



KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

HERGEBRUIK VAN PUBLIEKE DATA

MEER WETENSCHAP EN BETER OVERHEIDSBELEID



ADVIES



2018 Koninklijke Nederlandse Akademie van Wetenschappen (KNAW)
© Sommige rechten zijn voorbehouden / Some rights reserved

Voor deze uitgave zijn gebruiksrechten van toepassing zoals vastgelegd in de Creative Commons licentie. [Naamsvermelding 3.0 Nederland]. Voor de volledige tekst van deze licentie zie <http://www.creativecommons.org/licenses/by/3.0/nl/>

Koninklijke Nederlandse Akademie van Wetenschappen
Postbus 19121, 1000 GC Amsterdam
Telefoon + 31 20 551 0700
knaw@knaw.nl
www.knaw.nl

pdf beschikbaar op www.knaw.nl

Basisvormgeving: Edenspiekermann, Amsterdam
Engelse vertaling samenvatting: Balance, Maastricht
Opmaak en beeldbewerking: Ellen Bouma

ISBN 978-90-6984-728-3

Deze publicatie kan als volgt worden aangehaald: KNAW (2018). *Hergebruik van publieke data. Meer wetenschap en beter overheidsbeleid*. Amsterdam, KNAW.

HERGEBRUIK VAN PUBLIEKE DATA

MEER WETENSCHAP EN BETER OVERHEIDSBELEID

Koninklijke Nederlandse Akademie van Wetenschappen
December 2018

VOORWOORD

Wereldwijd streven onderzoekers ernaar data die met name voortkomen uit publiek gefinancierd onderzoek beter toegankelijk en herbruikbaar te maken. Hierbij geldt als uitgangspunt: ‘zo open als mogelijk, zo afgeschermd als noodzakelijk’.

Het Nationaal Plan Open Science, dat de KNAW en negen andere Nederlandse kennisorganisaties begin 2017 ondertekenden, is deel van de wereldwijde beweging om onderzoeksdata beter toegankelijk en herbruikbaar te maken. Het plan bouwt voort op eerdere initiatieven en stimuleert onderzoekers data en diensten van anderen waar mogelijk opnieuw te gebruiken. Voor de gezamenlijke en gecoördineerde inzet voor open science en het uitdragen van het belang ervan is het Nationaal Platform Open Science opgericht, en is Karel Luyben, oud-rector van de TU Delft, aangesteld als nationaal coördinator open science. In die rol zal hij zich de komende twee jaar inzetten om de Nederlandse ambities voor open science te realiseren en om de voortrekkersrol van ons land op dit terrein te versterken.

Nederland is dus op de goede weg, maar die weg is nog lang. De regering heeft de KNAW gevraagd aan te geven wat er nodig is om hergebruik van publieke data door onderzoekers te bevorderen. Publieke data worden verzameld door overheidsorganisaties zoals ministeries, gemeenten en waterschappen, en door private partijen die publieke taken uitvoeren. Het voorliggende adviesrapport beantwoordt deze vraag. Het geeft goede voorbeelden van overheidsterreinen waar nu al publieke data worden benut om beleid te versterken. Het geeft tevens aan dat er nog veel te winnen valt, door publieke data bewuster toegankelijk te maken voor onderzoek. Dit dient zowel het belang van de wetenschap als van de maatschappij.

Het rapport bevat enkele aanbevelingen. Aan de regering wordt aanbevolen een nationale *Chief Public Data Officer* (CPDO) in te stellen, die de regie voert op de toegankelijkheid en herbruikbaarheid van publieke data voor onderzoek. De CPDO zou overheidsorganisaties moeten aanzetten om (a) in de jaarverslagen op te nemen welke publieke data in het verslagjaar zijn verzameld, (b) datastewards aan te stellen met als taak publieke data beter toegankelijk en herbruikbaar te maken voor onderzoek, en (c) aan onderzoekers geen onnodige kosten in rekening te brengen, en transparant te zijn over de gehanteerde kostenstructuur.

Het voorliggende rapport beargumenteert mijns inziens niet alleen het belang van toegankelijkheid en hergebruik van publieke data, maar brengt dit ideaal door de concrete aanbevelingen ook daadwerkelijk dichterbij. Daarmee hoopt de KNAW, als partner van het eerste uur in het Nationaal Plan Open Science, de regering en de nationaal coördinator open science waardevolle instrumenten in handen te geven.

Wim van Saarloos
President

INHOUD

VOORWOORD	4
SAMENVATTING	7
SUMMARY	10
INLEIDING	13
Achtergrond	13
Vraagstelling en taakomschrijving commissie	14
Werkwijze commissie	15
Reikwijdte van dit adviesrapport	15
HET LANDSCHAP VAN OVERHEIDSADMINISTRATIES MET DATA VOOR	
KENNIS	16
PUBLIEKE DATA VOOR KENNIS	19
PRAKTIJKEN VAN TOEGANKELIJKHEID EN HERBRUIKBAARHEID VAN	
PUBLIEKE DATA	22
Inleiding	22
Findable	23
Accessible	25
Interoperable	31
Reusable	32
Conclusies	33
BEANTWOORDING VAN DE DEELVRAGEN UIT DE ADVIESVRAAG	34
AANBEVELINGEN	38
LITERATUURLIJST	41
BIJLAGEN	45
1. Instellingsbesluit van commissie ‘Toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek’	45
2. Experts die door de commissie zijn geraadpleegd	50
3. Reviewprocedure	51

SAMENVATTING

Om haar taken goed te kunnen uitvoeren, verzamelt de overheid van oudsher op reguliere basis data over personen, bedrijven, instellingen, markten, wegen, natuur, emissies, woningen, etc. Als onderzoekers betere toegang tot deze publieke data¹ hebben, kunnen zij daarmee nieuwe kennis produceren. Die kennis kan nuttig zijn voor het overheidsbeleid, voor het onderwijs, of voor andere toepassingen in de samenleving. Dit adviesrapport gaat over wat er nodig is om hergebruik van publieke data door onderzoekers te bevorderen.

Hergebruik van publieke data door onderzoekers dient een dubbel doel. Voor de wetenschap zijn publieke data van waarde omdat ze vaak een grote populatie en/of een lang tijdsbestek bestrijken, en doorgaans valt niemand uit de populatie van overheidsadministraties. Voor overheden zijn publieke data een potentiële bron van kennis. Kennis die licht werpt op het effect van hun beleid, en hen helpt om complexe taken uit te voeren. Zo'n complexe taak is bijvoorbeeld het hoeden van waarden als veiligheid, privacy, zelfbeschikking, solidariteit en eerlijke concurrentie. Over de effecten van een data-gedreven overheidsbeleid in een gedigitaliseerde samenleving op dit soort waarden is nog veel onduidelijk. Onderzoekers kunnen publieke data gebruiken om het beleid van een betrouwbare wetenschappelijke kennisbasis te voorzien, waarbij de brede context in het oog wordt gehouden.

Op zich beschikt Nederland over goede statistische informatie; overheden verzamelen die al lang en op een betrouwbare manier. Toch lijken nog niet alle mogelijkheden van die informatie te worden benut, blijkt uit een vergelijking met het buitenland. De

1 Niet te verwarren met 'publiek toegankelijke data'.

Scandinavische landen beschikken bijvoorbeeld over nationale registers van diagnoses, medicijngebruik, etc., die goed toegankelijk zijn voor onderzoekers. In Nederland bestaan soortgelijke longitudinale datasets ook, maar ze worden niet vanzelfsprekend beschikbaar gesteld voor onderzoek. Overheidsorganisaties maken data over personen en bedrijven vaak niet openbaar, om de privacy of bedrijfsgeheimen te beschermen.

Het Rijk, provincies, gemeenten, andere organisaties die publieke taken uitvoeren (zoals woningcorporaties), het CBS, maar ook onderzoekers zelf kunnen alle op hun eigen manier bijdragen aan een beter hergebruik van publieke data. De commissie ‘Toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek’ stelt dat het Rijk de primaire verantwoordelijkheid heeft voor het duurzaam beschikbaar stellen van data die voor publieke taken worden verzameld.

De commissie komt tot de volgende aanbevelingen.

- a. Rijksoverheid, maak interdepartementaal afspraken over de procedure waarmee en de condities waaronder overheidsadministraties data actief beschikbaar moeten stellen voor onderzoekers. Dit kan bijvoorbeeld via de Stuurgroep Data en Overheid, die ressorteert onder het ministerie van BZK. Stel een nationale *Chief Public Data Officer* (CPDO) aan, die een spilfunctie heeft in de totstandkoming van deze afspraken. Laat landelijk dekkende dataknooppunten de verstrekkers en aanvragers van publieke data faciliteren. Maak ook hierover interdepartementaal afspraken.
- b. Overheidsorganisaties, neem in uw jaarverslag op welke publieke data in het verslagjaar zijn verzameld, en hoe die voor onderzoek beschikbaar worden gesteld. Stimuleer hergebruik van data door deze FAIR te maken: *Findable, Accessible, Interoperable and Reusable*. Stel datastewards aan, die specifiek als taak hebben publieke data beter toegankelijk en herbruikbaar te maken. Naast expertise op het gebied van databeheer moeten deze datastewards gedegen kennis hebben van onder meer privacyvraagstukken en het auteursrecht. Breng onderzoekers die de data willen gebruiken niet méér in rekening dan de marginale kosten van de dataverstrekking zelf, en maak op voorhand duidelijk hoe de kosten berekend worden. Het is staande praktijk dat overheidsorganisaties aan onderzoekers kosten in rekening brengen voor het verstrekken van publieke data. Dat zij dit doen, is billijk. Er is op dit moment echter nog weinig helderheid over de kostenstructuur die zij hierbij hanteren.
- c. CBS, faciliteer zo goed mogelijk de toegankelijkheid en herbruikbaarheid van de data waar u zelf toegang toe heeft uit hoofde van uw statistische taak. CBS en ministerie van EZK, u onderzoekt momenteel hoe het juridische kader moet worden aangescherpt om private partijen met een publieke taak een juridische basis te bieden voor het leveren van hun data aan het CBS. Houd hierbij nadrukkelijk rekening met het belang van de beschikbaarheid van publieke data voor onderzoek.

- d. Onderzoekers, schroom niet om overheidsorganisaties te benaderen met een onderzoeksvraag, en die organisaties te verzoeken data uit hun administraties beter toegankelijk te maken. Uiteraard dient daarbij te worden voldaan aan de privacywetgeving, en moet onthulling van individuele gegevens worden voorkomen. Verwijs in publicaties naar de *uniform resource identifier* (URI) van de data-bronnen. Geef na afloop van een onderzoek feedback aan de overheidsorganisatie over de onderzoeksresultaten, de bijdrage aan de wetenschap, en het belang van de verstrekte data voor de onderbouwing van overheidsbeleid. Deel deze informatie bij voorkeur ook met de Chief Public Data Officer.

De commissie beveelt geenszins aan dat overheidsorganisaties méér data verzamelen voor onderzoek. Het gaat om hergebruik van reeds verzamelde publieke data. Zoals uit het adviesrapport blijkt, zijn op het gebied van milieu, arbeidsmarkt en onderwijs nu al goede voorbeelden te vinden van hoe overheidsbeleid kan worden versterkt door onderzoek. Deze voorbeelden kunnen als inspiratie dienen, zodat ook op andere terreinen kansen om publieke data in onderzoek te hergebruiken beter worden benut. Dit is in het belang van zowel de wetenschap als de maatschappij.

SUMMARY

To perform its tasks properly, government regularly collects data on individuals, enterprises, institutions, markets, roads, nature, emissions, homes, and so on. If researchers were given better access to public data,¹ they could use it to generate new knowledge. Such knowledge could be useful for government policy, for education, or for other purposes in society. This advisory report discusses what is needed to promote the reuse of public data by researchers.

The reuse of public data by researchers serves a dual purpose. For academic research, public data is valuable because it often covers a large population and/or a long period of time; in addition, people do not usually drop out of populations covered by government records. For government, public data is a potential source of knowledge – knowledge that sheds light on the effect of its policies, and helps it to carry out complex tasks. An example of a complex task is the protection of such values as security, privacy, self-determination, solidarity and fair competition. We do not really know how data-driven government policy in a digital society impacts these values. Researchers can use public data to give policy firm scientific underpinnings while bearing in mind the broader context.

The Netherlands has access to solid statistical information; public authorities have been collecting it for a long time and their work is reliable. Nevertheless, when we compare the Netherlands with other countries, we see that it is not exploiting all the opportunities that this information offers. For example, the Scandinavian countries have national registers of diagnoses, drug use and so on that are easily accessible to researchers. The Netherlands has similar longitudinal datasets, but they are not

1 Not to be confused with ‘publicly *accessible* data’.

necessarily made available to researchers. Government organisations often do not disclose data about persons and enterprises in order to protect privacy or trade secrets.

The national government, provincial and municipal authorities, other organisations that perform public tasks (such as housing corporations), Statistics Netherlands (CBS), but also researchers themselves can all contribute in their own way to improving the reuse of public data. The Committee on the Accessibility for Scientific Research of Data Administered in Public and Semi-public Record Systems argues that the national government bears primary responsibility for making data collected for public tasks permanently available.

The Committee makes the following recommendations.

