A network diagram consisting of various sized light blue circles connected by thin white lines, set against a solid blue background. The circles vary in size, with some being significantly larger than others, and they are interconnected in a complex, web-like structure.

Bedrijfstakonderzoek  
BTO 2020.058 | November 2020

**Inventarisatie  
behoefte naar een  
databank voor, en  
kennisdeling van  
stofinformatie**

Bedrijfstakonderzoek

**KWR**

Bridging Science to Practice



# Rapport

## Inventarisatie behoeftes naar een databank voor, en kennisdeling van stofinformatie

BTO 2020.058 | November 2020

Dit onderzoek is onderdeel van het collectieve Bedrijfstakonderzoek van KWR, de waterbedrijven en Vewin.

### Opdrachtnummer

402045/141

### Projectmanager

Patrick Bauerlein

### Opdrachtgever

BTO - Bedrijfsonderzoek

### Auteur(s)

Dirk Vries, Andrea Brunner, Tessa Pronk, Erik Emke

### Kwaliteitsborger(s)

Marja Lamoree

### Verzonden naar

Dit rapport is verspreid onder BTO-participanten.

Een jaar na publicatie is het openbaar.

### Keywords

stoffendatabank, DSFP, suspect screening, non-target screening, organische micro verontreinigingen

Jaar van publicatie  
2020

### Meer informatie

Dirk Vries  
T 030 60 69 671  
E [Dirk.Vries@kwrwater.nl](mailto:Dirk.Vries@kwrwater.nl)

PO Box 1072  
3430 BB Nieuwegein  
The Netherlands

T +31 (0)30 60 69 511  
F +31 (0)30 60 61 165  
E [info@kwrwater.nl](mailto:info@kwrwater.nl)  
I [www.kwrwater.nl](http://www.kwrwater.nl)



November 2020 ©

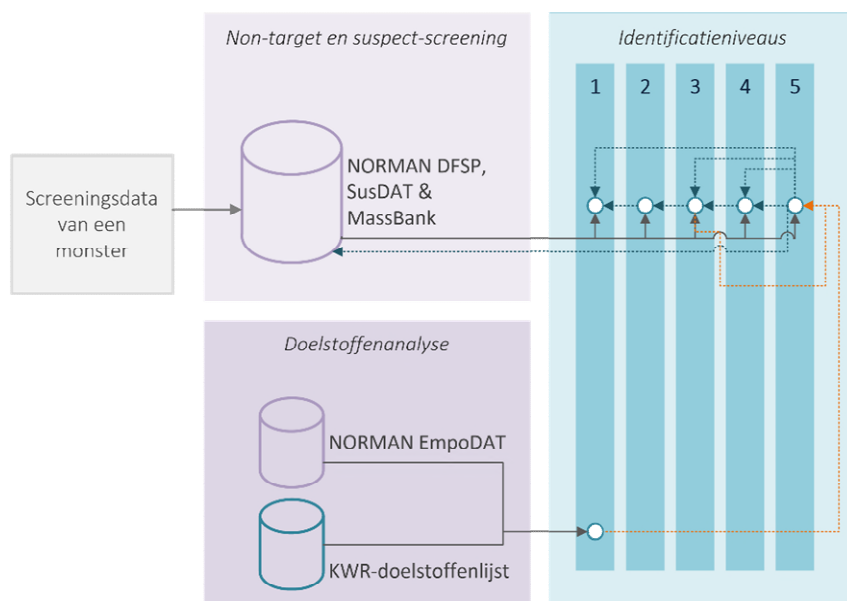
Alle rechten voorbehouden aan KWR. Niets uit deze uitgave mag - zonder voorafgaande schriftelijke toestemming van KWR - worden veelevoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier.

# Managementsamenvatting

## Aggregatie van doelstofgegevens vormt eerste stap in betere afstemming van doelstofanalyse- en screeningsgegevens tussen drinkwaterlaboratoria

**Auteur(s)** Dirk Vries, Andrea Brunner, Tessa Pronk, Erik Emke.

Het delen en uitwisselen van doelstofgegevens, suspectlijsten en gehanteerde referentiestandaarden is een eerste stap in het structureel en centraal opslaan en ontsluiten van stofgegevens uit diverse screenings en doelstofanalyses. Daardoor kunnen onderlinge afstemming tussen drinkwaterlaboratoria en prioritering van stoffen efficiënter en doeltreffender worden gefaciliteerd. Dit blijkt uit een inventarisatie die heeft geleid tot het kaderen van een gebruiksscenario van een databank. In opdracht van de Nederlandse drinkwaterbedrijven hebben de drinkwaterlaboratoria AquaLab-Zuid, HWL, het Vitens-laboratorium en WLN als voornaamste taak om de waterkwaliteit op diverse chemische en biologische parameters te screenen. Resultaten hiervan worden met name opgeslagen in eigen, hiervoor gecreëerde *in house* en leverancier-specifieke bestanden. Een dergelijke werkwijze staat efficiënte screening en identificatie in de weg, omdat bijvoorbeeld het ene laboratorium een stof wel identificeert en het andere niet, zonder dat ze hierover van elkaar op de hoogte zijn. De vraag of deze gegevens niet ten minste bij alle Nederlandse drinkwaterlaboratoria beschikbaar zouden moeten zijn, is in deze studie onderzocht. Daarnaast is de behoefte uitgesproken om stappen te zetten in het gebruik van (bestaande) databanken die toegankelijk en doorzoekbaar zijn en de eventuele ontwikkeling van een stoffendatabank indien bestaande oplossingen niet in de informatiebehoefte voorzien.



*Schematisch toekomstbeeld met de toepassing van NORMAN-databanken in het screenen en identificeren van stoffen op verschillende identificatieniveaus (van accurate massa – niveau 5 – tot volledige structuuropheldering – niveau 1). Blauwe pijlen geven het verder ophelderen van geanalyseerde stoffen aan, oranje pijlen tonen de weerlegging aan van de gegevens (door bijvoorbeeld een doelstofanalyse). Het gebruik van een KWR-doelstoffenlijst is een tussenstation (voor afkortingen, zie Tabel 3 in het rapport).*

**Belang:**

Door het delen van doelstofanalyse- en non-target screening (NTS)-gegevens ontstaat meer zicht op de laboratoriumtechnieken waarmee doelstoffen en suspects worden geanalyseerd en geïdentificeerd, op massaspectrum-fragmentatiespectra die van belang zijn, op het voorkomen van de stof en op de laboratoria die bepaalde stoffen hebben gemeten of geïdentificeerd. Deze kennis leidt tot een effectievere prioritering van (mogelijk) zorgwekkende stoffen, tot een snellere detectie of identificatie hiervan en tot een beter beeld hoe ze landelijk voorkomen

**Aanpak: inventarisatie van behoefte aan kennisdeling en delen van doelstofgegevens**

Bij de start van dit onderzoek is, tezamen met de drinkwaterlaboratoria, geïnventariseerd welke informatie daadwerkelijk gewenst is om gezamenlijk uit te wisselen, en welke oplossingen hiervoor op korte en lange termijn beschikbaar zijn. De inventarisatie heeft geleid tot het kaderen van een gebruiksscenario van een databank. Het gebruiksscenario omvatte het verzamelen en delen van doelstofgegevens, alsook een nadere verkenning van de mogelijkheden om gegevens in een Europees platform voor screeningsgegevens uit te wisselen, namelijk de databanken van het NORMAN-netwerk.

**Resultaten: behoefte aan delen van gegevens vanuit NTS, suspect- én doelstoffenanalyses**

Met de resultaten uit de verkenning is een eerste gebruiksscenario voor het delen en uitwisselen van stofgegevens opgesteld, namelijk die van het verzamelen van doelstofgegevens. Ook het verzamelen van suspectgegevens is nodig, maar dit bleek binnen dit project niet haalbaar. KWR heeft doelstofgegevens van ongeveer 1.000 stoffen geaggregeerd. Voor een deel hiervan kon de aggregatie niet eenduidig of volledig plaatsvinden omdat de identificatienamen van een stof inconsistent waren, of door het ontbreken van metagegevens. Verder moesten sommige laboratoria relatief veel inspanning leveren om deze gegevens te verzamelen. Daarnaast zijn de verschillende stappen die bij het uitwisselen van gegevens in NORMAN-databanken moeten worden doorlopen, in het voorliggend rapport uitgewerkt.

Deze studie leidt tot de volgende conclusies. Het is nodig om de doelstofgegevens, suspectlijsten en gehanteerde referentiestandaarden structureel en centraal op te slaan en te ontsluiten. Daarnaast moeten internationaal en nationaal erkende identificatienamen van een stof worden opgeslagen en gedeeld ten behoeve van consistentie en effectieve(re) 'matching' met een bestaande stof. Tot slot is het raadzaam om een adviesgroep op te richten die enerzijds zorgt voor de bevordering van het stroomlijnen van de stoffeninformatie tussen de drinkwaterlaboratoria en anderzijds de transitie in gang zet naar het gebruik van de databanken van het NORMAN-netwerk en de ervaringen hiervan deelt.

**Toepassing: onderhoud de doelstoffenlijst én werk toe naar het gebruik van het NORMAN-netwerk**

Drinkwaterlaboratoria kunnen op korte termijn zicht blijven houden op nieuwe, mogelijk opkomende, reeds geïdentificeerde stoffen in de Nederlandse (drink)watersector. Daarvoor is het wel van belang om op regelmatige basis de huidige lijst van doelstoffen bij te werken. Dit kan het beste structureel worden uitgevoerd door de Beheerscommissie Overdracht Kwaliteitsgegevens (BOK). Op die manier is het beheer centraal geregeld en krijgen de stoffen een landelijk gebruikte codering (namelijk SIKB- en REWAB-code) mee.

De aantrekkelijkheid om het NORMAN-netwerk te gebruiken, ligt in de beheerskosten en in de brede informatie- en kennisdeling van screeningsdata (NTS, suspects én doelstoffen). Leden van het NORMAN-netwerk kunnen zelf data invoeren, met een keuze voor inzage in verschillende beveiligingsniveaus. Op de langere termijn is de infrastructuur daarom aan te bevelen voor het opslaan en delen van screeningsdata, ook als het gevoelige watermonitoringsgegevens betreft.

**Het Rapport**

Dit onderzoek is beschreven in het rapport *Ontwikkeling van een centrale stoffendatabank voor drinkwater* (BTO 2020.058)

## Samenvatting

In Nederland hebben de drinkwaterlaboratoria Aqualab-Zuid, HWL, het Vitens-laboratorium en WLN als voornaamste taak om de waterkwaliteit op diverse chemische en biologische parameters te screenen en/of meten in opdracht van de Nederlandse drinkwaterbedrijven. Bij het screenen, analyseren en identificeren van stoffen worden veel gegevens opgeslagen in diverse databanken, met name leverancier-specifieke databanken voor het opslaan van massaspectra en andere stofgegevens. Dit heeft de vraag bij de Noordelijke drinkwaterbedrijven (Waterbedrijf Groningen en WMD) opgeroepen of gegevens die bij de analyse en identificatie van stoffen worden verzameld, niet ten minste onderling onder alle Nederlandse drinkwaterlaboratoria beschikbaar zouden moeten zijn. Daarnaast is de behoefte uitgesproken om stappen te zetten in de ontwikkeling van een databank die toegankelijk en doorzoekbaar is. De beantwoording van deze vragen is uitgevoerd door (a) een verkenning van gebruiksscenario's van een stoffendatabank, (b) een definitie en kadering van het meest relevante gebruiksscenario, (c) eerste stappen in verzameling van stofgegevens, en tenslotte (d) kennisdisseminatie over het gebruik van internationale databanken ten behoeve van gegevensuitwisseling in het NORMAN-platform.

Uit de verkenning blijkt dat het gebruiksscenario waarbij informatie over doelstoffen en suspects wordt gezocht, het meest relevant is om *nu* al gegevens onderling te delen en te ontsluiten. Een eerste stap is gezet met de verzameling van doelstoffen die door de verschillende labs worden gemeten. Daarnaast is de potentie van het gebruik van de databanken binnen het Europese netwerk NORMAN geïnventariseerd en beschreven, en is in detail ingegaan op het gebruik en potentieel van het Digital Sampling Freezing Platform (DSFP) van NORMAN. Bij het opvragen van de doelstofgegevens is gebleken dat sommige laboratoria relatief veel tijd en inspanning hebben moeten leveren om deze gegevens te verzamelen. KWR heeft vervolgens de stofgegevens geaggregeerd in één spreadsheet met ongeveer 1000 stoffen. Een deel van de gegevens kon niet eenduidig of volledig geaggregeerd worden door fouten in de naamgeving van een stof, het ontbreken van CAS-nummers of andere identificatiecodes van de stof, en het ontbreken van metagegevens (zoals voorkomen van een stof).

Ten behoeve van effectievere prioritering van (mogelijk) zorgwekkende stoffen, snellere detectie of identificatie van deze stoffen en een beter landelijk beeld van het voorkomen van deze stoffen wordt voorgesteld om (1) de doelstof- en suspectgegevens en gehanteerde referentiestandaarden structureel en centraal te gaan opslaan en ontsluiten op een wijze die aansluit bij de behoefte van de laboratoria en die de transitie naar het gebruik van NORMAN-databanken ondersteunt, (2) naast het CAS-registratienummer ook andere internationaal erkende identificatienamen van een stof<sup>1</sup> op te slaan en te delen, (3) een adviesgroep op te richten die enerzijds het stroomlijnen van de stoffeninformatie tussen de drinkwaterlaboratoria bevordert en anderzijds de transitie naar het gebruik van de databanken van het NORMAN-netwerk, DSFP en MassBank.eu in het bijzonder, in gang zet en de ervaringen daarvan deelt.

---

<sup>1</sup> Zoals SMILES en InChI.

