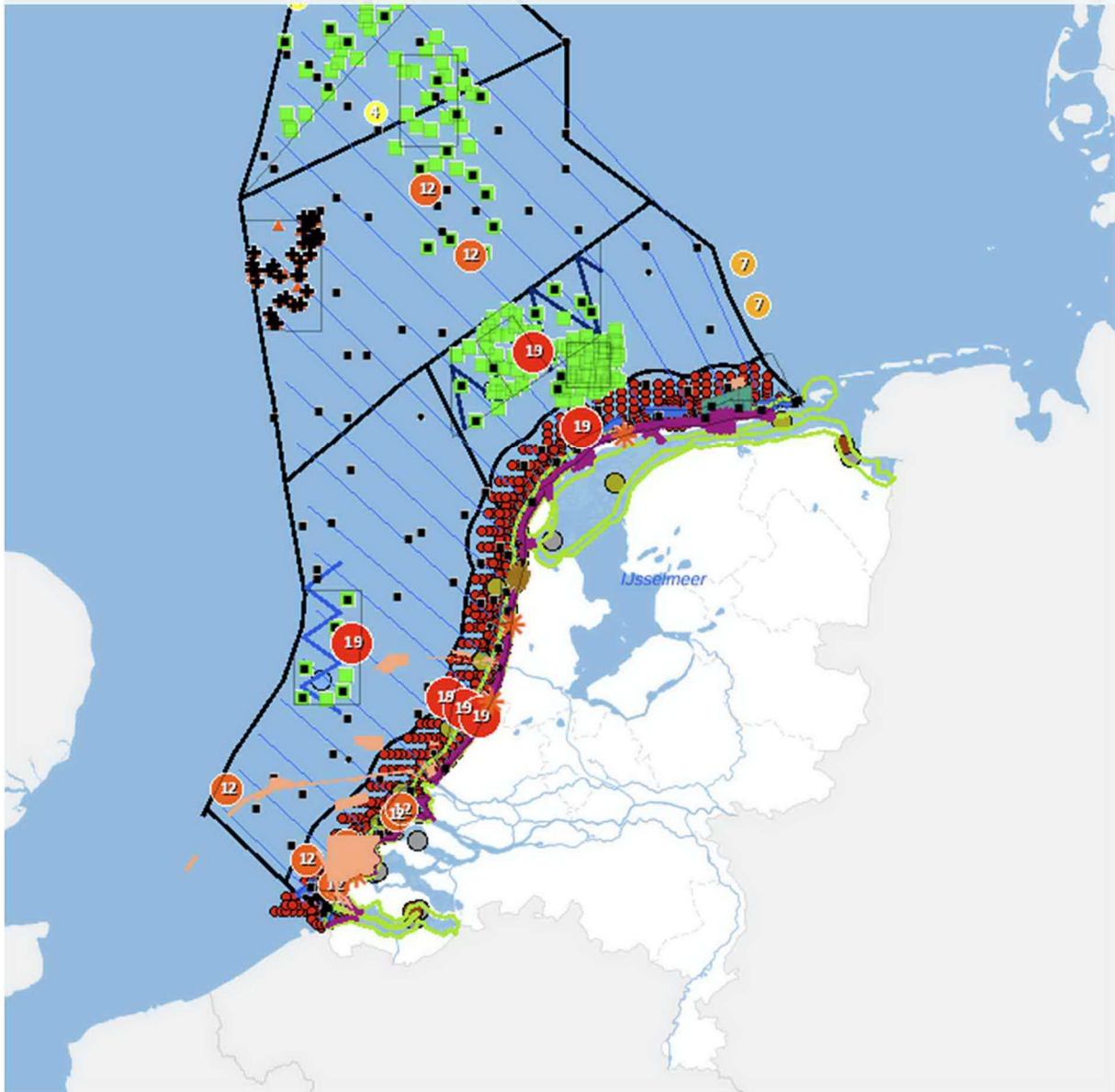


MONS ID137:

Plan van aanpak MONS Dataprotocol als onderdeel van het datamanagement voor het Noordzee Akkoord



29 oktober 2021

Auteur: Informatiehuis Marien

Deliverable: Plan van Aanpak voor opstellen MONS Dataprotocol.

Deadline: 1 november 2021

Titel: Plan van Aanpak opzet MONS Dataprotocol als onderdeel van het datamanagement voor het Noordzee Akkoord

Figuur voorkant:

Screenshot van de Noordzee Monitoringsagenda van het Informatiehuis Marien ([IHM Monitoringsagenda \(rijkswaterstaat.nl\)](http://ihm.rijkswaterstaat.nl))

Informatiehuis Marien (IHM) is een samenwerkingsverband van de ministeries van Infrastructuur en Waterstaat (Rijkswaterstaat), Landbouw, Natuur en Voedselkwaliteit, Defensie (Koninklijke Marine Dienst der Hydrografie). IHM maakt mariene data, informatie en onderzoeksgegevens over de Noordzee toegankelijk voor belangstellenden, professionals en overheden.

SAMENVATTING

IHM zal in 2022 een MONS Dataprotocol ontwikkelen voor het datamanagementbeheer van MONS.

Het Dataprotocol bouwt verder op het Protocol Mariene Data dat door IHM in 2013 is opgesteld. Hierbij worden de volgende principes gehanteerd:

1. Voor de ontsluiting van Data wordt zoveel mogelijk gebruik gemaakt van reeds bestaande en beproefde principes, services en structuren. De aanpak wordt ingebed in NL en Internationale Data- en Datakwaliteitsraamwerken.
2. Voor het uitwisselen van Data worden de FAIR-principes toegepast
3. Data bij de bron!
Het IHM principe blijft: Data worden ontsloten vanuit de databases bij de data experts.
De opdrachtverleners zijn verantwoordelijk voor het regelen van het databeheer via het Uitvoeringsbureau
4. MONS Governance is nodig voor goed werkend Dataprotocol
Datamanagement van dataleveranciers dient onderdeel te worden van de Overeenkomst voor uitbesteding (als Bijlage volgens de ARVODI voorwaarden - Pas na validatie wordt het product vrijgegeven voor betaling)
5. Service voor dataontsluiting
Een pilot wordt gestart voor het gebruik van Linked Data (RDF datamodel) voor data ontsluiting. Doel is om verschillende semantische standaarden te accepteren voor de uitwisseling van (Meta)data:
 - a. AQUO (uitwisseling binnen Nederland)
 - b. SeaDataNet/BODC (internationale ontsluiting Via EMODNET-GBIF- (EUR)Obis-LifeWatch)
 - c. ICES (via EMODNET, onder voorbehoud)
6. Data worden gepubliceerd via één Portaal

Het dataprotocol zal in een aantal runs (schrijfsessies) worden opgezet, waarbij aan het eind van elke sessie de input van Dataleveranciers wordt gevraagd (klankbordgroep). Ook zal er een klankbordgroep worden opgezet voor de nationale en internationale aansluiting bij relevante initiatieven.

Parallel aan deze activiteiten dienen de vereiste services en infrastructuur voor de ondersteuning en uitvoering van het MONS Dataprotocol opgezet te worden. Een eerste schets hiervoor is opgenomen in het Plan van Aanpak.

Een raming van de kosten voor de opzet van de infrastructuur is toegevoegd aan het Plan van Aanpak. Naast investering in data infrastructuur zal de uitbreiding van taken ca. 2 fte inzet vragen van IHM gedurende de looptijd van het programma.

Versie 1 van het Dataprotocol wordt opgeleverd december 2022.

INHOUDSOPGAVE

Samenvatting	ii
Inhoudsopgave.....	iii
1 Inleiding	1
2 Omschrijving van de organisatie	2
2.1 Opdracht en rol IHM	3
2.2 Governance voorwaarden voor datamanagement	3
3 Plan van aanpak Dataprotocol.....	5
3.1 Doel.....	5
4 Uitwerking van de deeltaken.....	6
4.1 Vereiste Onderwerpen van het MONS Dataprotocol:.....	6
4.1.1 Inleiding.....	6
4.1.2 Standaardisatie	6
4.1.3 Validatie	7
4.1.4 Organisatorische afspraken.....	7
4.1.5 Begrippen.....	7
4.2 Vereiste data-infrastructuur voor de ontsluiting van MONS data	8
4.2.1 Uitwisseling van (meta)data	8
4.2.2 Metadata	10
4.2.3 MONS Data Portaal.....	10
4.2.4 Serviceproviders voor ontsluiting van projectdata.....	10
4.2.5 Data Life Cycle Management: Archivering van data	11
4.2.6 Data publicatie	11
4.3 Betrokken partijen.....	11
4.4 Risicofactoren die kunnen leiden tot vertraging van de oplevering van het Dataprotocol	13
4.4.1 Bepalende stap: Governance	13
4.4.2 Bepalende stap: de ontwikkeling van de data-infrastructuur:.....	13
4.4.3 Kritisch proces: de mapping van AQUO-SDN/BODC standaarden.....	14
4.4.4 Belangrijk proces: uitwisseling van data (Peer to Peer).....	14
4.5 Planning van de activiteiten (Gantt chart)	15
4.6 Raming van de kosten	16
5 Bijlagen.....	18

1 INLEIDING

Een gezonde Noordzee is voor iedereen van belang. Nu al vraagt het mariene ecosysteem om bescherming en herstel op basis van bestaand gebruik. Het veranderende gebruik (in vorm en intensiteit) moet passen binnen de ecologische draagkracht van de Noordzee. Het Noordzee Akkoord (NZA) wil de uitdagingen van een veranderend gebruik oppakken en een nieuwe balans vinden en heeft hiervoor een grote behoefte een integraal en systematisch onderzoeks- en monitoringsprogramma dat de basis vormt voor kennis over het functioneren van de Noordzee.

