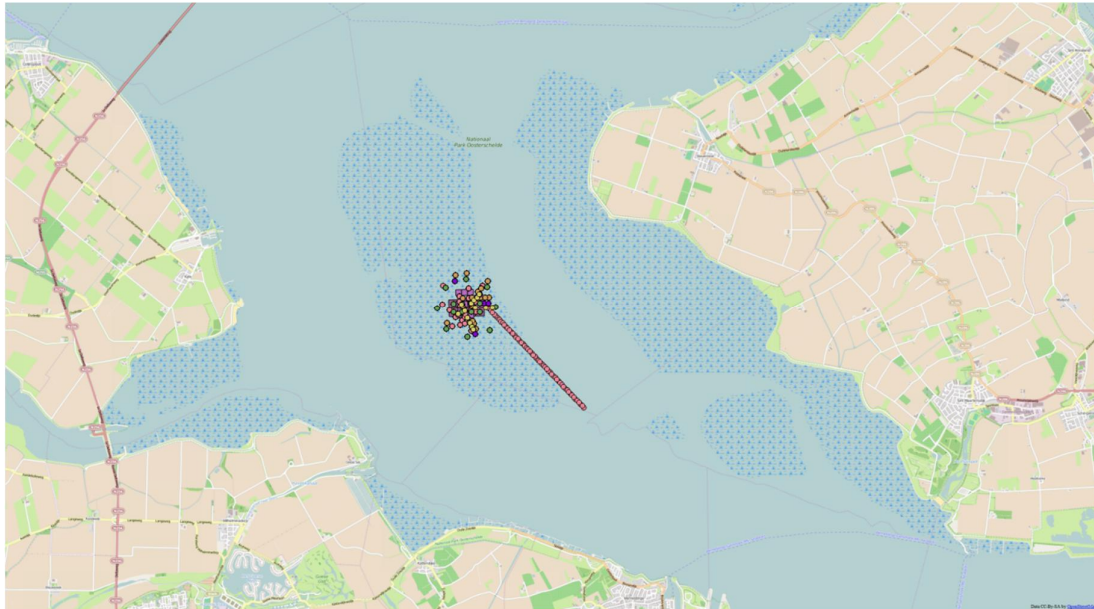
- 1 Geüniformeerd beschikbaar (in kustviewer via netCDF beschikbaar)
- 2 Metadata beschikbaar (in netCDFs voor kustviewer zijn CF² compliant metadata beschikbaar)
- 3 Metadata doorzoekbaar, toevoeging die voor PoC gerealiseerd moet worden
- 4 Gegevens beschikbaar in de kustviewer dient gerealiseerd te worden.
- 5 Gegevens voor ArcGIS bewerkingen beschikbaar door middel van een wcs, web coverage service, niet binnen kustviewer beschikbaar.

3.6.4 BWN

Building With Nature dataset is een dataset die deels (benthos deel) in het PMR-NCV datamodel is samengebracht in het kader van het Building With Nature programma. Status is dat deze data in de database beschikbaar is. Verdere opwerking heeft niet plaats gevonden. In Figuur 3.3 is een aantal onderdelen van de dataset weergegeven.

¹ Kustviewer is een lizard applicatie met de Google Earth plugin waarin gegevens uit diverse bronnen (lees OPeNDAP servers) via kml (Keyhole Markup Language) beschikbaar worden gemaakt.

² CF = Climate Forecasting Meta Data Convention, zie <http://cfconventions.org/> voor meer informatie



Figuur 3.3 Overzicht van locaties waar BWN data beschikbaar zijn

Naast benthos data bevat deze dataset ook sediment data op de hierboven aangegeven locaties.

3.6.5 KPP B&O Kustlijnzorg

In 2009 is naar aanleiding van een samenwerkingsovereenkomst tussen vier natuurbeschermingsorganisaties (NGO's) en het ministerie van Verkeer en Waterstaat, een middellange termijn onderzoeks- en monitoringprogramma opgesteld: "Ecologisch gericht suppleren, nu en in de toekomst". Doelstelling is meer inzicht te krijgen of, en in welke mate, zandsuppleties van invloed zijn op natuurwaarden en op welke wijze zandsuppleties in de nabije toekomst kunnen bijdragen aan de opgaven van veiligheid samen met natuurbehoud en -ontwikkeling. Het middellange termijn plan onderscheidt vier onderzoekscategorieën, gebaseerd op grootschalige vraagstukken die betrekking hebben op:

- het ecosysteem van de Nederlandse kust (of een deel daarvan zoals de Waddenzee);
- de ondiepe kustzone met een focus op de zone waarin gesuppleerd wordt
- het (droge) strand en de duinen
- het voedselweb in de ondiepe kustzone

In het convenant met de NGO's is afgesproken dat 1 % van het budget voorzandsuppleties, met een maximum van 400 k€ zal worden besteed aan onderzoek naar de ecologische effecten.

Dit project is onderdeel van een groter project waarbij OpenEarth datamanagement is toegepast voor de morfologische gegevens. De morfologische gegevens zijn ondergebracht in de kustviewer maar nog niet vindbaar gemaakt.

Figuur 3.4 toont de locaties van de bemonstering van benthische organismen. De data van de sedimentsamenstelling wordt nog verwerkt.



Figuur 3.4 Overzicht van locaties waar gegevens van B&O Kust en Ecologie in Aquo-format beschikbaar zijn

3.6.6 MCI Old data

MCI Old data is een collectie data die op CD/DVD beschikbaar zijn. Het enige wat met de gegevens is gedaan is dat deze op de repository zijn geplaatst. Het is niet mogelijk een indicatie te geven hoeveel tijd ermee gemoeid is om deze data online beschikbaar te maken. Daartoe zal eerst goed bekeken moeten worden wat het eigenlijk allemaal voorstelt. Voor nu is de data beschikbaar via de repository.

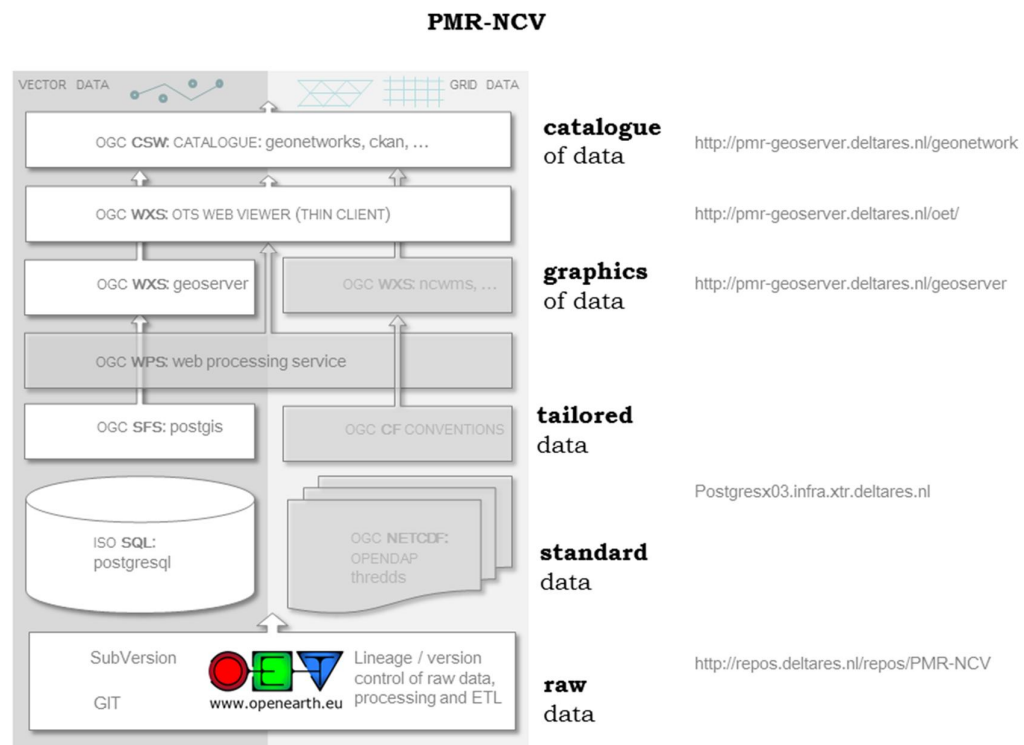
4 Implementatie op vier locaties

4.1 Deltares

Bij Deltares zijn inmiddels diverse OpenEarth stacks opgezet, niet alleen voor RWS projecten maar ook voor buitenlandse projecten. Voor het huidige project zijn twee stacks van toepassing, de stack ten behoeve van PMR-NCV en de stack waarin de overige mariene projectdata zijn opgeslagen.

4.1.1 Technische implementatie stack PMR-NCV

In Figuur 4.1 is schematisch weergegeven uit welke componenten de PMR-NCV stack bestaat en welke onderdelen zijn opgenomen (lichte vlakken) en welke onderdelen niet zijn geïmplementeerd, de grijze vlakken.



Figuur 4.1 PMR-NCV OpenEarth stack zoals geïmplementeerd bij Deltares

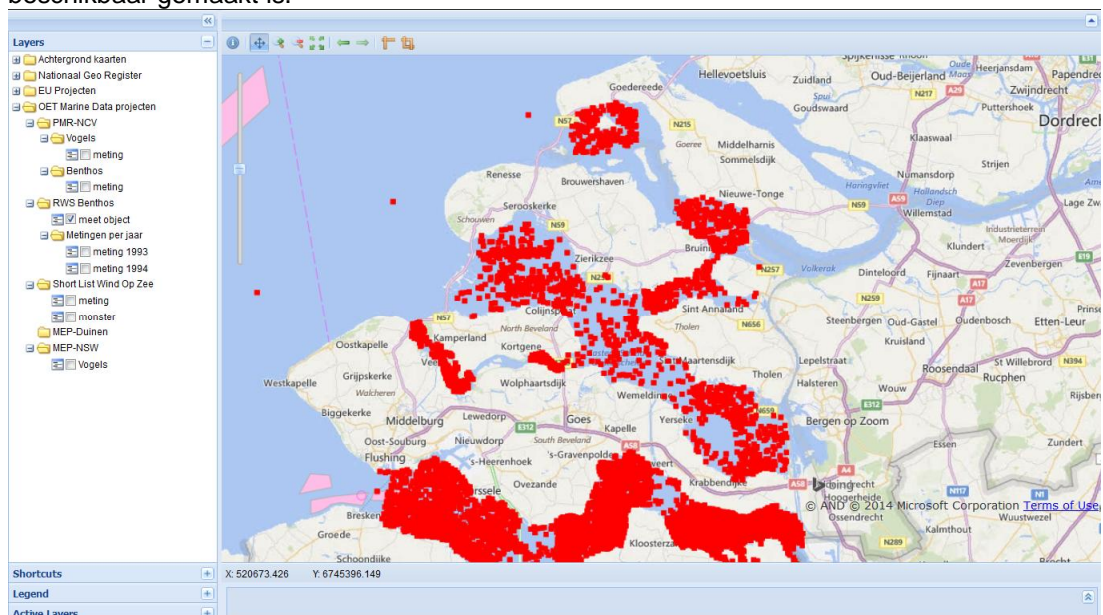
Rechts van iedere component is, indien van toepassing, het adres opgenomen waar de service beschikbaar is. In de paragraaf autorisatie wordt verder ingegaan in hoeverre de gegevens daadwerkelijk benaderbaar zijn en door wie.

De insteek in het PMR-NCV project is geweest dat alleen vector gegevens (puntdata) beschikbaar worden gemaakt. OPeNDAP /THREDDs en OGC-CF conventions alsmede ncWMS, die vooral geschikt zijn voor rasterdata, zijn in de PMR-NCV OpenEarth stack niet als dataopslag en bronnen beschikbaar.

Vector gegevens, en in dit geval alleen puntlocaties en de daarbij behorende observaties zijn via Aquo conforme services beschikbaar gemaakt via Geoserver. De gegevens zijn tevens doorzoekbaar gemaakt via GeoNetwork.

Aangezien de data niet via de DL ontsloten kunnen worden, is voor demonstratie doeleinden een basis viewer opgezet. Hiermee kunnen de lagen die beschikbaar worden gemaakt via Geoserver, worden weergegeven op een bepaalde standaard ondergrond.

Figuur 4.2 presenteert een screenshot van een van de datalagen die op deze manier beschikbaar gemaakt is.



Figuur 4.2 Voorbeeld van de een open source viewer waarmee gegevens via Geoserver kunnen worden getoond

Het betreft hier echter alleen het zichtbaar maken van locaties waar een bepaalde observatie is gedaan. Het is zonder uitbreiding van functionaliteit niet mogelijk om hiermee bijvoorbeeld een selectie uit te voeren en dan een tijdreeks of welke informatie dan ook op te halen. Parallel aan dit project is Deltares betrokken bij de zoektocht naar een geschikte viewer voor RWS. Er is zoveel beschikbaar op dit gebied, dat richtlijnen voor de keuze van een geschikte viewer meer dan welkom zijn.

4.1.2 Organisatorische invulling PMR-NCV

De OpenEarth stack is in termen van de XaaS (diverse mogelijkheden as a Service) volledig bij Deltares geplaatst. Met andere woorden, de database (SaaS) is op een server (HaaS) met een CentOS-operating systeem (PaaS) van Deltares geplaatst. Via Geoserver worden de data als een OGC service beschikbaar gemaakt; dit is daarmee een zogenaamde DaaS.

4.1.3 Autorisatie en online beschikbaarheid data tbv DL

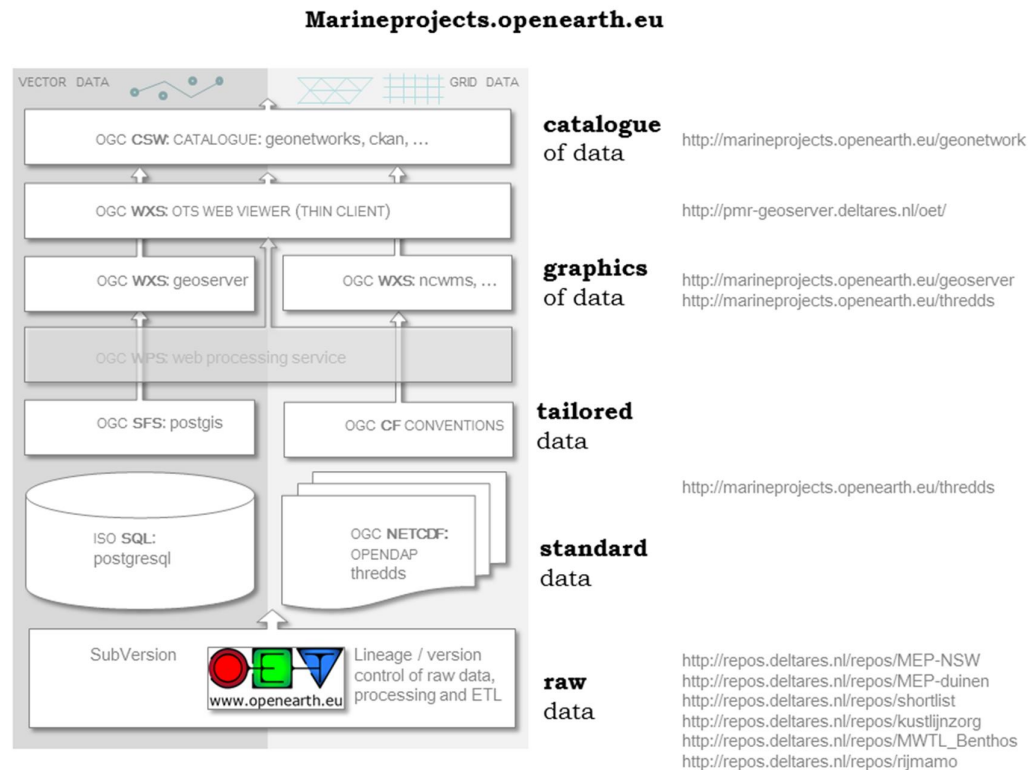
Voor het project PMR-NCV is ervoor gekozen dat de partners onderling de data zouden opwerken. Hiertoe zijn de volgende zaken ingericht met de daarbij behorende rechtenstructuur:

Stack onderdeel	rechten
Repository	Alle partners hebben lees, schrijf en update rechten
Database (PostgreSQL/PostGIS)	Alle partners hebben lees, schrijf en update rechten
Geoserver	Iedereen heeft lees rechten (layer preview). Deltares heeft volledige rechten.
GeoNetwork	Iedereen heeft lees rechten. Deltares heeft volledige rechten.

GeoNetwork en Geoserver vormen voor de distributielaag de ingangen om data beschikbaar te maken.

4.1.4 Technische implementatie tweede stack tbv overige mariene projecten

Figuur 4.3 geeft schematisch aan uit welke componenten de stack voor overige mariene projecten bestaat en welke onderdelen al dan niet gerealiseerd zijn (uit grijze delen zijn niet gerealiseerd).



Figuur 4.3 Mariene Projecten OpenEarth stack zoals geïmplementeerd bij Deltares

De figuur laat een groot aantal gerealiseerde onderdelen zien. Dit betekent niet dat ze allemaal gevuld zijn. Dat geldt bijvoorbeeld voor het rechter spoor voor rasterdata en bijbehorende software (OPeNDAP, THREDDs, ncWMS). Deze is beschikbaar in het kader van de verwachte aanvraag met betrekking tot het Rijn Maas Monding (RijMaMo-)project en de Proof Of Concept (PoC-) Lodingen. Naar verwachting zal volgend jaar een start worden gemaakt met de invulling van dit spoor om rastergegevens beschikbaar en vindbaar te maken.

4.1.5 Organisatorische invulling Mariene Projecten

De OET-stack is in termen van de XaaS (diverse mogelijkheden as a Service) volledig bij Deltares geplaatst. In andere termen, de database (SaaS) is op een server (HaaS) met een operating systeem (PaaS) van Deltares geplaatst. Via een bij Deltares geplaatste Geoserver wordt data als een OGC service beschikbaar gemaakt, de zogenaamde DaaS.

4.1.6 Autorisatie en online beschikbaarheid Mariene Projecten

In onderstaande tabel is opgenomen hoe de organisatorische invulling van de stack is geregeld.

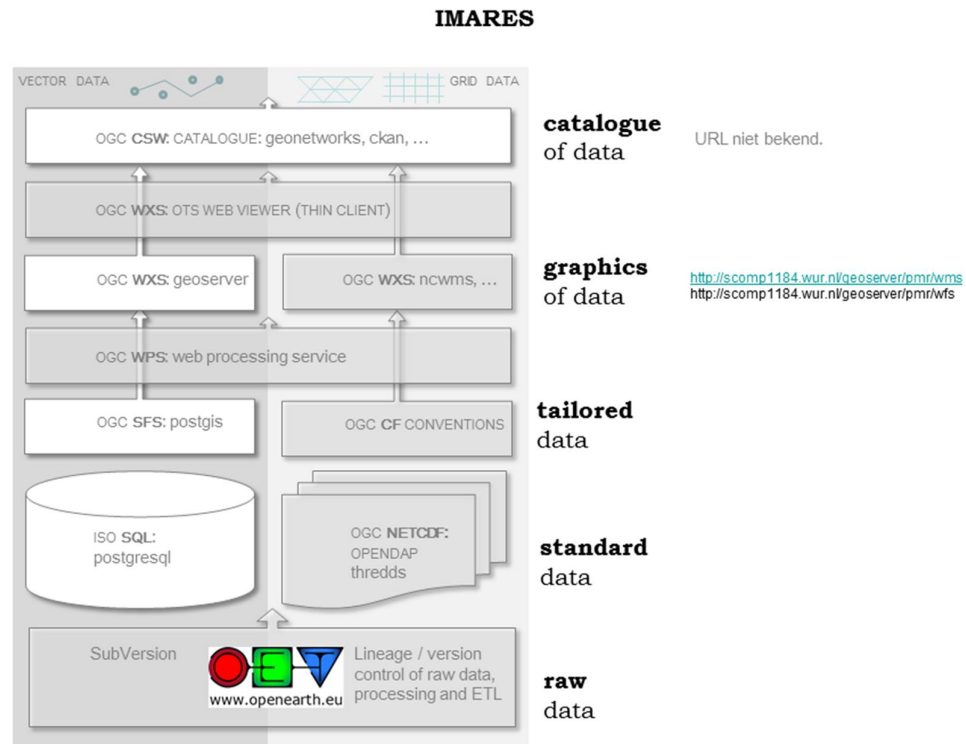
Stack onderdeel	rechten
Repository	Alle partners hebben lees, schrijf en update rechten
Database (PostgreSQL/PostGIS)	Alleen Deltares heeft alle rechten
Geoserver	Iedereen heeft lees rechten (layer preview). Deltares heeft volledige rechten.
GeoNetwork	Iedereen heeft lees rechten. Deltares heeft volledige rechten.

