



Kaders voor en kansen van open data

Samenvatting

Uitgesproken voorstanders van open data, welke (soms onder voorwaarden) vrij gebruikt, gewijzigd en gedeeld kunnen worden door iedereen voor ieder doel, zijn te vinden in de academische wereld, maar ook bij nationale en Europese overheden. De EU maakt het verplicht om steeds meer data te delen. Vanuit het onderzoek worden de voordelen van het delen van data steeds duidelijker. Ten slotte beloven ontwikkelingen in de kunstmatige intelligentie op korte termijn ongeëvenaarde mogelijkheden voor het vinden en analyseren van data, mits deze vindbaar, toegankelijk, uitwisselbaar en herbruikbaar, kortweg FAIR, zijn. Dit samenspel van ontwikkelingen leidt tot een kanteling van het speelveld ten faveure van het delen en exploiteren van veel meer data. Vanuit het oogpunt van wederkerigheid is het wenselijk dat waterorganisaties in gelijke mate hun data delen, waarbij een adagium als “open, tenzij” een goed uitgangspunt is. De publieke drinkwaterbedrijven kunnen hierin een trekkersrol spelen, aangezien zij niet beperkt worden door private bedrijfsmodellen.

Consequenties voor u

	Laag	Middel	Hoog	Beknopte uitleg
Impact				Open/FAIR data biedt veel kansen.
Zekerheid				Ontwikkeling reeds gaande





Trendbeschrijving en achtergrond

Inleiding

De term open data en de achterliggende ideeën zijn al bijna dertig jaar oud (Chignard, 2013). Met deze term worden data bedoeld welke vrij gebruikt, gewijzigd en gedeeld door iedereen voor elk doel ([Open Knowledge Foundation, 2023](#)). Desondanks blijft toegang tot relevante datasets voor bijvoorbeeld onderzoeksdoeleinden een uitdaging. Voortschrijdend inzicht (o.a. DWSI, 2011, 2014, 2017) laat bovendien zien dat data idealiter niet alleen open zijn, maar ook FAIR. FAIR data zijn data die *vindbaar, toegankelijk, uitwisselbaar* en *herbruikbaar* (*findable, accessible, interoperable, reusable*) zijn, met als belangrijke toevoeging dat deze ook voor machines/algoritmes toegankelijk moet zijn ([GO FAIR, 2023](#)).

De Europese Commissie is een uitgesproken voorstander van het delen van data, met name met betrekking tot de leefomgeving (EU, 2023). Ook de Nederlandse overheid geeft aan hier voorstander van te zijn in haar Visie Open Overheid (Ministerie BZK, 2013) en in haar Data Agenda Overheid (Overheidsbreed Beleidsoverleg Digitale Overheid, 2020) vermeldt zij te “streven naar het verder delen van publieke data, waar mogelijk in de vorm van open data”. Vlaanderen toont vergelijkbare ontwikkelingen. De Vlaamse regering keurde in 2011 een

conceptnota over open data goed en in 2020 gebeurde hetzelfde met het *actieplan open data 2020-2024* door het Stuurorgaan Vlaams Informatie- en ICT-beleid (Digitaal Vlaanderen, 2023). Desalniettemin heeft de Europese Commissie onlangs besloten om onder andere Nederland en België voor het Hof van Justitie van de Europese Unie te dagen omdat zij de EU-regels inzake open gegevens en het hergebruik van overheidsgegevens (Richtlijn EU 2019/1024, de zogenaamde richtlijn open gegevens) niet in nationaal recht hebben omgezet (EU, 2023b).

In de bovengenoemde gevallen gaat het bovenal om overheids- en omgevingsdata die de maatschappij en de economie tot baat kunnen zijn. Er zijn ook ontwikkelingen die specifiek op onderzoek zijn gericht. De Koninklijke Nederlandse Akademie van Wetenschappen deed in 2018 aanbevelingen om het hergebruik van publieke data door onderzoekers te bevorderen (KNAW, 2018). Het gaat hierbij om diverse data die ook voor de watersector van belang zijn, o.a. m.b.t. de omgeving (bijv. klimaat), natuur (bijv. ecologische processen), emissies (bijv. uitstoot verontreinigende stoffen), etc. Het *Ambitiedocument open wetenschap* (NPOS, Nationaal Programma Open Science, 2022) beschrijft leidende principes en strategische doelen voor het realiseren van open

wetenschap, met als doel om wetenschappelijke onderzoekspraktijken meer open collaboratief te maken, voor een grotere wetenschappelijke en maatschappelijke impact. Het NPOS wordt geleid en gefinancierd door NWO. In dit ambitiedocument neemt data, als een van de kennisproducten, een belangrijke rol in. Als strategisch doel wordt expliciet vermeld om in 2030 data en andere producten van kenniscreatie zoveel mogelijk open en FAIR te laten zijn.

Met de hierboven beschreven initiatieven nemen de betreffende overheden de belangrijke rol van initiatiefnemer en/of stimulators op zich die het bedrijfsleven en onderwijs in beweging brengen.

