

Rapport 2020.004 | 12 januari 2021

Spreadsheet- en tabelformat voor BRO- objecten

Toelichting en documentatie



Trefoil Hydrology

Services, Science & Software

Spreadsheet- en tabelformat voor BRO- objecten

Toelichting en documentatie

Rapport 2020.004 | 12 januari 2021

Opdrachtnummer
20107A

Versie
v0.9 (eindconcept)

Opdrachtgever(s)
M. (Marjan) Bevelander, Programmabureau BRO

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

Met dank aan
De begeleidingsgroep (zie bijlage II)

Trefoil Hydrology

Groenedijk 36
3544 AB Utrecht

+31 6 21690206
info@3hydro.nl

www.3hydro.nl

2020 © 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	Scope-afbakening	6
1.4	Sturing en communicatie	7
1.5	Opmerkingen, suggesties en aanvullende informatie	7
1.6	Leeswijzer	7
1.7	Versiehistorie	8
2	Van hiërarchische data naar tabellen	9
2.1	‘Platslaan’ van de hiërarchie	9
2.2	Groepering van entiteiten	9
2.3	Behoud en modellering van relaties	12
3	Formatprincipes en ontwerp	13
3.1	Hoofdkenmerken	13
3.2	De <i>data exemplification</i> en <i>data exchange</i> versie	14
3.3	Datatype, eenheid en referentieniveau	15
3.4	Overige aspecten en eigenschappen	16
	Literatuur	18
	Bijlage I: Vereenvoudigde domeinmodellen	19
	Bijlage II: Stuurgroep	21
	Bijlage III: Zomertijdproblemen BRO en Microsoft (en andere software)	23

1 Inleiding

1.1 Achtergrond

In dit document worden de opzet, principes en achtergronden beschreven van een spreadsheet- en tabelformat voor gegevens van en over de registratieobjecten van de Basisregistratie Ondergrond (BRO). Dit format is tot stand gekomen in opdracht van het programmabureau BRO van het Ministerie van Binnenlandse Zaken (BZK), op initiatief en verzoek van de platforms grondwaterkwantiteit en grondwaterkwaliteit van de meetnetbeheerders van de gezamenlijke provincies. Als standaardformat voor data-uitwisseling gebruikt de BRO Extensible Markup Language (XML) van het World Wide Web Consortium. De gegevensinhoud en opzet daarvan is uitgebreid gedocumenteerd in de inname- en uitgiftehandboeken van de verschillende registratieobjecten. XML is in de basis een vorm van platte tekst, en daarmee zowel bruikbaar voor communicatie van machine tot machine, als leesbaar voor de mens. De structuur is o.a. vanwege de hiërarchie en specifieke syntaxis van elementen, attributen en de andere structuren minder overzichtelijk en lastiger te interpreteren dan gegevens in eenvoudige tabelvorm. Om die reden, maar ook voor (gestandaardiseerde) opslag in relationele databases bestond er behoefte aan het hier beschreven spreadsheet- en tabelformat voor BRO-objecten.

Voor beide platforms zijn met name de registratieobjecten in het grondwaterdomein van belang, naast de eventuele gerelateerde objecten in bodem- en ondergrond of grondwatergebruik. Mede omdat die laatste op moment van schrijven nog in ontwikkeling zijn richt dit project zich op de volgende objecten:

- Grondwatersamenstellingsonderzoek (GAR)
- Grondwatermonitoringput (GMW)
- Grondwaterstandsonderzoek (GLD)
- Grondwatermonitoringnet (GMN)
- Geologische boorbeschrijving (BHR-G)

De domeinmodellen (entiteiten en attributen) daarvan zijn terug te vinden in de respectievelijke gegevenscatalogi (Standaardisatieteam, 2017, 2019a, 2019b, 2020b, 2020a), zie ook <https://basisregistratieondergrond.nl/werken-bro/producten-diensten/standaarden/catalogi>. Andere objecten kunnen ook omgezet worden in dit spreadsheetformat volgens de hier beschreven principes en keuzes, maar vanwege het handwerk dat daarmee gemoeid is valt dat buiten dit project.

1.2 Doelstelling

Het omzetten van de gegevens en gegevenscatalogi van de BRO in spreadsheetvorm dient verschillende doelen voor verschillende stakeholders rond de BRO tegelijk. De wensen en doelen voor de spreadsheetformat(s) zijn het op een overzichtelijke, eenvoudige en gestandaardiseerde manier kunnen:

1. Presenteren van de gegevensinhoud van de registratieobjecten;
2. Bekijken, bewerken en/of uitwisselen van de gegevens via spreadsheets;
3. Vertalen van de hiërarchische datamodellen naar (relationele) tabellen.

De spreadsheets dienen daarbij qua gegevensinhoud zoveel mogelijk één-op-één aan te sluiten bij dat van de BRO. Het streven is, met andere woorden, om het volledige domeinmodel met alle bijbehorende entiteiten, attributen en waardelijsten daarin op te nemen, zodat de spreadsheets naar wens inderdaad als uitwisselformat gebruikt kunnen

worden. Het is vanwege het bovenstaande niet nodig of zinvol om de gegevensinhoud van de spreadsheets zelf hier nog eens dunnetjes te beschrijven of toe te lichten. Doelstelling van deze documentatie is daarmee:

1. Toelichten van opzet en indeling van de spreadsheets en tabellen;
2. Toelichten van de principes en keuzes die daarbij gehanteerd en gemaakt zijn.

1.3 Scope-afbakening

In de vorige paragraaf is beschreven wat de achterliggende wensen en doelstellingen zijn m.b.t. dit project. Het is echter ook goed om helder te zijn over de vraag waar de invulling van deze doelen in dit kader ophoudt. Het project voorziet wat dat betreft een zogenoemd *Minimum Viable Product* (MVP) dat op zichzelf bruikbaar en levensvatbaar is, maar laat de eventuele verdere doorontwikkeling daarvan aan de markt. Ten aanzien van het streven om qua gegevensinhoud één-op-één aan te sluiten bij de BRO geldt daarbij de afbakening:

1. Het spreadsheetformat weerspiegelt de inhoud van de domeinmodellen in de gegevenscatalogi, niet die van de inname- en uitgiftehandboeken of het berichtenverkeer.
2. Afgeleide gegevens en attributen zijn per definitie redundant en worden niet opgenomen.
3. Gegevens en attributen waarvan de BRO slechts één enkele, vaste waarde toestaat voegen geen informatie toe en worden niet opgenomen.
4. Omwille van het overzicht worden verschillende BRO-entiteiten gecombineerd in één tabel of tabblad in de spreadsheets. Deze groepering vormt een stukje gegevensinhoud dat niet in de BRO voorkomt.