GeoNetwork en Geoserver vormen voor de distributielaag de ingangen om data beschikbaar te maken..

4.2 IMARES

4.2.1 Technische implementatie stack

Bij IMARES is op eigen hardware (WUR) een installatie uitgevoerd van de OpenEarth stack met de componenten PostgreSQL/PostGIS, Geoserver en GeoNetwork. Er is geen versiebeheersysteem (SubVersion) geïnstalleerd. IMARES heeft al een workflow voor data opwerking en het accent zou in dit geval dan ook liggen op het ontsluiten van die data via services uit de OpenEarth stack.



Figuur 4.4 OpenEarth stack zoals geïmplementeerd bij IMARES

In Figuur 4.4 staat de OpenEarth stack met alle componenten waaruit die conceptueel is opgebouwd. De selectie van bij IMARES geïnstalleerde componenten is in wit aangegeven.

4.2.2 Organisatorische invulling

De OpenEarth stack zou het IMARES mogelijk maken de mariene data die zij reeds in eigen (gesloten) databases opgeslagen hebben, te ontsluiten via de OpenEarth workflow. Hiertoe zouden zij een selectie van hun data in de geïnstalleerde PostgreSQL/PostGIS database opladen in het door het IHM voorgestelde uitwisselingsformaat (meetobject, monster, meting) en die vervolgens uitserveren via de geïnstalleerde Geoserver. Er is een benthos dataset in de PostgreSQL/PostGIS database geladen en als kaartlaag geconfigureerd in de Geoserver. Het betreft data uit de IMARES WOT schelpdierdatabase voor de volgende soorten en jaren:

- Ensis, Spisula, Venusschelp en Otterschelp
- 2002 t/m 2011 (10 jaar)
- Gebied: Nederlandse kustzone

4.2.3 Autorisatie en online beschikbaarheid data tbv DL

De PostgreSQL/PostGIS database is (uiteraard) niet direct van buiten toegankelijk. De Geoserver serveert de data in 7 lagen in WMS en WFS, maar niet conform het door IHM voorgestelde uitwisselingsformaat. De data is open beschikbaar via de URL die is aangegeven in Figuur 4.4. De status van deze lagen en de map service is onduidelijk daar IMARES gedurende het traject heeft aangegeven een ander spoor (SeaDataNet) te willen gaan volgen.

4.3 TU Delft / 3TU.Datcentrum

4.3.1 Technische implementatie stack

De zandmotor data is opgeslagen in een OpenEarth DataLab dat wordt gehost door 3TU.Datcentrum. Het datalab, met daarin zowel ruwe als opgewerkte data, is beschikbaar op <https://zandmotordata.nl>. Momenteel is de data alleen toegankelijk voor de gemeenschap die direct betrokken is bij de monitoring en/of het onderzoek rond de zandmotor. In een later stadium, vanaf 2016, zal de (opgewerkte) data geleidelijk open beschikbaar komen.

Het OpenEarth DataLab ondersteunt alle stappen die onderscheiden worden in de OpenEarth workflow:

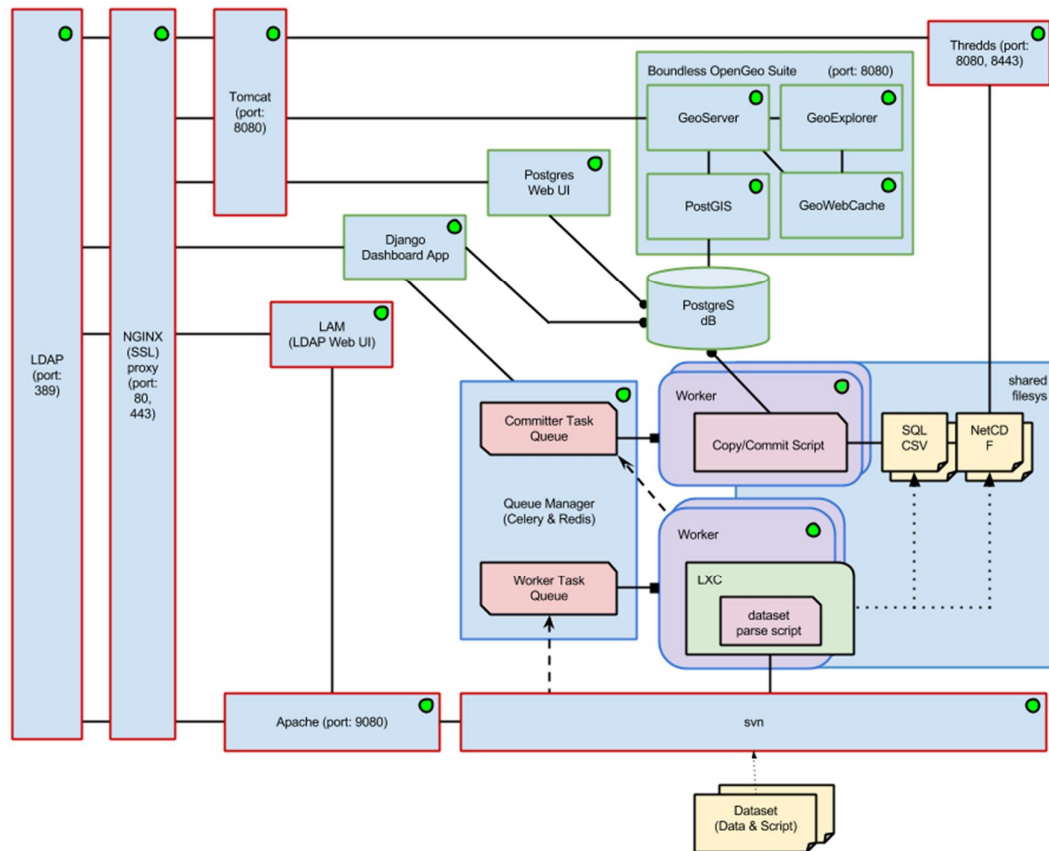
- 1 Opslag ruwe data in een subversion repository
- 2 Processing environments om opwerking scripts te draaien in de cloud (Python en Matlab worden ondersteund)
- 3 THREDDS/OPeNDAP server voor grid data in netCDF formaat
- 4 PostgreSQL/PostGIS server voor vector data
- 5 GeoNetwork voor catalogus functionaliteiten

In april 2014 is het zandmotor datalab in gebruik genomen. Bij in gebruikname waren de functionaliteiten voor netCDF (zowel processing environments als THREDDS/OPeNDAP server) volledig operationeel. In de periode daarna zijn de processing environments voor de PostgreSQL data base ontwikkeld.

Het systeem is zo opgezet dat de opwerking van de data volledig transparant en reproduceerbaar is. Dit opwerkingsproces vindt plaats in processing environments (workers genoemd in Figuur 4.5), door het runnen scripts. Vanuit deze environments worden de database records in de database geladen en de netcdf bestanden op de THREDDS server gezet. Doordat de processing environments uitsluitend gebruik maken van data en scripts uit de repository, is altijd te herleiden hoe de database records en netcdf bestanden tot stand gekomen zijn.

Bij 3TU Datacentrum is de volgende opzet gerealiseerd (zie Figuur 4.5). De Django dashboard app (midden links) is de webpagina waar alles bij elkaar komt. Hier staan de links naar de verschillende componenten en kan de gebruiker een data verwerking taak starten en monitoren.

Het OpenEarth DataLab zoals in gebruik voor de Zandmotor is gebaseerd op een CentOS 6.5 Linux distributie (PaaS). Het uitrollen van het systeem is geautomatiseerd. Hierdoor kan het DataLab relatief eenvoudig uitgerold worden, evenals beschikbare (veiligheids-) updates. Voor de ontwikkeling wordt een zogenaamde OTAP (ontwikkel, test, acceptatie, productie) procedure gevolgd, om de kwaliteit te waarborgen en daarmee het risico om het productieproces te hinderen, minimaal te houden. De overstap naar een nieuwere versie, of ander type, besturingssysteem vereist echter herziening van de automatische uitrol routine en uitgebreid testen.



Figuur 4.5 OpenEarth stack zoals geïmplementeerd bij Zandmotor DataLab

Figuur 4.5 geeft de componenten weer, inclusief de interacties, zoals deze voor het Zandmotor DataLab zijn geïmplementeerd. Links in de figuur staan de authenticatie (LDAP) en proxy (nginx) componenten. Rechts onder staat de versiebeheer component (svn) waar ruwe data en scripts binnenkomen. Daarboven staat een cluster van componenten die aan de workers (ook wel processing environments genoemd) gerelateerd zijn. Deze workers vormen de schakel om de ruwe data vanuit svn in de PostgreSQL database of THREDDSS server (netCDF) te zetten, na opwerking. De Boundless OpenGeo Suite bevat verschillende componenten om de data uit de PostgreSQL database aan te bieden. De componenten voor datagebruik bevinden zich rechtsboven in de figuur (Boundless OpenGeo Suite en THREDDSS server). De componenten voor ruwe data en verwerking bevinden zich rechtsonder in de figuur.

4.3.2 Organisatorische invulling

Het Zandmotor DataLab is gefinancierd met EFRO (Kansen voor West subsidie). Deze subsidie is door Rijkswaterstaat WVL verworven voor de zandmotor monitoring en waarborging van de data. Het datamanagement is uitbesteed aan Ecoshape, die de implementatie en hosting van het datalab door 3TU.Datacentrum hebben laten doen. Het 3TU.Datacentrum heeft de door Deltares samengestelde OpenEarth stack geïmplementeerd als datalab. Het 3TU.Datacentrum heeft de gebruikersondersteuning (zandmotor data helpdesk) uitbesteed aan de waterbouwkunde afdeling van TU Delft faculteit Civiele Techniek, waarvoor student assistenten ingezet worden.

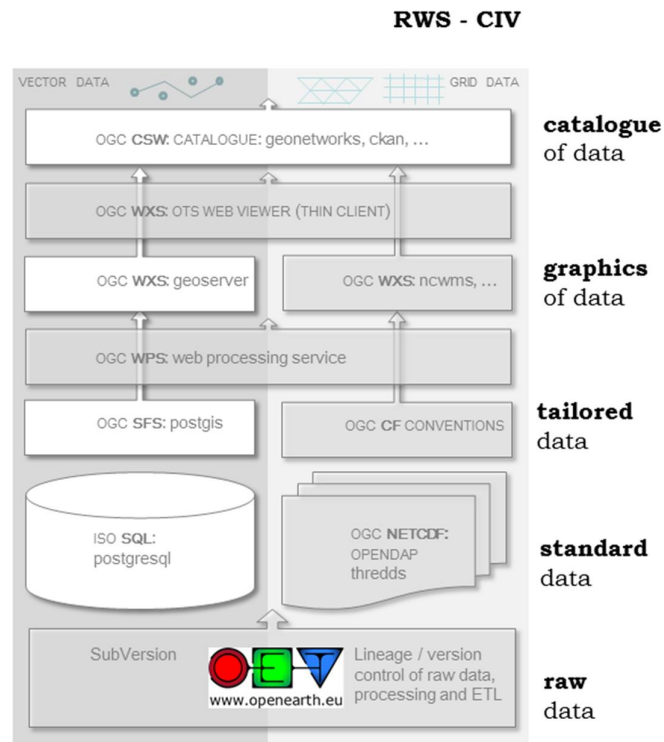
4.3.3 Autorisatie en online beschikbaarheid data tbv DL

Alle componenten van het Zandmotor DataLab zijn momenteel alleen toegankelijk voor geautoriseerde gebruikers. In een later stadium, vanaf 2016, zal de opgewerkte data publiek beschikbaar komen.

4.4 RWS-CIV

Voorafgaand aan de implementatie van de OpenEarth stack bij CIV is een PSA-traject doorlopen. In de Project Start-up Architecture worden de gevolgen van de inpassing van nieuw systemen in de RWS-IV infrastructuur besproken. De PSA gaf aan dat de OpenEarth stack met enkele aanpassingen in de RWS infrastructuur kan worden gerealiseerd (Dirksen, Flip en Ingeborg van Splunder, 2014. PSA Mariene Data v0.91).

4.4.1 Technische implementatie stack



Figuur 4.6 OpenEarth stack zoals geïmplementeerd bij RWS-CIV

De installatie bij RWS-CIV is identiek aan die van IMARES. Opgemerkt zij dat de installatie op een geïsoleerde server is uitgevoerd en dat het in een zekere quarantaine is geplaatst, daar geen van de software componenten in de RWS bouwstenen catalogus voorkomt. De installatie heeft een demo karakter waarmee RWS de OpenEarth stack van dichtbij kan bekijken.

In Figuur 4.6 staat de OpenEarth stack met alle componenten waaruit die conceptueel is opgebouwd. De selectie van geïnstalleerde componenten zijn in wit aangegeven.

4.4.2 Organisatorische invulling

De data voor deze OpenEarth stack installatie is door en bij Deltares op een schaduw stack opgewerkt. Een dump van die database zal bij RWS-CIV geïmporteerd worden (moet nog gebeuren, december 2014). Het betreft de MEP Nieuwe Duinen data.

4.4.3 Autorisatie en online beschikbaarheid data tbv DL

Zoals gezegd staat de stack bij RWS-CIV geïsoleerd en is die alleen toegankelijk via het interne RWS netwerk. Er is voor een selectie van Deltares medewerkers een account en telewerk token aangevraagd bij RWS zodat zij toegang kunnen krijgen tot het RWS interne netwerk en kunnen assisteren (op afstand) bij het configureren van de Geoserver.

Er is een vraag uitgezet bij RWS-CIV wat er nodig is om de stack ook voor de buitenwereld beschikbaar te maken, maar daar is tot op heden geen uitsluitsel over.

5 Afwegingskader beheerscenario's

In het voorgaande hoofdstuk is beschreven hoe de OpenEarth software stack op verschillende locaties draait. Het technisch functioneren van de OpenEarth stack is natuurlijk een eerste vereiste, maar voor het duurzaam in stand houden van het functioneren van de stack is meer nodig. Het gaat dan om de vraag hoe het beheer ervan georganiseerd kan worden. Dit hoofdstuk gaat in op de verschillende aspecten die bij de keuze van het beheer een rol spelen en biedt daarmee voor Rijkswaterstaat een afwegingskader om over het lange termijn beheer van data een gefundeerd besluit te nemen. Per paragraaf worden de punten van overweging samengevat.

5.1 Enkelvoudige versus meervoudige bronhouder

Het belangrijkste aspect bij het lange termijn beheer van data is de juridische eigendomsstatus van de data die in de stack zit. Als er een enkelvoudige eigenaar is, kan deze zelf besluiten waar data te (laten) beheren op basis van eigen criteria, en wie onder welke voorwaarden daartoe toegang krijgt (eenmalige kopie, kopie met updates, gebruiksrecht). Als er meerdere eigenaars zijn, moet alles gezamenlijk worden besloten. Vanuit haar rol als opdrachtgever is Rijkswaterstaat eigenaar van de PMR-NCV data, terwijl de data ingewonnen is door een ad hoc consortium. Het consortium kan Rijkswaterstaat adviseren, maar Rijkswaterstaat kan alleen beslissen. De Zandmotordata wordt ingewonnen door een consortium betaald door een scala aan financiers. In sommige gevallen kan een dataset gereduceerd worden tot een set van data met enkelvoudige bronhouders. In het geval van de Zandmotor is dat echter niet mogelijk, omdat bepaalde subsets ingewonnen zijn door een team van verschillende instituten. De data van de STRAINS- en MegaPEX-experimenten bijvoorbeeld zijn ingewonnen door onderzoekers van TU Delft, NOC Liverpool (UK) en University of Washington (USA) met instrumenten van deze drie instituten. De instrumenten zijn met een schip van Rijkswaterstaat uitgezet en opgehaald. Deze data is dus niet te splitsen tot een subset met enkelvoudige bronhouders. Voor archiefdoeleinden zou elke bronhouder een kopie kunnen krijgen, maar voor beslissingsdoeleinden moeten alle bronhouders in de loop blijven.

De meervoudige juridische status kan voorkomen worden als een partij -tegen betaling of gratis- afstand doet van het eigenaarschap van de data. Data vertegenwoordigen weliswaar waarde, maar ook een beheerlast. Dit wordt nogal eens vergeten. Een partij moet zich dus afvragen of de kosten van de beheerlast opwegen tegen de waarde van de data voor de eigen organisatie. Zeker als data na verloop van tijd openbaar wordt, brengt het juridische eigenaarschap vooral kosten met zich mee en geen baten ten opzichte van andere partijen. Meervoudige data vertegenwoordigt bovendien een grotere waarde als zij tezamen aangeboden worden met complementaire data.

Het regelen van eigendomsrecht vooraf is van groot belang. Waar mogelijk verdient de constructie van enkelvoudige bronhouder de voorkeur. Open data vergroten kans op hergebruik en daarmee de potentiële baten.

5.2 Waarde, onderhoudskosten en adaptieve opwerking

Bij het beoordelen van de mariene datasets van Wind op Zee speelde een soortgelijke vraag. Het blijkt dat deze data aardig wat achterstallig onderhoud vergt om de data en metadata geschikt te maken voor verwerking door middel van gestandaardiseerde dataformaten. Dit rechtvaardigt de vraag of dergelijke datasets waarde of kosten vertegenwoordigen. Voor datasets waar tijdens het verzamelen en de beoordeling zoveel zorg aan is besteed dat deze

automatisch verwerkbaar zijn, zoals het geval is met de data van PMR-NCV, is de balans van waarde in relatie tot kosten positief. Voor afgesloten datasets als die van Wind op Zee is die negatief.

Datasets waarvoor de balans onderhoudskosten versus waarde negatief is, hoeven niet per se als verloren beschouwd te worden als verdere subsidie voor beheer uitblijft. Mits deze data digitaal is (ondanks dat deze niet gestandaardiseerd is) is de opslag ervan relatief goedkoop. Men moet echter wel de verwachtingen temperen en de data als ruwe data beschouwen. De onderste module van de OpenEarth stack (versiebeheersysteem SubVersion) is er juist op gericht om dit soort ruwe data te archiveren, goedkoop mits met het juiste verwachtingspatroon.

Digitale data kan rustig afwachten in haar digitale opslagplaats, totdat er zich de gelegenheid aandient dat iemand de data onderzoekt, en deze deels een *upgrade* geeft naar een bepaalde standaard, of dat de upgrade gecombineerd kan worden met een vervolgproject, waarin de oude data een meerwaarde vertegenwoordigen. Als eindresultaat komen dan de nieuw verzamelde data en de oude data in opgewerkte vorm beschikbaar. Op deze wijze is er sprake van adaptieve opwerking. Dit soort “slapende” data kan het beste bewaard worden bij een stichting, onderzoeksinstituut of (universiteits-) bibliotheek, die aan de zorgplicht kan voldoen om de ruwe data te bewaren, en waar de kans op hergebruik groot is. De minimale investering die gedaan moet worden is de discovery metadata op orde brengen, zoals van wie was de data, met welk doel was die initieel ingewonnen, waar is de data nu en wat mag je er allemaal wel/niet mee.

In Nederland zijn er verschillende instellingen die beheer van data naar zich toehalen. Het Research Data Netherlands (RDNL) consortium waarin 3TU.Datacentrum, DANS en SURF deelnemen is een goed voorbeeld. 3TU.Datacentrum neemt gratis beperkte open datasets van onderzoekers over, zodat deze door anderen herbruikbaar zijn. De collecties moeten aan minimale eisen voor metadata voldoen. Voor grotere collecties zullen dergelijke partijen kosten in rekening moeten brengen. Het wordt interessant om te zien welk verdienmodel er gaat komen voor beheer van overgedragen datasets. Ook interessant wordt welke parallellen er komen met partijen die fysieke collecties beheren zoals museumstukken en boeken. Het onderbrengen van de als ondeelbaar geachte bibliotheek van het KIT³ kan een mooie analogie zijn. Het KIT kon geen subsidies meer verwerven om de bibliotheken te beheren, en had ook geen ander verdienmodel. Een deel van haar boekencollectie is overgedragen aan de Universiteit van Leiden, terwijl een ander deel naar de bibliotheek van Alexandrië in Egypte en een restje aan het publiek is uitgedeeld.