- a. National government: make inter-ministerial agreements on the procedure and conditions under which government administrators must actively make data available to researchers. One option would be to do so through the Data and Government Steering Committee, which reports to the Ministry of the Interior and Kingdom Relations. Appoint a national Chief Public Data Officer (CPDO), who will play a pivotal role in concluding these agreements. Let data nodes with national coverage facilitate those who provide and request public data. Make inter-ministerial agreements about this as well.
- b. Government organisations: state in your annual report which public data has been collected in the reporting year, and how that data is made available for research purposes. Encourage reuse by making data Findable, Accessible, Interoperable and Reusable (FAIR). Appoint data stewards whose specific task is to make public data more accessible and reusable. In addition to expertise in the field of data management, the data stewards must also have a thorough knowledge of privacy issues and copyright law, for example. Do not charge researchers who wish to use the data more than the marginal cost of data provision itself, and make it clear in advance how the costs will be calculated. It is common practice for government organisations to charge researchers for providing them with public data. It is only fair that they should do so, but there is little clarity at the moment about the cost structure that they use.
- c. Statistics Netherlands: wherever possible, facilitate access to and reuse of the data to which you yourself have access by virtue of your statistical duties. Statistics Netherlands and the Ministry of the Economic Affairs and Climate Policy: you are currently examining how to tighten up the legal framework in a way that gives private parties with a public task a firm legal basis for supplying their data to Statistics Netherlands. As you do so, be especially mindful of the importance of making public data available for research purposes.
- d. Researchers: do not hesitate to approach government organisations with a research question, and ask them to make data from their records more accessible. It goes without saying that all parties must comply with privacy legislation and

prevent the disclosure of individual data. In publications, refer to the uniform resource identifier (URI) of your data sources. At the end of a study, provide the relevant government organisation with feedback on your research results, the contribution your study has made to scholarship, and the importance of the data the government organisation has provided as evidence for government policy. You should preferably also share this information with the Chief Public Data Officer.

The Committee in no way recommends that government organisations should collect more data for research purposes. Its recommendations concern the reuse of public data that has already been collected. As the advisory report reveals, there are already good examples of how research can be used to underpin government policy in many areas, such as the environment, labour markets and education. These examples can inspire greater efforts to reuse public data in research in other areas, something that serves the interests of both academic research and society.

INLEIDING

Achtergrond

Om haar taken goed te kunnen uitvoeren, verzamelt de overheid van oudsher op reguliere basis data over personen, bedrijven, instellingen, markten, wegen, natuur, emissies, woningen, etc. In dit adviesrapport draait het om die data. Daarbij gaat het niet alleen om data verzameld door bijvoorbeeld ministeries, gemeenten, waterschappen, het Centraal Bureau voor de Statistiek (CBS), de planbureaus en de rijkskennisinstellingen, maar ook om data van private partijen die publieke taken uitvoeren. Denk aan woningcorporaties, pensioenfondsen en zorgverzekeraars. Dit adviesrapport duidt gemakshalve ook de private partijen die publieke taken uitvoeren aan als overheidsorganisaties.² De overheidsorganisaties leggen de data vast in administraties. Dit adviesrapport gebruikt voor die administraties de term overheidsadministraties,³ en voor de data in die administraties de term publieke data (niet te verwarren met publiek *toegankelijke* data).

Onderzoekers⁴ gebruiken niet alleen door henzelf verzamelde primaire data, ze maken in toenemende mate gebruik van door anderen verzamelde secundaire data (hergebruik). Een deel van de data die ze hergebruiken zijn publiek. Onderzoekers ervaren echter belemmeringen bij hergebruik van publieke data, bijvoorbeeld

2 De algemene en categorale ziekenhuizen hebben geen publieke taak en worden derhalve niet tot de overheidsorganisaties gerekend.

3 De commissie spreekt in dit adviesrapport korthedshalve niet van 'publieke en semi-publieke administraties', maar van 'overheidsadministraties'. De termen zijn hier bedoeld als inhoudelijk gelijk.

4 Met de termen 'onderzoeker(s)' en 'onderzoek' bedoelt de commissie: wetenschappelijk onderzoeker(s) respectievelijk wetenschappelijk onderzoek.

gebrekkige technische eigenschappen en/of beperkte hanteerbaarheid. Soms zijn data beschikbaar in een type bestand dat minder gemakkelijk te hanteren is voor wetenschappelijke analyse, bijvoorbeeld een tekstdocument, een audio- of een video-bestand. Ook zijn sommige grote datacollecties niet digitaal beschikbaar. In andere gevallen wil de overheidsorganisatie financieel volledig worden gecompenseerd voor de kosten om de data beschikbaar te stellen. Met alle restricties rond privacy en eisen aan documentatie is het beschikbaar stellen inderdaad geen sinecure. Door allerlei belemmeringen zijn publieke data in de praktijk dus niet altijd goed toegankelijk en herbruikbaar voor onderzoek.

De toegankelijkheid en herbruikbaarheid van data uit overheidsadministraties voor onderzoek is al beduidend beter geworden sinds onder andere de KNAW daarover adviezen uitbracht (KNAW, 2006 en update 2009; KNAW, 2012). Maar er is behoefte aan verdere verbetering.

Vraagstelling en taakomschrijving commissie

Het ministerie van OCW heeft de KNAW gevraagd advies uit te brengen over de wijze waarop de toegankelijkheid van data uit overheidsadministraties voor onderzoek verder kan worden verbeterd.⁵ Deelvragen zijn:

1. Zijn er internationaal vergelijkbare procedures of *best practices* waaraan Nederland een voorbeeld zou kunnen nemen?
2. In hoeverre volstaat het huidige aanbod van open data dat het kabinet ter beschikking stelt voor onderzoek? Indien dit niet volstaat, kunt u dan concrete voorbeelden noemen van data die open beschikbaar zouden moeten worden?
3. En in hoeverre volstaan voor onderzoek de huidige ontsluiting, kwaliteit en duurzame beschikbaarheid van data uit overheidsadministraties die geen open data zijn? Indien die niet volstaan, kunt u dan concrete voorbeelden noemen van data waarvan de toegankelijkheid, kwaliteit en duurzame beschikbaarheid niet optimaal zijn?
4. Wat zou de regering aanvullend kunnen doen om data uit overheidsadministraties toegankelijk te maken voor onderzoek?
5. Kunt u voorbeelden geven van beleidsterreinen waar kansen ontstaan om beleid te versterken met behulp van onderzoek, wanneer het advies dat u geeft in antwoord op bovenstaande deelvragen wordt opgevolgd?

De KNAW heeft de commissie 'Toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek' ingesteld, met de taak de deelvragen uit de brief van de staatssecretaris van OCW te beantwoorden.

⁵ Zie bijlage 1, die ook de adviesaanvraag omvat.

Werkwijze commissie

Aan het begin van het adviestraject is uitgebreid gezocht naar relevante literatuur. Uit de zoektocht bleek dat informatie en literatuur over overheidsadministraties niet of nauwelijks voorhanden zijn. Mede om deze reden heeft de commissie door middel van diepte-interviews met experts inzicht verkregen in de mate waarin en de condities waaronder data uit overheidsadministraties toegankelijk en herbruikbaar zijn voor onderzoek. Bijlage 2 geeft de namen van de geconsulteerde experts. Dit zijn onder meer onderzoekers die data uit overheidsadministraties gebruiken. Daarnaast zijn het experts die zijn betrokken bij de dataverstrekking aan onderzoekers. Verder zijn het ervaringsdeskundigen uit België, de Scandinavische landen en het Verenigd Koninkrijk. Vooral op basis van ervaringen door onderzoekers met het gebruik van publieke data heeft de commissie voorbeelden verzameld van Nederlandse overheidsadministraties waarvan de data goed dan wel niet of nauwelijks toegankelijk of herbruikbaar zijn voor onderzoek. Uit deze voorbeelden heeft de commissie er enkele geselecteerd die de tekst van het adviesrapport ondersteunen, met name van overheidsadministraties waarvan de data nationaal georganiseerd zijn. Die voorbeelden zijn in kaders in het adviesrapport opgenomen.

De commissie heeft als tweede stap haar bevindingen uitgeschreven in een conceptrapport met aanbevelingen om de toegankelijkheid en herbruikbaarheid van data uit overheidsadministraties verder te verbeteren. Ook heeft de commissie in dit rapport kort de vijf deelvragen uit de brief van de staatssecretaris beantwoord. Hierbij gaat de meeste aandacht uit naar het beantwoorden van de deelvraag wat de regering aanvullend zou kunnen doen om data uit overheidsadministraties beter toegankelijk en herbruikbaar te maken voor onderzoek.

Als derde stap heeft de commissie haar bevindingen mondeling toegelicht, onder meer aan vertegenwoordigers van het ministerie van OCW en van de interdepartementale Stuurgroep Data en Overheid, die ressorteert onder het ministerie van BZK.

Reikwijdte van dit adviesrapport

De doelstelling van dit KNAW-advies is dat de toegankelijkheid en herbruikbaarheid van data uit overheidsadministraties voor onderzoek verder wordt verbeterd. Daardoor kunnen onderzoekers meer en beter empirisch onderzoek doen, dat relevant kan zijn voor overheidsbeleid.

Gezien het brede scala van overheidsadministraties geeft dit adviesrapport geen uitputtend overzicht van die administraties zelf. Wel vermeldt het enkele verwijzingen naar catalogi van overheidsadministraties, en illustratieve voorbeelden. Het adviesrapport presenteert ook geen uitputtend overzicht van de wetenschapsgebieden die met overheidsadministraties te maken hebben. Wel worden hiervan voorbeelden gegeven.

HET LANDSCHAP VAN OVERHEIDSADMINISTRATIES MET DATA VOOR KENNIS

Het publieke bestel in Nederland is de afgelopen decennia langzaam veranderd. Binnen het bestel streven overheidsorganisaties naar meer bedrijfsmatig werken. Om dit te bereiken zetten zij informatie- en communicatietechnologie in. Vrijwel elke overheidsorganisatie verzamelt daartoe data, die zij als managementinformatie hanteert. Een gevolg van deze ontwikkeling is een expansie aan door overheidsorganisaties verzamelde data.

Wát een overheidsorganisatie registreert, heeft als het goed is altijd een publiek belang. Voor een goed begrip van de data uit overheidsadministraties is deze constatering cruciaal. Een voorbeeld van een dataverzameling met een publiek belang is de *Atlas Bestrijdingsmiddelen in Oppervlaktewater* (zie kader 1).

KADER 1. BELEID BESTRIJDINGSMIDDELEN OPPERVLAKTEWATER

Het Centrum voor Milieuwetenschappen (CML) van de Universiteit Leiden verzorgt samen met ingenieursbureau Royal Haskoning DHV sinds 2003 de Atlas Bestrijdingsmiddelen in Oppervlaktewater. Deze online atlas biedt instanties, bedrijven en particulieren op een toegankelijke wijze een betrouwbaar ruimtelijk beeld van de toestand van en trends in gewasbeschermingsmiddelen in de Nederlandse oppervlaktewateren. *Waar, welke en hoeveel* gewasbeschermingsmiddelen er in het water zijn gevonden, wordt in één oogopslag duidelijk. Metingen van onder meer de waterschappen leveren open, voor iedereen toegankelijke data op.

Voor agrarisch ondernemers zijn de gegevens uit de Atlas van groot belang, omdat zij bestrijdingsmiddelen gebruiken om hun oogst groter en zekerder te maken. De overheid wil juist kunnen verifiëren of de maatregelen die zij treft om duurzame teelt te bevorderen goed werken. Recentelijk is bijvoorbeeld het middel imidacloprid

bediscussieerd, dat zou bijdragen aan de verhoogde bijensterfte. Voor dit debat zijn de gemeten concentraties van deze stof in oppervlaktewater cruciaal. Dankzij de Atlas kunnen toegelaten stoffen op de Nederlandse markt worden geëvalueerd, en kunnen agrariërs richtlijnen en adviezen opdoen.

Niet alleen worden in het publieke bestel veel data verzameld, ook zetten steeds meer overheidssectoren stappen om een eigen dataknooppunt te ontwikkelen. Voorbeelden van dergelijke knooppunten zijn:

- **Publieke Dienstverlening Op de Kaart** (PDOK), een samenwerking van het Kadaster, de ministeries van Infrastructuur en Waterstaat (I&W), Binnenlandse Zaken en Koninkrijksrelaties (BZK) en Economische Zaken en Klimaat (EZK), Rijkswaterstaat (RWS) en Geonovum. Dit knooppunt ontsluit door Nederlandse overheden verzamelde data over de leefomgeving (zie kader 2).
- **DUO Open Onderwijsdata**, waarbij de Dienst Uitvoering Onderwijs (DUO) van het ministerie van OCW gegevens online beschikbaar stelt over het bekostigde onderwijs in Nederland.
- **Dataland**, een intergemeentelijk samenwerkingsverband op het gebied van vastgoedinformatie, met gegevens over bijna 9 miljoen gebouwde objecten op adresniveau.
- **Landelijk Register van Zorgaanbieders (LRZa)**, een landelijk register waarin alle zorgaanbieders zijn opgenomen. Het is een virtueel register, dat andere bronnen ontsluit maar de gegevens hieruit niet kopieert. Belangrijke bronnen zijn het Handelsregister van de Kamer van Koophandel (KvK), het AGB (Algemeen GegevensBeheer)-register van Vektis en het BIG (Beroepen in de Individuele Gezondheidszorg)-register van het Centraal Informatiepunt Beroepen in de Gezondheidszorg (CIBG).

KADER 2. PUBLIEKE DIENSTVERLENING OP DE KAART

Op 15 mei 2007 is de Europese kaderrichtlijn INSPIRE (Infrastructure for Spatial Information in Europe) van kracht geworden, met het doel de vindbaarheid, kwaliteit en toegankelijkheid te bevorderen van gegevens die gerelateerd zijn aan milieubeleid. De richtlijn is in 2009 verankerd in de Nederlandse wet- en regelgeving.

Circa vijf jaar geleden is het platform Publieke Dienstverlening Op de Kaart (PDOK) tot stand gekomen door een samenwerking tussen het Kadaster, de ministeries van I&W, BZK en EZK, RWS en Geonovum. PDOK stelt digitale informatie over de leefomgeving als dataservices en bestanden beschikbaar. De PDOK-diensten zijn gebaseerd op open data en daarom voor iedereen vrij beschikbaar. PDOK zorgt voor de ontsluiting van de geo-basisregistraties en andere kern-geodatasets zoals het Nationaal Wegen Bestand. De PDOK-diensten voldoen aan nationale en internationale standaarden, waaronder de Europese INSPIRE-standaard en de Nederlandse e-overheidstandaarden.