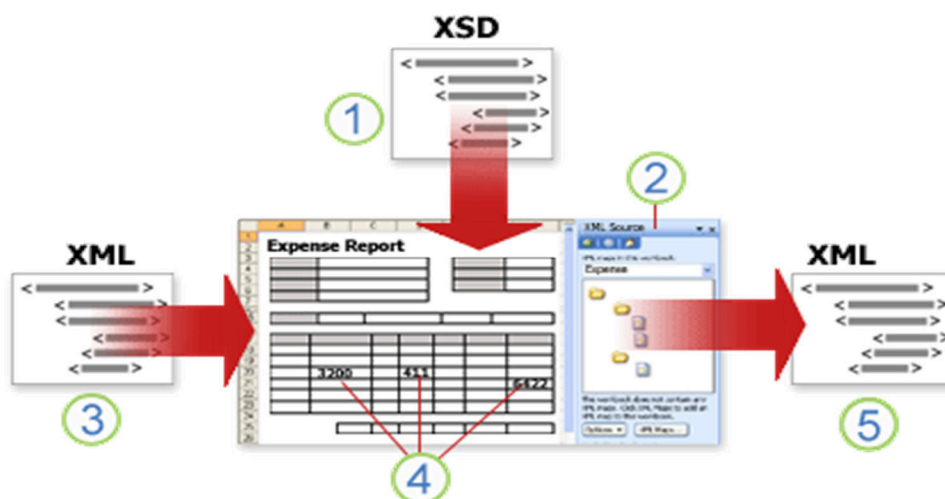
Ten aanzien van de doelstelling om de hiërarchische datamodellen te vertalen naar (relationele) tabellen geldt:

5. Het opstellen van zogenoemde DDL (*Data Definition Language*) scripts voor het aanmaken van de tabelstructuur in één of meer relationele databasesystemen valt buiten scope van dit project.

Ten aanzien van de doelstelling om gegevens te bekijken, bewerken en/of uitwisselen geldt:

6. Dit project voorziet niet in het uitwerken en opnemen van functionaliteit in de spreadsheets die deze doelstelling vereenvoudigt of ondersteunt.

Onder de laatstgenoemde functionaliteit valt bijvoorbeeld implementatie van de import en/of export van BRO-brondocumenten (XML-bestanden) en de validatie daarvan (o.a. via XSD-bestanden). Het programma Excel van Microsoft biedt de mogelijkheden daartoe wel (zie Figuur 1). Implementatie daarvan door een willekeurige marktpartij die ervaring heeft met automatisering via Microsoft Office kan interessant zijn en is relatief eenvoudig te realiseren.



Figuur 1: Illustratie van de mogelijkheden van Microsoft Excel voor import, export en validatie van XML-bestanden (bron: <https://support.office.com/en-us/article/overview-of-xml-in-excel>)

1.4 Sturing en communicatie

Dit project is zoals gezegd geïnitieerd door het platform meetnetbeheerders van de gezamenlijke provincies. Het ontwerp en de daarbij te maken keuzes zijn afgestemd met het vertegenwoordigers vanuit het platform en Rijkswaterstaat (zie bijlage I voor de samenstelling van de stuurgroep). Hiervoor zijn een tweetal overleggen georganiseerd met als onderwerp:

1. Ontwerpkeuzes en principes
2. De (concept)formats en documentatie

Communicatie over de gerealiseerde spreadsheet-formats in bredere zin vindt plaats via de BRO-website en/of nieuwsbrief, en/of door middel van presentatie en feedback tijdens een bijeenkomst van de BRO.

1.5 Opmerkingen, suggesties en aanvullende informatie

Het hier beschreven format staat open voor verdere verbetering en uitbreiding. De gerealiseerde formats zijn samen met de onderliggende documentatie beschikbaar op het GitHub-account van het platform meetnetbeheerders. De specificaties daarvan zijn:

Account: <https://github.com/PlatformMeetnetbeheerders>
Repository: BRO-Spreadsheet- en tabelformat
Description: Een spreadsheet- en tabelformat voor het bekijken en/of uitwisselen van gegevens van BRO-registratieobjecten

Heeft u opmerkingen, suggesties of behoefte aan aanvullende informatie? Neem dan contact op via Github en het account van het Platform Meetnetbeheerders, of via Jos von Asmuth (Jos.von.Asmuth@3hydro.nl, 030-6069512).

1.6 Leeswijzer

De opbouw van dit rapport is als volgt. In het eerstvolgende hoofdstuk 2 wordt ingegaan op omzetting van data met een hiërarchische structuur (zoals XML of JSON) naar data in de vorm van (relationele) tabellen. Om het aantal tabellen te beperken worden de domeinmodellen tegelijkertijd vereenvoudigd, door entiteiten te groeperen op

basis van hun kardinaliteit. Hoofdstuk 3 beschrijft vervolgens de eigenlijke opzet en ontwerp van de spreadsheets en achterliggende tabellen, en de keuzes die daarbij gemaakt worden. In de bijlagen is verdere informatie te vinden, waarbij Bijlage I de vereenvoudigde domeinmodellen bevat die de schakel vormen tussen de domeinmodellen van de BRO en de individuele tabellen in de spreadsheets of database.

1.7 Versiehistorie

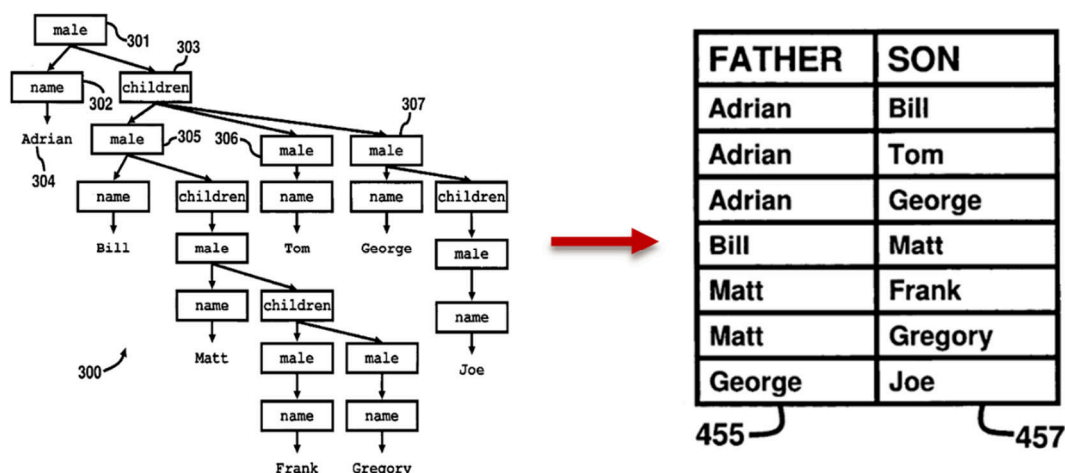
Datum	Versie	Omschrijving
Jan2021	0.9	Eindconcept documentatie en formats GAR en GMW
Sept 2020	0.8	Vervolguitwerking principes, domeinmodel GMW en conceptformats GAR, t.b.v. stuurgroepoverleg II
Juni 2020	0.5	Concept met ontwerpprincipes en achtergronden, t.b.v. stuurgroepoverleg I

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

2 Van hiërarchische data naar tabellen

2.1 'Platslaan' van de hiërarchie

Het vertalen van hiërarchische datamodellen (zoals de UML-schema's en XML-bestanden van de BRO) in (relationele) tabellen is een generiek maar niet eenvoudig oplosbaar probleem. Een kleine zoekactie in Google Scholar op de vraag 'how to convert hierarchical data to relational data' levert zo snel een lijst van 159.000 publicaties op, waaronder een groot aantal gepatenteerde methoden. US patent 2008/0172408 A1 (Meliksetian e.a., 2008) gebruikt een stamboom volgens de mannelijke lijn als illustratief voorbeeld van het 'platslaan' van een hiërarchische structuur in een tabel (Figuur 2). Naast het toegevoegde overzicht maakt het voorbeeld duidelijk dat ook relaties betekenis hebben. Bestaande, open tools lijken dit probleem zo snel ook niet erg bevredigend op te kunnen lossen (zie bijv. <https://www.convertcsv.com/xml-to-csv.htm>). De voor dit project gekozen aanpak richt zich dan ook niet zozeer op het vinden van generieke, volledig geautomatiseerde oplossingen voor dit probleem. De aanpak gaat uit van (expert)kennis van het domein, domeinmodel en werkveld, en een aantal 'handmatige' keuzes die daarbij gemaakt moeten worden.