Trendalert

Het doel van dit trendalert is om te signaleren dat een samenspel van ontwikkelingen het speelveld aan het kantelen is ten faveure van het delen van meer data. In het eerste deel van dit trendalert wordt stilgestaan bij de argumenten om wel of niet *vrijwillig* data te delen. Ook worden de nieuwe wettelijke kaders die vanuit EU-verband voor *verplichte* deling van data ontstaan kort beschreven. Vervolgens wordt in het tweede deel, dat een interview met een expert op het gebied van en uitgesproken voorstander van (bewuste) open data behelst, dieper ingegaan op het hoe en waarom van FAIR



en open data. Het trendalert eindigt met een beschouwing van de belangen en betekenis voor de watersector.

Datadeling

De watersector is traditioneel gezien enigszins terughoudend of in ieder geval behoedzaam met het delen van data, hoewel waterkwaliteitsdata van iedere winning door waterbedrijven in Nederland en België online beschikbaar worden gesteld. Toch zijn er diverse andere voorbeelden van watergerelateerde open data te noemen. Rijkswaterstaat biedt veel (o.a. watergerelateerde) data aan op basis van het principe “open, tenzij...” (Rijkswaterstaat, 2023, zie ook Rijkswaterstaat, 2022).

Verder deelt het Hoogheemraadschap de Stichtse Rijnlanden diverse datasets haar openbare dataportaal Infovijver (<https://data-hdsr.opendata.arcgis.com/>). Een ander voorbeeld is het EU-project WaterVerse (<https://waterverse.eu/>), waarin PWN meedenkt over en meebouwt aan een Europese dataspace voor waterdata. Het Waterkwaliteitsportaal biedt Nederlandse waterkwaliteitsdata voor oppervlakte- en grondwater aan (<https://www.waterkwaliteitsportaal.nl/>). De digitale delta (<https://digitaledeltaorg.github.io/>) is een ander voorbeeld. De World Bank stelt wereldwijde waterdata beschikbaar (<https://wbwaterdata.org/>), en NASA o.a. actuele grondwaterdroogtegegevens op basis van

satellietdata (<https://nasagrace.unl.edu/>). Andere, diverse repositories voor waterdata, waarin datasets ter beschikking worden gesteld, zijn te vinden via Re3data.org (<https://www.re3data.org/>). Meer voorbeelden zijn opgenomen aan het eind van dit trendalert.

(Vermeende) voor- en nadelen van data delen

Een uitgebreid overzicht van de vermeende voor- en nadelen van open data wordt gegeven o.a. door Wikipedia ([Wikipedia – Open data, 2023](#)); hier volgt een kort overzicht. Sommige argumenten zijn meer normatief van aard, zoals de gedachte dat op feiten geen auteursrecht zou kunnen rusten en bepaalde gegevens als “eigendom van de mensheid” zouden kunnen worden beschouwd (zoals bijvoorbeeld milieugegevens of genomen van organismen), of iets dichter bij de grond, dat gegevens die op kosten van de gemeenschap zijn verzameld ook beschikbaar zouden moeten zijn voor de gemeenschap. Andere zijn meer gericht op de praktische voordelen. Hieronder vallen o.a. de gedachte dat vrije beschikbaarheid van gegevens die onderliggend zijn aan een onderzoek de waarde van de onderzoeksresultaten vergroot en zorgt dat deze data behouden blijven, dat data een belangrijke factor voor sociaaleconomische ontwikkeling kunnen zijn, en dat in wetenschappelijk onderzoek de ontwikkeling van kennis versneld kan worden door betere toegang tot gegevens

(Pronk et al., 2015; Pronk, 2019). Tevens kan open data zorgen voor transparantie m.b.t. overheidsbesluiten, verbeterde besluitvorming door de overheid op basis van meer gegevens, en bijdragen aan burgerbetrokkenheid.

Hier kunnen argumenten tegenover worden gesteld tegen het delen van data. Opnieuw zijn sommige normatief van aard, zoals het uitgangspunt dat er geen duplicatie of aanvechting van activiteiten van de publieke sector op basis van overheidsfinanciering zou mogen plaatsvinden – anders gezegd: als een commerciële partij geld kan verdienen met bepaalde data, zou de overheid deze niet gratis ter beschikking mogen (laten) stellen – of dat het een inefficiënt en daarom onwenselijke besteding van overheidsgeld zou zijn als deze tot gegevens zou leiden die slechts voor een kleine groep gebruikers waarde oplevert. Misinterpretatie door onkundige gebruikers wordt ook gezien als een risico. Verder wordt gewaarschuwd voor de exploitatie van open gegevens zonder dat de gemeenschappen waar deze op worden betrokken er zelf baat bij hebben – bijvoorbeeld het publiceren van onderzoek op basis van gegevens over ontwikkelingslanden door rijke Westerse onderzoeksinstituten. Ten slotte heeft niet iedereen gelijke toegang tot digitale middelen en technologieën,



wat kan leiden tot een digitale kloof en ongelijke mogelijkheden om open data te benutten.