In het Nederlandse datalandschap heeft het CBS een bijzondere positie.⁶ Dit bureau heeft de data-infrastructuur voor overheidsorganisaties mede vormgegeven. Het functioneert als het expertisecentrum bij uitstek voor de inzet van big-data-analyses bij het vergaren van statistische informatie. Het CBS verzamelt data van meer dan tweehonderd overheidsinstellingen. Zo kan het overheden helpen bij statistische vraagstukken, bijvoorbeeld over de economische ontwikkeling van dorpen en steden (Ministerie EZK, 2018). Het CBS zegt over zijn positie in het datalandschap:

‘Om het CBS in staat te stellen betrouwbare statistieken te maken en tegelijk de administratieve lasten voor burgers en bedrijven te verlagen, heeft het CBS bij wet toegang gekregen tot publieke data. Zodoende worden data die al door andere overheidsorganisaties zijn vergaard, hergebruikt, verrijkt en gecombineerd voor statistische doeleinden. Dit heeft grote voordelen: het bespaart geld, de mate van detail van de cijfers is vergroot, de snelheid waarmee informatie beschikbaar komt, is toegenomen, de administratieve lasten zijn gereduceerd en de kwaliteit van de data én van de statistiek is verbeterd. Het CBS is hierdoor in essentie het dataknooppunt geworden van de overheid voor het maken van statistische informatie. Ook is er hierdoor bij het CBS veel kennis van en ervaring met dataverwerking en privacybescherming.’ (CBS, 2018, blz. 8 en 9)

6 In formele zin is het CBS gepositioneerd onder het ministerie van EZK. De formele positionering van het nationale statistische bureau verschilt sterk per land. Er zijn geen voorschriften voor: in sommige landen (waaronder dus Nederland) valt het onder het ministerie van Economische Zaken, in andere onder het ministerie van Financiën, dan weer direct onder de (minister-)president of het ministerie van Algemene Zaken. Duitsland heeft een getrapte systeem met eigen statistische bureaus bij de bundesländer.

PUBLIEKE DATA VOOR KENNIS

Publieke data zijn een bron van potentiële kennis voor overheidsbeleid. Onderzoekers kunnen op basis van die data wetenschappelijke kennis genereren die nuttig is voor het overheidsbeleid: kennis die opheldering geeft over de kwaliteit van uitvoering van publieke taken en bijdraagt aan het oplossen van *wicked problems* en aan integraal beleid. Wicked problems zijn kwesties waarbij normatieve consensus over het probleem en over de oplossing ontbreekt, en waar de kennis onzeker is. Dit geldt bijvoorbeeld voor het klimaatvraagstuk. Onderzoek aan de hand van publieke data kan de waarschijnlijkheid van toekomstige ontwikkelingen scherper in beeld brengen, waardoor het debat zuiverder kan worden gevoerd. In integraal beleid – zoals in het Interbestuurlijk Programma (IBP) – pakken het Rijk, de provincies, de gemeenten en de waterschappen grote maatschappelijke opgaven aan die zich lang niet altijd op eenzelfde schaalniveau bevinden (Overhedenoverleg van 14 februari 2018). Deze aanpak vereist niet alleen een gedeelde data-infrastructuur, maar ook nieuwe, gedeelde kennis over het meerschallige karakter van die opgaven en de dwarsverbanden daartussen. Vanuit de overtuiging dat een wendbare overheid blijvend op zoek is naar nieuwe beleidsinterventies en -instrumenten kan men ervoor kiezen om te experimenteren en de meest kansrijke ideeën door te ontwikkelen. Die werkwijze wordt ook wel omschreven als het ‘Strategielab’ (brief 28 maart 2017 van de secretarissen-generaal aan de kabinetsinformatuur). In die labs kunnen data als hulpmiddel worden ingezet ten behoeve van integraal beleid. Hoewel experimenteren met beleid bijzonder lastig is – ook ethisch – zouden onderzoekers op basis van die data wetenschappelijke kennis kunnen genereren die nuttig is voor het overheidsbeleid. Een goed voorbeeld daarvan staat in kader 3. Dit voorbeeld verduidelijkt om welk type experimenten het kan gaan, en laat ook de methodologische problemen zien: zelfselectie, geen controle op omgeving, etc.

KADER 3. EXPERIMENTEN MET DE PARTICIPATIEWET

Bij de evaluatie van gemeentelijke experimenten rond re-integratie in het arbeidsproces is vanaf het begin ingezet op de bereidheid van overheidsorganisaties om kennis te genereren voor overheidsbeleid. Op 22 februari 2017 is het *Tijdelijk besluit experimenten Participatiewet* (AMvB) vastgesteld, en gepubliceerd in het Staatsblad. Het besluit beoogt gemeenten de ruimte te geven te experimenteren met een eigen uitvoering van de Participatiewet, om deze effectiever te maken.

Onder de AMvB zijn zes gemeenten toegelaten die willen experimenteren met een nieuw soort begeleiding van cliënten naar werk of andere vormen van participatie. Ze doen dit bijvoorbeeld door hun cliënten een basisinkomen te geven, of hen vrij te stellen van de sollicitatieplicht. Positieve bejegening en maatwerk staan centraal, anders dan bij de standaardaanpak van verplichtingen en tegenprestatie, alsmede korting bij niet-naleving.

De gemeenten willen graag onderzoeken wat het effect van deze andere aanpak is op de re-integratie in de arbeidsmarkt.

Het ministerie van Sociale Zaken en Werkgelegenheid (SZW) heeft onderzoekers van de TiU en de RUG gevraagd de experimenten in een onderzoeksproject te evalueren. De betrokkenen stelden vast dat het CBS bij uitstek over zowel de vereiste juridische basis als beveiligde IT-omgeving beschikt om de data – afkomstig van verschillende bronnen – veilig met elkaar te kunnen koppelen. Vervolgens is samen met het CBS nagegaan hoe de koppeling van data kon worden gerealiseerd binnen de kaders van de Algemene verordening gegevensbescherming (AVG). De eerste gegevens voor de nulmeting zijn in februari 2018 door de gemeente Groningen aan het CBS geleverd. Het CBS maakt de data koppelbaar, om ze voor het project te analyseren in combinatie met overige CBS-data. De onderzoekers zijn op het moment van schrijven bezig met deze analyse. Zij zullen de resultaten via praktijkgerichte publicaties en internationale tijdschriften verspreiden.

Dit project laat gemeenten zien die vanaf het begin hebben ingezet op de koppelbaarheid van data. Zij waren bereid en gemotiveerd om experimenteel beleid gedegen extern te laten evalueren, en schrokken niet terug voor het overleg en de extra doorlooptijd die nodig zijn voor een juridisch en wetenschappelijk verantwoorde aanpak. In een vroeg stadium zochten zij afstemming met onderzoekers.

Het (her)gebruik van publieke data voor onderzoek behoeft niet alleen aanmoediging, maar ook kritische doordenking (o.a. O'Neil, 2016). Digitalisering en een data-gedreven overheid ('dataficering') leiden immers niet vanzelfsprekend tot beter overheidsbeleid. Data bieden allerlei mogelijkheden om beleid van een kennisbasis te voorzien, maar wat precies het effect hiervan is in de praktijk, is nog onduidelijk. Voorzichtigheid is daarom geboden. Ook roept de digitale transformatie van de samenleving fundamentele vragen op: bijvoorbeeld over privacybescherming, wie toegang heeft tot data, wat betrouwbare informatie is, hoe mensen de digitale transformatie kunnen bijbenen, en hoe deze transformatie zich verhoudt tot het functioneren van de democratie (Ministerie EZK, 2018). Daar komt nog het risico van een

vertekende voorstellingen van de werkelijkheid bij indien data niet in de juiste context worden beschouwd. Aandacht voor en een evenwichtige afweging van risico's en valkuilen die óók verbonden zijn met het beter toegankelijk en herbruikbaar maken van publieke data voor onderzoek zijn derhalve geboden.

Onderzoekers maken graag gebruik van data uit overheidsadministraties. Dit hergebruik kan bijdragen aan wetenschappelijke kennis (Welle Donker et al., 2017; Nederlandse gedragscode wetenschappelijke integriteit, 2018, #61). Immers:

- Afhankelijk van de wetenschappelijke vraagstelling kunnen de data uit overheidsadministraties reeds voorhanden datasets van de onderzoeker inhoudelijk sterk verrijken. Onderzoekers hoeven dan minder data (opnieuw) te verzamelen. Dit kan kostenbesparend werken.
- De data uit overheidsadministraties betreffen vaak een grotere populatie of een langer tijdsbestek dan die van surveyonderzoek en panels.
- In tegenstelling tot de steekproeven in longitudinale panels valt doorgaans niemand uit de populatie van overheidsadministraties.
- De kwaliteit van de data uit overheidsadministratie kan beter zijn dan in bijvoorbeeld surveyonderzoek, waar resultaten kunnen lijden onder vertekende geheugeneffecten van de respondenten.

Onderzoekers kunnen op basis van publieke data wetenschappelijke kennis genereren die nuttig is voor het overheidsbeleid. Bovendien vindt de aldus verkregen wetenschappelijke kennis ook haar weg naar het onderwijs en naar uiteenlopende toepassingen in de samenleving.

Hergebruik van publieke data heeft ook andere voordelen voor overheidsorganisaties. Uit de literatuur (Kronenburg et al., 2012; Algemene Rekenkamer, 2016) blijkt dat er door meer hergebruik kostenbesparingen zijn gerealiseerd bij de Rijksdienst voor het Cultureel Erfgoed (RCE; minder archiefruimte en minder receptionisten), de Dienst Uitvoering Onderwijs (DUO; een besparing van 5 fte) en het Nationaal Archief. Daarnaast kan de kwaliteit van publieke data beter worden als onduidelijkheden en fouten worden gerapporteerd aan de overheidsadministratie.

Internationale en nationale regelgeving laat ruimte voor onderzoek en statistiek met data. Er is ook ruimte voor onderzoek en statistiek met de meest privacygevoelige data, al worden dan wel strengere voorwaarden gesteld. Mits aan deze voorwaarden wordt voldaan, kunnen onderzoekers data uit overheidsadministraties dus hergebruiken.

Is er dan niets aan de hand? Dat is er wél. Want vrijwel elke onderzoeker heeft in de praktijk moeite om data uit overheidsadministraties in een zodanige vorm te krijgen dat ze voor onderzoek te hergebruiken zijn. Om erachter te komen waar dit aan ligt, is het nuttig systematisch na te gaan welke stappen een onderzoeker zet. Dit gebeurt in het volgende hoofdstuk.

PRAKTIJKEN VAN TOEGANKELIJKHEID EN HERBRUIKBAARHEID VAN PUBLIEKE DATA

Inleiding

Wereldwijd streven onderzoekers ernaar dat data die voortkomen uit onderzoek dat uit publieke middelen wordt gefinancierd beter toegankelijk en herbruikbaar worden gemaakt. Hierbij gelden de FAIR-principes: *Findable, Accessible, Interoperable and Reusable*. Onlangs is dit bevestigd in een workshop die is gehouden onder de patronage van de Europese leden van de International Council for Science (ICSU), waaronder de KNAW, in samenwerking met All European Academies (ALLEA). Op basis van de workshop is de verklaring *Open Data in Science in Europe* opgesteld, om te bevorderen dat publiek gefinancierde onderzoekers hun onderzoeksdata beschikbaar maken.⁷ Over de FAIR-principes zegt de verklaring het volgende:

‘The FAIR Principles (...) serve scientists as guidelines to make data available in such a way that they can be used for other research, and to allow the verification of research results to enhance their quality, robustness and reliability. Science academies help operationalise and implement the FAIR Principles.’ (European ICSU members, 2018)

De FAIR-principes werden voor het eerst officieel gepubliceerd in 2016 (Wilkinson et al., 2016). Er zijn meerdere omschrijvingen van de FAIR-principes in omloop. Op de

7 In tegenstelling tot *open data* staat *data hugging*. *Data hugging* kan worden omschreven als een al te grote terughoudendheid om data te delen. Een fundamentele angst van onderzoekers is dat de *credits* van het verzamelen van de data niet toereikend aan hen worden toegevoegd. Maar niet alleen onderzoekers, ook overheden zouden aan *data hugging* van *raw data* doen, zei Tim Berners-Lee in 2009 (zie de Ted-talk van Berners-Lee: *Stop Hugging Data*).

website van FORCE11⁸ is een uitgebreide versie te vinden. Die op de website van het Dutch Techcentre for Life Sciences (DTL)⁹ is helder en functioneel:

- *Findable*: de data en metadata zijn gemakkelijk vindbaar door zowel mensen als computers. Machine-leesbare metadata zijn essentieel voor automatische ontdekking van relevante datasets en -diensten.
- *Accessible*: beperkingen op het datagebruik, en protocollen voor het bevragen en kopiëren van data zijn expliciet gemaakt voor zowel mensen als machines.
- *Interoperable*: de computer kan de data interpreteren, zodat ze automatisch kunnen worden gecombineerd met andere data. Het gaat hier om bijvoorbeeld verschillen tussen hardware-ontwerpen, besturingssystemen, programmeertalen en communicatieprotocollen.
- *Reusable*: de data en metadata zijn zodanig goed omschreven voor zowel mensen als machines dat ze kunnen worden gerepliceerd of gecombineerd in toekomstig onderzoek.

Verschillende wetenschapsorganisaties en wetenschapsgebieden ontwikkelen inmiddels standaarden, tools en trainingen op basis van de FAIR-principes (Aerts, 2017; Uitkomsten van de ePLAN-workshop DIGITALIA, 2017; SURF, 2018). Het Nationaal Platform Open Science – waarin onder meer de KNAW participeert – bouwt voort op de vele initiatieven op het gebied van de toegankelijkheid van data die al gaande zijn, en stimuleert onderzoekers data en diensten van anderen waar mogelijk te hergebruiken (Nationaal Plan Open Science, 2017). Dergelijke initiatieven passen in het beleid van het Kabinet-Rutte III, dat erop is gericht dat open science en open access de norm worden in wetenschappelijk onderzoek (Regeerakkoord 2017-2021, 2017). Verbeterde toegankelijkheid en herbruikbaarheid van data die voortkomen uit onderzoek dat uit publieke middelen wordt gefinancierd, zijn aspecten van open science. De minister van OCW stelt investeringen in de digitale infrastructuur in het vooruitzicht om de ambities op het gebied van open science te kunnen verwezenlijken (Minister OCW, 9 maart 2018).