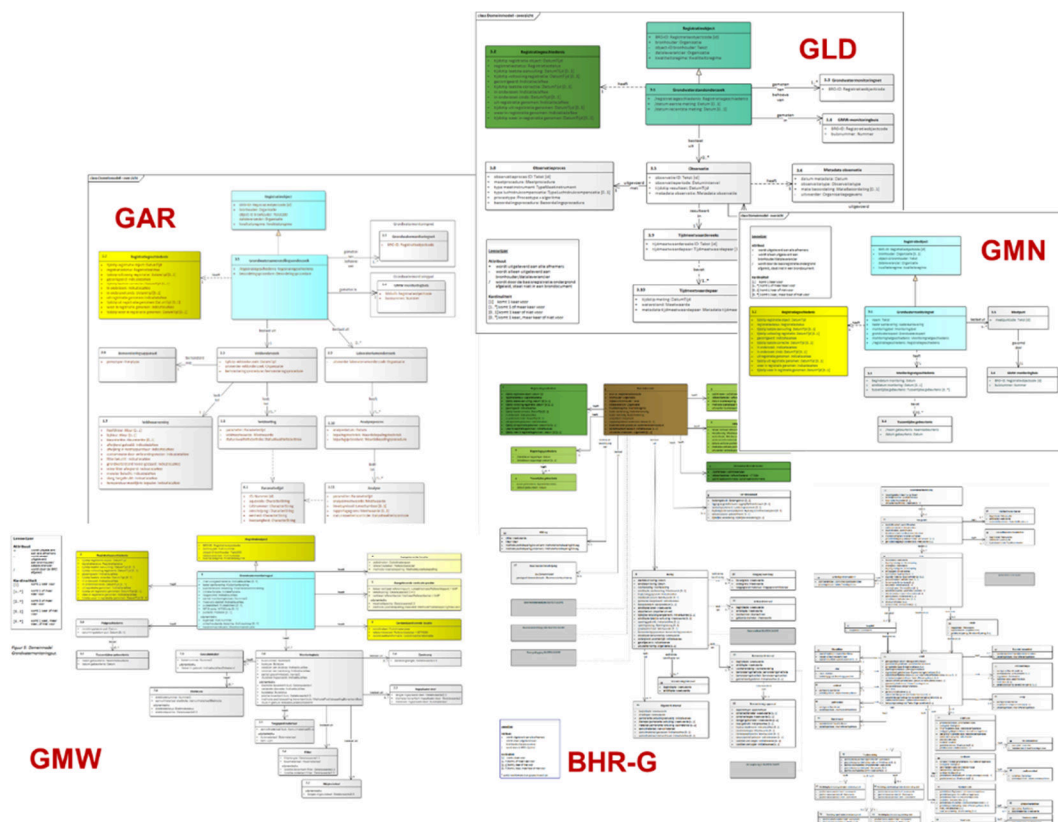


Figuur 2: Een stamboom volgens de mannelijke lijn als voorbeeld van het 'platslaan' van een hiërarchische structuur in tabellen (bron: Meliksetian e.a., 2008). Het voorbeeld illustreert het toegevoegde overzicht en laat goed zien dat ook relaties betekenis hebben.

2.2 Groepering van entiteiten

2.2.1 Vereenvoudiging domeinmodel

Zoals genoemd in paragraaf 1.2 is een belangrijk doel van de te vervaardigen formats het vergroten van het inzicht en overzicht over de gegevensinhoud van de desbetreffende BRO-registratieobjecten. Een belangrijk middel om dat doel te bereiken is het terugbrengen van het aantal elementen in de domeinmodellen (zie Figuur 3), door waar mogelijk en nuttig de verschillende elementen of entiteiten die de BRO onderscheidt te groeperen tot 'entiteitgroepen'. Nadeel van deze groepering is dat dit een stukje gegevensinhoud vormt dat als zodanig niet in de BRO voorkomt. Voordeel is dat, naast de reductie om complexiteit die op deze manier bereikt wordt, de groepering op dit onderdeel ook een kleine vertaling mogelijk maakt van de technische termen van de BRO-entiteiten naar eenvoudiger termen die zoveel mogelijk aansluiten bij de belevingswereld en processen in praktijk.



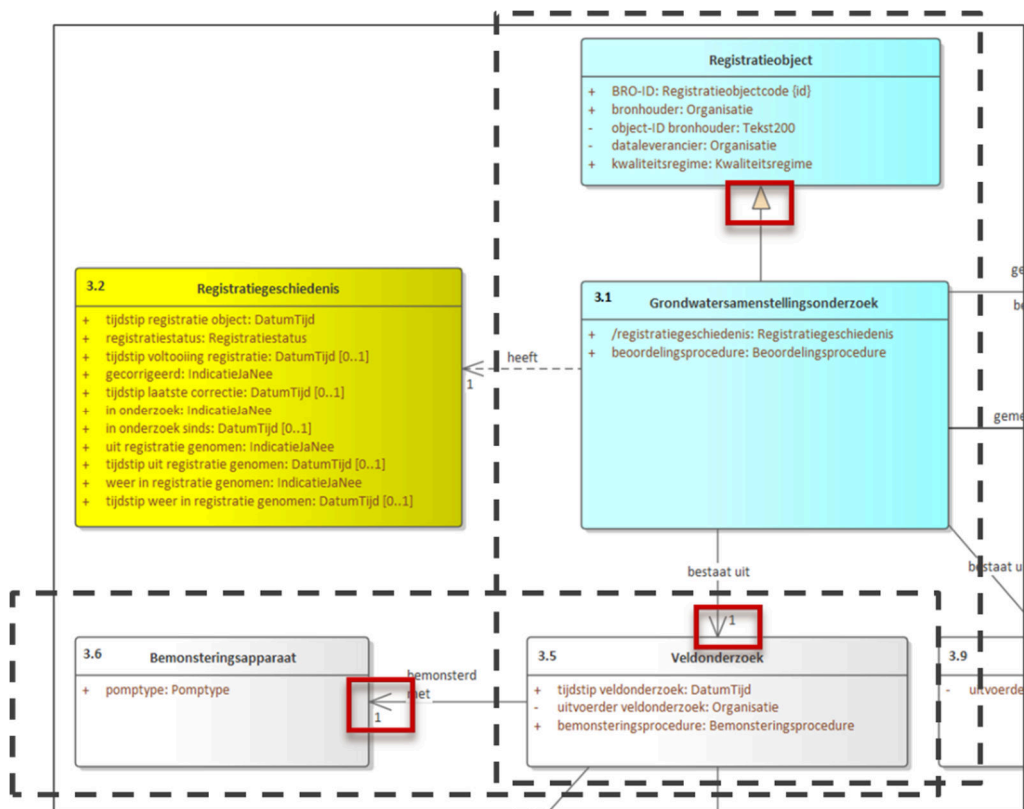
Figuur 3: Overzicht over de domeinmodellen of schema's van de verschillende voor dit project relevante BRO-registratieobjecten.