Ook worden er praktische argumenten voor het vragen van een geldelijke bijdrage voor data geopperd, zoals het feit dat er kosten worden gemaakt voor het verzamelen, opschonen, beheren en verspreiden van gegevens, die ergens van moeten worden betaald, dat inkomsten uit het publiceren van gegevens non-profitorganisaties in staat stellen andere activiteiten te financieren, en dat het openstellen van data prikkels weg zou nemen om te investeren in de verwerking van deze gegevens. Het vragen van een vergoeding is overigens niet strijdig met de FAIR-principes. Onderzoekers op hun beurt zijn vaak huiverig om data openbaar te stellen omdat daarmee anderen een mogelijke concurrerende publicatie kunnen schrijven, terwijl zij zelf tijd hebben moeten investeren voor het verkrijgen van de data. Ten slotte bestaan er mogelijke privacyrisico's bij het openbaar maken van bepaalde data, kan bij ontbreken van voldoende metadata de beoordeling van kwaliteit en kan geschiktheid van bepaalde data voor een zeker doel moeilijk.

Verplichte datadeling

Twee recente EU-richtlijnen verplichten het delen van diverse soorten data.

Open data

De EU-richtlijn 2019/1024 (EU, 2019) stelt dat “een Uniebrede lijst van datasets met een bijzonder potentieel om sociaal-economische baten te genereren, gekoppeld aan geharmoniseerde voorwaarden voor hergebruik, een belangrijke randvoorwaarde is voor grensoverschrijdende datatoepassingen en -diensten.” Het doel van het beschikbaar stellen van deze datasets is het stimuleren van innovatie en het creëren van sociaal-economische voordelen door het hergebruik van gegevens. In het kader van deze richtlijn heeft de EU een lijst opgesteld van datasets die openbare lichamen in de EU vanaf mei 2024 verplicht om deze data “via API's beschikbaar [te stellen] in machineleesbare formaten die beantwoorden aan de redelijke behoeften van hergebruikers.” (EU, 2023c). De lijst van data die het betreft omvat zes categorieën, te weten 1) geospatiale data; 2) aardobservatie en milieu; 3) meteorologische data; 4) statistiek; 5) bedrijven en eigendom van bedrijven, en 6) mobiliteit.

Drinkwaterrichtlijn

De EU-drinkwaterrichtlijn van 2020 (EU, 2020) vereist dat passende, actuele informatie beschikbaar is over water dat voor menselijke consumptie bestemd is. Deze informatie omvat minimaal de recentste monitoringresultaten (maximaal 1 jaar oud) voor 2 microbiologische parameters, 34 chemische stoffen

en/of stofgroepen en 18 indicatorparameters. Deze monitoringresultaten dienen beschikbaar te zijn om het publiek te informeren over de kwaliteit van het drinkwater en om ervoor te zorgen dat het voldoet aan de vereiste normen en richtlijnen voor menselijke consumptie.

Licenties

Bij het delen van data kunnen verschillende licenties gehanteerd worden, bijvoorbeeld een Creative Commons (CC) licentie (<https://creativecommons.org/>). Deze licenties variëren van werkelijk open data (CCO) tot data die gebruikt kunnen worden onder bepaalde voorwaarden zoals verplicht citeren (CCBY, meest populair), alleen voor niet-commercieel gebruik (CCBYNC), gelijk delen van het derivaat vereist (CCBYSA), of geen derivaten toegestaan (CCBYNCND). Een uitgebreider overzicht wordt aan het eind van dit trendalert geboden.

Interview

In het tweede deel van dit Trendalert worden de staat en perspectieven van open data in het kader van wateronderzoek beschouwd. Dit gebeurt aan de hand van een interview met een expert.



Dr. Rhea Verbeke is bevoegen onderzoeker op het gebied van membraantechnologie aan de KU Leuven (België) en de Johannes Gutenberg Universität Mainz (Duitsland). Zij is mede-initiatiefnemer van de [Open Membrane Database](#), waarin experimentele data over membranen vrij gedeeld wordt, en voorvechtster van open data, daar waar dit nuttig is. Op uitnodiging van de redactie heeft zij een artikel geschreven in de eerste editie van *Nature Water* ([Verbeke, 2023](#)) over het belang van Open en FAIR Data en wat er nodig is om data FAIR en Open te maken in een academische context. In een tweegesprek reflecteert zij op de belangen en kansen voor de drinkwatersector, van buitenaf als academicus en van binnenuit als onderzoeker op het gebied van watertechnologie.

U bent uitgesproken over het belang van het delen van data in de academische context, en hebt daar een artikel over gepubliceerd dat bijna leest als een manifest. Kunt u daar wat over vertellen? Wat is het belang van open data?

“Open data stelt ons in staat om betere wetenschap te bedrijven. Door alle data gestructureerd samen te brengen, kunnen we bijvoorbeeld een dieper inzicht krijgen in membranen, zoals de relatie tussen synthese, structuur en prestaties.” Dit soort inzichten zijn niet mogelijk met kleine datasets. Daarnaast is het

problematisch dat veel onderzoek publiekelijk wordt gefinancierd maar niet publiek toegankelijk is. Deze onwenselijke ethiek wordt omzeild door gebruik te maken van de Open Membrane Database, waar alle informatie vrij beschikbaar is. Het opzetten van deze database heeft inzicht gegeven in de staat van open data in het veld van membraantechnologie en in het algemeen. “De reden waarom ik daarmee begonnen ben is eigenlijk puur uit frustratie. Ik had een membraan ontwikkeld tijdens mijn doctoraat en wilde het graag vergelijken met andere membranen in de literatuur. Het probleem was dat de benodigde informatie verspreid was over verschillende artikelen in verschillende tijdschriften, vaak onleesbaar was en zonder beschikbare ruwe data. Bovendien werden de data niet op dezelfde manier verkregen, waardoor een accurate vergelijking lastig was.”