Voor afgesloten datasets als die van Wind op Zee zal ook financiering gevonden moeten worden om deze te beheren. Als die niet gevonden wordt, is er een scenario mogelijk dat bijvoorbeeld het ecologische deel wordt overgedragen aan een marien ecologisch instituut voor wie dat deel van de datacollectie meerwaarde vertegenwoordigt, en het fysische deel bijvoorbeeld aan een waterinstituut voor wie die gegevens meerwaarde vertegenwoordigen, en voor publiek toegankelijk gemaakt kan worden. Omdat deze data nog wel ruw is, is er wellicht nog een subsidie nodig als stimulans voor partijen om zich over delen te ontfemen. Een alternatief scenario is dat er ergens bij RWS een doos met Cd's blijft staan met alle ruwe data, die na verloop van tijd onleesbaar (en ongelezen) weggegooid moet worden. Voor het opschrijven van de aanpak om met data als die van Wind op Zee om te gaan, kan een Data herstelplan worden opgesteld. Hierin wordt gekeken welk deel van de data tegen welke kosten opgewerkt kan worden, en welke partijen dat het beste kunnen doen. Hierbij komen we bij de volgende twee overwegingen:

- is een dataset ondeelbaar, enkelvoudig of meervoudig?
- hoe gaan we om met versies tijdens het opwerken van een dataset?

³ <http://www.kit.nl/kit/bibliotheekcollectie-onder-dak/>

Data vertegenwoordigen waarde maar ook kosten. Deze balans moet goed afgewogen worden. Voor nieuwe projecten kunnen de kosten beperkt blijven mits tevoren een goed datamanagement plan is opgesteld, waarin maatregelen getroffen zijn die lange termijn borging mogelijk maakt. Lang wachten met het opwerken en ontsluiten van data van afgesloten projecten brengt het risico met zich mee dat de volle waarde niet meer achterhaald kan worden, omdat niet meer duidelijk is wat de data precies voorstellen. Adaptieve opwerking biedt het voordeel van financiering van het opwerken van data van afgesloten projecten uit vervolgprojecten, omdat de oude data daarin meerwaarde vertegenwoordigen.

5.3 Enkelvoudige versus meervoudige dataset

Een dataset kan als geheel meer waard zijn dan de samenstellende delen. De Zandmotor monitoring is zodanig ontworpen dat de verschillende databronnen elkaar aanvullen. Als de verschillende databronnen van meerdere eigenaren zijn, ontstaat zo de bovenvermelde situatie van meervoudig bronhouder. De Zandmotordata die momenteel bij 3TU staan, zijn ingewonnen door de onderzoekers. Hiernaast heeft Rijkswaterstaat nog haar eigen monitoringsprogramma, waarvan de data niet bij 3TU beheerd worden. De Rijkswaterstaat data zijn echter complementair aan de data van de onderzoekers. Ze zouden dan ook als één dataset beschouwd moeten worden. Voor een overall assessment van de Zandmotor zouden de onderzoeks- en monitoringsdata dus geaggregeerd moeten worden tot een meervoudige dataset. Als Rijkswaterstaat een kopie van haar Zandmotordata zou afstaan aan de onderzoekers, dan zou data overgaan van enkelvoudige tot meervoudige bronhouder. Een overweging om dit te doen zou het bijhouden van verschillende versies kunnen zijn. Een meervoudige dataset hoeft overigens niet fysiek op één locatie te staan. De PMR-NCV data zou in beheer verdeeld kunnen worden in biotische en abiotische data, beheerd door verschillende partijen met de juiste expertise. De hedendaagse webservices zouden van deze verspreide data een virtueel geheel kunnen maken.

Meervoudige datasets vertegenwoordigen extra waarde als de data complementair zijn. Door het gebruik van webservices zoals die door de OpenEarth stack gefaciliteerd worden, kunnen de data verspreid bewaard worden, maar toch als geheel aangeboden worden.

5.4 Versies van data, releases van data, foutmeldingen, bevroren basisdata

Van oudsher wordt bij Rijkswaterstaat data als ruwe data ingewonnen en opgewerkt tot basisdata. De basisdata wordt in het algemeen uitgeleverd aan derden als een statische (bevroren) set. Ondanks alle energie die gestoken wordt in kwaliteitscontrole, zal men bij gebruik van de data op fouten stuiten, die al dan niet terug gerapporteerd worden. Veel data vallen onder een beheer- en onderhoudscyclus waarbij fouten eens in de zoveel tijd gerepareerd worden, en er een nieuwe release van de basisdata komt. Voor sommige data zijn er echter wettelijke kaders waardoor deze niet zonder meer veranderd mogen worden in een nieuwe release. Dit zijn bijvoorbeeld data die gebruikt zijn voor juridische beslissingen, zoals ontruiming of bouwbesluiten, zodat de onderbouwing van de beslissing naderhand traceerbaar blijft en gereconstrueerd kan worden. Deze data worden zelden (laagfrequent) bijgewerkt, en dan alleen als daarvoor nieuwe ruwe data ter beschikking komen. Een voorbeeld zijn de WTI data die gebruikt worden voor wettelijke toetsing van dijken.

Soms duurt verwerking van data zo lang, dat nieuwe data al beschikbaar zijn voordat fouten in de oude data aan het licht komen. Een voorbeeld is de jongste versie van het Actueel Hoogtebestand Nederland, AHN2. In de meeste gevallen is het echter aannemelijk dat er bij intensief gebruik vroeg of laat fouten gevonden worden. De vraag voor het beheer is dan of er

werkelijk sprake is van een fout, en zo ja, hoe daarmee omgegaan moet worden. Simpelweg stellen dat basisdata onveranderlijk is, zodat daarvoor geen procedure ontworpen hoeft te worden, miskent het probleem. Er zal dus nagedacht moeten worden over versiebeheer van basisdata, zeker in het licht van het toenemende opleidingsniveau van gebruikers, gecombineerd met het groeiende (her-)gebruik van data. De meest elementaire vorm is helpdeskwater.nl, waar geconstateerde fouten gemeld kunnen worden. Beheer zonder enige mogelijkheid tot feedback is niet meer van deze tijd.

OpenEarth gaat uit van een iets andere aanpak, waarbij meer verantwoordelijkheid bij de gebruiker komt te liggen, en minder bij de beheerder/bronhouder. De filosofie hierachter is dat gebruik van andermans data op rechten berust, maar ook plichten met zich meebrengt. In OpenEarth worden de ruwe data en de bewerkingssoftware onder versiebeheer opgeslagen. Hierin kunnen alle gebruikers zelf aanpassingen doen. Dit leidt tot nieuwe versies. Eens in de zoveel tijd worden alle aanpassingen beoordeeld, al dan niet meegenomen, en wordt er een nieuwe release uitgebracht. Merk op dat versies een hogere rolatiesnelheid hebben dan releases. Merk ook op dat versies niet geaccordeerd zijn door de bronhouder, terwijl releases dat wel zijn. Deze aanpak is overgenomen uit de softwareontwikkeling (closed en open) waar ook een snelle versiecyclus, en een langzame releasecyclus hand in hand gaan. Er is nog een extra cyclus mogelijk, de zogenaamde *branch*. Hierbij komt er een vertakking voor het omgooien van de verwerking van bepaalde ruwe data. Deze vertakking kan na verloop van tijd *merged* worden met de basisdata, of kan als aparte versie uitgegeven worden. Een voorbeeld is dat AHN2 beschikbaar gesteld wordt als gefilterde set, of als ongefilterde set.

Het toestaan van feedback van gebruikers op (versies van) data gaat makkelijker als gebruikers zich moeten registreren om toegang tot de data te hebben.

Een voorbeeld uit de dagelijkse praktijk, dat verder gaat dan slechts het toestaan van feedback is het actief vragen van feedback over producten of diensten via facebook d.m.v.



Voor feedback op data zou de koppeling met een platform waarvoor men zich moet authenticeren ook voordelen bieden. Het zou een deel van het dagelijkse-sleur werk van helpdeskwater.nl kunnen afvangen. Dit brengt ons bij de volgende overweging.

Het versioneren van data is onmisbaar om de kwaliteit van de data in de toekomst te kunnen garanderen. Feedback van gebruikers moet gefaciliteerd worden. Periodiek vaststellen van versies in releases moet gemeengoed worden.

5.5 Open data vs. data onder embargo: authenticatie

Voor het hosten van niet-open data moeten twee soorten data beheerd worden. Naast de data zelf moet ook het beheer van authenticatie belegd worden. Deze authenticatie moet gekoppeld worden aan databronnen en moet het ongeautoriseerde gebruik van niet-open data voorkomen. Voor het beheer van authenticatiegegevens moet er dus een partij zijn die een authenticatie-database beheert. Dit aspect kan worden genegeerd als alleen open data worden verspreid.

Als alternatief voor het "ouderwets" beheren van een authenticatie database, zijn er ontwikkelingen gaande voor standaarden om authenticatie gegevens universeel uitwisselbaar te maken, net zoals er voor de data zelf uitwisselingsstandaarden zijn. Een oplossing voor het uitwisselen van authenticatiegegevens specifiek voor onderzoekers is ORCID, of OpenID¹ voor het internet in het algemeen. In deze standaarden krijgt iedereen op het internet tijdelijke inlogcodes die een url zijn. Een instelling kan rechten uitdelen aan OpenID accounts door urls toe te wijzen aan bronnen. Hierbij hoeft een instelling geen wachtwoord voor een gebruiker te registreren, maar kan simpelweg checken of een gebruiker zo'n tijdelijke url kan verschaffen. Als die bestaat dan is iemand wie hij zegt dat hij is. KNMI gebruikt OpenID al om klimaatdata uit te wisselen met partners¹. Voor het delen van mariene data is OpenID nog in ontwikkeling, en momenteel nog geen optie.

Voor meervoudige data die open is, zouden de partners elk hun deel van de data kunnen publiceren op eigen servers, die dan door middel van links tot één virtuele dataset geaggregeerd kunnen worden. Voor mariene projecten is dit geen optie. Tijdens het inwinnen van data t.b.v. vergunningsverlening mag data over het algemeen nog niet openbaar zijn, ofwel als eis van overheidswege, ofwel vanuit de vergunningsaanvrager. Pas na vergunningsverlening is data openbaar. Voor mariene data moet het beheer van authenticatiegegevens dus een integraal onderdeel van de oplossing zijn. Het koppelen van databestanden aan een authenticatiedatabase is voor mariene projecten dus een vereiste.

Een bronhouder zou de data en authenticatie kunnen beheren namens alle andere bronhouders. Dat dient dan te gebeuren in een omgeving waarin alle andere bronhouders eenvoudig bij de data kunnen. De systemen van Rijkswaterstaat zijn typisch ingericht voor het intern beheren van data en gebruikers. Ook bij andere bronhouders, zoals bedrijven en universiteiten, is dit doorgaans het geval. Het toekennen van rechten aan externe partijen aan systemen binnen een corporate firewall is meestal geen wenselijke oplossing, en vaak zelfs geen mogelijke oplossing. Binnen instellingen wordt gewoonlijk een Enterprise service bus (ESB) gebruik die gebruikersrechten aan (data)bronnen toekent. Deze ESB gaat uit van een interne gebruikersdatabase. Voor het verlenen van toegang aan externen moet er een aparte gebruikersdatabase voor externen ingericht worden, plus mogelijkheden om door de firewall heen te gaan van zowel data als logingegevens. Een voorbeeld hiervan is de wiki-omgeving die Deltares host voor beheer van documentatie van software voor partners (bijv. Rijkswaterstaat RWSoS), en de omgeving die Deltares heeft ingericht om met partners aan open source projecten te werken (Delft3D, OpenEarthTools). Deltares heeft flink geïnvesteerd in een database met externe users die voldoet aan de vigerende privacy wetgeving plus een IT-omgeving die buiten de corporate firewall staat.

Voor Rijkswaterstaat data die deel uitmaakt van een meervoudige dataset zou het op dergelijke wijze extern beschikbaar maken van interne resources een technische oplossing zijn. De ervaring leert echter dat de veiligheidseisen die Rijkswaterstaat aan IT stelt zo hoog liggen, dat deze optie in feite een illusie is. Dat geldt trouwens evenzeer voor de meeste bedrijven en universiteiten. Voor het hosten van data met meervoudige bronhouders wordt in het algemeen gebruik gemaakt van externe web service bussen, in plaats van enterprise service bussen. Als data onder embargo staat, is het in het algemeen het meest effectief om alles extern te hosten in een aparte omgeving, en wordt huur betaald voor gebruik van de externe omgeving. Commerciële partijen die dergelijke externe oplossingen aanbieden voor meervoudige datasets, zijn bijv. ArcGISOnline van Esri. Het 3TU.Datacentrum heeft flink geïnvesteerd om authenticatie toe te voegen aan het OpenEarth DataLab. SURF is nu partner van 3TU in Research Data Netherlands; SURF beheert de uitwisseling van authenticatie tussen alle universiteiten en hogescholen in Nederland.

Het beheren van een onderlinge authenticatiedatabase is een onderdeel van mariene data dat niet onderschat moet worden. Zaken als het automatisch afhandelen van 'wachtwoord vergeten'-gevallen, het opschonen van personeel dat uit dienst is gegaan bij een partner, of het voldoen aan privacyreglementen vergen een hoge graad van professionaliteit. Het verdient aanbeveling dat te laten doen door een partij die daar aantoonbaar ervaring mee heeft. De complexiteit en bijkomende kosten van authenticatiesystemen kunnen verminderd worden door data open te maken, ook al brengt dat sommige andere nadelen met zich mee.

Voor niet open data is het (al dan niet automatisch) bijhouden van een authenticatiedatabase essentieel. Voor werkelijk open data is dit niet nodig.

5.6 Data expertise, helpdesk

Helpdeskwater.nl heeft niet alleen als functie om gerapporteerde fouten te registreren, maar ook om inhoudelijke vragen van gebruikers te beantwoorden. Bij beheer van data hoort dat er direct of indirect beroep gedaan kan worden op expertkennis van mensen die bij de inwinning of verwerking van de data betrokken waren. Voor de monitoringsdata van Rijkswaterstaat werkt deze helpdesk goed. Voor Zandmotor is er een soortgelijke helpdesk ingericht, die daarnaast ook een taak heeft bij het invoeren van de data. Voor langdurig beheer van projectdata die namens Rijkswaterstaat worden ingewonnen door externe partijen, is het inrichten van een dergelijke helpdesk een ingewikkelder verhaal. Het voortdurend beheer van data door middel van een Service Level Agreement (SLA) bij de partijen die de data hebben ingewonnen, maakt de lijnen kort voor inhoudelijke vragen. Als RWS de data zelf in beheer zou nemen zou ze vragen kunnen doorverwijzen naar de originele partijen. Er zal een afspraak (dunne SLA) moeten komen met deze partijen moeten komen hoeveel vragen ze beantwoorden.

Bij het inrichten van een helpdesk moet er mee rekening gehouden worden dat ook inhoudelijke vragen beantwoord moeten worden. Uitbesteding van de helpdeskfunctie behoort tot de mogelijkheden.

5.7 Vinden van data en opvragen van data zonder DL: voordeel van standaarden

Bij aanvang van dit project was de opdracht om data zo op te werken dat deze automatisch zou koppelen aan de datadistributielaag (DL) van Rijkswaterstaat. De OpenEarth stack bevat de stekkers (standaarden) die aansluiten bij de DL: WFS-Aquo. Een op een willekeurige plek gehoste OpenEarth stack kon zo (i) metadata leveren aan de DL over welke data (wat, waar, wanneer) er allemaal in die stack zit, en (ii) kon vervolgens de data zelf doorleveren, eventueel na transformatie (Figuur 2.10). Helaas heeft de DL van Rijkswaterstaat vertraging opgelopen, waardoor de koppeling nooit daadwerkelijk getest is, en alleen op papier opgeleverd is. Op dit moment is er dus geen mogelijkheid om aan de DL een centrale vraag te stellen die bronnen uit verschillende OpenEarth stacks combineert. De DL kan alleen interne bronnen van Rijkswaterstaat aan. Zelfs de OpenEarth stack gehost binnen Rijkswaterstaat is niet gekoppeld aan de DL, terwijl de stack daar wel de juiste standaarden voor bevat.

Het ontbreken van de DL betekent echter niet dat de in dit project gerealiseerde OpenEarth stacks nu onbruikbaar zijn. Integendeel. Voor de OpenEarth stack en de koppeling met de DL is er gelukkig gekozen om nationale en internationale standaarden te gebruiken: WFS-Aquo. Dit maakt het mogelijk om een reeks aan gangbare open source en commerciële tools te koppelen aan de OpenEarth stacks. Hiermee kunnen de stacks één voor één bevestigd en afgetapt worden.

Daarnaast is het zo dat er overheidsbreed ook al een datadistributielaagtraject liep (zij het minder geavanceerd dan de DL door afwezigheid van Aquo semantiek en het ontbreken van een transformatiemogelijkheid). Deze twee trajecten zijn in Hoofdstuk 1 al even aangestipt: het Nationaal Georegister NGR⁴, en het opendata portaal van de Nederlandse overheid⁵. Zij hosten beide een web-based catalogus voor web services, waarbij NGR metadata doorlevert aan het opendata portaal. Beide catalogi bevatten data van Nederlandse overheidspartijen, andere partijen kunnen hun data niet via deze weg beschikbaar maken. Wel kunnen andere partijen deze overheidscatalogi integraal overnemen, en samenvoegen met andere catalogi met behulp van de OGC CSW standaard voor het uitwisselen van catalogi. Gehoste OpenEarth stacks zouden door Rijkswaterstaat aan een van deze catalogi doorgegeven kunnen worden. Het delegeren van het daadwerkelijk hosten van een service door een andere partij is niet ongebruikelijk. De Dienst der Hydrografie bijvoorbeeld heeft de data web service voor haar grid uitbesteed aan GIS-leverancier Caris⁶. Het is wel zaak dat er een SLA opgesteld wordt voor de gehoste OpenEarth stacks.

Zolang er nog geen SLA is met Rijkswaterstaat over elders gehoste web services, is er geen koppeling met centrale catalogi als NGR en opendata portaal mogelijk. In het geval dat er verschillende stacks zijn op verschillende locaties, wordt het vinden van een dataset niet eenvoudiger. Ten behoeve van een demo voor CIV van dit project heeft Deltares de catalogus component in de OpenEarth stack geactiveerd. De Geonetwork software, exact dezelfde als NGR, is gebruikt om metadata op te slaan van de verschillende mariene projecten. Hierdoor is het toch mogelijk om de metadata van de mariene projecten doorzoekbaar te maken. Deze metadata kan vanuit de catalogus doorgeleverd worden aan andere catalogi.

Met het wegvallen van de koppeling tussen de OpenEarth stacks is niet alleen de zoek- en download functionaliteit van de DL vervallen. Ook een interne RWS webviewer voor de data is nu niet beschikbaar voor de data. De datafeeds kunnen wel één-voor-één, handmatig in andere webviewers geladen worden, zoals de mapviewer⁷. Voor demonstratiedoeleinden voor CIV is er binnen dit project in enkele dagen een open source webviewer geconfigureerd zodat deze de mariene projectdata kon tonen, met wat data van NGR als achtergrond (Figuur 4.2).

Alhoewel Rijkswaterstaat dus door vertraging van de DL niet de kracht van OpenEarth kan benutten voor koppeling aan eigen processen, is de data toch makkelijk benaderbaar met alternatieve tools. Dit is het resultaat van het gezamenlijk besluit om gangbare open standaarden te gebruiken voor deze koppeling in het voorgaande pilot project. Als in de voorafgaande pilot voor dataleverantie niet voor open standaarden was gekozen, dan hadden we nu met lege handen gestaan. Nu is het zaak de stacks met de mariene data bij een geschikte partij te hosten. Daarover gaat de laatste overweging.

Het ontbreken van de distributielaag beperkt de ontsluiting van de opgewerkte mariene data. Het gebruik van standaarden in de OpenEarth stack maakt ontsluiting via viewers toch mogelijk.