2.2.2 Rol van kardinaliteit

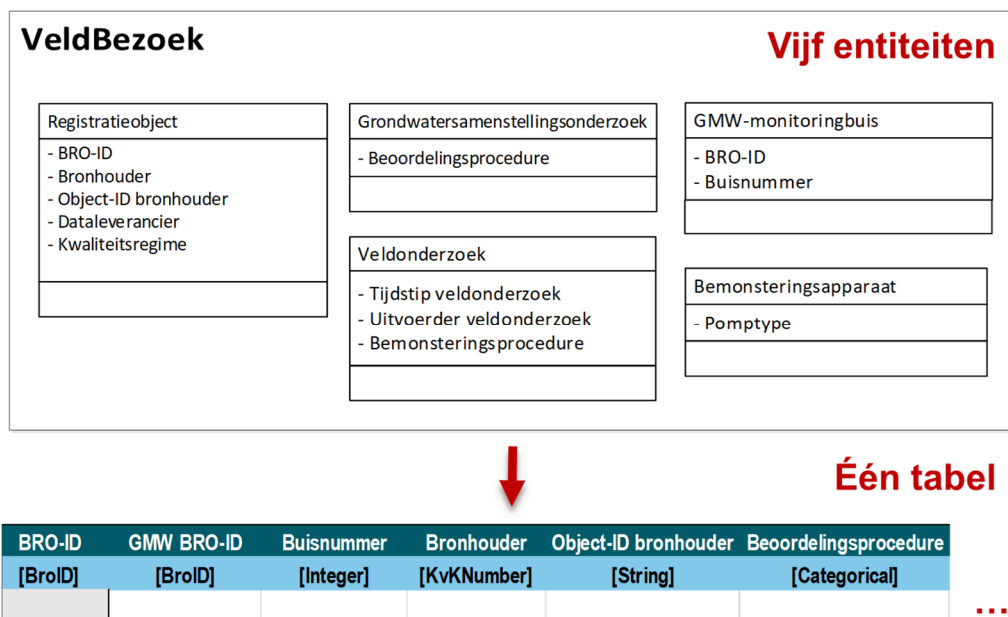
Het combineren van de gegevens van verschillende tabellen tot één tabel is een standaardoperatie voor relationele databasebeheersystemen in het algemeen. De desbetreffende operatie wordt binnen de Standard Query Language (SQL) een *Join* genoemd, waar verschillende varianten van bestaan (zie bijv. [https://nl.wikipedia.org/wiki/Join_\(SQL\)](https://nl.wikipedia.org/wiki/Join_(SQL))). Een belangrijk nadeel van het samenvoegen van tabellen is dat daardoor redundantie en/of vermenigvuldiging van het aantal gegevens en vrijheidsgraden in de datastructuur kan ontstaan, die ongewenst is en versieconflicten in de gegevens in de hand werkt.

De elementen of entiteiten in een hiërarchisch datamodel hebben een hiërarchische relatie, met als eigenschap het aantal elementen dat voor kan komen daarin (de zogenoemde kardinaliteit, zie bijv.

<https://nl.wikipedia.org/wiki/Kardinaliteit>). Bij entiteiten die een één op één relatie hebben (genoteerd als '1..1', of '0..1' indien de entiteit niet verplicht aanwezig is) geldt het bovengenoemde nadeel niet. De eigenschappen of attributen van die entiteiten kunnen vertaald worden naar kolommen in één en dezelfde tabel, zonder dat het aantal gegevens of mogelijke gegevenscombinaties daarbij vergroot wordt. Figuur 4 laat als voorbeeld een detail uit het domeinmodel van registratieobject grondwatersamenstellingsonderzoek (GAR) zien, waaruit duidelijk wordt dat op deze manier tenminste vier entiteiten gegroepeerd kunnen worden zonder dat dit de kwaliteit of structuur van het oorspronkelijke datamodel aantast. In deze figuur zijn de relaties tussen GAR en de GMW-monitoringbuis, en die tussen GAR en GMN buiten beeld. De relatie tussen GAR en de GMW-monitoringbuis vormt een vijfde entiteit in de groepering. De relatie tussen GAR en GMN beschouwen we hier ook figuurlijk als buiten beeld of buiten scope, omdat deze in principe zelf redundant en optioneel is (op moment van schrijven wordt een werkafspraken opgesteld die dit formeel bevestigt). Via de beschreven groepering ontstaat een entiteitgroep 'Veldadministratie', met als 'kolommen in de tabel' zoals gezegd de verschillende attributen van de individuele entiteiten (Figuur 5).



Figuur 4: Detail uit het domeinmodel van registratieobject grondwatersamenstellingsonderzoek (GAR), met voorbeeld van de groepering van verschillende entiteiten (zwarte stippellijn) die een 0..1 of 1..1 relatie (rode vierkanten) hebben. De entiteit Registratiegeschiedenis wordt afgeleid door de BRO en is buiten scope. Buiten beeld vormt de relatie tussen GAR en de GMW-monitoringbuis een vijfde element in deze groepering.



Figuur 5: Illustratie van de entiteitsgroep 'VeldBezoek' die ontstaat n.a.v. de in Figuur 4 weergegeven groepering op kardinaliteit. De verschillende attributen van deze entiteiten kunnen vertaald worden naar verschillende kolommen in één en dezelfde tabel.

2.3 Behoud en modellering van relaties

De in de vorige paragraaf beschreven groepering leidt tot vereenvoudigde domeinmodellen die op zichzelf hiërarchisch zijn en blijven, en die zoals gezegd waar nuttig een eigen en beter herkenbare terminologie hebben gekregen. Een overzicht van de vereenvoudigde domeinmodellen van de verschillende BRO-objecten die aldus verkregen zijn, is te vinden in bijlage I. Voor het opslaan van gegevens van de entiteitgroepen in een spreadsheet of database, is het nodig om de hiërarchische structuur daarvan op te splitsen en te verspreiden over verschillende tabellen en/of tabbladen in de spreadsheet.

De relaties tussen de entiteitgroepen worden hiermee in principe verbroken, maar ze zijn wel van belang voor de gegevens en gegevensinhoud. De relaties kunnen waar nodig behouden en gemodelleerd worden door sleutels of ID's toe te kennen of toe te voegen aan de attributen van de entiteitgroepen (bijvoorbeeld een 'Labonderzoek ID' in geval van het vereenvoudigde domeinmodel van grondwatersamenstellingsonderzoek).

Bij relationele datamodellen wordt daarbij onderscheid gemaakt in de zogenoemde *Primary Key*, die *gegevens* of regels in een tabel uniek identificeert, en *Foreign Keys*, die verwijzen naar de gegevens in een andere tabel (zie bijv. https://en.wikipedia.org/wiki/Foreign_key en/of de SQL standaard (ISO, 2016)). De 'BRO-ID' kan in principe dienst doen als *Primary Key* op het niveau van het registratieobject, ware het niet dat gegevens die nog niet daadwerkelijk geregistreerd zijn daarmee uitgesloten zouden worden. We voegen daarom ook een eigen ID of *Surrogate Key* toe aan de attributen van het registratieobject zelf (bijvoorbeeld een 'localGarID'). Omdat de laagste orde entiteitgroepen (d.w.z. entiteitgroepen onderaan de hiërarchie) geen ondergeschikte entiteiten hebben, is toevoeging van een eigen ID daar in principe niet noodzakelijk.

