

Memo

Aan

Erik de Rooij, Flip Dirksen, Geri Wolters, Jan-Maarten Verbree, Jeroen Gerrits, Jurgen Boerboom, Reinout van Rees, Sander Loos, (cc Bob van der Staak, Bram de Graaf)

Datum	Kenmerk	Aantal pagina's
27 maart 2018	11202186-000-DSC-0008	5
Van	Doorkiesnummer	E-mail
Stef Hummel	+31(0)6 1019 8112	Stef.Hummel@deltares.nl

Onderwerp

Besprekingsverslag bijeenkomst DD werkgroep 19 maart j.l.

Uitbreidingen van de DD-API met verwachtingen

De diverse voorstellen in memo 11200534-DSC-0024-DD-API-voorstel-verwachtingen worden besproken (waarbij Stef indoeën van toepassing aangeeft wat de ideeën hierover zijn bij het tweede agendapunt - de DD-API bij de MKK-DL van RWS).

De *forecastReferenceTime* wordt hernoemd in **analysisTime**.

Dit keyword wordt in de response niet voor in elk event gezet, maar slechts 1 keer per tijdserie, en wordt op het zelfde niveau als "start" en "end" gezet.

Voor uit meerdere verwachtingssommen samengestelde tijdreeksen kan dan alleen een analysisTime voor de samengestelde serie worden terug gegeven.

Ook de ensemble member index (als die er is) wordt op het hoogste niveau (dus naast start en end) gezet.

Sander presenteert de opzet in Hydronet. Hier bestaan naast een TimeSeries ook ModelTimeSeries, met een keyword modelDate, en een EnsembleTimeSeries. Een vergelijkbare opzet dus.

Het keyword "prediction" heeft een te hoog astrologiegehalte en wordt hernoemd in "forecast".

Het keyword "institution" zal niet op hetzelfde niveau als "source" gezet worden, maar zal onderdeel van "source" gaan uitmaken.

Onder het keyword "source" gaan we geen "type" gebruiken, maar het keyword "process" om aan te geven of de data een forecast of measurement betreft. Het keyword "source" zal ook een keyword "modelName" gaan bevatten met informatie over het model, zowel de naam als het versienummer (e.g. HIRLAM v7.2). Het "version" attribuut vervalt dus.

Vooralsnog focussen we binnen de DD op het terug geven van tijdreeksen, dus we gaan voorlopig niet volledige data sets (volledige netCDF-files) retourneren. Ook parkeren we voorlopig het ontsluiten van roosterdata.

Een ieder denkt na over hoe ensembles ontsloten zouden moeten worden:

- (met name) hoe krijgt een gebruiker een volledige ensemble set terug (file downloaden)

- hoe haalt hij een specifieke member op (voor de hand liggend is een filterparameter in de /timeseries-request)

Vervolgactie

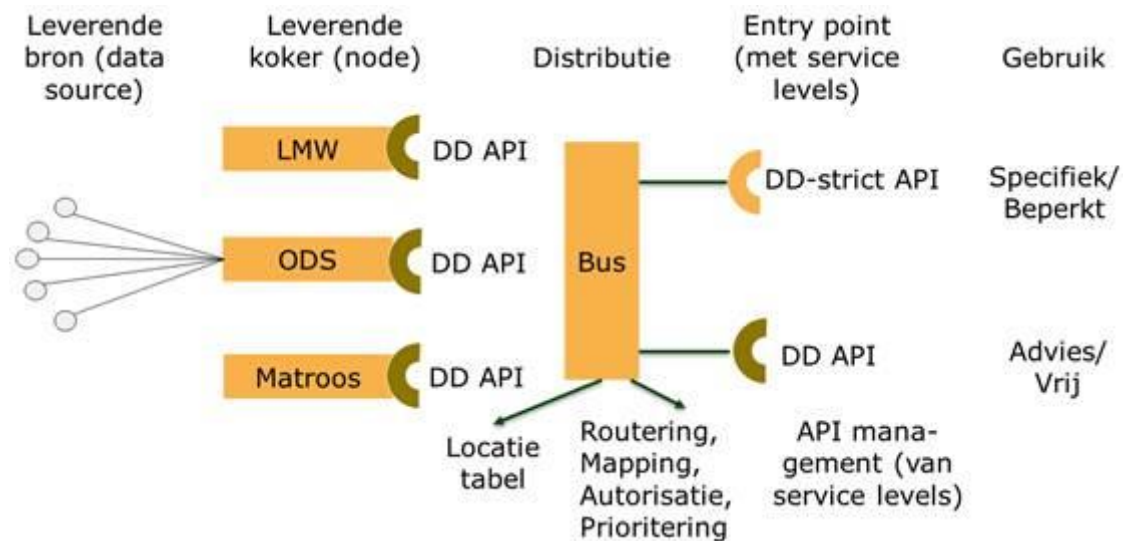
Stef gaat nieuwe JSON voorbeelden maken waarin bovenstaande wijzigingen verwerkt zijn zodat de hele werkgroep nog eens goed kan kijken naar het ontwerp.