Dit adviesrapport ziet publieke data als een deelverzameling van alle data die voor onderzoek worden gebruikt en meent dat de hierboven beschreven FAIR-principes evengoed van toepassing zijn op data uit overheidsadministraties. Het vervolg van dit hoofdstuk gaat nader in op de mate waarin publieke data FAIR kunnen zijn.

Findable

Aan het begin van het onderzoek moet de onderzoeker achterhalen welke data uit overheidsadministraties het meest geschikt zijn om de wetenschappelijke vraag te beantwoorden. De onderzoeker kan niet putten uit een compleet overzicht van publieke data, want zo'n overzicht is niet voorhanden.

8 Zie <https://www.force11.org/group/fairgroup/fairprinciples>

9 Zie <https://www.dtls.nl/fair-data/fair-data/>

Afhankelijk van het wetenschapsgebied en de onderzoeksvraag kan de onderzoeker zoeken in de lappendeken van overheidsadministraties met datasets. Een startpunt voor de zoektocht naar geschikte data is wellicht het portaal *data.overheid.nl* dat is ingericht in het kader van het open-databaseleid van de overheid.¹⁰ De daar aangetroffen datasets zijn voorzien van beschrijvende informatie (metadata), waardoor ze in principe gemakkelijker vindbaar zijn. Vanwege de bescherming van persoons- en bedrijfsgegevens staan datasets met dit soort gegevens niet vermeld op het portaal. Dit is voor onderzoek dat gebruikmaakt van gegevens over personen of bedrijven een belangrijke beperking.

De onderzoeker kan datasets ook benaderen via de dataknooppunten die eerder in dit rapport zijn beschreven. Verder is bij bijvoorbeeld het Rijksinstituut voor Volksgezondheid en Milieu (RIVM) en de Rijksdienst voor Ondernemend Nederland (RVO) een groot aantal datasets te vinden. Het geheel aan websites overziend, constateert de commissie dat de mate waarin data toegankelijk zijn zeer verschilt. Bijvoorbeeld het RIVM biedt data over de leefomgeving aan die door iedereen te downloaden zijn. Veel databestanden bevatten echter uitsluitend geaggregeerde gegevens. Brondata zijn sowieso vaak niet beschikbaar. Overigens laat de organisatie, inhoud en functionaliteit van met name de RVO-website naar het oordeel van de commissie nog een aanzienlijke ruimte voor verbetering.

Het zoekproces naar geschikte data voor een onderzoek kan een tijdrovende kwestie zijn. Als publieke data een *persistent identifier* zoals een *digital object identifier* (DOI) hebben en de metadata goed zijn beschreven, verbetert en versnelt dit het zoekproces aanzienlijk.

Het CBS fungeert niet alleen als hét dataknooppunt voor de overheid, maar verleent ook microdata-diensten aan onder meer Nederlandse universiteiten en wetenschappelijke organisaties. De datacatalogus van het CBS is zeer uitgebreid. De wettelijke taak van het CBS is het maken van betrouwbare statistieken; dit omvat ook data gereedmaken en data beschikbaar maken. Tegelijk heeft het CBS als taak de administratieve lasten voor burgers en bedrijven te verlagen. Het CBS stelt een schat aan data uit overheidsadministraties beschikbaar aan de wetenschappelijke wereld. Er zijn echter ook veel databestanden waarover het CBS niet beschikt. Momenteel onderzoeken het ministerie van EZK en het CBS op welke wijze het juridische kader voor het

10 Het kabinet streeft ernaar zoveel mogelijk van de data uit overheidsadministraties beschikbaar te stellen als zogenoemde 'open data'. In 2016 en 2017 heeft het ministerie van BZK de overheidsbrede 'open data inventarisatie' uitgevoerd. De publiek beschikbare data staan vermeld op het portaal <http://data.overheid.nl>. De data op dit portaal zijn dezelfde die de overheidsorganisaties beschikbaar hebben; ze hebben dezelfde kwaliteit en betrouwbaarheid. Het ministerie van BZK is verantwoordelijk voor de afstemming tussen overheidsorganisaties die data leveren.

CBS¹¹ moet worden aangescherpt om private partijen met een wettelijke/publieke taak (zoals ziektekostenverzekeraars/Vektis, woningcorporaties, etc.) en private partijen met een niet-wettelijke taak maar wel met een semipubliek karakter (openbaar vervoer, openbare telecommunicatie en andere nutsvoorzieningen, etc.) een goede juridische basis te bieden voor het leveren van hun data aan het CBS.

Accessible

De onderzoeker die gebruik wil maken van publieke data (bijvoorbeeld persoons- of bedrijfsgegevens) zal al gauw ontdekken dat die data óf niet openbaar zijn óf uitsluitend in geaggregeerde vorm óf uitsluitend onder gecontroleerde veiligheidscondities beschikbaar zijn. Kader 4 geeft een goed voorbeeld van onderzoek waarbij overheidsorganisaties het beschikbaar stellen van data zelfs op wijkniveau gevoelig vinden.

KADER 4. UITDAGINGEN BIJ GEBRUIK VAN DATABRONNEN UIT DE JEUGDGEZONDHEIDSZORG

Het project *Datagestuurd preventief werken* ontwikkelt een model om problemen bij gezinnen op tijd te identificeren of zelfs te voorkomen. 'Een zorgprofessional of leraar heeft vaak een onderbuikgevoel, maar weet dat gevoel niet precies te onderbouwen', vertelt Patricia Prüfer, projectleider bij CentERdata. 'Samen met Sterk Huis, de gemeente Tilburg en het Data Science Centre van Tilburg University zoeken wij naar de onderbouwing van dat gevoel. We hebben een lijst van indicatoren die een potentieel onveilige situatie voor kinderen voorspellen. De achterliggende data van deze indicatoren verzamelen we bij verschillende bronnen: gemeente, CBS, GGD, GGZ, leerplicht-ambtenaar, politie, scholen en eerdere onderzoeken. Door de data te onderzoeken met hulp van *machine learning*, willen we de giftige cocktail van risicofactoren opsporen.' De grootste uitdagingen voor het onderzoek blijken praktisch van aard. 'Iedereen is enthousiast, iedereen steunt het onderzoek. Maar als je om de gegevens vraagt, vindt men ze toch te gevoelig om beschikbaar te stellen, ook al analyseren we de data niet op persoonsniveau maar op wijkniveau', vertelt Prüfer. Wat ook blijkt: de datakwaliteit is lang niet altijd optimaal. Prüfer adviseert: 'Zoek de juiste contacten met bevoegdheden om data te delen, reserveer veel tijd voor gesprekken en voorlichting, zorg voor een hoger aggregatieniveau als data anders niet gedeeld mogen worden. Wees transparant en deel je kennis, zo kweek je vertrouwen. Om data steeds bruikbaar en het model steeds beter te maken, moet je de professionals die met een registratiesysteem werken, trainen.' Eind dit schooljaar moet de pilot het eerste voorspellingsmodel opleveren. 'We willen interactieve sessies organiseren met beleidsmedewerkers, leerkrachten en ouders. Door hen te vragen of ze herkennen wat de analyses opleveren kan het model waar nodig bijgesteld en verbeterd worden. We zien het onderzoek als één levend systeem van dataproducten, analisten en eindgebruikers. De hele keten is lerend',

11 Dit juridische kader bestaat uit de Wet op het Centraal Bureau voor de Statistiek en het Besluit Gegevensverwerving CBS.

aldus Prüfer. Uiteindelijk is het doel om het model op landelijk niveau uit te rollen. Prüfer: 'Vaak zien onderzoekers (maar ook zorgprofessionals) alleen wat ze al kennen. Zoals een huisarts die ontkent dat er huiselijk geweld wordt gepleegd door patiënten van zijn praktijk. Data science combineert bestaande informatie. Dat biedt een kans om tunnelvisie te doorbreken en open te kijken naar wat er echt aan de hand is.'

(Uit: *E-data&research*, juni 2018b)

Uit de gesprekken met experts heeft de commissie opgetekend dat onderzoekers belemmeringen van diverse aard ervaren die het daadwerkelijke hergebruik van data hinderen. Voorbeelden zijn:

- De overheidsorganisatie brengt naar voren dat er wet- en regelgeving in de weg staat van het beschikbaar stellen van de data. Daarbij kunnen privacyaspecten van de data als argument worden gehanteerd voor het beperken van de toegang tot die data (zie kader 5). Ook sinds de invoering van de AVG in mei 2018 blijft er bij overheidsorganisaties onzekerheid welke data zij wel, en welke zij niet beschikbaar kunnen stellen in het licht van internationale en nationale regelgeving (Frankowski et al., 2015).¹²
- Er zijn overheidsorganisaties die consequent nagaan welke wet- en regelgeving wel en welke niet van toepassing is bij het beschikbaar stellen van data. Een goed voorbeeld daarvan is het RIVM, dat onderzoek heeft laten doen naar de eisen die aan open data worden gesteld (zie kader 6).
- De overheidsorganisatie verlangt in ruil voor het beschikbaar stellen van de data een financiële vergoeding van de onderzoeker. Dat overheidsorganisaties dit doen, is billijk. De vergoedingen zijn bedoeld om een deel van de uitgaven voor de eigen infrastructuur en de administratieve en organisatorische handelingen te dekken. Maar overheidsorganisaties hanteren hierbij zeer verschillende bekostigingsmodellen. Laatstgenoemd punt werd door een aantal experts onder de aandacht gebracht van de commissie. Ook bijvoorbeeld het CBS brengt voor het gebruik van zijn diensten kosten in rekening bij de onderzoeker. Enerzijds kan de

¹² Die onzekerheid is er overigens ook bij onderzoekers. Een procedure die onderzoekers kunnen volgen om te beoordelen of het koppelen van gezondheidsgegevens van meerdere instellingen voldoet aan de relevante wetgeving is beschreven in de *Koppelcode Gezondheidsgegevens* (Biolink&BBMRI-NL, 2015): 'Voor het beantwoorden van een wetenschappelijke onderzoeksvraag kan het nodig zijn gegevensbestanden van verschillende organisaties aan elkaar te koppelen. Hiervoor kunt u bijvoorbeeld gebruik maken van de 'Koppelcode'. Deze Koppelcode is door verschillende Nederlandse organisaties die met zulke gegevensbestanden werken, zoals ziekenhuizen of biobanken, opgesteld en getekend. Onderdeel van deze Koppelcode is een lijst met vragen die de onderzoeker zich moet stellen om inzichtelijk te maken of aan de relevante wetgeving wordt voldaan. Het invullen van deze vragenlijst kan ook voor u nuttig zijn. Voor hulp bij het beantwoorden van de vragen of bij het beoordelen of aan de relevante wetgeving wordt voldaan, kunt u contact opnemen met de METC, nWMO-toetsingscommissie of Research Office van uw instelling'. Voor meer hulp kunnen onderzoekers een beroep doen op de ELSI Servicedesk voor ethische, juridische en maatschappelijke vraagstukken over *personalized medicine & next generation sequencing*.

onderzoeker de kosten die het CBS in rekening brengt als een belemmering ervaren voor de toegang tot publieke data, naast de belemmering die de onderzoeker ervaart van de restricties die op sommige datasets zitten (omdat ze onder de CBS-wet vallen). Anderzijds is de onderzoeker blij met het CBS, want vanwege de maatregelen van het CBS om persoons- en bedrijfsgegevens te beschermen, zijn onderzoekers in staat op microniveau toch van die gegevens gebruik te maken.

- De overheidsorganisatie heeft intern niet de juiste (kwalitatief) of onvoldoende (kwantitatief) menskracht voorhanden om de data beschikbaar te stellen.
- De overheidsorganisatie is niet bereid data te delen, omdat de kwaliteit van de data te gering zou zijn (zie kader 7).
- Wetenschappelijke kennis staat of valt bij de (mogelijkheden voor) verificatie en replicatie van de gevonden uitkomsten. Onderzoekers mogen publieke data dus niet argeloos gebruiken (KNAW, 2018b). Hun kwaliteitsbeoordeling richt zich vooral op de zorg die de overheidsadministratie heeft betracht bij het vergaren van de data, en op de metadata. De beoordeling op deze aspecten is bedoeld om de toegankelijkheid van de data voor hergebruik te verbeteren (SURF, 2010).
- De overheidsorganisatie wil geen inzage geven in de data, uit vrees dat er claims vanuit de samenleving volgen als de onderzoeksresultaten op basis van die data zouden worden gepubliceerd. Denk aan gezondheidsschade vanwege de aardbevingen in Groningen (zie kader 5) of vanwege fijnstofconcentraties langs snelwegen.
- De overheidsorganisatie is niet intrinsiek gemotiveerd. Zij vraagt zich af waarom zij eigenlijk data beschikbaar zou willen stellen aan de wetenschap.