3 Formatprincipes en ontwerp

3.1 Hoofdkenmerken

M.b.t. de principes en ontwerp van het spreadsheet- en tabelformat kunnen we gebruik maken van beschikbaar materiaal en ervaring met een eerder spreadsheetformat dat ontwikkeld is voor hydrologische objecten (von Asmuth en Vonk, 2017). Analoog daaraan heeft ook het hier beschreven format de volgende kenmerken (Figuur 6):

- Interne definitie van het format (naam, versie, type, inhoud)
- Interne definitie van variaties in het format (datatype, data-inhoud, objecttype, locatie-ID label of sleutel)
- Interne definitie van kolommen of velden (datatype, eenheid, referentie)
- Aanvullende definities, documentatie en toelichting zijn beschikbaar via de gegeven link
- Beschikbaar in Excel-, CSV- en HDF5 -variant

The diagram illustrates the structure of the HydroMonitor - Observation Well Excel format, divided into three main sections with annotations:

- Format- en objecttype definitie:** This section defines the format and object type. It includes a table with the following data:

Format Name	HydroMonitor - open data exchange format
Format Version	1.0
Format Definition	http://www.kwrwater.nl/downloads/hydromonitor_data_exchange_format_v1.0.pdf
File Type	Excel
File Contents	Header Metadata Data
Object Type	ObservationWell
Object ID fields	Name FilterNo
- Data- definitie:** This section defines the data structure. It includes a table with the following data:

Name	NITGCode	OLGACode	FilterNo	StartDateTime	XCoordinate	YCoordinate	SurfaceLevel
[String]	[String]	[String]	[Integer]	[ExcelDate]	[m]	[m]	[m+ref]
B16B0114	B16B0114	16BP0114	1	8-10-1980 0:00	191741	538215	-0,19
B16B0114	B16B0114	16BP0114	1	8-11-1980 0:00	191741	538215	-0,197
B16B0114	B16B0114	16BP0114	3	28-1-1981 0:00	191741	538215	0,04
B16B0114	B16B0114	16BP0114	3	27-2-1981 0:00	191741	538215	-0,19
B16B0114	B16B0114	16BP0114	3	27-3-1981 0:00	191741	538215	-0,197
B16B0114	B16B0114	16BP0114	3	28-4-1981 0:00	191741	538215	-0,197
- Object-identificatie:** This section identifies the object. It includes a table with the following data:

Name	FilterNo	DateTime	Head
[String]	[Integer]	[ExcelDate]	[m+ref]
B16B0114	1	8-10-1980 0:00	-0,99
B16B0114	1	8-11-1980 0:00	-1,00
B16B0114	1	28-1-1981 0:00	-1,05
B16B0114	1	27-2-1981 0:00	-1,01
B16B0114	1	27-3-1981 0:00	-1,06
B16B0114	1	28-4-1981 0:00	-0,98
B16B0114	1	27-5-1981 0:00	-0,95

Additional annotations include 'Object-eigenschappen' (Object properties) and 'Object-(meet)gegevens' (Object (measurement) data).

Figuur 6: Voorbeeld van het HydroMonitor – Observation Well Excel format, met toelichting op de opzet en verschillende onderdelen daarvan (bron: von Asmuth en Vonk, 2017).

Qua inhoud en definitie van de attributen die voorkomen in de datadefinitie in de kop van de individuele tabellen verwijzen we hier nogmaals naar de respectievelijke gegevenscatalogi van de BRO (Standaardisatieteam, 2017, 2019a, 2019b, 2020b, 2020a) en sluiten één-op-één daarbij aan. Omwille van de eenvoud combineren we hier daarnaast het datatype en indien relevant de beoogde eenheid en referentieniveau, net als het HydroMonitor format, in de tweede regel van de kop of datadefinitie. Voor toelichting daarop zie paragraaf 3.3 en/of von Asmuth en Vonk (2017).

3.2 De data exemplification en data exchange versie

Qua gegevensinhoud sluit het spreadsheetformat zoals toegelicht in hoofdstuk 1 zoveel mogelijk één-op-één aan bij dat van de BRO. Dit is afgezien van de hier toegepaste vereenvoudiging en groepering, die een stukje gegevensinhoud vormen dat niet in de BRO voorkomt. Hier zijn met name de keuzes die mogelijk zijn en gemaakt moeten worden m.b.t. de opzet en het ontwerp van de spreadsheets en tabellen van belang. Het gaat in dit kader al met al om de volgende opties en keuzes in:

- Taalgebruik (Engels of Nederlands)
- Tabeloriëntatie (landscape of portrait)
- Opmaak en opzet (bijv. kleurgebruik en indeling)
- Groepering en termen daarvoor
- Weergave van één of meer objecten
- Elementen van de gegevenscatalogi (bijv. toelichting, definities en regels)

Voor het maken van deze keuzes kunnen algemene criteria zoals overzichtelijkheid toegepast worden, voor een ander deel is wat 'optimaal' is afhankelijk van het doel waarvoor de formats in praktijk gebruikt worden. In paragraaf 1.2 worden een drietal doelstellingen voor het format genoemd. Deze doelstellingen zijn echter terug te brengen tot twee hoofddoelen, namelijk het verhogen of verkrijgen van:

1. (Menselijk) inzicht en overzicht over de gegevensinhoud van domeinmodellen in het algemeen;
4. (Technisch) format en/of datamodel voor gegevens van specifieke objecten in het bijzonder.

Omdat het hierbij om duidelijk gescheiden doelen gaat, is het ook mogelijk om het ontwerp en de aanpak zo te kiezen dat deze het meest gunstig zijn voor één van beide. Omdat er in dat geval geen compromis nodig is, worden in het kader van dit project twee versies van het spreadsheetformat gerealiseerd, die beide een eigen aanpak en opmaak hebben die gericht is op de twee doelstellingen afzonderlijk, te weten:

Data exemplification versie:

- | | |
|-------------------------------|---|
| • Doelstelling | Inzicht en overzicht over de data-inhoud en datadefinitie |
| • Taalgebruik: | Nederlands |
| • Tabeloriëntatie: | <i>portrait</i> |
| • Aantal objecten: | één |
| • Elementen gegevenscatalogi: | uitgebreid (bijv. toelichting, definities en regels) |

De keuze voor Nederlands als taal zorgt bijvoorbeeld voor een extra stukje leesgemak voor Nederlandstaligen, de keuze voor *portrait* tabellen met slechts één object maakt het daarbij mogelijk om een groter aantal elementen uit de gegevenscatalogi op te nemen en weer te geven.

Data exchange versie:

- | | |
|-------------------------------|---|
| • Doelstelling | Uniforme en eenvoudige data-opslag en uitwisseling in tabelvorm |
| • Taalgebruik: | Engels (lowerCamelCase) |
| • Tabeloriëntatie: | landscape |
| • Aantal objecten: | meer |
| • Elementen gegevenscatalogi: | beperkt (bijv. datatype) |

Voor deze versie valt de keuze bijvoorbeeld op *landscape* tabellen om de gegevens van meerdere objecten per sheet op te kunnen nemen, en Engelstalige en zogenoemd *lowerCamelCase* attribuutnamen die gangbaar zijn voor gebruik als format en/of datamodel.