Inzet van de DD-API bij de 'missiekritische data' web service van RWS

Stef en Jurgen presenteren de door RWS gewenste werkwijze m.b.t. het gebruik van de DD-API.

Overzicht van het systeem

RWS/CIV gaat de DD-API implementeren op haar distributie-laag, en gaat de DD-API gebruiken voor het bevragen van de toeleverende databronnen, zie onderstaand plaatje (waarbij de 'Bus' staat voor de distributie-laag).



Voor operationele systemen wil men echter naast de reguliere DD-API een beperktere versie aanbieden ('DD-Strict-API'), die geen discovery-functie biedt, maar waarin men rechtstreeks de gewenste tijdserie dient te benaderen.

Het DD-strict end point:

- Zet de request-url om in een DD-API-call naar het bronsysteem
 - Geeft de response die uit het bronsysteem komt ongewijzigd door naar de vraagsteller
- Het idee is dat daarbij niet eerst via een query een uuid wordt opgevraagd, maar dat men in de url volledig specificeert waarom het gaat.

De bus verzorgt de routing van de binnenkomende call naar het betreffende bronsysteem. Dat doet hij ook voor de binnenkomende 'reguliere' DD-API calls. Voor laatstgenoemde geldt dan wel dat in bepaalde gevallen extra informatie moet worden meegegeven om te kunnen bepalen wat de bron is. Voorgesteld wordt om een filterparameter 'processName' te

introduceren, met als waarden b.v. measurement en forecast (de mogelijke waarden zijn gelijk aan die van het veld 'source.process' in bovengenoemd voorstel voor de verwachtingen).
Het end point in 'DD-strict'

De `/locations` end point in 'DD-strict' ziet er als volgt uit:

- `/locations` (welke locaties?)
- `/locations/hoekvanholland/parameters` (welke grootheden aldaar?)
- `/locations/hoekvanholland/parameters/waterlevel/timeseries` (welke tijdstippen?)
- `/locations/hoekvanholland/parameters/waterlevel/timeseries?start=...&end=...`

(Over het keyword 'parameters' is bij de voorbereiding reeds lang gediscussieerd. In feite gaat het bij een parameter om identificerende string van wat in de huidige DD-API het `observationType` is. Zie de discussie/afwegingen verderop.)

De groep vraagt zich af waarom er een apart end point wordt gedefinieerd, en waarom niet gewoon `/timeseries?< filter-parameters>` wordt gebruikt. Omdat men binnen de MKK-DL niet met uuid's wil werken moet deze functionaliteit immers sowieso toegevoegd worden aan de DD-API, aangezien DD-strict-url gemapped moet kunnen worden naar zo'n `/timeseries?<filter-parameters> url`.

Jurgen en Stef geven aan dat de wens is van de MKK-DL-projectleider, om te zorgen dat het 'pad' naar de beschikbare gegevens zo expliciet mogelijk wordt. Daarnaast is het een veel gebezigde en gepromote aanpak voor RESTful-API's.

Er ontstaat discussie over de naam **DD-strict**. De mening van de groep is dat het hier gaat om een web service van RWS, die weliswaar vertaald wordt in een DD-API-request, maar daarmee geen onderdeel van de DD-API-definitie hoeft te zijn (en volgens de groep ook niet moet zijn). Verzoek aan RWS is dus om voor de base-url van het `/locations` end point een andere name te bedenken. (In het resterende deel van het verslag wordt nog wel de term DD-strict gehanteerd.)

Overigens heeft de groep geen moeite met het toevoegen van extra filter-parameters aan het `/timeseries` end point, zolang de uuid-benadering ook maar blijft bestaan, en het implementeren van de extra filter-parameters optioneel is. Laatstgenoemde is het geval; systemen die geen toeleverancier van MKK-DL zijn hoeven ze niet te implementeren.