KADER 5. MOEILIJK TOEGANKELIJKE DATA OVER SCHADEMELDINGEN IN GRONINGEN

In september 2015 heeft de Nationaal Coördinator Groningen onderzoekers van de RUG gevraagd onderzoek te doen naar de beleving van de aardbevingen in Groningen. De onderzoekers werkten samen met de Groninger gemeenten, de provincie Groningen, de Veiligheidsregio Groningen, de GGD Groningen en de afdeling Onderzoek & Statistiek van de gemeente Groningen, en binnen de RUG met het Kenniscentrum Aardbevingen en Duurzame Ontwikkeling (KADO). Ook zijn openbare data over bodembeweging uit bronnen van het KNMI gebruikt. In een panel en via de GGD zijn mensen bevraagd. Verder gebruikten de onderzoekers gegevens uit Lifelines (longitudinale gezondheidsdata over 167.000 personen), gegevens die het CBS heeft verzameld (bijvoorbeeld over woningwaarde, de sociaaleconomische positie van inwoners en van buurten, en de mate van urbanisatie) en gegevens van het Kadaster (over kenmerken van gebouwen). Het onderzoek gaat onder meer over de impact op bewoners van schade aan gebouwen. Gegevens over schademeldingen zijn dan ook essentieel voor het onderzoek, maar de data van het Centrum Veilig Wonen (CVW) – het bedrijf dat schademeldingen registreerde – bleken slechts zeer beperkt beschikbaar te worden gesteld aan de onderzoekers. Begin 2018 is de Tijdelijke Commissie Mijnbouwschade Groningen (TCMG) van start gegaan. Daar zijn de data uit het schaderegister nu

ondergebracht. Onderzoekers van de RUG zijn in gesprek met de TCMG over de toegankelijkheid van data uit het schaderegister.

KADER 6. RIVM HAALT ZIJN DATA DOOR DATAWASSTRAAT

Het rapport *RIVM open data* (Van Loenen c.s., 2016), opgesteld op verzoek van het RIVM, gaat in op de 'datawasstraat' van het RIVM. Die voorziet in samenhangende afwegingskaders, processen en een raamwerk voor datamanagement. Het RIVM kan de datawasstraat gebruiken om te bepalen of data als open data gepubliceerd mogen worden, en in welke vorm dit mogelijk is. Maar het gebruikt de datawasstraat ook om te bepalen of er nog elementen van datamanagement zijn die moeten worden verbeterd. De datawasstraat is niet bedoeld als een geautomatiseerd systeem; vooral het juridische aspect is dermate complex dat het uitwerken daarvan kennisintensief mensenwerk blijft.

De 'beslisboom open data' is onderdeel van de datawasstraat en geeft invulling aan het juridische aspect. Om data te kunnen vrijgeven als open data zijn voor het RIVM verschillende factoren van belang. Die hebben te maken met achtereenvolgens: hoe om te gaan met de bescherming van persoonsgegevens, eventuele aansprakelijkheid voor de gevolgen van het vrijgeven van de data, intellectuele eigendomsrechten en andere wettelijke kaders die van belang zijn, zoals de speciale positie van milieudata en van gegevens die onder het nieuwe Nagoya Protocol vallen.

KADER 7. VOEDSEL- EN WARENAUTORITEIT WEIGERT HERGEBRUIK INSPECTIEGEGEVENS

Sinds de Nederlandse Voedsel- en Warenautoriteit (NVWA) in 2014 een start maakte met de Horeca Inspectiekaart zijn inspectieresultaten nog niet beschikbaar als open data. Slechts 5% van deze resultaten is openbaar gemaakt. Sinds 2016 publiceert de NVWA weliswaar inspectieresultaten in een pdf-bestand, maar verzoeken tot hergebruik op basis van de Wet hergebruik overheidsinformatie stuiten op bezwaren van de NVWA. Aanvankelijk om 'technische redenen' en later omdat naar de mening van de NVWA de verzochte gegevens 'uiteindelijk het resultaat [zijn] van een door een natuurlijke persoon verrichte handeling'.

(Van website Open State Foundation, augustus 2018)

De onderzoeker kan de belemmeringen benoemen in het rechtstreekse contact met de overheidsorganisatie die de data heeft verzameld. Voor een aantal van die belemmeringen zal een *modus operandi* kunnen worden gevonden die voor beide partijen acceptabel is. Maar het overlegproces is in veel gevallen zeer tijdrovend en de afloop niet altijd bevredigend.

In het licht van de wet- en regelgeving op het gebied van persoonsgegevens kan het CBS onderzoekers bijstaan: sinds 2004 biedt het CBS aan onderzoekers verbonden

aan geautoriseerde instellingen de mogelijkheid via een beveiligde internetverbinding vanuit hun eigen locatie (*remote access*) te werken met microdatabestanden.¹³ De CBS-wet geeft het CBS een handvat om onderzoekers in staat te stellen in een beveiligde omgeving met persoonsgegevens te werken, waarbij de privacy van de deelnemers is gewaarborgd. Een goed voorbeeld van onderzoekers die samenwerken met het CBS is het project Nationaal Cohortonderzoek Onderwijs (zie kader 8). Hierbij wordt gebruikgemaakt van gegevens die bij de Dienst Uitvoering Onderwijs (DUO) beschikbaar zijn en gegevens van het CBS over bijvoorbeeld de achtergrond van leerlingen en hun ouders. Het project Nationaal Cohortonderzoek Onderwijs (NCO) maakt het mogelijk om zowel vragen van het ministerie als vragen uit het veld (scholen) te beantwoorden. Ook kunnen onderzoekers in het NCO op maat gemaakte aanbevelingen doen aan scholen. Die zijn op hun beurt betere data gaan aanleveren; een duidelijke win-winsituatie.

KADER 8. NATIONAAL COHORTONDERZOEK ONDERWIJS

Vanaf 2016 geeft het Nationaal Regieorgaan Onderwijsonderzoek (NRO) het project Nationaal Cohortonderzoek Onderwijs (NCO) vorm. Dit onderzoek naar cohorten stelt de reeds aanwezige data over leerlingen centraal (i.c. schoolloopbaangegevens van DUO) en koppelt deze aan andere relevante data, met name achtergrondgegevens van leerlingen en hun ouders die bij het CBS beschikbaar zijn. Om ervoor te zorgen dat de privacy van de leerlingen wordt geëerbiedigd, worden de data opgeslagen bij het CBS en alleen aan onderzoekers beschikbaar gesteld via de bij het CBS gebruikelijke veilige routes. Als het mogelijk is data te aggregeren of aan te passen op een zodanige wijze dat ze niet meer tot individuele leerlingen herleidbaar zijn, worden ze opgeslagen bij DANS, het Nederlands instituut voor permanente toegang tot digitale onderzoeksgegevens.

Voor allerlei onderzoeksvragen is aanvullend materiaal nodig. Dat geldt in de eerste plaats voor de voortgangsgegevens van leerlingen, zoals die opgeslagen zijn in de leerlingvolgsystemen van scholen. Besturen in het primair en voortgezet onderwijs zijn bereid deze gegevens beschikbaar te stellen. Maar daarnaast is het ook nodig nieuwe data te verzamelen, omdat niet alles in registers beschikbaar is. Hierbij gaat het bijvoorbeeld om gegevens over burgerschap en sociaal-emotionele ontwikkeling. Ook zijn er internationale onderzoeken zoals het Programme for International Student Assessment (PISA, een grootschalig internationaal vergelijkend onderzoek dat wordt uitgevoerd onder auspiciën van de OECD), waarvoor nieuwe data moeten worden verzameld. De opzet van het NCO is draagvlak te creëren bij de scholen om aan dit soort onderzoek mee te werken. Bij wijze van tegenprestatie krijgen de scholen een overzicht van waar hun leerlingen zich bevinden in het vervolgonderwijs (vanaf 2017) en van de voortgang van hun leerlingen (vanaf 2018) vergeleken met die van referentiescholen. Deze rapporten kunnen een belangrijke rol spelen in de kwaliteitszorg van de scholen.

13 Voor details, zie <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen>

De dataverzameling en het combineren van gegevens maken het mogelijk om gericht onderzoek te doen en ook om additionele vragen uit te zetten en te beantwoorden.

De infrastructuur ODISSEI (zie kader 9) is een ander goed voorbeeld van onderzoekers die samenwerken met het CBS. Daarbij worden grootschalige en longitudinale dataverzamelingen gekoppeld met CBS-data, waarbij de privacy van de deelnemers is geborgd.

KADER 9. ODISSEI: TOEGANG TOT GROOTSCHALIGE DATAVERZAMELINGEN VOOR DE SOCIALE WETENSCHAPPEN

De Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI) ontwikkelt een duurzame wetenschappelijke infrastructuur voor de sociale wetenschappen in Nederland. Via ODISSEI krijgen onderzoekers in de sociale wetenschappen toegang tot grootschalige en longitudinale dataverzamelingen die zijn gekoppeld met CBS-data op de Cartesius supercomputer bij SurfSARA. Dit virtuele netwerk maakt het mogelijk nieuwe, discipline-overstijgende onderzoeksvragen te beantwoorden en bestaande vragen op een nieuwe manier te onderzoeken. ODISSEI is een gezamenlijk initiatief van de projectleiders van grootschalige dataverzamelingen, het CBS en NWO, en is opgenomen in de Nationale Roadmap Grootschalige Wetenschappelijke Infrastructuur 2016-2020. ODISSEI-deelnemers krijgen een financiële tegemoetkoming in de kosten voor het gebruik van de microdatabestanden. Er wordt een gefedereerde data-infrastructuur gerealiseerd.

Om de privacy van personen en bedrijfsgeheimen te beschermen, zijn technologische maatregelen (gegevensbescherming via ontwerp en via standaardinstellingen) in combinatie met organisatorische maatregelen nodig. Zo kan worden gegarandeerd dat data over personen of bedrijven alleen worden gebruikt voor onderzoek, zonder kans op onthulling van individuele gegevens en zonder dat meer mensen toegang hebben dan strikt noodzakelijk is (Nederlandse gedragscode wetenschappelijke integriteit, 2018). Er is een scala van maatregelen beschikbaar of in ontwikkeling.

- Onder strikte voorwaarden toegang krijgen tot de microdata bij het CBS via een beveiligde internetverbinding vanuit de eigen locatie (*remote access*).
- In de arm nemen van een *trusted third party* (TTP). Dit is een instantie die bijvoorbeeld broncode, datacertificaten en sleutels voor derden in bewaring neemt, transacties vastlegt en daardoor uitwisseling van informatie transparant maakt voor alle betrokken partijen: de overheidsadministratie, de onderzoekers en de toezichthouder. De TTP moet volstrekt onafhankelijk zijn. Er zijn diverse TTP-dienstverleners die hun diensten, tegen betaling, aanbieden. (SURFsara, 2016)
- Algoritmische oplossingen, zoals encryptie van de data of het delen van parameters en modellen die de data beschrijven (onder andere op basis van *machine learning*), zonder de data zelf te delen.
- Technieken op basis van *secure multiparty computation*, waardoor meerdere partijen gezamenlijk aan gefedereerde data kunnen rekenen alsof ze een gedeelde

database hebben en tegelijkertijd elkaars data nooit kunnen inzien.

- Gefedereerde analyses. Soms is het niet mogelijk alle data van alle cohorten bij elkaar te brengen. De oplossing is om per cohort de analyses uit te voeren en de resultaten te delen. De associaties zijn meestal niet privacygevoelig en kunnen worden gedeeld, waarna via meta-analyses toch grootschalige analyses kunnen worden uitgevoerd.
- Werken op een hoger aggregatieniveau dan het persoons- of bedrijfsniveau. Deze maatregel zal niet op alle onderzoeksopzetten aansluiten, maar kan voor bepaalde typen onderzoek een goede oplossing zijn.
(Uit onder meer: *E-data&research*, juni 2018a)

Interoperable

Als bijvoorbeeld de ene overheidsorganisatie een bedrijf registreert onder een bepaald type identificatienummer, en dat bedrijf wordt door een andere overheidsorganisatie geregistreerd onder een ander type identificatienummer, dan is het koppelen van de data een moeizame aangelegenheid voor de onderzoeker. Doorgaans maakt de onderzoeker de data meer interoperabel door met die overheidsorganisaties over de definities en identificerende gegevens af te stemmen. Dit is de inhoudelijke dimensie van interoperabiliteit.

Daarnaast is er de technische dimensie van interoperabiliteit. In deze dimensie maakt de onderzoeker data meer interoperabel door een ict-standaard (of data-format) te gebruiken waarmee ze leesbaar zijn voor onderzoekers (eenheid van taal) en hun computers (machine-leesbaar). De keuze voor een standaard hangt af van wat gangbaar is in een wetenschapsgebied en wat een onderzoeksinstituut beschikbaar heeft. GO-FAIR¹⁴ ondersteunt het maken van standaarden. Om te leren die standaarden toe te passen, neemt de onderzoeker bijvoorbeeld deel aan een Essential Skills-workshop van het Netherlands eScience Center of aan een Bring Your Own Data (BYOD)-workshop van het Dutch Techcenter for Life Sciences (DTL, zie kader 10). Data-, competence of vergelijkbare centra in de eigen onderzoeksinstelling ondersteunen de onderzoeker bij diens activiteiten om de data technisch meer interoperabel te maken. Bij deze centra werken ook datastewards, die onder andere verantwoordelijk zijn voor het beheer van de onderzoeksgegevens. Daarnaast bieden op nationaal niveau faciliteiten zoals SURF en het Netherlands eScience Center ondersteunende diensten aan.

14 Zie <https://www.go-fair.org/>

KADER 10. DE BRING YOUR OWN DATA-WORKSHOP

Tijdens een *Bring Your Own Data* (BYOD)-workshop werken functionarissen bij overheidsorganisaties, onderzoekers in een bepaald wetenschapsgebied, databasebeheerders en FAIR data-experts gezamenlijk aan specifieke datasets. De data-experts assisteren bij het verbeteren van de FAIRness van de onderzoeksdata. Het belangrijkste doel van de workshop is om te leren hoe datasets FAIR te maken zijn met behulp van linked-datatechnologie, en de datasets te combineren met andere FAIR datasets om een wetenschappelijke vraag te beantwoorden.

(Van website Dutch Techcente for Life Sciences, april 2018).