⁴ www.nationaalgeoregister.nl/

⁵ data.overheid.nl

⁶ <http://www.nationaalgeoregister.nl/geonetwork/srv/dut/search?#j94aa70da-9787-4195-b78f-ccc0ccbc6a6c>

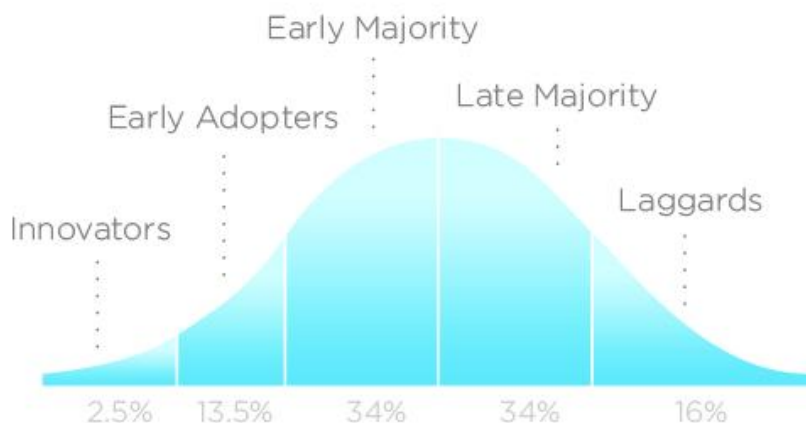
verwijst naar http://inspire.caris.nl/server/services/ows/view/map/bathy_nl

⁷ <http://www.rijkswaterstaat.nl/apps/geoservices/mapviewer2i/>. Er waren zijn er nog technische hobbels om de data in te laden in de beoogde centrale opvolger: geoweb. Hiervoor is binnen dit project contact gelegd met de leverancier.

5.8 Hosting van een OpenEarth stack met mariene data

In dit project is de OpenEarth software stack geschikt gemaakt voor koppeling aan de DL van Rijkswaterstaat. Deze koppeling berust geheel op webservices. Deze koppeling kon uiteindelijk niet getest worden, waardoor er vragen blijven over de volwassenheid van deze techniek. Is dit nog een ver doorgevoerde pilot, of is het klaar voor gebruik? Oftewel, zijn web mapping services in het algemeen, en de OpenEarth stack in het bijzonder, al zover ontwikkeld dat deze als volwassen beschouwd kunnen worden? Deze vraag is relevant voor de overweging of een dergelijke techniek ergens gehost kan worden met de vereisten aan uptime zoals de CIV die stelt, of dat er een service-as-is of best-effort met de ontwikkelaars afgesproken wordt.

Voor adoptie van nieuwe techniek geldt een ontwikkelingscyclus. Voor elke fase van ontwikkeling bestaan er verschillende partijen voor wie dat tot de *comfort-zone* behoort om te gebruiken, of te verkopen. Een gangbare typering is de product levenscyclus van Rogers⁸. Een voorbeeld is de smartphone die inmiddels niet meer weg te denken is uit de maatschappij. De eerste smartphone kwam uit in 2007. In 2010 werd het eerste wikipedia artikel erover geschreven⁹, en inmiddels die het doorgedrongen tot in de haarvaten van de maatschappij. Figuur 5.1 typeert de eerste gebruikers als *innovators*, met daarna de *early adopters* en *early majority*. Wie nu nog geen smartphone heeft is *late majority* of een *laggard*.



INNOVATION ADOPTION LIFECYCLE

Figuur 5.1: Product levenscyclus adoptie curve van Rogers [bron: wikipedia].

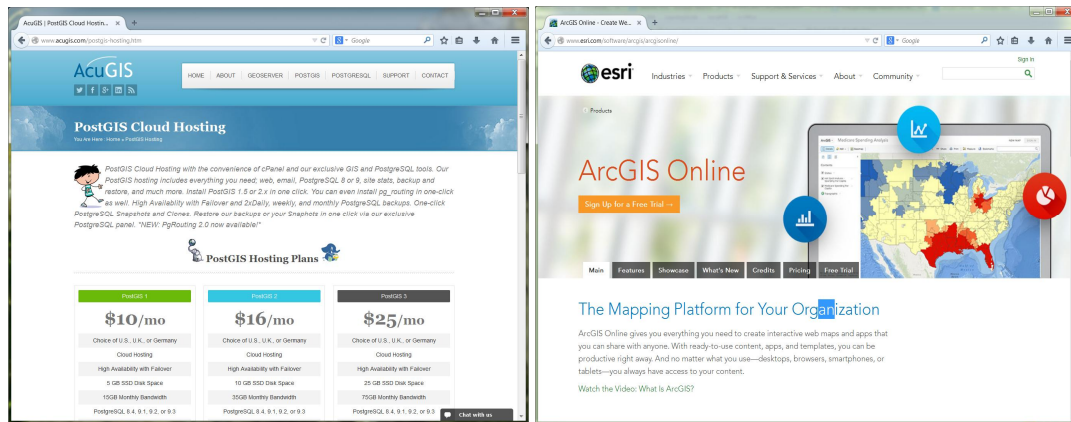
Voor web services zoals in de OpenEarth stack kan hetzelfde concept toegepast worden. Het wikipedia artikel over web mapping services dateert van 2006¹⁰, dus van voor de 1^e smartphone. Het artikel is bescheiden als het vermeldt dat “web mapping today is still being developed” en “advent of web mapping can be regarded as a major new trend in cartography”. Via web services kan iedereen namelijk inmiddels geodata bekijken op zijn smartphone. Voor het huidige project is een demo gegeven waarin de mariene projectdata op een telefoon te bekijken waren, zelfs sneller dan via het bekabelde Rijkswaterstaat intranet (Figuur 4.2). We kunnen dus stellen dat web services, de basis van de OpenEarth stack,

⁸ http://en.wikipedia.org/wiki/Early_adopter

⁹ http://en.wikipedia.org/wiki/IPhone_%281st_generation%29

¹⁰ http://en.wikipedia.org/wiki/Web_mapping

inmiddels volwassen zijn. Dat partijen nu nog niet meedoen aan web services kunnen we typeren als *late majority* of zelfs *laggards*. Voor een commerciële GIS partij als Esri is ArcGISonline bijvoorbeeld inmiddels de hoofdontwikkeling¹¹ ten faveure van hun desktop software. Esri biedt online software aan (SaaS), die kan draaien op PaaS/HaaS servers van Amazon. De gebruikers hoeven zich alleen om de data te bekommeren, de DaaS. Er zijn ook partijen die op commerciële basis open source software aanbieden. De onderste bouwstenen uit de OpenEarth stack zijn bijvoorbeeld te huur via een provider als AcuGIS¹² (Figuur 5.2). De software is weliswaar open source en dus gratis, maar een commercieel bedrijf biedt tegen betaling het gewenste service level niveau. Doordat AcuGIS open source gebruikt, leidt een SLA met hen niet tot *vendor lock in*, en kan er altijd omgeschakeld worden naar een betere aanbieder. Dit Amerikaanse bedrijf biedt gelukkig ook Europese hosting aan, een eis voor hosten van Nederlandse overheidsdata. AcuGIS verhuurt precies het stukje van de OpenEarth stack waarin in dit project mariene data gehost zijn: PostgreSQL/PostGIS/Geoserver. De bedragen, € 25,- per maand voor 25 GB aan data liegen er niet om: dit is uitontwikkelde mainstream technologie voor bodemprijzen.



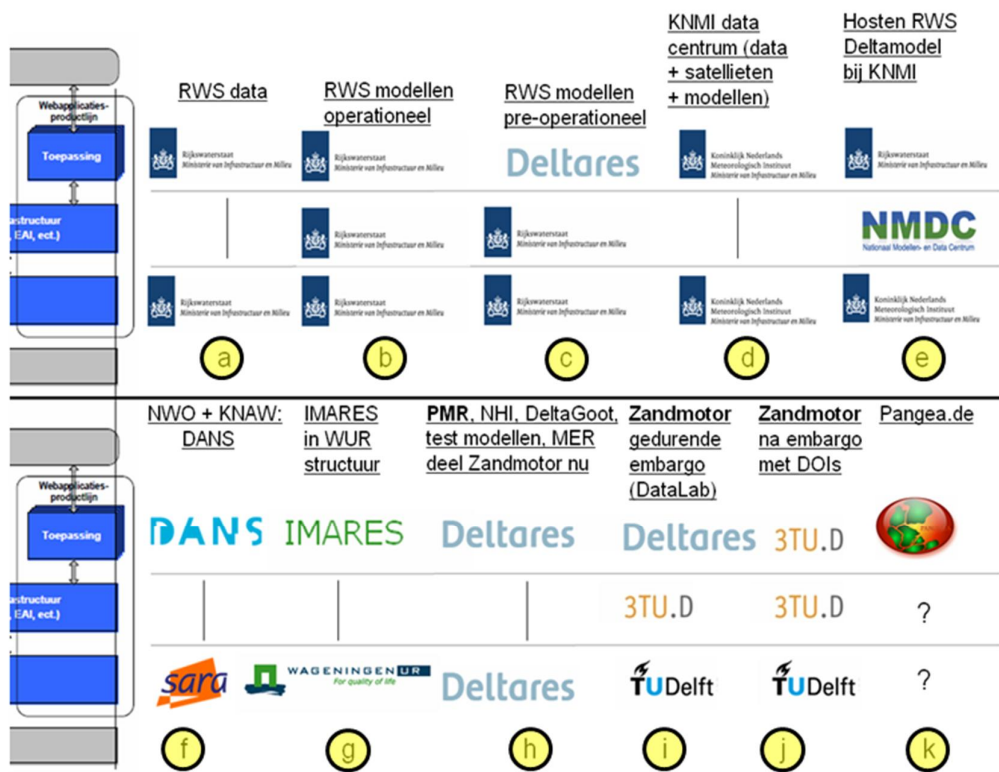
Figuur 5.2 Commerciële cloud hosting van een stukje van de OpenEarth stack door <http://www.acugis.com> voor €25 per maand (links) of het huren van een online web-based GIS omgeving bij Esri gehost bij Amazon (rechts)

Het bestaan van goedkope commerciële aanbieders van een deel van de OpenEarth stack doet de vraag rijzen waarom het beheer van mariene data zo duur is en het maken van koppelingen zoveel vertraging oploopt. Het antwoord is dat een lege database met web services op zich (SaaS) ook niet zoveel kost. In technische termen zijn dit de bedragen voor slechts het afnemen van een SAAS en de onderliggende HaaS en PaaS. De echte kosten voor beheer van mariene data zitten in het maken van een DaaS: in een database opnemen met een solide datamodel, het mappen van de grootheden parameters op Aquo, het inrichten van Geoserver zodanig dat de resulterende WFS Aquo compliant is en de goede metadata doorlevert aan DL. Daarin schuilt het meeste werk. Als data geprotocolleerd is kan deze stap relatief goedkoop zijn. Voor mariene data op projectbasis is altijd maatwerk vereist, en bestaat data voor een lange periode alleen als ruwe data en niet als gestandaardiseerde data. Het vullen van een database en het up-to-date houden ervan zijn dus duur voor mariene projecten. Een goed gevulde DaaS kost dus veel meer dan de € 25,- per maand voor een lege HaaS/PaaS/SaaS. Het maken van afspraken voor het ter duurzame beschikking archiveren en dissemineren van data is DaaS: Data as a Service.

¹¹ Pers.com.Jan-Willem van Eck, ESRI, op maps4society meeting.

¹² <http://www.acugis.com/postgis-hosting.htm>

Het bestaan van spotgoedkope SaaS oplossingen voor gestandaardiseerde data, en het bestaan van gekwalificeerd personeel dat deze SaaS kan benutten om een goede DaaS te garanderen, dient een rol te spelen bij het overwegen van het aangaan van een SLA voor het hosten van mariene data. Aan de een kant verdient het aanbeveling commerciële hosting van de OpenEarth SaaS te gebruiken om kosten te sparen en uptime te garanderen tegen de laagste kosten. Aan de andere kant is de grootste uitdaging het vinden van goed geschoold personeel ten behoeve van een DaaS. In het voorgaande pilot rapport is een aantal voorbeelden gegeven van combinaties van het afnemen van gestapelde services door verschillende partijen. Hier herhalen we de figuur om de volgende SLA overwegingen in perspectief te plaatsen.



Een SLA kan voor een DaaS afgesloten worden, waarbij een aanbieder zelf de onderliggende SaaS en PaaS/HaaS regelt. Voor een optimaal beheer van mariene data is het zorg dat elke component uitgevoerd wordt door de partij die daar het beste toe geëigend is. De vertraging die is opgelopen bij het uitvoeren van het project mariene data kan verklaard worden omdat een aantal partijen een deel van de stack voor hun rekening nam die buiten hun comfort-zone lag. Bij het verdelen van taken en verantwoordelijkheden is het zaak elke partij optimaal in te schakelen, en daarvoor is 'door de DaaS' heen kijken naar de onderliggende SaaS en PaaS/HaaS aan te raden. We kunnen uit dit mariene project concluderen dat voor Rijkswaterstaat het in beheer nemen van externe projectdata niet tot de comfortzone behoort. De CIV is optimaal ingericht om geprotocolleerde data en interne data in beheer te nemen. Voor het maatwerk voor het in beheer nemen van door anderen ingewonnen data bleken de afgelopen drie jaar (pilot en project mariene data) niet de benodigde experts en capaciteit beschikbaar. Uiteraard zou dit kunnen wijzigen als een beleidsaanpassing de experts hier de ruimte voor geeft. In de huidige situatie luidt ons advies dat voor duurzaam beheer van externe mariene projectdata, zonder overgangsregelingen, het afsluiten van een SLA met

een externe partij voor RWS de verstandigste keus lijkt. (NB Dit geldt uiteraard slechts voor dit type op projectbasis ingewonnen meervoudige datasets waarvoor de interne oplossingen van Rijkswaterstaat niet voldoen).

DaaS: voor het van metadata voorzien van data van mariene projecten zijn specialisten op het gebied van ecologie en hydromorfodynamiek nodig. Deze projecten lopen maar een aantal jaar, waarvoor maatwerk vereist is. Bij meerjarige monitoring is er de mogelijkheid om protocollering, standaardisatie en automatisering toe te passen, waarna minder specialisten nodig zijn. De juiste specialisten voor projectduur, die hier tijd aan mogen besteden, zitten in het algemeen bij de partijen die ook de metingen uitvoeren zoals kennisinstellingen (Deltares, IMARES, NIOZ), kennisintensieve ingenieursbureaus (Bureau Waardenburg), maar ook bij algemene ingenieursbureaus (Grontmij, Arcadis), en tenslotte ook bij Rijkswaterstaat zelf (CIV!). Voor het daadwerkelijk opnemen van de gevonden metadata in relationele databases is echter meer expertise vereist dan Aquo codes aanwijzen en in CSV files opnemen. Het blijkt dat deze data experts schaars zijn bij alle partijen. De hoogste concentraties van deze experts zitten bij de kennisinstellingen. Voor het beheer van data kunnen we dan ook adviseren dat deze het beste belegd kan worden bij kennisinstellingen. Deze conclusie past bij het feit dat voor gronddata bijvoorbeeld TNO het landelijk beheer voert. Voor simpele datatypes in bulkhoeveelheden zouden ook ingenieursbureaus een optie zijn.

- Geprotocolleerde data: SLA met ingenieursbureaus
- Projectgebonden maatwerk: SLA met kennisinstellingen

SaaS: De OpenEarth stack is ontwikkeld door kennisinstellingen (Deltares, TU Delft). Een groot deel van de componenten van de OpenEarth stack zijn bestaande open source producten. Deltares heeft als integrator deze software als een stack opgebouwd die als één geheel functioneert. Er zijn commerciële aanbieders die delen van deze stack ook aanbieden. Echter, in die commerciële stacks ontbreekt in het algemeen het deel voor versiebeheer van ruwe data. Voor data die door consortia panklaar als WFS-Aquo verzameld wordt, zou het verkennen van een spotgoedkope aanbieder een aanbeveling zijn. Veel data bestaat echter voor een groot deel en voor een lange subperiode, uit ruwe data. Hier is nog geen commerciële aanbieder voor, maar wel semi-publieke aanbieders. De 3TU kan de OpenEarth stack met authenticatie inmiddels als stack aanbieden. De kennisinstellingen als Deltares zelf zijn minder ingericht voor het aangaan van SLA. Deltares draait bijvoorbeeld voor Rijkswaterstaat RWSoS software alleen testinstanties, en geen operationele instanties. Voor het aanbieden van SaaS services voor data kunnen we dan ook adviseren dat deze het beste belegd kan worden bij een professioneel datacentrum als 3TU voor complexe data, ofwel bij een commerciële aanbieder voor panklare data. Voor data waarvoor geen SLA hoeft te worden afgesloten omdat service-as-is voldoende is, zoals de PMR-NCV data tijdens de project uitvoering, is er ook de optie om deze zelf te hosten, de OpenEarth stack is daarvoor beschikbaar voor iedere partij.

- Service-as-is: een project consortium partner
- SLA: Geprotocolleerde data: commerciële aanbieder
- SLA: Projectgebonden maatwerk: professioneel datacentrum als 3TU.Datacentrum

PaaS/HaaS: Het aanbieden van software om data te hosten heeft onderliggende servers nodig. Als een SaaS commercieel afgenomen wordt, wordt de PaaS automatisch ook geregeld. Het is hierbij zaak te letten op de juridische status hiervan (Europese datacentra), en op zaken als back-up. Een SaaS aanbieder kan - als deze zijn zaakjes goed op orde heeft, makkelijk switchen tussen PaaS aanbieders. Het aanschaffen van hardware om software op te draaien is inmiddels verouderd. Eigen hardware wordt niet optimaal ingezet ('s nachts) en moet na een paar jaar vervangen worden. Het huren van hardware in een private

cloud (eigen IT dienst) of extern is altijd voordeliger. Alleen in het geval van enorme datasets zoals griddata en remote sensing data kan het aanschaffen van eigen storage voordeliger zijn. Dit heeft vooral te maken met de kosten die bij gehuurde IT in rekening gebracht wordt voor bandbreedte. Voor alle overige data is het huren van servercapaciteit de beste oplossing. De goedkoopste aanbieder is over het algemeen een aanbieder die grote volumes aankan. Wereldspelers als Amazon, Google cloud services en Microsoft Azure leveren over het algemeen de goedkoopste oplossingen. De Nederlandse overheid zou als geheel ook haar buying power kunnen gebruiken om goedkoop servers aan te bieden. Men moet zich wel realiseren dat de snelheid van optuigen waarschijnlijk trager is door alle tussenliggende schakels dan als men direct bij een grote commerciële speler afneemt. Een optie is ook dat een Nederlandse overheidspartij in bulk inkoop bij een commerciële partij. SURF zou dat bijvoorbeeld kunnen doen.

- Punt- en lijndata: virtuele servers, fysieke locatie irrelevant mits volgens NL regelgeving. Centraal inkopen kan, maar let op procesvoortgang.
- Griddata: eigen hardware plaatsen nabij gebruikers (bandbreedte)

Voordat er overgegaan wordt om een SLA voor mariene data vorm te geven, dient er nog een vraag beantwoord te worden of er überhaupt een SLA afgesloten dient te worden. Voor niet alle projectdata wegen de kosten van een SLA op om de data te borgen. Het volgende stappenplan zou doorlopen kunnen worden voordat er besloten wordt om tot een SLA over te gaan:

- Data als verloren beschouwen (beheerskosten >> waarde)
- Data ter overname aanbieden, zonder zorgplicht
- Data ter adoptie aanbieden, incl. zorgplicht
- Data bewaren in service-as-is repository, adaptieve ontsluiting stimuleren
- Data bewaren in service-as-is repository, t.b.v. eigen zorgplicht, wachtend op subsidie
- Data herstelplan opstellen
-

De ambitie om data te hosten moet samenvallen met de mogelijkheden. Via het sluiten van Service Level Agreements kunnen de mogelijkheden van externe partijen benut worden.

6 Conclusies en aanbevelingen

Naast de punten van overweging, die in het vorige hoofdstuk zijn opgesomd, volgen onderstaand de conclusies en aanbevelingen uit het project als geheel.

Van PMR-NCV naar Zandmotor datalab

PMR-NCV is de tijd ver vooruit geweest op het gebied van beschikbaar maken van project gerelateerde open data. Achteraf zit de meeste inspanning in het opwerken van de data van de 1 * status van Tim Berners-Lee naar de 3 *** of 4 **** status. Bottleneck hierbij is vooral gedegen afspraken ten aanzien van aan te leveren data. Dat is tevens het belangrijkste leerpunt van alle binnen het project onder handen genomen project gerelateerde biotische en niet biotische data.

De aanpak binnen het zandmotordata lab is een voorbeeld van hoe datalevering aangepakt zou kunnen worden. Binnen de Zandmotor zijn vaste formaten voorgeschreven die door het systeem kunnen worden opgepakt. Het nadeel is dat zodra formaten niet voldoen, het feitelijk onmogelijk is om deze data anders dan alleen in ruwe vorm in het zandmotor datalab op te nemen. Dan zal handmatig een aantal stappen moeten worden doorlopen om de data beschikbaar te maken.