Om dit inefficiënte proces aan te pakken, besloot Rhea om zelf een oplossing te creëren. De Open Membrane Database is zeer waardevol gebleken voor onderzoekers en bedrijven, die alle relevante data op één plek kunnen vinden. Dit is ook handig voor bedrijven, omdat commerciële membranen gegroepeerd zijn in plaats van dat ze afzonderlijke datasheets moeten raadplegen. Echter, hoewel de Open Membrane Database veel gebruikers heeft en veel wordt bezocht, draagt niemand bij aan het platform. Het oorspronkelijke idee was dat

wanneer onderzoekers artikelen publiceren en hun data voldoen aan de vereisten van de database, ze deze zouden indienen op het platform. Zo zou de database continu aangevuld worden met nieuwe gegevens en dus up-to-date zijn. Een van de belangrijke leerpunten met betrekking tot de status van open data bleek echter dat iedereen graag toegang wil krijgen tot de gegevens, maar dat er weinig mensen bereid zijn om bij te dragen. “Er is geen langetermijnoplossing als het veld zelf niet meedoet.”

Vergelijkbare overwegingen gelden wellicht voor het meer praktisch georiënteerde onderzoek waar waterbedrijven en instituten als KWR zich mee bezighouden?

Het belangrijkste uitgangspunt is dat water van essentieel belang is voor het leven en de industrie, aangezien water nodig is voor de productie van verschillende goederen. Het wordt erkend als een mensenrecht door de VN en de Sustainable Development Goals. Gezien het vitale belang van water als een openbaar goed, is het moeilijk te accepteren dat watergerelateerd onderzoek niet openbaar en niet open toegankelijk is. Dit standpunt heeft een filosofisch-ethische basis; het negeren van dit aspect op de lange termijn zal tot ernstige problemen leiden, gezien de vele andere grote problemen m.b.t. waterkwaliteit en -



beschikbaarheid waarmee we al geconfronteerd worden.

Bovendien is het kortzichtig om te denken dat de waterproblematiek lokaal geïsoleerd is. “We kunnen waterproblematiek niet bekijken in een lokaal kader. Dat moet opengetrokken worden, en dan is het eerlijk en efficiënter wanneer de data gedeeld wordt.” Als een drinkwaterbedrijf een probleem heeft met een component, kan een ander drinkwaterbedrijf mogelijk hetzelfde probleem hebben. Het gebrek aan communicatie leidt ertoe dat ze onafhankelijk van elkaar naar oplossingen zoeken, wat tijds- en grondstofinefficiënt is. Dit geldt ook voor het vinden van oplossingen voor opkomende verontreinigende stoffen zoals PFAS, waarbij het gebrek aan communicatie resulteert in vertraagde oplossingen.

Geldt dit voor alle data of denkt u dat er specifieke soorten data zijn waar dit voor geldt? Zijn er ook datasets bij waterbedrijven waarvan er weinig meerwaarde is als die gedeeld wordt?

Het is belangrijk om te kunnen schakelen tussen een breed overzicht en gedetailleerde informatie. Daarbij is het van belang om alleen nuttige data te delen. Om te bepalen welke data precies relevant is, is de betrokkenheid van de watersectorgemeenschap in

België en Nederland essentieel. Het probleem is echter dat weinig mensen zich geroepen voelen om hieraan deel te nemen, omdat het brede perspectief van open data nog niet volledig doorgedrongen is tot het publiek. Een ideaalbeeld zou zijn dat elk bedrijf dat water behandelt, informatie deelt over de samenstelling en kenmerken van de instroom en uitstroom van water.

De wetgevingen op het gebied van water kunnen een grote verschuiving teweegbrengen in hoe we naar water kijken, en dat is interessant. Het idee dat bedrijf X zijn werkwijze kan delen en dat andere bedrijven daarvan kunnen leren bij het opzetten van nieuwe installaties, is zeer aantrekkelijk. Dit geldt niet alleen voor Europa, maar ook voor ontwikkelingslanden en voor water in het algemeen. Zelfs voor water dat momenteel niet behandeld hoeft te worden, maar in de toekomst wel. Het delen van data zou de efficiëntie in deze processen aanzienlijk kunnen verbeteren. Bovendien, als we in staat zouden zijn om grondwater, rivierwater en oppervlaktewater op lange termijn te monitoren met behulp van verschillende parameters, dan zouden we nieuwe installaties veel robuuster kunnen maken. De fluctuaties over langere periodes in o.a. waterkwaliteit kunnen dan bijvoorbeeld in rekening gebracht worden tijdens het design van de nieuwe installatie. Nu is deze informatie vaak niet beschikbaar en wordt een installatie

gebouwd op basis van punt metingen van nabij gelegen locatie.

Over welke data hebben we het dan?