3.3 Datatype, eenheid en referentieniveau

Zoals gezegd combineren we hier het datatype en indien relevant de beoogde eenheid en referentieniveau in de tweede regel van de kop of datadefinitie (Figuur 6). Deze werkwijze en de hieronder opgenomen toelichting daarop zijn met enkele kleine verbeteringen of aanpassingen overgenomen van het HydroMonitor format (von Asmuth en Vonk, 2017). De datatypen en gebruikte notatie en terminologie hierbij zijn echter niet één-op-één toepasbaar op relationele databasesystemen, en variëren daarbij ook tussen databasesystemen onderling nog enigszins.

De datadefinitie omvat het datatype en indien relevant de beoogde eenheid en referentieniveau daarbij. Deze elementen van de datadefinitie worden gecombineerd weergegeven in een enkele cel, waarbij voor numerieke (*floating point*) gegevens waarop een eenheid en/of referentieniveau van toepassing zijn de volgende notatie wordt gebruikt:

[<teken><eenheid>+<referentieniveau>]

en voor overigens gegevens:

[<naam datatype>]

Het format kent naast numerieke gegevens met eenheid en referentieniveau de volgende datatypen:

Datatype	Omschrijving
[String]	Losse tekst
[Integer]	Heel getal (nummer of volgorde)
[%]	Percentage
[Categorical]	Element uit een vaste waarde- of categorielijst
[Date]	Datum, al dan niet met eigen notatie (zie onder)
[DateTime]	Datum en tijd, al dan niet met eigen notatie (zie onder)
[EndDateTime]	Einddatum en tijd van een periode (bijv. onttrekking), al dan niet met eigen notatie (zie onder)

M.b.t het referentieniveau worden de volgende opties en afkortingen daarvan gehanteerd:

Afkorting	Referentieniveau	Omschrijving
[ref]	ReferenceLevel	Nationaal en/of overkoepelend referentieniveau
[surflev]	SurfaceLevel	Maaiveldhoogte
[welltop]	WellTopLevel	Bovenkant van de peilbuis en/of meetpunt
[gaugetop]	GaugeTopLevel	Bovenkant van de peilschaal en/of meetpunt

Datum- en tijdnnotatie, tijdzone, zomer- en wintertijd:

Het format waarmee datums en tijdstippen getoond en/of genoteerd dienen te worden hangt af van het type bestand dat gebruikt wordt:

- **[ExcelDate]** – Het interne datumnummer van Excel, voor Excel-bestanden. De (standaard)datumweergave in Excel zelf is afhankelijk van de regionale instellingen van Windows. De datumweergave wordt standaard dus niet vastgelegd in het bestand, de weergave zelf is voor de gegevensinhoud ook niet van belang. Controleer zo nodig of een datum in Excel ook echt een datumnummer is door bijv. ‘1’ ofwel één dag bij de datum op te tellen. Excel legt in het algemeen de tijdszone niet vast en rekent ook niet om van winter- naar zomertijd of terug. We beperken ons voor dit format **anders dan de BRO** (zie ook onder) m.b.t. de tijdseenheid tot Coördinated Universal Time + 1 uur (ofwel **wintertijd (UTC+1)**). Een BRO-conforme weergave is in dat geval wel mogelijk in Excel, door als weergaveformat ‘yyyy-mm-ddThh:mm"+01:00" te gebruiken bij kolommen of cellen met datums en tijdstippen.
- [MatlabDate]** - ofwel het interne datumnummer van MATLAB voor HFD5-bestanden, eveneens anders dan de BRO beperkt tot **wintertijd (UTC+1)** als tijdseenheid.
- **[JJJJ-MM-DD] of [JJJJ-MM-DD UU:MM:SS]** - voor CSV-bestanden. De hier gekozen DatumTijd-notatie is niet conform de BRO-specificaties, **omdat Excel de BRO-notatie niet kan lezen of interpreteren**. We beperken ons ook voor dit format anders dan de BRO tot **wintertijd (UTC+1)** als tijdseenheid.

Let op: Het gebruik van zomer- en/of wintertijd kan **verwarring en problemen opleveren in zowel de BRO als de programmatuur van Microsoft**, en daarbuiten (zie 0).

3.4 Overige aspecten en eigenschappen

3.4.1 Kwaliteitsregime

Ten aanzien van het datatype, de waarden die toegestaan zijn en/of het verplicht zijn van te registreren gegevens maakt de BRO in het algemeen onderscheid in twee verschillende kwaliteitsregimes, te weten:

1. **IMBRO/A-kwaliteit** – hierbij gaat het om historische gegevens, d.w.z. gegevens die ingewonnen zijn vóór inwerkingtreding van het desbetreffende registratieobject van de BRO.
2. **IMBRO-kwaliteit** – dit regime geldt voor alle nieuw ingewonnen gegevens, d.w.z. gegevens die ingewonnen zijn ná inwerkingtreding van het desbetreffende registratieobject van de BRO

We beperken ons om praktische redenen hier tot de gegevens en gegevensdefinities met IMBRO-kwaliteit. Het onderscheid tussen IMBRO en IMRBO/A is met name t.a.v. de waardelijsten in het *data exchange* format niet eenvoudig te maken. Het relatief eenvoudige oplossing hiervoor is het realiseren van afzonderlijke versies van het *data exchange* format voor beide kwaliteitsregimes.

3.4.2 Waardelijsten

Omwille van het gemak, het inzicht en overzicht over de gegevensinhoud van de BRO én ter voorkoming van foutieve waarden zijn de waardelijsten van de registratieobjecten (volgens IMBRO-kwaliteit) in beide spreadsheetformats opgenomen als:

- **Drop down menus** – in de gegevenscellen van de desbetreffende attributen
- **Afzonderlijke tabellen** - in een aparte *Categoricals* of waardenlijsten tabblad of sheet

Deze oplossing is in die zin een uitbreiding en verbetering van het oorspronkelijke HydroMonitor-format (von Asmuth en Vonk, 2017). Het valt te overwegen om de waardelijsten in toekomstige versies van het *data exchange* format niet op te nemen als afzonderlijke tabellen en tabblad, maar hard te coderen in de drop down menus zelf.

3.4.3 Verplichte en niet-verplichte entiteiten en attributen

Wederom als uitbreiding en verbetering van het oorspronkelijke HydroMonitor-format wordt in dit format ook onderscheid gemaakt tussen verplichte en niet-verplichte entiteiten en attributen door respectievelijk het desbetreffende tabblad en de desbetreffende attribuutdefinitie groen te kleuren (Figuur 7). Alhoewel het maken van een visueel onderscheid aantrekkelijk is, is het niet mogelijk om het feitelijke onderscheid hiermee sluitend weer te geven omdat de kardinaliteit van de registratieobjecten van de BRO ingewikkelder zijn dan dat. We kiezen in het geval van een combinatie tussen een niet-verplichte entiteitgroep en wél verplichte attributen daarvan (die alleen verplicht zijn indien de entiteitgroep zelf aanwezig is) om die laatste blauw en dus als wel verplicht te kleuren.