Punten versus de rest

Tot op heden is de aandacht vooral uitgegaan naar de observaties op puntlocaties. De MEP projecten duinen geven echter ook gegevens in raster (luchtfoto's en lidar beelden) en overige vectorformaten (lijnen en polygonen). Deze passen niet of alleen via specifieke handelingen in het datamodel van de OpenEarth stack. Deze voorbeelden leren dan ook dat een gefixeerd data model niet altijd handig is en mogelijk zelfs maar ten dele bruikbaar, bijvoorbeeld alleen voor puntdata.

De stack is daarentegen vanwege de modulaire opbouw flexibel genoeg om met andere formaten dan puntdata om te gaan.

Standaarden

De OpenEarth stack heeft in diverse demo's en projecten laten zien hoe data te delen is, inclusief eventuele mappings naar standaarden waartoe het van oorsprong niet is opgezet. Het OBIS formaat is daarvan een voorbeeld.

De Aquo-standaard voldoet goed voor de 'standaard' observaties. Er zijn echter diverse keuzes gemaakt die niet geheel stroken met de nationale en internationale standaardnotities voor organismen. Het bijhouden van een nationale standaard voor te gebruiken soortnamen zou beter gecoördineerd kunnen worden zoals het voorbeeld van de plantennamen aantoont. Er zijn organisaties in Nederland die de standaardlijsten voor Nederland bijhouden. Voor vegetatie is dat bijvoorbeeld FLORON, voor ongewervelden is dat EIS. In internationaal verband betreft het de veel genoemde WoRMS en PESI lijsten die door INSPIRE worden gebruikt, evenals de CF en EIONET standaarden.

MEP-Duinen heeft naast de stikstofmetingen ook habitattypen als data beschikbaar. Deze zijn niet opgenomen in de Aquo-standaard.

Data distributielaag

In het projectplan staat een aantal punten die betrekking hebben op aansluiting van de OpenEarth stacks op de DL. Binnen de looptijd van het project is geen aansluiting bij de DL

gerealiseerd. De DL is nog niet opgeleverd voor de niet maatschappelijk vitale processen. De DL is vooralsnog alleen in het intranet van RWS beschikbaar.

Een offerte van Deltares voor de aansluiting van de OpenEarth stack op de DL is uiteindelijk niet gehonoreerd.

Het niet beschikbaar zijn van de DL voor Deltares heeft zodanige impact gehad op de verloop van het proces dat de onderdelen van het Plan van Aanpak (thema's b¹³ en d¹⁴) niet of slechts beperkt ingevuld konden worden. Hierover heeft veelvuldig overleg met RWS plaatsgehad.

OpenEarth stack en andere initiatieven

De OpenEarth stack wordt op diverse manieren ingezet om te voldoen aan de vragen die er zijn, op zowel nationaal als internationaal gebied. De bijdrage aan zowel EMODNet Chemistry en Biology alsmede de bijdrage aan de Digitale Atlas Natuurlijk Kapitaal zijn hiervan voorbeelden.

Mariene projectdata na 2014

Het huidige project is eind 2014 afgelopen. Voor 2015 wordt een KPP Beheer en Onderhoud Mariene data met Deltares afgesloten. Deltares voert daarbij het nodige beheer en onderhoud van hard- en software uit. Tevens is Deltares beschikbaar voor een beperkt aantal dagen om ondersteuning te verlenen bij vragen over de data. Eventuele fouten kunnen daarbij worden gecorrigeerd. De ruimte binnen het KPP is echter niet bedoeld om alle data die nog niet is opgewerkt verder te verwerken.

Mariene Data in de toekomst

Er is zowel binnen als buiten Deltares nog een heel scala aan data en dataformaten aanwezig die in aanmerking komen om vindbaar gemaakt te worden. Hoog op de lijst staan de lodingen en andere morfologische gegevens die nu via o.a. kustviewer beschikbaar zijn en/of alleen op de OPeNDAP server van Deltares staan.

Ook meteorologische en hydrodynamische gegevens uit MATROOS horen bij deze set. Het betreft vrijwel uitsluitend rastergegevens die in het huidige project buiten beschouwing zijn gelaten. De ontwikkelde services zijn niet alleen bruikbaar voor mariene data, maar voor data in de volle breedte, waarbij het aan de gebruiker is om de inhoud te bepalen.

Vindbaar maken lijkt hier een doel te zijn, het gaat echter verder dan dat. Als data vindbaar is en beschikbaar in internationaal gestandaardiseerde formats dan is het mogelijk om daarop gestandaardiseerde processen aan te bieden zodat online kalibratie van modellen mogelijk is. Modellen zouden via simpele internetprotocollen gegevens moeten kunnen opvragen ten behoeve van kalibratie en validatie. Dat kan met de services zoals deze door de OpenEarth stack worden geleverd.

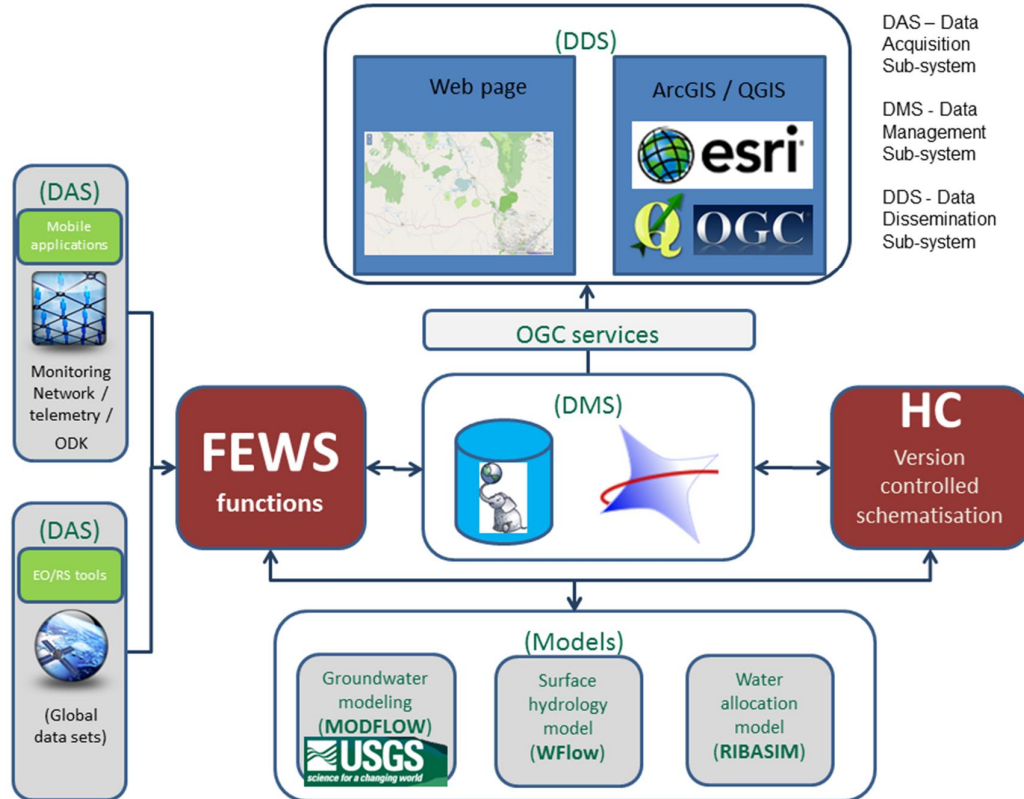
Data in de toekomst

Data hoort een meer centrale plaats te krijgen, al dan niet in de cloud, zodat het kan worden gebruikt en/of aangevuld door een heel scala aan modules. In Figuur 6.1 is de architectuur

¹³ Thema technische doorontwikkeling Deltares Open Earth stack en koppeling aan Data-Distributielaag

¹⁴ inpassen in bestaand IV RWS en beheerlandschap.

opgenomen van een Water Informatie Systeem dat in 2015 binnen een Integraal Waterbeheer project opgezet gaat worden.



Figuur 6.1 OpenEarth stack als onderdeel van Water Information System (WIS)

De DMS bestaat hier uit de OpenEarth stack aangevuld met een heel scala aan zelfstandig opererende modules. In deze figuur is te zien dat FEWS hierin een rol gaat spelen, alsmede het HC (HydroConnect) concept. FEWS wordt hier in de stand alone configuratie gebruikt als een van de mogelijkheden om gestandaardiseerd data te verwerken en te centraliseren naar zowel database als OPeNDAP server. Verder kan FEWS de data weer uit deze centrale opslag halen en gebruiken voor de aansturing van modellen. Deze modellen op hun beurt kunnen door HC op een gestandaardiseerde manier en voorzien van versiebeheer worden geschematiseerd. In het DDS (webviewer) kan vervolgens met behulp van OGC standaarden de gegevens worden gevisualiseerd en/of aanvullende informatie worden opgehaald uit de gegevens (tijdreeksen, animaties).

7 Literatuur

Tim Berners-Lee 5 stars of linked open data, <http://inkdroid.org/journal/2010/06/04/the-5-stars-of-open-linked-data/>

Dirksen, Flip en Ingeborg van Splunder, 2014. PSA Mariene Data v0.91

Hendriksen, G., Van der Kolff, G.H., Van Splunder, I. Borst, K. en J. Staeb, 2014 Handreiking datamanagement Versie 1.0: Datamanagement (mariene) projectmetingen in Rijkswateren

Rijkswaterstaat, Ministerie I&M: Bouwstenen Catalogus DID IAP Generiek Hard en Software Infra, 1 februari 2012.

A Gedetailleerd overzicht opwerking data per project

A.1 PMR-NCV

project	thema	datasets	% opgeslagen in Repository	% in relational Database/netcdf- CF	% OGC WFS/WMS	% Aquo conform	% catalogus	
PMR- NCV	Benthos (totaal)	Benthos	85	85	85	85	85	
		CTD	85	75	75	75	75	
		SECCHI	85	75	75	75	75	
		Sediment	85	75	75	75	75	
PMR- NCV	Vis (totaal)	vismetingen	75	75	75	75	75	
		maaginhoud	75	75	75	75	75	
		snijgegevens	75	75	75	75	75	
		benthos	75	75	75	75	75	
	trawls	CTD						
		vismetingen						
		maaginhoud						
		Secchi 2009						
		Secchi 2010						
		Secchi 2011						
		Vismetingen 2009						
		Vismetingen 2010						
		Vismetingen 2011						
		Waterdiepte 2009						
ctd, dieren 2009								
ctd, dieren 2010								
PMR- NCV	Vogels (totaal)	vogels	100	100	100	100	100	
PMR- NCV	Abiotiek (totaal)	uitvoer	100	100	100	100	100	
		invoerdata						
		validatie data						
PMR- NCV	Gebruik (totaal)	geodatabase_gebruik	100	75	75	75	75	
			85.45454545	80.45454545	80.45454545	80.45454545	80.454545	

A.2 Zandmotor

Zie Tabel 3.2 Overzicht zandmotor datasets en status Zandmotor

A.3 Wind op zee

datasets	% opgeslagen in Repository	% in relational Database/netcdf-CF	% OGC WFS/WMS	% Aquo conform	% catalogus
Verspreiding kleine mantelmeeuw	100	100	100	0	1
Verspreiding vislarven (ruw_IMARES)	100	0	0	0	1
Verspreiding vislarven (ruw_IMARES)	100	0	0	0	1
Verspreiding vislarven (ruw_IMARES)	100	0	0	0	1
Verspreiding zeevogels en bruinvissen(basis)	100	100	100	100	1
Verspreiding zeevogels en bruinvissen (ruw_BUWA)	100	0	0	0	1
Verspreiding zeevogels (ruw_IMARES)	100	0	0	0	1
Verspreiding zeevogels (ruw_IMARES)	100	0	0	0	1
Verspreiding zeevogels(ruw_IMARES)	100	0	0	0	1
Verspreiding zeevogels (ruw_IMARES)	100	0	0	0	1
Verspreiding zeevogels (ruw_IMARES)	100	0	0	0	1
Verspreiding bruinvissen(ruw_IMARES)	100	0	0	0	1
Meetstandaard onderwatergeluid					
Bron onderwatergeluid (OWG)					
Effect OWG op vislarven					
Effect OWG zeezoogdieren					
	100	17	17	8	1

A.4 MEP NSW

datasets	Tabel	% opgeslagen in Repository	% in relational Database/tcdf	% OGC WFS/MS	% Aquo conform catalog	% in us
01. Demersale vis	CTD_NSWDEM_apr14.csv	100	-	-	-	
01. Demersale vis	location_NSWDEM_jan13.csv	100	100	100	100	100
01. Demersale vis	observation_benthos_NSWDEM_jan13.csv	100	100	100	99	99
01. Demersale vis	observation_meet_NSWDEM_jan13.csv	100	100	100	99	99
01. Demersale vis	observation_snij_NSWDEM_jan13.csv	100	100	100	99	99
01. Demersale vis	observation_stomach_NSWDEM_jan13.csv	100	100	100	99	99
02. Pelagische vis	CTD_downcast_NSWPEL_apr2011_per_extracted-CTD_towed-	100	0	0	0	0
02. Pelagische vis	body_net_NSWPELapr2007_apr2011_okt2011.csv-CTD_towed-	100	0	0	0	0
02. Pelagische vis	body_net_NSWPELapr2007_apr2011_okt2011.csv	100	0	0	0	0
02. Pelagische vis	observation_benthos_NSWPEL_jan13.csv	100	100	100	99	99
02. Pelagische vis	observation_meet_NSWPEL_jan13.csv	100	100	100	99	99
03. Gillnet	location_NSWGILL_jan13.csv	100	100	100	100	100
03. Gillnet	observation_benthos_NSWGILL_jan13.csv	100	100	100	100	100
03. Gillnet	observation_meet_NSWGILL_jan13.csv	100	100	100	99	99
04. DIDSON	NSWDIDSON_data_2011.csv	100	100	100	100	100
05. Vogels	MDB met eigen datamodel volgens ESAS richtlijnen	100	100	0	0	0
06. Benthic communities	OWEZ_marine_benthos.xls	100	100	100	100	100
07. Bivalves	NZW 2007 environmental_2007 RefL lander.csv	100	0	0	0	0
07. Bivalves	NZW 2007 environmental_2007 OWEX lander.csv	100	0	0	0	0
07. Bivalves	NZW 2007 juvenile bivalves_geogr positions stations.csv	100	100	100	100	100
07. Bivalves	NZW 2007 juvenile bivalves_identified species.csv	100	100	100	100	100
07. Bivalves	NZW 2007 juvenile bivalves_numbers per boxcore.csv	100	100	100	100	100
08. Macrobenthos	NZW 2011 macrobenthos_boxcore_geographic positions stations.csv	100	100	100	100	100
08. Macrobenthos	NZW 2011 macrobenthos_boxcore_per species per station.csv	100	100	100	100	100
08. Macrobenthos	NZW 2011 macrobenthos_boxcore_totals per station.csv	100	100	100	100	100
08. Macrobenthos	NZW 2011 macrobenthos_Triple-D_biomass_production per species.csv	100	50	0	0	0
08. Macrobenthos	NZW 2011 macrobenthos_Triple-D_geographic positions stations.csv	100	50	0	0	0
08. Macrobenthos	NZW 2011 macrobenthos_Triple-D_n per species per station.csv	100	50	0	0	0
09. Recapture sole and cod		100	50	0	0	0
10. Flight patterns		100	0	0	0	0
11. Harbour seals		100	0	0	0	0
12. Harbour porpoise acoustics		100	0	0	0	0
		100	68	58	58	58

A.5 MEP bestaande duinen

thema	datasets	% opgeslagen in Repository	% in relational Database	% OGC WFS/WM conformance	% Aquos conform catalogus	% in netcdf
planten	knolorchissen	100	0	0	0	0
abiotiek	biomassa/	100	0	0	0	0
abiotiek	grondwater/ 3 gwg rasters	100	0	0	0	0
vegetatie						
e	habitattypen/ 11 rasters	100	0	0	0	0
biotiek	nauwe korfslak/	100	0	0	0	0
biotiek	specifieke soorten/ 2 tabellen	100	0	0	0	0
vegetatie	vegetatie-opnamen/ 2 tabellen en turbovege					
e	gegevens	100	0	0	0	0
vegetatie						
e	vegetatiestructuur/ 4 rasters	100	0	0	0	0
abiotiek	wind/ 3 rasters 1 tabel	100	0	0	0	0
abiotiek	zandspray/	100	0	0	0	0
abiotiek	zoutspray goeree/ 5 rasters	100	0	0	0	0
abiotiek	zoutspray voorne/ 3 rasters	100	0	0	0	0
		100	0	0	0	0

A.6 MEP Nieuwe duinen

thema	datasets	% opgeslagen in Repository	% in relational Database/new tcdf-CF	% OGC WFS/WMS	% Aquo conform	% catalogus
abiotiek	grondwaterstand 2010, 2011, 2012, 2013	100	0	0	0	0
biotiek	vegetatie start vanaf 2014	0	0	0	0	0
Profiel	diepte	100	100	100	0	0
voeroever en strand	ligging GHW	100	100	100	0	0
voeroever en strand	hoogte strand	100	100	100	0	0
voeroever en strand	zandvolume strand	0	0	0	0	0
Morfologie	hoogteligging vlakdekkend	100	100	100	0	0
Morfologie	hoogteligging transecten (n= 16)	0	0	0	0	0
Morfologie	accumulatie deflatie	50	0	0	0	0
Grondwater	grondwaterstanden	70	0	0	0	0
Grondwater	zoetzoutgrensvlak	0	0	0	0	0
Grondwater	grondwatermodellering	0	0	0	0	0
Grondwater	grondwaterkwaliteit	100	0	0	0	0
Vegetatie en flora	structuur	0	0	0	0	0
Vegetatie en flora	locale ruigtes	0	0	0	0	0
Vegetatie en flora	vegetatietypen	0	0	0	0	0
Vegetatie en flora	aandachtssoorten hogere planten en GK-orchis (in vallei)	0	0	0	0	0
Vegetatie en flora	aantal territoria / broedparen	0	0	0	0	0
Recreatie	recreatiedruk / betreding	0	0	0	0	0
Beheer	Beheersmaatregelen	0	0	0	0	0
		36	20	20	0	0

A.7 ECOLIMS-BENTHOS

dataset	Tabel	% opgeslagen in relational database	% in WFS/WFS-MS	% conform catalogus	%
noordzee	MWTL_Noordzeebenthos_20062010_GRONTMIJ_herzien_20120426.xls	100	85	85	50
delta					
monsters	MONSTERS_MWTL_2012	100	100	100	100
delta					
monsters	RECORDS_MWTL_2012	100	92	92	50
delta					
monsters	MONSTERS_MWTL_2011	100	100	100	100
delta					
monsters	RECORDS_MWTL_2011	100	94	93	50
delta					
monsters	MONSTERS_BIOMON_MWTL_1992_2010	100	0	0	0
delta					
monsters	RECORDS_BIOMON_MWTL_1992_2010	100	100	100	100
delta					
monsters	SOORTEN_BIOMON_MWTL_1992_2010				
		100	82	81	64

A.8 KPP B&O Kustlijnzorg (ecologie)

datasets	Tabel	% in relational				
		% opgeslagen in Repository	Database/netcdf- CF	% OGC% WFS/WMS conform	Aquo% catalogus	
bodemdieren2009	monsters	100	100	0	0	0
	data					
bodemdieren2010	n/m2	100	100	0	0	0
	monsters	100	100	0	0	0
bodemdieren2011	dicht,bio,div	100	100	0	0	0
	soorten	100	100	0	0	0
	sediment	100	0	0	0	0
	monsters	100	100	0	0	0
bodemdieren2012	dicht,bio,div	100	100	0	0	0
	soorten	100	100	0	0	0
	sediment	100	0	0	0	0
	monsters	100	100	0	0	0
		100	78.57143	0	0	0

B Uitwisselmodel Mariene Data

EINDCONCLUSIES werksessie uitwisseling MARIENE DATA gehouden op 6 dec 2013

Domeinwaarden zijn te vinden op <http://domeintabellen-idsw.nws.nl>

bron: http://www.aquo.nl/meer_lezen/aquo-onderdelen/aquo-uitwisselformaten/csv-formaat/ en rapport-specificatie-csv-encoding-um-aquo-metingen.pdf
soms wordt gekozen voor de code en soms voor de omschrijving, dit is om praktische redenen gedaan, soms is het ene handig, soms het andere