# Inhoud

<b>Rapport</b>	<b>2</b>
<b><i>Managementsamenvatting</i></b>	<b>3</b>
<b>Samenvatting</b>	<b>5</b>
<b>Inhoud</b>	
<b>1 Achtergrond en aanpak</b>	<b>8</b>
1.1 Aanleiding	8
1.2 Initiatieven in het BTO om stofgegevens centraal te ontsluiten	9
1.3 Inventarisatie van diverse gebruiksscenario's en het delen van stofgegevens	11
1.4 Terminologie voor stofidentificatie	12
<b>2 Gebruiksscenario's voor de identificatie van stoffen</b>	<b>14</b>
2.1 Enquête	16
2.2 Uitvraag gebruiksscenario's	16
2.3 Definitie van gebruiksscenario	20
<b>3 Verzameling van doelstoffen en suspects gemeten door drinkwaterlaboratoria</b>	<b>21</b>
3.1 Aanpak	21
3.2 Verwerking data	21
3.2.1 Databestanden klaar maken voor verwerking	21
3.2.2 CAS-registratienummers en andere identificatienamen toekennen	23
3.2.3 Gegevens van stoffen van de laboratoria koppelen	23
3.2.4 Handmatige identificatie van stoffen	23
3.3 Resultaat	24
3.4 Conclusie	24
<b>4 NORMAN-netwerk en databanken</b>	<b>26</b>
4.1 Het NORMAN-netwerk van Europese laboratoria, onderzoeksinstituten en universiteiten	26
4.2 Suspect list exchange	26
4.3 NORMAN Digital Sample Freezing Platform	27
4.4 MassBank.eu	28
4.5 Conclusie	28
<b>5 Verkennen van upload van NTS-gegevens en analyse met DSFP</b>	<b>29</b>

5.1	Introductie	29
5.2	Werkwijze	29
5.2.1	Inlog en contactgegevens instituut	29
5.2.2	Interne standaarden	30
5.2.3	Conversie naar mzML	30
5.2.4	Upload van gegevens	30
5.2.5	Suspect-screening	31
5.2.6	Suspectlijsten	32
5.2.7	Conclusie	32
<b>6</b>	<b>Conclusies, aanbevelingen en perspectief</b>	<b>33</b>
6.1	Conclusies en aanbevelingen	33
6.2	Perspectief	34
<b>I</b>	<b>Bijlage Vragenlijst</b>	<b>37</b>
	Antwoorden op vragenlijst over de informatiebehoefte van drinkwaterlaboratoria (incl. KWR)	37



# 1 Achtergrond en aanpak

## 1.1 Aanleiding

In de Europese Unie zijn er meer dan 20.000 registreerde chemicaliën, gedestilleerd uit meer dan 100.000 dossiers<sup>2</sup>. Deze chemicaliën zijn in het kader van de REACH-wetgeving geadmistreerd. Wereldwijd zijn er miljoenen unieke stoffen. Niet alle geproduceerde stoffen zijn bekend, of staan geregistreerd in een databank. Voor structuurinformatie en het opzoeken van verschillende naamconventies van stoffen zijn er diverse databanken, zoals PubChem (111 miljoen unieke stoffen) en ChemSpider (90 miljoen unieke stoffen).

Een deel van deze stoffen en hun afbraakproducten kunnen terecht komen in het aquatische milieu<sup>3</sup>, bijvoorbeeld door ineffektieve (afval)waterzuiveringen, lozingen en afspoeling van landbouwgrond. Om deze redenen is informatie afkomstig van monitoring, screening, risicobeoordeling en zuivering van groot belang om te kunnen prioriteren en adequate maatregelen te kunnen treffen.

In Nederland hebben de drinkwaterlaboratoria AquaLab-Zuid, HWL, het Vitens-laboratorium en WLN als voornaamste taak om de waterkwaliteit op diverse chemische en biologische parameters te screenen en/of meten in opdracht van de Nederlandse drinkwaterbedrijven. Bij het screenen, analyseren en identificeren van stoffen worden verschillende gegevens opgeslagen in diverse bestanden. Met name commerciële pakketten geleverd door leveranciers van analyse-apparatuur worden gebruikt voor de stoffenscreening, zoals Compound Discoverer (Thermo Fisher Scientific) of Metaboscape (Bruker). Bij identificatie wordt via macro's of handmatig de geïdentificeerde stof in een apart bestand opgeslagen. Gegevens zoals welke stoffen geïdentificeerd zijn en met welke betrouwbaarheid, met welke methode en rapportagegrens, en in welke watermatrix als ook welke spectra in non-target screening (NTS) zijn gemeten en suspect-stoffenlijsten, worden niet op structurele basis onderling gedeeld. Dit heeft als gevolg dat een onbekend spectrum (of stof) bij het ene laboratorium wel geïdentificeerd kan zijn en bij een ander laboratorium niet, zonder dat de laboratoria daarvan op de hoogte zijn. Een dergelijke werkwijze staat efficiënte screening en identificatie in de weg.

Omdat data alleen via aparte, in-huis databanken toegankelijk zijn, kunnen kennis en metingen niet op eenvoudige wijze gebundeld opgevraagd worden. Het verzamelen en op orde brengen van data vraagt om deze reden veel tijd en is kostbaar. Daarnaast speelt het aspect van eigenaarschap en mogelijk bedrijfsgevoelige informatie: er zijn bij de waterbedrijven en drinkwaterlaboratoria zorgen over het publiek worden van waterkwaliteitsgegevens. Deze redenen

Bij doelstofanalyses (target analyses) wordt een monster alleen op doelstoffen (targets) geanalyseerd en gekwantificeerd. Dit betekent dat er op basis van een meting van een doelstof in een standaardmengsel een respons kan worden bepaald. Door standaarden in verschillende verdunningen te meten kan in een monster door gelijkenis met de bekende accurate massa, mogelijk bijbehorend fragmentatie-massaspectrum en retentietijd in de chromatografiekolom de concentratie worden bepaald.

Bij non-target screening (NTS) wordt gekeken naar 'onbekende onbekenden' en 'bekende onbekenden'. Van de laatste categorie, de zogeheten suspects, wordt verwacht dat ze in een monster aanwezig zijn en er zijn massaspectra of andere kenmerken bekend die in databases zijn op te zoeken. Deze zoektocht wordt ook 'suspect screening' genoemd. Bij de 'onbekende onbekenden' wordt getracht met gelijkenis-zoekacties op hydrofobiciteit, fragmentatiepatronen, etc. (similarity searches) nadere identificatie te bewerkstelligen.

<sup>2</sup> Bron: <https://echa.europa.eu/information-on-chemicals/registered-substances>, 8 september 2020.

<sup>3</sup> Sjerps, R., Emke, E. en Maassen, M. (2016), BTO-rapport 2016.027, Haalbaarheidsstudie stoffendatabase voor de Nederlandse drinkwatersector"

hebben tot de wens geleid om digitaal, de stoffen die worden gemeten door de drinkwaterlaboratoria in opdracht van de (drink)waterbedrijven, toegankelijk, beveiligd, gestandaardiseerd en doorzoekbaar te maken voor ten minste de drinkwaterlaboratoria. Waterbedrijf Groningen en WMD hebben in 2019 daarom het initiatief genomen om de kennisdeling op het onderwerp van (onbekende) stoffen te bevorderen en ondersteunen via bedrijfsonderzoeksgelden. Vanwege de aard van het onderwerp, namelijk het delen en beheren van stoffengegevens, hebben de twee waterbedrijven de begeleiding van dit project gedelegeerd naar WLN.

## 1.2 Initiatieven in het BTO om stofgegevens centraal te ontsluiten

Het idee voor een centrale opslag, verwerking en ontsluiting van informatie over stoffen is niet nieuw, en er zijn de afgelopen jaren verschillende initiatieven gestart:

- In 2016 is in het BTO een haalbaarheidsstudie<sup>4</sup> uitgevoerd naar de ontwikkeling van een centrale stoffendatabank/-portaal. Onderzoekers en belanghebbenden uit de sector hebben in dat project een breed gedragen pakket van eisen voorgesteld voor de opzet van een stoffendatabank/-portaal, waarin gegevens over bekende en onbekende stoffen worden gedeeld tussen de drinkwaterbedrijven en hun laboratoria. Vanuit het LMO (Labmanagersoverleg) is aan de Contactgroep Chemie (KWR, Vitens lab, AqualabZuid, WLN, HWL en De Watergroep) gevraagd om de behoeften voor een dergelijke stoffendatabank verder te inventariseren en op basis daarvan met een projectvoorstel te komen. De focus is hierbij gelegd op het verzamelen van (non-target) screeningsdata, als ook data van geïdentificeerde stoffen, op een centrale plek.
- Een ander idee raakt aan onderzoek bij KWR waarbij het gebruik en de ontsluiting van waterkwaliteitsgegevens of informatie over stoffen centraal staat. In het WiCE-project Water Wise<sup>5</sup> wordt gepleit voor een databank waarbij stofinformatie, regelgeving, data van zuiveringen en microbiële gegevens op eenduidige wijze worden opgeslagen en toegankelijk zijn voor webapplicaties (front-end) die deze gegevens kunnen aggregeren en via modellen naar informatie kunnen vertalen. Daarnaast heeft KWR in 2020 de webtool AquaPriori ontwikkeld. Deze applicatie voorspelt wat de verwijdering van een zuivering voor een bepaalde organische microverontreiniging zal zijn aan de hand van een duizendtal gedefinieerde stoffeigenschappen die op basis van de SMILES-identificatie worden berekend. De applicatie is gekoppeld aan een databank met stofgegevens (chemische structuur en verschillende standaard identificatienamen) waarin nu al reeds meer dan 2500 stoffen staan. De identificatienamen worden opgehaald van PubChem.

Er bestaan al diverse nationale en internationale stoffendatabanken en –portalen, zoals die van het NORMAN-netwerk (<https://www.norman-network.com/nds/>), de REWAB-databank, de zoekinterface van RIVM voor risico's van stoffen, Chemspider (<http://www.chemspider.com/>), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) en het US EPA Chemistry Dashboard (<https://comptox.epa.gov/dashboard>). Zie een niet-uitputtend overzicht in Tabel 1. Dat roept de vraag op of het wel zinvol is om een nieuwe databank op te zetten. Nu zijn de genoemde databanken gevuld met een bepaald doel voor ogen, waardoor ze niet perse volledig aansluiten op de gebruiksbehoeftes van (andere) bedrijven, laboratoria, toezichthouders en kennisinstituten in de watersector. Het NORMAN-netwerk en de REWAB-databank uitgezonderd, is in geen van de genoemde databanken de mogelijkheid om data van de Nederlandse drinkwatersector te ontsluiten en er grip op te houden. Daarnaast is de herkomst van data in sommige van deze databanken onbekend. Ten slotte is het zo dat data die niet eenduidig geïnterpreteerd kunnen worden en niet voorzien zijn van een kwaliteitsborging, van beperkte waarde blijken.

---

<sup>5</sup> BTO 2020.033 WiCE-rapport *Water Wise – a decision support system for safe water use and reuse*, juni 2020, KWR.

Tabel 1: overzicht van verschillende databanken die stoffeninformatie bevatten. Gebruikte afkortingen: GW: grondwater, OW: oppervlaktewater, DW: drinkwater, RW: ruw water, PW: proceswater, RG: rapportagegrens.

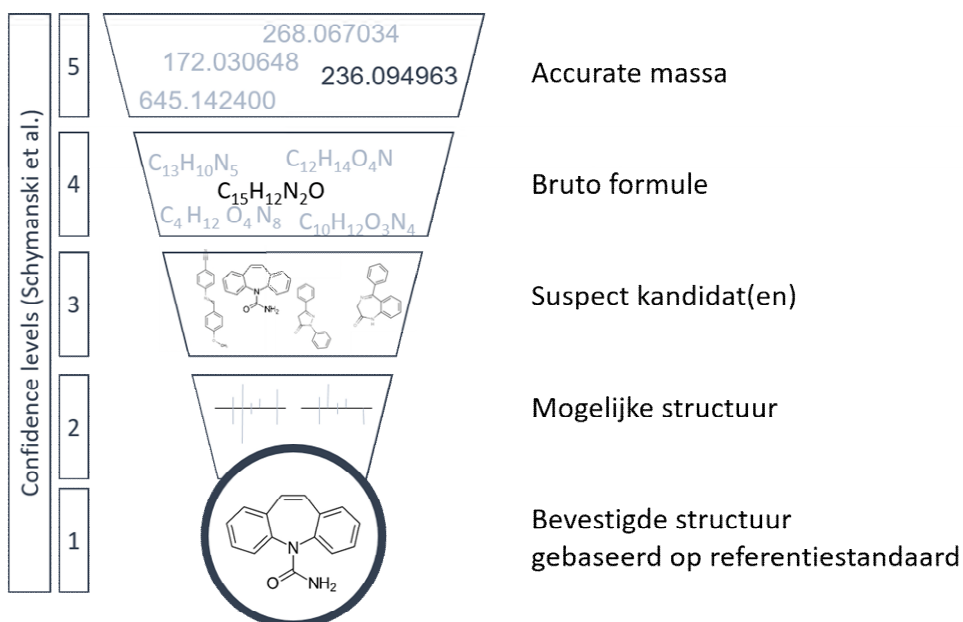
Databank	Beheerder	Aantal stoffen	Chemisch-fysische eigenschappen	Toxiciteitsgegevens	Metadata
SusDAT	NORMAN netwerk	>110000	Formule, massa, m/z	PNEC (QSAR)	
Waterkwaliteitsportaal	Informatiehuis Water		Concentratie	-	Meetplaats, datum, matrix (ow, gw)
REWAB	KWR	>1600*	Concentratie (geaggregeerd per jaar)	-	Meetplaats, jaar, matrix (gw, dw, ow, rw)
Zoeksysteem risico's van stoffen	RIVM	>5000	Structuur	Diverse humaan en ecotoxisch relevante normen en gevaarsindelingen	Categorie van stoffen (Zeer Zorgwekkende Stoffen-lijst, Stofklasse)
Chemistry Dashboard	US EPA	882000	Formule, massa, log Kow, etc	<i>In vitro</i> toxiciteit, ADME, bioactivity, environmental fate/transport, human & eco hazard	
Chemspider	Royal Society of Chemistry	92 mio**	Formule, massa, log Kow, etc	beperkt	publicaties
PubChem	NIH	111mio	Formule, massa, log K <sub>ow</sub> , Collision Cross Section, Kovacs RTI, spectrale informatie: MS, UV, IR		publicaties, transformatie-producten

\* Incl. de parameters die wettelijk verplicht zijn te meten of melden in het kader van de KRW en Drinkwaterbesluit.