	A	B	C	D	E	F	G
1	labResearchID	parameter	analysisMeasurementValue	limitSymbol	reportingLimit	qualityControlStatus	
2	[ID]	[ID]	[Numeric]	[Categorical]	[Numeric]	[Categorical]	
3	1	1270	5		2	goedgekeurd	
4	1	441	0.1	LT	0.1	afgekeurd	
5	1	1522	12.2			onbeslist	
6	2	4445	0.7		0.1	goedgekeurd	
7							
8							
	FieldVisit	Observations	Measurements	LabResearch	Analysis	Parameters	Categoricals

Figuur 7: Voorbeeld van het onderscheid tussen verplichte en niet-verplichte entiteiten en attributen via (groen)kleuring van de desbetreffende attribuutdefinitie of tabblad (bron: data exchange file van registratieobject GAR).

3.4.4 Dynamische attributen en gebeurtenissen

Ten aanzien van de in beginsel veranderlijke en/of uitbreidbare gegevens binnen het grondwaterdomein is er binnen de BRO voor gekozen om deze te modelleren als:

1. **Afzonderlijke registratieobjecten** – in het geval van de toevoeging van nieuwe metingen aan de meetreeksen binnen het grondwaterstand- en grondwatersamenstellingsonderzoek
2. **Tussentijdse gebeurtenissen** - in het geval van wijziging van de eigenschappen van de grondwatermonitoringput en grondwatermonitoringnet

De registratie van de gewijzigde eigenschappen en gegevens gebeurt in het laatste geval via zogeheten aanvullende brondocumenten, die gekoppeld zijn aan zogeheten tussentijdse gebeurtenissen. Het datamodel van de aanvullende brondocumenten valt echter niet onder de desbetreffende gegevenscatalogus, maar onder de berichtencatalogus. We bakenen om praktische redenen het spreadsheetformat hier af tot de gegevenscatalogus, met *eventName* en *eventDate* als attributen van de tussentijdse gebeurtenissen of *interimEvent*.

Literatuur

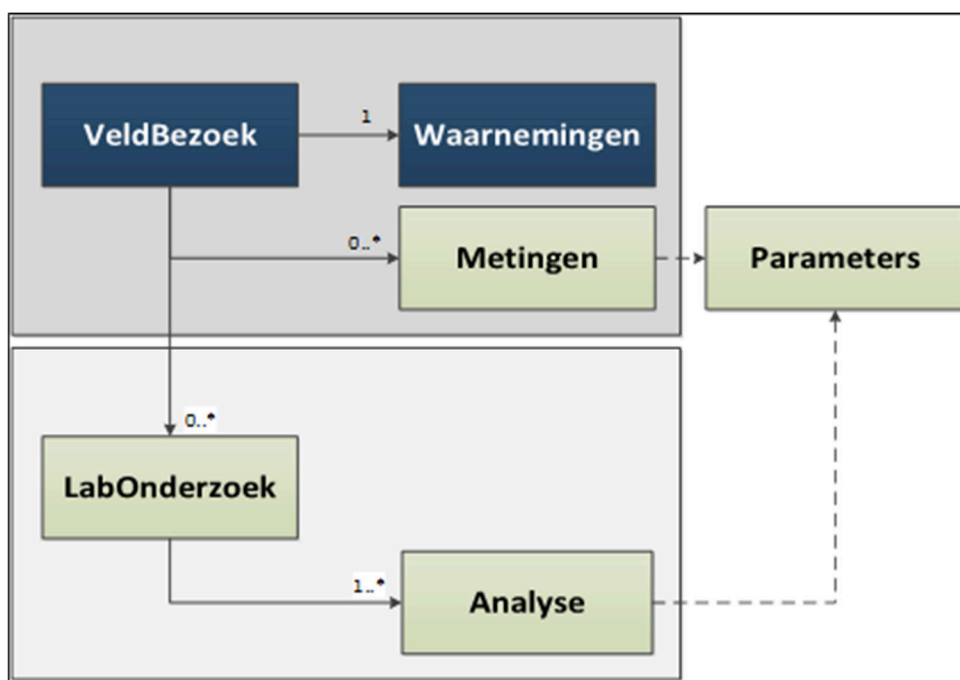
- ISO** (2016) ISO/IEC 9075-2:2016, Information technology - Database languages - SQL - Part 2: Foundation (SQL/Foundation); Technical Committee : ISO/IEC JTC 1/SC 32 Data management and interchange, ISO copyright office, Geneva, Switzerland.
- Meliksetian, D., G. Mihaila, S. Padmanabhan en N. Zhou** (2008) Converting recursive hierarchical data to relational data; Pub. No.: US 2008/0172408 A1, United States Patent Application Publication Annapolis.
- Platform meetnetbeheerders** (2018) Kwaliteitsborging grondwaterstands- en stijghoogtegegevens: Protocol voor datakwaliteitscontrole (QC) (versie 2.0); Rapport PMB2018, Platform meetnetbeheerders grondwaterkwantiteit van de gezamenlijke provincies, Arnhem.
- Standaardisatieteam** (2017) Basisregistratie Ondergrond (BRO) Catalogus Grondwatermonitoringput, versie 1.0; Programmabureau BRO, Ministerie van I&M, Den Haag.
- (2019a) Basisregistratie Ondergrond (BRO) Catalogus Grondwatermonitoringnet, versie 0.99; Programmabureau BRO, Ministerie van I&M, Den Haag.
- (2019b) Basisregistratie Ondergrond (BRO) Catalogus Grondwatersamenstellingsonderzoek, versie 0.99; Programmabureau BRO, Ministerie van BZK, Den Haag.
- (2020a) Basisregistratie Ondergrond (BRO) Catalogus Booronderzoek Geologische boormonsterbeschrijving, versie 0.99; Programmabureau BRO, Ministerie van BZK, Den Haag.
- (2020b) Basisregistratie Ondergrond (BRO) Catalogus Grondwaterstandonderzoek, versie 0.99; Programmabureau BRO, Ministerie van BZK, Den Haag.
- 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 I: Vereenvoudigde domeinmodellen

Inleiding

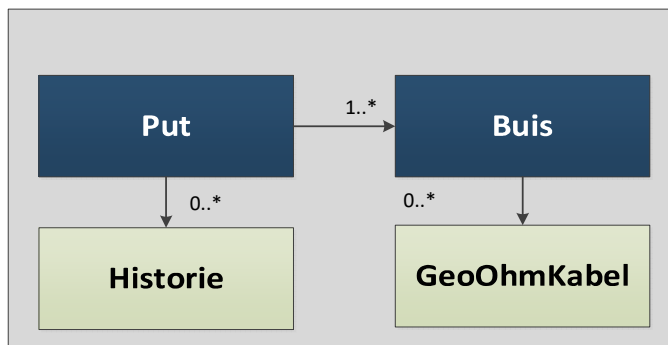
Deze bijlage bevat de vereenvoudigde en gegroepeerde domeinmodellen, zoals die tot stand zijn gekomen via het in paragraaf 2.2 beschreven proces. De entiteitgroepen die aldus ontstaan en in de schema's opgenomen zijn, worden hier niet verder toegelicht of gedefinieerd. De definitie van een entiteitgroep omvat in essentie niet meer dan de gekozen term of naam daarvoor, en de samenstellende onderdelen of onderliggende entiteiten van de registratieobjecten van de BRO zelf. Welke entiteiten bij een bepaald entiteitgroep horen is terug te vinden in de bestanden van de respectievelijke spreadsheetformats, die te vinden zijn op de GitHub-site van het platform meetnetbeheerders.