Het programma Monitoring-Onderzoek-Natuurherstel-Soortbescherming (MONS) is hiervoor in het leven geroepen en heeft als doel de centrale vraag te beantwoorden hoe het veranderende gebruik van de Noordzee past binnen de ecologische draagkracht van de Noordzee.

In het MONS-programma gaan veel data verzameld worden, door een groot aantal partijen. Daarnaast zal veel onderzoek plaatsvinden op basis van nieuw verzamelde en bestaande data, wederom door een veelheid aan partijen. Deze dataverzameling en het daaruit voortvloeiende onderzoek geeft uiteindelijk handelingsperspectieven voor goed beheer van de Noordzee.

In het NZA is afgesproken dat alle data en onderzoeksresultaten volgend uit het Noordzee Akkoord worden ontsloten volgens het 'openbaar tenzij' principe.

Het datamanagementbeheer zal zo open mogelijk tot stand worden gebracht, maar de rechten van de opstellers van de gegevens dienen te worden beschermd. Dit houdt in dat de nodige aandacht aan de kwaliteit van de gegevens en de traceerbaarheid van de bronnen moet worden geschonken.

De afgelopen jaren heeft het Informatiehuis Marien (IHM) ervaring opgebouwd in het bundelen, regisseren en beschikbaar maken van data van het Nederlandse mariene milieu. Deze expertise wordt binnen MONS ingezet om een nationaal platform op te zetten voor het (FAIR) beschikbaar maken en houden van mariene data. Het IHM zal hiervoor, in overleg met de NZA-stakeholders (opdrachtgevers en dataleveranciers), een infrastructuur opzetten voor de binnen MONS te verzamelen en te gebruiken data, inclusief afspraken over relevante standaarden, formats en ontsluiting in een Dataprotocol.

MONS biedt een unieke kans om binnen Nederland Open Data op grote schaal en breed gedragen gestandaardiseerd en FAIR te ontsluiten. IHM grijpt deze kans aan om standaardisatie van mariene Data dicht bij de bron te laten en mee te helpen bruggen te slaan tussen verschillende datastandaarden en data-verzamelaars. In samenwerking met het MONS consortium kan gedurende het programma verder worden gebouwd aan een licht, duurzaam, datamanagement raamwerk gebaseerd op 'Best Practices' voor de ontsluiting en meervoudig gebruik van Open Data binnen en tussen verschillende domeinen.

Het MONS programma is eind 2021 gestart met het uitzetten van 'no-regret' projecten: projecten die essentieel geacht worden voor een snelle en adequate start van het Programma. In dit kader is het IHM een opdracht verleend om een Plan van Aanpak te ontwikkelen voor de opzet van een Dataprotocol (Bijlage 1).

Het resultaat van de uitvoering van dit project is een plan van aanpak dat als vertrekpunt zal dienen voor het opstellen van het MONS Dataprotocol.

2 OMSCHRIJVING VAN DE ORGANISATIE

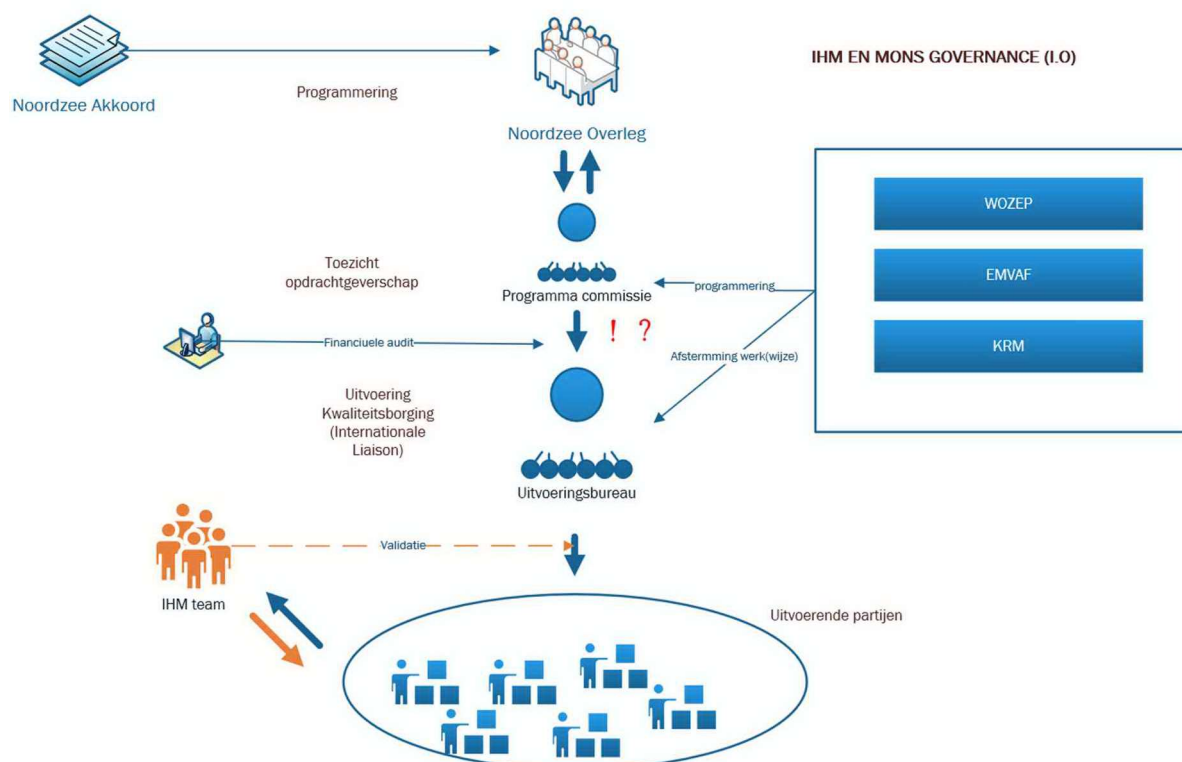
Het programma Monitoring-Onderzoek-Natuurherstel-Soortbescherming (MONS) heeft als doel de centrale vraag te beantwoorden hoe het veranderende gebruik van de Noordzee past binnen de ecologische draagkracht van de Noordzee als ondersteuning voor het behalen van de doelen van het Noordzee Akkoord. Het Noordzee Overleg (NZO) ziet toe op de adequate uitvoering van het Noordzee Akkoord en het toetst de (effectieve) besteding van de beschikbare middelen, waaronder MONS.

De organisatie van MONS ziet er als volgt uit:

NZO: is verantwoordelijk voor de uitvoering van het Noordzee Akkoord en opdrachtgever van het MONS programma. Neemt besluiten over de programmering op basis van voorstellen van de Programmacommissie.

Programmacommissie (PC): fungeert als tussenportaal tussen het uitvoeringsbureau MONS en het NZO. Zij legt de verbinding tussen het NZO (dat programmeert) en het Uitvoeringsbureau (UB) MONS (dat het programma uitvoert). Zij legt gedurende de uitvoering van het MONS programma verantwoording af aan het NZO. De PC stemt af met de Stuurgroep WOZEP, Comité van Toezicht EMVAF en Stuurgroep Marien milieu (t.a.v. KRM en OSPAR), vanwege de onderlinge samenhang en bijdragen voor de uitvoering van het MONS programma.

Uitvoeringsbureau: voert het MONS programma uit als opdrachtnemer van de PC. Het zorgt ervoor dat het MONS programma resultaten oplevert, d.w.z. de beantwoording van de kennisvragen zoals in het programma opgenomen op basis van de uitkomsten van het onderzoek.



Figuur 1. Organogram van de MONS organisatie en het IHM [bron: NZO Governance MONS Programma]

2.1 Opdracht en rol IHM

Het IHM is in het MONS programmadocument¹ aangewezen als uitvoerder van paragraaf 4.10.1 – Ontwikkelen en implementeren van duurzaam databeheer voor effectief gebruik en beheer van de Noordzee. IHM voedt het UB met informatie die nodig is voor projectmanagement (zie Figuur 1).

Het IHM zal, in overleg met de NZA-stakeholders, een infrastructuur vormgeven voor de binnen MONS te verzamelen en te gebruiken data, inclusief afspraken over relevante standaarden, formats en ontsluiting. Vervolgens moet deze infrastructuur worden onderhouden. De inzet van het IHM moet worden ingebed in een governance structuur (2.2) om te zorgen dat opdrachtnemers verantwoordelijk worden voor dataontsluiting binnen deze infrastructuur.

Het IHM zal fungeren als platform voor het uitwisselen van data, door een portal functie te creëren waarin alle MONS data vindbaar zijn. Hiervoor zal het IHM een geschikte data infrastructuur (laten) ontwerpen. Deze dient aan te sluiten op de afspraken die worden gemaakt in het MONS Dataprotocol.

In het Dataprotocol worden afspraken vastgelegd over de te gebruiken standaarden en formats, validatie van opgeleverde gegevens, waardoor de interface met het IHM kan worden onderhouden. Op dit moment biedt het IHM in een 'Open Data Viewer' voor alle bronhouders een overzicht van de beschikbare data. Dit werkt als aanjager voor het ontsluiten van data.

2.2 Governance voorwaarden voor datamanagement

Voor de ontsluiting van geharmoniseerde/ gestandaardiseerde datasets is het van belang dat datastandaardisatie vroeg in de productieketen van data geregeld wordt: bij de dataleverancier. De beste manier om dit te bereiken is om datamanagement te integreren in de governance van het programma. Dit kan op relatief eenvoudige wijze worden bereikt door datamanagement deel te maken van de Overeenkomst met de dataleverancier.

De procedure is als volgt (zie ook Box 1).