\*\* Niet alle stoffen in deze database voldoen aan dezelfde validiteitscriteria vanwege de commerciële basis van de zoekmachine. Een deel van de opgenomen stoffen in Chemspider is toegevoegd door leveranciers. Hiervan is een deel nog nooit gesynthetiseerd.

### 1.3 Inventarisatie van diverse gebruiksscenario's en het delen van stofgegevens

Met vloeistofchromatografie gekoppeld aan hoge resolutie massaspectrometrie (LC-HRMS) kan uit de accurate massa en het isotopenpatroon de chemische formule van een stof worden bepaald. De combinatie van accurate massa (MS1) van het geïoniseerde ion en bijbehorende retentietijd (*retention time index*: RTI) wordt ook wel feature genoemd. Gebaseerd op MS2 fragmentatie-spectra kan door middel van zoeken in bibliotheken van fragmentatiespectra ('spectral libraries') (en 'in silico' fragmentatiespectra de structuur van een stof worden opgehelderd en afhankelijk van de beschikbaarheid van een referentiestandaard de structuur worden bevestigd of ontkracht. Het gebruik van MS2 data dient hier twee doelen: de identificatie zelf en de verhoging van het betrouwbaarheidsniveau van op basis van accurate massa geïdentificeerde verbindingen. Het proces van identificatie leidt tot een opheldering van de samenstelling van een stof die met verschillende identificatieniveaus wordt getypeerd, de zogenaamde Schymanski-schalen<sup>6</sup>, zie Figuur 1. Afhankelijk van het doel van de analist of onderzoeker – het gebruiksscenario – worden verschillende databanken en gegevensbronnen geraadpleegd.



Figuur 1: de verschillende niveaus van identificatie bij het screenen van stoffen volgens Schymanski et al. (2014).

Bij dit onderzoek zijn de gebruiksscenario's geïnterpreteerd, en is met behulp van de inventarisatie een scenario ('use case') geselecteerd waarbij de behoefte het grootst is om gegevens over stoffen onderling te delen. Met de financiële middelen die er waren bleek een (eenvoudige) database met gegevens over doelstoffen en NTS niet haalbaar. In plaats daarvan is de studie werk beperkt tot het verzamelen en delen van gegevens over doelstoffen. Als stip op de horizon is het uploaden van gegevens en gebruik van twee internationale databanken van het NORMAN-netwerk benoemd, namelijk het uploaden en automatische screenen voor suspects met het Digital Sample Freezing Platform (DSFP), en het uploaden en zoeken van MS2 fragmentatie data in MassBank. De functionaliteiten van het DSFP zijn verkend door het uploaden van data afkomstig uit een ringonderzoek dat uitgevoerd is door het chemisch laboratorium van KWR. Tenslotte zijn de doelstof-gegevens verzameld en gedeeld onder de waterlaboratoria als eerste stap om tot kennisdeling te komen.

<sup>6</sup> Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Emma L. Schymanski, Junho Jeon, Rebekka Gulde, Kathrin Fenner, Matthias Ruff, Heinz P. Singer, and Juliane Hollender, *Environmental Science & Technology* 2014 48 (4), 2097-2098. DOI:

## 1.4 Terminologie voor stofidentificatie

In het voorliggende rapport zijn diverse termen gebruikt die toegepast worden in het chemische domein ten behoeve van stofidentificatie. Deze nomenclatuur is in Tabel 2 opgenomen, tezamen met enkele karakteristieken.

Tabel 2: nomenclatuur voor verschillende (standaard-)identificaties van stoffen

Identificatie	Uniek	Niet unieke stoffen (voorbeelden)	Organisatie	Karakteristieken
<b>(Preferred*) IUPAC-naam</b>	Ja*	N.v.t.	IUPAC (International Union of Pure and Applied Chemistry)	Conventie om op gestructureerde wijze een stof te benoemen
<b>InChI (International Chemical Identifier)</b>	Nee	Complexe metallo-organische stoffen <sup>7</sup>	InChI Trust	De InChI-identificatiecode is vergezeld van een versleutelde code bestaande uit 27 karakters (InChI-key). De wijze van versleuteling is niet uniek en databank-afhankelijk.
<b>CAS-registratienummer</b>	Nee	<a href="https://pubchem.ncbi.nlm.nih.gov/#query=5864-38-0&amp;tab=substance">https://pubchem.ncbi.nlm.nih.gov/#query=5864-38-0&amp;tab=substance</a>	CAS (Chemical Abstracts service)	Numerieke code bestaande uit 6 tot maximaal 10 getallen die in 3 groepen is opgebouwd (notatie: XXX-XXX-X). Het laatste getal dient om de overige getallen te checken op validiteit.
<b>(Canonical) SMILES</b>	Nee**	De (canonieke) SMILES is niet uniek voor stereo-isomeren, zoals bijvoorbeeld cis-trans isomerie.	Niet meer in beheer bij een organisatie. Conventies voor SMILES-naamgeving zijn aangeboden als een open standaard op OpenSMILES.	Conventie om van een ASCII-codering (m.b.v. een computerprogramma) een molecuulstructuur te reconstrueren. Een canonieke SMILES is uniek voor identificatie van een specifieke stof binnen een databank, maar niet geschikt als unieke identificatie daarbuiten.
<b>REWAB-identificatienummer</b>	N.b.	N.b.	Lijst in beheer bij KWR, RIVM is naamgevende instantie.	

<sup>7</sup> Zie voor een voorbeeld van een niet-unieke InChI: <https://pubchem.ncbi.nlm.nih.gov/#query=ZZXUAMCWBFFCJX-OUCXYWSSA-L>

<b>SIKB- registratienummer</b>	N.b.	N.b.	Informatiehuis Water (IHW), waterkwaliteitsparameters staan geregistreerd bij de 'Domeintabellen'	Het SIKB-nummer valt onder de Nederlandse Aquo- standaard, en is daarom in gebruik bij de Nederlandse watersector.
------------------------------------	------	------	---------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------

\* Uniek indien het een Preferred IUPAC-naam (PIN) betreft.

\*\*Isomerische SMILES worden gebruikt om onderscheid te maken tussen stereo-isomeren.

## 2 Gebruiksscenario's voor de identificatie van stoffen

Voordat stofgegevens van de waterlaboratoria zijn verzameld en opgenomen in een stoffenlijst (of databank), is als eerste stap een enquête verricht. Vervolgens is met behulp van de antwoorden en gesprekken met vertegenwoordigers van de waterlaboratoria een uitvraag verricht om te komen tot een gebruiksscenario dat het meeste kans van slagen heeft voor het verzamelen van stofgegevens.

Diverse informatiebronnen en databanken worden gebruikt voor identificatie van stoffen, zie Tabel 3 voor een niet-uitputtend overzicht.

Tabel 3: informatiebronnen voor identificatie van stoffen, met link en beschrijving.

Databank, of informatiebron	Link	Korte beschrijving of opmerking	Genoemd in enquête (E), of in definitie van gebruiksscenario (G)
Chemspider	<a href="https://www.chemspider.com/">https://www.chemspider.com/</a>	Databank en zoekmachine met daarin informatie over de structuur en naamgeving van een stof	E
Zoeksysteem risico's van stoffen (RIVM)	<a href="https://rvszoeksysteem.rivm.nl/">https://rvszoeksysteem.rivm.nl/</a>	Actuele, geautoriseerde informatie over risico's van stoffen voor mens en milieu.	E
NORMAN substance database (SusDat)	<a href="https://www.norman-network.com/nds/susdat/">https://www.norman-network.com/nds/susdat/</a>	Een lijst van tegenwoordig meer dan 111.000 stoffen samengevoegd uit alle NORMAN suspect-lijsten	E en G
Wikipedia	<a href="https://en.wikipedia.org/">https://en.wikipedia.org/</a>	Publiek toegankelijke ('vrije') encyclopedie.	E
ECHA REACH database	<a href="https://echa.europa.eu/information-on-chemicals/registered-substances">https://echa.europa.eu/information-on-chemicals/registered-substances</a>	Stoffen die bij de European Chemicals Agency (ECHA) staan geregistreerd.	E
Norman Massbank	<a href="https://massbank.eu/MassBank/">https://massbank.eu/MassBank/</a>	Databank met massaspectra van bekende, onbekende en voorlopig geïdentificeerde stoffen	E en G
NIST Chemistry WebBook	<a href="https://webbook.nist.gov/chemistry/">https://webbook.nist.gov/chemistry/</a>	Databank en zoekmachine met daarin informatie over de structuur en naamgeving van een stof, opgezet door	E

		National Institute of Standards and Technology (US Dept. of Commerce)	
<b>Toelatingen bestrijdingsmiddelen (CTBG)</b>	<a href="https://toelatingen.ctgb.nl/nl/authorisations">https://toelatingen.ctgb.nl/nl/authorisations</a>	Databank met daarin de toelatingen van gewasbeschermingsmiddelen, biocides en additieven.	E
<b>Pubchem</b>	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>	Publiek toegankelijke databank met informatie over de structuur en naamgeving van een stof.	E
<b>DSFP</b>		In DSFP kunnen ruwe meetgegevens opgeslagen en op massa's, bruto formule of stofnaam worden doorzocht, als gewenst ook met filtering op retentietijd door middel van een retentietijdindex (RTI).	G
<b>EmpoDat</b>	<a href="https://www.norman-network.com/nds/empodat/">https://www.norman-network.com/nds/empodat/</a>	NORMAN chemical occurrence database, databank met geo-informatie betreffende monitoring van stoffen en bio-monitoring.	G
<b>NORMAN Ecotoxicology database</b>	<a href="https://www.norman-network.com/nds/ecotox/">https://www.norman-network.com/nds/ecotox/</a>	Platform voor systematische verzameling en evaluatie van ecotoxiciteitsstudies voor geharmoniseerde afleiding van milieukwaliteitsnormen	G
<b>CompTox Chemicals Dashboard (US EPA)</b>	<a href="https://comptox.epa.gov/dashboard">https://comptox.epa.gov/dashboard</a>	Compilatie van chemische, fysische, toxiciteits en environmental fate informatie afkomstig van web pagina's, databases en bronnen, waaronder Amerikaanse federale en nationale bronnen en internationale instanties	E en G
<b>Toxcast database (US EPA)</b>		Databank met <i>in vitro</i> toxiciteitsgegevens	G
<b>MASST</b>	<a href="https://masst.ucsd.edu/">https://masst.ucsd.edu/</a>	Openbaar beschikbare MS2-databank	G
<b>myLibrary (Thermo Fisher Scientific)</b>	Online toegankelijk, maar beveiligd	MS databank van eigen opgenomen spectra die met andere laboratoria kan worden gedeeld	
<b>Eigen suspect-databanken</b>	niet online toegankelijk	Suspect-databanken van waterlaboratoria die in eigen beheer zijn.	E en G



## REWAB

Niet genoemd, maar zeker relevant in het kader van het opslaan van waterkwaliteitsgegevens in Nederland, is de REWAB-databank (Registratie Waterkwaliteit gegevens Bedrijven). REWAB is een essentiële pijler onder het proces van registreren, archiveren en digitaliseren van de (openbare) waterkwaliteitsgegevens. De drinkwaterbedrijven zijn wettelijk verplicht om jaarlijks (i) het uitgevoerde meetprogramma inclusief meetgegevens en (ii) het voorgenomen meetprogramma voor het komende jaar te laten beoordelen door de VROM-inspectie (IL&T). Dit is een verplichting op grond van artikel 26, lid 2 en 3: Bijlage A van het Drinkwaterbesluit en Bijlage 3, 4, en 5 van de Drinkwaterregeling. Het digitaal aanleveren van de meetgegevens aan IL&T vindt plaats met behulp van de REWAB software. Naast deze wettelijke verplichting hechten de drinkwaterbedrijven ook belang aan het goed beheer van gegevens en deze incidenteel, voor bijvoorbeeld BTO, beschikbaar te stellen. Het genereren en opslaan van meetgegevens ten aanzien van wettelijke parameters wordt uitgevoerd door de drinkwaterlaboratoria. KWR faciliteert het samenbrengen van de meetgegevens in de REWAB-databank. Het beheer van de parameterlijst wateronderzoek binnen REWAB behoort tot het takenpakket van de Beheerscommissie Overdracht Kwaliteitsgegevens (BOK).

### 2.1 Enquête

De enquête bevat vragen met betrekking tot:

- Gebruik van zoekfuncties in chemische databanken voor het identificeren en/of opzoeken van chemische informatie van een stof. Zie Tabel 3 voor een overzicht;
- Het type informatie dat wordt gezocht (bijv. toxicologische informatie, signaleringslocatie, etc.);
- De wijze van presentatie van informatie: is deze voldoende, en waar gaat de voorkeur naar uit (in de vorm van bijvoorbeeld informatiebladen *'fact sheets'*, of voorkeur voor presentatie van doelgerichte informatie).
- De frequentie van raadplegen van externe informatiebronnen.

De vragenlijst is door ten minste één vertegenwoordiger van een drinkwaterlaboratorium ingevuld, in totaal vulden 7 personen de vragenlijst in. Uit de enquête kwam naar voren dat:

- Bij het raadplegen van stoffen voornamelijk de ChemSpider databank en de 'NORMAN substance database' (NORMAN SusDat) worden gebruikt. Daarnaast wordt er veelvuldig gezocht naar risico's van stoffen in het RIVM-zoeksysteem voor risico's van stoffen. Andere databanken worden minder vaak geraadpleegd.
- De frequentie van het zoeken naar stofgegevens is ca. een aantal keren per maand, alleen bij KWR is dat (bij enkele onderzoekers) een aantal keren per dag.
- Het merendeel van de respondenten geeft aan dat de voorkeur uitgaat naar internetpagina's of zoekmachines die gerichte, maar zo compleet mogelijke informatie bieden. Daarnaast is er de behoefte dat gegevens overzichtelijker worden gepresenteerd, onduidelijk is om welke gegevens dit gaat.

De antwoorden op de vragenlijsten staan vermeld in Bijlage I.

