



informatiehuis
WATER

inzicht voor morgen

De data-pipeline van de Aquo naar Wetterskip Fryslân

... en terug

Informatiehuis Water | 28 maart 2024



Wijnand van Riel

Data-architect Aquo
Informatiehuis Water



Marcel Adema

Data Scientist
Wetterskip Fryslân



Met **stip op 1** bij thema Datagedreven werken!



Paar noties van beoordelaars: “Dit **lijkt een saai** onderwerp. Het is wel een heel belangrijk onderwerp.” 😊 Mooie uitdaging om de presentatie dus niet saai te maken.

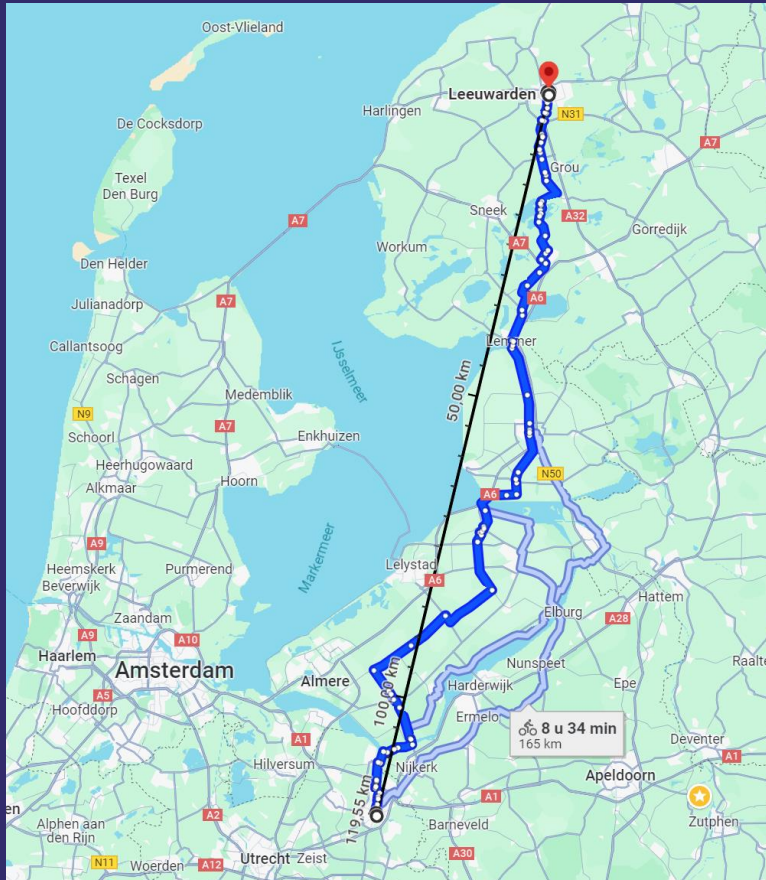


“Enige presentatie over **datakwaliteit**”.



Agenda

- Introductie (Wijnand)
- Pipeline Aquo-datalakehouse WsF (Marcel)
- Vraagstukken/incidenten vanuit het gebruik van de Aquo-API (Marcel/Wijnand)
- Van incident naar datakwaliteit en datamanagement (Wijnand)





Introductie Aquo

- Transitie Aquo
- In een klap FAIR: alles, informatieobjecten, begrippen, domeinwaarden, in één keer vindbaar, toegankelijk, uitwisselbaar, herbruikbaar
- Toegankelijkheid en uitwisselbaarheid mbv de AquoWiki-API nam zichtbaar toe. Het Wetterskip is daar een mooi voorbeeld van
- Het gebruik van de AquoWiki-API brengt een heel nieuwe dynamiek in de Aquo-standaard met zich mee.
- Bij het gebruik van API's is namelijk minder menselijke interventie aanwezig.
- Dat betekent dat aan we aan de Aquo-kant nog meer de processen op orde moeten hebben om kwaliteit en consistentie te garanderen.
- FAIR zijn is prachtig, maar zegt nauwelijks iets over datamanagement en datakwaliteit





Gebruik van de Aquo (API) data binnen WF



Waterkwaliteitsdashboard



Meerdere bronnen -> integrale data



Oplossing -> Datalakehouse (Azure)



Datalakehouse?

- Waterinfodag 2023:



Shit, ik wil een datapakhuis

12:30 - 13:15 • Dexter 11 • Jan Schotenboer | Wetterskip Fryslân, Lisette van Beusekom | Royal HaskoningDHV

Wie écht datagedreven wil werken, ontkomt er niet aan: het centraler organiseren en beheren van data en informatie ten behoeve van dashboards en rapportages. Ook het Wetterskip zag dit. Vorig jaar is gestart met het ontwikkelen van een centrale data-omgeving in Azure. Inmiddels staat het Fundament en draaien de eerste toepassingen. Een rechte lijn van ...

[Meer informatie](#)



16
maart



Datalakehouse?

You
Wat is een datalakehouse?

ChatGPT
Een "datalakehouse" is een n analyse. Het combineert de f geïntegreerde oplossing.

Een datalake is een opslagpl formaten en van verschillenc van grote hoeveelheden geg

You
En in 1 zin?

ChatGPT
Een datalakehouse is een geïntegreerde gegevensarchitectuur die zowel de flexibiliteit van een datalake als de analytische kracht van een datawarehouse combineert, waardoor organisaties gestructureerde en ongestructureerde gegevens kunnen opslaan, integreren en analyseren binnen één platform.

Aan de andere kant is een da gegevens die zijn georganise wordt vaak gebruikt voor het geoptimaliseerd voor rappor

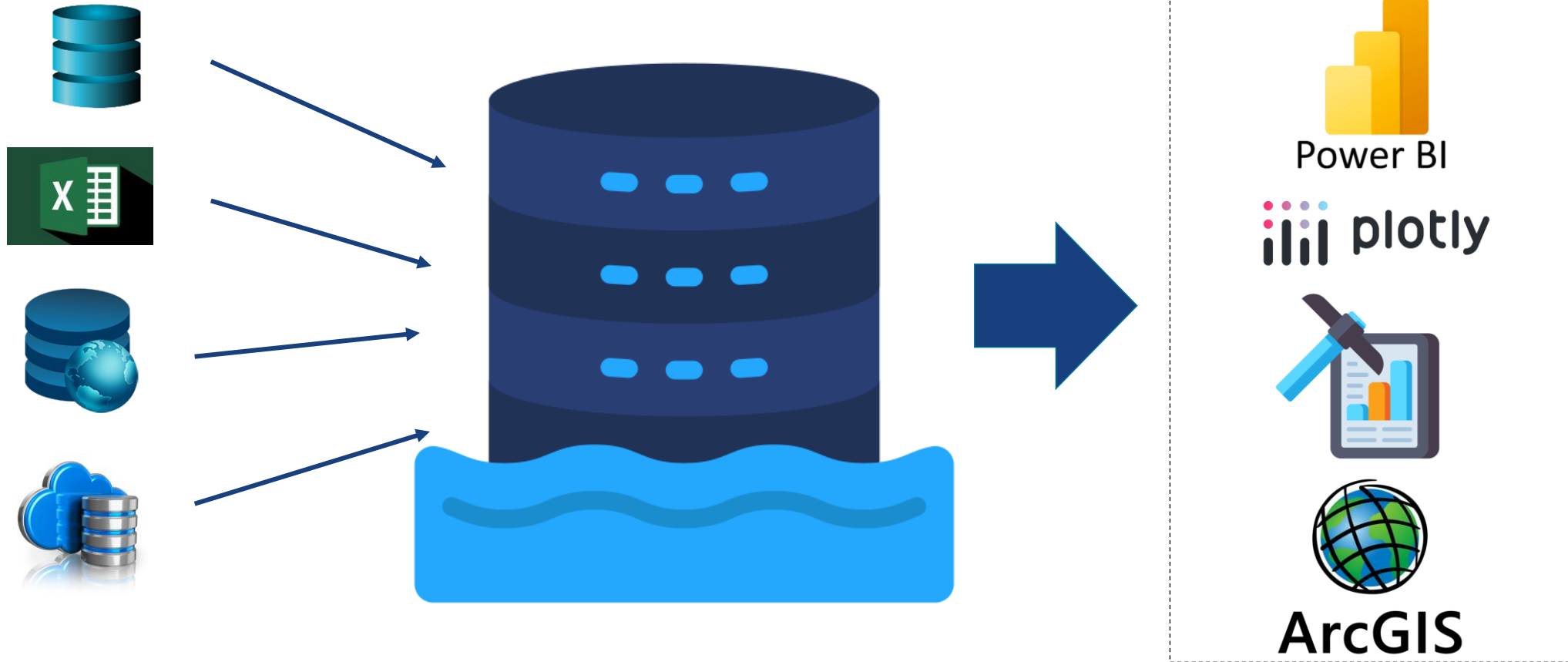
Een datalakehouse combine geïntegreerde omgeving te l gegevens kunnen worden op

organisaties zowel traditionele analytische taken als meer geavanceerde analyses, zoals machine learning en kunstmatige intelligentie, uitvoeren op één enkel platform.

Het concept van een datalakehouse is nog in ontwikkeling en evolueert voortdurend naarmate organisaties streven naar meer geïntegreerde en schaalbare oplossingen voor gegevensbeheer en analyse.



Datalakehouse?



