

Rapport 2021.008 | 18 februari 2022

Quickscan QC- implementatie

Data-aspecten, scenario's, fasering en
conclusies



Trefoil Hydrology

Services, Science & Software

Quickscan QC-implementatie

Data-aspecten, scenario's, fasering en conclusies

Rapport 2021.008 | 18 februari 2022

Opdrachtnummer
21101A

Versie
v0.7 (concept)

Opdrachtgever(s)
Platform meetnetbeheerders grondwaterkwantiteit provincies
Rijkswaterstaat Water, Verkeer en Leefomgeving
Programmabureau BRO, ministerie van BZK

Auteur(s)
J. (Jos) von Asmuth, Trefoil Hydrology

Met dank aan
Stuurgroep en leveranciers (zie bijlage B en C)

Trefoil Hydrology

Groenedijk 36
3544 AB Utrecht

+31 6 21690206
info@3hydro.nl

www.3hydro.nl

2022 © Trefoil Hydrology



Uitgegeven onder de Creative Commons Attribution 4.0 International Public License (CC-BY). Deze uitgave mag, onder voorwaarde van naamsvermelding, worden veeleenvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, en openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.
Trefoil Hydrology is niet aansprakelijk voor eventuele schade voortvloeiend uit toepassing van de inhoud van dit rapport of onderzoek.

Inhoud

1	Inleiding	5
1.1	Achtergrond	5
1.2	Doelstelling	5
1.3	Leeswijzer	5
1.4	Versiehistorie	5
2	Randvoorwaarden	6
2.1	Inleiding	6
2.2	Uitgangspunten	7
2.3	Wettelijk kader	7
2.4	Normatief kader	7
2.5	Praktisch kader	8
3	Data-inhoud	9
3.1	Inleiding	9
3.2	BRO-gegevens	9
3.3	Bron-gegevens	10
3.4	Non-BRO-gegevens	10
3.5	Correcties en keuringen	10
4	Dataformats	12
4.1	Inleiding	12
4.2	Bron- en HDF5-format	12
4.3	Aanlevering en import	13
4.4	Retourlevering en export	14
4.5	Sensorgegevens	14
5	Databeheer	16
5.1	Inleiding	16
5.2	Oplossingen voor versiebeheer	17
5.3	Scenario I: Vast versiebeheer met gedeeltelijke uitwisseling	18
5.4	Scenario II: Flexibel versiebeheer met volledige uitwisseling	18
5.5	Scenario III: Flexibel en gedeeltelijk versiebeheer	19
5.6	Scenario IV: Onafhankelijk en minimaal versiebeheer	19
6	Datastroom	20
6.1	Inleiding	20
6.2	Scenario I: regie QC-Wizard	20
6.3	Scenario II: regie leverancierssysteem	21
6.4	Scenario III: regie opslag	22
6.5	Scenario IV: Integratie	23
6.6	Opslag	23

7	Fasering	24
7.1	Inleiding	24
7.2	Korte termijn: BRO-migratie historie en voorlopig oordeel	24
7.3	Langere termijn: historie en volledig oordeel	24
7.4	Lange termijn: operationele data en volledig oordeel	24
8	Conclusies en aanbevelingen	25
8.1	Organisatie	25
8.2	BRO-implementatie	25
8.3	QC- implementatie	25
8.4	Vervolg en aanbevelingen	25
	Literatuur	26
	Bijlage A: Stuurgroepleiden	27
	Bijlage B: Leveranciers	28
	Bijlage C: Quickscan-release en disclaimer	29
	Quickscan-release data en applicaties (met disclaimer) <bijwerken>	29
	Tijdelijke ontheffing wettelijke plichten GLD	30
	Dataverzameling, -conversie, -migratie en -kwaliteitscontrole	31
	Operationele release data en applicaties, voorlopig oordeel	32

1 Inleiding

1.1 Achtergrond

Dit rapport beschrijft...

<verder uitwerken>

1.2 Doelstelling

De doelstelling van het project is ...

<verder uitwerken>

1.3 Leeswijzer

De opbouw van het rapport is...

<verder uitwerken>

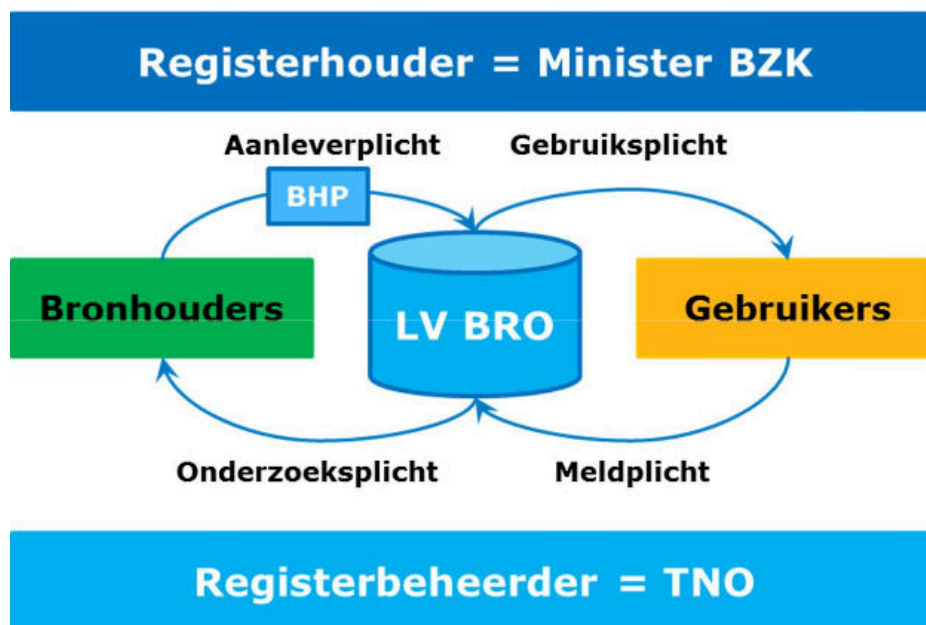
1.4 Versiehistorie

Datum	Versie	Omschrijving
xxx 2022	1.0	Definitief
xxx 2022	0.9	Eindconcept
Feb 2022	0.7	Concept t.b.v. databaseheeropties en -formats: <ul style="list-style-type: none">- H4: herzien en aangevuld- H5: toegevoegd- H7: aangescherpt
Juli 2021	0.65	Concept t.b.v. leverancieroverleg: <ul style="list-style-type: none">- H2: aangevuld- H3,6,7: aangescherpt- H4: herzien
Juni 2021	0.6	Ruw concept t.b.v. belronde bronhouders

Tabel 1: Overzicht over de verschillende versies die verschijnen of verschenen zijn van dit document

2 Randvoorwaarden

2.1 Inleiding



Figuur 1: Wettelijke plichten voor bestuursorganen rond en in samenhang met de Basisregistratie Ondergrond (Programmabureau BRO, 2019)

Verschillende partijen hebben een verschillende belang en/of rol bij gegevens over de ondergrond, en de processen daaromheen. De komst van de BRO heeft daarbij consequenties voor alle partijen in de keten van bron tot BRO, naar gebruik en weer retour. De BRO stimuleert daarbij juist de verbinding en samenwerking in de keten, waar dat tot voor kort veel lastiger te realiseren was. Die verbinding kun je in technische zin interpreteren, omdat het gaat om geautomatiseerd dataverkeer (zie o.a. Figuur 1), maar ook in organisatorische zin, omdat de dataverbindingen ook zorgen voor verbindingen en samenwerking tussen mensen en organisaties.