### 2.2 Uitvraag gebruiksscenario's

Om de gebruiksscenario's beter in beeld te krijgen is, na de enquête, een aparte uitvraag per email verricht. Enkele typische kennisvragen zijn opgesteld die leidend zijn in een gebruiksscenario, zoals bijvoorbeeld 'Waar zijn de

suspects gevonden?'. Deze vragen zijn verstuurd naar analisten bij de drinkwaterlaboratoria met het verzoek de lijst van vragen aan te vullen. Vervolgens zijn de kennisvragen gegroepeerd volgens drie uitgangspunten, namelijk:

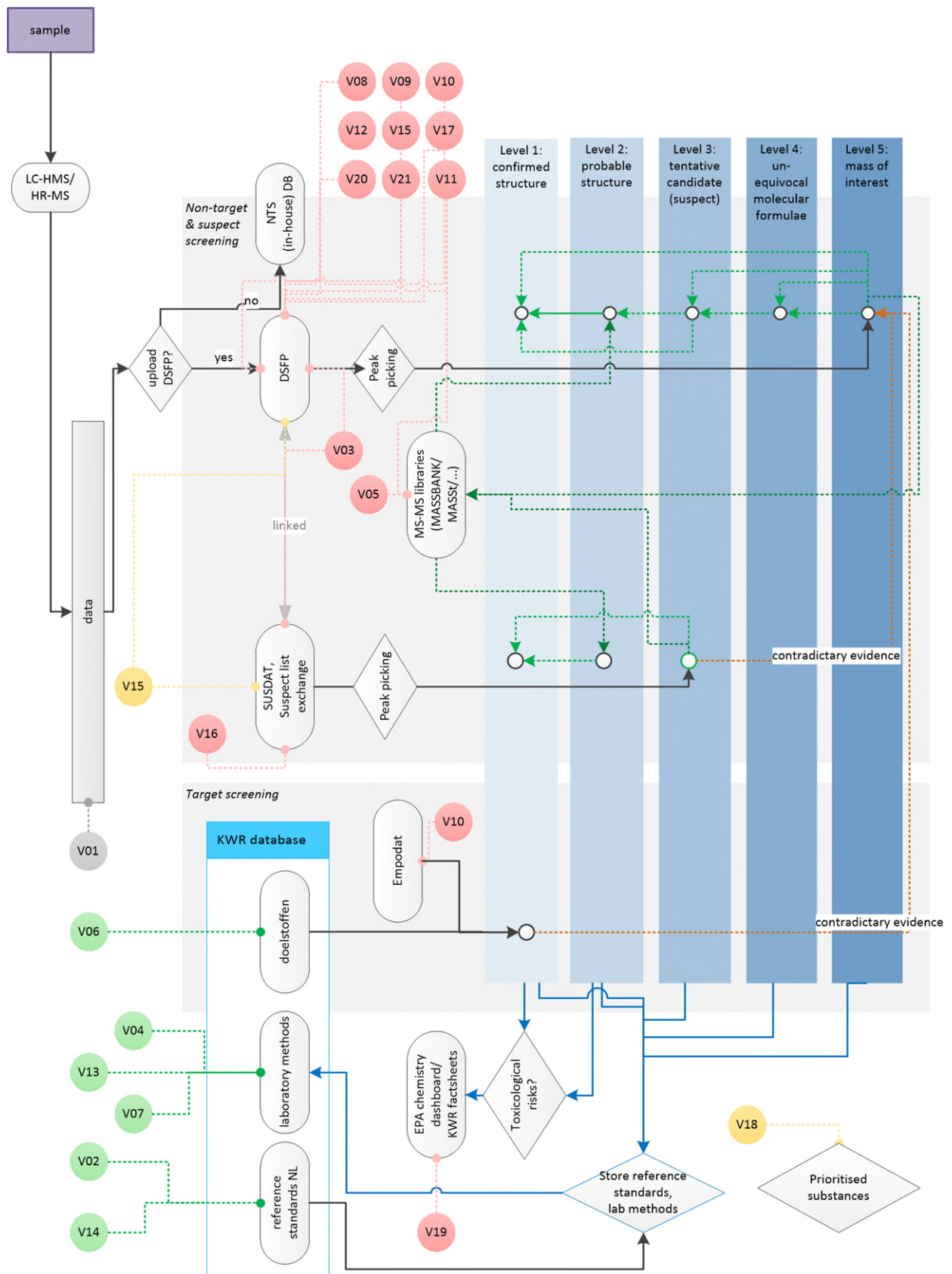
1. Het gebruiken van **bestaande, internationaal opgezette databanken** zoals de NORMAN substance database of het NORMAN DSFP. Voor een suspect en NTS-doeleinde, is het gebruik van DSFP een logische keus. Voor deze functionaliteit is het noodzakelijk om per meetserie een RTI-standaard ingesloten te hebben. Meer informatie staat in Hoofdstukken 4 en 5.
2. Gegevens die **niet eenvoudig beschikbaar zijn in internationale databanken**, maar mogelijk wel waardevol. Bijvoorbeeld suspects die op bepaalde locaties zijn gemeten of zuiveringsefficiënties die bepaald zijn voor (zorgwekkende) stoffen. De opzet van een (Nederlandse) databank kan mogelijk uitkomst bieden. Een deel van de gebruiksscenario's in deze categorie van kennisvragen vergt nadere uitwerking.
3. **Externe, veelal leverancier-specifieke bibliotheken**. In veel gevallen zijn er bijvoorbeeld lokaal bibliotheken beschikbaar met MS/MS-data. Deze zijn veelal apparaat-specifiek, maar er zijn ook niet-commerciële bibliotheken beschikbaar zoals bijvoorbeeld de NORMAN Massbank.

In Tabel 4 staan de resultaten van de uitvraag, gerangschikt op de categorie zoals in de lijst hierboven en inclusief de databanken die gebruikt kunnen worden bij het beantwoorden van de kennisvragen. Figuur 2 toont schematisch de samenhang van gebruiksscenario's die onderdeel uitmaken van het screeningsproces.

Tabel 4: overzicht van gebruiksscenario's, kennisvragen en databanken die relevant zijn bij de screening van stoffen. Cat.: categorie, SD: Norman SusDat.

Cat.	Gebruiksscenario	Vraag	Vragen behorende bij scenario	Databank
1	Inzicht in welke suspects waar worden gescreend, zodat bij dit laboratorium om nadere informatie kan worden gevraagd, en/of het laboratorium de stof kan analyseren.	V03	Welk (Nederlands) laboratorium heeft welke suspects gemeten?	DSFP/SD
	Inzicht in non-target/suspect screening van een stof, nuttig voor verdere afhandeling en/of prioritering	V08	Is een screeningresultaat (onbekende stof) in een ander lab gevonden?	DSFP
	Inzicht in welke suspects waar worden gemeten, nuttig voor verdere afhandeling en/of prioritering.	V09	Waar zijn de suspects gevonden?	DSFP/SD
	Inzicht in de features van NTS-data	V11	Wat zijn de gegevens per feature (genormaliseerde retentietijd, accurate massa, identificatieniveau volgens Schymanski-schema L1-L5, etc.)	DSFP
	Inzicht in het identificatieniveau van een stof	V12	Welke labs hebben welke stoffen al geïdentificeerd (targets=L1-niveau)?	DSFP
	Inzicht in het voorkomen van een stof, deze informatie kan nuttig zijn voor prioritering	V17	In welk type water (grondwater, oppervlaktewater, drinkwater) zijn suspects en/of bekende stoffen gemeten?	DSFP

	Inzicht in stofcategorieën, deze informatie kan nuttig zijn voor prioritering	V20	Tot welke stofklasse hoort een (target)stof?	Ruime keuze metadata voor DSFP
	Inzicht in onzekerheid van identificatie van een stof	V21	Zoekopdrachten met een "window" rond de accurate massa	DSFP
	Inzicht in welke suspects worden gemeten, nuttig voor verdere afhandeling en/of prioritering	V16	Wat zijn de (huidige) suspect-lijsten bij de drinkwaterlaboratoria in Nederland?	SD
	Inzicht in doelstofanalyses	V10	Welke targets worden er gemeten door welke labs, en hoe vaak?	NORMAN Empodat
	Inzicht in toxiciteit van een stof	V19	Welke informatie m.b.t. humane toxiciteit is er van een (nieuw aangetoonde) stof?	NORMAN Ecotoxicology module, Chemistry Dashboard (US EPA), Toxcast (US EPA)
	Inzicht in identificatie van een stof (wereldwijd)	V22	Is een onbekende stof ooit eerder ergens in de wereld aangetroffen?	Mass-t
<b>2</b>	Inzicht in methodieken/standaarden die door laboratoria in gebruik zijn om stoffen te meten, nuttig voor verdere afhandeling	V06	Welke doelstoffen worden in de screeningsmethoden gebruikt (referentiestandaarden)?	<i>N.b.</i>
		V13	Er is een stof gevonden in literatuur/media, maar deze stof kan door ons lab niet gemeten worden. Welk lab heeft de analysemethode in huis?	evt. mogelijk in DSFP/ Mass-t
		V14	Is er een referentiestandaard voor de bevestiging van een suspect? (van Schymanski-level 2 en 3 naar level 1)	<i>N.b.</i>
		V04, V07	Wat zijn de target, suspect en NTS - methoden die in het lab worden gebruikt?	<i>N.b.</i>
		V02	Welk (Nederlands) laboratorium heeft een methode om een bepaalde (bekende) stof te meten	<i>N.b.</i>
	Inzicht in veel voorkomende features van NTS-data	V15	Welke features moeten geprioriteerd worden voor moeilijk te identificeren stoffen?	Veel features worden bijgehouden in diverse, interne databanken van de drinkwaterlaboratoria
Inzicht in suspect-screening en prioritering	V18	Wat zijn alle, door de bedrijfstak geprioriteerde, stoffen waarvoor suspect screening mogelijk is?	Use case Vitens	
<b>3</b>	Inzicht in de beschikbaarheid van spectrale bibliotheken, zodat er effectiever een gelijkenis (match) is	V05	Wat zijn de gebruikte spectrale bibliotheken, en worden deze intern gemaakt of worden commerciële gebruikt?	NORMAN MassBank, myLibrary (Thermo Fisher Scientific)



Figuur 2: workflows voor screening van stoffen met daarin een voorstel om gegevens voor doelstof-screening structureel en centraal op te slaan (KWR databank). Kennisvragen zijn met bolletjes aangegeven en zijn gepresenteerd in Tabel 2..

### 2.3 Definitie van gebruiksscenario

In overleg met vertegenwoordigers van de drinkwaterlaboratoria is het volgende gebruiksscenario naar voren gekomen als eerste stap om tot gegevensuitwisseling te komen:

*Er is voldoende draagvlak voor het ontwikkelen van een basisfunctionaliteit voor het delen van gegevens over onbekende en bekende (doel)stoffen en suspect-lijsten. Het heeft de voorkeur om in eerste instantie te werken met een eenvoudige spreadsheet, die kan worden aangevuld door de labs. Na ontsluiting van de lijst via een beveiligde omgeving kan na een evaluatieperiode worden besloten om de lijst door upload naar de NORMAN suspect list exchange te delen en het doorzetten van de gegevens naar een databank zoals het DSFP-platform (Hoofdstuk 4) verder te automatiseren.*

## 3 Verzameling van doelstoffen en suspects gemeten door drinkwaterlaboratoria

### 3.1 Aanpak

Aan de drinkwaterlaboratoria is gevraagd om verschillende gegevens aan te leveren ten aanzien van doelstoffen ('targets') en nog niet geïdentificeerde stoffen ('suspects'). De uitvraag voor suspects leverde te weinig op om te verwerken tot een databestand. De uitvraag voor targets is gedaan om inzicht te krijgen in (a) de analyse van een bepaalde stof door een laboratorium, (b) de gehanteerde analysemethodiek en (c) het voorkomen van een stof in oppervlakte- of grondwater. Hiertoe zijn de gegevens zoals weergegeven in Tabel 5 opgevraagd.

Tabel 5: overzicht van gevraagde gegevens met betrekking tot geanalyseerde doelstoffen.

<b>Identificatie van stof</b>	<b>Naam, afkorting van stofnaam (optioneel), REWAB-identificatienummer (optioneel), SIKB-identificatienummer (optioneel), CAS-registratienummer</b>
<b>Methodiek</b>	Meetmethode, detectiegrens, eenheid van detectiegrens
<b>Voorkomen</b>	Geaccrediteerd in oppervlaktewater (ja/nee), geaccrediteerd in grondwater (ja/nee), gevonden in grondwater (ja/nee), gevonden in oppervlakte water (ja/nee)

### 3.2 Verwerking data

#### 3.2.1 Databestanden klaar maken voor verwerking

Van HWL, WLN, Vitens en KWR zijn vervolgens data ontvangen. In Tabel 7 staan de aantallen unieke parameters en het totaal aantal rijen met parameters dat is aangeleverd per laboratorium. De parameters werden in alle gevallen het meest volledig omschreven door een naam, en bij HWL, WLN en het Vitens-laboratorium deels ook door CAS-registratienummers. De gegevensset van KWR bevatte ook parameters die chemische stofgroepen of -categorieën aanduiden, zoals 'geur' en parameters die aangeven dat ze gemeten zijn na een bepaalde voorbehandeling zoals 'na filtratie'. Deze parameters zijn zoveel mogelijk verwijderd. Ook de parameters waarvan aangegeven was dat ze niet meer gemeten worden zijn verwijderd. In de gevallen dat er in de dataset geen Engelse naam gegeven was, is dit aangevuld met de gegevens uit de kolom Nederlandse n Voor HWL werden er drie bestanden aangeleverd die samengevoegd zijn tot één bestand. Bij Vitens zaten de gegevens voor diverse matrices (drinkwater, afvalwater, etc.) in één enkele kolom. Deze informatie is per matrix in een aparte kolom gezet. De kolomnamen die de laboratoria aanhouden in hun bestanden zijn omgezet naar standaard kolomnamen zoals te zien is in Tabel 6. In Tabel 7 is een overzicht te zien van de informatie die aanwezig was in de door de laboratoria aangeleverde bestanden.

Tabel 6: Gestandaardiseerde namen op basis van de door de laboratoria aangeleverde originele naam, en de betekenis daarvan.