De geconsulteerde experts hebben diverse soorten belemmeringen voor de interoperabiliteit onder de aandacht gebracht van de commissie. Voorbeelden zijn:

- Data zijn niet gedigitaliseerd of computer-leesbaar. Het medium dat is gebruikt voor de data (papier, ponskaarten, digitaal, geannoteerd) staat de interoperabiliteit in de weg. Bijvoorbeeld de langjarige historische data over rekruten die het ministerie van Defensie heeft verzameld (handgeschreven op papier) zijn interessant voor onderzoek naar veranderingen in de gezondheid van jonge mannen, maar niet machine-leesbaar. Vanuit het perspectief van het onderzoek is digitalisering van historische data alleen mogelijk onder voorwaarde dat de context van de oorspronkelijke verzameling goed wordt beschreven. Dat dit hoge kosten met zich meebrengt, staat buiten kijf. In bijvoorbeeld het Verenigd Koninkrijk en Zweden zijn dergelijke digitaliseringsprojecten wel met succes gerealiseerd.
- Data zijn wel gedigitaliseerd maar toch niet herbruikbaar. Ze hebben bijvoorbeeld de vorm van een pdf, in plaats van Excel of iets vergelijkbaars.
- Datasets van gemeenten zijn niet gemakkelijk doorzoekbaar omdat er meerdere ict-systemen naast elkaar bestaan. De commissie heeft overigens het vermoeden dat hetzelfde geldt voor ministeries, provincies en uitvoeringsorganisaties.

Reusable

Nadat de onderzoeksresultaten openbaar zijn gemaakt, deponeert de onderzoeker de onderzoeksdata en de metadata – maar niet de brondata – doorgaans bij een instituut voor permanente toegang tot digitale onderzoeksgegevens, zoals DANS (zie ook de Nederlandse gedragscode wetenschappelijke integriteit, 2018, #35). De brondata zijn de publieke data die beschikbaar zijn bij de overheidsorganisatie. Tot de metadata behoort alles wat duidelijk maakt hoe de onderzoeksdata zijn verzameld en bewerkt uit de brondata. Onder metadata valt bijvoorbeeld documentatie over de gehanteerde onderzoeksmethodologie, codeboeken en machine-instellingen. De metadata moeten zodanig goed zijn omschreven dat het onderzoek in principe kan worden gerepliceerd, of gecombineerd met andere onderzoeksresultaten. Als de metadata ontbreken of onvolledig zijn, is dit een belemmering voor het hergebruik van publieke data, want dan zijn de onderzoeksdata niet goed te traceren tot de brondata.

Conclusies

Er zijn bij overheidsorganisaties geen formele belemmeringen voor het beschikbaar stellen van data aan onderzoekers.

Wel hebben onderzoekers te maken met grotere en kleinere praktische belemmeringen indien zij data uit overheidsadministraties willen hergebruiken. Welke belemmeringen er precies zijn, hangt mede af van het wetenschapsgebied waarin de onderzoekers werken. Uit de gesprekken met experts is naar voren gekomen dat het de moeite loont te investeren in een goede relatie met de overheidsorganisatie: er moet vertrouwen bij de overheidsorganisatie worden opgebouwd.

Het volgende hoofdstuk beantwoordt kort de deelvragen van de staatssecretaris van OCW. Het daaropvolgende hoofdstuk geeft een aantal aanbevelingen die de toegankelijkheid en herbruikbaarheid van publieke data voor onderzoek verder kunnen verbeteren.

BEANTWOORDING VAN DE DEELVRAGEN UIT DE ADVIESVRAAG

Het ministerie van OCW heeft de KNAW gevraagd¹⁵ advies uit te brengen over de wijze waarop de toegankelijkheid van data uit overheidsadministraties voor onderzoek verder kan worden verbeterd. De adviesaanvraag bestaat uit deelvragen, die in dit hoofdstuk opeenvolgend worden beantwoord.

Zijn er internationaal vergelijkbare procedures of best practices waaraan Nederland een voorbeeld zou kunnen nemen? (deelvraag 1)

De commissie heeft zich tot België, de Scandinavische landen en het Verenigd Koninkrijk beperkt omdat het hergebruik van publieke data door onderzoekers daar het sterkst is geïnstitutionaliseerd.

Uit de consultatie van ervaringsdeskundigen die betrokken zijn bij de dataverstrekking aan onderzoekers in deze landen blijkt dat de technische mogelijkheden voor hergebruik van publieke data voor onderzoek met name rond vraagstukken van gezondheid, ziekte en cognitieve en schoolse ontwikkeling beter zijn dan die in Nederland. Dit ondanks het feit dat er in Nederland de afgelopen twintig jaar wel vooruitgang is geboekt. Voor wat betreft de mogelijkheden voor onderzoek naar economische, sociale en geografische vraagstukken meenden diverse experts echter dat Nederland juist als *best practice* zou kunnen dienen voor andere landen, vooral dankzij de rol die het CBS voor de wetenschap speelt.

Uit de internationale inventarisatie zijn best practices naar voren gekomen waarvan Nederland zou kunnen leren. De commissie geeft twee voorbeelden.

Het eerste voorbeeld betreft de Scandinavische landen. Die beschikken over nationale registers met data over voorgeschreven geneesmiddelen die via apotheken worden afgegeven (Furu et al., 2009). Ook beschikken zij over nationale registers

15 Zie bijlage 1.

met data over diagnoses en behandelingen van patiënten.¹⁶ Vergelijkbare registers bestaan niet in Nederland; de Nederlandse registers zijn niet landelijk dekkend en gaan minder lang terug in de tijd. De gezondheidsregisters in de Scandinavische landen bevatten een persoonsidentificatienummer (PIN) dat koppeling mogelijk maakt aan censusgegevens, data van scholen, overlijdensregisters, etc. Koppeling vereist aparte toestemming van de nationale ethische commissies en van de administraties die de data verzamelen. Ook blijken de Scandinavische landen – meer dan Nederland – mogelijkheden te bieden voor de financiering van onderzoek op basis van publieke data. Voorbeelden van recente initiatieven zijn de *Nordic Programme on Health and Welfare* (2014-2018) om onderzoek op basis van gekoppelde gezondheids- en sociaaleconomische gegevens te faciliteren, de *Nordic Information for Action eScience Center* (2014-2018) om data uit biobanken en gezondheidsregisters beter te kunnen gebruiken, en de *Nordic Microdata Access Network* (2015-2017) om voor onderzoeksprojecten de toegang tot data uit registers te faciliteren (Palmgren, 2017). Het tweede voorbeeld betreft het Verenigd Koninkrijk. Dit land heeft – meer dan Nederland – diensten in het leven geroepen voor de analyse van tijdreeksen op basis van censusgegevens en data uit nationale gezondheidsregisters.¹⁷ In Nederland is het weliswaar mogelijk op basis van sociale, economische, demografische en gezondheidsgegevens uit de CBS-registers longitudinale datasets samen te stellen, maar de resulterende datasets worden niet vanzelfsprekend beschikbaar gesteld voor hergebruik. Wel worden syntaxen gedeeld.

In hoeverre volstaat het huidige aanbod van open data dat het kabinet ter beschikking stelt voor onderzoek? Indien dit niet volstaat, kunt u dan concrete voorbeelden noemen van data die open beschikbaar zouden moeten worden? (deelvraag 2)

De wetenschap kan weinig met het begrip ‘open data’ zoals het kabinet dit invult. Het woord ‘open’ heeft de connotatie ‘gratis’ en/of beschikbaar op internet. Maar in de praktijk zijn niet alle open data gratis en/of beschikbaar op internet. Sommige publieke data worden voor iedereen beschikbaar gesteld, meestal op internet. Data over de leefomgeving die overheidsorganisaties verzamelen zijn hiervan een voorbeeld. Daarentegen stellen overheidsorganisaties data over personen, bedrijven en andere organisaties niet voor iedereen beschikbaar, om de privacy van personen of bedrijfsgeheimen te beschermen. Het CBS heeft bij wet toegang gekregen tot publieke data om betrouwbare statistieken te maken. Het verleent

16 Een voorbeeld is het Zweedse *National Patient Register*, zie <https://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish>

17 Voor bijzonderheden, zie <https://www.gov.uk/government/organisations/home-office/about/statistics> en <http://www.ucl.ac.uk/iehc/research/epidemiology-public-health/research/celsius>

onderzoekers onder passende voorwaarden toegang tot een groot aantal van die data.¹⁸ Het gaat hier om juridische voorwaarden (toegang wordt alleen verleend aan volgens de CBS-wet gemachtigde instellingen, onder contract en geheimhoudingsverklaringen), technische voorwaarden (de strikt gepseudonimiseerde data blijven binnen een netwerkomgeving van het CBS, waarop geautoriseerde onderzoekers slechts via een streng beveiligde procedure kunnen inloggen), statistische voorwaarden (alleen geaggregeerde gegevens mogen worden geëxporteerd), publicatieplicht en vergoeding voor de marginale kosten die het CBS maakt voor het mogelijk maken van deze onderzoeksfaciliteit.

En in hoeverre volstaan voor onderzoek de huidige ontsluiting, kwaliteit en duurzame beschikbaarheid van data uit overheidsadministraties die geen open data zijn? Indien die niet volstaan, kunt u dan concrete voorbeelden noemen van data waarvan de toegankelijkheid, kwaliteit en duurzame beschikbaarheid niet optimaal zijn? (deelvraag 3)

Vanuit het perspectief van de wetenschap zijn de voordelen van publieke data helder, of het wel of niet open data zijn, maakt niet uit. Onderzoekers maken graag gebruik van publieke data. Met name de beperkte interoperabiliteit van de publieke data is echter een belemmering voor hergebruik. Hier zou een voorbeeld kunnen worden genomen aan de inspanningen van onderzoekers om onderzoeksdata meer interoperabel te maken. Dit doen zij door een ict-standaard (of dataformat) te gebruiken waarmee de data leesbaar zijn voor onderzoekers (eenheid van taal) en hun computers (machine-leesbaar). Navolging van dergelijke standaarden voor publieke data kan leiden tot minder (uitvoerings)kosten voor de overheid en grotere effectiviteit van het onderzoek dat op basis van die data wordt uitgevoerd.

Op dit moment zijn sommige publieke data relatief slecht beschikbaar voor onderzoek. Het Rijk, de provincies, de gemeenten en andere organisaties die publieke taken uitvoeren, hebben elk hun eigen verantwoordelijkheden. De commissie is van mening dat het Rijk het laatste woord heeft voor het duurzaam beschikbaar stellen van data die voor publieke taken worden verzameld. Daartoe behoren ook de data die door bijvoorbeeld woningcorporaties worden gegenereerd: dit zijn weliswaar private organisaties, maar ze voeren publieke taken uit.

Wat zou de regering aanvullend kunnen doen om data uit overheidsadministraties toegankelijk te maken voor onderzoek? (deelvraag 4)

De belangrijkste aanbeveling aan de regering is via een *Chief Public Data Officer* (CPDO) te monitoren of data uit overheidsadministraties voor onderzoek beter

¹⁸ Voor de *Dienstencatalogus microdata services* van het CBS, zie https://www.cbs.nl/-/media/_pdf/2017/48/171201_dienstencatalogus_2018.pdf

toegankelijk en herbruikbaar worden.¹⁹ Dit doet de CPDO mede op basis van de informatie die onderzoekers verstrekken aan het einde van hun onderzoek.

Indien na enkele jaren blijkt dat data uit overheidsadministraties niet of nauwelijks beter toegankelijk en herbruikbaar worden, kan de regering overwegen in alle wetgeving een dataparagraaf op te nemen waarin de toegankelijkheid en herbruikbaarheid voor onderzoek wordt geregeld van de data die in het kader van die wetgeving worden verzameld. Een goed voorbeeld is de regeling die in 2015 van kracht is geworden bij de decentralisatie van de jeugdzorg.

Kunt u voorbeelden geven van beleidsterreinen waar kansen ontstaan om beleid te versterken met behulp van onderzoek, wanneer het advies dat u geeft in antwoord op bovenstaande deelvragen wordt opgevolgd? (deelvraag 5)

Voorbeelden van overheidsterreinen waar kansen worden benut om beleid te versterken op basis van onderzoek met publieke data zijn het milieubeleid (zie kader 1), het arbeidsmarktbeleid (zie kader 3) en het onderwijsbeleid (zie kader 8). Op deze en andere overheidsterreinen kunnen en moeten kansen om publieke data in onderzoek te hergebruiken beter worden benut, met wetenschappelijke en maatschappelijke meerwaarde als gevolg. Het is in dit adviesrapport geenszins de bedoeling dat overheidsorganisaties meer data verzamelen voor onderzoek. Het gaat om hergebruik van reeds verzamelde publieke data.

19 In bijvoorbeeld Estland, Frankrijk, Griekenland, Ierland, Japan, Korea, Mexico, Nieuw Zeeland, Polen en het Verenigd Koninkrijk is een *chief data officer* aangesteld met coördinerende en monitorende taken ten aanzien van dataketens van de publieke sector (OECD, 2018).

AANBEVELINGEN

VOOR DE REGERING:

Een aantal overheidsorganisaties is om uiteenlopende redenen terughoudend in het beschikbaar stellen van data voor onderzoek. Daarbij lijken sommige redenen meer steekhoudend dan andere. Beleg daarom de verantwoordelijkheid voor de toegankelijkheid en herbruikbaarheid van data uit overheidsadministraties expliciet bij een nationale *Chief Public Data Officer* (CPDO), die de volgende functies heeft:

- Klankbordfunctie.
Meedenken met overheidsorganisaties om het hergebruik van publieke data te bevorderen.
- Ombudsfunctie.
Intermediëren bij vastgelopen dataverzoeken, en daarbij bemiddelen tussen overheidsadministraties, onderzoekers en eventuele andere personen en instanties die publieke data hergebruiken. Bij alle overheidsadministraties zijn functionarissen gegevensbescherming aangesteld in het licht van de AVG. De CPDO kan deze functionarissen ondersteunen, en eraan bijdragen dat zij niet onnodig besluiten tot *datahugging*.
- Monitorfunctie.
Monitoren of data uit overheidsadministraties voor onderzoek beter toegankelijk en herbruikbaar worden. De CPDO doet dit mede op basis van de informatie die onderzoekers aan de overheidsadministratie verstrekken over de onderzoeksresultaten, de bijdrage aan de wetenschap, een het belang van de verstrekte data voor de onderbouwing van overheidsbeleid.
- Initiërende en stimulerende functie.

