

Arjan El Fassed, directeur Open State Foundation:

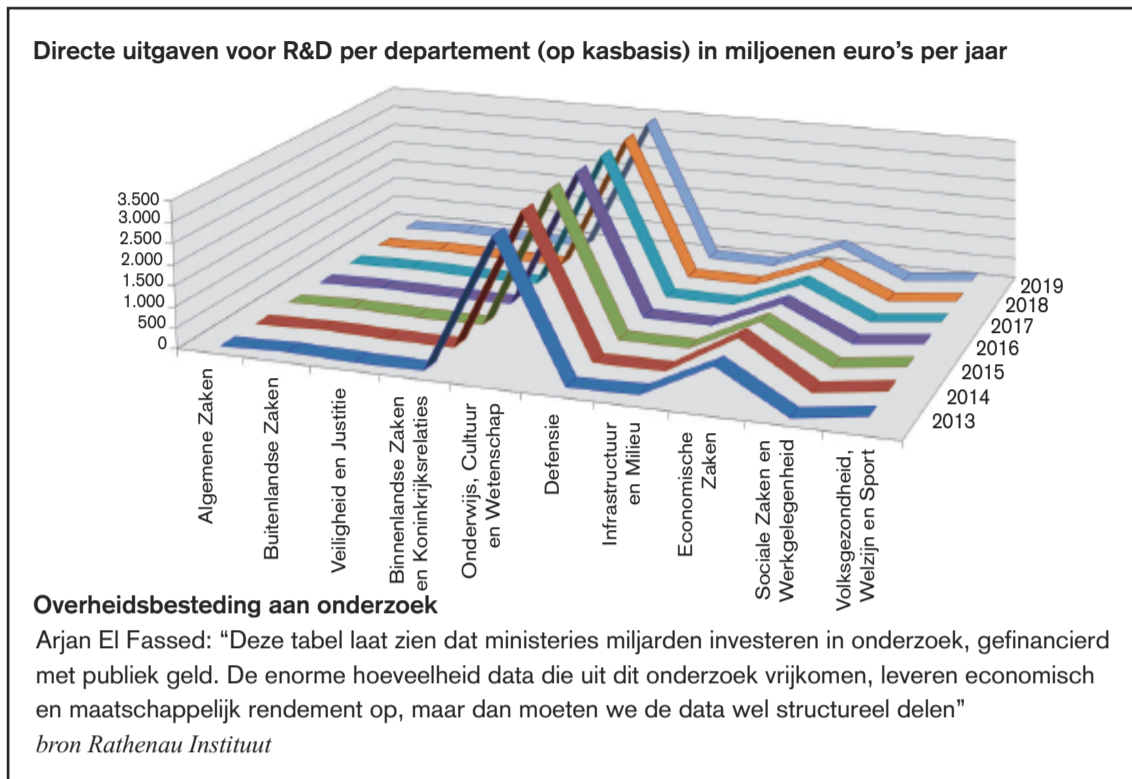
‘Kennnis is macht, mits gedeeld’

Overheden huren vaak externe adviesbureaus in voor onderzoek. Het resultaat van die onderzoeken wordt beschikbaar gesteld, maar de onderliggende data vaak niet. Wat betekent dit in de praktijk? Arjan El Fassed

Met bijna 5 miljard euro per jaar geven overheden opdracht voor grote en kleine onderzoeken. Deze onderzoeken herbergen een schat aan data die met publieke middelen zijn bekostigd en die hergebruikt zouden kunnen worden. Maar doordat onderliggende data niet beschikbaar zijn, is het voor derden onmogelijk de kwaliteit van deze onderzoeken te verifiëren en te valideren, terwijl het resultaat van die onderzoeken van grote invloed is op overheidsbeleid.

Wake up call

In de wetenschap is data delen al jaren gemeengoed, zoals onder kwantitatieve sociologen, of is dat recent geworden, zoals onder archeologen. Onder psychologen was de affaire Stapel een wake-up call. In een gesloten datacultuur bestaan echter geen mogelijkheden data te controleren en kan manipulatie met onderzoeksgegevens gemakkelijk onopgemerkt blijven. Bovendien betekent een gebrek aan open data cultuur ook, dat nuancering ontbreekt. Toen twee jaar geleden het rapport Werk aan de Wijk van het Sociaal en Cultureel Planbureau (SCP) verscheen, kopte De Volkskrant de-



zelfde dag: ‘Aanpak Vogelaarwijken mislukt’. Een te snel getrokken conclusie, want in een tijd van econo-

mische crisis was het juist een succes dat zwakkere wijken niet verder achteruit gaan.

Democratisch gat

Op de website van het SCP stond het desbetreffende rapport (als PDF-bestand) en een bijlage met een uitleg over de gebruikte databestanden. Zo was voor het onderzoek gebruik gemaakt van de Woonmilieudatabase, de Integrale Veiligheidsmonitor, de Leefbaarometer en het Woononderzoek Nederland (WoON). Weliswaar kon je wel doorklikken naar een uitleg over deze bestanden, maar kon je niet bij de data zelf. En dat terwijl al deze bestanden voortkomen uit onderzoek gedaan in opdracht van en gefinancierd door ministeries (VROM, Binnenlandse Zaken en Justitie) en diensten als de Rijksplanologische Dienst (RPD) en

het Centraal Bureau voor de Statistiek (CBS).

Het ontbreken van data voortkomend uit onderzoek in opdracht van de overheid betekent naast een democratisch gat ook dat duplicatie plaats vindt. Tel eens alle onderzoeken op dat jaarlijks in opdracht van gemeenten gedaan wordt. Denk bijvoorbeeld aan onderzoeken op het gebied van publieke dienstverlening of het terrein van decentralisaties. Dat behelst veel data in de vorm van survey onderzoek. Deze data zouden op een eenvoudige manier hergebruikt kunnen worden, ware het niet dat deze data niet beschikbaar komt.