Gestandaardiseerde naam	Originele naam	Betekenis
HWL_naam	NAME	Parameternaam gehanteerd door het laboratorium
HWL_Type	ANALYSIS	Soort Parameter geanalyseerd
HWL_Eenheid	HWL_REPORT_UNITS	Eenheid waarin gerapporteerd wordt
HWL_Q_OW	Q_OW	Kwaliteitsstandaard aanwezig voor oppervlaktewater
HWL_Q_GW	Q_GW	Kwaliteitsstandaard aanwezig voor grondwater
HWL_Q_PW	Q_PW	Kwaliteitsstandaard aanwezig voor proceswater
HWL_Q_DW	Q_DW	Kwaliteitsstandaard aanwezig voor drinkwater
WLN_naam	Omschrijving	Parameternaam gehanteerd door het laboratorium
WLN_Methode	Methode (file 1)	Methode van meten
WLN_RG	RG	Rapportagegrens
WLN_Eenheid	Eenheid	Eenheid waarin gerapporteerd wordt
WLN_Gebruik.Q	Gebruik.Q (file 1)	Kwaliteitsstandaard aanwezig
WLN_gw_Methode	Methode (file 3)	Methode van meten in grondwater
WLN_ow_Methode	Methode (file 2)	Methode van meten in oppervlaktewater
WLN_Q_OW	Gebruik.Q (file 2)	Kwaliteitsstandaard aanwezig voor oppervlaktewater (0 is nee, 1 is ja)
WLN_Q_GW	Gebruik.Q (file 3)	Kwaliteitsstandaard aanwezig voor grondwater (0 is nee, 1 is ja)
VIT_naam	Componentnaam	Parameternaam gehanteerd door het laboratorium
VIT_RG	Rap.Grens	Rapportagegrens
VIT_Type	Analysenaam	Soort Parameter geanalyseerd
VIT_Methode	Meettechniek	Methode van meten
VIT_Eenheid	Eenheid	Eenheid waarin gerapporteerd wordt
VIT_Q_OW	Oppervl-water	Kwaliteitsstandaard aanwezig voor oppervlaktewater
VIT_Q_GW	Grondwater	Kwaliteitsstandaard aanwezig voor grondwater
VIT_Q_PW	Proceswater	Kwaliteitsstandaard aanwezig voor proceswater
VIT_Q_AW	Afvalwater (kolom 'Matrix')	Kwaliteitsstandaard aanwezig voor afvalwater
VIT_Q_DW	Drinkwater	Kwaliteitsstandaard aanwezig voor drinkwater
KWR_naam	EN_REPORTED_NAME	Parameternaam gehanteerd door het laboratorium
KWR_RG	LIMIT	Rapportagegrens
KWR_Methode	ANALYSIS	Methode van meten
KWR_Eenheid	UNIT	Eenheid waarin gerapporteerd wordt

Tabel 7: Informatie met betrekking tot chemische metingen gerangschikt per laboratorium. Waar geen informatie beschikbaar is, is een lege cel. Zie Tabel 6 voor een nadere beschrijving van de betekenis van de informatielabels.

	HWL	WLN	Vitens	KWR
<b>LabNaam</b>	HWL_naam	WLN_naam	VIT_naam	KWR_naam
<b>Methode</b>		WLN_Methode	VIT_Methode	KWR_Methode
<b>Rapportagegrens</b>		WLN_RG	VIT_RG	KWR_RG
<b>Type</b>	HWL_Type		VIT_Type	
<b>Eenheid</b>	HWL_Eenheid	WLN_Eenheid	VIT_Eenheid	KWR_Eenheid
<b>Quality</b>		WLN_Gebruik.Q		
<b>Methode GW</b>		WLN_gw_Methode		
<b>Methode OW</b>		WLN_ow_Methode		
<b>Quality OW</b>	HWL_Q_OW	WLN_Q_OW	VIT_Q_OW	
<b>Quality GW</b>	HWL_Q_GW	WLN_Q_GW	VIT_Q_GW	
<b>Quality PW</b>	HWL_Q_PW		VIT_Q_PW	
<b>Quality AW</b>			VIT_Q_AW	
<b>Quality DW</b>	HWL_Q_DW		VIT_Q_DW	

### 3.2.2 CAS-registratienummers en andere identificatienamen toekennen

De namen in de bestanden zijn via Chemistry Dashboard<sup>8</sup> gekoppeld aan standaard identificatienamen (*identifiers*), voor zover deze herkend werden. Chemistry Dashboard zoekt ook op synoniemen van namen. Ook zijn de parameters gekoppeld met AQUO (<http://www.aquo.nl/over-aquo/aquo-onderdelen/aquo-parameterlijsten/>) met gestandaardiseerde Nederlandse naam, SIKB-code en afkorting van parameters. In AQUO staan ook standaard namen voor metingen aan combinaties van stoffen. Daarnaast zijn parameters, waar mogelijk, gekoppeld aan REWAB codes.

### 3.2.3 Gegevens van stoffen van de laboratoria koppelen

Daarna zijn de parameters van alle laboratoria aan elkaar gekoppeld, zodat een parameter op een rij staat met de gegevens van de vier labs erachter. Dit is gedaan op basis van de CAS-registratienummers uit het Chemistry Dashboard of waar deze niet beschikbaar waren, de eigen CAS-registratienummers. Waar laboratoria meerdere namen voor een parameter aanhouden zijn deze behouden, waardoor een parameter in meerdere regels kan staan. Op deze manier kan elk lab op basis van eigen gehanteerde namen parameters terugvinden in de lijst.

In de bestanden van de laboratoria waren parameters soms ook meerdere keren aanwezig door bijvoorbeeld registratie van meerdere meetmethoden. Deze zijn wel teruggebracht tot een enkele rij, waarbij de verschillende meetmethoden in een cel bij elkaar staan. Ook de bijbehorende rapportagegrenzen zijn op deze manier in één cel bij elkaar gebracht.

### 3.2.4 Handmatige identificatie van stoffen

Voor parameternamen die niet herkend werden via het opzoeken in het CompTox Chemistry Dashboard (zie Tabel 3, Hoofdstuk 2) en waarbij geen CAS-registratienummer vermeld was, zijn de parameternamen nog getest op aanwezigheid in de gekoppelde AQUO omschrijving bij stap 2, en de namen van de andere laboratoria die reeds verwerkt waren. Als dit zo was, is het CAS-registratienummer van deze naam gebruikt. Daarna is op de rest per parameternaam online het CAS-registratienummer gezocht, in diverse bronnen zoals verkopers van chemische standaarden, Wikipedia en PubChem. Als twee bronnen hetzelfde CAS-registratienummer hadden voor de parameter is deze als voorkeur opgenomen. Deze parameters zijn ook nog in het Chemistry Dashboard geconsulteerd om de aanvullende gegevens bij te voegen. Hierbij vervielen weer enkele CAS-registratienummers omdat deze niet herkend werden door deze tool. Parameters die uiteindelijk niet te koppelen waren (geen CAS-

<sup>8</sup> [https://comptox.epa.gov/dashboard/dsstoxdb/batch\\_search](https://comptox.epa.gov/dashboard/dsstoxdb/batch_search)



registratienummer kunnen achterhalen, geen CAS-registratienummer dat bij een ander lab bekend was) zijn als aparte rijen wel alsnog toegevoegd. Er is gecontroleerd of de gegeven CAS-registratienummers valide waren. Het laatste nummer moet namelijk een resultaat zijn van de eerdere nummers in de CAS code, en er is een vast format. Indien dit niet klopte is de correcte CAS-registratie op basis van de naam die het laboratorium hanteert via internet (Google) opgezocht. Als laatste stap is gecontroleerd of er standaard IUPAC- of Aquo-namen waren met verschillende CAS-registratienummers. In dat geval is er mogelijk een verkeerd of oud CAS-registratienummer gebruikt. Dit was niet het geval.

### 3.3 Resultaat

Het samenbrengen van alle informatie resulteerde in een lijst met 1135 unieke parameters geïdentificeerd op CAS-registratienummer, met per rij een stof met een aantal standaard identifiers en daarna met gegevens van de vier laboratoria (waar gemeten). De lijst is in totaal 1508 rijen lang. In totaal 221 parameters konden niet geïdentificeerd worden. Deels komt dit omdat het geen stoffen zijn, maar bijvoorbeeld parameters zoals DOC. Daarnaast komen sommige parameters toch nog dubbel voor vanwege de verschillende namen die door de laboratoria zijn toegekend aan dezelfde parameters binnen een laboratorium. Daardoor beslaan deze parameters meerdere rijen. Deze zijn behouden om labs de mogelijkheid te geven op hun eigen gehanteerde naam te zoeken in de databank. In totaal worden er 38 parameters door alle vier de laboratoria gemeten. Tabel 8 geeft een overzicht van de parameters die per laboratorium uniek aanwezig waren in de bestanden, en het aantal parameters dat uiteindelijk niet geïdentificeerd kon worden. In Figuur 3 is als voorbeeld een screenshot te zien van de eerste regels van de algemene informatie per parameter in het geïntegreerde databestand.

Tabel 8: Resultaten in aantallen parameters, totaal, uniek gemeten in het laboratorium en niet geïdentificeerd met een CAS nummer.

	Aantal rijen aangeleverde spreadsheet	Parameters (op laboratoriumnaam)	Uiteindelijk niet geïdentificeerde parameters	Parameters met CAS nummer (uniek)	Parameters met CAS nummer uniek voor het laboratorium
<b>HWL</b>	431	340	15	325	53
<b>WLN</b>	543	527	35	492	98
<b>Vitens</b>	7295	949	99	849	211
<b>KWR</b>	1528	520 (na verwijdering van niet-chemische parameters)	72	448	209

A	B	C	D	E	F	G	H	I	J	K
	CASnr	AquoOmschrijving	AquoCode	IUPAC_NAME	PREFERRED_NAME	SIKbcode	REWABcode.s.	SMILES	INCHIKEY	MONOISOTOPIC_MASS
1	[7143-09-1]	NA	NA	Methyl (1R,2R,3S, Ecgonine methyl ester	NA	NA	NA	COC(=O)[C@H](O)N		199.12084
2	[100-02-7]	4-nitrofenol	4NO2Fol	4-Nitrophenol	4-Nitrophenol	235	8461	OC1=CC=C(C=C1)[N+](=O)[O-]	BTJIUGUIF	139.026943
3	[100-41-4]	ethylbenzeen	C2yBen	Ethylbenzene	Ethylbenzene	728	1089	CCC1=CC=CC=C1	YNQLUTRI	106.0782503
4	[100-42-5]	styreen	styrn	Ethenylbenzene	Styrene	1503	1088	C=CC1=CC=CC=C1	PPBRXRYQ	104.0626003
5	[100-61-8]	methylaniline	C1yAn	N-Methylaniline	N-Methylaniline	3324	1700	CNC1=CC=CC=C1	AFBPFQWN	107.0734993
6	[100-75-4]	nitrosopiperidine	NOpprdne	1-Nitrosopiperidin	N-Nitrosopiperidine	2433	2141	O=NN1CCN1	UWSDONI	114.079313
7	[100-88-9]	cyclamaat	cycmt	Cyclohexylsulfam	Cyclamic acid	5887	2280	OS(=O)(=O)C1=CC=CC=C1	HCAJEUSC	179.0616145

Figuur 3: Screenshot van de gemeenschappelijke informatie voor elke geïdentificeerde parameter.

### 3.4 Conclusie

Het bleek dat geen van de laboratoria consequent internationale standaard identificatienamen (zoals InChI, SMILES, of IUPAC-naam) hanteerde, en ook CAS nummers niet consequent werden gebruikt waardoor het koppelen werd bemoeilijkt. Voor een klein deel van de aangeleverde parameters is dit ook niet gelukt (zie Tabel 7). Ook voor de

analyse van groepen van chemicaliën zijn geen standaard namen gebruikt. Alleen WLN gebruikte voor een deel hiervan Aquo codes.

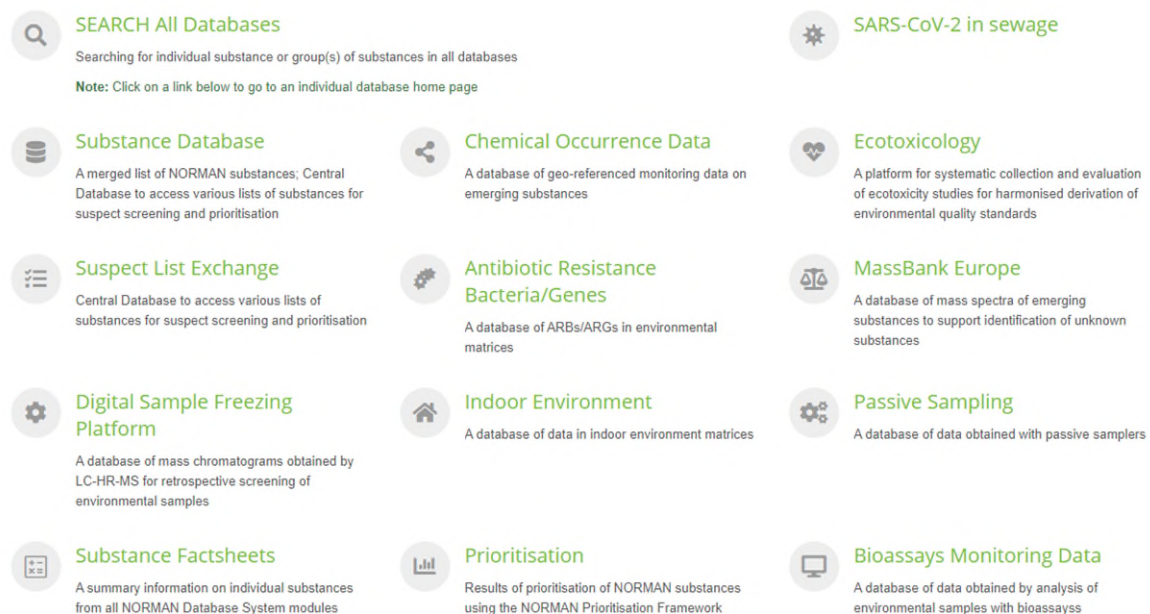
Elk laboratorium heeft een deels unieke set van parameters die gemeten worden, of in het verleden gemeten zijn (zie Tabel 8). De geaggregeerde stoffenlijst kan gebruikt worden om een idee te krijgen van het soort parameters en welke parameters er gemeten worden in diverse matrices (oppervlaktewater, grondwater, etc.). De lijst is niet geschikt om te kijken welke parameters in welke matrix (oppervlaktewater, grondwater, etc.) gevonden zijn, omdat deze informatie ontbreekt.