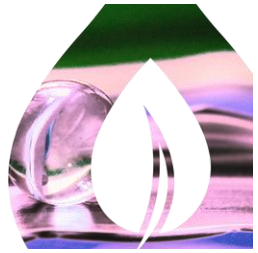
Data

Data-integratie
in Datalakehouse

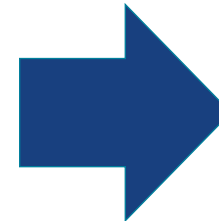
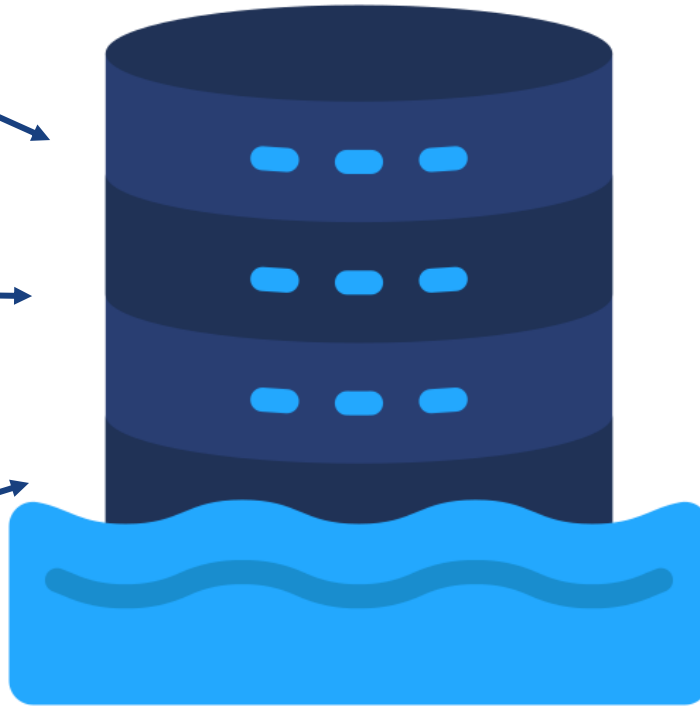
360° Visualisatie



Datalakehouse - waterkwaliteit



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat



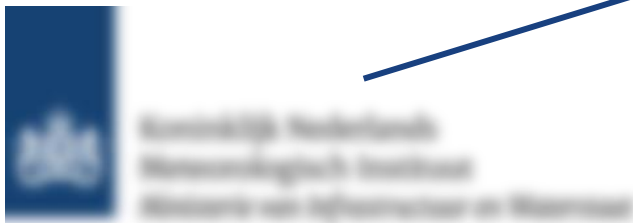
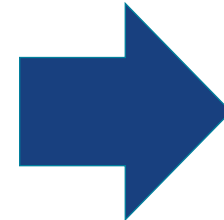
Data

Data integratie

360° Visualisatie

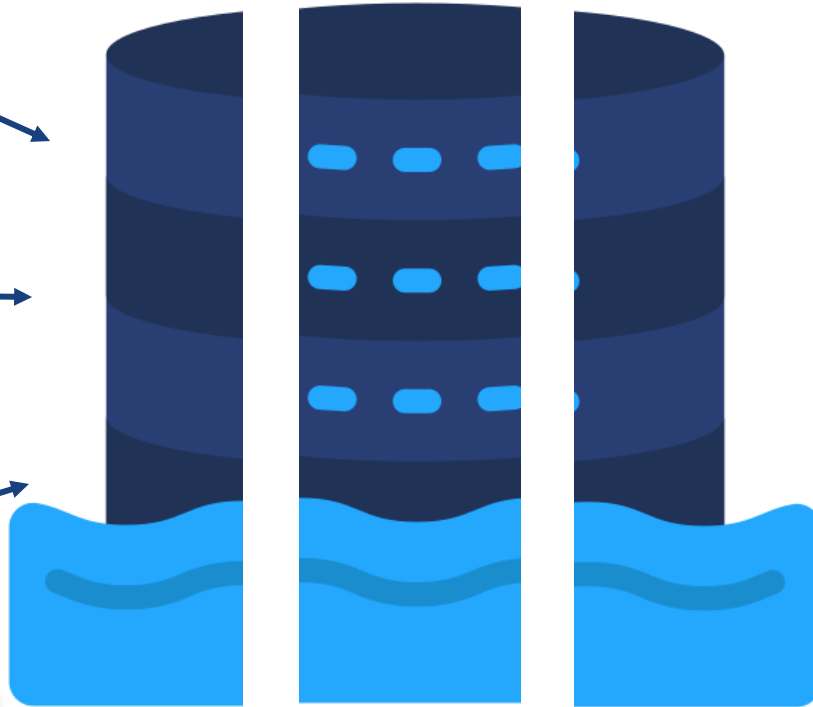
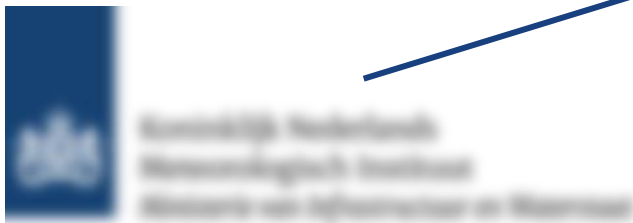


Datalakehouse - waterkwaliteit





Datalakehouse - waterkwaliteit

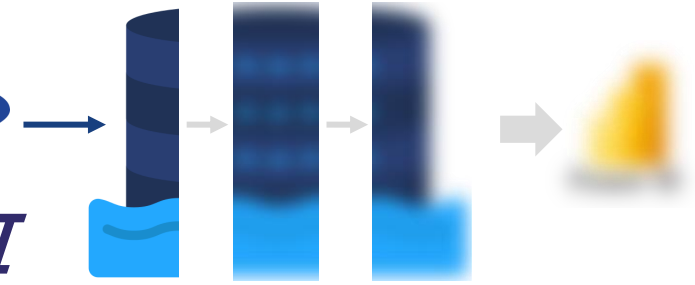


1: Ophalen data Aquo API

2: Transformeren API data

3: Maken dataproduct

4: Data visualiseren in Power BI



Stap 1: Ophalen data Aquo API

Accessing-MediaWiki-Aquo-Code-Examples



`Breder::Id-ecbe0e11-3cdc-4661-903f-f9382b32a343` : Zoekt naar een specifieke eigenschap met een identifier.

Hier is een ontleding van een voorbeeld-URL en de componenten: <https://www.aquo.nl/index.php?title=Speciaal:Vragen&q=%5B%5BCategorie%3ADomeinwaarden%5D%5D+%5B%5BBreder%3A%3AId-ecbe0e11-3cdc-4661-903f-f9382b32a343%5D%5B%5BBegin+geldigheid%3A%3A%3E1+januari+2000%5D%5D%5B%5BEind+geldigheid%3A%3A%3E1+juli+2020%5D%5D%5B%5Bwijzigingsdatum%3A%3A%3E1+juli+2021%5D%5D%5B%5BStatus%3A%3AG>

Parameter

Compartment

Waarnemingssoort

Eenheid

Hoedanigheid

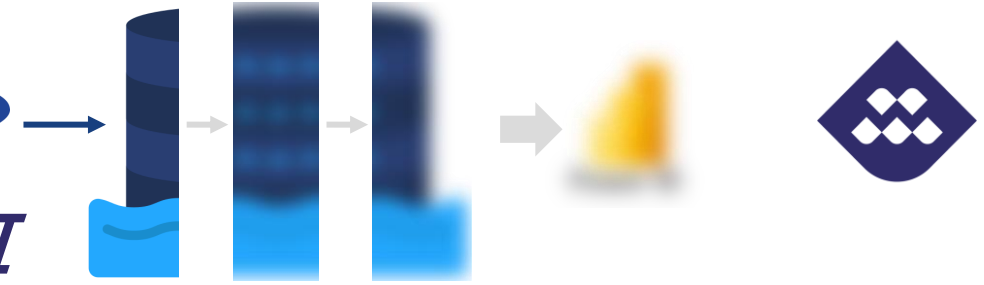
- `Begin geldigheid::>1 januari 2000` : Filtert resultaten die een "Begin geldigheid" hebben na 1 januari 2000.
- `Eind geldigheid::>1 juli 2020` : Filtert resultaten die een "Eind geldigheid" hebben na 1 juli 2020.
- `wijzigingsdatum::>1 juli 2021` : Filtert resultaten met een wijzigingsdatum na 1 juli 2021.

◦ `title=Speciaal:Vragen` : Specificeert dat het een speciale pagina-aanvraag is voor "Vragen" (Ask).

3. Query Parameters (&):

◦ `q=` : De zoekquery parameter. De waarde die volgt, is de eigenlijke zoekopdracht, geëncodeerd voor gebruik in een URL.