Naamgeving `observationType` / `parameter`

Zoals gezegd is hier in de voorbereiding bij RWS reeds veel over gediscussieerd. De bestaande term `observationType` gaat te veel uit van metingen (daar is eerder in DD-API-traject ook al flink over gediscussieerd), terwijl het ook om verwachtingen gaat, en meer in het algemeen over modelresultaten. Kandidaten voor een betere naamgeving moeten gezocht worden in de terminologie van bestaande standaarden, zoals:

- `observedProperty` (OGC's O&M, en dus Aquo's IM-metingen)
- `quantity` (SI-stelsel, en b.v. voor een deel bij netcdf-CF)
- `parameter` (veel gebruikt in de mariene hoek, b.v. SeaDataNet en EMODnet)

Na enige discussie blijkt *parameter* de favoriet te zijn. Bezwaar is dat het niet conform Aquo is, maar het voordeel is dat het een door iedereen begrepen term is.

`observationType` wordt dus vervangen door `parameter`. Het is van belang om goede voorbeelden op te stellen van parameters, zowel in de response van (vanaf nu dus) het

`/parameters` end point, als in het gebruik van filteren op parameter binnen het `/timeseries` end point.

Optionele onderdelen in de response

Binnen de het gerichte gebruik van de DD-API vanuit de 'DD-strict' calls zijn een aantal onderdelen in de response niet nodig of zelfs verwarrend. Voorstel is om aan de `/timeseries` end point parameters mee te kunnen geven die aangeven dat in de response deze onderdelen onderdrukt moeten worden. Jurg en Stef zullen voorbeelden uitwerken.

Vervolgactie

Stef gaat url- en response-voorbeelden maken waarin bovenstaande wijzigingen verwerkt zijn zodat de hele werkgroep er goed naar kan kijken.

Specificatie van een uitgebreidere DD-API voor ecologische data

De ervaring met de DD-API op AquaDesk heeft geleerd dat DD-API-1.0 niet voldoende geschikt is voor ecologische data (metingen van organismen, chemische stoffen, fysiologische waarden). Ecologische metingen zijn veel meer divers dan waterkwaliteitsmetingen: Er worden meer parameters gemeten, er worden meer eenheden gebruikt, er worden verschillende meetstandaarden gebruikt, etcetera. Er wordt derhalve meer informatie omtrent metingen vastgelegd dan in de resource objecten van de DD-API zijn gespecificeerd, informatie die noodzakelijk kan zijn voor bepaalde beoordelingssystemen.

Geri heeft daarom op verzoek van o.a. HHNK, HydroLogic en RWS een voorstel uitgewerkt voor een uitbreiding op de DD-API voor ecologische data.

Hij presenteert de achtergrond en opzet van deze 'DD-ECO-API' (de presentatie is bijgevoegd bij dit verslag).

De werkgroepleden vinden het een zeer interessante ontwikkeling, waarover ze graag mee willen denken en discussiëren.

Vervolgacties

De werkgroepleden overleggen op het thuisfront wie er namens hun partij deel wil nemen aan de uitwerking van de DD-ECO-specificaties.

Stef en Geri zorgen vervolgens dat hier bijeenkomsten en/of web conferences voor worden belegd. Om geïnteresseerden een goed beeld te geven van de wens en opzet zal op korte termijn een memo worden rondgestuurd met een wat uitgebreidere beschrijving van wat er in de presentatie staat.

Status authenticatie-pilot en vervolgstappen

Bob sluit remote aan bij de vergadering.

Zowel Erik als Reinout hebben problemen ervaren met het op basis van OpenID Connect 'doorloggen van machine naar machine':

- De gebruiker logt in op machine A waar de metadata-catalogus staat.
- Machine A haalt een token bij de AuthenticationProvider (AP)
- De gebruiker wil data zien die op machine B gaat, dus vanuit machine A wordt met dat token een get-request gedaan naar machine B.
- Het token is echter alleen afgegeven voor machine A, dus op B komt een foutmelding.

Datum
27 maart 2018

Ons kenmerk
11202186-000-DSC-0008

Pagina
5/5

Bob geeft aan dat e.e.a. wel binnen HydroNet werkt. Hij zal voor het eind van de week tijd nemen om op te schrijven hoe het daar is geïmplementeerd.

Vervolgacties

Erik en Reinout mailen voor de volledigheid de beschrijving van hun probleem naar Bob.

Bob beschrijft hoe e.e.a. binnen HydroNet werkt.

In de communicatie hierover wordt ook Geri meegenomen.

Als dit inderdaad een goeie oplossing is plannen we met alle partijen een orde-2-daagse mini-sprint om het doorloggen van systeem naar systeem daadwerkelijk te implementeren.