Bijvoorbeeld, stel dat bedrijf X een nieuwe installatie bouwt. Ze delen gedetailleerd elke stap van het proces. Ze monitoren het ruwe water gedurende een periode van bijvoorbeeld twee jaar. Ze delen de gegevens van het ruwe water, de specificaties van de verschillende behandelingsopties en de kwaliteit van het gezuiverde water. Dit geeft inzicht in de efficiëntie van de gebruikte technieken en de kwaliteit van het gezuiverde water. Deze kennis kan een ander bedrijf of een tweede installatie van hetzelfde bedrijf helpen. Hoewel het water nooit exact hetzelfde zal zijn vanwege variaties, kan het wel een idee geven welke richting gevolgd moet worden. Dit wordt soms betwist in het veld, waar mensen zeggen dat water altijd verschilt en niet hetzelfde kan zijn. Maar het kan wel helpen om een algemene richting te bepalen, zeker met betrekking tot nieuwe contaminanten.

Er zijn echter beperkingen waar rekening mee moet worden gehouden en het is belangrijk om te begrijpen waarom het op dit moment nog niet gebeurt. De huidige digitale infrastructuur is vrijwel niet aanwezig of onvoldoende ontwikkeld, hoewel er vooruitgang wordt



geboekt. Bovendien gaat het niet alleen om het delen van open data; het gaat erom dat mensen daadwerkelijk iets kunnen doen met de data. Het is niet voldoende om zomaar alle data op een server te plaatsen. Men moet weten wat de data betekent, hoe deze is verkregen, wanneer deze is verkregen, inclusief alle bijbehorende metadata. Daarom zijn de FAIR-dataprincipes zo belangrijk. Het is niet zinvol dat iedereen lukraak alle data deelt. Het gaat erom nuttige en bruikbare open FAIR-data te hebben.

Er moet een onderscheid gemaakt worden tussen het menselijke vlak en het technische vlak. Als het allemaal alleen om data zou gaan, zouden algoritmes het beter kunnen doen dan mensen. Maar er is natuurlijk ook het menselijke aspect dat een rol speelt, evenals de mindset. Het feit dat iedereen data wil ontvangen, maar niemand wil delen, is begrijpelijk en tegelijkertijd frustrerend. Het is waar dat het veel werk vereist van mensen om bij te dragen, vooral in de opstartfase wanneer alles nog niet helemaal duidelijk is en de digitale infrastructuur nog uitgebouwd moet worden. Er moet worden nagedacht over welke data gedeeld wordt, hoe de data verwerkt wordt, enzovoort. Ook als bepaalde analyses nodig zijn, moeten die worden uitgevoerd, wat extra werk betekent voor de deelnemer of de persoon die de data uploadt, zonder dat daar iets tegenover staat. De directe return op investement op korte termijn is voor de betreffende

personen of bedrijven soms klein. “Daarom denk ik ook dat de publieke drinkwaterbedrijven hierin echt een trekkersrol kunnen spelen.” Zij hebben geen last van de beperkingen van private bedrijfsmodellen.

En dan, terugkomend op welke data er gedeeld moet worden, begrijpt Rhea dat het uiteraard niet zo vanzelfsprekend is. Uiteindelijk zijn het de mensen in het veld die moeten beslissen wat handig en nuttig is voor hen. Er zijn echter ook zaken waar minder discussie over nodig is, bijvoorbeeld gebruikte meettechnieken en zuiveringstechnieken, en data over Emerging Contaminants, zoals PFAS. Iedereen heeft ermee te maken en iedereen moet ze tot op zekere hoogte verwijderen. Er zijn al enkele databanken beschikbaar, zoals MassBank (<https://massbank.eu/MassBank/About>), een grote spectrale massadatabank met veel van deze verontreinigingen, die ook open is. Ook het delen van informatie over waar deze verontreinigingen zich bevinden en welke technieken gebruikt kunnen worden om ze te verwijderen, lijkt haar van groot belang. Als je bijvoorbeeld kunt aantonen dat een membraanfiltratie de concentratie kan verlagen van 98 ppm naar 2 ppm, kan een ander bedrijf dat ook met die verontreiniging te maken heeft hierop inspelen. Zo kan er efficiënter en sneller omgegaan worden met verontreinigingen waar we nog niet veel ervaring mee hebben. Voor bepaalde zaken is het dus vrij eenduidig.

Hoeveel data en welke data we nodig hebben, kan echter vaag zijn, en daarvoor hebben we de betrokkenheid van de gemeenschap nodig. Mensen zullen mogelijks opbrengen dat randeffecten en zeer case-specifieke factoren de data beïnvloeden, en dat is waar, maar toch kan de data volgens Rhea iets bijdragen. Het zal nooit perfect zijn. “Het idee is om het beter te maken dan het nu is.”

Welke rol kunnen de recente EU-richtlijn m.b.t. drinkwater (2020/2184) en de uitvoeringsverordening over hoogwaardige datasets (2023/138) spelen bij het beschikbaar krijgen van drinkwaterdata?

Richtlijnen en wetten kunnen niet alleen als stimulans dienen, maar ook helpen bij een snellere verandering van de mindset. Het kapitalistisch systeem kan soms bedrijven belemmeren om bijvoorbeeld data te delen, terwijl een wet bepaald gedrag kan stoppen en in een nieuwe richtingen kan aansturen. Het is belangrijk dat de wet begrepen wordt door de mensen in het veld, zodat ze het nut en de voordelen ervan inzien. “Ik hoop dat hier aandacht aan wordt besteed, omdat begrip essentieel is voor optimale kwaliteit van de data.”

Het openbaren van welke datasets van derden zou volgens u de drinkwaterbedrijven tot nut zijn?