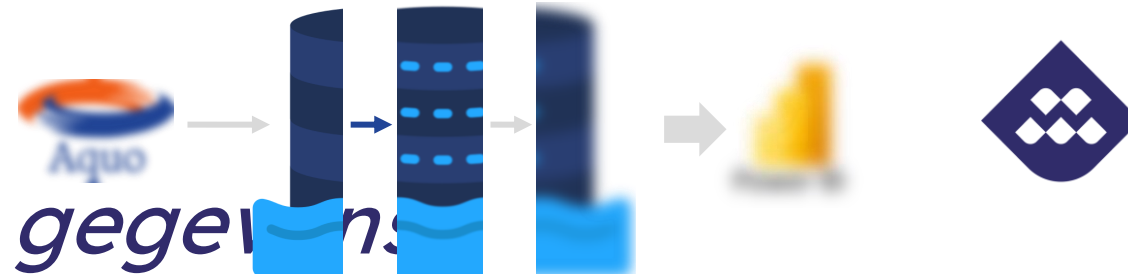
- `%5B%5B` : URL-encoding voor `[[` , wat een begin van een semantische zoekquery in MediaWiki markeert.
- `Categorie:Domeinwaarden` : Zoekt naar pagina's in de categorie "Domeinwaarden".
- `%5D%5D` : URL-encoding voor `]]` , wat het einde van een zoekparameter markeert.
- `Breder::Id-ecbe0e11-3cdc-4661-903f-f9382b32a343` : Zoekt naar een specifieke eigenschap met een



Stap 1: Ophalen data Aquo API

← → ∨ ↑ root > LANDING > extern > aquo > waarnemingssoort

Naam	Toegangslaag	Toegangslaag laatst gewijzigd	Laatst gewijzigd	Blobtype	Inhoudstype	Grootte	Status	Verwijc
waarnemingssoort_0.csv	Dynamisch (afgeleid)		3/24/2024 10:36 AM	Blok-blob	application/octet-stream	175.12 KiB	Actief	
waarnemingssoort_1000.csv	Dynamisch (afgeleid)		3/24/2024 10:39 AM	Blok-blob	application/octet-stream	179.35 KiB	Actief	
waarnemingssoort_10000.csv	Dynamisch (afgeleid)		3/24/2024 11:05 AM	Blok-blob	application/octet-stream	177.68 KiB	Actief	
waarnemingssoort_10500.csv	Dynamisch (afgeleid)		3/24/2024 11:07 AM	Blok-blob	application/octet-stream	178.18 KiB	Actief	
waarnemingssoort_11000.csv	Dynamisch (afgeleid)		3/24/2024 11:09 AM	Blok-blob	application/octet-stream	177.89 KiB	Actief	
waarnemingssoort_11500.csv	Dynamisch (afgeleid)		3/24/2024 11:10 AM	Blok-blob	application/octet-stream	178.78 KiB	Actief	
waarnemingssoort_12000.csv	Dynamisch (afgeleid)		3/24/2024 11:13 AM	Blok-blob	application/octet-stream	179.05 KiB	Actief	
waarnemingssoort_12500.csv	Dynamisch (afgeleid)		3/24/2024 11:14 AM	Blok-blob	application/octet-stream	177.69 KiB	Actief	
waarnemingssoort_13000.csv	Dynamisch (afgeleid)		3/24/2024 11:16 AM	Blok-blob	application/octet-stream	161.61 KiB	Actief	
waarnemingssoort_13500.csv	Dynamisch (afgeleid)		3/24/2024 11:17 AM	Blok-blob	application/octet-stream	41.99 KiB	Actief	
waarnemingssoort_1500.csv	Dynamisch (afgeleid)		3/24/2024 10:40 AM	Blok-blob	application/octet-stream	177.86 KiB	Actief	
waarnemingssoort_2000.csv	Dynamisch (afgeleid)		3/24/2024 10:42 AM	Blok-blob	application/octet-stream	178.42 KiB	Actief	
waarnemingssoort_2500.csv	Dynamisch (afgeleid)		3/24/2024 10:44 AM	Blok-blob	application/octet-stream	177.82 KiB	Actief	
waarnemingssoort_3000.csv	Dynamisch (afgeleid)		3/24/2024 10:45 AM	Blok-blob	application/octet-stream	177.84 KiB	Actief	
waarnemingssoort_3500.csv	Dynamisch (afgeleid)		3/24/2024 10:46 AM	Blok-blob	application/octet-stream	177.60 KiB	Actief	
waarnemingssoort_4000.csv	Dynamisch (afgeleid)		3/24/2024 10:48 AM	Blok-blob	application/octet-stream	178.59 KiB	Actief	
waarnemingssoort_4500.csv	Dynamisch (afgeleid)		3/24/2024 10:50 AM	Blok-blob	application/octet-stream	177.83 KiB	Actief	
waarnemingssoort_500.csv	Dynamisch (afgeleid)		3/24/2024 10:38 AM	Blok-blob	application/octet-stream	177.42 KiB	Actief	
waarnemingssoort_5000.csv	Dynamisch (afgeleid)		3/24/2024 10:51 AM	Blok-blob	application/octet-stream	178.04 KiB	Actief	
waarnemingssoort_5500.csv	Dynamisch (afgeleid)		3/24/2024 10:52 AM	Blok-blob	application/octet-stream	178.85 KiB	Actief	
waarnemingssoort_6000.csv	Dynamisch (afgeleid)		3/24/2024 10:54 AM	Blok-blob	application/octet-stream	178.92 KiB	Actief	
waarnemingssoort_6500.csv	Dynamisch (afgeleid)		3/24/2024 10:55 AM	Blok-blob	application/octet-stream	178.87 KiB	Actief	
waarnemingssoort_7000.csv	Dynamisch (afgeleid)		3/24/2024 10:57 AM	Blok-blob	application/octet-stream	179.07 KiB	Actief	
waarnemingssoort_7500.csv	Dynamisch (afgeleid)		3/24/2024 10:58 AM	Blok-blob	application/octet-stream	178.75 KiB	Actief	
waarnemingssoort_8000.csv	Dynamisch (afgeleid)		3/24/2024 11:00 AM	Blok-blob	application/octet-stream	179.51 KiB	Actief	
waarnemingssoort_8500.csv	Dynamisch (afgeleid)		3/24/2024 11:01 AM	Blok-blob	application/octet-stream	160.37 KiB	Actief	
waarnemingssoort_9000.csv	Dynamisch (afgeleid)		3/24/2024 11:02 AM	Blok-blob	application/octet-stream	178.84 KiB	Actief	
waarnemingssoort_9500.csv	Dynamisch (afgeleid)		3/24/2024 11:04 AM	Blok-blob	application/octet-stream	178.05 KiB	Actief	



Stap 2: Transformeren API gegevens

- Mapping van kolomnamen, “Codes” -> “id”
- Selectie van kolommen

Notebooks

- ▶ experiments
- ▶ extern
- ▶ aquo
 - aquo_compartiment_landing_to_raw
 - aquo_compartiment_raw_to_base
 - aquo_eenheid_landing_to_raw
 - aquo_hoedanigheid_landing_to_raw
 - aquo_parameter_landing_to_raw
 - aquo_parameter_raw_to_base
 - aquo_parameterhierarchie_to_base
 - aquo_waarnemingssoort_landing_to_raw
 - aquo_waarnemingssoort_raw_to_base
 - aquo_watertype_to_base
 - aquo_watertype_to_raw
- ▶ aquokit
- ▶ cbs
- ▶ handelsregister
- ▶ waterkwaliteitsportaal

Not started

Blok 3: Transform

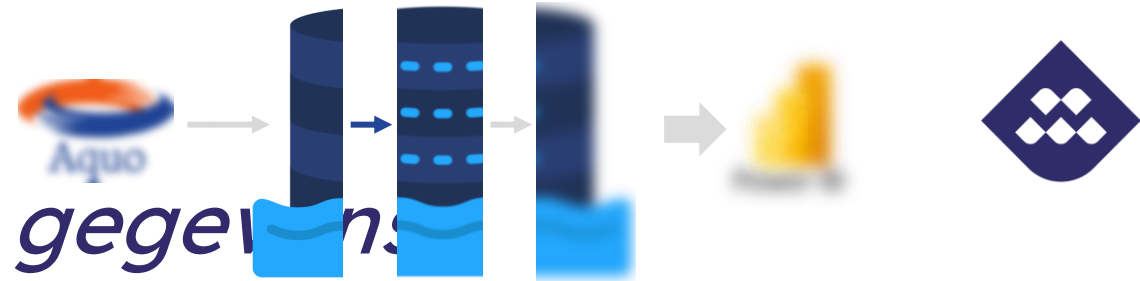
Mapping parameter

```

1 df_parameter = (df_par
2   .withColumn("_wijzigingsdatum", _HUIDIG_TIJDSTIP)
3   .withColumn("id", concat(lit('parameter_'), col('Codes')))
4   .withColumn("identificatie", col('Codes'))
5   .withColumn("titel", concat(col('Codes'), lit(' - '), col('Omschrijving')))
6   .withColumn("code", col('Codes'))
7   .withColumn("naam", col('Omschrijving'))
8   .withColumn("versie_id", concat_ws("_", lit('parameter'), col('Codes'), split(col('Gerelateerd'), "Id-")[1]))
9   .withColumn("versie_uuid", split(col('Gerelateerd'), "Id-")[1])
10  .withColumn('startdatum', convert_to_date_udf(col('Begin_geldigheid')))
11  .withColumn('einddatum', convert_to_date_udf(col('Eind_geldigheid')))
12  .withColumn('status', when((col("einddatum") > current_date()) & (col("startdatum") < current_date()), "Actief")
13    .otherwise(lit('Inactief')))
14  .withColumn('cas_nummer', col('CASnummer'))
15  .withColumn('groep', col('Groep'))
16  .select("_wijzigingsdatum", "id", "identificatie", "titel", "code", "naam", "versie_id", "versie_uuid", "startdatum", "einddatum", "status", "cas_nummer", "groep")