## 4 NORMAN-netwerk en databanken

### 4.1 Het NORMAN-netwerk van Europese laboratoria, onderzoeksinstituten en universiteiten

NORMAN is een Europees netwerk van laboratoria, onderzoeksinstituten en universiteiten voor milieumonitoring. Doel van het netwerk is de uitwisseling van informatie over opkomende milieurelevante stoffen en het versterken van inspanningen naar validatie en harmonisatie van gemeenschappelijke meetmethoden en monitoringtools. Daardoor zou beter kunnen worden voldaan aan de eisen van risicobeoordelaars en risicomangers. NORMAN is specifiek bedoeld om de kennisuitwisseling tussen onderzoeksteams uit verschillende landen op het gebied van opkomende stoffen te bevorderen. Onder andere organiseert NORMAN de ontwikkeling en het onderhoud van verschillende webgebaseerde databanken voor het verzamelen en evalueren van gegevens over opkomende stoffen in het milieu in het NORMAN Database System (Figuur 4). Deze databanken bevatten gegevens over doelstoffen (Substance Factsheets) en doelstofmetingen (Chemical Occurrence Data), suspectlijsten (Substance Database, Suspect List Exchange), ruwe NTS data (DSFP en MassBank Europe), toxiciteit (Ecotoxicology en Bioassay Monitoring Data), en gespecialiseerde modules voor bepaalde matrices (Indoor Environment en Passive Sampling) en gevaren (Antibiotic Resistance Bacteria/Genes en SARS-CoV-2-in rioolwater).

KWR vertegenwoordigt de Nederlandse waterlaboratoria in het NORMAN-netwerk en vertaalt nieuwe inzichten uit NORMAN onderzoek, het Joint Programme of Activities, naar de Nederlandse drinkwatersector. Onlangs zijn ook de waterlaboratoria HWL, Vitens en WLN lid van NORMAN geworden. Dit maakt het aantrekkelijk om de NORMAN infrastructuur te gebruiken voor doelstof-, suspect- en NTS data-analyse en -opslag.



Figuur 4. Schematische weergave van de modules van de NORMAN databanken.

### 4.2 Suspect list exchange

De NORMAN Suspect List Exchange (NORMAN-SLE) werd in 2015 opgericht als een centraal toegangspunt voor NORMAN-leden (en anderen) om suspectlijsten te vinden die relevant zijn voor NTS in het kader van milieumonitoring. De NORMAN-SLE bevat alle individuele lijsten die deel uitmaken van NORMAN SusDat, de

gefuseerde NORMAN Substance Database (DOI: 10.5281 / zenodo.2664077), zoals bijvoorbeeld de STOFF-IDENT Databank van waterrelevante stoffen, een lijst met REACH geregistreerde stoffen, en verschillende doelstof- en suspectlijsten.

Bij KWR wordt voor suspect screening routinematig onder andere de NORMAN SusDat databank gebruikt. KWR heeft rond 35000 stoffen uit de NORMAN SusDat inclusief chemische structuren als suspectlijst in “.masslist” formaat voor Compound Discoverer (Thermo Fisher Scientific) beschikbaar gemaakt. Daardoor kan in Compound Discoverer automatisch naar matches met *in silico* voorspelde spectra van deze stoffen worden gezocht. In het NORMAN DSFP (zie 4.3) kan NORMAN SusDat of een selectie van NORMAN suspect(s) en/of -lijsten worden gebruikt voor een automatische suspect screening.

Suspectlijsten die bij laboratoria in gebruik zijn kunnen aan de NORMAN-SLE worden toegevoegd. KWR heeft al twee suspectlijsten met NORMAN gedeeld. Deze bevatten in Nederland gedetecteerde stoffen (KWRSJERPS, 10.5281/zenodo.2621941) en stoffen die in Nederlands water worden verwacht door hun productievolumes en/of emissieroutes (KWRSJERPS2, 10.5281/zenodo.2648817). Chemische data van de stoffen in NORMAN SusDat zoals structuur en toxicologische informatie is via de NORMAN Substance Factsheets en het Chemistry Dashboard van de US Environmental Protection Agency (<https://comptox.epa.gov/dashboard>) beschikbaar. Toxiciteitsdata maken ook deel uit van de ecotoxicologie-module.

### 4.3 NORMAN Digital Sample Freezing Platform

Gegevens over het voorkomen van stoffen zijn bij NORMAN opgeslagen in EMPODAT (chemical occurrence database) als het doelstoffen betreft, en in het DSFP als het suspect en NTS data betreft. Het DSFP is een databank voor de digitale archivering van NTS data, i.e. vloeistofchromatografie hoge resolutie massaspectrometrie (LC-HRMS) data, en voor de retrospectieve screening van monsters. De databank bevat ruwe LC-HRMS data, i.e. massachromatogrammen en geassocieerde fragmentatiespectra. Het DSFP werd ontwikkeld voor de retrospectieve suspect screening van duizenden milieuverontreinigende stoffen met de ambitie om een Europese - en mogelijk wereldwijde - standaard te worden. Het DSFP omvat alle recente ontwikkelingen m.b.t. suspect en NTS binnen het NORMAN netwerk. Dit betekent dat suspect screening mogelijk is met de suspectlijsten uit SusDat, gebaseerd op massa of bruto formule en een gespecificeerde massa afwijking. Voorlopig geïdentificeerde suspect kandidaten kunnen verder worden bevestigd gebaseerd op retentietijdindices, en MS2 data uit Massbank. Daardoor maakt het DSFP een volledige suspect screeningsanalyse mogelijk, van eigen data en openbare data die door anderen op internet gezet is en gedeeld. Dit faciliteert het vergelijken van resultaten en het monitoren van ruimtelijke en temporele trends van stoffen. Het potentieel ervan werd aangetoond door de monitoring van 670 antibiotica en 777 REACH-chemicaliën van de Joint Black Sea Surveys 6 (JBSS<sup>9</sup>).

Het uploaden van ruwe meetgegevens naar het DSFP-platform maakt een retrospectieve screening naar suspects mogelijk. Bovendien kan het delen van de gegevens via het platform ervoor zorgen dat in de ( nabije) toekomst het identificatieniveau van meer features naar een hoger plan kan worden gebracht, als ook het effectiever prioriteren van onbekende stoffen die in meerdere bronnen voorkomen. Tijdens de upload naar het DSFP kan tussen verschillende privacy-niveaus worden gekozen. Daardoor kunnen data doelgericht met technici van een lab of met alle Nederlandse water laboratoria worden gedeeld of zelfs helemaal openbaar gemaakt worden. Ook kan de ruimtelijke resolutie van de data worden gedefinieerd, bijvoorbeeld door het vervangen van de exacte locatie op basis van GPC coördinaten door een provincie- of een riviernaam. Dit is een belangrijke aspect vanwege de gevoeligheid van drinkwater-gerelateerde meetgegevens.

---

<sup>9</sup> NORMAN digital sample freezing platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in “digitally frozen” environmental samples. Nikiforos A. Alygizakis, Peter Oswald, Nikolaos S. Thomaidis, Emma L. Schymanski, Reza Aalizadeh, Tobias Schulze, Martina Oswaldova, Jaroslav Slobodnik, TrAC Trends in Analytical Chemistry, <https://doi.org/10.1016/j.trac.2019.04.008>.

In Hoofdstuk 5 zijn de functionaliteiten van het DSFP voor de retrospectieve screening van NTS-data van watermonsters verkend. Daarbij is gebleken dat de upload van NTS gegevens nog al moeizaam is. Een toekomstige automatisering van de upload zou de drempel voor de drinkwaterlaboratoria om het DSFP te gebruiken aanzienlijk verkleinen. Een vervolproject zou daarom ook de mogelijkheden voor automatisering moeten verkennen.

#### 4.4 MassBank.eu

Zoals de meeste NTS data-analyse workflows gebruikt ook het DSFP, data van MS2 fragmentatiespectra uit een bibliotheek met massaspectra om de betrouwbaarheid van de geïdentificeerde suspects te verhogen tot niveau 2 of 3 en uiteindelijk de structuur van de suspects te bevestigen. Daarvoor is het DSFP aan MassBank.eu (<https://massbank.eu/MassBank/>) gekoppeld, een databank met massaspectra van bekende, onbekende en voorlopig geïdentificeerde stoffen. De 2020.09 versie van MassBank bevat 88168 unieke spectra van een totaal van 14838 stoffen gemeten op QTOF en Orbitrap instrumentatie, geüpload door universiteiten en onderzoeksinstituten. KWR heeft met 207 entries aan MassBank bijgedragen. De upload wordt gedaan met hulp van de R package RMassBank<sup>10</sup>. Dit package levert een workflow om MS2-bestanden te verwerken en MassBank-records op te bouwen. De functies omvatten geautomatiseerde extractie van MS2-spectra, formuletoewijzing aan MS2-fragmenten, hercalibratie van MS2-spectra met toegewezen fragmenten, opschonen van spectra, geautomatiseerd ophalen van samengestelde informatie uit internetdatabanken en exporteren naar MassBank.

De beschikbaarheid van een massaspectrum is nodig om een voorlopig geïdentificeerd suspect te bevestigen. Hoe meer stoffen met MS2 spectra in MassBank beschikbaar zijn hoe betrouwbaarder de identificatie. Daarom is het wenselijk dat de drinkwaterlaboratoria hun MS2 spectra van doelstoffen, suspects en ook onbekende verbindingen naar MassBank uploaden. Dit lijkt een efficiënte en makkelijke manier om NTS data te delen en te hergebruiken. In een vervolproject zou de upload en het gebruik van MassBank kunnen worden verkend.

#### 4.5 Conclusie

Binnen het NORMAN-netwerk is er een uitgebreide infrastructuur aan databanken beschikbaar voor, enerzijds het opzoeken van chemische, toxicologische en gegevens over het voorkomen van een stof, en anderzijds het opslaan van screeningsdata en -resultaten. Deze databanken worden door NORMAN onderhouden. Hierdoor is het gebruik van dit netwerk vanuit oogpunt van kosten aantrekkelijk, en wordt de databank door de leden van het NORMAN-netwerk met data gevoed. Data kunnen openbaar worden gedeeld of van verschillende beveiligingsniveaus voor inzage, worden voorzien. Dit maakt de NORMAN-infrastructuur een aanbevelenswaardige keuze voor het opslaan en delen van screeningsdata, ook als het gevoelige water-monitoringsgegevens betreft.

---

<sup>10</sup> <https://www.bioconductor.org/packages/release/bioc/html/RMassBank.html>

## 5 Verkennen van upload van NTS-gegevens en analyse met DSFP

### 5.1 Introductie

Het DSFP<sup>11</sup> is opgezet door het NORMAN-netwerk, met als doel om data van LC-HRMS en gaschromatografie-hoge resolutie massaspectrometrie (GC-HRMS) op een eenduidige en uniforme manier op te slaan en doorzoekbaar te maken. Het platform heeft het potentieel om hét archief van stoffen te worden. DSFP heeft koppelingen met andere NORMAN-databanken zoals 'MassBank', de 'Suspect list exchange' en de 'Ecotoxicology studies' (zie Tabel 2).

In het DSFP data upload en analyse workflow kunnen onbewerkte LC-HRMS data van verschillende instrumenten (leverancier-onafhankelijk) worden omgezet in mzML. Daarna worden MS1 (full scan) en MS2 (fragmentatie spectra) spectra en chromatografische informatie over duizenden pieken van elk monster geëxtraheerd. De 'digitaal bevroren' monsters kunnen achteraf worden gescreend op de aanwezigheid van vrijwel elke verbinding die geanalyseerd kan worden met LC-MS met behulp van een combinatie van informatie over (i) de exacte massa, (ii) het voorspelde retentietijdvenster in het chromatogram, (iii) de isotoop fit en (iv) de fragmentionen.

Om deze functionaliteit te testen is er gebruik gemaakt van data afkomstig uit een ringonderzoek dat uitgevoerd is door het chemisch laboratorium van KWR.

### 5.2 Werkwijze

In de handleiding<sup>12</sup> staat beschreven welke stappen er worden geadviseerd om data te uploaden naar het DSFP. Zonder lidmaatschap van het NORMAN netwerk is het mogelijk om op de demoversie in te loggen (gebruikersnaam 'Demo' en wachtwoord 'Demo'). Vervolgens kan via de Help-functie de handleiding (PDF) worden bekeken. Om daadwerkelijk een archief te starten, is lidmaatschap van NORMAN vereist.

#### 5.2.1 Inlog en contactgegevens instituut

Als eerste zal er door de beheerder van het DSFP een gebruikers account moeten worden aangemaakt. Hiervoor is er een formulier beschikbaar in Excel waarin de basisgegevens van een instituut en informatie over de gebruikte analytische apparatuur kan worden aangegeven.

---

<sup>11</sup> <https://norman-data.net/>

<sup>12</sup> Handleiding DSFP beschikbaar na inloggen op: <https://norman-data.net/Demo/>

**NORMAN Digital Sample Freezing Platform (Log-in Page)**

Show  entries Search:

**You have access to the following DSFP repositories:**

- [DSFP Demo](#)
- [DSFP KWR \(Netherlands\)](#)

Showing 1 to 2 of 2 entries Previous  Next

Figuur 5: Inlogscherm van het NORMAN DSFP. Keuze repository (Archief)

### 5.2.2 Interne standaarden

Om gebruik te maken van de retentietijdindex binnen DSFP is het mogelijk om een kalibratiemix aan te vragen bij de beheerders van DSFP en is het raadzaam deze altijd met een meetserie te analyseren. Deze mix bestaat uit 18 stoffen die positief en 18 die negatief geïoniseerd kunnen worden.

Main Page Batch Mode **Contribute** More tools Help

Basic-Information Sample-MetaData Instrumental-MetaData **RTI\_Calibration** Spiked-Compounds

Upload-Files

I have not injected these calibrants OR I am using GC-HRMS

You have to input the RT of the calibrants. Let the cells empty for compounds without RT.