Dit maakt dat deze mensen en organisaties ook meer dan voorheen met elkaar te maken krijgen, van elkaars werk gebruik kunnen maar ook moeten maken, en elkaar dus ook moeten kunnen vinden en begrijpen. Dat is niet altijd vanzelfsprekend, omdat de gehanteerde methoden, concepten en begrippen (en het onderwerp QC is daarbij zeker geen uitzondering) ingewikkeld zijn, zeker voor de niet-technici onder ons. We moeten elkaar dus vinden en begrijpen, en de provincies en Rijkswaterstaat pakken die handschoen als bronhouders gezamenlijk verder op. Samenwerking is daarbij uitermate positief, maar kan ook wrijving en afstemmingsproblemen leiden. We specificeren hier dan ook graag allereerst een aantal punten die juist niet een doel of uitgangspunt zijn, namelijk het:

- overnemen van de rol van marktpartijen
- beoordelen of benadelen van specifieke marktpartijen
- bewegen van marktpartijen om 'om niet' hieraan bij te dragen

Voor samenwerking is in het algemeen wel meer uniformiteit en afstemming nodig. We zien als doel of uitgangspunt dus we! het:

- reguleren van de markt waar nodig, opdat:
 - voldaan wordt aan wettelijke plichten en eisen
- stimuleren van de markt waar mogelijk, met betrekking tot:
 - eenduidigheid, uniformiteit en standaardisatie
 - openheid
 - efficiëntie
 - samenwerking
 - delen van begrippen, inzichten, methoden en resultaten

2.2 Uitgangspunten

In overkoepelende en meer technische zin is implementatie van het QC-protocol van de provincies en RWS (Platform meetnetbeheerders, 2018) het uitgangspunt. Onderliggende uitgangspunten van dit protocol, zijn:

- transparantie en reproduceerbaarheid
- uniformiteit en standaardisaties
- systematische, gedeelde aanpak
- gegarandeerde basiskwaliteit
- samenwerking en koppeling van systemen
- onafhankelijke en open data en datakwaliteit

2.3 Wettelijk kader

De (Rijks)overheid moet, net als alle burgers, nu eenmaal ook zelf voldoen aan alle wettelijke kaders die gelden voor haar gegevens, en dat wederom vanaf de bron in het veld, via bronhouderportaal en BRO, naar gebruik en weer retour via bijvoorbeeld terugmeldingen en da daarbij behorende meldplicht. Het stimuleren en reguleren van de markt is daarom nodig, met als doel te voldoen aan die wettelijke kader. Al met al zijn met betrekking tot de gegevens in de BRO, tenminste de volgende, onderstaande wettelijke kaders en regelingen van belang.

De Wet BRO

De komst van de Basisregistratie Ondergrond bestendigt en formaliseert al deze rollen en relaties via de wet BRO, die ook een wettelijke basis en verplichting geeft aan deze processen, in de vorm van de leverplicht, gebruiksplicht, meldplicht en onderzoeksplicht (Figuur 1).

De Kaderrichtlijn Water (KRW)

<verder uitwerken>

Privaatrecht

<Vergunningen en contracten, verder uitwerken>

2.4 Normatief kader

Normen

<NEN/ISO, verder uitwerken>

Protocollen

< QA/QC verder uitwerken>

De IMBRO en IMBRO/A-standaard

< BRO, verder uitwerken>

De IMBRON- en IMBRON/A-standaard (praktijkstandaard)

< Te ontwikkelen i.s.m. Geonovum, verder uitwerken>

2.5 Praktisch kader

Uitdagingen en kansen

< verder uitwerken>

Realiseerbaarheid en toepasbaarheid

< verder uitwerken>

Efficiëntie

< verder uitwerken>

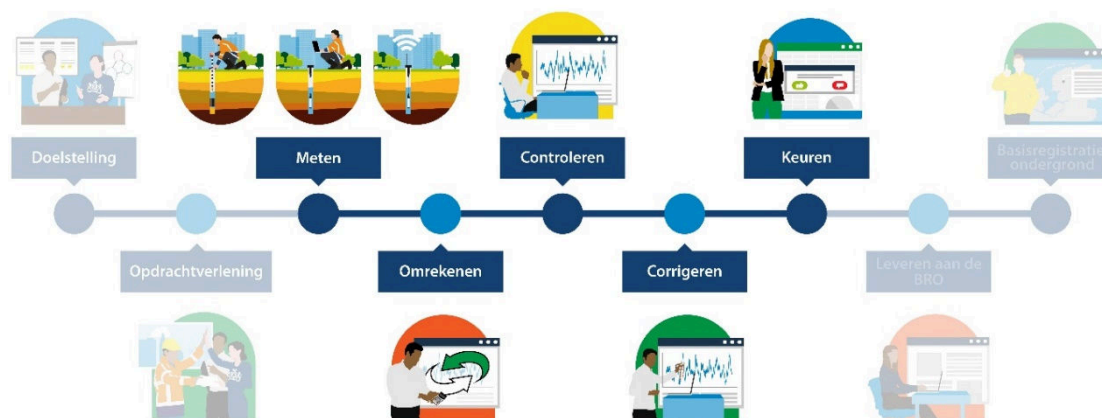
Rol van de QC-Wizard

< hulpmiddel, toetsing van kwaliteit en richtlijnen, verder uitwerken>

3 Data-inhoud

3.1 Inleiding

De gegevensinhoud van de BRO is logischerwijze niet uitputtend. Bij elk van de registratieobjecten speelt een scope-afbakening, die wordt beschreven en vastgelegd in de zogenoemde scope-documenten. De gegevensinhoud van de BRO is gericht op het registreren van unieke en authentieke gegevens, die geschikt zijn voor hergebruik en die van 'landelijk' belang zijn, d.w.z. voor meer partijen dan de individuele bronhouder of andere stakeholder alleen. De consequentie van deze scope-afbakening is dat uiteraard ook gegevens daarbuiten vallen. Die gegevens zijn dus niet landelijk van belang, maar kunnen dat wel zijn voor uitoefening van de processen en plichten van individuele bronhouders. Het platform meetnetbeheerders van de provincies heeft al langer geleden de handschoen opgepakt om de kwaliteit van grondwaterdata en de inwinning daarvan te borgen van 'bron tot BRO', met als aanleiding destijds de Kaderrichtlijn Water (Von Asmuth en Van Geer, 2013; Platform meetnetbeheerders, 2018).



Figuur 2: De stappen in het bewerkingsproces, vanaf de originele meting tot aan de definitieve gegevens (bron: (Platform meetnetbeheerders, 2018; Standaardisatieteam, 2020)).

<Afbakening Bron en BRO-gegevens, verder uitwerken>

3.2 BRO-gegevens

< Verder uitwerken van onderstaande punten>

BRO-gegevens (zie catalogi):

1. GMN
2. GMW
3. GLD
4. GAR
5. BHR

N.B.:

- De inhoud van GAR valt grotendeels buiten scope van dit project, er zijn wel relaties tussen GAR, GMN en GMW en GLD.

- De inhoud van BHR valt grotendeels buiten scope van dit project, er zijn wel relaties én er is overlap van gegevens (die dus niet uniek zijn in de BRO zelf) tussen BHR en GMW.

3.3 Bron-gegevens

< Bron-gegevens hebben een wel verwijzing naar een BRO-ID, verder uitwerken van onderstaande punten, omschrijving en definitie volgt, o.a. bij standaardisatie >

Bron-gegevens:

1. Bron-gegevens GMN
2. Bron-gegevens GMW
3. Bron-gegevens GLD
4. Bron-gegevens GAR
5. Bron-gegevens BHR

Onderscheid in kwaliteitsregime historisch en actueel net als BRO (bijv. Bron/A- en Bron-kwaliteit)

3.4 Non-BRO-gegevens

< Non-BRO-gegevens hebben geen verwijzing naar een BRO-ID, verder uitwerken van onderstaande punten, omschrijving en definitie volgt, o.a. bij standaardisatie >

Non-BRO-gegevens (dus zonder verwijzing naar BRO-ID):

1. Peilschalen (voor zover geen peilbuis)
2. Minder dan IMBRO/A kwaliteit

3.5 Correcties en keuringen

< Verder uitwerken >

4 Dataformats

4.1 Inleiding

De functionaliteit van de QC-Wizard m.b.t. het importeren en converteren van gegevens die aangeleverd worden in verschillende formats is beperkt. De kernfunctionaliteit van de Wizard is QC, uitbreiding met functionaliteit om als 'stekkerdoos' of 'data hub' te gaan fungeren is zelfs onwenselijk, aangezien functionaliteit op dit gebied beter door een nader te selecteren marktpartij ontwikkeld kan worden. Vanuit het QC-traject wordt bovendien om dergelijke redenen gestreefd naar de ontwikkeling van een 'praktijkstandaard voor brongegevens' die zo mogelijk de data-inhoud van zowel de Bron- en BRO-gegevens als de correcties en keuringen in zijn geheel bestrijkt. Op moment van schrijven is tegelijkertijd echter de migratie van gegevens van DINO naar de BRO in volle gang. De QC-Wizard dient om die reden waar mogelijk toch ook al zoveel mogelijk bruikbaar te zijn daarbij.

Beide hierboven genoemde doelen en uitgangspunten zijn helaas moeilijk verenigbaar met elkaar. Er bestaat nog geen open datastandaard of vrij beschikbare software op gebied, de dataformats die de QC-Wizard nu of in de nabije toekomst wel ondersteunt zijn daarom met name aan de import-kant een compromis (zie paragraaf 4.3 en 4.4). Qua implementatie is en blijft het streven wel om dit op een zo uniform mogelijke wijze te doen, of al dan niet te kiezen voor import en uitwisseling op ad hoc basis.