```

Stap 2: Transformeren API gegevens



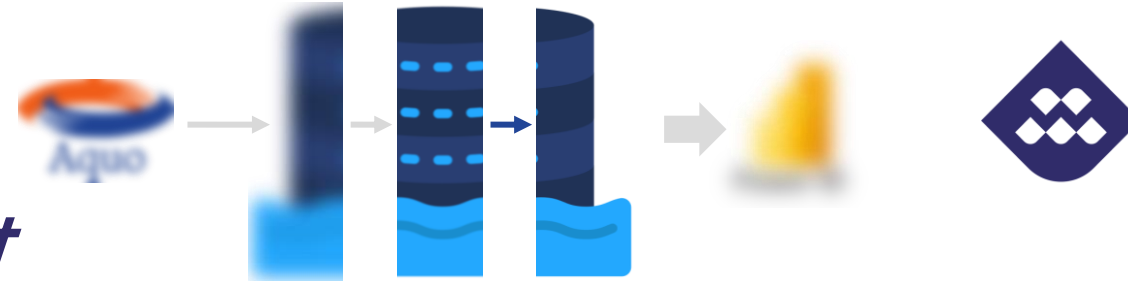
view Table Chart Export results

view	Groep	Codes	Omschrijving	CASnummer	Id	Begin_geldigheid	Eind_geldigheid	Gerelateerd	Wijzigingsdatum
Id-18475181-...	ChemischeStof	O2	zuurstof	7782-44-7	1701	8 juni 2011	1 januari 2100	undefined	13 december 2023 12:09:05



Table Chart Export results

_wijzigingsdatum	id	identificatie	titel	code	naam	versie_id	versie_uuid	startdatum	einddatum	status	cas_nummer	groep
2024-02-21 10:57:30.414	parameter_O2	O2	O2 - zuurstof	O2	zuurstof	parameter_O2	undefined	2011-06-08	2100-01-01	Actief	7782-44-7	ChemischeStof



Stap 3: Maken dataproduct

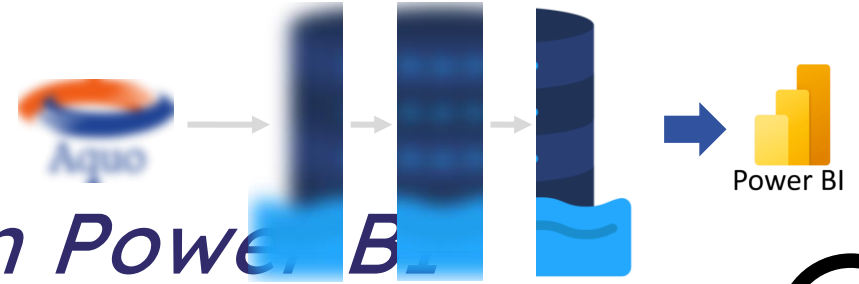
```

78 df_parameter = df_parameter.select(
79     '_wijzigingsdatum',
80     'id',
81     'identificatie',
82     'titel',
83     'code',
84     'naam',
85     'versie_id',
86     'versie_uuid',
87     'startdatum',
88     'einddatum',
89     'status',
90     'cas_nummer',
91     'groep',
92 )
93
94 df_compartiment = df_compartiment.select(
95     '_wijzigingsdatum',
96     'id',
97     'identificatie',
98     'titel',
99     'code',
100    'naam',
101    'uuid',
102    'startdatum',
103    'einddatum',
104    'status',
105    'groep',
106 )
107
108 df_waarnemingssoort= df_waarnemingssoort.select(
109     '_wijzigingsdatum',
110     'id',
111     'identificatie',
112     'titel',
113     'code',
114     'naam',
115     'uuid',
116     'startdatum',
117     'einddatum',
118     'status',
119     'parameter_id',
120     'compartiment_id',
121     'eenheid_id',
122     'hoedanigheid id',

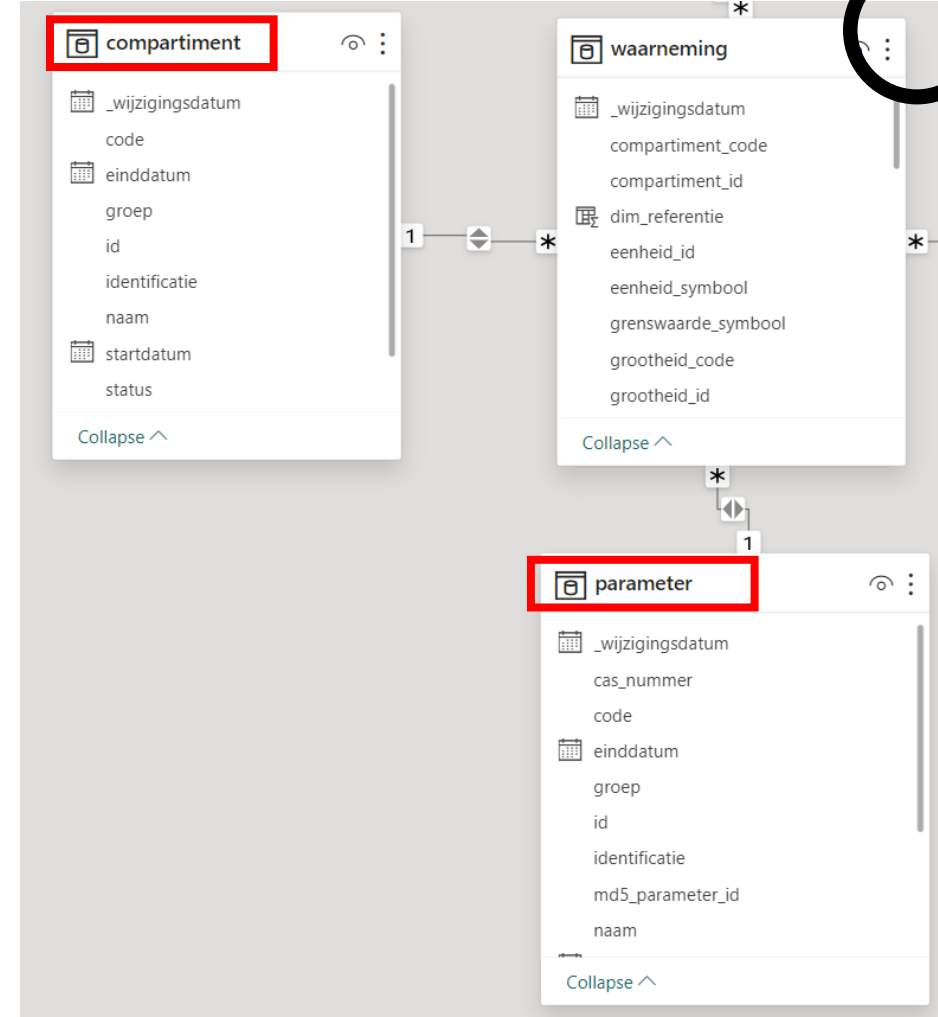
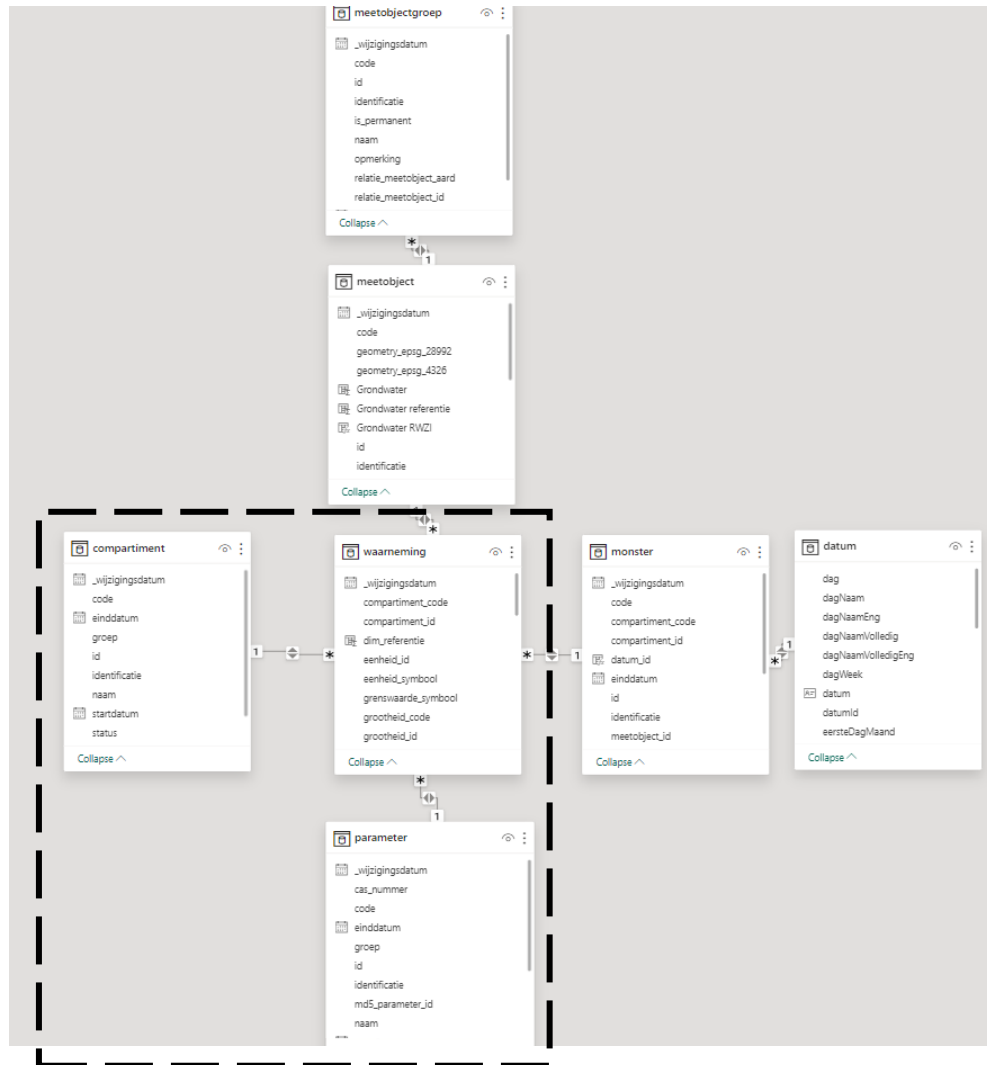
```