- Bij het sluiten van een Overeenkomst worden vooraf afspraken vastgelegd over verschillende aspecten van te leveren data (Data Factsheets²), zoals welke data (in welk detail en format) aangeleverd worden, wie er bij de procesketen betrokken is (datamanagement/provenance - wie doet wat en voor hoe lang?) en toegangsbeheer (moratorium – wie krijgt wanneer toegang tot de data).
- Na afronding van het (deel)project wordt gecontroleerd of de data volgens de afspraken en specificaties van Data Factsheet opgeleverd zijn (data validatie). Daarna ontvangt het Uitvoeringsbureau (UB) bericht dat tot betaling kan worden overgegaan.
- Deze afspraken kunnen contractueel worden vastgelegd in een Overeenkomst als Bijlage³ volgens de ARVODI voorwaarden en/of NOKs.
- IHM voert alleen de datavalidatie uit en stelt het UB op de hoogte van de uitkomst. Hierdoor blijft de inmenging van het IHM in het UB beperkt.

¹ Het MONS-programma zoals vastgesteld door het Noordzee akkoord op 8 september 2021

² De gegevens van de Data Factsheets vormen de basis voor de Metadata

³ ARVODI: 1.2 Bijlage: een aanhangsel bij de Overeenkomst dat na parafering door beide Partijen deel uitmaakt van de Overeenkomst

Het werken met de Data Factsheets is al opgezet binnen het Wind Op Zee Ecologisch Programma (WOZEP), en is efficiënt gebleken. Andere grote projecten zoals Programmatische Aanpak Grote Wateren (PAGW) hebben deze aanpak al geadopteerd voor hun datamanagementbeheer.

Box 1. Consequenties van ARVODI uitbestedingsvoorwaarden:

Enkele relevante artikelen uit Algemene Rijksvoorwaarden voor het verstrekken van opdrachten tot het verrichten van diensten (ARVODI, 2018):

Artikel 1 Begrippen

1.2 Bijlage: een aanhangsel bij de Overeenkomst dat na parafering door beide Partijen deel uitmaakt van de Overeenkomst;

Artikel 4 Beoordeling en acceptatie

4.1 Opdrachtgever beoordeelt de resultaten van de Diensten binnen een termijn van 30 dagen na de dag van levering. Indien Opdrachtgever de resultaten als voldoende beoordeelt, accepteert hij deze door middel van een kennisgeving aan Opdrachtnemer.

4.2 Indien Opdrachtgever de resultaten van de Diensten als onvoldoende beoordeelt, stuurt hij Opdrachtnemer een kennisgeving van niet-acceptatie.....

Artikel 24 Intellectuele eigendomsrechten

24.1 Tenzij anders overeengekomen komen alle auteursrechten die kunnen worden uitgeoefend – waar en wanneer dan ook – ten aanzien van de resultaten van de verrichte Diensten toe aan Opdrachtgever. Deze intellectuele eigendomsrechten worden op grond van de Overeenkomst door Opdrachtnemer op het moment van het ontstaan daarvan aan Opdrachtgever overgedragen, welke overdracht door Opdrachtgever reeds nu voor alsdan wordt aanvaard.

24.2 Alle databankenrechten die kunnen worden uitgeoefend – waar en wanneer dan ook – ten aanzien van de resultaten van de verrichte Diensten komen toe aan Opdrachtgever. Deze intellectuele eigendomsrechten worden op grond van de Overeenkomst door Opdrachtnemer op het moment van het ontstaan daarvan aan Opdrachtgever overgedragen, welke overdracht door Opdrachtgever reeds nu voor alsdan wordt aanvaard.

Uitgangspunt: de drie financierende ministeries blijven juridisch de opdrachtgevers voor dienstverlening (IenW, LNV, EZK). Het uitvoeringsbureau zorgt voor de Overeenkomsten en voert de beoordeling en acceptatie van de overeengekomen diensten uit (art. 4).

De Data Factsheet wordt onderdeel van een Overeenkomst tussen een dataleverancier en de Opdrachtgever na parafering (art. 1.2).

In het kader van door het IHM gehanteerde principe van 'data bij de bron' en artikel 24.2 wordt de opdrachtgever na acceptatie van de dienst (art. 4.1) in principe bronhouder en wordt verantwoordelijk voor het onderhoud en beheer van de data(set).

Deze diensten kunnen door de opdrachtgever in eigen beheer worden uitgevoerd, of worden uitbesteedt aan derden.

Dit dient vastgelegd te worden in de Data Factsheet en uitgewerkt in een Service Level Agreement.

Deze procedure heeft verschillende doelen:

- De opdrachtgever krijgt een objectief en reproduceerbare beoordeling van het voldoen aan de opgegeven specificaties van data en wordt hierdoor voor een ontlast.
- Daarnaast krijgt het voldoen aan bepaalde data specificaties van de oplevering van data als eindproduct een duidelijker, meer dwingend, karakter.
- De gegevens van de Data Factsheet vormen bovendien een goede basis voor de metadata beschrijving van de dataset.
- De additionele inspanning van dataleveranciers wordt zo gering mogelijk gehouden.

3 PLAN VAN AANPAK DATAPROTOCOL

3.1 Doel

Het IHM is gevraagd om een Plan van Aanpak op te zetten voor de opzet van een Dataprotocol (Bijlage 1). Dit Dataprotocol is nodig om er voor te zorgen dat de eisen aan datamanagement voor al het MONS-onderzoek eind 2022 helder zijn.

Het Dataprotocol bevat de instructies t.a.v. datamanagement die gelden bij door MONS uitgevoerd onderzoek. Basis voor het Dataprotocol is een *infrastructuur voor de binnen MONS te verzamelen en te gebruiken data, inclusief afspraken over relevante standaarden en formats en ontsluiting*. Dit houdt in dat het Dataprotocol en de data-infrastructuur parallel dienen te worden ontwikkeld. Hier wordt in het Plan van Aanpak rekening mee gehouden.

Het plan van aanpak voor het opstellen van het Dataprotocol is onderverdeeld in deeltaken (zie hoofdstuk 4):

1. Indeling en vereiste onderwerpen van het Dataprotocol
2. Vereiste data-infrastructuur en services voor de ontsluiting van MONS data
3. Betrokken partijen
4. Risicofactoren die kunnen leiden tot vertraging van het plan van aanpak
5. Planning van de activiteiten
6. Kostenraming voor het opstellen van het Dataprotocol

Het eindproduct van dit no-regret project is een plan van aanpak voor het opstellen van het MONS Dataprotocol in 2022 en een eerste beschrijving van de daarbij behorende infrastructuur en governance voorwaarden.

4 UITWERKING VAN DE DEELTAKEN

4.1 Vereiste Onderwerpen van het MONS Dataprotocol:

Het Protocol Mariene Data van IHM (bijlage 1) dient als uitgangspunt voor de opzet van het Dataprotocol MONS, aangevuld met informatie van het datamanagement van WOZEP en PAGW. Binnen het Dataprotocol wordt gebruik gemaakt van de terminologie van NORA, het datakwaliteitsstelsel, en INSPIRE. Vanwege het internationale karakter van de monitoring- en onderzoeksactiviteiten op de Noordzee wordt het MONS Dataprotocol in het Engels opgesteld. De volgende hoofdstukindeling wordt voorzien.

1. Inleiding
2. Standaardisatie
3. Validatie
4. Organisatorische afspraken
5. Begrippen

Een lijst met te adresseren topics binnen een hoofdstuk is in trefwoorden weergegeven in Bijlage 3.

4.1.1 Inleiding

In de inleiding worden (algemene) principes van data ontsluiting en datamanagement geïntroduceerd en het Dataprotocol nationaal en internationaal ingekaderd. Ook worden de verschillende rollen binnen het datamanagement ingeleid, waar in hoofdstuk 3 van het Dataprotocol op wordt aangehaakt. Ook wordt gespecificeerd op welke wijze en bij welke partijen het beheer van het Dataprotocol duurzaam belegd wordt.

4.1.2 Standaardisatie

Standaardisatie van (meta)datasets is van eminent belang voor eenduidige interpretatie van Data die uitgewisseld worden binnen en tussen domeinen. Er zal een lijst worden opgesteld van een aantal technische en semantische verplichte open standaarden waarin de data en metadata dienen of kunnen worden aangeleverd. De lijst met standaarden van Forum Standaardisatie (<https://www.forumstandaardisatie.nl/open-standaarden>) vormt hiervoor de basis. Deze lijst is geënt op nationaal gebruik. Veel dataleveranciers die werkzaamheden uitvoeren op de Noordzee zijn echter meer ingericht op internationale samenwerking en gebruiken voor de uitwisseling van data internationale standaarden. Dit heeft met name betrekking op de semantische standaarden (woordenboeken). In Nederland is de AQUO standaard een verplichte (semantische) standaard. Binnen Europa zijn veel gebruikte standaarden:

- de ICES standaard voor de uitwisseling van vis/visserij gegevens, en
- de SeaDataNet/BODC standaard voor de uitwisseling van fysisch-chemische en biologische (niet-vis) gegevens voor opname in EMODNET als data-aggregator.

Deze standaarden (AQUO, SDN-BODC, ICES) zijn op elkaar af te stemmen door middel van vertaaltabellen (mapping). Deze mappings zijn al gedeeltelijk voorhanden. Er is een minimale inspanning nog om deze bruikbaar te maken voor MONS.

- In het verleden is een begin gemaakt met de mapping tussen AQUO en SDN/BODC. Deze wordt binnenkort aangevuld in een RWS-VNSC project. De voorbereidingen hiervoor zijn in gang gezet en er vindt overleg plaats om voor binnen 1 jaar een voor MONS bruikbare mapping ter beschikking te hebben.
- Binnen het Europese EMODNET initiatief zijn er projecten opgezet om een koppeling gemaakt met ICES. Dit betekent dat het in theorie mogelijk is om via de SDN-BODC

standaard ook ICES data volgens de AQUO standaard te koppelen. Dit zal zo spoedig mogelijk worden onderzocht.