N.B.:

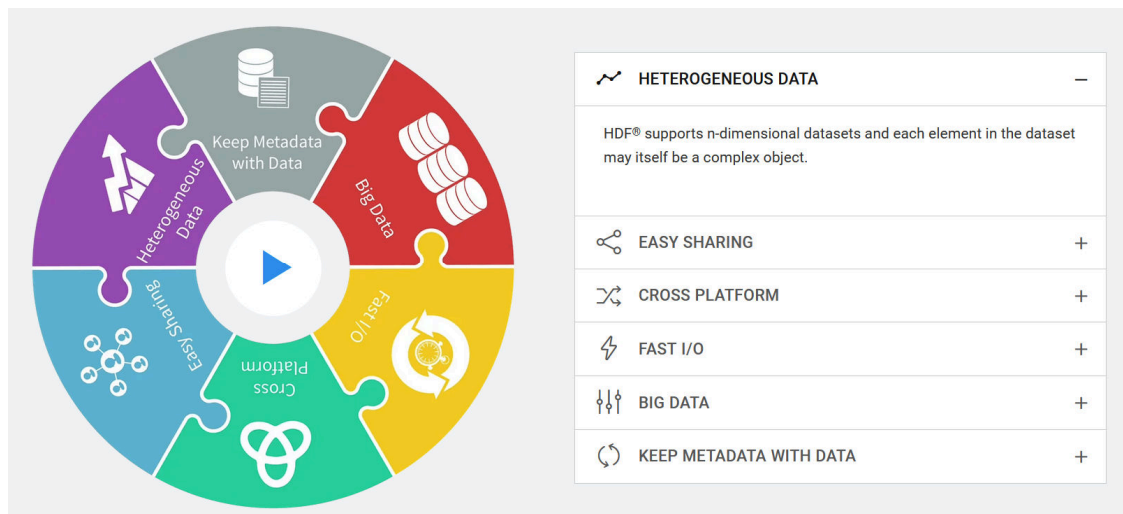
- We maken in dit hoofdstuk geen onderscheid in ondersteuning van dataformats voor import via een user interface, dan wel aan- of retourlevering van bestanden zonder dat. Reden daarvoor is de tijdelijkheid van dat onderscheid, aangezien de data in toestand van migratie zijn en de QC-Wizard nog in toestand van ontwikkeling.
- Inzet van andere programmatuur (waaronder ArtDiver of MenyanthesOS) geeft meer mogelijkheden en flexibiliteit voor toepassing van de QC-Wizard op andere databronnen en formats. Mocht ondersteuning van aanvullende formats of opties toch nodig zijn, dan dienen de consequenties daarvan in onderling overleg bekeken te worden.

4.2 Bron- en HDF5-format

De QC-Wizard zelf gebruikt voor het eigen datamodel en dataopslag het zogeheten bron-format, dat is gebaseerd op het spreadsheet- en tabelformat voor BRO-objecten (Von Asmuth, 2020). De opzet en principes van dit format zijn dusdanig gekozen dat het naadloos aansluit bij de datamodellen van de BRO zelf aan de ene kant, terwijl het aan de andere kant ook ruimte laat voor bron-gegevens. De bron- of aanvullende gegevens dienen en worden daarbij dan tegelijkertijd ook netjes afgebakend en 'gemapt' op de BRO-gegevens. Het spreadsheet- en tabelformat format is bovendien eenvoudig overdraagbaar en vertaalbaar in verschillende varianten en platforms, zoals (zie ook von Asmuth en Vonk (2017)) :

- **CSV-format** – deze kunnen eenvoudigweg verkregen kunnen worden door o.a. conversie van de beschikbare xlsx-voorbeeldbestanden;
- **HDF5-format** – dit is een open format waarvan de afkorting staat voor Hierarchical Data Format, de verdere eigenschappen van HDF5 worden hieronder nader toegelicht;
- **DDL-script** – deze afkorting staat voor Data Definition Language, en DDL dient in het algemeen voor het aanmaken van tabel-structuren in databasesystemen zoals Oracle of PostgreSQL.

De bron-bestanden van de QC-Wizard zijn zelf in feite een variant van het HDF5-format, waarmee dit format de voorkeur verdient voor data-uitwisseling omdat het a) de data-inhoud van de QC-Wizard per definitie volledig afdekt en er b) geen verdere import- of export-routines nodig zijn om het aan te maken. HDF5 is zoals gezegd een open format dat ondersteund wordt door een groot aantal softwareplatforms en programmeertalen, inclusief Python en Matlab (zie https://en.wikipedia.org/wiki/Hierarchical_Data_Format). Het hiërarchische karakter van het format sluit goed aan bij dat van het XML-format dat de basis vormt van alle BRO-berichtenverkeer of de hiërarchische UML-schema's die op hun beurt weer de basis zijn van alle BRO-catalogi. HDF is oorspronkelijk ontwikkeld door het National Center for Supercomputing Applications, en wordt nu ondersteund door de non-profit organisatie HDF Group (<https://www.hdfgroup.org>). Voor een illustratie over met voordelen en toelichting op het HDF-format, zie *Figuur 3*.



Figuur 3: Toelichting op en voordelen van het HDF-format (bron: <https://www.hdfgroup.org>).

4.3 Aanlevering en import

De formats die de QC-Wizard ondersteunt voor aanlevering of import (al dan niet via de user interface) zijn zoals gezegd een compromis en afhankelijk van de oorsprong en inhoud van de data. We splitsen de verschillende opties hieronder om die reden uit al naar gelang daarvan:

DINO- en BRO-gegevens

1. Bestanden met 'datadumps' door TNO vanuit DINO (*.csv)
2. BRO-registratie- en uitgifteberichten (*.xml)

Bron-gegevens

3. Bestanden en bijbehorende gegevens van gangbare merken (druk)sensoren (o.a. *.idc, *.mon)

Bron- en non-BRO-gegevens, correcties

4. ArtDiver-bestanden (*.adi, gedeeltelijk)
5. MenyanthesOS bestanden (*.men, tijdelijk)

N.B.: Omdat men-bestanden onvoldoende BRO-bestendig zijn, zal de import daarvan in zowel de QC-Wizard als MenyanthesOS komen te vervallen (mogelijk al binnenkort). Het vraagstuk van BRO-bestendigheid speelt wat dat betreft ook voor adi-bestanden (Von Asmuth en Van Vliet, 2020; Van Vliet e.a., 2021)).

Bron-, non-BRO- en BRO-gegevens, correcties en keuringen

6. Bron-bestanden (*.bron)

Naar verwachting volgen:

7. BRO-correctieberichten (*.xml)
8. Spreadsheetformat voor BRO-objecten (*.xlsx)
9. Spreadsheetformat voor bron-gegevens (*.xlsx)

N.B.: Het spreadsheet- en tabelformat voor BRO-objecten is ontwikkeld in opdracht van het programmabureau BRO (Von Asmuth, 2020). Documentatie daarover met nadere toelichting is samen met voorbeeldbestanden te vinden op <https://github.com/QC-Protocol>. **<nu nog platform meetnetbeheerders>.**

4.4 Retourlevering en export

Waar het gaat om de retourlevering of export van resultaten van de QC-Wizard, dan zijn nut en noodzaak om daarin compromissen te sluiten of voor ad hoc oplossingen te kiezen minder groot. Er is nog geen bestaand format of open datastandaard die de relevante data-inhoud in zijn geheel bestrijkt, het (in ontwikkeling zijnde) datamodel van de QC-Wizard zelf kan daarom als vertrek- of uitgangspunt dienen. Voor wat betreft het BRO-aandeel in de gegevens, fungeert de BRO zelf uiteraard als standaard. Communicatie met en/of in de taal van de BRO is voor de QC-Wizard om uiteenlopende redenen van belang, reden waarom import en export van BRO-berichten (ook) daar geïmplementeerd is.