	Name	RT	CAS	Formula	Ion
1	Guanykurea	1.500	141-83-3	C2H6N4O	103.0614
2	Amitrol	1.750	61-82-5	C2H4N4	85.0509
3	Histamine	1.440	51-45-6	C5H9N3	112.0869
4	Chlormequat		999-81-5	C5H13CN	122.0731
5	Methamidophos	2.530	10265-92-6	C2H8NO2PS	142.0006
6	Vancomycin	6.010	1404-90-6	C66H75O2N9O24	724.7224
7	Cefoperazone	10.030	62893-19-0	C25H27N9O8S2	646.1497
8	Trichlorfon (Dylox)	9.010	52-68-6	C4H8Cl3O4P	256.9299
9	Butocarbomix	11.470	34681-10-2	C7H14N2O2S	191.0849
10	Dichlorvos	13.160	62-73-7	C4H7Cl2O4P	220.9532
11	Tylosin	12.860	1401-69-0	C46H77NO17	916.5264
12	TCMTB	10.090	21564-17-0	C9H6N2S3	238.9766
13	Rifaximin	17.040	80621-81-4	C43H51N3O11	786.3596
14	Spinosaad A (Spinosaad A)	17.090	131929-60-7	C41H65NO10	732.4681
15	Emamectin B1a	18.700	121124-29-6	C49H75NO13	886.5311
16	Avermectin B1a (Abamectin)		71751-41-2	C48H72O14	890.5260
17	Nigericin		28380-24-7	C40H68O11	725.4834
18	Ivermectin B1a		70288-86-7	C48H74O14	892.5436

You have to input the RT of the calibrants. Let the cells empty for compounds without RT.

	Name	RT	CAS	Formula	Ion
1	Amitrole		61-82-5	C2H4N4	83.0363
2	Benzoic acid	10.900	65-85-0	C7H6O2	121.0295

Figuur 6: Invulscherm retentietijden voor de index

### 5.2.3 Conversie naar mzML

Vervolgens moeten de ruwe datafiles zoals deze door de acquisitiesoftware van de fabrikant zijn opgenomen omgezet worden naar een mzML file. Dat kan door gebruik te maken van software van Proteowizard<sup>13</sup> (MSConvert). Daarmee worden met een ruwe databestand op basis van de MS1 data accurate massa's bepaald en de bijbehorende retentietijden. De combinatie van massa's en retentietijden worden "features" genoemd. Op die manier is er een eenduidig format beschikbaar waar DSFP mee kan werken. Door de beheerder is er een handleiding geschreven voor de volgende fabrikanten: Bruker, Agilent, Waters, Thermo Scientific en SCIEX.

### 5.2.4 Upload van gegevens

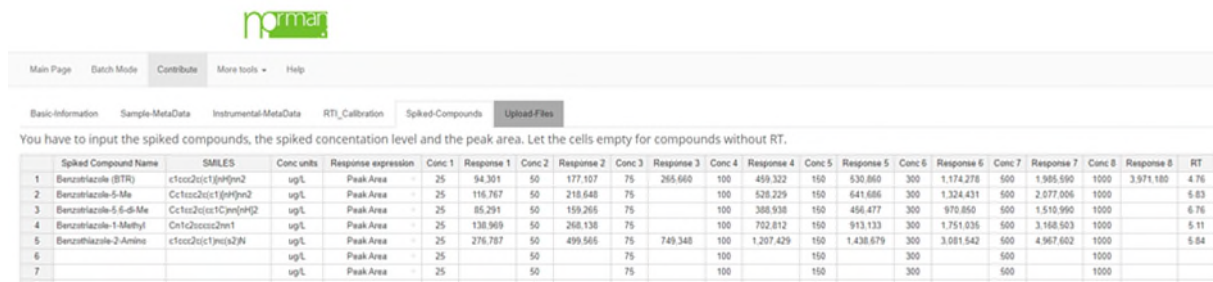
Voorafgaand aan het uploaden wordt er gevraagd naar meta-informatie met betrekking tot het monster: de locatie van bemonstering onder vermelding van GIS-coördinaten, de wijze van monsternamen, en specifieke informatie over

<sup>13</sup> <http://proteowizard.sourceforge.net/>

de matrix. Een deel van de metadata is verplicht, de mate van detail vergroot de waarde voor hergebruik. Er kan een eventuele opmerking per serie aan worden meegegeven als referentie voor de eigen bijdrage om eenvoudig weer een specifiek monster of serie op te zoeken.

Vervolgens moeten van het gemeten kalibratiemengsel de actuele retentietijden op worden gegeven zodat de retentietijdindex van alle *features* berekend kan worden.

Tevens is er de mogelijkheid om een semi-kwantitatieve berekening uit te laten voeren op de features door een zelf samengestelde kalibratiecurve te meten. Vervolgens kunnen hier de resultaten worden ingevoerd per standaard.

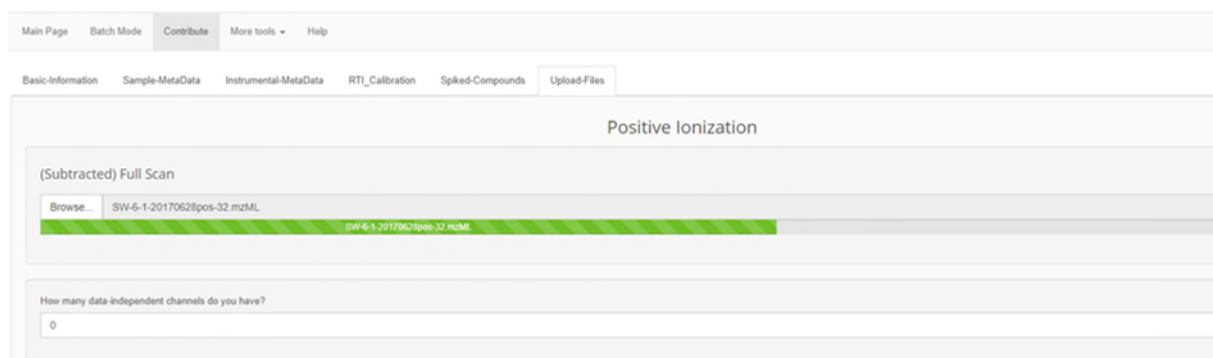


The screenshot shows the 'normal' software interface with a table for calibration. The table has columns for Spiked Compound Name, SMILES, Conc units, Response expression, and eight concentration-response pairs (Conc 1-8, Response 1-8), plus an RT column. The data is as follows:

Spiked Compound Name	SMILES	Conc units	Response expression	Conc 1	Response 1	Conc 2	Response 2	Conc 3	Response 3	Conc 4	Response 4	Conc 5	Response 5	Conc 6	Response 6	Conc 7	Response 7	Conc 8	Response 8	RT
1 Benzotriazole (BTR)	c1cc2c(c1[nH]n2	ug/L	Peak Area	25	54.301	50	177.107	75	265.660	100	459.322	150	530.860	300	1.174.278	500	1.985.590	1000	3.971.180	4.76
2 Benzotriazole-5-Me	Cc1cc2c(c1[nH]n2	ug/L	Peak Area	25	116.767	50	218.648	75		100	528.229	150	641.686	300	1.324.431	500	2.077.006	1000		5.83
3 Benzotriazole-5,6-di-Me	Cc1cc2c(c1C)c[nH]2	ug/L	Peak Area	25	85.291	50	159.265	75		100	388.938	150	456.477	300	970.860	500	1.510.990	1000		6.76
4 Benzotriazole-1-Methyl	Cn1c2ccsc2n1	ug/L	Peak Area	25	138.969	50	268.138	75		100	702.812	150	913.133	300	1.751.035	500	3.168.503	1000		5.11
5 Benzotriazole-2-Amino	c1cc2c(c1)n[nH]2N	ug/L	Peak Area	25	276.787	50	499.565	75	749.348	100	1.207.429	150	1.438.679	300	3.081.542	500	4.967.602	1000		5.84
6		ug/L	Peak Area	25		50		75		100		150		300		500		1000		
7		ug/L	Peak Area	25		50		75		100		150		300		500		1000		

Figuur 7: Invulscherm voor een kalibratiemix

Vervolgens kan per keer één mzML-bestand worden geüpload.



The screenshot shows the 'normal' software interface with the 'Upload-Files' tab selected. The title is 'Positive Ionization'. Below it, there is a section for '(Subtracted) Full Scan' with a 'Browse' button and a file path: 'SW-6-1-20170628pos-32.mzML'. A green progress bar is visible below the file path. At the bottom, there is a question 'How many data-independent channels do you have?' with a text input field containing the number '0'.

Figuur 8: Upload van een mzML file

### 5.2.5 Suspect-screening

Op basis van een naam, CAS-registratienummer of InChI-naam kan er gezocht worden in een of meerdere monsters of een suspect aanwezig is. Verder kan bijvoorbeeld op de accurate massa van een precursor of brutoformule worden gezocht. Dit maakt het mogelijk ook naar stoffen te screenen die nog niet geïdentificeerd zijn bij een gebruiker, maar wel in de databank zitten en misschien ook al geïdentificeerd zijn door andere gebruikers. Verder kan aangegeven worden binnen welke massa-afwijking gezocht kan worden en welke range in de berekende retentietijdindex er gezocht moet worden. Door vervolgens aan te geven in welke monsters gezocht moet worden zal er op basis van de ingestelde criteria een resultaat verschijnen.

In het onderstaande voorbeeld is gekozen voor propazine. Wat opvalt is dat de berekende retentietijdindex (RTI: 614) behoorlijk afwijkt van de werkelijke RTI (867). Met een verhoging van de tolerantie-drempelwaarde voor RTI komt de juiste feature voor propazine in beeld.



Contributed Samples Results Chromatograms Interactive Map

Copy Print Download View Columns Show 100 entries Search:

Select per page

	Retention time [min]	Mass of ion [m/z]	Intensity	MS/MS available	Proposed substance	Level of confirmation of identification	RTI	Qualifier Ions [mz/RT (min)/Intensity]
All	All	All	All	All	All	All	All	All
<input type="checkbox"/> KWR_LC-ESI-QFT_River water intakepoint Lek channel_Nieuwegein_Netherlands_21.06.2017_Norman Interlab_N062817	16.4	230.1166	10412969	No	-	Exact mass Plausible RT	867	188.0704/1108.762/8.495 152.1180/3935.22/8.823 146.0225/4303.194/13.138 79.00576/858.5617/22.367
<input type="checkbox"/> KWR_LC-ESI-QFT_River water intakepoint Lek channel_Nieuwegein_Netherlands_21.06.2017_Norman Interlab_N062817	16.87	230.1162	78653	No	-	Exact mass Plausible RT	894.1	188.0704/1108.762/8.495 152.1180/3935.22/8.823 146.0225/4303.194/13.138 79.00576/858.5617/22.367
<input type="checkbox"/> KWR_LC-ESI-QFT_River water intakepoint Lek channel_Nieuwegein_Netherlands_21.06.2017_Norman Interlab_N062817	12.73	230.1171	7933	No	-	Exact mass Plausible RT	654.2	188.0704/1108.762/8.495 152.1180/3935.22/8.823 146.0225/4303.194/13.138 79.00576/858.5617/22.367

Figuur 9: Resultaat zoeken op een suspect in de datafile

### 5.2.6 Suspectlijsten

Door gebruik te maken van de batch mode is het mogelijk om uit diverse suspect-lijsten te kiezen uit de SLE<sup>14</sup> (Suspect List Exchange) van Norman. Zie in het onderstaande voorbeeld de match van het monster met S05 (Sjerps et al. suspect list)

	A	B	C	D	E	F	G
1	Compounds	Formula	CAS	SMILES	InChi	InChiKey	Identification_Proofs
2	DEET	C12H17NC	CAS_RN: 1	c1(C(N(CC	InChi=1S/i	MMOXZB	Mass accuracy, plausible RT
3	N-Acetylaminoantipyri	C13H15N3	CAS_RN: 5	c1(=O)NC	InChi=1S/i	OIAGWXK	Mass accuracy, plausible RT
4	Terbutylazine	C9H16ClN	CAS_RN: 5	c1(nc(nc(r	InChi=1S/i	FZXISNSW	Mass accuracy, plausible RT
5	Caffeine	C8H10N4C	CAS_RN: 5	c12c(n(c(=	InChi=1S/i	RYVVLZVU	Mass accuracy, plausible RT
6	Phenazone	C11H12N2	CAS_RN: 6	CN1N(C(=	InChi=1S/i	VEQOALN	Mass accuracy, plausible RT
7	Carbamazepine epoxi	C15H12N2	CAS_RN: 3	NC(=O)N2	InChi=1S/i	ZRWWEEV	Mass accuracy, plausible RT
8	Tramadol	C16H25NC	CAS_RN: 2	COC1=CC=	InChi=1S/i	TVYLLZQT	Mass accuracy, plausible RT
9	Valsartan	C24H29N5	CAS_RN: 1	CCCCC(=O	InChi=1S/i	ACWBQPN	Mass accuracy, plausible RT
10	Irbesartan	C25H28N6	CAS_RN: 1	C12(C(N(C	InChi=1S/i	YOSHYTLC	Mass accuracy, plausible RT
11	Oxprenolol	C15H23NC	CAS_RN: 6	O(c1cccc:	InChi=1S/i	CEMAWM	Mass accuracy, plausible RT
12	N,N'-Diethylcarbanilic	C17H20N2	CAS_RN: 6	CCN(C(=O	InChi=1S/i	PZIMIYVO	Mass accuracy, plausible RT

Figuur 10: Match van features in het monster met suspectlijst 5 (<https://doi.org/10.5281/zenodo.2621941>) uit de SLE

### 5.2.7 Conclusie

Het principe om een gestandaardiseerde procedure te maken voor het detecteren van features in ruwe databestanden is vanuit het oogpunt van efficiëntie en effectiviteit zeer waardevol. Merk op dat het DSFP ten tijde van deze rapportage nog in ontwikkeling is. Zo kwamen er uit de test stapsgewijs enkele fouten naar voren, die echter adequaat en op korte termijn waren opgelost in overleg met de ontwikkelaar van de database (Nikiforos Alygizakis). Het verdient daarom aanbeveling met enkele testbestanden de upload-procedure te doorlopen.