De analytische bibliotheken van bijvoorbeeld Thermo Fisher zijn waardevolle bronnen voor het identificeren van componenten, maar ze worden momenteel tegen hoge prijzen verkocht en zijn onwaarschijnlijk openbaar toegankelijk te worden. Dit vormt een uitdaging voor drinkwaterbedrijven in ontwikkelingslanden die niet over de middelen en apparatuur beschikken om deze analyses uit te voeren.

Een ander probleem is de traagheid van het proces, vooral met betrekking tot opkomende verontreinigende stoffen. Het publiceren van onderzoek achter een betaalmuur belemmert de toegang en vertraagt de vooruitgang. Wanneer een drinkwaterbedrijf een onbekende component in een spectrum detecteert, kost het tijd om te achterhalen wat het is. Het kan jaren duren voordat het onderzoeksartikel achter de betaalmuur wordt opgemerkt, de producent van de stof wordt gecontacteerd en een referentie-exemplaar wordt verkregen. Ondertussen blijft de verontreiniging ongemond, omdat er geen accurate detectiemethoden zijn.

Het gebrek aan toegang tot wetenschappelijke literatuur belemmert ook de ontdekking van overlappend onderzoek. Het is onaanvaardbaar dat potentieel schadelijke stoffen, zoals kankerverwekkende metabolieten, worden ontdekt aan universiteiten zonder

dat drinkwaterbedrijven of het publiek toegang hebben tot die informatie. Dit vertraagt de ontwikkeling en implementatie van nieuwe technologieën, zoals die voor PFAS-verontreiniging.

Rhea benadrukt de zware gevolgen van dit systeem en roept op tot samenwerking en waardering voor gedeelde kennis en het algemeen belang. Ze erkent echter dat dit geen gemakkelijke taak is gezien de huidige belangenconflicten, maar beklemtoont dat we binnenkort geen andere acceptabele keuze meer hebben gezien de uitdagingen die de klimaat- en watercrises met zich meebrengen.

Op welke wijze en in welke mate is standaardisatie van data nodig?

FAIR data en standaardisatie zijn essentieel voor het waarborgen van bruikbare en toegankelijke gegevens. Alleen open data is niet voldoende, omdat algoritmes en AI de gegevens niet kunnen lezen en toepassen zonder een goed opgezette digitale infrastructuur en gebruiksvriendelijke interfaces. Het is belangrijk dat de gemeenschap dit ondersteunt, aangezien zij uiteindelijk degenen zijn die de data willen gebruiken. Het is een complex proces, maar het biedt veel voordelen en heeft het potentieel om ons ver te brengen.

Wat zou u ons tot slot willen meegeven?

“Ik wil voorkomen dat mensen open en FAIR data zien als iets dat *moet* gebeuren. Want het *moet* helemaal niet. Het moet alleen maar als het nuttig is. Als het niet nuttig is voor het veld, laat het ons niet doen. Er zijn heel veel andere problemen waar we waar we ons mee kunnen bezighouden. Als het wel nuttig is voor het veld en het algemeen belang, laten we dan samen nadenken over hoe we open en FAIR data kunnen implementeren.”

Eerder in het gesprek zei u ook al dat het veld moet aangeven wat dan de belangrijke data zijn. Hoe gaan we dat doen? Moeten we met het veld een workshop organiseren? Ziet u voor zich hoe we dat hoe we dat moeten realiseren?

Een consortium vormen met vertegenwoordigers uit de academische wereld, de private drinkwaterbedrijven en de publieke sector zou nuttig zijn om verschillende meningen te horen. Een tweedaags symposium zou een goede gelegenheid zijn om met elkaar in gesprek te gaan en de toegevoegde waarde van samenwerking te bespreken. Het is inefficiënt hoe we momenteel opereren, waarbij verschillende landen afzonderlijk proberen omgaan met zaken zoals de zuivering van PFAS. Een regionaal initiatief, zoals tussen de Lage Landen, kan een goede start zijn, gezien onze gedeelde



waterwegen en zeeën. Het is belangrijk dat mensen inzien dat samenwerking een tool is om processen efficiënter te maken, om van elkaar te leren en minder terug te vallen op inefficiënte trial-and-error processen. Het delen van data (en ook van algoritmes, analyses, etc.) is geen abstract concept, maar een manier om gezamenlijk tot mooie resultaten te komen, vooral in het licht van de watercrisis. “Het is tijd om onze prioriteiten opnieuw te evalueren.”

Relevantie

De grote uitdagingen van de komende decennia hangen veelal samen met industriële ontwikkeling en vervuiling, populatiegroei met de daarbij verwachte watervraag, en ons veranderende klimaat en de daaraan gekoppelde hydrologie. De relevante fysische processen trekken zich vanzelfsprekend niets aan van bestuurlijke en organisatorische grenzen. Water stroomt daar ongehinderd doorheen en steeds vaker in steeds kleinere of juist veel grotere hoeveelheden dan voorheen. De natuurlijke schaal voor deze problematiek is die van stroomgebieden. In praktische zin is voor de beschouwing van water in stroomgebieden - voor het begrijpen en kunnen voorspellen van waterkwaliteit en -kwantiteit, voor het (her)inrichten watersystemen en waterverdelingen - data van vele en diverse instanties

die zich in het stroomgebied bevinden nodig. De gemakkelijkste manier om deze data beschikbaar te maken is door deze open te maken. Zij komt dan niet alleen beschikbaar voor alle bestuurlijke en uitvoerende entiteiten in een stroomgebied, maar ook voor onderzoekers die nieuwe en onverwachte inzichten uit de data kunnen halen, en voor bedrijven die diensten op basis van deze data kunnen aanbieden en hiermee waarde kunnen toevoegen. Let op dat dit laatste wel aanvullende competenties vergt die niet vanzelfsprekend zijn.