BRO-gegevens

10. BRO-registratieberichten (*.xml)

Bron-, non-BRO- en BRO-gegevens, correcties en keuringen

11. Bron-bestanden (*.bron, hdf5 format)

Naar verwachting volgen:

12. BRO-correctieberichten (*.xml)
13. Spreadsheetformat voor BRO-objecten (*.xlsx)
14. Spreadsheetformat voor Bron-gegevens (*.xlsx)

4.5 Sensorgegevens

Dir	Bestand	NITGCode	FilterNr	LoggerSerienr	LoggerDiepte	ControleDatumTijd	ControleMeting
[String]	[String]	[String]	[Integer]	[String]	[m-bb]	[ExcelDate]	[m-bb]
okt-11	31GZHxxx01_10,08,2011_00,00,00_00.IDC	B31F2155	1	3046			
okt-11	3D-KOLLAND_27,07,2011_00,00,00_00.IDC	B39B1571	1	11334			
jun-11	40524_29,03,2011_16,00,00_00.IDC	B32G0387	1	5319			
okt-11	6D_26,07,2011_00,00,00_00.IDC	B39B1570	1	11344			
okt-11	716-B_22,07,2011_00,00,00_00.IDC	B31D1431	1	11356			
okt-11	7D-KOLLAND_09,08,2011_00,00,00_02.IDC	B39B1572	1	11346			
okt-11	7D-KOLLAND_27,07,2011_00,00,00_02.IDC	B39B1572	1	11346			
okt-11	BV_gw1_22,07,2011_00,00,00_00.IDC	B31G2303	1	11345			
okt-11	BZ_gw1_28,07,2011_00,00,00_00.IDC	B32G1154	1	11347			
okt-11	D3_26,07,2011_00,00,00_02.IDC	B39B1569	1	11342			
okt-11	H310666.IDC	B31H0666	1	10917			
okt-11	HB_gw1_04,08,2011_00,00,00_00.IDC	B39E2778	1	11335			
okt-11	HB_gw2_04,08,2011_00,00,00_02.IDC	B39E2779	1	11336			
okt-11	MK_gw1_26,07,2011_00,00,00_00.IDC	B32G1155	1	11332			
okt-11	NP_gw1_06,08,2011_00,00,00_00.IDC	B31F2299	1	11331			
okt-11	NP_gw2_06,08,2011_00,00,00_00.IDC	B31F2300	1	11324			

Figuur 4: Voorbeeldoverzicht van sensorgegevens, met bijbehorende relatie met de put (NITGCode) en buis (FilterNr) waar de metingen gedaan zijn én de handmatige controlemetingen (bron: provincie Utrecht).

In opdracht van provincie Utrecht en in samenwerking met RHDHV is inmiddels ervaring opgedaan met het volgen van optie 3. Het gaat daarbij dus om het verzamelen en gebruiken van de onderliggende sensorbestanden, als brondata voor o.a. de tijdmeetwaardeparen die onderdeel zijn van registratieobject Grondwaterstandonderzoek (GLD). Sensorbestanden bevatten doorgaans meer dan de (omgerekende) tijdmeetwaarden alleen. Het format en de inhoud en betekenis van die gegevens is echter bekend, zie daarvoor de desbetreffende leveranciers en hun documentatie. Daarmee zijn sensorbestanden in het algemeen ook bruikbaar als brondata. Naast de sensorgegevens zelf, is echter ook de bijbehorende relatie met de put (NITGCode, of een alternatief zoals BroID) en het buisnummer (FilterNr) waar de metingen gedaan zijn van belang. Bovendien zijn de handmatige controlemetingen nodig, zie verder de overige aandachtspunten hieronder.

Aandachtspunten:

- Brongegevens sensoren en loggers (merk, type, bereik, etc.)
- Brongegevens luchtdruk en -compensatie
- Brongegevens omrekening (bekend en/of beschikbaar in de databestanden bij bekende merken zoals Diver en Keller)

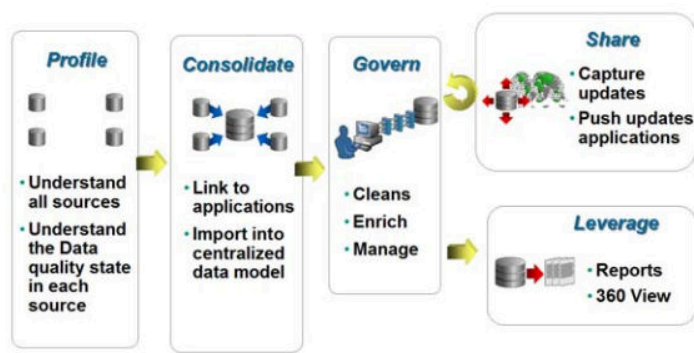
<verder uitwerken>

5 Databeheer

5.1 Inleiding

Het is uiteraard verstandig om waar mogelijk 'het wiel niet zelf uit te vinden', maar gebruik te maken van reeds bestaande oplossingen. Dat databeheer in enigerlei vorm binnen alle sectoren en organisaties plaatsvindt, is daarbij natuurlijk een open deur. Datakwaliteit en QC spelen ook in het algemeen een belangrijke rol in de informatiearchitectuur en kernprocessen bij databeheer, en dat al dan niet via 3rd party en/of afzonderlijke business intelligence systems (Figuur 5, Figuur 6).

<verder uitwerken>



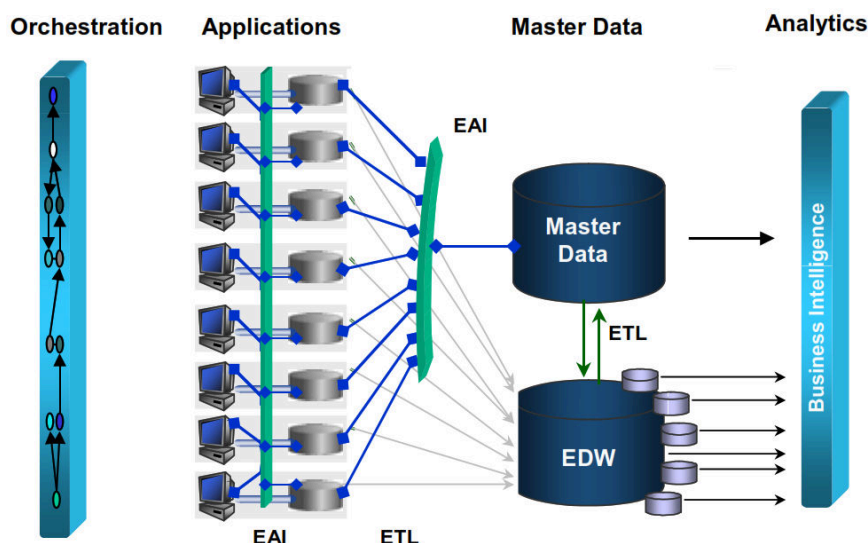
These are the key processes for any MDM system.

- Profile the master data. Understand all possible sources and the current state of data quality in each source.
- Consolidate the master data into a central repository and link it to all participating applications.
- Govern the master data. Clean it up, deduplicate it, and enrich it with information from 3rd party systems. Manage it according to business rules.
- Share it. Synchronize the central master data with enterprise business processes and the connected applications. Insure that data stays in sync across the IT landscape.
- Leverage the fact that a single version of the truth exists for all master data objects by supporting business intelligence systems and reporting.

Figuur 5: Kernprocessen voor beheer van (master)data volgens Oracle, met belangrijke rollen voor datakwaliteit en QC (Butler en Stackowiak, 2010).

5.2 Oplossingen voor versiebeheer

Ideal Information Architecture



Figuur 6: De ideale informatie-architectuur volgens Oracle (Butler en Stackowiak, 2010). QC kan binnen dit raamwerk beschouwd worden als Business Intelligence, de bijbehorende data-uitwisseling als ETL-proces.

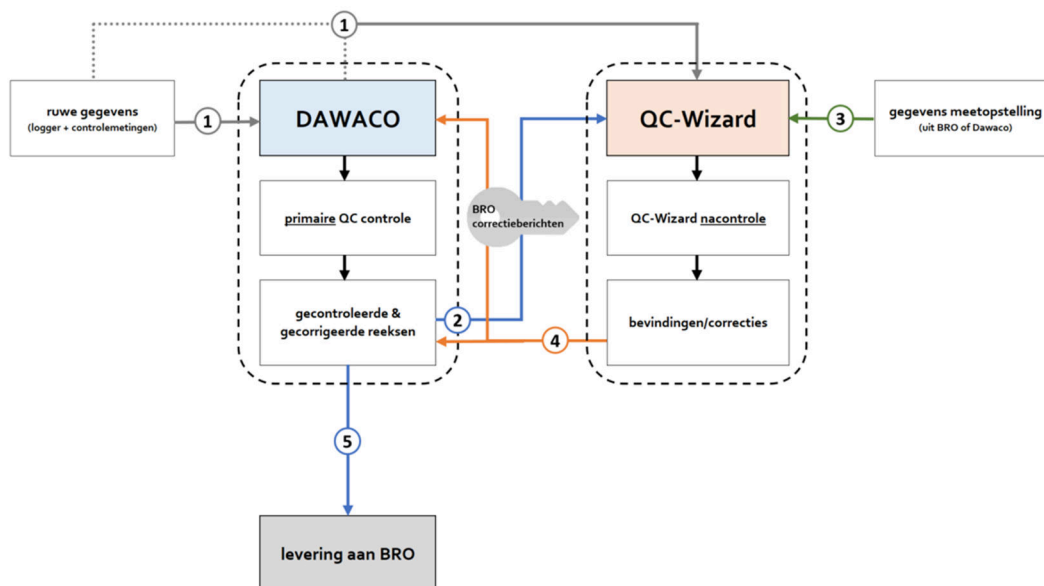
Data-uitwisseling (het uitgeven, wijzigen en weer importeren van gegevens) valt binnen het bredere raamwerk van het master data management (Butler en Stackowiak, 2010) onder de noemer Extract, Transform and Load (ETL)-proces (zie bijv. https://nl.wikipedia.org/wiki/Extraction,_Transformation_and_Load). Het zorgen voor een *single version of the truth* (het voorkomen van versieconflicten dan wel de omgang met wijzigingen en correcties) staat centraal hierin, en dat geldt uiteraard ook waar het gaat om de beoogde data-uitwisseling tussen databaseersystemen zoals Dawaco en de QC-Wizard.