<sup>14</sup> <https://www.norman-network.com/nds/SLE/>

## 6 Conclusies, aanbevelingen en perspectief

### 6.1 Conclusies en aanbevelingen

We komen tot de volgende conclusies:

- Uit de verkenning blijkt dat het gebruiksscenario waarbij informatie over doelstoffen en suspects wordt gezocht, het meest relevant is om nu al gegevens onderling te delen en te ontsluiten. Het belang van deze gegevensverzameling is (a) effectievere prioritering van (mogelijk) zorgwekkende stoffen in meetcampagnes, (b) snellere detectie of identificatie van deze stoffen, (c) een beter landelijk beeld van het voorkomen van deze stoffen en (d) efficiënter uitwisselen van referentiestandaarden en gebruikte analysemethoden.
- Het verzamelen van doelstofgegevens blijkt voor enkele laboratoria een tijdrovende en arbeidsintensieve klus omdat de gebruikte systemen niet gericht zijn op het filteren van de gevraagde informatie.
- Bij het aggregeren van de doelstofgegevens naar één lijst blijkt dat voor sommige stoffen de identificatie van de stof niet eenduidig is of niet achterhaald kan worden. De reden is dat enkele laboratoria internationale standaarden (zoals InChI, SMILES, IUPAC voorkeursnaam) en de nationale SIKB-standaard voor stofnaamgeving niet structureel verzamelen of registreren.
- Het DSFP biedt voor de Nederlandse drinkwaterbedrijven een kans om diverse screeningsgegevens op te zoeken én (internationaal) te ontsluiten omdat DSFP als doel heeft om screeningsdata (LC-HRMS en GC-HRMS) op een eenduidige en uniforme manier op te slaan en doorzoekbaar te maken. Het platform bevat naast zoekfunctionaliteit een gestandaardiseerde procedure voor het detecteren van features in ruwe data zonder daarbij afhankelijk te zijn van software die door fabrikanten geleverd wordt bij de laboratoriumanalyse-apparatuur. Het DSFP biedt ook de mogelijkheid om naar onbekende verbindingen te zoeken en deze te delen. Het uploaden van gegevens naar DSFP vergt, zeker de eerste keer, echter een groot aantal handelingen en sluit mogelijk niet aan bij de werkhandelingen van laboratorium-analisten.

Op basis van resultaten in dit onderzoek komen we tot de volgende aanbevelingen:

- (1) het toevoegen van doelstofgegevens aan de bestaande spreadsheet, te evalueren. Het bijwerken van een spreadsheet kan, zonder verdere validatie, leiden tot de introductie van dubbelingen, (menselijke) fouten bij de invoer van CAS-nummers of andere (meta-)gegevens. Het is raadzaam om bij frequent gebruik van het spreadsheet invoer van gegevens te automatiseren via bijvoorbeeld een webformulier en daaraan gekoppelde databank (zie ook 6.2 Perspectief). De gevalideerde doelstoffenlijst zou op regelmatige basis op de server van de NORMAN chemical substance search database (SusDat) moeten worden neergezet. Een mogelijkheid is om deze taak door de BOK (Beheerscommissie Overdracht Kwaliteitsgegevens) te laten uit te voeren. Dit heeft als voordeel dat stoffen meteen een SIKB- en REWAB-code meekrijgen;
- (2) naast het CAS-nummer ook andere internationaal erkende identificatienamen, zoals IUPAC-naam en InChI, van een stof op te slaan en te delen;
- (3) te starten met het verzamelen van suspect-gegevens, zodat ook met deze gegevens effectievere prioritering, voorkomen van suspects en verdere identificatie wordt bevorderd;

- (4) een stuur- en adviesgroep op te richten die enerzijds het stroomlijnen van de stoffeninformatie tussen de drinkwaterlaboratoria bevordert en anderzijds de transitie naar het gebruik van de databanken van het NORMAN-netwerk, het DSFP en Massbank in het bijzonder, in gang zet en de ervaringen daarvan deelt. Ook deze taak zou door de BOK kunnen worden opgepakt, omdat voor de REWAB-databank 'nieuwe' parameters ook via de BOK aangemeld en geregistreerd worden.
- (5) in samenwerking met de ontwikkelaars het ontsluiten van data naar DSFP en MassBank verkennen, evalueren en dusdanig inrichten of automatiseren zodat het aansluit bij de gebruikte software voor screening en identificatie van stoffen in de Nederlandse situatie.

## 6.2 Perspectief

Met hoge resolutie-massaspectrometrie (LC-HRMS en GC-HRMS) worden doelstofanalyses en suspect en NTS uitgevoerd ten behoeve van analyse en identificatie van organische microverontreinigingen in watermonsters. Door de gegevens van deze metingen te registreren wordt voorkomen dat een arbeidsintensieve identificatie van een stof opnieuw moet worden uitgevoerd. Dit overstijgt de eigen laboratoriumbelangen: ook andere organisaties in de watersector zijn gediend met effectievere prioritering van stoffen, snellere detectie of identificatie en een beter landelijk beeld van het voorkomen van (mogelijk) zorgwekkende stoffen.

### Stip op de horizon

Voor de Nederlandse drinkwaterlaboratoria bieden internationale databanken voor MS<sup>2</sup>-spectra, gegevens over suspects en volledig geïdentificeerde stoffen de mogelijkheid om informatie over de gemeten en geanalyseerde stoffen te ontsluiten. Databanken als het NORMAN DSFP, EMPODAT en MassBank bieden uitgebreide mogelijkheden hiertoe. Het uploaden van de gegevens vergt op dit moment echter, zeker voor de eerste keer en ook omdat DSFP nog in ontwikkeling is, een groot aantal handelingen. Het voordeel van het gebruik van deze databanken is evident: doorontwikkeling en onderhoud van deze databanken is vanuit NORMAN geregisseerd, informatie is breed toegankelijk en (inter)nationaal beschikbaar. De keuze van NTS data-analysesoftware beïnvloedt de screeningsresultaten. Daarom is het wenselijk het gebruik van data en methodiek te standaardiseren, bijvoorbeeld door het gebruik van het open-source softwarepakket patRoon<sup>15</sup>. Daardoor zouden NTS resultaten ook beter uitwisselbaar worden.

### Keuze voor de korte termijn

Op de korte termijn staan de drinkwaterlaboratoria voor de volgende keuze:

- (1) doorgaan met de huidige werkwijze, dat wil zeggen, het niet (meer op structurele basis) delen van doelstof- en suspect-screeningsgegevens. Deze werkwijze wordt vanuit oogpunt van effectiviteit en kennisdeling afgeraden.
- (2) toewerken naar een situatie waarin zoveel mogelijk gebruik wordt gemaakt en geleerd van de mogelijkheden om gegevens op te sturen naar DSFP en andere (NORMAN) databanken, zoals EMPODAT. Een dergelijk besluit vergt inzicht van de noodzaak hiervan bij de laboratoria en drinkwaterbedrijven en vraagt om tijd, middelen en opleiding. Dit proces kan worden geïnitieerd en begeleid door een speciaal voor deze taak opgerichte stuur- en adviesgroep en zou in een project gefaciliteerd kunnen worden door KWR. Ook kan het opsturen van data naar DSFP, EMPODAT en MassBank verder geautomatiseerd en vereenvoudigd worden (zie punt 5 in paragraaf 6.1).
- (3) toewerken naar een situatie waarbij DSFP en andere NORMAN-databanken in de standaard analyse- en screeningsprocedure van de laboratoria worden opgenomen, én doorgaan met het toevoegen van

<sup>15</sup> <https://doi.org/10.21203/rs.3.rs-36675/v1>

(nieuwe) stoffen in de doelstoffenlijst. Afhankelijk van de gebruikservaring van het werken met een spreadsheet kan worden besloten om de spreadsheet te vertalen naar een webapplicatie met webformulier en zoekmachine. Voor een dergelijke web-databank hebben onderzoekers en ontwikkelaars bij KWR inmiddels ervaring opgedaan met Gondwana (optimalisatie-platform voor distributienetwerken) en AquaPriori (applicatie om zuiveringsefficiëntie voor groot aantal stoffen te bepalen o.b.v. QSARs). De databank van AquaPriori kan eenvoudig worden voorzien van een aparte, beveiligde website met webformulier en zoekmachine, als ook uitgebreid met meta-gegevens zoals het voorkomen van de stof, rapportagegrens en analysemethode. In Hoofdstuk 2, (Figuur 2), is een dergelijke workflow gepresenteerd. Het transitieproces kan door een speciaal voor deze taak opgerichte stuur- en adviesgroep begeleid worden en zou in een project gefaciliteerd kunnen worden door KWR.

### Risico-afweging en kosteninschatting

Voor de keuze die hierboven geschetst staat toont Tabel 9 een analyse van het kwalitatieve risico, kosten en baten van de verschillende opties.

Tabel 9: kwalitatieve risico-, kosten en batenanalyse bij het delen van doelstof- en suspect-stofgegevens.

Voorgestelde werkwijze	Risico's	Kosten	Baten/belang
<b>Doorgaan met de huidige werkwijze</b>	Geen overzicht van suspects, missende of ontbrekende informatie met betrekking tot stofidentificatie.	Niet direct meetbaar. Geschat wordt dat enkele duizenden euro's (enkele werkdagen) extra onderzoekstijd per jaar nodig is om de gevraagde informatie boven tafel te krijgen.	Geen directe (meetbare) kosten.
<b>Bekend raken met DSFP en MassBank en/of Mass-t.</b>	Kennisvragen die nu al van belang zijn (bijvoorbeeld: is de stof eerder gemeten door een ander lab?) kunnen nog niet snel en efficiënt worden opgelost.	<ul style="list-style-type: none"> <li>Bekend raken met databanken en het op een server plaatsen van data: enkele dagen (per analist en KWR-onderzoekers). Voorverkenning nodig, en delen van ervaringen en evaluaties (zie adviesgroep).</li> <li>Adviesgroep: één tot enkele dagen per jaar per vertegenwoordiger van een drinkwaterlaboratorium en KWR-onderzoeker(s).</li> <li>Begeleidingsuren, uren voor documentatie en verslaglegging (KWR).</li> </ul>	Middellange termijn: volledige beschikking over doelstof, suspect en NTS-gegevens die van belang zijn voor de Nederlandse situatie. Kennis over welke onbekende stoffen, waar zijn gemeten.
<b>KWR-databank voor referentiestandaarden, doelstoffen en suspects én</b>	<ul style="list-style-type: none"> <li>Ontwikkeling en onderhoud/ondersteuning van de webapplicatie is</li> </ul>	<ul style="list-style-type: none"> <li>Adviesgroep: enkele dagen per jaar per vertegenwoordiger van een</li> </ul>	<ul style="list-style-type: none"> <li>Korte termijn: toegang tot informatie over doelstoffen en suspects.</li> </ul>

<b><i>transitie naar DSFP en MassBank.</i></b>	duurder dan eerder ingeschat. <ul style="list-style-type: none"><li>• Gebruik van de webapplicatie sluit onvoldoende aan bij de procedures van analisten.</li></ul>	drinkwaterlaboratorium en KWR-onderzoeker(s). <ul style="list-style-type: none"><li>• KWR: begeleiding, vertalen van spreadsheet naar webapplicatie, export naar DSFP-/MassBank-<i>ready</i> input en documentatie</li></ul>	<ul style="list-style-type: none"><li>• Middellange termijn: volledige beschikking over doelstof, suspect en NTS-gegevens die van belang zijn voor de Nederlandse situatie. Kennis over welke onbekende stoffen, waar zijn gemeten.</li></ul>
------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

# I Bijlage Vragenlijst

## Antwoorden op vragenlijst over de informatiebehoefte van drinkwaterlaboratoria (incl. KWR)

- 1 Voor de waterbedrijven maakt u gebruik van verschillende databronnen om informatie over een stof te zoeken. Welke informatiebronnen zijn dat?

Antwoorden	Scores
Chemspider	86%
Risico's van stoffen (RIVM)	71%
NORMAN substance database	57%
Wikipedia	57%
Chemical Classification and Information Database (EPA)	29%
ECHA REACH database ( <a href="https://echa.europa.eu/information-on-chemicals/registered-substances">https://echa.europa.eu/information-on-chemicals/registered-substances</a> )	29%
Massbank	14%
WebBook (NIST)	14%
Toelatingen bestrijdingsmiddelen (CTBG)	14%
ToxNet	14%
Pubchem	14%
Eigen suspect databanks	14%

- 2 Voor de waterbedrijven zoekt u met name informatie over (beoordeel wat u relevant vindt):

	Relevant (waarvan zeer relevant)
Zuiveringsefficiëntie van een niet eerder gemeten stof, chemische naam onbekend (accurate massa bekend).	43% (14%)
Zuiveringsefficiëntie van een stof, waarvan de chemische naam bekend is.	71% (43%)
Signaleringslocatie een stof (naam is onbekend)	43% (29%)
'Informatieblad' van een stof, met bijv. vermeld: stofeigenschappen, stofcategorie, zuiveringsefficiëntie, gezondheidsrisico's	57% (43%)
Signaleringslocatie van een stof (naam is bekend)	71% (14%)
Toxicologische informatie van een stof.	100% (71%)
Informatie over normen en richtlijnen voor een stof	100% (71%)
Informatie van die stof uit BTO-rapporten.	100% (57%)

- 3 Het liefst zoekt u allerlei gegevens over een gezochte stof via een doorverwijspagina.  
**Nee: 34%, 66%: interessant (waarvan 80% mogelijk interessant).**
- 4 Het liefst heeft u de presentatie van gegevens over een gezochte stof zo compleet mogelijk van één aspect (bijvoorbeeld voorkomen van een stof), inclusief bron- en meta-informatie.  
**Ja, waarvan 71% mogelijk interessant.**
- 5 Hoe vaak zoekt u, gemiddeld gezien, informatie over specifieke stoffen?  
**Enkele keren per maand: 86%, enkele keren per (werk)dag: 14%.**
- 6 Als data overzichtelijker en gemakkelijker te raadplegen zou zijn, zou u dan vaker gebruik maken van een stoffenkennisbank?  
**Ja: 71%, Nee/waarschijnlijk niet: 29%. Toelichting 'waarschijnlijk niet': Zeer waarschijnlijk staan er veel**

**minder stoffen in [een databank met gevalideerde metadata], dan een grote databank [als bijvoorbeeld PubChem]. Deze zullen daarom als eerste worden geraadpleegd.**