Om data daadwerkelijk bruikbaar te maken moet zij naast open ook FAIR zijn. Recente ontwikkelingen in AI, en specifiek op het gebied van grote taalmodellen als chatGPT, maken het nu al mogelijk om aan een chatbot te vragen om bepaalde datasets bij elkaar te halen en daar analyses op uit te voeren d.m.v. door de chatbot geschreven visualisatiescripts. Op het moment van schrijven werkt dit nog niet vlekkeloos, maar naar verwachting zal het niet lang duren voordat dit wel het geval is. Door zoveel mogelijk relevante data open, FAIR en machinetoegankelijk te maken, komen met behulp van deze nieuwe techniek de data, analyses daarop, en inzichten die eruit verkregen kunnen worden binnen handbereik van velen. En daar kunnen de watersector en de maatschappij alleen maar van profiteren, doordat beslissingen meer en meer op zoveel mogelijk relevante

informatie kan worden gebaseerd. Een randvoorwaarde is wel dat de data van voldoende kwaliteit zijn. Verder is het type licentie waaronder open data wordt verstrekt voor automatisch verwerken van data ook heel belangrijk.

Soms is het van tevoren niet duidelijk welke data gaat bijdragen aan het beantwoorden van welke vraag, maar dat wil niet zeggen dat deze bijdrage er niet zal zijn. Vanuit het oogpunt van wederkerigheid is het bovendien wenselijk dat alle organisaties in een domein (zoals water) in gelijke mate hun data delen. Daarbij kan een adagium als “open, tenzij” gehanteerd door Rijkswaterstaat een goed uitgangspunt zijn.

Tot slot

Wie zelf aan de slag wil gaan met het openstellen van data, kan bijvoorbeeld terecht bij [het Open Data Handbook](#) (2023) voor meer kennis en aanbevelingen.

AI-verklaring

Bij het redigeren van dit trendalert is gebruik gemaakt van ChatGPT. De afbeelding op de eerste pagina van dit trendalert is gegenereerd met DALL-E.

Open waterdata

Genoemd in te tekst:

<https://data-hdsr.opendata.arcgis.com/>



<https://www.waterkwaliteitsportaal.nl/>
<https://digitaledeltaorg.github.io/>
<https://wbwaterdata.org/>
<https://nasagrace.unl.edu/>
<https://www.re3data.org/>

Aanvullende bronnen:

<https://www.waterinfo.be/>
https://www.opengeodata.nrw.de/produkte/umwelt_klima/wasser/
<https://environment.data.gov.uk/water-quality/view/landing>
<https://www.waterqualitydata.us/>
<https://data.gov.au/dataset?groups=environment>
<https://opendata.swiss/en/organization/bundesamt-fur-umwelt-bafu>
<https://catalog.data.gov/dataset?tags=water>
<https://www.data.gouv.fr/en/datasets/?tag=eau>
<https://datos.gob.mx/busca/dataset?q=agua>
<https://open.canada.ca/data/en/dataset?q=water>
<https://eosc-portal.eu/>

Voorbeelden van datalicensies

Open Data Commons Open Database License (ODbL): Deze licentie is specifiek ontworpen voor het delen van databanken. Het stelt gebruikers in staat om de data te gebruiken, aan te passen en te delen, zolang ze de

oorspronkelijke bron vermelden en eventuele aanpassingen openbaar maken.

Creative Commons-licenties: Creative Commons biedt verschillende licenties die flexibiliteit bieden bij het delen van content, waaronder data. Bijvoorbeeld de CC-BY (Attribution) licentie, die gebruikers toestaat om de data te delen en aan te passen, op voorwaarde dat de oorspronkelijke bron wordt vermeld.

GNU General Public License (GPL): Deze licentie wordt vaak gebruikt bij het delen van software, maar kan ook worden toegepast op databestanden. De GPL vereist dat eventuele afgeleide werken ook onder dezelfde licentie worden verspreid.

Open Database License (ODbL): Deze licentie is vergelijkbaar met de ODbL van Open Data Commons en stelt gebruikers in staat om de data te delen en aan te passen, met vermelding van de oorspronkelijke bron.

Public Domain Mark: In sommige gevallen kan de data als 'public domain' worden vrijgegeven, wat betekent dat er geen auteursrechtelijke bescherming op rust en dat gebruikers vrij zijn om de data te gebruiken, aan te passen en te delen zonder beperkingen.

Data Commons Attribution License: Deze licentie stelt gebruikers in staat om de data te delen en aan te passen, op voorwaarde dat de oorspronkelijke bron wordt vermeld.