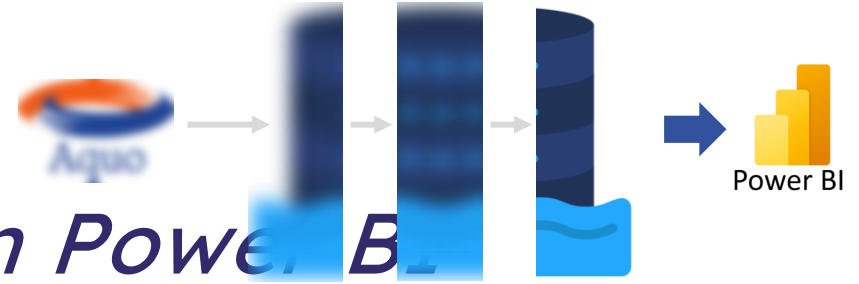
← → ∨ ↑ enriched > wetterskip > watersysteembeheer

Name	Last Modified	Content Type
compartment	9-11-2023 15:24:22	Folder
krw_monitoring	11-1-2024 11:39:59	Folder
krw_norm	12-1-2024 15:07:02	Folder
krw_waarneming	13-11-2023 16:28:08	Folder
meetobject	9-11-2023 15:22:56	Folder
meetobjectgroep	9-11-2023 15:19:25	Folder
monster	9-11-2023 15:23:01	Folder
parameter	9-11-2023 15:24:18	Folder
parameterhierarchie	13-11-2023 16:28:23	Folder
projectie_meetobject	13-11-2023 16:28:27	Folder
waarneming	9-11-2023 15:23:09	Folder
waarnemingssoort	14-2-2024 08:37:50	Folder
waterlichaam	13-11-2023 16:27:43	Folder
waterlichaam_simplified	22-11-2023 10:51:34	Folder
watertype	13-11-2023 16:28:03	Folder