In de eerste opzet is voorzien dat Data worden geaccepteerd zodra deze gestandaardiseerd zijn volgens de volgende semantische standaarden:

1. AQUO (verplichte open data standaard volgens 'pas toe of leg uit principe' binnen de NL water sector (sensu lato)
2. ICES voor Europese vis-ecologische en visserij gegevens.
3. SeaDataNet – BODC standaard voor de Europese ontsluiting van fysisch-chemische en (niet-vis) biologische data
4. De facto (as is) gestandaardiseerde data

Er zal worden onderzocht of er een service kan worden opgericht om de facto gestandaardiseerde datasets kunnen worden ontsloten via een ETL⁴ service naar RDF. Hiervoor wordt een haalbaarheidsstudie uitgevoerd.

4.1.3 Validatie

Vóór het ophalen van datasets zal worden gecontroleerd of deze op de juiste manier gestandaardiseerd (AQUO, SDN-BODC) en FAIR zijn, en voldoen aan de specificatie die vastgelegd zijn in de Data Factsheets. Bij succesvolle validatie wordt de dataset geaccepteerd als deel - of eindproduct (deliverable) en zal dit worden teruggekoppeld met het UB. Deze kan dan overgaan tot uitbetaling (zie sectie 2.2).

Er zijn verschillende applicaties beschikbaar (bij IHM, RWS, IHW) die kunnen worden aangepast voor de Validatie. De keuze van de applicaties kan pas worden gemaakt als er een beslissing wordt genomen over de wijze van ontsluiting van de MONS data.

4.1.4 Organisatorische afspraken

Hierbij wordt de opzet van het Protocol Mariene Data van het IHM gevolgd (Bijlage 2). Er worden afspraken gemaakt over de taken bij de verschillende belanghebbenden, zoals welke partij de data beheert, online zet (service verleners), toegang tot de data verleent, archiveert etc. De gehele levenscyclus van de data wordt hierbij betracht.

Ook wordt de inbedding van de Data Factsheet in de projectorganisatie (governance) beschreven.

4.1.5 Begrippen

Voor eenduidige interpretatie van het Dataprotocol worden gebruikte begrippen gedefinieerd. Hierbij wordt zoveel mogelijk de taal van het Raamwerk Kwaliteitsstelsel en INSPIRE gebruikt die passen binnen de verschillende ISO normen. Voor de vertaling van de begrippen zal worden aangesloten bij AQUO. Indien hiervan wordt afgeweken, wordt de bron vermeld.

⁴ ETL: Extraction, Transformation Load tool

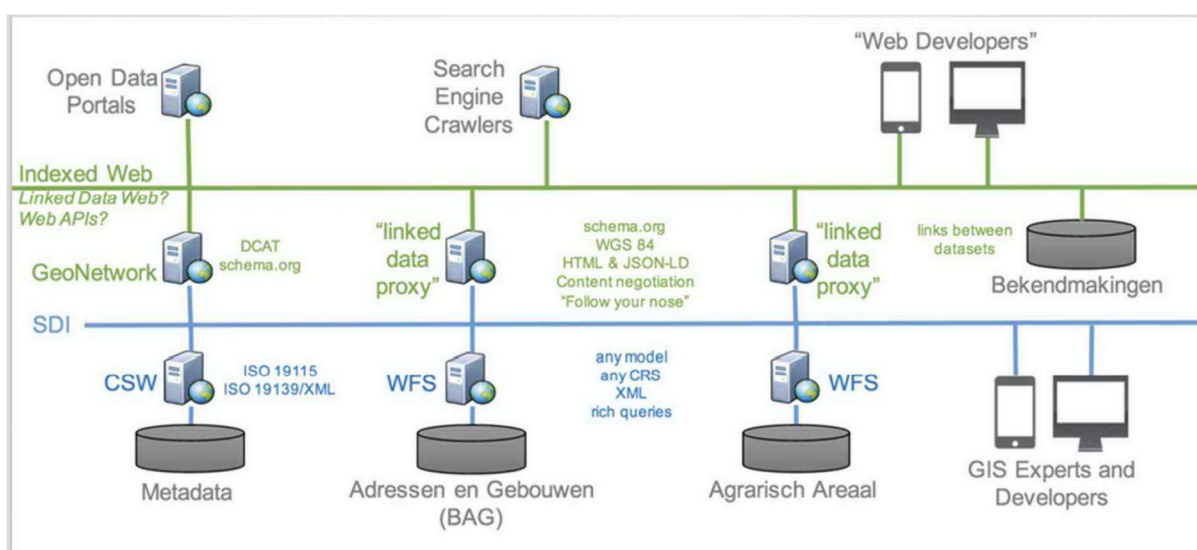
4.2 Vereiste data-infrastructuur voor de ontsluiting van MONS data

Voor de FAIR ontsluiting van MONS data, waarvoor regels en afspraken worden vastgelegd in het Dataprotocol is een data-infrastructuur nodig. Het is cruciaal dat deze goed op elkaar zijn afgestemd. Daarom wordt voorgesteld om deze parallel aan elkaar te ontwikkelen.

4.2.1 Uitwisseling van (meta)data

Voor de technische uitwisseling van de data wordt voorgesteld om gebruik te maken van:

1. de OGC Standaard zoals nu gebruikelijk is met WMS, WFS services voor het delen van geodata en WCS voor metadata.
2. Het RDF⁵ datamodel (Linked Data) een relatief nieuwe methode die data kan uitwisselen in verschillende data en metadata standaarden (APIs)



Figuur 2. Schematische voorstelling van de combinatie van OGC services en Linked data (RDF) voor het delen van Geo Informatie (Bron presentatie OGC vs Open/Linked data, GeoCat)

De OGC standaard wordt bij IHM momenteel succesvol gebruikt als technische standaard (service) voor het uitwisselen van Geodata. De combinatie met het RDF data model maakt het mogelijk om Linked Data hieraan te koppelen. Linked Data zijn gestructureerde gegevens die gelinkt zijn aan andere gegevens en daardoor beter bruikbaar zijn in semantische queries. De methode is gebaseerd op de techniek van HTTP-URI's en RDF. Linked Data kunnen worden gelezen door mensen via internetpagina's en geautomatiseerd door computers (bron: Wikipedia).

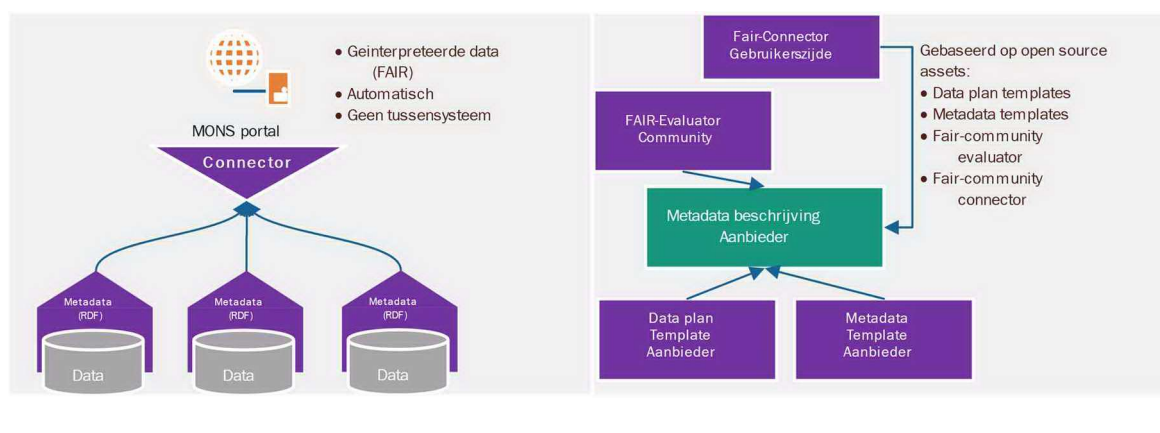
Een schematische voorstelling van de koppeling is weergegeven in Figuur 4. Een korte omschrijving van de toepassing van RDF binnen MONS is weergegeven in Box 2.

⁵ Resource Description Framework, zie bijlage voor meer informatie.

Box 2. Korte omschrijving van de toepassing van RDF voor de FAIR ontsluiting van datasets volgens een de facto standaard.

Door een extern bedrijf wordt bij dataleveranciers een applicatie geïnstalleerd om datasets (in een de facto (eigen) standaard) om te zetten naar het RDF datamodel (Resource Description Framework, Figuur 4 **Fout! Verwijzingsbron niet gevonden.**) met daarin een ingebouwde 'automatische' omzetting naar een datastandaard zoals bijvoorbeeld AQUO (via Data-templates). Metadata worden geïntegreerd toegevoegd en kunnen op soortgelijke wijze als de data omgezet worden naar een andere standaard (via Metadata templates).

Het uitwisselen van de RDF data gebeurt met een generieke API (FAIR-connector). Ook is er een validatie-tool ontwikkeld die kan toetsen of de dataset FAIR is ontsloten (FAIR-evaluator. Voor meer informatie: <https://www.go-fair.org/fair-principles/>)



Voor de ontwikkeling van de service wordt gebruik gemaakt van INSPIRE guidelines ([Guidelines for the RDF encoding of spatial data \(inspire-eu-rdf.github.io\)](https://inspire-eu-rdf.github.io/)).

De toepassing van deze structuur zal de inspanning voor het gestandaardiseerd ontsluiten van data door dataleveranciers aanzienlijk vereenvoudigen en versnellen. Door het opnemen van verschillende datastandaarden in de data-template van de RDF zouden data 'on the go' in verschillende datastandaarden uitgewisseld kunnen worden. Bijvoorbeeld in de AQUO standaard voor nationale toepassingen en de SeaDataNet/BODC standaard voor internationale aansluiting met data-aggregatoren zoals EMODNET, (Eur)Obis, GBIF en LifeWatch.

RDF is reeds succesvol toegepast in een Pilot van IHW over de FAIR ontsluiting van Data (Pilot Droogte en wateroverlast FAIR-Data, <https://www.ihw.nl/op-naar-een-landelijke-datasamenwerking/>).