Enkele voorbeelden van initiërende en stimulerende activiteiten die bijdragen aan de toegankelijkheid en herbruikbaarheid van publieke data voor onderzoek:

- Bij onderzoekers verkennen welke datasets beter toegankelijk en herbruikbaar zouden moeten worden gemaakt. Een goed gearticuleerde vraag – zoals de vraag naar gegevens over de ontwikkeling van kinderen die beschikbaar zijn bij Centra voor Jeugd en Gezin – kan een overheidsorganisatie motiveren te investeren in verbetering van toegankelijkheid en herbruikbaarheid van die datasets. Daarbij is sturing vanuit de wetenschap welkom: waaraan heeft de wetenschap behoefte? Wat zou bijvoorbeeld ‘de dataset van de maand’ zijn?
- Een pilot opzetten met onderzoekers, die op basis van data uit specifieke overheidsadministraties een goed gearticuleerde wetenschappelijke vraag formuleren. Met die administraties overleggen op welke wijze onderzoekers toegang kunnen krijgen tot de data.

VOOR ALLE OVERHEIDSORGANISATIES:

- Maak interdepartementaal – bijvoorbeeld via de Stuurgroep Data en Overheid die ressorteert onder het ministerie van BZK – afspraken over de procedure waarmee en condities waaronder overheidsadministraties data die zij verzamelen actief beschikbaar stellen voor hergebruik door onderzoekers. Roep overheidsorganisaties op actief bij te dragen aan het formuleren van deze procedure en condities. Betrek hierbij ook de private organisaties die publieke data verzamelen, zoals woningcorporaties.
- Neem in de jaarverslagen op welke publieke data in het verslagjaar zijn verzameld en op welke wijze die voor onderzoek beschikbaar worden gesteld.
- Streef ernaar dat de verzamelde data volgens de FAIR-principes (Findable, Accessible, Interoperable and Reusable) beter toegankelijk en herbruikbaar worden gemaakt voor onderzoek. Met name de beperkte interoperabiliteit van de publieke data is een belemmering voor hergebruik. Het beter toegankelijk en herbruikbaar maken van publieke data voor onderzoek dient bij voorkeur plaats te vinden in landelijk dekkende dataknooppunten, die de verstrekkers en aanvragers van publieke data faciliteren (KNAW, 2018b). Ook hierover dienen interdepartementaal afspraken te worden gemaakt, bijvoorbeeld via de Stuurgroep Data en Overheid.
- Stel meer datastewards aan die publieke data beter toegankelijk en herbruikbaar maken voor onderzoek. Naast expertise op het gebied van databaseer moet de datasteward gedegen kennis hebben van onder meer privacyvraagstukken en het auteursrecht.
- Wees transparant over het bekostigingsmodel voor het verstrekken van publieke data voor onderzoek en breng niet meer kosten in rekening dan de marginale kosten van de dataverstrekking zelf. Het is voor onderzoekers nu vaak onduidelijk welke bedragen er moeten worden betaald, omdat overheidsorganisaties diverse bekostigingsmodellen hanteren.

VOOR HET CBS:

Faciliteer zo goed mogelijk de toegankelijkheid en herbruikbaarheid van data voor onderzoek. Benut daarbij uw positie als poortwachter die toegang heeft tot alle data van overheidsorganisaties voor zover die nodig zijn voor uw statistische taak.

VOOR HET MINISTERIE VAN EZK EN HET CBS:

Neem het belang van de beschikbaarheid voor onderzoek expliciet mee bij het aanscherpen van het juridische kader om private partijen met een wettelijke/publieke taak een basis te bieden voor het leveren van door hen verzamelde data aan het CBS. Dit juridisch kader is overigens ook van belang voor private partijen met een niet-wettelijke taak maar wel met een semipubliek karakter.

VOOR DE ONDERZOEKERS DIE VAN DATA UIT OVERHEIDSADMINISTRATIES GEBRUIKMAKEN:

Schroom niet om overheidsorganisaties te benaderen met een onderzoeksvraag, en verzoek die organisaties data uit hun administraties toegankelijk te maken. Uiteraard dient daarbij te worden voldaan aan de privacywetgeving en onthulling van individuele gegevens dient te worden voorkomen. Verwijs in de publicatie met de onderzoeksresultaten in ieder geval naar de *uniform resource identifier* (URI) van de databronnen, en geef na afloop van het onderzoek feedback aan de overheidsorganisatie over de onderzoeksresultaten, over de bijdrage aan de wetenschap, en over het belang van de verstrekte data voor de onderbouwing van overheidsbeleid. Deel deze informatie bij voorkeur ook met de CPDO.

LITERATUURLIJST

- Aerts, P.J.C. (2017). FAIR: Facts and Implementations
- Algemene Rekenkamer (2016). Trendrapport open data 2016
- Australian Government's Department of Communications and the Arts (2016). Open government data and why it matters. A critical review of studies on the economic impact of open government data
- AWTI (2016). Durven delen
- BBMRI-ERIC (2017). The EU General Data Protection Regulation. Answers to Frequently Asked Questions
- Biolink & BBMRI-NL (2015). Naar gekoppelde toetsing van koppilverzoeken
- Brief 28 maart 2017 van de secretarissen-generaal aan de kabinetsinformatieurbureau
- CBS (2018). Strategisch meerjarenprogramma 2019-2023
- CBS en Kadaster (2018). Eindrapport pilot open data stelsel
- Center for data innovation (2017). The state of data innovation in the EU
- Centraal Planbureau (2018). R&D meer datagedreven. Toegang data onderdeel innovatiebeleid
- Centre for international governance innovation (2018). Open data endgame. Countering the digital consensus
- Chief Science Advisor to the Prime Minister of New Zealand (2017). Enhancing evidence-informed policy making
- Computable (5 juni 2018). Data: het nieuwe goud en de toekomst is federatief
- Computable (5 juni 2018). Nederland moet een data-federatie oprichten
- CWTS, Elsevier, Universiteit Leiden (2017). Open data. The researcher perspective
- de Volkskrant (18 april 2018). Datahandel
- de Volkskrant (17 juli 2018). De moderne ambtenaar is een nerd
- EC (2017). Sustainable European research infrastructures
- EC (2018). On access to and preservation of scientific information
- Ecorys (2018). Internationale ranglijsten over open overheid
- E-data&research (juni 2018a). Veilig hoog-dimensionale en CBS-data combineren
- E-data&research (juni 2018b). Veilig opgroeien door machine learning
- EUA (2017). Recommendations on research data management and text & data mining
- European ICSU members (2018). Open data in science in Europe

European Open Science Cloud (2017). Uitkomsten van de ePLAN-workshop DIGITALIA op 22 december 2016 bij SURF in Utrecht. Idealen, opinies en ambities vanuit de Nederlandse wetenschap

Expertwerkgroep effectmeting (2012). Durf te meten

Frankowski, A., Van der Steen, M., Meijer, A., Van Twist, M. (2015). De publieke waarde(n) van open data

Furu, K., Wettermark, B., Andersen, M., Martikainen, J.E., Almarsdottir, A.B., Sørensen, H.T. (2009). The Nordic countries as cohort for pharmacoepidemiological research. *Basic & Clinical Pharmacology & Toxicology*, 106, 86-94

Her Majesty's Treasury, Verenigd Koninkrijk (2018). The economic value of data. Discussion paper

iBestuur (oktober 2017). Special Open data

Jong, I. de, Twist, M. van, Bressers, D., Schram, J. (2018). Gevoel voor getallen

KING (2017). Nulmeting open data gemeenten definitief

KNAW (2006, update 2009). De toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek

KNAW (2012). Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens

KNAW (2016a). Ethische en juridische aspecten van informatica-onderzoek

KNAW (2016b). Citizen Science. De betrokkenheid van burgers in het wetenschappelijk proces. Journalistiek verslag van het KNAW-symposium Citizen Science

KNAW (2018a). Replication studies. Improving reproducibility in the empirical sciences

KNAW (2018b). Big data in wetenschappelijk onderzoek met gegevens over personen

Kronenburg, T., Monasso, T., Boschker, E., Thaens, M. (2012). De waarde van open data. Keuzes en effecten van open-datastrategieën voor publieke organisaties

KVAB (2017). Datawetenschappen en gezondheidszorg

Loenen, B. van, Welle Donker, F.M., Ploeger, H.D. (2016). RIVM open data. Delft: Kenniscentrum open data. Faculteit Bouwkunde, TU Delft

Minister BZK (12 oktober 2015). Brief aan de Tweede Kamer (TK) over actieve beschikkingstelling overheidsinformatie

Minister BZK (30 november 2015). Brief aan TK over Nationale open data agenda (NODA)

Minister BZK (21 maart 2016). Reactie op derde trendrapport open data van de Algemene Rekenkamer

Minister BZK (22 juni 2016). Brief aan TK met stand van zaken uitvoering Nationale open data agenda (NODA)

Minister BZK (7 juni 2017). Brief aan TK over uitvoering open-databeleid

Minister BZK (13 juli 2018). Brief aan TK over Agenda Digitale Overheid NL DIGIbeter

Minister EZ (13 december 2012). Brief aan TK over eindrapport expertwerkgroep effectmeting

Minister OCW (28 juni 2016). Brief aan TK over big data in onderwijs en wetenschap

Minister OCW (19 januari 2017). Brief aan TK over voortgang open science

Minister OCW (21 februari 2017). Brief aan TK met aanbieding van nationaal plan open science

Minister OCW (30 november 2017). Brief aan TK over investeringen in wetenschap en onderzoek

Minister OCW (9 maart 2018). Brief aan TK met uitwerking investeringen wetenschap en onderzoek

Minister VWS (18 juli 2014). Brief aan TK over voortgang open data in het zorgdomein

Minister VWS (15 november 2018). Brief aan TK met aanbieding van de brief Data laten werken voor gezondheid

Ministerie BZK (2017). Rapportage tussentijdse zelfevaluatie Nationaal Actieplan 2016-2018

Ministerie BZK (3 september 2018). Vliegende start voor Rijks Data Science Programma (nieuwsbericht)

Ministerie EZK (2018). Nederlandse Digitaliseringsstrategie. Nederland Digitaal. Hier kan het. Hier gebeurt het

Ministerie VWS (2009). Reactie op RGO-advies Van gegevens verzekerd resp. biobanken

Mons, B. (2018). Data stewardship for open science. Implementing FAIR principles

Nationaal Plan Open Science (2017)

Nature genetics (editorial, 29 maart 2016). FAIR principles for data stewardship

Nederlands Dagblad (26 mei 2018). Naar hergebruik van onderzoeksgegevens

Nederlandse gedragscode wetenschappelijke integriteit (2018)

NESTA (2018). Reclaiming the city. Personal data, trust and the new commons

NFU&SURFsara (2016). Haalbaarheidsstudie nationale TTP-dienst

NRC Handelsblad (6 oktober 2017). Moeizame weg naar open overheid

NRC Handelsblad (1 december 2017). Hoe de overheid 100 miljoen uitgaaf aan drie ict-mislukkingen

NRC Handelsblad (3 februari 2018). Gevecht om waterdata

NRC Handelsblad (12 mei 2018). U gaat frauderen. Dat zegt de computer

NRC Handelsblad (6 september 2018). Pijnstillers diclofenac moet uit vrije verkoop wegens bijwerkingen

NRC Next (22 maart 2018). De computer berekent hoe ziek je wordt

NWA-routeworkshop Big data (2016). Toegankelijke en verantwoorde waardecreatie uit big data

NWO (2017). Topwetenschap vereist topinfrastructuur. Adviesrapport nationale digitale infrastructuur

OECD (2016). Research ethics and new forms of data for social and economic research

OECD (2018). Open Government Data Report. Enhancing policy maturity for sustainable impact

O'Neil, C. (2016). Weapons of math destruction

Open Data Institute (2018). Attitudes towards data sharing Europe

Open State Foundation (2016). De toekomst van hergebruik van overheids- en collectiedata

Overhedenoverleg van 14 februari 2018. Programmastart IBP. Samen meer bereiken als één overheid. Rijk, gemeenten, provincies en waterschappen starten met een interbestuurlijk programma en een gezamenlijke agenda

Palmgren, J. (ed., 2017). Nordic biobanks and registers—A basis for innovative research on health and welfare. Oslo:NordForsk, Policy Paper 2/2017

Regeerakkoord 2017-2021 (2017). Vertrouwen in de toekomst

RGO (2009). Van gegevens verzekerd

Roos, A.F. (2018). Mergers and competition in the Dutch healthcare sector

Ruppert, E. (2018). Sociotechnical imaginaries of different data futures. An experiment in citizen data

Science Business (10 oktober 2017). How data fuels Estonia's economy

Staatssecretaris EZK, Staatssecretaris BZK en Minister J&V (6 augustus 2018). Brief aan de Eerste Kamer over Nederlandse digitaliseringsstrategie

SURF (2010). Over kwaliteit van onderzoeksdata

SURF (2018). FAIR data – Advanced use cases

SURFsara (2016). Haalbaarheidsstudie Nationale TTP-dienst. Inventarisatie rapport

Tallinn declaration on eGovernment (6 oktober 2017)

The World Wide Web Foundation (2017). Open data barometer 4th edition. Global report

- Trouw (8 augustus 2018). Een dossier van en voor de patiënt
- Twist, M.J.W. van, Schulz, M., Steen, M. van der, Ferket, J. (2013). De deltacommissaris. Een kroniek van de instelling van een regeringscommissaris voor de Nederlandse delta
- UK Authority (22 augustus 2018). Treasury hints at retreat from open data
- VSNU (2017). Digital Society Research Agenda. Leading the way through cooperation in a Digital Society
- Welle Donker, F., Loenen, B. van, Korthals Altes, W. (2017). Maatschappelijke kosten-batenanalyse open data. Delft: OTB-Onderzoek voor de gebouwde omgeving. Faculteit Bouwkunde, TU Delft
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva, Santos L.B. da, Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J.N., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., Schaik, R. van, Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., Lei, J. van der, Mulligen, E. van, Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K.J., Zhao, J. (2016), The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data

BIJLAGEN

Bijlage 1. Instellingsbesluit van commissie ‘Toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek’

Het bestuur van de KNAW, overwegende dat de staatssecretaris van OCW bij brief van 6 september 2017 heeft verzocht advies uit te brengen over de wijze waarop de toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek verder kan worden verbeterd, besluit tot het instellen van de commissie ‘Toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek’, hierna te noemen ‘de commissie’.