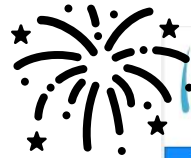


Stap 4: Data visualiseren in Power BI





Stap 4: Data visualiseren in Power BI

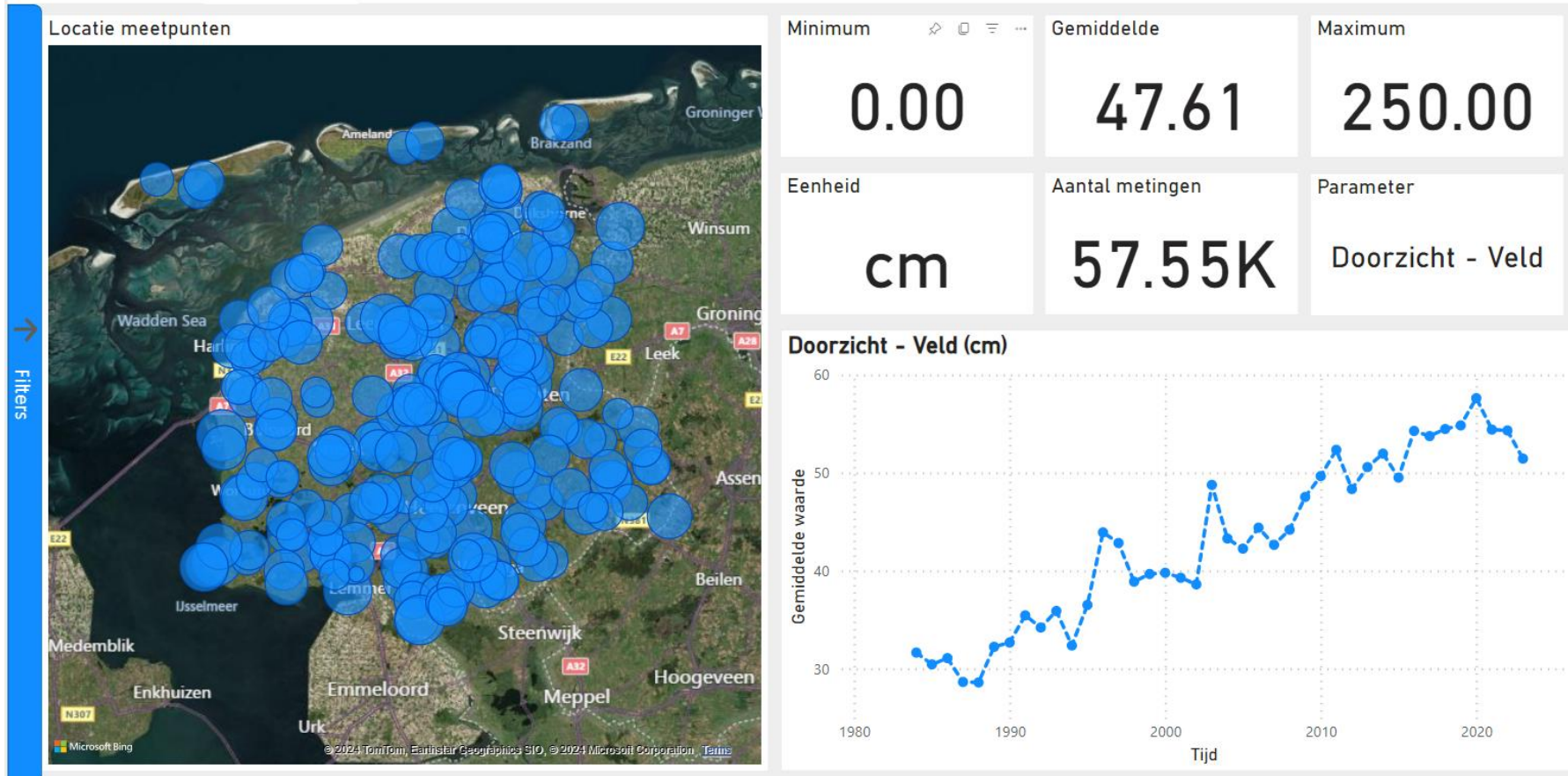


Dashboard Dawaco - Waterkwaliteit & Monitoring

V3.9 12-12-2023

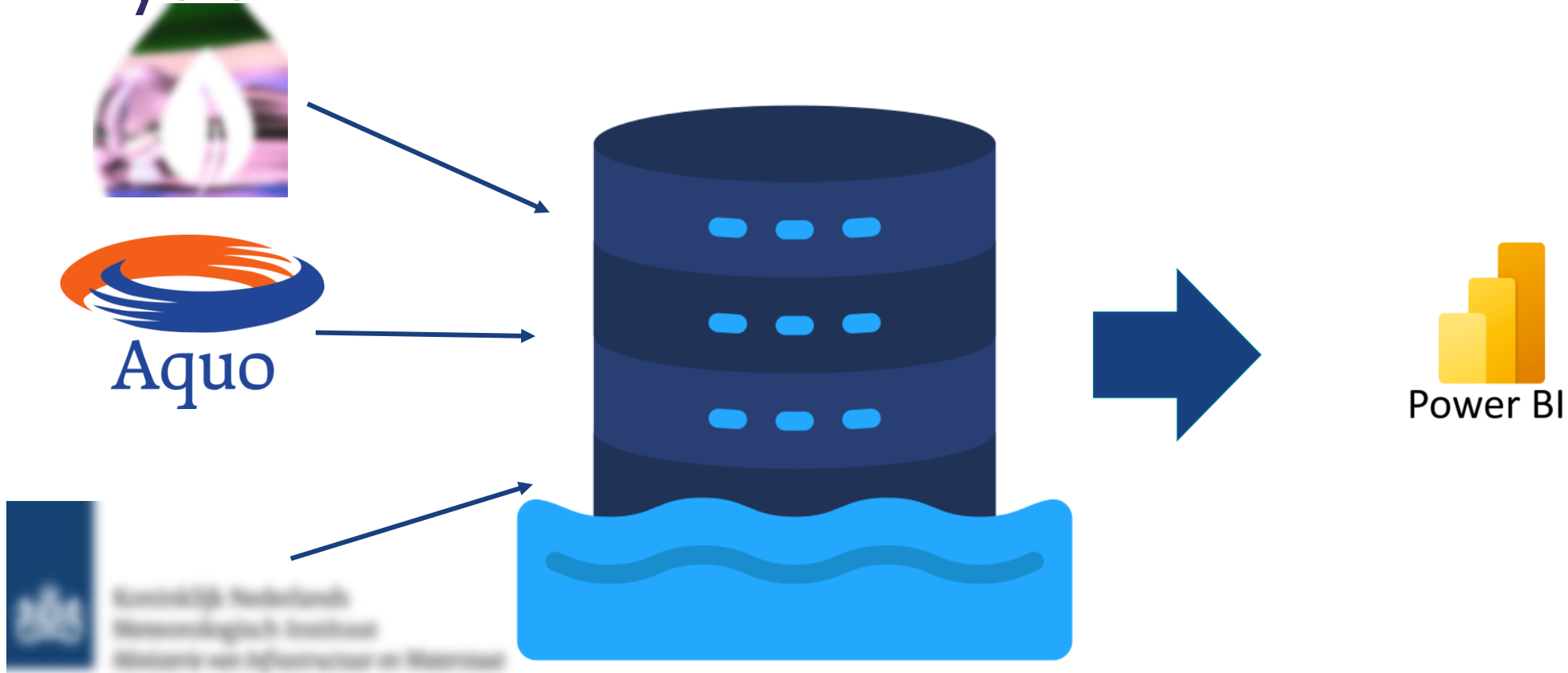
Laatste meetdatum: 14-3-2024

Handleiding



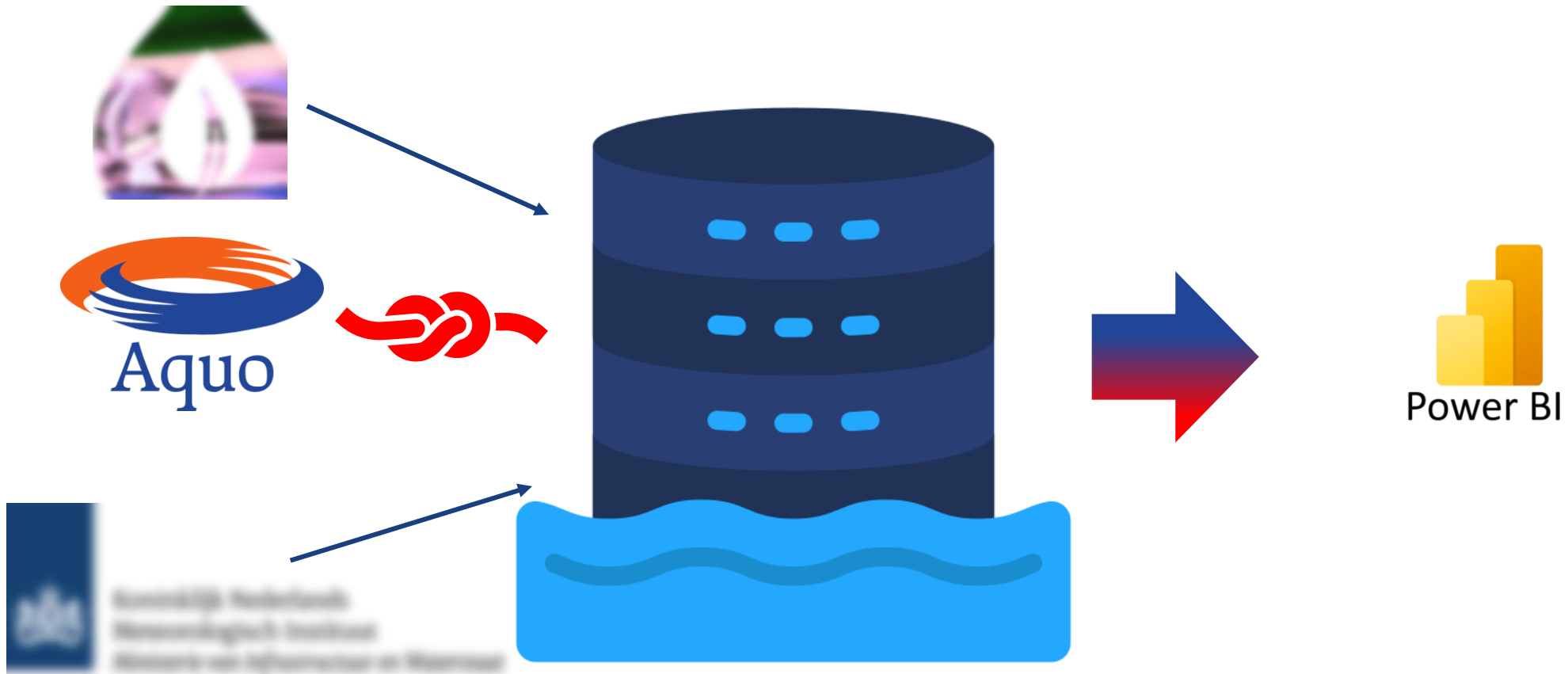


De data-pipeline van de Aquo naar Wetterskip Fryslân





Kink in de kabel...





```
1 foundDuplicatesIds = df_parameter.groupBy(df_parameter.id).agg(count('id').alias('idcount')).filter(col('idcount')>1)
2 display(foundDuplicatesIds)
3
4 duplicateIdRow = df_parameter.where(col('id')== 'parameter_C1yAoC3yHxSA')
5 display(duplicateIdRow)
```

✓ 13 sec - Command executed in 13 sec 448 ms by su_tvanwerkhoven on 3:20:09 PM, 1/24/24

> Job execution Succeeded Spark 2 executors 8 cores

[View in monitoring](#) [Open Spark UI](#)

View **Table** Chart [Export results](#)

id	idcount
parameter_C1yAoC3yHxSA	2

View **Table** Chart [Export results](#)

_wijzigingsdatum	id	identificatie	titel	code	naam	ve
2024-01-24 15:19:57.373	parameter_C1yAoC3yHxSA	C1yAoC3yHxSA	C1yAoC3yHxSA - N-[3-(Dimethyl...	C1yAoC3yHxSA	N-[3-(Dimethylamino)Propyl]Trid...	pi
2024-01-24 15:19:57.373	parameter_C1yAoC3yHxSA	C1yAoC3yHxSA	C1yAoC3yHxSA - Carboxymethyl...	C1yAoC3yHxSA	Carboxymethyldimethyl-3-(((3,3...	pi



Vraagstukken/incidenten vanuit het gebruik van de Aquo-API

2401-0094 **Dubbele codes Para...** Opslaan ★ ↻ Escaleren Aanmaken ▾ Meer ▾

ALGEMEEN INFORMATIE KOPPELINGEN BIJLAGEN (6) TIJDREGISTRATIE (1)

▶ **Stefan Jong** ⋮
WS Fryslân
06 - 43161885

▶ **Dubbele codes Parameter domeinta...** 💡 ⋮
Informatiehuis Water - E-mail
Aquo-standaard - Documentatie

▶ **Object/locatie** ⋮
Soort

Planning 🗨️

Impact ⋮
Prioriteit ⋮
Doorlooptijd ⋮
Streefdatum 📅

Afhandeling

Behandelaar 👤 ⋮
Status ⋮
Gereageerd 📅
Afgemeld 📅

Nienke Oosting 26 januari 2024 08:49
Recent zijn er nieuwe waarden toegevoegd aan de domeintabel Parameter. Klopt het dat de codes voor ID 7115 en 7114 identiek zijn? Voor alle andere parameters geldt namelijk dat de codes uniek zijn, zie bijlage

Maak zichtbaar voor aanmelder **B I U** 🖼️ 📎 🗑️ 🔄 📄

Acties, E-mailberichten, Bijlagen, zichtbaar voor aanmelder tonen

Nienke Oosting 1 februari 2024 09:42
📧 2 Afmelden - Informatiehuis Water

Nienke Oost... onzichtbaar voor aanmelder 1 februari 2024 09:42
Beste Stefan,
Hartelijk dank voor je oplettendheid. Per abuis is er inderdaad een fout ingeslopen. Mijn excuses daarvoor. Ik heb contact gehad met de indiener om haar op de hoogte te stellen van deze fout. De fout is



Data-kwaliteit: een voorbeeld van uniciteit

- ID7114 met CASnummer C1yAoC3yHxSA
- ID7115 met CASnummer C1yAoC3yHxSA maar met een andere omschrijving ...