Master data management heeft tegelijkertijd betrekking op zeer complexe en pluriforme data die ondergebracht worden in zogeheten Data Warehouses (zie <https://nl.wikipedia.org/wiki/Datawarehouse>), terwijl het bij het onderwerp van dit rapport gaat om gestructureerde gegevens met een (relatief) eenvoudige en uniforme data-inhoud. Voor de taak van versiebeheer voldoen in dergelijke gevallen ook meer specifieke oplossingen, die werken volgens één van de onderstaande principes:

- **Check-out/Check-in** – hierbij kan de uitgegeven subset van gegevens door één persoon tegelijk gewijzigd worden; versieconflicten worden in dat geval voorkomen door de desbetreffende gegevens in de master database tijdelijk te bevriezen.
- **Commit-and-merge** - hierbij kunnen meerdere takken of *branches* van de gegevens tegelijk afgesplitst en bewerkt worden door verschillende personen; versieconflicten worden in dat geval voorkomen doordat de software de wijzigingen en verschillen specifiek in beeld brengt waarmee de databasebeheerder die naar believen zelf kan beoordelen en kan opnemen in de master database (de stam of *trunk*, zie Figuur 8)

Voorbeelden van versiebeheerssoftware die werkt volgens het commit-and-merge-principe zijn Data Version Control (DVC), of SubVersion (SVN) en GIT als het gaat om broncode van software (wat doorgaans hoofdzakelijk tekst en dus in feite ook eenvoudige en uniforme informatie is). Het QC-protocol en de QC-Wizard gaan daarbij uit van het principe van 'databorging van bron tot BRO', wat zoveel inhoudt dat de bron-gegevens als onveranderlijke

en gedeelde basis gelden (en als zodanig ook vastgesteld dienen te worden). Het apart en specifiek opslaan en in beeld brengen en uitwisselen van correcties en wijzigingen op de gegevens, is dan ook een kernrichtlijn uit het QC-protocol en belangrijk stuk van de oplossing voor data-uitwisseling en sluitend versiebeheer (zie ook Figuur 7).



Figuur 7: 'Praatplaat' voor uitwisseling van bron-gegevens en correcties tussen Dawaco en databaseersoftware in het algemeen (bron: overleg Dawaco - QC-Wizard, De Boorder, 2021).

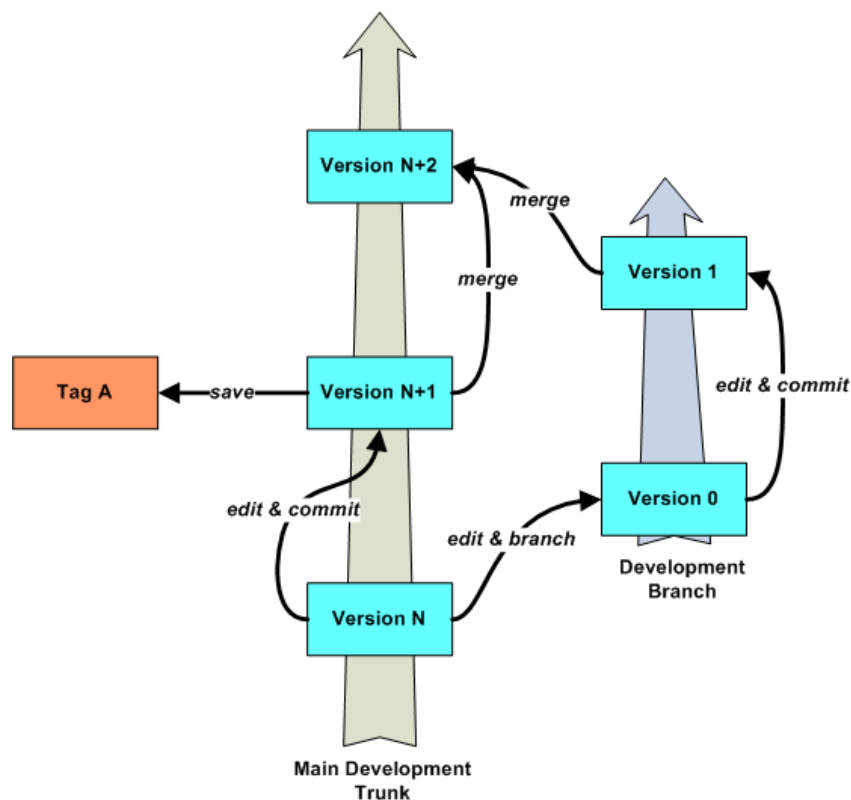
<verder uitwerken>

5.3 Scenario I: Vast versiebeheer met gedeeltelijke uitwisseling

<Er is één mastersysteem, waarmee een deel van de gegevens uitgewisseld wordt via het check in-check out-principe.>

5.4 Scenario II: Flexibel versiebeheer met volledige uitwisseling

< In principe dienen alle gegevens uitgewisseld te kunnen worden (ook de tussentijdse wijzigingen en correcties). Verschillende versies (branches) worden samengevoegd door uitwisseling en eventuele beoordeling van de verschillen via het commit-and-merge -principe.>



Figuur 8: Versiebeheer volgens het commit-and-merge-principe waarmee verschillende personen tegelijk meerdere takken of branches van de gegevens kunnen afsplitsen en bewerken (bron: <https://www.clear.rice.edu>), zie ook Data Version Control (<https://dvc.org>).

5.5 Scenario III: Flexibel en gedeeltelijk versiebeheer

< De mens bepaalt welke data de master is, deze is dus flexibel. Een alledaags en concreet voorbeeld is te vinden in de manier waarop o.a. MS Word omgaat met 'Track changes'. Behoud van alle tussentijdse wijzigingen en correcties, of het ontstaan van versieconflicten, is niet per definitie gegarandeerd, dat is aan de gebruiker.>

5.6 Scenario IV: Onafhankelijk versiebeheer met minimale uitwisseling

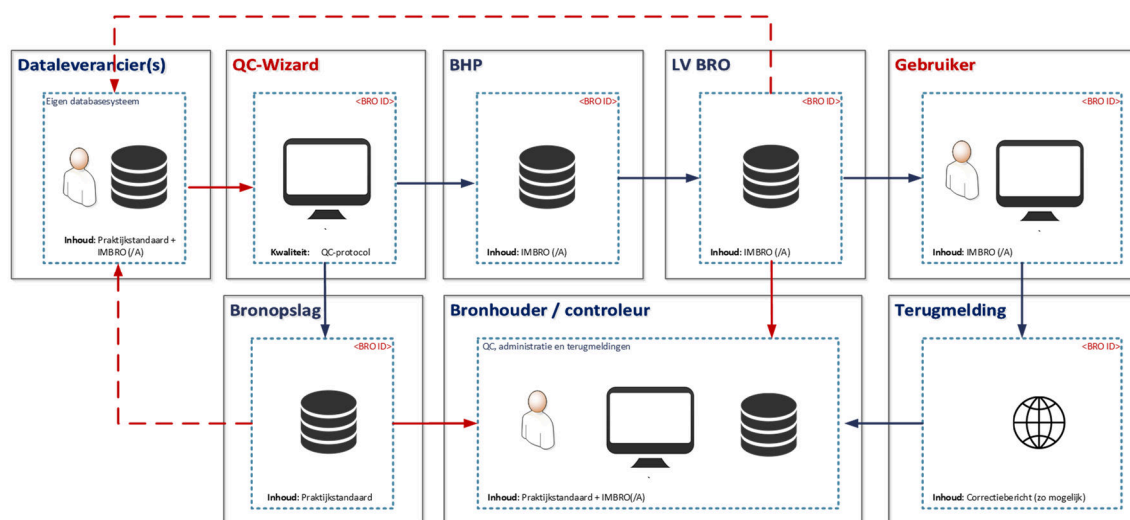
< De correcties en wijzigingen worden gereconstrueerd m.b.v. intelligentie in de software zelf (i.e. databeheer- en/of QC-software en/of versiebeheersystemen zoals DVC, SVN, GIT). De route die provincie Utrecht, RHDHV en Trefoil Hydrology samen volgen bij de migratie en QC komt in principe hierop neer.>