Artikel 1. Taakopdracht

De commissie heeft tot taak de deelvragen te beantwoorden uit de brief van de staatssecretaris van OCW. De brief is als bijlage achtergevoegd.

Het advies wordt opgesteld in het Nederlands. De KNAW publiceert het advies als adviesrapport.

De commissie draagt zorg voor de aanbidding van het conceptadvies aan het bestuur vóór november 2018.

Artikel 2. Samenstelling

Tot commissielid worden op persoonlijke titel benoemd:

Voorzitter van de commissie:

- Prof. dr. Pearl Dykstra, Empirische sociologie, EUR

De overige leden van de commissie:

- Prof. dr. Dorret Boomsma, Gedragsgenetica, VU
- Prof. dr. Lude Franke, Systeemgenetica, UMCG
- Prof. dr. Wilco Hazeleger, Directeur eScience center / buitengewoon hoogleraar WUR
- Prof. dr. Pieter Hooimeijer, Human geography and demography, UU
- Prof. dr. Bert-Jaap Koops, Regulation and technology, TiU
- Prof. dr. Johan Mackenbach, Maatschappelijke gezondheidszorg, Erasmus MC

Als portefeuillehouder namens het bestuur treedt op prof. dr. Maarten Prak.

Vanuit het bureau van de KNAW zal dr. Ans Vollering de commissie ondersteunen.

Artikel 3. Integriteit en kwaliteit

De commissieleden hebben voorafgaand aan de eerste vergadering van de commissie kennisgenomen van de *Code ter voorkoming van oneigenlijke beïnvloeding door belangenverstrengeling* en dit in een schriftelijke verklaring bevestigd. De commissieleden hebben kennisgenomen van de *Handleiding adviezen KNAW*. Van het in deze handleiding beschreven beleid voor beoordeling van het conceptadvies wordt niet afgeweken.

Artikel 4. Werkplan

De commissie stelt een werkplan op met haar werkwijze en de communicatie- en implementatiestrategie. De commissievoorzitter bespreekt het werkplan met de voorzitter van de meest relevante adviesraad of adviesraden.

Artikel 5. Reiskostenvergoeding

De KNAW keert de commissieleden een reiskostenvergoeding uit, maar geen andere vergoedingen.

Artikel 6. Geheimhouding

De leden van de commissie nemen geheimhouding in acht ten aanzien van alle informatie die hun bij de uitvoering van dit besluit ter ore komt en waarvan vermoed kan worden dat deze vertrouwelijk is.

Aldus vastgesteld door het bestuur van de Koninklijke Nederlandse Akademie van Wetenschappen op 21 december 2017 te Amsterdam.

Namens het bestuur van de KNAW,
Mr. M. Zaanen
Algemeen directeur van de KNAW

Adviesaanvraag.



Ministerie van Onderwijs, Cultuur en
Wetenschap

>Retouradres Postbus 16375 2500 BJ Den Haag

Koninklijke Nederlandse Akademie van Wetenschappen
Mevrouw Prof. Dr. J.F.T.M. van Dijk
Postbus 19121
1011JV AMSTERDAM

Kennis
Rijnstraat 50
Den Haag
Postbus 16375
2500 BJ Den Haag
www.rijksoverheid.nl

Contactpersoon
Pim Kantebeen
T +31 6 31 74 92 15
p.kantebeen@minocw.nl

Onze referentie
1229421

Datum **06 SEP. 2017**

Betreft Advies over de toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek

Geachte mevrouw Van Dijk,

Het kabinet stelt zoveel mogelijk van de gegevens uit overheidsadministraties beschikbaar als open data. In onze brief 'Big Data in Onderwijs, Cultuur en Wetenschap' schreven de minister van OCW en ik aan de Tweede Kamer over het belang van open data voor datagedreven onderzoek.¹ In de wetenschap ontstaan steeds meer inzichten door de nieuwe mogelijkheden om data te verwerken en in combinatie met andere databestanden te analyseren. Voor de overheid betekent dit betere toegang tot gegevens over en bredere kennis van de Nederlandse samenleving. Dit kan het werk van beleidsmakers en besluitvormers in de publieke sector bij het aangaan van maatschappelijke uitdagingen ten goede komen.

De overheid stelt echter niet alle gegevens uit overheidsadministraties als open data beschikbaar. Dat geldt bijvoorbeeld voor informatie die de overheid in vertrouwen heeft verworven, zoals gegevens die betrekking hebben op natuurlijke personen of concurrentiegevoelige gegevens. Zulke gegevens kunnen alleen voor onderzoek worden verwerkt in veilige omgevingen, waaruit informatie die herleidbaar is tot individuen of afzonderlijke bedrijven niet naar buiten kan komen. Algemene inzichten of data die niet meer herleidbaar zijn tot individuele personen of bedrijven kunnen vanuit zo'n omgeving wél worden gedeeld.

Artikel 33 van de Wet op het Centraal bureau voor de statistiek biedt veel ruimte om publieke en semi-publieke gegevens voor statistische doeleinden te verwerven. Die gegevens kunnen dan beschikbaar worden gesteld aan geautoriseerde instellingen, waaronder instellingen voor wetenschappelijk onderzoek. Daardoor zijn die instellingen in de gelegenheid onderzoek te doen met de microdatabestanden van het Centraal Bureau voor de Statistiek (CBS). Het CBS ziet door middel van een veilige omgeving toe op de vertrouwelijkheid van informatie die herleidbaar is tot individuen of afzonderlijke bedrijven.

Als gevolg van een recente taakstelling op het CBS was het stopzetten van de verzameling en verwerking van een aantal (vanuit Europa niet-verplichte)

¹ Tweede Kamerstuk 31288-545 van 28 juni 2016

KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN	
No.	73299
Ingek.	08 SEP 2017
Actie:	Kopie naar:
B2(0)	DIR

statistieken helaas onvermijdelijk. Begin 2016 heeft u, samen met de VSNU en NWO, in een brief aan de minister van EZ gewezen op mogelijke consequenties voor het wetenschappelijk onderzoek dat gebruik maakt van deze statistieken. Het ministerie van OCW deelde destijds uw zorgen in de brief van juni 2016 die de SG van OCW aan de DG CBS stuurde. Inmiddels zijn de gesprekken hierover afgerond en zijn veel ministeries, waaronder het ministerie van OCW, bereid gebleken een groot deel van de niet-verplichte statistieken als maatwerk te continueren en te financieren.

Onze referentie
1229421

Het CBS werkt momenteel aan zijn Meerjarenprogramma (MJP) voor de statistische informatievoorziening, waarbij wensen van verschillende organisaties worden meegenomen. Ik heb vertrouwen in de kwaliteit en het proces waarmee het CBS het MJP tot stand brengt. Naast het produceren van statistieken, heeft het CBS inmiddels een belangrijke rol in het beschikbaar stellen van overheidsdata aan het Nederlandse onderzoek. Gezien het toenemende belang van data voor de wetenschap, blijf ik alert op goede toegang, kwaliteit en duurzame beschikbaarheid van overheidsdata voor onderzoek, nu en in de toekomst, via huidige structuren maar wellicht ook nieuwe.

Gegeven de geschetste ontwikkelingen en context vraag ik de KNAW bij deze om uiterlijk in december 2018 advies uit te brengen over de wijze waarop de toegankelijkheid van data uit publieke en semi-publieke administraties voor onderzoek verder kan worden verbeterd. In het bijzonder bestaat mijn adviesvraag uit de volgende deelvragen:

1. Hoe kunnen data uit publieke en semi-publieke administraties zo goed en zorgvuldig mogelijk toegankelijk worden gemaakt voor onderzoek? Hierbij gelden de uitzonderingsgronden en beperkingen uit artikel 10 van de Wet openbaarheid van bestuur alsmede de wettelijke voorwaarden gesteld voor de omgang met persoonsgegevens als randvoorwaardelijk. Het kan daarbij leerzaam zijn na te gaan op welke wijze data uit publieke en semi-publieke administraties toegankelijk worden gemaakt voor onderzoek in enkele andere landen die vanaf 25 mei 2018 onder dezelfde Europese privacywetgeving vallen. Zijn er internationaal vergelijkbare procedures of *best practices* waaraan Nederland een voorbeeld zou kunnen nemen?
2. In hoeverre volstaat het huidige aanbod van open data dat het kabinet ter beschikking stelt voor onderzoek? Indien dit niet volstaat, kunt u dan concrete voorbeelden noemen van data die open beschikbaar zouden moeten worden?
3. En in hoeverre volstaan voor onderzoek de huidige ontsluiting, kwaliteit en duurzame beschikbaarheid van data uit publieke en semi-publieke administraties die geen open data zijn? Indien die niet volstaan, kunt u dan concrete voorbeelden noemen van data waarvan de toegankelijkheid, kwaliteit en duurzame beschikbaarheid niet optimaal zijn?
4. Wat zou de regering aanvullend kunnen doen om data uit publieke en semi-publieke administraties toegankelijk te maken voor onderzoek?
5. Kunt u voorbeelden geven van beleidsterreinen waar kansen ontstaan om beleid te versterken met behulp van wetenschappelijk onderzoek, wanneer het advies dat u geeft in antwoord op bovenstaande deelvragen wordt opgevolgd?

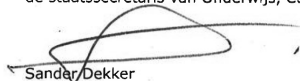
Omdat Nederland de ambitie heeft leidend te zijn in open science, verzoek ik u bij de beantwoording van mijn adviesvraag rekening te houden met het Nationaal Plan Open Science, in het bijzonder met de lopende afstemming binnen de werkgroep 'onderzoeksdata optimaal geschikt maken voor hergebruik'. Ik verzoek u in uw advies ook rekening te houden met de Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI, voorheen NDSW).

Onze referentie
1229421

Tenslotte verzoek ik u bij de beantwoording van de bovenstaande deelvragen te kijken naar het Nationaal Cohortonderzoek Onderwijs als *good practice* voor het hergebruik van data van de semi-overheid voor onderzoek en af te stemmen met belangrijke partijen die overheidsdata voor onderzoek ontsluiten en verrijken.

Met vriendelijke groet,

de staatssecretaris van Onderwijs, Cultuur en Wetenschap,



Sandra Dekker

Bijlage 2. Experts die door de commissie zijn geraadpleegd

Onderzoekers die data uit overheidsadministraties gebruiken:

Pieter Bakx, EUR

Peter van Bodegom, UL

Michiel Bots, UMCU

Henri de Groot, VU

Bastiaan van Loenen, TUD

Barend Mons, UL

Tom Postmes, RUG

Patricia Prüfer, TiU

Anne-Fleur Roos, EUR

Henk Scholten, Spatial Information laboratory (SPINlab) van VUA

Karien Stronks, AMC

Eddy Van Doorslaer, EUR

Rolf van der Velden, UM

Arnoud Verhoeff, GGD Amsterdam

Frederika Welle Donker, TUD

Experts die betrokken zijn bij de dataverstrekking aan onderzoekers:

Astrid Boeijen, CBS

Sylvia Commandeur, GGD GHOR

Peter Doorn, DANS

Ondine Engelse, Nederlands Centrum Jeugdgezondheid, NCJ

Khadija El Hamdaoui, Rijksdienst voor Identiteitsgegevens, RvIG

Inez Joung, RIVM

Ruurd Schoonhoven, CBS

Huib van de Stadt, CBS

Rob van de Velde, Geonovum

Ervaringsdeskundigen met data uit overheidsadministraties in België, de Scandinavische landen, en het Verenigd Koninkrijk:

Ron Dekker, CESSDA

Tom Emery, NIDI

Maarten van Ham, TUD

Juho Härkönen, European University Institute, Florence

Øystein Kravdal, Norwegian Institute of Public Health, Oslo

Philip Van Avermaet, Vlaamse overheid

Geert Van Grootel, Vlaamse overheid

Overige experts:

Mark Bressers, ministerie van EZK

Oscar Delnooz, voorheen ministerie van EZK

Arjan El Fassed, voorheen Open State Foundation
Erik Fledderus, SURF
Jaron Haas, Ministerie van BZK
Nora El Haider, Open State Foundation
Geertjan Holtrop, ministerie van EZK
Tom Kunzler, Open State Foundation
Anja Lelieveld, Ministerie van BZK
Karel Luyben, nationaal coördinator open science
Bram de Rijk, Ministerie van BZK
Lucas van der Meer, ODISSEI
Rudi Turksema, Algemene Rekenkamer
Arre Zuurmond, Ombudsman Metropool Amsterdam

Bijlage 3. Reviewprocedure

Een conceptversie van dit rapport is conform de Handleiding Adviezen KNAW (2017) beoordeeld door:

- Koen Becking, voorzitter CvB, TiU.
- Chantal Kemner, hoogleraar Biologische ontwikkelingspsychologie, UU.
- Brenda Penninx, hoogleraar Psychiatrische epidemiologie, VU Medisch Centrum.
- Yvonne van Rooy, voorzitter Nederlandse Vereniging van Ziekenhuizen (NVZ).
- Mark van Twist, hoogleraar Bestuurskunde, in het bijzonder bestuurs- en beleidsadvisering in publiek-private context, EUR.

Daarnaast is die versie becommentarieerd vanuit de Raad voor Geesteswetenschappen, de Raad voor Medische Wetenschappen, de Raad voor Natuur- en Technische Wetenschappen, en de Sociaal-Wetenschappelijke Raad.

De KNAW is de beoordelaars en de leden van de adviesraden veel dank verschuldigd. Zij dragen geen verantwoordelijkheid voor de inhoud van het rapport.