[N-\[3-\(Dimethylamino\)Propyl\]Tridecafluorhexaansulfonamide](#)

vs_

[Carboxymethyldimethyl-3-\[\(3,3,4,4,5,5,6,6,7,7,8,8,8-tridecafluorooctyl\)sulfonyl\]amino\]propylammonium hydroxide](#)



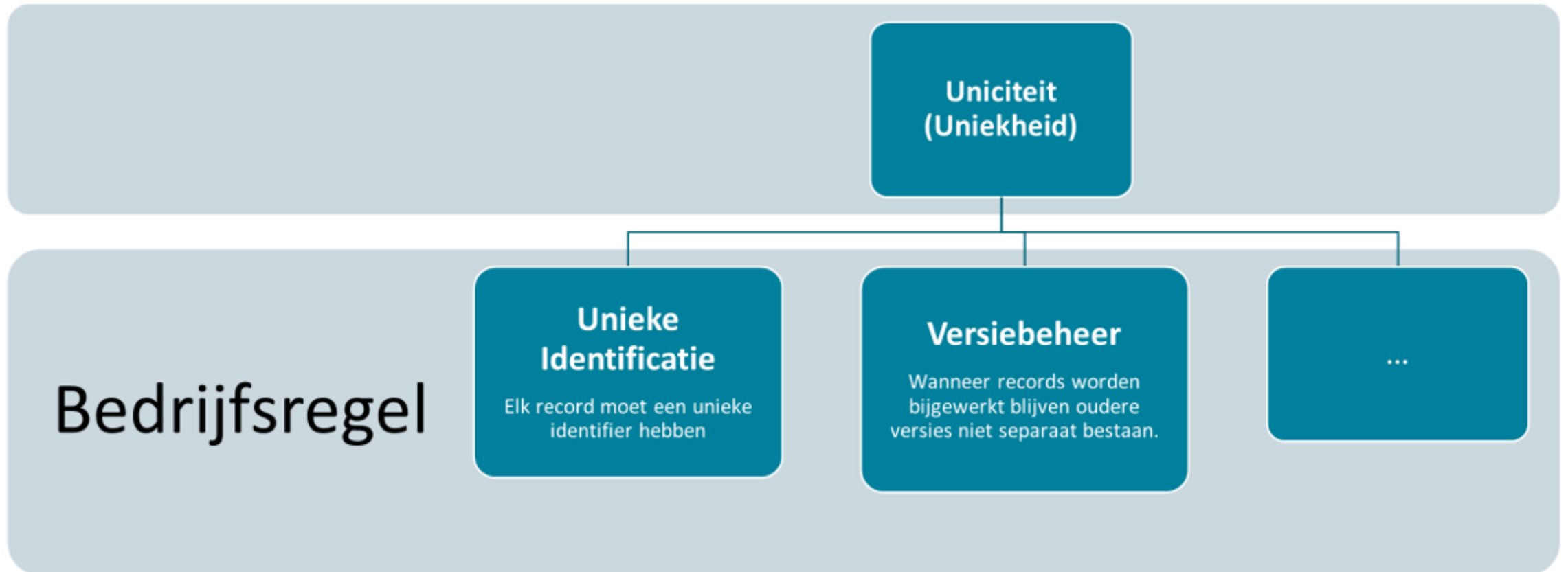
Bedrijfsregels voor uniciteit

- **Unieke identificatie:**
 - Iedere domeinwaarden moet binnen de domeintabel een unieke ID, code en omschrijving hebben
 - implementatie: ID en Code zijn verplicht en niet wijzigbaar
 - implementatie: ...
- **Versiebeheer:**
 - Wanneer records worden bijgewerkt, zorg ervoor dat oudere versies niet als separate records blijven bestaan
- **Logging:**
 - Houd logs bij van alle wijzigingen om te helpen bij het traceren en verwijderen van duplicaten.
- **En meer...**





Uitzoomen stap 1:





Data kwaliteit

Uniciteit = 'dimensie' van data-kwaliteit

Andere dimensies zijn:

- Betrouwbaarheid (voor ons erg belangrijk)
- Nauwkeurigheid
- Volledigheid
- Consistentie (lastig)
- Relevantie
- Actueel
- Herleidbaar



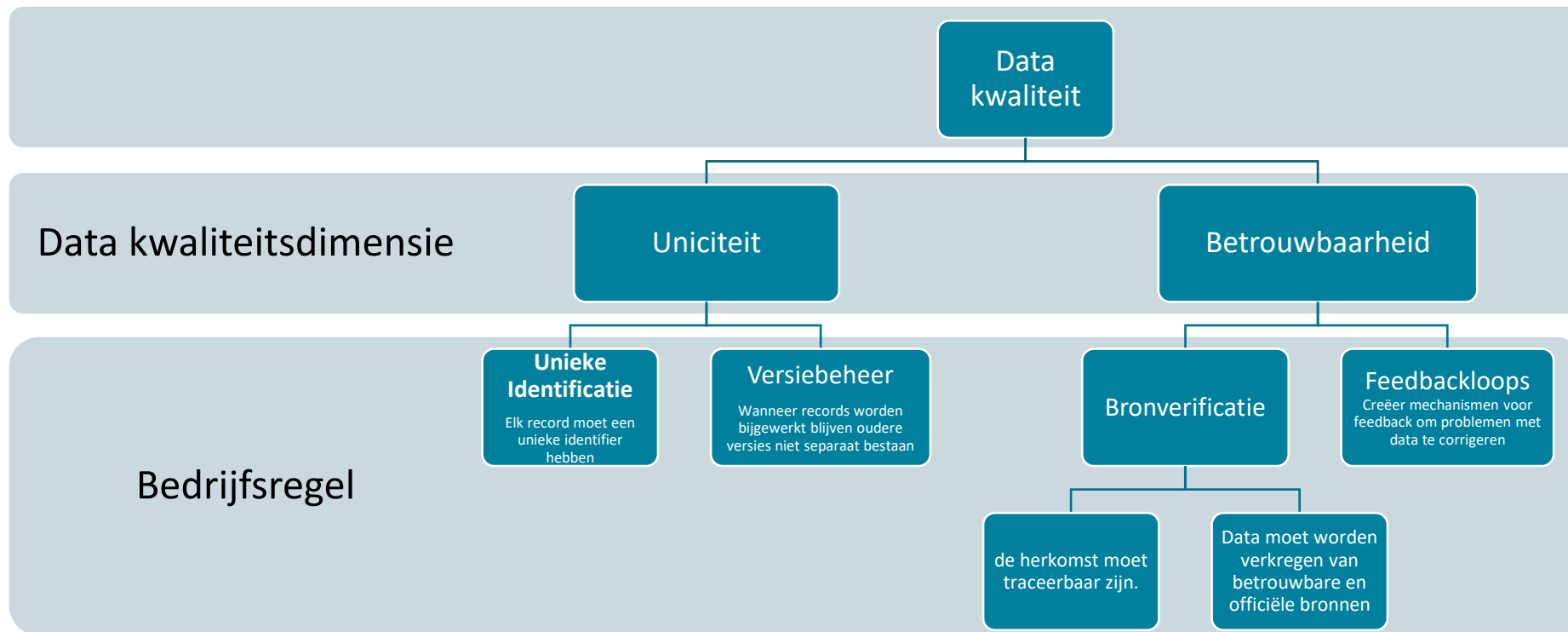


Dimensies van datakwaliteit

- Uniekheid Data is vrij van onnodige duplicatie. Elke data-eenheid moet uniek zijn. Dit helpt bij het voorkomen van verwarring en inconsistenties.
- Nauwkeurigheid: correct en vrij van fouten is. Nauwkeurige data vertegenwoordigt de werkelijkheid of een betrouwbare bron zonder afwijkingen.
- Volledigheid: Het gaat erom dat geen informatie ontbreekt en dat alle relevante gegevens beschikbaar zijn voor de gebruiker.
- Consistentie: data is over verschillende systemen, rapporten en processen heen uniform. Er zijn geen tegenstrijdigheden in de data.
- Betrouwbaarheid: de mate waarin data als geloofwaardig en betrouwbaar kan worden beschouwd. Betrouwbare data komt van een bron die als betrouwbare autoriteit wordt erkend.
- Relevantie: de mate waarin data toepasselijk is voor de behoeften van de gebruiker
- Actueel: Data moet actueel zijn en binnen een acceptabele tijd na verzameling beschikbaar zijn.
- Herleidbaar: de mogelijkheid om de oorsprong van data, inclusief de bronnen, verwerking en wijzigingen die het heeft ondergaan, te volgen.



Uitzoomen stap 2: Data kwaliteit





Bedrijfsregels en protocollen: Uniciteit

- **Unieke Identificatoren (GUID)**
 - Elk record moet een unieke identifier hebben,
- **Combinatie van attributen**
 - Een combinatie van meerdere velden (bijvoorbeeld naam, geboortedatum en adres) moet uniek zijn om een enkele entiteit te identificeren.
- **Check bij invoer**
 - Bij de invoer van nieuwe data moet een controle plaatsvinden om zeker te stellen dat het nieuwe record niet duplicaat is met bestaande records.
- **Protocol bij dubbele records**
 - Als een duplicaat wordt gedetecteerd, moet er een gestandaardiseerd proces zijn voor hoe hiermee om te gaan, bijvoorbeeld samenvoegen of verwijderen.
- **Versiebeheer**
 - Wanneer records worden bijgewerkt, zorg ervoor dat oudere versies niet als separate records blijven bestaan.
- **Regelmatige data opschoning**
 - Plan periodieke audits en data opschoningsactiviteiten om duplicaten te identificeren en te verwijderen
- **En meer ...**





Bedrijfsregels: Betrouwbaarheid

- **Bronverificatie**
 - Data moet worden verkregen van betrouwbare en officiële bronnen, en de herkomst moet traceerbaar zijn.
- **Validatieregels**
 - Er moeten regels zijn die data valideren op juistheid voordat deze wordt ingevoerd of geüpdatet in de database.
- **Regelmatige audit en review**
 - Plan periodieke audits van de data om de betrouwbaarheid te controleren en te waarborgen.
- **Versiegeschiedenis en logging:**
 - Behoud een gedetailleerde versiegeschiedenis en log wijzigingen om de herleidbaarheid van aanpassingen te verzekeren.
- **Herhaalbaarheid van processen:**
 - Wijzigingsproces is vastgesteld en gedeeld (impactmatrix, EG's TW, CCvD en updatecyclus)
 - Zorg dat processen voor data-invoer en -bewerking gestandaardiseerd en herhaalbaar zijn, om consistentie te garanderen.
- **Feedbackloops**
 - Creëer mechanismen voor feedback van gebruikers om problemen met data te identificeren en te corrigeren.
- **En meer ...**