6 Datastroom

6.1 Inleiding

< Borging data van bron tot BRO, verder uitwerken >

6.2 Scenario I: regie QC-Wizard



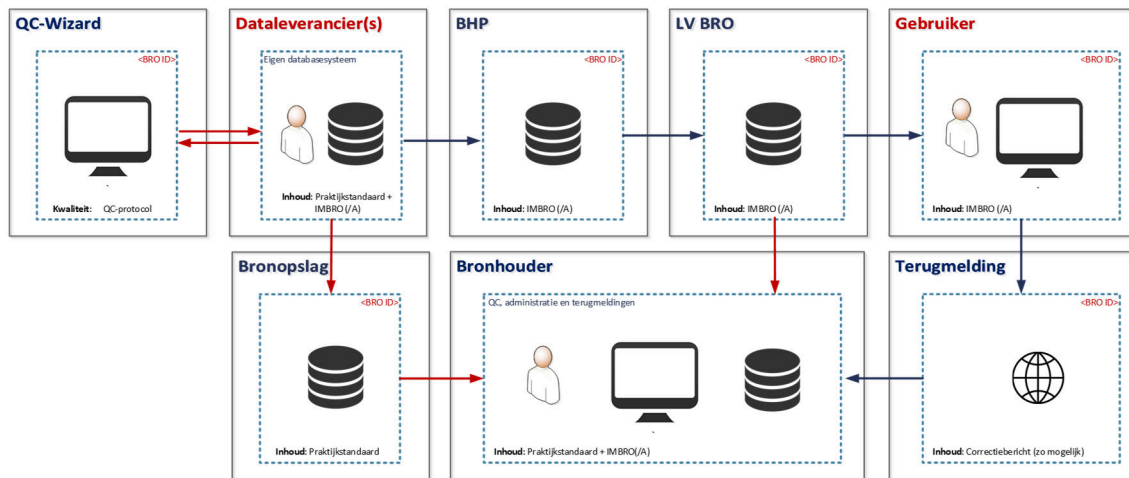
Figuur 9: Overzicht over de datastroom, rollen, systemen en architectuur in scenario I. Hier levert de QC-Wizard aan Bronhouderportaal en Bronopslag.

Aandachtspunten (rode lijnen):

1. Het beheer- of leverancierssysteem moet gegevens aanleveren aan de QC-Wizard (op een ondersteunde of nog te ondersteunen wijze).
2. Als er een beheer- of leverancierssysteem is dat up-to-date moet blijven, moet deze de aangebrachte correcties ophalen uit de BRO (en/of brondata uit de Bronopslag).
3. De bronhouder / controleur moet de leveringen bewaken en monitoren, bewaking van de kwaliteit ervan is geborgd door de QC-Wizard zelf.

< Verder uitwerken >

6.3 Scenario II: regie leverancierssysteem



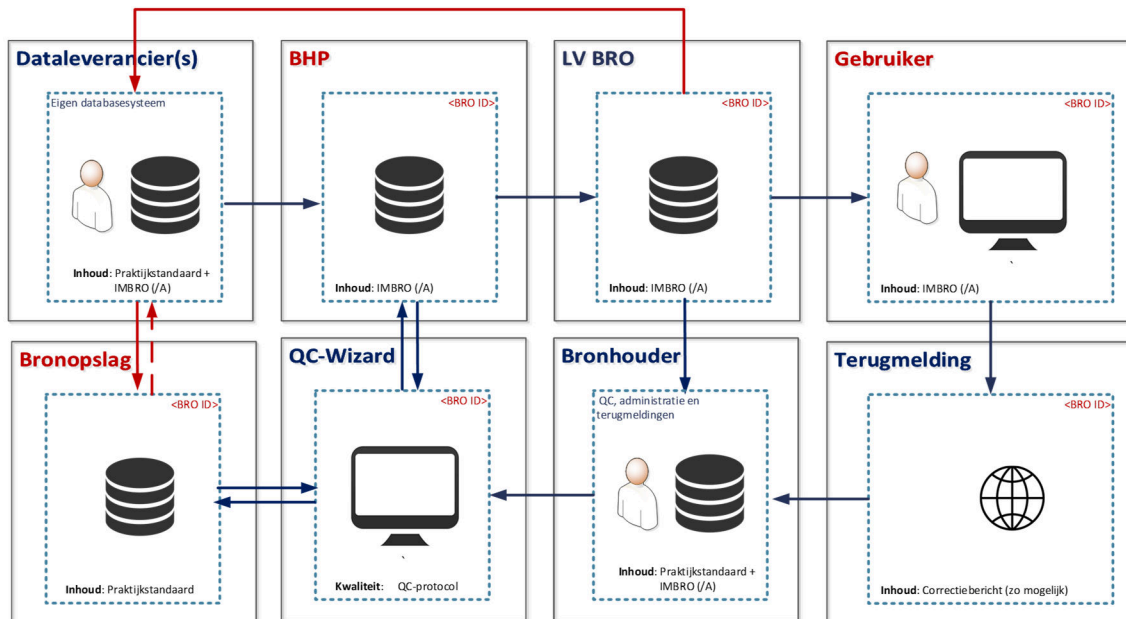
Figuur 10: Overzicht over de datastream, rollen, systemen en architectuur in scenario II. Hier levert het beheer- of leverancierssysteem aan Bronhouderportaal en Bronopslag.

Aandachtspunten (rode lijnen)

1. Het beheer- of leverancierssysteem moet gegevens aanleveren aan de QC-Wizard (op een ondersteunde of nog te ondersteunen wijze).
2. De QC-Wizard dient de aangebrachte correcties en keuringen aan te leveren aan het beheer- of leverancierssysteem.
3. De bronhouder / controleur moet de leveringen bewaken en monitoren, inclusief bewaking van de kwaliteit ervan.
4. Het beheer- of leverancierssysteem dient de aanlevering van data aan de bronopslag te ondersteunen, conform het datamodel daarvoor (op termijn de praktijkstandaard).

< Verder uitwerken >

6.4 Scenario III: regie opslag



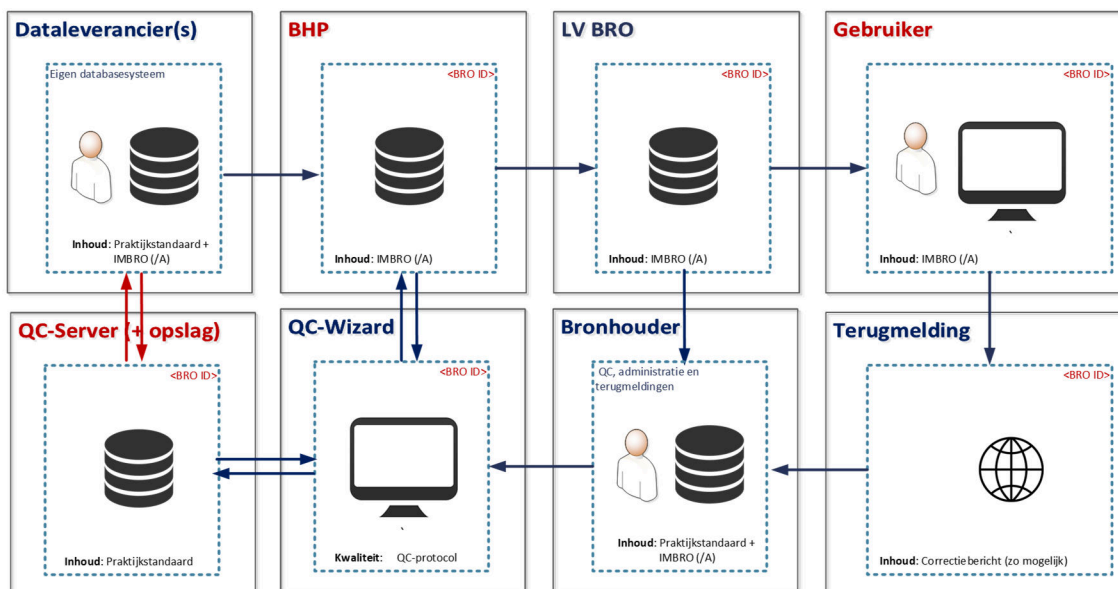
Figuur 11: Overzicht over de datastroom, rollen, systemen en architectuur in scenario III. Hier ligt de regie bij bronopslag en bronhouderportaal, zowel het beheer- als leverancierssysteem communiceren daarmee.