Om de haalbaarheid van de combinatie van OGC en Linked Data binnen MONS te bestuderen is door IHM in november 2021 een Pilot Haalbaarheidsstudie uitgezet, gefinancierd door LNV. De resultaten worden in maart verwacht. In deze studie zal het gebruik RDF en Linked Data voor de ontsluiting en uitwisseling van MONS data worden geëvalueerd.

Vertegenwoordigers van het projectconsortium zullen worden betrokken bij de besluitvorming over de haalbaarheid van het datamodel en de ontplooiing daarvan binnen MONS bij een positief besluit. Indien, op basis van deze studie, deze combinatie niet

haalbaar wordt geacht kan er teruggevallen worden op uitwisseling van data door middel van alleen de OGC services.

De uitwisseling van data wordt aangestuurd via API's. Hierbij zal worden aangesloten bij de Digitale Delta API familie. Voor de 'standaard' uitwisseling van data zullen scripts ter beschikking worden gesteld op het MONS Data Portaal (via registers en de Open Data Viewer).

4.2.2 Metadata

De kwaliteit van een dataset is onder andere afhankelijk van de kwaliteit van de Metadata beschrijvingen. Hiervoor zal een gebruiksvriendelijke service worden ingericht door integratie van informatie die gedurende verschillende stappen tijdens de oplevering van data wordt vertrekt.

De Data Factsheet zal zodanig worden vormgegeven dat de verstrekte informatie ook kan worden gebruikt voor de Metadata beschrijving. Aanvullende Metadata van Geodata worden via een interface geïmporteerd vanuit de WMS en WFS. Metadata worden zo gestructureerd dat vanuit een zelfde set aan gegevens de Metadata tussen verschillende standaarden kunnen worden uitgewisseld (RDF).

Voor de opslag van Metadata dient op zijn minst tijdelijk een Catalogue Service for the Web (CSW) server ingericht te worden. Hierdoor wordt een Geonetwerk gecreëerd en kunnen Metadata aan de kaarten en gegevens worden gekoppeld. De CSW server dient ook voor de aansturing van de Geoviewer.

4.2.3 MONS Data Portaal

Er is een portaal nodig voor de presentatie van de beschikbare data: zowel voor de data die binnen MONS gegenereerd zijn als reeds beschikbare data die betrekking hebben op de Noordzee. Deze kan voortbouwen op de reeds voorhanden zijnde faciliteiten bij het IHM (Open Data Viewer). Gedurende het project zal aansluiting worden gezocht met andere portalen zoals Emodnet, ICES, OSPAR en Nationaal Georegister.

De portaalfunctie wordt in samenspraak met het UB opgebouwd. Het is de bedoeling dat er een register komt met beschikbare data met daaraan gekoppeld de API code en/of OGC url die nodig is voor de uitwisseling van de data.

4.2.4 Serviceproviders voor ontsluiting van projectdata

De opdrachtgevers zijn verantwoordelijk voor de ontsluiting van data. Dat wil niet zeggen dat ze de verschillende deelprocessen in eigen beheer moeten/hoeven uit te voeren. Dit kan ook worden uitbesteed aan derden. Afspraken hierover worden vastgelegd in de Data Factsheet.

Opdrachtgevers hebben hiervoor meestal al afspraken gemaakt of in hun eigen organisatie belegd (eigendom, beheer, opslag). Voor data die niet bij de opdrachtgever of opdrachtnemer kunnen worden beheerd zal een MONS geoserver worden opgetuigd, beheerd door Deltares. Vergoeding van de kosten gebeurt op basis van nacalculatie.

4.2.5 Data Life Cycle Management: Archivering van data

Bij het opzetten van het Dataprotocol dient de gehele levenscyclus van Data in acht te worden genomen.



Figuur 3. Life Cycle Management (bron: <https://www.ontrack.com/en-gb/blog/how-to-protect-your-data-through-its-lifecycle-part-two>)

Alhoewel aspecten van de archivering van data voldoende aandacht zullen krijgen gedurende de loop van het project (back-upprocedures, veilige opslag), dient de echt langdurige opslag (decennia) en het ter beschikking stellen van deze data ook geregeld te worden. In samenspraak met het MONS consortium (aansturing, uitvoerende partijen en direct belanghebbenden) wordt hiervoor gedurende het project een strategie ontwikkeld. Deze ambitie wordt vastgelegd in het MONS-IHM Dataprotocol.

4.2.6 Data publicatie

Een Wetenschapper wordt in de regel afgerekend op de citatie-index van zijn publicaties. Het is nu gebruikelijk dat ook onderliggende datasets gepubliceerd kunnen (of zelfs moeten) worden. Ook hiervoor zijn citatie-indices.

Een voorwaarde hiervoor is een unieke en persistente link naar de dataset (DOI). In overleg met de dataleveranciers en onderzoekers zal een service gekozen of ingericht worden.

In samenspraak met het MONS consortium (aansturing, uitvoerende partijen en direct belanghebbenden) wordt hiervoor gedurende het project een strategie ontwikkeld. De ambitie wordt vastgelegd in het Dataprotocol.

4.3 Betrokken partijen

Het MONS Dataprotocol wordt zoveel mogelijk gebaseerd op reeds bestaande afspraken, principes, diensten, methodes en standaarden die in de praktijk hun effectiviteit hebben aangetoond. De inhoudelijke experts van deze praktijken zullen worden geconsulteerd voor afstemming en fine-tuning van de verschillende componenten van het datamanagementbeheer en het Dataprotocol.

De belangrijkste partijen die zijn betrokken bij de opstelling van het Dataprotocol:

MONS Uitvoeringsbureau (RWS-WVL).

Integratie van datamanagement in de aansturing van het Programma: MONS

Vormgeving van de Data Factsheet: WOZEP (RWS-WVL/ Witteveen en BOS)

IHW / AQUO

Verschillende services die nodig zijn voor het data-managementbeheer binnen MONS zijn ontwikkeld door IHW/AQUO. Deze dienen op maat gemaakt te worden voor MONS. Hiervoor zal de samenwerking met IHW worden geïntensiveerd.

Consortium mapping AQUO-SDN/BODC standaarden.

In samenwerking met RWS, VLIZ, Deltares en VNSC (Vlaams Nederlandse Schelde Commissie) is in 2021 een update van de AQUO-SDN/BODC mapping geïnitieerd. Deze is mapping is van belang voor de interoperabiliteit van datasets die zijn gestandaardiseerd volgens een van deze standaarden.

Deltares

Voor het datamanagement van MONS zal een lichte data infrastructuur nodig zijn. Hierbij kan worden voortgebouwd op de infrastructuur die Deltares voor RWS projectdata in beheer heeft. Er dienen een aantal nieuwe componenten opgezet te worden:

1. Een portaal met inzicht en toegang tot de beschikbare data (register) die door het programma gegenereerd worden
2. Een portaal met beschikbare externe data die van belang kunnen zijn
3. Een viewer om data met een geo-component te presenteren.
4. Een OGC-CSW server voor de (tijdelijke) opslag van Metadata.

Daarnaast dient te worden onderzocht of bestaande services opgeschaald kunnen worden te ondersteuning van het MONS programma, zoals de MONS Geoserver voor de (tijdelijke) opslag van projectdata.

IHM werkt al jaren samen met Deltares op dit gebied. Deltares is bij veel projecten betrokken voor het datamanagement en werkt al jaren samen met IHM aan de ontsluiting van mariene geodata. IHM wil graag de opgedane expertise blijven aanspreken voor de opzet van het datamanagementbeheer binnen MONS.

WOZEP (RWS-WVL/ Witteveen en BOS)

Voor de implementatie van het gebruik van Data Factsheets wordt de expertise die hiermee is opgedaan binnen WOZEP gebruikt. Hiervoor is overleg met WOZEP nodig.

GEONOVUM en Nationaal Georegister

De uitrol van RDF als datamodel voor de ontsluiting van Data zal indien mogelijk worden uitgevoerd volgens de INSPIRE Guidelines. Hiervoor zal afstemming plaatsvinden met GEONOVUM.

Een strategie voor de beschrijving en internationale inbedding van de MONS metadata zal in overleg met GEONOVUM en Nationaal Georegister worden ontwikkeld

Klankbordgroep operationeel (Opdrachtnemers)

Om het Dataprotocol gedragen te laten worden door dataleveranciers, dienen deze zo vroeg mogelijk bij de opzet van het Dataprotocol te worden betrokken.

Binnen MONS worden projecten in principe openbaar aanbesteedt en is er geen zicht op wie dataleveranciers worden. Een oplossing hiervoor is het installeren van een klankbordgroep voor het Dataprotocol die breed aangekondigd wordt.

Tijdens de overleggen wordt gezocht naar Best-Practices van opdrachtnemers die de ontsluiting van data verbeteren of makkelijker maken.

Klankbordgroep regie ontsluiting van mariene data

Het Noordzee Akkoord biedt een kans om meer regie aan te brengen in de nationale en internationale FAIR ontsluiting van gestandaardiseerde mariene data. Om dit te bereiken wordt een klankbordgroep opgericht om daar vanuit Nederlands perspectief naar toe te werken.

Klankbordroep moet aansluiting zoeken met inhoudelijke experts van de CIO's en CDO van relevante Nederlandse en Internationale Partijen.

Zoals:

- Departementen (BZK; IenW, LNV, Defensie, EZK)
- Wetenschappelijke sector NWO, KNAW
- Kennisinstituten
- Geonovum/GI beraad (in het kader van het SDI gebeuren binnen NL)
- Tactische beraad EU Informatie
- UN-GGIM
- UNESCO-IOC
- KRM-DIKE
- INSPIRE

4.4 Risicofactoren die kunnen leiden tot vertraging van de oplevering van het Dataprotocol

4.4.1 Bepalende stap: Governance

Een stringente voorwaarde voor effectief datamanagementbeheer is de inbedding in de projectaansturing door het Uitvoeringsbureau (Figuur 1). Dit kan op een organisatorisch en administratief zeer lichte wijze uitgevoerd. Voorwaardes hierbij zijn dat projecten uitgezet worden met gebruikmaking van de ARVODI voorwaarden (zie sectie 2.2) en het MONS Dataprotocol samen met de Data Factsheet opgenomen kunnen worden als (geparafeerde) bijlage. Hierdoor worden deze onderdeel van de Overeenkomst en kan hierop gestuurd worden bij afwijking van de specificaties (alleen uitbetaling bij succesvolle validatie van de aangeleverde data en Metadata).