Uitzoomen stap 3: Datakwaliteit management

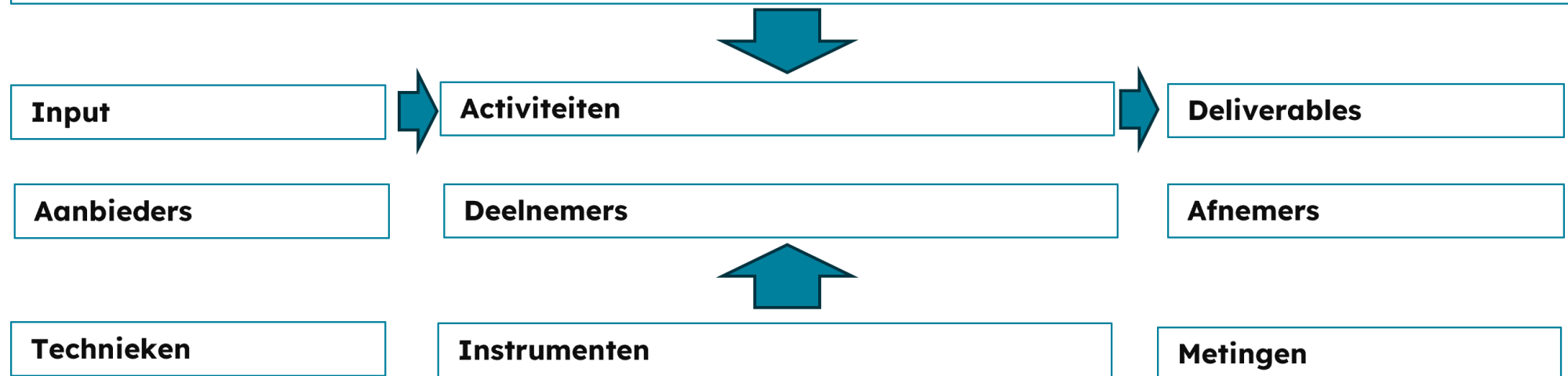
- Met bedrijfsregels en dimensies zijn we er nog niet
- Wat is het doel wat we willen bereiken met Datakwaliteit
- Welke activiteiten moeten we gaan plannen en welke activiteiten voeren we al uit
- Welke data-kwaliteitsproducten hebben we al (moeten we aanpassen) en willen we aanbieden
- Wie gaat dat doen?
- Kortom, we hebben een framework nodig



Datakwaliteit management

Definitie

Doelstellingen





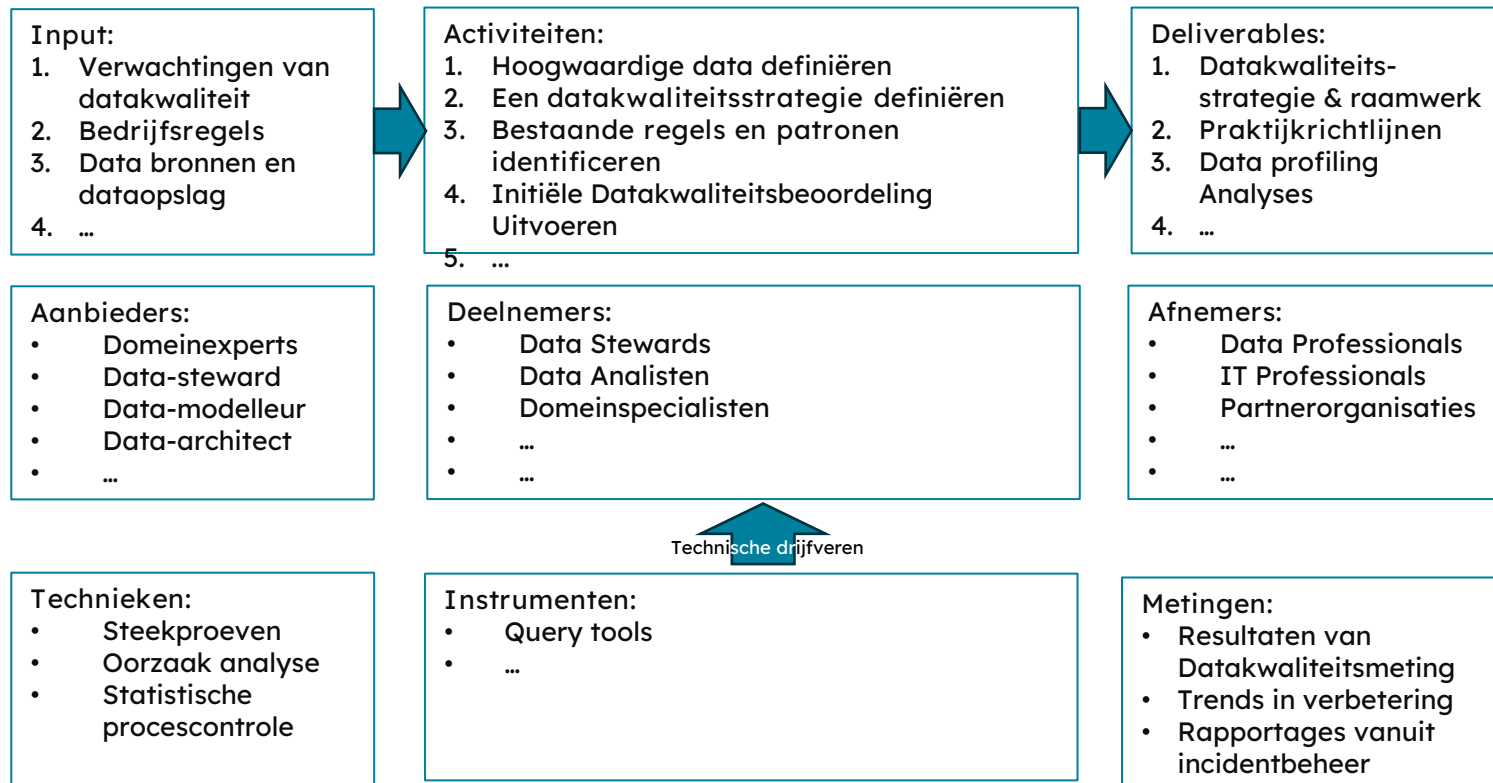
Datakwaliteit management

Definitie: Het plannen, implementeren en beheren van activiteiten om te verzekeren dat data geschikt is voor gebruik en voldoet aan de behoeften van onze gebruikers.

Doelstellingen:

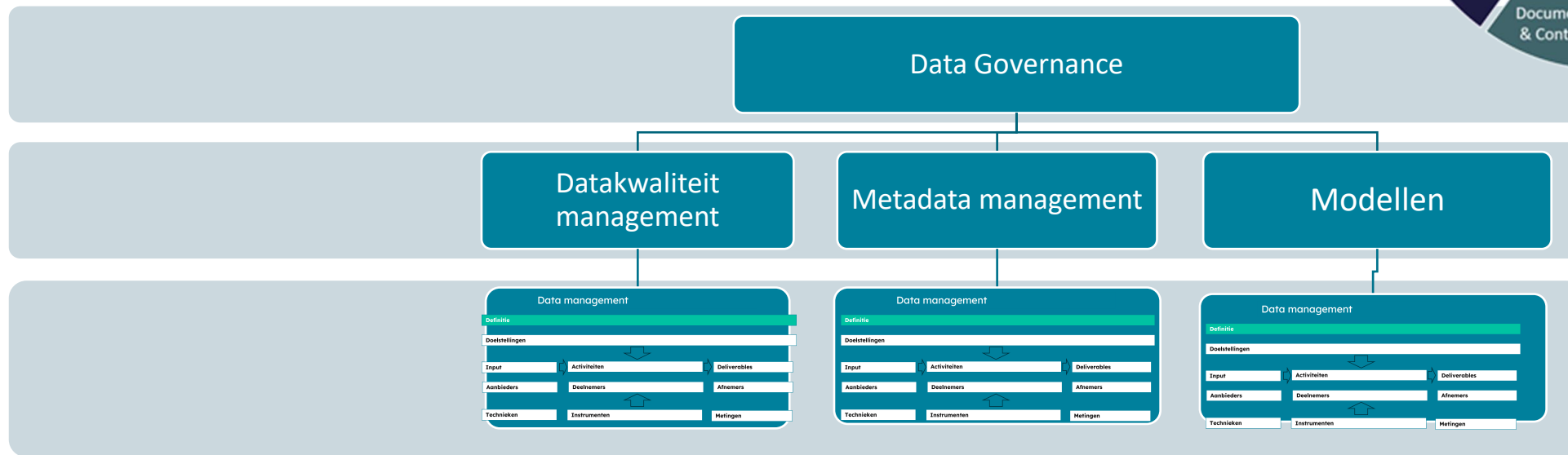
1. Ontwikkelen van een gecontroleerde aanpak om data geschikt te maken voor het beoogde doel en vereisten van de gebruiker
2. Onberispelijke data
3. ...

Zakelijke drijfveren





Uitzoomen stap 4: Data governance







Onze ambitie

Wij willen binnen de watersector de onbetwiste en centrale facilitator van waterinformatie zijn!



Onze Servicedesk staat voor je klaar!

Heb je een vraag? Of heb je ondersteuning nodig? Neem contact op met de Servicedesk!