Aandachtspunten (rode lijnen):

1. Het beheer- of leverancierssysteem moet gegevens aanleveren aan bronhouderportaal en bronopslag, conform het datamodel daarvoor (op termijn de praktijkstandaard).
2. De bronhouder / controleur gebruikt de QC-Wizard onafhankelijk van het leverancierssysteem, via BHP en Bronopslag.
3. Om het beheer- of leverancierssysteem up-to-date te houden, moet deze de aangebrachte correcties ophalen uit de BRO (en/of brondata uit de Bronopslag).

< Verder uitwerken >

6.5 Scenario IV: Integratie



Figuur 12: Overzicht over de datastroom, rollen, systemen en architectuur in scenario III. Hier is de bronopslag vervangen door een bron- of QC-Server, die communiceert via open API's.

Aandachtspunten (rode lijnen):

1. Zowel leveranciers als de QC-Wizard kunnen QC-functies en functionaliteit (incl. opslag) toepassen en integreren in hun eigen systemen via bijv. (al dan niet gedeeltelijk) open API's op een QC-Server
2. De leverancierssystemen dienen de QC-API's te ondersteunen, conform het koppelvak en datamodel daarvoor (op termijn XML-documenten volgens de praktijkstandaard)

<Realisatie van deze optie binnen de geldende deadline van aanlevering van oktober 2021 wordt niet haalbaar geacht, verder uitwerken>

6.6 Opslag

< De opslag is generiek bedoeld, het zou ook een bestaand of nog door een willekeurige leverancier te vervaardigen systeem kunnen zijn. Dit ondanks het feit dat het platform meetnetbeheerders inmiddels ook een eigen server heeft. Verder uitwerken>

7 Fasering

7.1 Inleiding

< Toelichting verschil Soll en IST-situatie

Toelichting verschil datamigratie en operationele inwinning

Toelichting verschil voorlopige en volledige implementatie en oordeel

Toelichting Bron/A en Bron-kwaliteit

Pleidooi Learning by doing: van alleen theorieles leer je nooit autorijden!>

<verder uitwerken>

7.2 Korte termijn: BRO-migratie historie en voorlopig oordeel

< Voor aanlevering van gegevens aan zowel de BRO als de Bronopslag, is op korte termijn niet noodzakelijkerwijs ook zogeheten 'machine-to-machine' communicatie nodig. Het beheer- of leverancierssysteem kan bijv. GLD-gegevens direct en zelfstandig aan Bronhouderportaal en BRO leveren, waarna de QC-Wizard op enige termijn de beoordeling en levering van gegevens met een volledig oordeel zou kunnen verzorgen, verder uitwerken>

7.3 Langere termijn: historie en volledig oordeel

< Op langere termijn (conform het QC-protocol één jaar) is levering van gegevens met een volledig oordeel nodig. Dat geeft meer ruimte voor implementatie van het bijbehorende dataverkeer. Verder uitwerken>

7.4 Lange termijn: operationele data en volledig oordeel

< Op de lange termijn zal de inhoud en het format voor aanlevering van brongegevens moeten voldoen aan de nog te ontwikkelen praktijkstandaard. Ook de architectuur van de QC-Wizard en de bronopslag zal herzien en verder doorontwikkeld worden. Een optie daarbij is om de gegevens en praktijkstandaard Verder uitwerken>

8 Conclusies en aanbevelingen

8.1 Organisatie

- 8.1.1 Draagvlak en consensus
- 8.1.2 Overleg en besluitvorming
- 8.1.3 Koers en fasering

8.2 BRO-implementatie

- 8.2.1 Gegevenscatalogi
- 8.2.2 Data-migratie
- 8.2.3 Brondocumenten en webservices

8.3 QC- implementatie

- 8.3.1 Beoordeling en algoritmen
- 8.3.2 ICT-architectuur en fasering
- 8.3.3 Koppelingen en beheersystemen
- 8.3.4 Back-End en bouwstenen
- 8.3.5 Front end en user interfaces

8.4 Vervolg en aanbevelingen

< verder uitwerken >

Literatuur

- Butler, D. en B. Stackowiak** (2010) Master Data Management; An Oracle White Paper; Oracle Corporation, Redwood Shores, CA.
- Platform meetnetbeheerders** (2018) Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: Protocol voor datakwaliteitscontrole (QC) (versie 2.0); Rapport PMB2018, Platform meetnetbeheerders grondwaterkwantiteit provincies, Arnhem.
- Programmabureau BRO** (2019) Rollen en taken in het BRO-proces, in uw organisatie; Presentatie, Ministerie van BZK, Den Haag, www.basisregistratieondergrond.nl.
- Standaardisatieteam** (2020) Basisregistratie Ondergrond (BRO) Catalogus Grondwaterstandonderzoek, versie 1.0; Programmabureau BRO, Ministerie van BZK, Den Haag.
- Van Vliet, F., M. Harkema, C. Poulie, m.J. Van Sijl en M. Berntssen** (2021) BRO-bestendigheid ArtDiver en Objectenbeheer Vitens; Projectnummer 20.47.39, Artesia Water Research Unlimited, Schoonhoven
- Von Asmuth, J. en F. Van Vliet** (2020) Conversie datamodel ArtDiver t.b.v. de QC-Wizard; (beknopte) toelichting en documentatie (v0.9); Rapport 2020.002, Trefoil Hydrology & Artesia, Utrecht.
- Von Asmuth, J.R.** (2020) Spreadsheet- en tabelformat voor BRO-objecten; toelichting en documentatie.; Rapport 2020.004, Trefoil Hydrology, Utrecht.
- Von Asmuth, J.R. en F.C. Van Geer** (2013) Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: op weg naar een landelijke standaard Quality Control of groundwater level and groundwater head measurements: towards a national standard (in Dutch); Rapportnr. KWR 2013.027, KWR Watercycle Research Institute / TNO, Nieuwegein / Utrecht.
- von Asmuth, J.R. en E. Vonk** (2017) HydroMonitor - open data exchange format (toelichting en definitie); Report KWR 2016.062, KWR Watercycle Research Institute, Nieuwegein.

Bijlage A: Stuurgroepleden

<verder uitwerken>

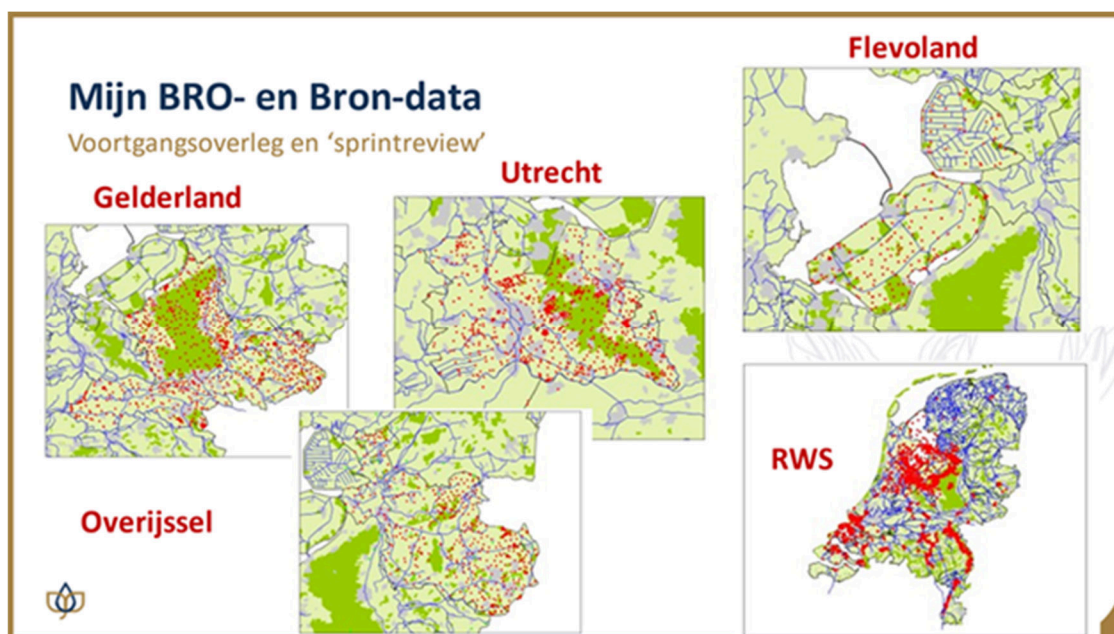
Bijlage B: Leveranciers

<verder uitwerken>

Bijlage C: Quickscan-release en disclaimer

Quickscan-release data en applicaties (met disclaimer) <bijwerken>

Verschillende provincies en RWS komen nu richting het punt dat hun (ruwe) database van BRO-putten (GMW) en meetreeksen (GLD) voldoende op orde is. Het gaat daarbij om de meetreeksen uit DINO (tot zover beschikbaar) of uit eigen databeheersysteem, en dat aangereikt met Bron-data (eveneens voor zover beschikbaar). Hiermee komen de QC-problemen die spelen beter in beeld, wat al betrekkelijk veel inzicht en houvast geeft (in de vorm van een letterlijke QC-Quickscan dus).