Ingeschat risico: aanwezig

Gevolgen: groot voor wat betreft mogelijkheden en efficiëntie;

Alternatief: geen. Zonder inbedding in de MONS Governance is er geen ander instrument beschikbaar om de standaardisatie van data dwingend aan te sturen. De consequentie is dan dat het Dataprotocol haar officiële status verliest en het karakter krijgt van een niet bindend Memorandum of Understanding.

Conclusie: projectmanagement kan dataontsluiting in gevaar brengen. Van belang dat dit goed wordt ingebed.

4.4.2 Bepalende stap: de ontwikkeling van de data-infrastructuur:

Is het gebruik van het RDF datamodel mogelijk binnen MONS?

Dit wordt binnenkort onderzocht in een haalbaarheidsstudie die door IHM samen met LNV wordt uitgezet zodat het antwoord op deze vraag vroeg in het proces beschikbaar komt (gepland in maart 2022, zie paragraaf 4.2.2).

Mocht uit de Pilot studie blijken dat deze aanpak niet haalbaar is dan zal dit niet leiden tot vertraging van de oplevering van het Dataprotocol. Het gebruik van RDF en Linked Data is een innovatieve toevoeging op de reeds bestaande conventionele manieren van data ontsluiting.

Dit heeft echter wel bepalende gevolgen voor de verdere invulling van de infrastructuur (geen gebruik van Linked Data) en service verlening voor de ontsluiting van data bij de Bronhouder. Als Linked Data niet inpasbaar zijn betekent dit dat er op termijn geïnvesteerd dient te worden in innovatie, en er een grotere inspanning (dus kosten) nodig is voor de ontsluiting van de MONS data.

Ingeschat risico: klein

Gevolgen: groot voor wat betreft mogelijkheden en efficiëntie;

Alternatief: conventionele reeds bestaande aanpak van data- ontsluiting

Conclusie: primaire proces van ontsluiting van data komt niet in gevaar; vertraging mogelijk

4.4.3 Kritisch proces: de mapping van AQUO-SDN/BODC standaarden.

Dit proces is essentieel voor de nationale (AQUO) en internationale (SDN/BODC) aansluiting (interoperabiliteitsissue) van MONS data en neemt voor dataleveranciers de spanning weg tussen een keuze in deze semantische standaarden.

Vertraging in dit externe proces leidt tot vertraging in de FAIR ontsluiting (en interoperabiliteit) van MONS Data.

Het is echter belangrijk dat dit proces op relatief korte termijn bruikbare resultaten oplevert. De gevoeligheid is dat dit proces buiten MONS wordt uitgevoerd en er geen directe aansturing vanuit MONS governance mogelijk is. Wel zijn er mogelijkheden voor monitoring van de voortgang via participatie van IHM in het project.

Een mogelijke oplossing voor het voorkomen van vertraging is door het bijdragen aan de financiering van de mapping.

Ingeschat risico: aanwezig

Gevolgen: vertraging van de ontsluiting van data.; (tijdelijke) beperking van de interoperabiliteit.

Alternatief: financiële ondersteuning voor de mapping van de twee standaarden

Conclusie: primaire proces van ontsluiting komt niet in gevaar; maar wordt vertraagd

4.4.4 Belangrijk proces: uitwisseling van data (Peer to Peer)

Bij de huidige opzet van het datamanagement (beheer) wordt er van uitgegaan dat data bij de bron blijven en daar FAIR ontsloten worden. Het Mons Portaal faciliteert in het peer-to-peer uitwisselen van data. Een voordeel is eenvoudig beheer uitgevoerd door bronhouders. Bij grote datastromen kan dit leiden tot vertraging in de levering. Er is nog geen zicht in het volume van de datastromen en het is niet mogelijk om daarop te anticiperen. Er zijn verschillende manieren om dit op te lossen:

1. Koop grotere capaciteit in voor de lokale servers.
2. Maak gebruik van een client-server approach waarbij het Portaal een functie krijgt als cache van de data die op de lokale servers staat.

Optie 1 is een simpele pragmatische oplossing. Optie 2 zorgt voor een centraal aangestuurde garantie voor de performance met betrekking tot de uitwisseling van (meta)Data.

In het Dataprotocol en de Data Factsheet kunnen eisen gesteld worden aan de performance van de servers. Bronhouders dienen dan de capaciteit van de servers hierop in te richten.

Voor optie 2 dient een centrale server zorgvuldig ingericht te worden met bijvoorbeeld afspraken over toegang (licentiebeheer) en spelen licenties, en verversing van de cache (om verouderde kopieën van een dataset te voorkomen⁶). De voorgenomen strategie voor de oplossing van dit probleem dient te worden overlegd met het UB.

Ingeschat risico: aanwezig

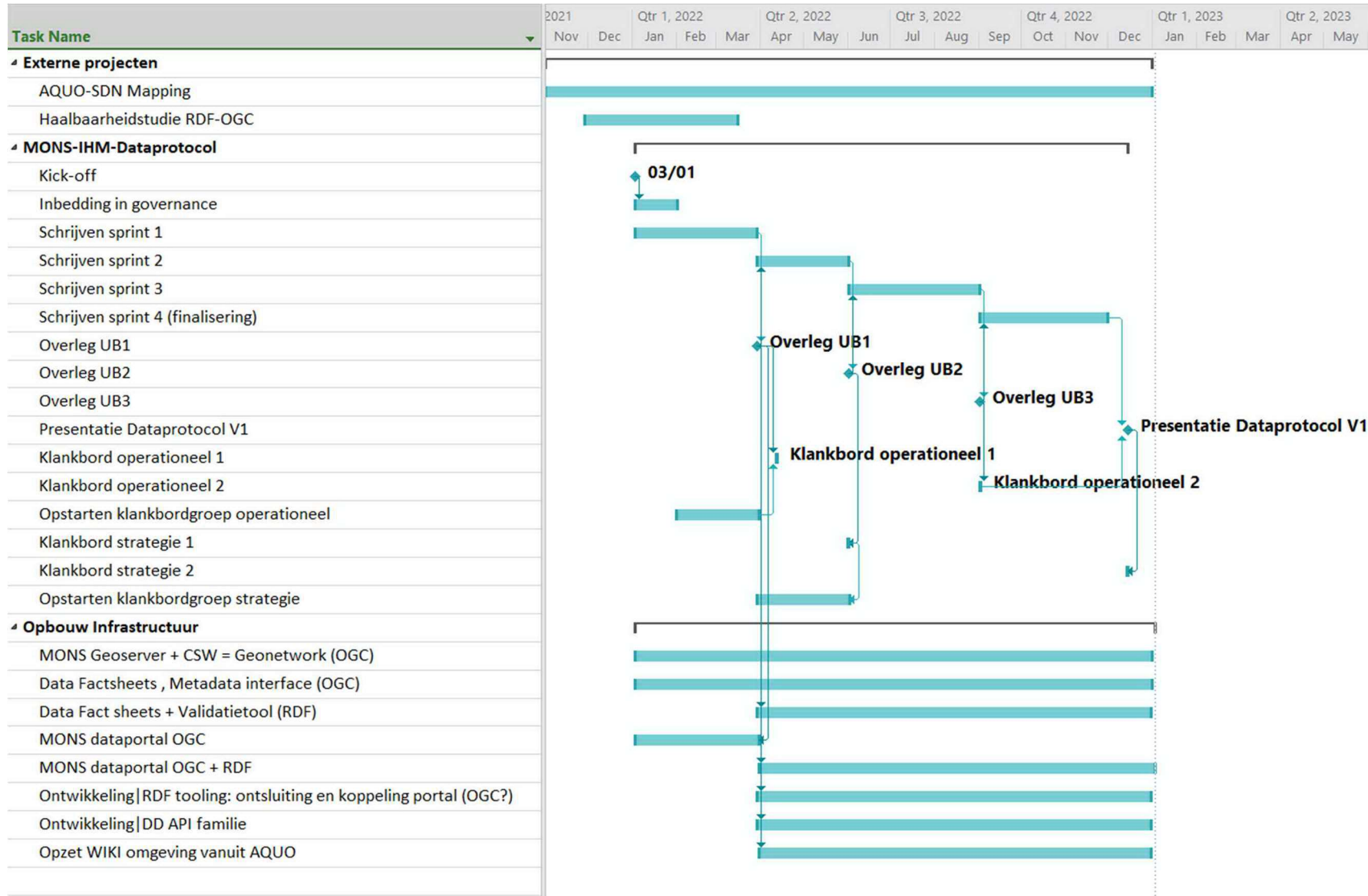
Gevolgen: vertraging van de ontsluiting van data; (tijdelijke) beperking van de interoperabiliteit.