Inspire
UM AQUO
protocol MD
mogelijk zinvolle uitbreiding protocol MD
mogelijk zinvolle uitbreiding anders

nivo	veldnaam	international	cvo	opmerking / toelichting
1 Business-gegevensset- versie	de "130" Inspire velden			Deze metadata gaat geleverd worden via CSW of OPeNDAP. Hierin staan ook een deel van de in het protocol Mariene Data verplicht gestelde velden zoals met name: naamInwinnendeInstantie, naamDataEigenaar, naamDatabeheerder, naamDataprovider, auteursrechten (creative commons licentie, publiek domein Mark, Geo Gedeeld). Andere verplichte velden uit het protocol Mariene Data staan in tabel 2,3 en 4.
2 meetobject	Identificatie	id	v	Id nummer van meetobject (of locatie of trek) waar "monster" en "meting" naar verwijzen
	Meetpuntomschrijving	locationdescripto	o	naam van meetobject, locatie of trek bijv Noordwijk-12
	Waterbeheerdercode		o	optioneel veld, wordt veel gebruikt door waterschappen.
	Geometripunt (x)		c	de x en y in wgs84. Evt kan ander coördinatenstelsel ook. De EPSG code moet aangegeven worden in tabel 1. Bij een trek of een raai wordt het middelpunt aangegeven. Indien polygonen gewenst zijn, kan dat in principe ook met een speciale string worden gecommuniceerd in het veld "geometrie", maar het is niet zeker dat de ontvangende partij dit begrijpt. Ook is het een mogelijkheid om over te stappen op xml ipv dit csv model. Het csv model wordt in de praktijk echter het meest gebruikt en is voor chemie en biologie doorgaans voldoende.
	Geometripunt (y)		c	
	Geometrie	geometry	c	(vlak, raster, polygoon) dit is een uitbreiding. Voor uitleg zie hierboven
3 Monster	EPSG code	srid	v	bijv. etrs89, wgs84, eds0, RD
	Identificatie	id	v	unieke identificatie monster
	Meetpunt.identificatie		v	verwijzing naar tabel 2 meetobject (de naam van dit veld komt voort uit de historie)
	compartiment.code		v	~28 domeinwaarden zoals: oppervlaktewater, organisme (biota) etc
	orgaan.code		c	indien organisme dan verplicht ~32 domeinwaarden zoals: spierweefsel, dielijk weefsel etc
	Biotaxon.naam		c	indien organisme dan verplicht ~duizenden domeinwaarden zoals Abra, Abra alba, etc.
	begindiepte_m		v	monsteremingsdiepte t.o.v. referentievlak
	einddiepte_m		o	monsteremingsdiepte t.o.v. referentievlak indien het een mengmonster over een dieptereange betreft
	referentievlak.code		v	~28 domeinwaarden zoals: tov mean sea level, tov NAP, tov bodem, tov waterspiegel
	monsterbepalingsmethode.code		o	~23 domeinwaarden zoals: NEN-EN-ISO 5667-3, etc
	bemonsteringsmethode.code		o	~53 domeinwaarden zoals: NENs en RWSVs
	monstercriterium.code		o	De domeintabel bestaat uit twee subtabellen van de domeintabel Hoedanigheid: BiologischeKenmerk 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.
	# niet opnemen #			UM Aquo vermeldt optioneel plaatsbepalingsapparaat, maar hier wordt iets uitzonderlijks mee bedoeld. Niet opnemen dus.
Veldapparaat.omschrijving		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.	
monsteremingsdatum	date	v	datum	
monsteremingstijd	time	o	tijd	
tijd_UTCOffset	timeoffset	c	indien afwijkend van UTC	

4 Meting	Monster.identificatie	v	verwijzing naar tabel 3 monster
	Meetpunt.identificatie	v	verwijzing naar tabel 2 meetobject (de naam van dit veld komt voort uit de historie)
	Grootheid.code	c	~187 domeinwaarden, subdeel van domeintabel parameter je hebt of een grootheid of een typering
	Typering.code	c	~57 domeinwaarden, subdeel van domeintabel parameter je hebt of een grootheid of een typering
	# niet opnemen #	o	UM Aquo vermeld optioneel parameter.groep, maar redundant met parameter.code en .omschrijving
	Parameter.code	c	~3000 domeinwaarden (bij chemische stoffen en objecten verplicht)
	Parameter.omschrijving	c	~3000 domeinwaarden, de omschrijving leest soms makkelijker dan de code (bij chemische stoffen): bij biotaxon is dit verplicht
	Eenheid.code	v	~167 domeinwaarden
	Hoedanigheid.code	v	~415 domeinwaarden verdeeld in referentievlakken, fracties, equivalenten, referentiebases, biologische kenmerken en combinaties hiervan (verplicht veld, in noodgevallen kan ook NVT worden ingevuld)
	# niet opnemen #	o	UM aquo schrijft compartiment.code voor. Dit is een herhaling van het compartiment bij monster omdat in de historische praktijk meetgegevens uitgewisseld werden zonder verder info over het monster. Het is dus niet voor submonsters: Bij submonsters overleggen met IHV/IHM, hoe deze het beste uitgewisseld kunnen worden.
	waardebepalingmethode.code	o	~50 domeinwaarden verdeeld in percentielen en statistiek
	Waardebepalingmethode.omschrijving	o	~600 domeinwaarden, vooral NENs en ISOs. Huismethodes dienen opgenomen te worden als "other: Naam van de huismethode incl nr en versie"
	Waardebepalingstechniek.code	o	~150 domeinwaarden, laboratoriumprincipes, geven globaal idee van de techniek
	Begindatum	v	Datum waarover de waarneming geldig is. In de praktijk meestal de monsternemingsdatum
	Begintijd	o	
	Einddatum	o	Vooraf nodig bij tijdsproportionele bemonsteringen.
	Eindtijd	o	
	Tijd_UTCOffset	c	indien afwijkend van UTC
	Limietsymbool	v	"<" indien waarneming beneden de rapportagegrens is, ">" indien waarneming boven het bereik is.
	Numeriekewaarde	c	De feitelijke waarde!
	Alfanumeriekewaarde	c	indien geen numerieke waarde is gegeven
	Kwaliteitsoordeel.code	o	~13 domeinwaarden. Uit mappingwerkgroep is gebleken dat alleen waarde 0 en 3 betekenis hebben.
	Accreditatiestatus.omschrijving	o	ISO17025:geen;ander
	Meetonzekerheid.bias.waarde	o	
	Meetonzekerheid.reproduceerbaarheid.waarde	o	2 sigma waarde
	Meetonzekerheid.eenheid.code	o	%, ug/L, mg/L, pH etc
	Meetonzekerheid.referentie.waarde	o	indien hierboven % dan hier referentiewaarde opgeven
	Meetonzekerheid.referentie.eenheid	o	indien hierboven waarde ingevuld, dan hier de eenheid: ug/L, pH etc
	reserve1	o	navragen of er nog meer ISO verplicht of relevant zijn (zoals: bewaartermijn overschreden?)
	reserve2	o	
reserve3	o		

C Aquo als standaard voor mariene data

C.1 Ervaringen met betrekking tot het werken met Aquo.

C.1.1 Inleiding

In de opdracht aan Deltares voor de uitvoering van PMR-NCV is als voorwaarde voor de oplevering van data gesteld dat dit in een voor RWS acceptabel formaat moest gebeuren. In het verloop van het project zijn achtereenvolgens een Pilot datamanagement en een vervolg hierop geformuleerd en uitgevoerd. Voor het laatstgenoemde project betreft een van de onderdelen communicatie met Informatie Huis Water (IHW), de organisatie die Aquo beheert. Het bleek namelijk al snel dat het lastig zou worden om voor het project een volledige standaard te kunnen gebruiken.

In de uitvoering van Datamanagement Mariene Projecten heeft met diverse organisaties overleg plaatsgevonden over de wijze waarop monitoringdata het best beschikbaar gemaakt kunnen worden. In dit traject is o.a. overleg geweest met IHW waaruit een lijst van minimaal benodigde velden is gekomen.

C.1.2 Omissies

Voor de diverse metingen in de verschillende percelen van PMR-NCV (benthos, vis, vogels, abiotiek, gebruik) zijn diverse methoden gebruikt waarbij het niet altijd mogelijk bleek om hiervoor een passende Aquo-code te vinden. In sommige gevallen zijn deze methoden ook niet meer te achterhalen vanwege het feit dat het 'historische' gegevens zijn. Wat bij Aquo domeintabellen opvalt, is dat vaak de term 'niet bekend' of 'niet van toepassing' niet voorkomt. Automatisch gebruik van de services die Aquo biedt zijn tevens niet altijd geharmoniseerd. In de ene tabel komt cijfercode (getal type) voor in andere wordt weer gebruik gemaakt van code (tekst type). Dat maakt het gebruik lastiger dan strikt noodzakelijk is.

De volgende twee tabellen laten zien wat hierboven genoemde omissies zijn.

Domein

Bemonsteringsapparaat	ID	Omschrijving	groep	cijfercode
	99	Niet van toepassing		99
	98	Niet bekend		98

Bemonsteringsmethode	ID	Omschrijving	groep	Code
	98	Niet bekend		NB

Bovenstaande tabel is voorgelegd aan servicedesk@ihw.nl op 26-11-2014.

Het antwoord van de servicedesk is hieronder weergegeven in de vorm van een copy van de email ontvangen op 3 december 2014.

Hierbij berichten wij dat uw wijziging genummerd W-1411-0036, is afgehandeld.

Uw vraag luidde:

Bijgaand een verzoek voor het aanvullen van een aantal domeintabellen.

Vanwaar dit verzoek? Voor Rijkswaterstaat wordt op dit moment gewerkt aan het 'open maken' van diverse soorten monitoringgegevens. Het geval wil dat dit project is gestart met een groot aantal dataset die in het verleden (voor 2012) zijn opgeleverd aan Rijkswaterstaat

waardoor een aantal essentiële gegevens niet beschikbaar zijn. Deze zijn ondergebracht in het wijzigingsvoorstel dat is bijgevoegd.

Waarom deze twee domeinen? Het antwoord is dat voor vrijwel alle opgeleverde datasets geldt dat bemonsteringsapparaat of bemonsteringsmethode niet bekend is of niet van toepassing is, terwijl met onze opdrachtgever is afgesproken deze op te nemen in een Aquo conforme manier. Vandaar deze voorstellen.

Ons antwoord luidt:

Dit RfC betreft een aanvraag voor:

- 1) tabel Bemonsteringsapparaat: verzoek om de waarden 'Niet van toepassing' en 'Niet bekend' toe te voegen.*
- 2) tabel Bemonsteringsmethode: verzoek om de waarde 'Niet bekend' toe te voegen.*

Deze aanvraag wordt afgewezen omdat de aangevraagde waarden geen echte valide domeinwaarden voor de betreffende tabellen zijn (ze voldoen niet aan de definitie van het domein), maar ze vallen onder de zogenaamde technische waarden.

Technische waarden kunnen wel binnen een applicatie/database in gebruik zijn, maar ze vallen buiten de Aquo-standaard.

Ter verdere informatie:

A - De attributen waaraan deze tabellen gekoppeld zijn, zijn optioneel in de uitwisseling

- 1) In het UM Aquo-metingen in de klasse MonsterObject zijn de attributen veldApparaat (waaronder het bemonsteringsapparaat valt) en bemonsteringsMethode optioneel te gebruiken.*
- 2) Ook in het voor UM Aquo-metingen beschikbare csv uitwissel formaat zijn deze velden niet verplicht.*

B - Uitwisseling van de aangevraagde waarden is op zich wel mogelijk. Beide domeintabellen zijn van het type codelist, dus in de uitwisseling m.b.v. een UM Aquo xml-bestand mag ook gebruik gemaakt worden van een niet in de domeintabel opgenomen waarde. Die waarde moet dan wel vooraf worden gegaan door de tekst: "other:".

De tekst die na "other:" vermeld staat is vrije tekst, daarop wordt géén controleslag uitgeoefend, behalve dan dat er nog tenminste twee tekens als tekst verwacht worden (het mag dus niet leeg zijn).

C.2 Observaties van terrestrische soorten (vegetatie)

Observaties van terrestrische soorten maken vooral deel uit van de MEP Bestaande duinen en MEP Nieuwe duinen monitoringprogramma's.

Bij MEP Bestaande duinen (met o.a. vegetatie karteringen op Goeree, Voorne en Solleveld/Kapittelduinen) is een habitatype bepaald op basis van twee veel gebruikte vegetatieclassificatie systemen, te weten het systeem van het Staatsbosbeheer en het classificatiesysteem volgens de Flora van Nederland (Schaminée en Stortelder). Zowel de classificatiesystemen als de habitattypen komen niet voor in de Aquo-systematiek. Doordat er onvoldoende tijd was om deze zaken voor te leggen (gegevens pas in december 2014 opgeleverd) is deze casus nog niet aan servicedesk IHW (Aquo) voorgelegd.

Binnen het Nieuwe duinen monitoringprogramma worden onder andere losse vegetatie waarnemingen opgenomen. Hier blijkt dat de systematiek binnen de Aquo-lijst niet geheel correct te zijn of uit te gaan van de vorige standaardlijst van de Nederlandse Flora. Hieronder

een samenvatting van de geconstateerde onterechte toewijzingen aan geslacht, familie en/of verkeerd gespelde soortnamen.

Soortnaam	Opmerking
<i>Abies veitchii</i> 1861	dubbeling van <i>Abies veitchii</i>
<i>Ankistrodesmus subtilis</i>	moet zijn <i>Ankistrodesmus subtilis</i> (4e letter een i)
<i>Arrhenatherum elatius</i> ssp. <i>Elatius</i>	als Species aangeduid, moet zijn Subspecies
<i>Asteromonas</i>	als Species aangeduid, moet zijn Genus
<i>Bidens radiata</i>	als Phylum aangeduid, moet zijn Species
<i>Bromus racemosus</i> ssp. <i>Racemosus</i>	als Species aangeduid, moet zijn Subspecies, Taxonouder wordt <i>Bromus racemosus</i>
<i>Closterium littorale</i> var. <i>Littorale</i>	als Species aangeduid, moet zijn Varietas
<i>Cosmarium difficile</i> var. <i>messikommeri</i>	als Species aangeduid, moet zijn Varietas, Taxonouder wordt <i>Cosmarium difficile</i>
<i>Euastrum richteri</i>	Taxonouder <i>Euastropsis</i> , moet waarschijnlijk zijn <i>Euastrum</i>
<i>Gloeotila longicellularis</i>	als Genus aangeduid, moet zijn Species, Taxonouder wordt <i>Gloeotila</i>
<i>Linaria</i>	Taxonouder <i>Veronica</i> moet zijn <i>Plantaginaceae</i>
<i>Mucidosphaerium</i>	als Species aangeduid, moet zijn Genus
<i>Nitellopsis obtusa</i>	Taxonouder <i>Nitella</i> moet zijn <i>Nitellopsis</i>
<i>Oocystis borgei</i>	als Genus aangeduid, moet zijn Species
<i>Persicaria lapathifolia</i>	Taxonouder <i>Polygonum</i> , moet zijn <i>Persicaria</i>
<i>Persicaria maculosa</i>	Taxonouder <i>Polygonum</i> , moet zijn <i>Persicaria</i>
<i>Schizomeris leibleinii</i>	als Genus aangeduid, moet zijn Species

Deze lijst is nog niet aan IHW voorgelegd. Oorzaak is de achterstand in de verwerking van gegevens.

C.3 Observaties van soorten uit het mariene domein (Fauna)

Bij de start van het monitoringproject PMR-NCV is met de consortiumpartners overeengekomen dat voor de naamgeving van soorten in het mariene domein de lijst die door WoRMS beschikbaar wordt gesteld de leidraad vormt. De omzetting naar Aquo verloopt soepel, al zijn er enkele knelpunten.

Het toevoegen van een soort in Aquo is omgeven door een gehele procedure. De procedure is te vinden op http://www.aquo.nl/meer_lezen/update/indienen-wijziging/#aanvragen-taxon. In het kort komt het erop neer dat inhoudelijk specialisten van het laboratorium van Rijkswaterstaat de aanvraag behandelen. “*Het aanvragen van een nieuw taxon in de TWN-lijst kan alleen worden gedaan door de contactpersoon binnen uw organisatie via de tool TWN-lijst. Weet u niet wie de contactpersoon is binnen uw organisatie, neemt u dan contact op met de servicedesk.*”

Met betrekking tot de grootste dataset met soortgegevens, PMR-NCV, is met behulp van de TWN-lijst van 20-11-2014 de volgende vergelijking gemaakt met de lijst die is bijgehouden binnen het project PMR.

In totaal zijn er binnen het PMR project 595 unieke soorten gevonden. Van die unieke soorten komen er 78 niet voor in de TWN-lijst. De volledige lijst is als Bijlage C.3.1 ingevoegd.

Hieronder worden de 78 soorten opgesomd, die niet in TWN-lijst van 20-11-2014 zijn gevonden. De oorzaak hiervan is niet nader onderzocht. Wel is uitgesloten, dat het om de

verwisseling van hoofd/kleine letters gaat. Dit is in de vergelijking allemaal 'lower case' gemaakt. De hieronder opgenomen aphiaid geeft de mogelijkheid om de exacte beschrijving van de soort in te zien op WoRMS. Als de aphiaid wordt gebruikt in de URL:

<http://www.marinespecies.org/aphia.php?p=taxdetails&id=aphiaid>

dan komt men direct op de juiste pagina. De eerste twee gevallen zijn wel in de PMR database opgenomen maar zijn ofwel geen geaccepteerde soort of er zit een typo in de naam. Aphiaids zijn wel correct. De afkorting "f" staat voor "false".

Parameterdescription	Number of observations	In TWN-biotaxon list	aphiaid	opmerking
PELECYPODA	544	f	105	Not accepted species
<i>plathyhelminthes</i>	81	f	793	Typo in de naam, moet zijn <i>Platyhelminthes</i>
<i>Ctenodrilidae</i>	8	f	905	
<i>Pisces</i>	4	f	11676	
MACTRACEA	16	f	14634	
<i>Ascarididae</i>	169	f	22862	
<i>Pontoporeia</i>	3	f	101744	
<i>Cheirocratus sundevalli</i>	8	f	102798	
PAGUROIDEA	88	f	106687	
<i>Galatheidae</i>	20	f	106733	
<i>Hyas</i>	2	f	106903	
<i>Homarus gammarus</i>	3	f	107253	
<i>Hyas araneus</i>	1	f	107322	
<i>Macropodia tenuirostris</i>	30	f	107346	
<i>Callinectes sapidus</i>	3	f	107379	
<i>Liocarcinus pusillus</i>	2	f	107393	
<i>Necora puber</i>	29	f	107398	
<i>Philocheras fasciatus</i>	6	f	107559	
<i>Palaemon serratus</i>	7	f	107616	
<i>Pandalus borealis</i>	9	f	107649	
<i>Pandalus montagui</i>	6	f	107651	
<i>Pseudocuma (Pseudocuma) simile</i>	8	f	110628	
<i>Diphasia</i>	1	f	117228	
<i>Hydrallmania falcate</i>	13	f	117890	
<i>Anthura gracilis</i>	3	f	118467	
<i>Idotea balthica</i>	8	f	119039	
<i>Siriella armata</i>	16	f	120208	
<i>Ophiura sp.</i>	1381	f	123084	
<i>Echinidae</i>	11	f	123160	
<i>Spatangus purpureus</i>	5	f	124418	
<i>Alcyonium digitatum</i>	3	f	125333	
<i>Sardinella maderensis</i>	30	f	126423	
<i>Gymnammodytes semisquamatus</i>	3	f	126754	
<i>Pegusa lascaris</i>	12	f	127156	

<i>Atherina (Hepsetia) boyeri</i>	6	f	127429
<i>Eunice</i>	12	f	129278
<i>Pectinaria (Lagis) koreni</i>	616	f	130595
<i>Hesionura augeneri</i>	160	f	130646
<i>Phyllodoce (Anaitides) groenlandica</i>	112	f	130668
<i>Phyllodoce (Anaitides) lineata</i>	8	f	130671
<i>Phyllodoce (Anaitides) mucosa</i>	2120	f	130679
<i>Phyllodoce (Anaitides) rosea</i>	120	f	130680
<i>Autolytus edwardsii</i>	8	f	131254
<i>Nymphon rubrum</i>	8	f	134705
<i>Scyphozoa</i>	1	f	135220
<i>Rhizostoma pulmo</i>	19	f	135299
<i>Cyanea lamarckii</i>	62	f	135302
<i>Chrysaora hysoscella</i>	64	f	135304
<i>Aurelia aurita</i>	98	f	135306
<i>Larus fuscus</i>	11784	f	137142
<i>Sula bassana</i>	414	f	137180
<i>Loligo</i>	32	f	138139
<i>Sepia</i>	34	f	138477
<i>Aeolidiella glauca</i>	1	f	138711
<i>Loligo forbesi</i>	5	f	140270
<i>Loligo vulgaris</i>	7	f	140271
<i>Nassarius incrassatus</i>	2	f	140503
<i>Sepia officinalis</i>	54	f	141444
<i>Chamelea gallina</i>	368	f	141907
<i>Fabricia stellaris stellaris</i>	8	f	146433
<i>Alloteuthis subulata</i>	99	f	153131
<i>Psetta maxima</i>	277	f	154473
<i>Fabriciinae</i>	56	f	154918
<i>Tharyx marioni</i>	536	f	155177
NATANTIA	36	f	181484
<i>Caprelloidea</i>	22	f	196121
<i>Tellininae</i>	336	f	225468
TELLINACEA	70	f	246044
<i>Hinia nitidus</i>	233	f	248247
<i>Chelidonichthys lucernus</i>	792	f	274877
<i>Liparis liparis liparis</i>	216	f	293624
<i>Scoloplos (Scoloplos) armiger</i>	2	f	334772
<i>Loligo subulata</i>	224	f	341892
<i>Loligo forbesii</i>	9	f	416668
<i>Chroicocephalus ridibundus</i>	1830	f	465358
<i>Echinidea</i>	3	f	510534