Grondwatersamenstellingsonderzoek (GAR)



Figuur 8: Vereenvoudigd domeinmodel van registratieobject grondwatersamenstellingsonderzoek (GAR), met entiteitgroepen en hun onderlinge kardinaliteit (0.., 1, 1..*, etc.). De samenstellende entiteiten per entiteitgroep zijn terug te vinden in de bestanden van de respectievelijke spreadsheetformats.*

Grondwatermonitoringput (GMW)



Figuur 9: Vereenvoudigd domeinmodel van registratieobject Grondwatermonitoringput (GMW), met entiteitgroepen en hun onderlinge kardinaliteit (0.., 1, 1..*, etc.). De samenstellende entiteiten per entiteitgroep zijn terug te vinden in de bestanden van de respectievelijke spreadsheetformats.*

De registratieobjecten :

- Grondwatermonitoringnet (GMN)
- Grondwaterstandsonderzoek (GLD)
- Geologische boorbeschrijving (BHR-G)

Konden binnen het kader van dit project niet meer op eenzelfde manier uitgewerkt worden. Het verdient aanbeveling om dat in een ander kader alsnog te doen, om aldus ook voor alle objecten binnen het grondwaterdomein de bijbehorende spreadsheetformats en tabelstructuur beschikbaar te hebben.

Bijlage II: Stuurgroep

Dit document en de bijbehorende producten zijn tot stand gekomen in opdracht van het programmabureau BRO van het Ministerie van Binnenlandse Zaken (BZK). De uitwerking is begeleid door de platforms grondwaterkwantiteit en grondwaterkwaliteit van de meetnetbeheerders van de gezamenlijke provincies, in samenwerking met Rijkswaterstaat. De volgende personen hadden zitting in de stuurgroep, die ten behoeve van dit project is geformeerd:

- N. (Nanko) de Boorder, provincie Noord-Holland (projectleider)
- H. (Henny) Kempen, provincie Gelderland
- H. (Harry) van Manen, Rijkswaterstaat WVL
- J. (Jack) van Velthuisen, Omgevingsdienst Midden- en West-Brabant
- J. (Janco) van Gelderen, provincie Utrecht en Informatiehuis Water
- C. J. (Ronnie) Hollebrandse, provincie Zeeland
- E. W. (Elroy) Crucq, provincie Zeeland
- M. (Mathijs) Oudega, provincie Fryslan
- J. (Joost) Gooijer, provincie Overijssel

Als agendalid en 'organisator achter de schermen' was betrokken:

- K. (Kor) Gerritsma, Strict, Ketenmanager BRO

Bijlage III: Zomertijdproblemen BRO en Microsoft (en andere software)

Inleiding

Het gebruik van zomer- en wintertijd kan gemakkelijk problemen en verwarring opleveren. Om een tijdsaanduiding eenduidig en correct te kunnen interpreteren, is het nodig om vast te leggen welke tijdszone en tijdseenheid bedoeld worden. Verwarring tussen zomer- en wintertijd treedt zowel op binnen de besturingssystemen en programmatuur van Microsoft (MS Windows, Excel, Outlook, etc.) als binnen de BRO. Omdat eventuele verwarring niet eenvoudig zichtbaar is, is aannemelijk dat het probleem niet tot de BRO en Microsoft beperkt is. Het verdient aanbeveling om de wijze waarop zomer- en wintertijd gebruikt worden in de verschillende hardware- en softwaresystemen die een rol spelen in de datastroom van bron tot BRO in kaart te brengen (Platform meetnetbeheerders, 2018).

BRO en bronhouderportaal

De BRO gebruikt bij tijdregistraties in principe altijd de notatie 'JJJJ-MM-DDTUU:MM:SS+UU:MM', waarbij uitgegaan wordt van Coördinated Universal Time (UTC) en een tijdzone in uren en minuten meegegeven kan worden via de laatste '+UU:MM' in de notatie. Nadeel van deze notatie is dat deze niet gangbaar is, zodat bijv. Excel de BRO-notatie niet kan lezen of interpreteren.

Let op: In het kader van dit project bleek dat deze notatie en definitie, c.q. het gebruik van de tijdszone om het verschil tussen zomer- en wintertijd aan te geven, ook van de kant van de BRO niet altijd goed en eenduidig doorgevoerd te worden. Zo ontbreekt bijv. de tijdszone in voorbeeldberichten of voorbeeldbrondocumenten van de BRO. Het is bovendien de vraag hoe de LV-BRO berichten zonder tijdszone interpreteert. Het bronhouderportaal en de LV-BRO blijken de tijdsaanduiding m.b.t. zomer- en wintertijd (al dan niet soms) ook verschillend te interpreteren. Een en ander is gemeld via de servicedesk, op moment van schrijven wordt hier verder naar gekeken (melding ID = BM20 09 00029).

Microsoft Excel en Outlook

In Microsoft Excel wordt de tijdzone c.q. het verschil tussen zomer- en wintertijd niet expliciet aangegeven. De Excel-functie '=NOW()' (of '=NU()' in de Nederlandstalige versie) vraagt bijv. de lokale computertijd op, die doorgaans in de zomer (in Nederland) op zomertijd staat. Excel past echter geen klokverschuiving toe en rekent niet om van zomer- naar wintertijd, waarmee het verder min of meer impliciet uitgaat van wintertijd (UTC+1). Gebruik van de functie '=NOW()' (of je horloge of een andere klok) levert daarmee in de zomer een foutieve tijd op (met tijdsverschuiving van één uur). Ter illustratie van dit principe is een voorbeeld uitgewerkt in Excel (Figuur 10). Het bestand '2020-09-07 Voorbeelduitwerking Excel-probleem zomer- en wintertijd.xlsx' is te vinden op de GitHub-site van het platform meetnetbeheerders.

Datum	Tijd (uur)	Datum + Tijd (getal)	Datum + Tijd (weergave)	Zomer/Winter
29-03-2020 00:00	1	43919.04	29-03-2020 01:00	Wintertijd
29-03-2020 00:00	2	43919.08	29-03-2020 02:00	Overgang zomertijd
29-03-2020 00:00	3	43919.13	29-03-2020 03:00	Zomertijd
29-03-2020 00:00	4	43919.17	29-03-2020 04:00	Zomertijd
29-03-2020 00:00	5	43919.21	29-03-2020 05:00	Zomertijd
29-03-2020 00:00	6	43919.25	29-03-2020 06:00	Zomertijd
29-03-2020 00:00	7	43919.29	29-03-2020 07:00	Zomertijd

Figuur 10: Screenshot van bestand '2020-09-07 Voorbeelduitwerking Excel-probleem zomer- en wintertijd.xlsx' in Excel, dat een voorbeeld bevat van het optreden van een foutieve tijdsverschuiving van een uur vanwege het verschil in winter- en zomertijd, omdat de lokale computertijd in zomertijd staat.

Terwijl in Microsoft Excel het verschil tussen zomer- en wintertijd dus niet expliciet wordt aangegeven, gebeurt dat in Microsoft Outlook wel, maar foutief (Figuur 11).

Subject: Spreadsheet- en tabelformat BRO-objecten: (eind)concepten

When: donderdag 10 september 2020 14:00-15:00 (UTC+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna.

Where: Microsoft Teams Meeting

Figuur 11: Screenshot van een uitnodiging in Microsoft Outlook. Outlook geeft netjes aan dat het zelf wintertijd (UTC+1) hanteert, maar herkent niet dat de computertijd in september in Nederland op zomertijd (UTC+2) staat.