Meer mogelijk

Welke mogelijkheden en nuanceringen ontstaan, als de uitkomsten onderling vergelijkbaar zouden zijn?

En valt er niet van alles te besparen als onderzoeksdata herbruikbaar zouden zijn? Hoe betrouwbaar zijn onderzoeken eigenlijk en hoeveel overlap zit tussen de onderzoeken? Het beschikbaar maken van onderzoek data stimuleert ook nog eens innovatie. Uit onderzoek in onder andere Denemarken en het Verenigd Koninkrijk blijkt dat het delen van onderzoek data een enorm economisch voordeel kan opleveren voor midden en klein bedrijven, waardoor ook economisch rendement te behalen valt.

Goed nieuws

Kortom: het openstellen van onderzoekdata creëert een economische en maatschappelijke meerwaarde, stimuleert innovatie en hergebruik, en zorgt voor sneller profijt van wetenschappelijke ontdekkingen. Het goede nieuws is dat Algemene Rijksvoorwaarden voor het verstrekken van opdrachten tot het verrichten van diensten (ARVODI) bestaan. Met deze voorwaarden wordt de overheid eigenaar van data die voortkomen uit onderzoek waar ze zelf opdracht toe heeft gegeven. Sommige ministeries schrappen deze bepaling echter uit contracten. En als het al in de contracten staat, wordt het niet altijd nageleefd. Ook inkopers van onderzoek op decentraal niveau zijn zich vaak niet bewust dat overheden dergelijke voorwaarden kunnen opleggen. Dat is zonde en hiermee gaat veel waarde verloren en wordt de overheid en de belastingbetaler op kosten gejaagd.

Arjan El Fassed is directeur Open State Foundation
openstate.eu/nl

Vijf tips voor het delen van data

Open State Foundation maakt met open data en hergebruik publieke informatie digitaal transparant. Vijf tips voor onderzoekers:

1. Weet wat je hebt. Doe een data inventarisatie.
2. Maak de data toegankelijk via open machine-leesbare formaten.
3. Maak gebruik van open licenties, bij voorkeur cc0.
4. Zorg ervoor dat de data vindbaar is.
5. Communiceer erover en stimuleer het hergebruik ervan.

Dutch Techcentre for Life Science

Expertiseplatform voor de life sciences

Met het Dutch Techcentre for Life Sciences (DTL) is Nederland een expertiseplatform van onderzoeksorganisaties uit de life sciences rijker. Ruben Kok, directeur DTL, vertelt. Rutger Nugteren



Ruben Kok
foto Thijs Rooijmans

“Met inmiddels 35 partners willen we een duurzaam en samenhangend netwerk vormen van lokale expertisegroepen en hun geavanceerde onderzoeksfaciliteiten, biobanken en databanken. We brengen experts uit het brede veld van de life sciences bij elkaar,” aldus Ruben. “We helpen bijvoorbeeld biomedische en klinische wetenschappers, maar ook onderzoekers uit de (agro)genomics sector, voedingsonderzoek en biotechnologie, actief bij het vinden van

zowel expertise als infrastructuur. Ook bundelen we de kennis die binnen de partners aanwezig is ten aanzien van gemeenschappelijke uitdagingen, zoals data stewardship (rentmeesterschap) en methoden en standaarden voor het combineren van data.”

Alle partners beseffen dat bepaalde

aspecten van onderzoek beter in gezamenlijkheid gedaan kunnen worden. Ruben: “We leren veel van elkaar en voorkomen dat iedereen het wiel opnieuw uitvindt. Synergie ontstaat ook door verschillende disciplines bij elkaar te brengen. DTL werkt hierin nauw samen met onderzoeksfinanciers zoals de NWO en ZonMw.”

Data4lifesciences

Een ander voorbeeld is de samenwerking tussen UMC's, NFO, SURF en DTL. Samen werken ze aan ‘Data4lifesciences’, een data- en ICT-programma, met als doel een gezamenlijke onderzoeksdata infrastructuur. “De rol van DTL ligt met name op het versterken van het interne expertise-netwerk van de

UMC's en op de aansluiting met instituten en initiatieven buiten de UMC's. Biobanken en cohorten, experimentele faciliteiten, databanken uit onderzoek en zorg: allemaal noodzakelijke ingrediënten voor een samenhangende infrastructuur die essentieel zal zijn voor toekomstig onderzoek op het vlak van gezondheid. Dit vereist een hoogwaardige computationele omgeving die niet gebouwd is op grote centrale capaciteit, maar meer op een aanpak van gedistribueerde data analyse. DTL buigt zich hier graag over.

Wilt u zich aansluiten bij het DTL platform? Heeft u vragen over DTL of over FAIR data? Neem dan contact op met Ruben Kok.
ruben.kok@dtls.nl

FAIR-data aanpak

DTL hanteert in haar werk de FAIR-data aanpak, afgeleid van de internationaal ontwikkelde FAIR data principles (datafairport.org). FAIR staat voor Findable (vindbaar), Accessible (toegankelijk), Interoperable (uitwisselbaar) en Reusable (herbruikbaar). “Door data goed te beschrijven, verbanden te leggen en betekenis en context aan datasets toe te voegen, worden datasets extra waardevol omdat ze met informatie uit andere databronnen verrijkt kunnen worden. Hierdoor kunnen nieuwe inzichten worden verworven uit data, en kan nieuwe kennis worden verkregen.”