<i>Hydrocoloeus minutus</i>	484	f	567449	
<i>Sternula albifrons</i>	410	f	567480	

Gebruikte websites

http://www.aquo.nl/meer_lezen/aquo-onderdelen/aquo-domeintabellen/

<http://domeintabellen-idsw.rws.nl/DomeinWaardenEdit.aspx?Edit=Lezen>

<http://www.aquo.nl/tools/twn-lijst/>

<http://www.verspreidingsatlas.nl/planten>

<http://marinespecies.org/>

C.3.1 Volledige soortenlijst PMR-NCV

Aantal: is het aantal observaties in PMR-NCV database; biotaxon: geeft aan of de soort in de lijst biotaxon van Aquo is gevonden (t = true, f = false); aphiaid: is het identificatienummer wat door WoRMS aan betreffende soort is toegekend. Deze id kan worden ingevuld in de URL <http://www.marinespecies.org/aphia.php?p=taxdetails&id=aphiaid> waarbij dan aphiaid in het deel id=aphiaid moet worden gevuld met de genoemde ID.

parameterdescription	aantal	biotaxon	aphiaid
<i>Animalia</i>	1780	t	2
<i>Mollusca</i>	16	t	51
POLYPLACOPHORA	16	t	55
GASTROPODA	46	t	101
<i>Bivalvia</i>	268	t	105
PELECYPODA	544	f	105
<i>Littorinidae</i>	3	t	119
<i>Hydrobiidae</i>	22	t	120
<i>Rissoidae</i>	3	t	123
<i>Mytilidae</i>	579	t	211
<i>Pterioida</i>	2	t	212
<i>Veneroïda</i>	6	t	217
<i>Mactridae</i>	3	t	230
PLATHYHELMINTHES	81	f	793
<i>Annelida</i>	15	t	882
POLYCHAETA	1473	t	883
OWENIIDA	24	t	897
TEREBELLIDA	8	t	900
<i>Paraonidae</i>	32	t	903
<i>Ctenodrilidae</i>	8	f	905
<i>Spionidae</i>	480	t	913
<i>Capitellidae</i>	22	t	921
<i>Arenicolidae</i>	19	t	922
<i>Opheliidae</i>	8	t	924
<i>Phyllodocidae</i>	37	t	931
<i>Aphroditidae</i>	13	t	938
<i>Polynoidae</i>	32	t	939
<i>Syllidae</i>	24	t	948
<i>Amphinomidae</i>	3	t	960
<i>Eunicidae</i>	3	t	966
<i>Dorvilleidae</i>	8	t	971
<i>Pectinariidae</i>	8	t	980
<i>Ampharetidae</i>	54	t	981
<i>Terebellidae</i>	96	t	982
NEMERTEA	3208	t	1053
CRUSTACEA	80	t	1066

DECAPODA	568	t	1130
<i>Isopoda</i>	8	t	1131
TANAIDACEA	88	t	1133
<i>Amphipoda</i>	494	t	1135
CUMACEA	195	t	1137
GAMMARIDEA	141	t	1207
<i>Ctenophora</i>	10	t	1248
<i>Sipuncula</i>	8	t	1268
<i>Anthozoa</i>	37	t	1292
<i>Pycnogonida</i>	16	t	1302
INSECTA	35	t	1307
HYDROZOA	147	t	1337
ACTINIARIA	8052	t	1360
NUDIBRANCHIA	32	t	1762
PHORONIDA	632	t	1789
<i>Echinodermata</i>	9	t	1806
<i>Gnathostomata</i>	154	t	1828
<i>Asciacea</i>	2	t	1839
OLIGOCHAETA	1425	t	2036
CHAETOGNATHA	128	t	2081
<i>Actinopterygii</i>	4	t	10194
<i>Pleuronectiformes</i>	81	t	10331
<i>Pisces</i>	4	f	11676
<i>Cephalopoda</i>	3	t	11707
<i>Phoxichilidiidae</i>	8	t	14469
MACTRACEA	16	f	14634
<i>Nereididae</i>	51	t	22496
<i>Ascarididae</i>	169	f	22862
<i>Actiniidae</i>	21	t	100653
<i>Actinia</i>	15	t	100694
<i>Lampetra fluviatilis</i>	21	t	101172
<i>Petromyzon marinus</i>	90	t	101174
<i>Caprellidae</i>	768	t	101361
<i>Aoridae</i>	136	t	101368
<i>Corophiidae</i>	8	t	101376
<i>Melitidae</i>	8	t	101379
<i>Gammaridae</i>	208	t	101383
<i>Haustoriidae</i>	2	t	101384
<i>Oedicerotidae</i>	16	t	101400
<i>Microdeutopus</i>	32	t	101471
<i>Corophium</i>	96	t	101489
<i>Atylus</i>	152	t	101497

<i>Gammarus</i>	42	t	101537
<i>Microprotopus</i>	8	t	101561
<i>Photis</i>	8	t	101563
<i>Leucothoe</i>	32	t	101580
<i>Melita</i>	88	t	101679
<i>Pontocrates</i>	80	t	101702
<i>Bathyporeia</i>	672	t	101742
<i>Pontoporeia</i>	3	f	101744
<i>Stenothoe</i>	16	t	101770
<i>Urothoe</i>	397	t	101789
<i>Ampelisca brevicornis</i>	52	t	101891
<i>Amphilocheus neapolitanus</i>	136	t	101968
<i>Microdeutopus anomalus</i>	64	t	102043
<i>Corophium acherusicum</i>	8	t	102082
<i>Corophium arenarium</i>	80	t	102087
<i>Corophium sextonae</i>	8	t	102100
<i>Corophium volutator</i>	40	t	102101
<i>Atylus falcatus</i>	536	t	102126
<i>Atylus swammerdami</i>	736	t	102131
<i>Apherusa ovalipes</i>	32	t	102172
<i>Gammarus locusta</i>	56	t	102281
<i>Haustorius arenarius</i>	64	t	102317
<i>Microprotopus maculatus</i>	240	t	102380
<i>Parajassa pelagica</i>	8	t	102440
<i>Leucothoe incisa</i>	974	t	102460
<i>Leucothoe lilljeborgi</i>	24	t	102462
<i>Leucothoe spinicarpa</i>	2	t	102470
<i>Orchomenella nana</i>	94	t	102691
<i>Tryphosella sarsi</i>	64	t	102771
<i>Megaluropus agilis</i>	272	t	102783
<i>Abludomelita obtusata</i>	744	t	102788
<i>Cheirocratus sundevalli</i>	8	f	102798
<i>Melita palmata</i>	80	t	102843
<i>Periculodes longimanus</i>	200	t	102915
<i>Pontocrates altamarinus</i>	1152	t	102916
<i>Pontocrates arenarius</i>	8	t	102918
<i>Synchelidium haplocheles</i>	48	t	102924
<i>Bathyporeia elegans</i>	1264	t	103058
<i>Bathyporeia guilliamsoniana</i>	504	t	103060
<i>Bathyporeia pelagica</i>	472	t	103066
<i>Bathyporeia pilosa</i>	104	t	103068
<i>Bathyporeia sarsi</i>	56	t	103073

<i>Bathyporeia tenuipes</i>	32	t	103076
<i>Stenothoe marina</i>	240	t	103166
<i>Stenothoe monoculoides</i>	2	t	103169
<i>Urothoe brevicornis</i>	1504	t	103226
<i>Urothoe poseidonis</i>	4936	t	103235
<i>Asciodiella aspersa</i>	1	t	103718
<i>Ciona intestinalis</i>	2	t	103732
<i>Botryllus schlosseri</i>	8	t	103862
<i>Mustelus</i>	26	t	105732
<i>Mustelus mustelus</i>	18	t	105822
<i>Raja clavata</i>	27	t	105883
<i>Pleurobrachia pileus</i>	36	t	106386
<i>Anomura</i>	30	t	106671
BRACHYURA	511	t	106673
CARIDEA	95	t	106674
PAGUROIDEA	88	f	106687
Galatheidae	20	f	106733
Paguridae	207	t	106738
Majidae	3	t	106760
Portunidae	249	t	106763
Crangonidae	1161	t	106782
Processidae	20	t	106791
Pagurus	57	t	106854
Hyas	2	f	106903
<i>Liocarcinus</i>	124	t	106925
<i>Hippolyte</i>	8	t	106987
<i>Crangon</i>	165	t	107007
<i>Pontophilus</i>	9	t	107011
<i>Palaemon</i>	12	t	107032
<i>Callinassa</i>	211	t	107072
<i>Pisidia longicornis</i>	336	t	107188
<i>Porcellana platycheles</i>	24	t	107190
<i>Diogenes pugilator</i>	6360	t	107199
<i>Pagurus bernhardus</i>	3221	t	107232
<i>Homarus gammarus</i>	3	f	107253
<i>Cancer pagurus</i>	111	t	107276
<i>Corystes cassivelaunus</i>	1486	t	107277
<i>Thia scutellata</i>	3966	t	107281
<i>Ebalia cranchii</i>	1	t	107294
<i>Achaeus cranchii</i>	13	t	107313
<i>Hyas araneus</i>	1	f	107322
<i>Macropodia rostrata</i>	16	t	107345

<i>Macropodia tenuirostris</i>	30	f	107346
<i>Callinectes sapidus</i>	3	f	107379
<i>Carcinus maenas</i>	9852	t	107381
<i>Liocarcinus depurator</i>	474	t	107387
<i>Liocarcinus holsatus</i>	17670	t	107388
<i>Liocarcinus marmoreus</i>	25	t	107390
<i>Liocarcinus navigator</i>	3136	t	107392
<i>Liocarcinus pusillus</i>	2	f	107393
<i>Necora puber</i>	29	f	107398
<i>Portumnus latipes</i>	2881	t	107400
<i>Pilumnus hirtellus</i>	2	t	107418
<i>Eriocheir sinensis</i>	21	t	107451
<i>Hemigrapsus penicillatus</i>	281	t	107454
<i>Pinnotheres pisum</i>	232	t	107473
<i>Athanas nitescens</i>	24	t	107486
<i>Hippolyte varians</i>	20	t	107518
<i>Crangon allmanni</i>	147	t	107551
<i>Crangon crangon</i>	3711	t	107552
<i>Philocheras fasciatus</i>	6	f	107559
<i>Philocheras trispinosus</i>	348	t	107562
<i>Palaemon serratus</i>	7	f	107616
<i>Palaemonetes varians</i>	12	t	107624
<i>Pandalus borealis</i>	9	f	107649
<i>Pandalus montagui</i>	6	f	107651
<i>Processa parva</i>	620	t	107690
<i>Callianassa subterranea</i>	148	t	107729
<i>Bodotriidae</i>	24	t	110378
<i>Diastylis</i>	16	t	110398
<i>Bodotria arenosa</i>	16	t	110440
<i>Bodotria pulchella</i>	8	t	110444
<i>Bodotria scorpioides</i>	64	t	110445
<i>Iphinoe trispinosa</i>	8	t	110462
<i>Cumopsis goodsir</i>	40	t	110465
<i>Diastylis bradyi</i>	452	t	110472
<i>Diastylis rathkei</i>	24	t	110487
<i>Pseudocuma (Pseudocuma) simile</i>	8	f	110628
<i>Diphasia</i>	1	f	117228
<i>Sertularia</i>	1	t	117234
<i>Hydractinia echinata</i>	2	t	117644
<i>Hydrallmania falcata</i>	13	f	117890
<i>Sertularia cupressina</i>	9	t	117913
<i>Tubularia indivisa</i>	13	t	117994

<i>Ione thoracica</i>	16	t	118218
<i>Pseudione hyndmanni</i>	8	t	118240
<i>Idotea</i>	16	t	118454
<i>Anthura gracilis</i>	3	f	118467
<i>Eurydice pulchra</i>	8	t	118852
<i>Idotea balthica</i>	8	f	119039
<i>Idotea chelipes</i>	8	t	119042
<i>Idotea linearis</i>	42	t	119046
<i>Mysidae</i>	72	t	119822
<i>Gastrosaccus</i>	8	t	119859
<i>Gastrosaccus spinifer</i>	2006	t	120020
<i>Mesopodopsis slabberi</i>	320	t	120072
<i>Praunus flexuosus</i>	24	t	120177
<i>Schistomysis kervillei</i>	828	t	120203
<i>Schistomysis spiritus</i>	49	t	120206
<i>Siriella armata</i>	16	f	120208
<i>Asteroidea</i>	6	t	123080
ECHINOIDEA	36	t	123082
<i>Ophiura sp.</i>	1381	f	123084
<i>Echinidae</i>	11	f	123160
<i>Ophiuridae</i>	212	t	123200
<i>Echinocardium</i>	45	t	123426
<i>Ophiura</i>	4	t	123574
<i>Asterias rubens</i>	12599	t	123776
<i>Psammechinus miliaris</i>	180	t	124319
<i>Echinocardium cordatum</i>	7530	t	124392
<i>Spatangus purpureus</i>	5	f	124418
<i>Ophiura albida</i>	7991	t	124913
<i>Ophiura ophiura</i>	12589	t	124929
<i>Ophiura sarsii</i>	24	t	124934
<i>Amphipholis squamata</i>	12	t	125064
<i>Ophiothrix fragilis</i>	50	t	125131
<i>Alcyonium digitatum</i>	3	f	125333
<i>Clupeidae</i>	9	t	125464
<i>Ammodytidae</i>	40	t	125516
<i>Mugilidae</i>	60	t	125546
<i>Ammodytes</i>	374	t	125909
<i>Gobius</i>	19	t	125988
<i>Pomatoschistus</i>	1812	t	125999
<i>Syngnathus</i>	15	t	126227
<i>Anguilla anguilla</i>	138	t	126281
<i>Belone belone</i>	204	t	126375

<i>Alosa fallax</i>	72	t	126415
<i>Clupea harengus</i>	2284	t	126417
<i>Sardina pilchardus</i>	18	t	126421
<i>Sardinella maderensis</i>	30	f	126423
<i>Sprattus sprattus</i>	687	t	126425
<i>Engraulis encrasicolus</i>	69	t	126426
<i>Gadus morhua</i>	353	t	126436
<i>Merlangius merlangus</i>	8690	t	126438
<i>Pollachius virens</i>	6	t	126441
<i>Raniceps raninus</i>	6	t	126442
<i>Trisopterus luscus</i>	309	t	126445
<i>Trisopterus minutus</i>	138	t	126446
<i>Ciliata mustela</i>	216	t	126448
<i>Enchelyopus cimbrius</i>	6	t	126450
<i>Gaidropsarus vulgaris</i>	3	t	126458
<i>Gasterosteus aculeatus</i>	81	t	126505
<i>Osmerus eperlanus</i>	444	t	126736
<i>Ammodytes marinus</i>	177	t	126751
<i>Ammodytes tobianus</i>	624	t	126752
<i>Gymnammodytes semisquamatus</i>	3	f	126754
<i>Hyperoplus lanceolatus</i>	732	t	126756
<i>Callionymus lyra</i>	5970	t	126792
<i>Callionymus maculatus</i>	6	t	126793
<i>Callionymus reticulatus</i>	387	t	126795
<i>Trachurus trachurus</i>	625	t	126822
<i>Aphia minuta</i>	180	t	126868
<i>Crystallogobius linearis</i>	9	t	126878
<i>Gobius niger</i>	27	t	126892
<i>Neogobius melanostomus</i>	21	t	126916
<i>Pomatoschistus lozanoi</i>	1594	t	126925
<i>Pomatoschistus microps</i>	120	t	126927
<i>Pomatoschistus minutus</i>	1279	t	126928
<i>Dicentrarchus labrax</i>	333	t	126975
<i>Chelon labrosus</i>	18	t	126977
<i>Mullus surmuletus</i>	189	t	126986
<i>Pholis gunnellus</i>	129	t	126996
<i>Scomber scombrus</i>	60	t	127023
<i>Pagellus acarne</i>	3	t	127057
<i>Spondyliosoma cantharus</i>	3	t	127066
<i>Zoarces viviparus</i>	177	t	127123
<i>Arnoglossus laterna</i>	6139	t	127126
<i>Hippoglossoides platessoides</i>	3	t	127137

<i>Limanda limanda</i>	17437	t	127139
<i>Microstomus kitt</i>	67	t	127140
<i>Platichthys flesus</i>	7673	t	127141
<i>Pleuronectes platessa</i>	19828	t	127143
<i>Scophthalmus maximus</i>	21	t	127149
<i>Scophthalmus rhombus</i>	457	t	127150
<i>Buglossidium luteum</i>	4172	t	127153
<i>Pegusa lascaris</i>	12	f	127156
<i>Solea solea</i>	11745	t	127160
<i>Oncorhynchus mykiss</i>	3	t	127185
<i>Salmo salar</i>	21	t	127186
<i>Agonus cataphractus</i>	855	t	127190
<i>Myoxocephalus scorpius</i>	576	t	127203
<i>Taurulus bubalis</i>	6	t	127204
<i>Cyclopterus lumpus</i>	45	t	127214
<i>Liparis liparis</i>	261	t	127219
<i>Chelidonichthys cuculus</i>	3	t	127259
<i>Chelidonichthys lucerna</i>	276	t	127262
<i>Entelurus aequoreus</i>	26	t	127379
<i>Syngnathus acus</i>	242	t	127387
<i>Syngnathus rostellatus</i>	809	t	127389
<i>Atherina (Hepsetia) boyeri</i>	6	f	127429
<i>Arenicola</i>	304	t	129206
<i>Heteromastus</i>	5	t	129214
<i>Eunice</i>	12	f	129278
<i>Glycera</i>	200	t	129296
<i>Microphthalmus</i>	96	t	129313
<i>Magelona</i>	240	t	129341
<i>Nephtys</i>	4851	t	129370
<i>Nereis</i>	602	t	129379
<i>Ophelia</i>	48	t	129413
<i>Eteone</i>	3308	t	129443
<i>Eumida</i>	1400	t	129446
<i>Anaitides</i>	664	t	129455
<i>Harmothoe</i>	256	t	129491
<i>Sigalion</i>	8	t	129594
<i>Malacoceros</i>	32	t	129614
<i>Marenzelleria</i>	328	t	129615
<i>Polydora</i>	96	t	129619
<i>Scolecopsis</i>	288	t	129623
<i>Spio</i>	920	t	129625
<i>Streblospio</i>	16	t	129627

<i>Autolytus</i>	56	t	129659
<i>Streptosyllis</i>	8	t	129678
<i>Lanice</i>	183	t	129697
<i>Polycirrus</i>	24	t	129710
<i>Ampharete acutifrons</i>	8	t	129775
<i>Ampharete finmarchica</i>	24	t	129778
<i>Aphrodita aculeata</i>	80	t	129840
<i>Arenicola defodiens</i>	28	t	129867
<i>Arenicola marina</i>	67	t	129868
<i>Capitella capitata</i>	2801	t	129876
<i>Heteromastus filiformis</i>	2680	t	129884
<i>Notomastus latericeus</i>	2536	t	129898
<i>Chaetopterus variopedatus</i>	8	t	129914
<i>Aphelochaeta marioni</i>	1452	t	129938
<i>Chaetozone setosa</i>	8	t	129955
<i>Ophryotrocha gracilis</i>	8	t	130023
<i>Protodorvillea kefersteini</i>	16	t	130041
<i>Glycera alba</i>	24	t	130116
<i>Glycera lapidum</i>	16	t	130123
<i>Glycera oxycephala</i>	32	t	130126
<i>Glycera tridactyla</i>	368	t	130130
<i>Kefersteinia cirrata</i>	8	t	130164
<i>Microphthalmus aberrans</i>	16	t	130168
<i>Microphthalmus listensis</i>	104	t	130172
<i>Microphthalmus sczelkowi</i>	48	t	130174
<i>Microphthalmus similis</i>	176	t	130176
<i>Ophiodromus flexuosus</i>	8	t	130187
<i>Podarkeopsis helgolandica</i>	8	t	130197
<i>Syllidia armata</i>	40	t	130198
<i>Magelona filiformis</i>	16	t	130268
<i>Magelona mirabilis</i>	1816	t	130271
<i>Magelona papillicornis</i>	4440	t	130272
<i>Nephtys assimilis</i>	88	t	130353
<i>Nephtys caeca</i>	778	t	130355
<i>Nephtys cirrosa</i>	8896	t	130357
<i>Nephtys hombergii</i>	3984	t	130359
<i>Nephtys longosetosa</i>	48	t	130364
<i>Eunereis longissima</i>	2088	t	130375
<i>Platynereis dumerilii</i>	40	t	130417
<i>Ophelia limacina</i>	200	t	130494
<i>Travisia forbesii</i>	568	t	130512
<i>Scoloplos armiger</i>	4575	t	130537