Alternatief: uitbreiding capaciteit lokale servers of Client-server approach met een cache

Conclusie: primaire proces van ontsluiting komt niet in gevaar; maar wordt vertraagd

⁶ Dataset is geupdated, maar de oude versie staat nog in de cache en wordt uitgewisseld

4.5 Planning van de activiteiten (Gantt chart)

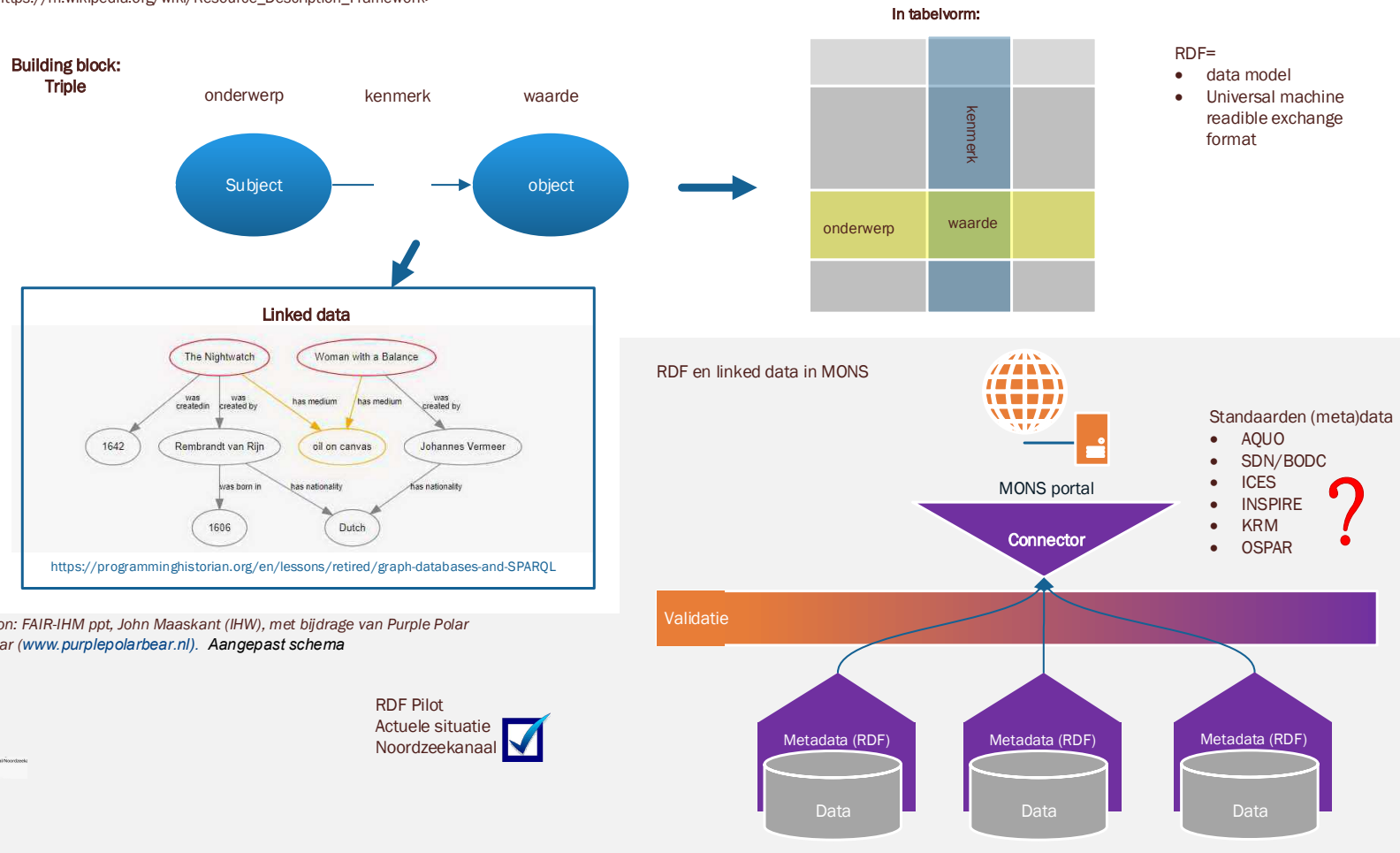


Resource Description Framework

Resource Description Framework of RDF is a standard of the World Wide Web Consortium (W3C), originally designed as a metadata model, but increasingly used as a format to exchange data in general.

With the RDF model, statements can be made about the characteristics of resources on the web (resources) in the form of a three-part subject-predicate-object structure (in RDF terms a triple). The subject is in essence the resource being described. The predicate is which characteristic or aspect of that source is being described. The object is what the value of that characteristic is.

<https://nl.wikipedia.org/wiki/Resource_Description_Framework>



Figuur 4. Schematische voorstelling van de toepassing van RDF in MONS

5 BIJLAGEN

Bijlage 1. MONS_137 Projectbeschrijving Datamanagement.....	19
Bijlage 2. Protocol Mariene Data IHM	22
Bijlage 3. Schets van het MONS Dataprotocol	23

MONS ID137

Datamanagement

1. Inleiding; MONS en het NZA

Een gezonde Noordzee is voor iedereen van belang. Nu al vraagt het mariene ecosysteem om bescherming en herstel op basis van bestaand gebruik. Het veranderende gebruik (in vorm en intensiteit) moet passen binnen de ecologische draagkracht van de Noordzee. De ecologische draagkracht is randvoorwaardelijk voor het individuele en cumulatieve gebruik van de Noordzee door verschillende functies. Om de Noordzee gezond te maken en te houden is een extra inspanning noodzakelijk.

Beleidsontwikkeling voor bescherming, herstel en duurzaam gebruik wordt bemoeilijkt door een structureel gebrek aan kennis. Dit geldt zeker voor de soorten die het meest gevoelig zijn voor de grote transitie op de Noordzee. Deze soorten zijn daarmee indicatoren voor de te verwachte majeure veranderingen die in en op de Noordzee plaats gaan vinden. Het Noordzee Akkoord (NZA) wil de uitdagingen van een veranderend gebruik oppakken en een nieuwe balans vinden. Het NZA schetst een grote behoefte aan een integraal en systematisch onderzoeks- en monitoringsprogramma dat de basis vormt voor kennis over het functioneren van de Noordzee. Het programma Monitoring-Onderzoek-Natuurherstel-Soortbescherming (MONS) heeft als doel de centrale vraag te beantwoorden hoe het veranderende gebruik van de Noordzee past binnen de ecologische draagkracht van de Noordzee.

Daartoe is een integraal en systematisch monitoringsprogramma nodig dat zich richt op de fysische, chemische en biologische basisparameters voor het functioneren van het ecosysteem en op (de variatie in) het voorkomen van vogels, vleermuizen, bodemdieren, vissen en zeezoogdieren.

Het MONS-programma geeft een eerste uitwerking van het onderzoek dat de komende tien jaar zal worden uitgevoerd om de kennisvragen zoals die geformuleerd zijn in het NZA te kunnen beantwoorden.

De uitwerking van dit programma verloopt in fasen. In de eerste fase is een lijst met 'no regret onderzoeksprojecten' in kaart gebracht waarvoor het onderzoek al op korte termijn kan worden uitgevoerd, gestart en al deels afgerond voor het einde van het jaar 2021. Het hier beschreven project is een van die 'no regret' projecten, zoals vastgesteld in het NZO op 9 juni 2021.

2. Datamanagement, visualisatie en communicatie

Draagkracht, effecten van drukfactoren, en mogelijkheden voor natuurherstel en soortenbescherming zijn belangrijke NZA-onderwerpen waarvoor het gebruik van innovatieve monitoring- en meettechnieken van groot belang is. Het MONS-programma zal veel data opleveren. Goed beheer van de MONS-data draagt bij aan een goede opstelling en uitvoering van het Noordzee Akkoord. Kwalitatief goede data die gemakkelijk toegankelijk zijn voor eenieder vergroot ook het vertrouwen van NZA-stakeholders in opstelling en uitvoering van het NZA.

In het MONS-programma gaan veel data verzameld worden, door een groot aantal partijen. Daarnaast zal veel onderzoek plaatsvinden op basis van nieuw verzamelde en bestaande data, wederom door een veelheid aan partijen. Deze dataverzameling en het daaruit voortvloeiende onderzoek faciliteert uiteindelijk een goed beheer van de Noordzee. In het NZA is afgesproken dat alle data en onderzoeksresultaten volgend uit het Noordzee Akkoord worden ontsloten volgens het 'openbaar tenzij' principe. Het

datamanagementbeheer zal zo open mogelijk tot stand worden gebracht, maar de rechten van de opstellers van de gegevens dienen te worden beschermd. Dit houdt in dat de nodige aandacht aan de kwaliteit van de gegevens en de traceerbaarheid van de bronnen moet worden geschonken.

Een efficiënte organisatie van deze datastromen en het databeheer is een voorwaarde voor het gebruik ervan voor de in het NZA beschreven transitie. Zo'n organisatie kan ook zorg dragen voor de geautomatiseerde opwerking tot 'dashboards' waar data wordt opgewerkt tot bruikbare informatie voor het beheer van de Noordzee. Voor het maximaal benutten van de binnen MONS verzamelde en gebruikte gegevens voor inzicht en beheer is regie noodzakelijk.

De afgelopen jaren heeft het Informatiehuis Marien (IHM) een sterke positie verworven in het bundelen, regisseren en beschikbaar maken van data van het Nederlandse mariene milieu. Ditzelfde concept kan worden toegepast op het NZA door middel van een structurele versterking van het IHM. Daarmee kan het IHM uitgroeien tot een gezaghebbend nationaal platform voor het (FAIR) beschikbaar maken en houden van mariene data. Het IHM kan dan, in overleg met de NZA-stakeholders, een infrastructuur opzetten voor de binnen MONS te verzamelen en te gebruiken data, inclusief afspraken over relevante standaarden en formats en ontsluiting. Vervolgens moet deze infrastructuur worden onderhouden en moeten alle stakeholders aan de afspraken worden gehouden.

Naast de regie-kant door IHM is er ook inspanning vereist door de dataleveranciers. In de huidige werkwijze van het IHM wordt data preferentieel bij de bron (meestal de verzamelende instantie) opgeslagen. Dit betekent dat ook aan de bron-kant infrastructuur moet worden opgezet, de overeengekomen standaarden en formats moeten worden geïmplementeerd, en de interface met het IHM moet worden onderhouden. Soms is beheer door de bron onmogelijk of onwenselijk. In zulke situaties moet dataopslag zo worden georganiseerd dat ook na afloop van het project databases en meta-data goed bewaard en ontsloten blijven. Dit zal veelal betekenen dat IHM de data zelf bewaart.