Meer informatie

- Simon Chignard (2013) Une brève histoire de l'Open Data. ParisTech Review. <http://www.paristechreview.com/2013/03/29/origines-open-data/>, bezocht op 16 maart 2023.
- Tim Davies, Stephen B. Walker, Mor Rubinstein en Fernando Perini (redacteuren) (2019) The State of Open Data Histories and Horizons. <https://www.africanminds.co.za/state-of-open-data/>, bezocht op 10 maart 2023.
- Digitaal Vlaanderen (2023) Het beleid achter open data <https://www.vlaanderen.be/digitaal-vlaanderen/onze-oplossingen/open-data/het-beleid-achter-open-data>, bezocht op 6 april 2023.
- DWSI (2011) Leven in het net. Denktanksessie. <https://www.dwsi.nl/leven-in-het-net/>, bezocht op 10 augustus 2023.
- DWSI (2014) Big Data: the art of the possible. Trendalert.
- DWSI (2017) Data als dwarsverband. Trendalert.
- EU (2019) RICHTLIJN (EU) 2019/1024 VAN HET EUROPEES PARLEMENT EN DE RAAD van 20 juni 2019 inzake open data en het hergebruik van overheidsinformatie. <https://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:32019L1024&from=EN>, bezocht op 3 februari 2023
- EU (2020) RICHTLIJN (EU) 2020/2184 VAN HET EUROPEES PARLEMENT EN DE RAAD van 16



- december 2020 betreffende de kwaliteit van voor menselijke consumptie bestemd water. <https://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:32020L2184&from=EN#d1e2675-1-1>, bezocht op 6 februari 2023
- EU (2023a) Open data. <https://digital-strategy.ec.europa.eu/en/policies/open-data>, bezocht op 6 april 2023.
 - EU (2023b) The European Commission decides to refer Belgium, Bulgaria, Latvia and the Netherlands to the Court of Justice of the European Union for failing to enact EU rules on open data and public sector data re-use, <https://digital-strategy.ec.europa.eu/en/news/european-commission-decides-refer-belgium-bulgaria-latvia-and-netherlands-court-justice-european>, bezocht op 6 april 2023.
 - EU (2023c) UITVOERINGSVERORDENING (EU) 2023/138 VAN DE COMMISSIE van 21 december 2022 tot vaststelling van een lijst met specifieke hoogwaardige datasets en de regelingen voor publicatie en hergebruik van die gegevens. <https://eur-lex.europa.eu/legal-content/NL/TXT/HTML/?uri=CELEX:32023R0138&from=EN>, bezocht op 3 februari 2023
 - GO FAIR (2023) FAIR Principles. <https://www.go-fair.org/fair-principles/>, bezocht op 10 maart 2023.
 - KNAW (2018) Hergebruik van publieke data – meer wetenschap en beter overheidsbeleid. [https://storage.knaw.nl/2022-06/2019-advies-hergebruik-van-publieke-data%20\(3\).pdf](https://storage.knaw.nl/2022-06/2019-advies-hergebruik-van-publieke-data%20(3).pdf)
 - Ministerie BZK (2013) Visie Open Overheid. https://data.overheid.nl/sites/default/files/uploaded_files/visie-open-overheid%20kopie.pdf, bezocht op 6 april 2023.
 - Nationaal Programma Open Science (2022) Open Science 2030 in the Netherlands. NPOS2030 Ambition Document And Rolling Agenda. <https://www.openscience.nl/npos/>, bezocht op 15 juni 2023
 - <https://zenodo.org/record/7433767#.ZA9L33bMKUK>, bezocht op 13 maart 2023
 - Open Knowledge Foundation (2023) Open Definition - DEFINING OPEN IN OPEN DATA, OPEN CONTENT AND OPEN KNOWLEDGE, <http://opendefinition.org/>, bezocht op 10 maart 2023.
 - Open Data Handbook (2023) <http://opendatahandbook.org/>, bezocht op 10 maart 2023.
 - Overheidsbreed Beleidsoverleg Digitale Overheid (2020) NL DIGITAAL Data Agenda Overheid. <https://open.overheid.nl/documenten/ronl-29467235-a671-4432-9aec-ba5c72b3ffd1/pdf>, bezocht op 6 april 2023.
 - Pronk, T.E., Wiersma, P.H., van Weerden, A, en Schieving, F. (2015) A game theoretic analysis of research data sharing. PeerJ 3:e1242 <https://doi.org/10.7717/peerj.1242>
 - Pronk, T.E. (2019) The Time Efficiency Gain in Sharing and Reuse of Research Data. Data Science Journal, 18(1), p.10. DOI: <http://doi.org/10.5334/dsj-2019-010>
 - Rijkswaterstaat (2022) Open overheid: "Winst voor iedereen". <https://www.magazinesrijkswaterstaat.nl/zakelijkeninnovatie/2022/01/wet-open-overheid>, bezocht op 15 juni 2023.
 - Rijkswaterstaat (2023) Open Data Rijkswaterstaat. <https://www.rijkswaterstaat.nl/zakelijk/open-data>, bezocht op 6 april 2023.
 - Verbeke, R. (2023) FAIR and Open Data requires proper incentives and a shift in academic culture. *Nature Water* 1, 7–9. <https://doi.org/10.1038/s44221-022-00012-1>
 - Wikipedia – Open data (2023) https://en.wikipedia.org/wiki/Open_data, bezocht op 10 maart 2023.

Keywords

open data, FAIR data