<i>Owenia fusiformis</i>	1912	t	130544
<i>Aricidea minuta</i>	72	t	130564
<i>Paraonis fulgens</i>	672	t	130582
<i>Pectinaria (Lagis) koreni</i>	616	f	130595
<i>Pholoe minuta</i>	440	t	130603
<i>Eteone flava</i>	56	t	130613
<i>Eteone foliosa</i>	32	t	130614
<i>Eteone longa</i>	88	t	130616
<i>Eumida sanguinea</i>	32	t	130644
<i>Hesionura augeneri</i>	160	f	130646
<i>Hesionura elongata</i>	16	t	130649
<i>Phyllodoce (Anaitides) groenlandica</i>	112	f	130668
<i>Phyllodoce (Anaitides) lineata</i>	8	f	130671
<i>Phyllodoce (Anaitides) mucosa</i>	2120	f	130679
<i>Phyllodoce (Anaitides) rosea</i>	120	f	130680
<i>Poecilochaetus serpens</i>	648	t	130711
<i>Gattyana cirrhosa</i>	32	t	130749
<i>Harmothoe imbricata</i>	8	t	130769
<i>Harmothoe impar</i>	56	t	130770
<i>Lepidonotus squamatus</i>	8	t	130801
<i>Malmgreniella lunulata</i>	1688	t	130816
<i>Psammodrillus balanoglossoides</i>	32	t	130859
<i>Sigalion mathildae</i>	520	t	131072
<i>Sthenelais boa</i>	592	t	131074
<i>Malacoceros fuliginosus</i>	32	t	131131
<i>Marenzelleria viridis</i>	88	t	131135
<i>Marenzelleria wireni</i>	8	t	131136
<i>Polydora ciliata</i>	16	t	131141
<i>Polydora cornuta</i>	544	t	131143
<i>Pseudopolydora pulchra</i>	208	t	131169
<i>Pygospio elegans</i>	848	t	131170
<i>Scolelepis bonnieri</i>	1928	t	131171
<i>Scolelepis foliosa</i>	40	t	131173
<i>Scolelepis squamata</i>	256	t	131177
<i>Spio filicornis</i>	496	t	131183
<i>Spio gonioccephala</i>	792	t	131184
<i>Spio martinensis</i>	2440	t	131185
<i>Spiophanes bombyx</i>	6592	t	131187
<i>Streblospio benedicti</i>	360	t	131191
<i>Streblospio shrubsolii</i>	648	t	131193
<i>Autolytus brachycephalus</i>	8	t	131252
<i>Autolytus edwardsii</i>	8	f	131254

<i>Proceraea cornuta</i>	40	t	131362
<i>Syllis gracilis</i>	8	t	131435
<i>Lanice conchilega</i>	3444	t	131495
<i>Nymphon</i>	8	t	134591
<i>Achelia echinata</i>	16	t	134599
<i>Ammothella longipes</i>	8	t	134614
<i>Nymphon rubrum</i>	8	f	134705
<i>Anoplodactylus petiolatus</i>	72	t	134723
<i>Scyphozoa</i>	1	f	135220
<i>Rhizostoma pulmo</i>	19	f	135299
<i>Cyanea lamarckii</i>	62	f	135302
<i>Chrysaora hysoscella</i>	64	f	135304
<i>Aurelia aurita</i>	98	f	135306
<i>Chlidonias</i>	22	t	137042
<i>Gavia</i>	12	t	137057
<i>Clangula hyemalis</i>	6	t	137071
<i>Melanitta fusca</i>	110	t	137072
<i>Melanitta nigra</i>	2576	t	137073
<i>Somateria mollissima</i>	1628	t	137074
<i>Alca torda</i>	12	t	137128
<i>Uria aalge</i>	228	t	137133
<i>Larus argentatus</i>	15146	t	137138
<i>Larus canus</i>	1886	t	137141
<i>Larus fuscus</i>	11784	f	137142
<i>Larus marinus</i>	1858	t	137146
<i>Larus minutus</i>	258	t	137148
<i>Larus ridibundus</i>	180	t	137149
<i>Rissa tridactyla</i>	480	t	137156
<i>Sterna albifrons</i>	124	t	137157
<i>Sterna hirundo</i>	5618	t	137162
<i>Sterna paradisaea</i>	2	t	137165
<i>Sterna sandvicensis</i>	14554	t	137166
<i>Stercorarius parasiticus</i>	4	t	137172
<i>Stercorarius skua</i>	14	t	137174
<i>Phalacrocorax carbo</i>	8950	t	137179
<i>Sula bassana</i>	414	f	137180
<i>Podiceps cristatus</i>	110	t	137182
<i>Gavia arctica</i>	2	t	137186
<i>Gavia immer</i>	2	t	137187
<i>Gavia stellata</i>	1698	t	137188
<i>Fulmarus glacialis</i>	14	t	137195
<i>Acanthocardia</i>	12	t	137732

<i>Cerastoderma</i>	52	t	137735
<i>Epitonium</i>	51	t	137943
<i>Hydrobia</i>	21	t	138081
<i>Loligo</i>	32	f	138139
<i>Spisula</i>	151	t	138159
<i>Nassarius</i>	36	t	138235
<i>Crassostrea</i>	59	t	138297
<i>Patella</i>	1	t	138312
<i>Petricola</i>	8	t	138332
<i>Ensis</i>	128760	t	138333
<i>Abra</i>	16	t	138474
<i>Sepia</i>	34	f	138477
<i>Sepiola</i>	227	t	138483
<i>Tellina</i>	194	t	138533
<i>Venerupis</i>	34	t	138647
<i>Aeolidiella glauca</i>	1	f	138711
<i>Buccinum undatum</i>	25	t	138878
<i>Crepidula fornicata</i>	326	t	138963
<i>Acanthocardia paucicostata</i>	28	t	138993
<i>Cerastoderma edule</i>	7885	t	138998
<i>Parvicardium minimum</i>	6	t	139010
<i>Donax vittatus</i>	9090	t	139604
<i>Epitonium clathratulum</i>	12	t	139718
<i>Hydrobia ulvae</i>	16	t	140126
<i>Loligo forbesi</i>	5	f	140270
<i>Loligo vulgaris</i>	7	f	140271
<i>Lutraria lutraria</i>	564	t	140295
<i>Mactra stultorum</i>	40	t	140299
<i>Spisula elliptica</i>	117	t	140300
<i>Spisula solida</i>	1344	t	140301
<i>Spisula subtruncata</i>	10399	t	140302
<i>Montacuta ferruginosa</i>	2376	t	140371
<i>Mya arenaria</i>	7619	t	140430
<i>Mytilus edulis</i>	7873	t	140480
<i>Nassarius incrassatus</i>	2	f	140503
<i>Nassarius nitidus</i>	2641	t	140509
<i>Nassarius reticulatus</i>	11409	t	140513
<i>Euspira catena</i>	25	t	140528
<i>Euspira pulchella</i>	4760	t	140539
<i>Crassostrea gigas</i>	135	t	140656
<i>Ostrea edulis</i>	13	t	140658
<i>Petricola pholadiformis</i>	1317	t	140730

<i>Ensis directus</i>	16877	t	140732
<i>Ensis ensis</i>	40	t	140733
<i>Phaxas pellucidus</i>	12	t	140737
<i>Barnea candida</i>	67	t	140767
<i>Alvania lactea</i>	25	t	141203
<i>Scrobicularia plana</i>	993	t	141424
<i>Abra alba</i>	28277	t	141433
<i>Sepia officinalis</i>	54	f	141444
<i>Sepiolo atlantica</i>	2	t	141454
<i>Macoma balthica</i>	12981	t	141579
<i>Tellina fabula</i>	8250	t	141587
<i>Tellina tenuis</i>	2102	t	141595
<i>Tornus subcarinatus</i>	118	t	141690
<i>Turritella communis</i>	8	t	141872
<i>Chamelea gallina</i>	368	f	141907
<i>Fabricia stellaris stellaris</i>	8	f	146433
<i>Aora typica</i>	224	t	146895
<i>Epitonium clathrus</i>	12	t	146905
<i>Paraonis fulgens</i>	416	t	146932
<i>Tellimya ferruginosa</i>	2947	t	146952
<i>Venerupis senegalensis</i>	2987	t	146955
<i>Mysta picta</i>	8	t	147026
<i>Arenaria interpres</i>	2	t	147431
<i>Haematopus ostralegus</i>	398	t	147436
<i>Morus bassanus</i>	608	t	148776
<i>Anas platyrhynchos</i>	34	t	148791
MYSIDACEA	906	t	149668
<i>Echiichthys vipera</i>	3601	t	150630
<i>Eutrigla gurnardus</i>	219	t	150637
<i>Sander lucioperca</i>	3	t	151308
<i>Hediste diversicolor</i>	408	t	152302
<i>Lagis koreni</i>	403	t	152367
<i>Nemertea</i>	400	t	152391
<i>Alloteuthis subulata</i>	99	f	153131
<i>Psetta maxima</i>	277	f	154473
<i>Hippocampus guttulatus</i>	3	t	154776
<i>Fabriciinae</i>	56	f	154918
<i>Autolytus langerhansi</i>	392	t	155048
<i>Tharyx marioni</i>	536	f	155177
<i>Phyllodocinae</i>	8	t	155424
<i>Ectopleura larynx</i>	18	t	157933
<i>Anas acuta</i>	2	t	158939

<i>Anas clypeata</i>	20	t	158941
<i>Anas crecca</i>	2	t	158943
<i>Tringa nebularia</i>	2	t	158968
<i>Tringa totanus</i>	2	t	158970
<i>Numenius arquata</i>	74	t	159037
<i>Numenius phaeopus</i>	2	t	159040
<i>Calidris alpina</i>	4	t	159047
<i>Calidris alba</i>	46	t	159084
<i>Cygnus cygnus</i>	2	t	159089
<i>Cygnus olor</i>	16	t	159090
<i>Melanitta perspicillata</i>	2	t	159095
<i>Mergus serrator</i>	138	t	159098
<i>Pluvialis squatarola</i>	4	t	159137
<i>Aythya fuligula</i>	2	t	159164
<i>Anas penelope</i>	30	t	159168
<i>Aythya marila</i>	34	t	159172
<i>Branta bernicla</i>	14	t	159175
<i>Branta canadensis</i>	6	t	159176
<i>Bucephala clangula</i>	40	t	159197
<i>Ensis magnus</i>	5	t	160539
<i>Demospongiae</i>	2	t	164811
<i>Enteropneusta</i>	2	t	178738
<i>Venerupis corrugata</i>	32	t	181364
NATANTIA	36	f	181484
Caprelloidea	22	f	196121
Macropodia	516	t	205077
Hinia	58	t	205688
<i>Salmo trutta trutta</i>	15	t	223866
Tellininae	336	f	225468
<i>Aythya ferina</i>	6	t	232037
<i>Tadorna tadorna</i>	52	t	232042
<i>Alitta succinea</i>	232	t	234850
<i>Alitta virens</i>	160	t	234851
<i>Pestarella tyrrhena</i>	221	t	238027
<i>Myrianida langerhansi</i>	304	t	238198
<i>Myrianida prolifera</i>	24	t	238200
TELLINACEA	70	f	246044
<i>Hinia nitidus</i>	233	f	248247
<i>Atherina presbyter</i>	9	t	272030
<i>Symphodus melops</i>	12	t	273571
<i>Chelidonichthys lucernus</i>	792	f	274877
<i>Liparis liparis liparis</i>	216	f	293624

<i>Phyllodoce groenlandica</i>	328	t	334506
<i>Scoloplos (Scoloplos) armiger</i>	2	f	334772
<i>Loligo subulata</i>	224	f	341892
<i>Kurtiella bidentata</i>	5496	t	345281
<i>Loligo forbesii</i>	9	f	416668
<i>Platalea leucorodia</i>	360	t	416678
<i>Anser anser</i>	6	t	416682
<i>Monopseudocuma gilsoni</i>	8	t	422916
<i>Chroicocephalus ridibundus</i>	1830	f	465358
<i>Echinidea</i>	3	f	510534
<i>Alauda arvensis</i>	42	t	558531
<i>Hydrocoloeus minutus</i>	484	f	567449
<i>Sternula albifrons</i>	410	f	567480


D Mapping SeaDataNet - Aquo

Opname Mapping Seadatanet – Aquo in Aquo DS

Leeswijzer


De opname van de seadatanet bestaat uit twee bestanden:

Voorzet opname in DS (xlsx)

 **sdn_mapping_in_ds.xlsx**

een xls (“sdn_mapping_in_ds.xls”) met de Aquo elementen (uitsluitend codes) beperkt tot die Aquo elementen die daadwerkelijk gemapt worden. Dit document bevat tevens een voorzet voor de tabel- en kolomdefinities en structuren zoals deze in de domeintabellenservice zullen worden opgenomen.

Worddocument Opname Mapping Seadatanet (docx)

 **Opname Mapping Seadatanet.docx**

dit worddocument met uitgangspunten

Rode teksten indiceren nog te beantwoorden vragen of acties die nog gedaan moeten worden voor het afgerond kan worden.

Tabeldefinities

De tabeldefinities zijn opgesteld in tabblad 1 van het excel document.

Kolomnamen alleen aanpassen op tabblad 1, werken door in formules in overige tabbladen ten behoeve van csv inlezen bij opname.

Volgorde

De mapping zal altijd uitgaan van het enkelvoudige element in de Seadatanet kolom, naar het meervoudig opgesplitste element in Aquo.

Indien een element niet opgenomen is in Seadatanet is een lege mapping met uitsluitend het Aquo element nutteloos. De lijst krijgt de omvang van de complete lijst van domeinwaarden van seadatanet en Aquo waarin men in het kader van een mapping geïnteresseerd is. Latere toevoegingen zijn altijd mogelijk. Dubbelingen dienen opgelost te worden aan de gebruikerszijde van de mapping. Door dubbelingen kan het voorkomen dat een eenduidige één of wéér mapping niet mogelijk is. De combinatie van de Aquo-code / SDN code dient wel uniek te zijn, waarbij één van de elementen eventueel wel dubbel mag voorkomen. Dit doet zich op het moment uitsluitend tot de mapping naar de tabel “kwaliteitsoordeel”.

Narrow / Broader / Bad match

De mapping tussen SDN en Aquo domeinwaarden is gekozen op basis van expert judgement. Een deel van de domeinwaarden heeft aan de Aquo of de SDN zijde een bredere dan wel nauwere definitie. Dit is bijna onvermijdelijk.

Verschillende tabellen

Indien een SDN tabel over verschillende Aquo tabellen is verspreid, wordt er per Aquo tabel een aparte mapping aangemaakt. Dit is het geval bij de P05 tabel die over respectievelijk de Aquo domeintabellen meetapparaat, bemonsteringsapparaat en plaatsbepalingsapparaat is verdeeld.

Geen Eenheid in SDN parameter mapping.

Rijkswaterstaat meet chemische stoffen in een bepaalde eenheid. De wens van Rijkswaterstaat was om in de mapping op te nemen welke eenheid dat in gebeurt. Deze voorkeurseenheid zal echter niet als kolom worden opgenomen omdat de SDN zijde van deze mapping geen informatie over voorkeurseenheden geeft.

De lossen mapping tussen tabel Unit en Eenheid zal wel worden opgenomen. Voorbeeld:

UGPL <-> ug/l

Omschrijvingen Aquo parameters nooit in mapping

Aan de Aquo zijde van een mapping zal nooit meer informatie opgenomen worden dan de code van de betreffende domeinwaarde. Dit vanwege de verwijkskolom functionaliteit die eenduidigheid tussen domeinwaarden helpt borgen. Verder is deze informatie met automatisering te koppelen door de betreffende domeintabel te koppelen op de code. Daarnaast zou het opnemen van niet code informatie altijd leiden tot dubbel beheer op deze informatie. In Aquo registreren we een domeinwaarde altijd maar één keer. Het IHW is zich er van bewust dat dit de leesbaarheid van de mapping verkleint. In het kader van het doel van de mapping (automatisering) is dit echter te rechtvaardigen in relatie tot de nadelen die het opnemen van extra informatie voor de beheersbaarheid van de mapping introduceert.

Check: Kolomnaam lengte van de op te nemen tabellen.

(Mapping)tabelnamen dienen korter te zijn dan <60 tekens. SDN codes **vallen ruim** binnen de gebruikelijk 12 tekens voor codes in Aquo.

Check: Controleren Aquo codes op geldigheid

De gebruikte Aquo codes in de aangeleverde mappings zijn allen **geldig**.

Meervoudige hoedanigheden

In de opvolger van UM Aquo, het IM Metingen, zullen hoedanigheden meervoudig voorkomen. HoedanigheidCombinaties als het Nnf (uitgedrukt in Stikstof / na filtratie) zullen worden opgesplitst. Omdat deze combinaties uit vaste combinaties tussen verschillende groepen bestaan. (**Gecontroleerd**) wordt er vooruitkijken vast rekening mee gehouden in de mapping.

In het gebruik in UM-Aquo metingen zal tot die tijd de hoedanigheid als combinatie moeten worden uitgewisseld. Er is dus een aanvullende mapping nodig van de meervoudige hoedanigheden naar de enkelvoudige combinatie.

Voor deze mapping is dit beperkt tot één regel: HoedanigheidKorrelgroottefractie Dk0063 en hoedanigheid Referentiebasis 'dg' wordt samen 'DK0063dg'. Deze combinatie is voor dit doel **aangemaakt en opgenomen**.

Hoedanigheid/Monstercriterium

Het in de mapping opgenomen compartiment 'sediment <63um' bestaat niet. Dit betreft compartiment 'BS' en monstercriterium 'Dk0063'.

In de mapping wordt het monstercriterium als optionele kolom toegevoegd. Mogelijke waarden beperking zich (dit is een afspraak, niet afgedwongen door domeintabellenservice) tot waarden uit de tabel Hoedanigheid in de groep 'Korrelgroottefractie'.

Hoedanigheid

De hoedanigheden die betrekking hebben op de referentiebasis, zoals 'dg' (t.o.v. drooggewicht) en 'nf' (na filtratie). Komen in een aparte optionele kolom. Omdat niet valt uit te sluiten dat er andere soorten hoedanigheden of combinaties gekozen zullen worden wordt niet gedefinieerd uit welke groep deze hoedanigheden dienen te komen..

Status/Kwaliteitsoordeel Mapping

De mapping zoals voorgesteld in het brondocument introduceert nogal wat ambiguïteit. Bij het gebruik van deze mapping dient daar op gelet te worden/

Losse besluiten

- In de mappings worden alleen codes opgenomen (dus ook geen datum laatste wijziging SDN zijde)
- Tabelkoppen aan Aquo zijde van de mapping krijgen postfix "code" in de titel.

- Tabelnaam in DS krijgt de naam van de SDN tabel, met prefix Mapping_SDN_<tabelnummer> (dus "P05") en de naam van de betreffende Aquo tabel als dit noodzakelijk is voor het onderscheid zoals in de P05 tabel die uit drie Aquo tabellen bestaat, anders niet.
- Alle mapping waarden opnemen. Ook wanneer dit ambuiguiteit oplevert. Combinaties van sleutels dienen wel uniek te zijn (twee keer dezelfde mapping mag niet voorkomen).
- Mappings die dubieus lijken (bredere mapping op nauwere waarde en visa versa) worden as-is opgenomen.
Mappings zijn bedoeld om twee kanten op te mappen. Dubbelingen worden aan beide zijden geaccepteerd.

Er wordt gekeken nog welke elementen nodig zijn voor de soortentellingen in de P01 mapping.