Voor het opstellen van het MONS-IMH-Dataprotocol wordt de volgende aanpak gevolgd. Bij het opzetten van onderzoekstrajecten en het uitzetten van opdrachten worden duidelijke afspraken gemaakt over datamanagement. Dit kan worden aangestuurd vanuit het IHM, mits daarvoor structurele versterking wordt belegd vanuit het Noordzee Akkoord. Binnen één jaar na aanvang van de onderzoekstrajecten in het kader van het Noordzee Akkoord zullen databeheerders en –managers gezamenlijk een Dataprotocol moeten opstellen met afspraken over datastandaardisatie, kwaliteitscontrole, dataopslag en ontsluiting. Afronding van de verschillende trajecten zal het databeheer en uitrol van ontsluiting betekenen. Dit speelt voor elk project met name in de opdracht- fase en bij het beschikbaar komen van data, en is dus gedurende de hele looptijd van MONS (maar feitelijk ook daarna) relevant.

Centrale kennisvraag:

Een infrastructuur opzetten voor de binnen MONS te verzamelen en te gebruiken data, inclusief afspraken over relevante standaarden en formats en ontsluiting.

3. Plan van aanpak opstelling MONS-IMH Dataprotocol

Het MONS-project 137 omvat drie onderdelen: datamanagement, visualisatie en communicatie. De inhoud van dit project wordt beschreven in paragraaf 4.10 van het MONS-programma. Als no regret project is er voor gekozen om hiervan het onderdeel datamanagement op te pakken.

De werkwijze om te komen tot een goed beheer van de MONS-data is als volgt. Er zal door het Informatiehuis Marien (IHM) een Dataprotocol worden ontwikkeld dat nodig is om ervoor te zorgen dat het datamanagement voor al het MONS-onderzoek en -monitoring vanaf 2022 op orde is. Dit Dataprotocol bevat de instructies t.a.v. datamanagement die gelden bij door MONS uitgevoerd onderzoek (zie hierboven).

Als voorbereiding op de opstelling van het MONS-IMH-Dataprotocol uiterlijk eind 2022 wordt hier als no regret project gestart met het in kaart brengen welke onderwerpen zouden moeten worden opgenomen in het Dataprotocol. Hierbij wordt gebruik gemaakt van bovenstaande informatie, de informatie in paragraaf 4.10 van het MONS-programma en van onderliggende informatie die is gebruikt voor de opstelling van paragraaf 4.10. Verder wordt aangegeven welke partijen nodig zijn voor de opstelling van het protocol. Ook wordt aangegeven welke factoren een rol kunnen spelen bij een ongewenste vertraging bij de opstelling van het protocol en worden mogelijkheden beschreven hoe vertragingen tijdig kunnen worden voorkomen. Tot slot wordt een indicatie gegeven van de kosten voor opstelling van het protocol.

Het resultaat van de uitvoering van dit project is een plan van aanpak dat als vertrekpunt zal dienen voor het opstellen van het MONS-IMH-Dataprotocol.

4. Tijdpad en samenhang met andere (MONS) projecten

Het resultaat dient voor 1 november te worden aangeboden aan de opdrachtgever, i.c. het ministerie van I&W. Contactpersonen bij I&W zijn Guido Hommel en Ricardo van Dijk

Bijlage 2. Protocol Mariene Data IHM

Het Protocol Mariene Data is te downloaden via:

https://www.informatiehuismarien.nl/publish/pages/109348/protocol_november_2013_1805.pdf

Bijlage 3. Schets van het MONS Dataprotocol

Het Dataprotocol wordt in de Engelse taal opgesteld

Inleiding

In de inleiding worden (algemene) principes van data ontsluiting en datamanagement geïntroduceerd en het Dataprotocol nationaal en internationaal ingekaderd. De volgende topics worden in dit hoofdstuk behandeld.

- De ontsluiting van Open Data en datamanagement in nationale en internationale context / Visie op Data ontsluiting
- Doel van het Dataprotocol
- De omgeving
- Data
 - Datatypes
 - Observations and Measurements
 - Rights, Restrictions and Responsibilities
 - Modeling
 - Algorithms
 - Metadata?
 - Levenscyclus van data
 - Data archivering onder de aandacht brengen, beleggen bij dataeigenaar, databeheerder of service verlener.
- Rollen, taken en verantwoordelijkheden (Tasks and responsibilities)
 - Beschrijving van de rollen en verantwoordelijkheden van actoren bij de ontsluiting van Data binnen MONS (volgens Protocol Mariene Data, Bijlage 2).
 - Het beleggen van het duurzaam beheer van het Dataprotocol en bijbehorende services en infrastructuur bij partij(en).

Standaardisatie

- Data
 - Data dienen gestandaardiseerd aangeleverd te worden volgens vooraf vastgelegde specificaties van de Data Factsheets)
 - Geaccepteerde technische standaarden:
 - Geaccepteerde semantische datastandaarden: AQUO; SDN-BODC
 - Vóór de oplevering worden data gevalideerd.
 - Alleen gegevens die aan de validatieregels voldoen, worden geaccordeerd als (deel)product.
 - RDF data-templates worden beheerd door AQUO
- Metadata
 - Dataleveranciers verplichten zich tot het leveren van rijke metadata bij de datasets (FAIR) als onderdeel van de Data Factsheet. Een template hiervoor wordt leidend. Er komt een minimum lijst, gebaseerd op het Nederlandse Metadata profiel voor data en services⁷, met aanvullingen die nodig zijn voor verschillende andere standaarden (AQUO, SDN,INSPIRE)
 - Nederlandse metadataprofiel(en) beheerd door GEONOVUM vormen de basis voor het opstellen van het metadata-template.
 - De metadata maken deel uit van de validatie.

⁷ <https://www.geonovum.nl/geo-standaarden/metadata>

- RDF Metadata template(s) worden beheerd door AQUO

Metadata gegevens worden in eerste instantie opgeslagen op een eigen CSW server. Er wordt zo spoedig mogelijk onderzocht of een link te leggen is met het Nationaal Georegister (PDOK). Er wordt een gebruikersvriendelijke service opgezet voor de invoer van metadata. Zo worden gegevens geëxtraheerd van de Data Factsheet en indien mogelijk Geoservers.

Er wordt een service opgezet om (meta)data in verschillende standaarden uit te wisselen (bijv. AQUO, SDN, INSPIRE, KRM, OSPAR)
- Internationale standaarden and guidelines:
 - Inbedding van de gebruikte datastandaarden (AQUO, SDN) in internationale context: ISO 9001 (processes), ISO 17001 (ICT, security), ISO 19115 (Metadata), NEN 3610 semantiek NL, INSPIRE
- Interoperability
 - De eenduidige uitwisseling van data wordt gefaciliteerd door het gebruik van datastandaarden en uitwisselservices.
 - OGC als basis
 - REST API / SOAPS (Digitale Delta API?)
 - Harmonised Data Transport Formats (SDN, AQUO)
 - INSPIRE richtlijnen voor serviceverlening zijn leidend.
- Provenance
 - De links naar data en metadata dienen eenduidig en persistent te zijn (FAIR).

De actoren die een rol spelen bij de verschillende processen van de data leverantie en beheer zullen worden geregistreerd aan de hand van de Data Factsheets. Hierbij dient de AVG in acht te worden genomen.
 - Als een van de revenuen voor dataleveranciers wordt een service voor het registreren van een DOI voor datasets opgeboden om publicatie van de data.
- Governance
 - Licensing (geheimhouding, moratorium, intern extern)
 - Er dient overeenstemming te worden bereikt over een gebruikerslicentie die past binnen ARVODI, AQUO en SDN. De gebruikerslicentie dient opgenomen te worden in de Data Factsheet.
 - Voor toegangsbeheer tot data zijn verschillende methodes voorhanden (bijvoorbeeld Client-Server Certificate Authorisation). Daar zal een pragmatische keuze in gemaakt worden. De bottom-line is dat de bronhouder alleen de data ontsluit die hij/zij openbaar wil maken.
 - Metadata zijn altijd openbaar
- Data lifecycle management
 - Data dienen tijdens en na het project beschikbaar te zijn, zie ook 0.

Validatie (Data quality control)

Vóór oplevering worden data en metadata gecontroleerd of ze voldoen aan FAIR principes, technische, semantische of inhoudelijke datakwaliteitsnormen en de opgegeven specificaties (Data Factsheets) bij aanvang van een project. Pas na goedkeuring wordt betaling van het (deel)project vrijgegeven door het Uitvoeringsbureau.

- Er is een scala aan kwaliteitscontroles mogelijk⁸. De validatie zal echter beperkt worden tot de technische en semantische kwaliteit van de (meta)data.
- Inhoudelijke kwaliteitsbeoordeling op outlyers in de data heeft niet de voorkeur van het IHM omdat hiermee mogelijk subjectiviteit wordt geïntroduceerd. Outlyers kunnen bovendien ecologische importantie hebben die door filtering worden gemist.

Organisatorische afspraken

In deze sectie worden afspraken vastgelegd die van belang zijn voor het datamanagement(beheer) van MONS zoals in hoofdstuk 3 van het Protocol Mariene Data - Bijlage 1). Aspecten van de gehele levenscyclus van data wordt hierin betrokken (inclusief de archivering en publicatie (DOI) van data.

Begrippenlijst

Eenduidige uitleg van de gebruikte begrippen (zie Protocol Mariene Data). Hierbij wordt aangehaakt op gebruikte terminologie van het Kader Datakwaliteitsstelsel, Nora Architectuur en semantische standaarden.

⁸ Handreiking (Kader) Datakwaliteitsstelsel; beschikbaar via intranet zoekterm 'Datakwaliteitsstelsel'.