Figuur 13: Eenvoudige overzichtskaartjes met locaties van de putten van verschillende bronhouders op het moment van de 'Quickscan-release' van de bijbehorende BRO- en DINO- (en/of Bron)-gegevens (bron: voortgangsoverleg QC-Stuurgroep).

Disclaimer

- Het doel van de ter beschikking gestelde gegevens én van de bijbehorende QC-Wizard- en MenyanthesOS-versies is het doen van een letterlijke 'quickscan' op eventueel optredende QC-problemen, en nadrukkelijk niet meer dan dat.
- De gegevens kunnen in deze vorm vragen oproepen waar een goed antwoord op nodig is, bijvoorbeeld:
 - In de datasets kunnen meerdere putten zonder meetreeksen voorkomen. Daar speelt bij mee dat verschillende, individuele putten in DINO zijn opgesplitst in meerdere putten in de BRO. Bij de reeksen in deze data gaat het om DINO-reeksen die dus (nog) niet zijn opgeknipt, waardoor de andere opgedeelde putten (nog) geen meetreeksen bevatten. De opsplitsing betekent overigens niet per definitie dat het daadwerkelijk verschillende dan wel vervangen putten betreft. In de beoordeling en verwerking zoals die gedaan is door de provincie Overijssel zijn bijv. veel van de opsplitsingen weer ongedaan gemaakt.
- De vrijgegeven QC-Wizard- en MenyanthesOS-versies bevatten functionaliteit die nog ad hoc, onvoldoende getest en onvoldoende 'BRO-bestendig' zijn voor gebruik zonder intensieve begeleiding.

Operationeel gebruik daarvan (met name de BRO-export) wordt op dit moment nadrukkelijk nog afgeraden en is geheel voor eigen risico.

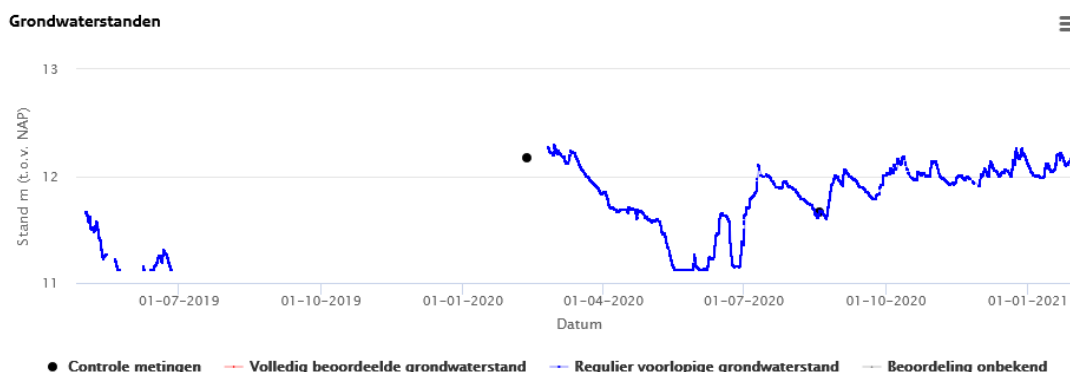
Tijdelijke ontheffing wettelijke plichten GLD

Er zijn een aantal redenen te geven waarom het voldoen aan de wettelijke plichten m.b.t. registratieobject grondwaterstandonderzoek (GLD) aan de ene kant moeilijk realiseerbaar is en aan de andere kant tot praktische problemen kan leiden. Het gaat daarbij om:

- het tijdig afgerond en op orde hebben van de onderliggende migratie van de putten (GMW);
- het tijdig afgerond en op orde hebben van de aanlevering van GLD-gegevens vanuit DINO;
- het tijdig in technisch-softwarematige en administratieve zin klaar zijn voor levering van actuele GLD-gegevens;
- de complexiteit die optreedt bij de gebruiksplicht vanwege de fragmentatie in de landelijke dataset van GLD- en GMW-gegevens, die samenhangt met de migratie en de opsplitsing van historische en actuele meetreeksen daarbij; voor aanlevering van historische gegevens staat wettelijk gezien een periode van 5 jaar, waarmee deze situatie lang aan kan houden.
- het tijdig op orde hebben van de fouten en problemen die optreden in de data, zodat de meld- en onderzoekplicht ook praktisch uitvoerbaar blijven.

Op de leverplicht die wettelijk gezien per 1 januari 2021 in zou moeten gaan is overigens al ontheffing verleend, en dat aan provincies en RWS tot 1 oktober 2021. Bij het opbouwen van de eerste datasets komen inmiddels op het oog een nog verder te onderzoeken aantal fouten en QC-problemen aan het licht. Bij tenminste een deel daarvan gaat het daarbij om consistentieproblemen (bijv. QC3c: Droogval sensor), waarvan in beginsel met zékerheid gezegd kan worden dat deze gegevens foutief zijn. Het bewust aanleveren van foutieve gegevens strookt uiteraard niet met de uitgangspunten van de meld- en onderzoekplicht uit de wet BRO.

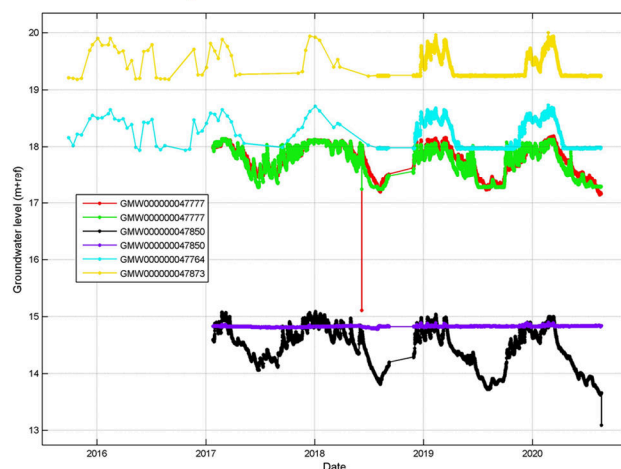
Het feit dát er QC-problemen optreden kan daarbij het migreren van de desbetreffende gegevens naar de BRO vertragen en compliceren. Dit speelt niet alleen bij GLD, maar leert ook de ervaring bij de put- of GMW-migratie. Een werkafspraken of andere tijdelijke ontheffing van de wettelijke plichten zou daarin verlichting kunnen brengen, en problemen in praktijk kunnen voorkomen.



Figuur 14: : Voorbeeld van GLD-gegevens waarin QC-problemen zichtbaar zijn (QC3c: Droogval sensor, bron: één van de eerste aan de BRO aangeleverde meetreeksen (bron: BROloket)

Gegevens Gelderland: eerste 6 reeksen (uit DINO)

Voortgang Quicksan QC-implementatie



Figuur 15: Voorbeeld van GLD-gegevens waarin QC-problemen zichtbaar zijn (o.a. QC3c: Droogval sensor, bron: Quicksan-release data Gelderland, voortgangsoverleg QC-Stuurgroep)

Dataverzameling, -conversie, -migratie en -kwaliteitscontrole

Object	Bron	BRO	Aandachtspunten
Put (GMW)	Datadump DINO Eigen data	REST-Uitgifteservice (GMW_PPO Documenten)	Gebeurtenissen deels o.b.v. 'best guess' Aanvullende data deels (nog) niet gemigreerd
Stand (GLD)	Datadump DINO Eigen data	<nog niet gemigreerd>	Combinatie bron- en volledige data Vervangen of gesplitste putten Standen t.o.v. BB of MV Afronding cm DINO Opmerkingen DINO
Net (GMN)	Niet in DINO Evt. eigen data	<niet gemigreerd>	Meetnetindeling o.b.v. aanpak Gelderland
Boring (BHR)	Datadump DINO Eigen data	<nog niet gemigreerd>	Pdf's en non-BRO-data, Keuze registratieobject (BHR-G/GT)

Tabel 2: Overzicht van registratieobjecten met gegevensbronnen en aandachtspunten bij de dataverzameling, -conversie, -migratie en -kwaliteitscontrole

<verder uitwerken>

Operationele release data en applicaties, voorlopig oordeel

< Voor bronhouders, leveranciers en controleurs die geen eigen applicatie hebben, dient de QC-Wizard geschikt gemaakt te worden voor het geven, migreren en leveren van gegevens met een voorlopig oordeel, verder uitwerken>