

Vergaderjaar 2015–2016

31 288

Hoger Onderwijs-, Onderzoek- en Wetenschapsbeleid

Nr. 545

BRIEF VAN DE MINISTER EN STAATSSECRETARIS VAN ONDERWIJS, CULTUUR EN WETENSCHAP

Aan de Voorzitter van de Tweede Kamer der Staten-Generaal

Den Haag, 28 juni 2016

Big Data, het Internet der Dingen, 3D-printen en robotisering: digitalisering zet razendsnel door. Iedere nieuwe golf absorbeert hierbij het water van de vorige. Voor een groot deel zijn dit autonome ontwikkelingen die zich voltrekken buiten de invloedssfeer van OCW, maar deze digitalisering heeft ook impact op onze werkterreinen. Zij biedt kansen, maar er kunnen soms ook risico's aan verbonden zijn. Daarom volgen wij dergelijke ontwikkelingen op de voet en kijken wij naar de betekenis ervan voor onderwijs, cultuur en wetenschap.

Op 5 juni 2015 zonden wij uw Kamer een brief en het rapport «Big Data in onderwijs en wetenschap». Dit rapport was een eerste verkenning van het gebruik van Big Data. Het leverde ons daarnaast ook een eerste beeld op van de impact van Big Data en de vraagstukken voor de beleidsterreinen van OCW door middel van een viertal essayistische verkenningen: twee utopieën en twee dystopieën.¹

Inmiddels hebben wij het begrip Big Data nader onderzocht. We zijn het gesprek aangegaan met experts en onze bestuurlijke partners over wat de betekenis van Big Data is voor onderwijs, cultuur en wetenschap. We hebben daarbij vijf thema's geïdentificeerd die voor OCW relevant zijn: *infrastructuur op orde, voldoende deskundigheid, verbreden open access en open data, datagedreven onderwijsonderzoek en ethiek*. Met deze brief gaan wij nader in op de betekenis van Big Data voor onderwijs, cultuur en wetenschap, en specifiek op deze vijf thema's. Daarmee doen wij onze toezegging gestand om er nader op terug te komen.

¹ Kamerstuk 31 288, nr. 472.

Betekenis

Er is geen eenduidige definitie van het begrip Big Data. In plaats daarvan zijn er meerdere definities in omloop die verschillende aspecten van deze ontwikkeling benadrukken. De Wetenschappelijke Raad voor het Regeringsbeleid (WRR) structureerde deze definities onlangs door een onderscheid te maken tussen definities die betrekking hebben op de data zelf, op de analyse van deze data en op het gebruik ervan:²

Data	Analyse	Gebruik
<ul style="list-style-type: none">• Volume: Het gaat om grote hoeveelheden data• Variatie: Het gaat om ongelijksoortige data (gestructureerd en «ongestructureerd»)• Velocity: Het gaat om data die met grote snelheid wordt ingewonnen	<ul style="list-style-type: none">– Datagedreven: Dus op zoek naar patronen voorafgaand aan het opstellen van hypothesen– Vaak voorspellend op basis van real time data	<ul style="list-style-type: none">– Ontschotten: Data uit het ene domein toepassen in het andere domein– Beslisinformatie: Toepassen kennis op geaggregeerd niveau op beslissingen nu in concrete gevallen (individuen, objecten), vaak in combinatie met realtime voorspellen.

Maar wat betekenen Big Data en datagedreven onderzoek voor onderwijs, cultuur en wetenschap? In de wetenschap is het nut van datagedreven onderzoek het meest evident met voorbeelden als de deeltjesversneller Large Hadron Collider en de radiotelescoop Square Kilometre Array, die gigantische hoeveelheden data genereren. Of denk aan het Human Brain Project waarbij het menselijk brein door middel van Big Data wordt gesimuleerd. Wetenschappelijke en technische doorbraken leiden tot betere sensoren die fijnmaziger meten of dieper kijken. Sensoren worden bovendien steeds sneller en goedkoper. Nog nooit heeft de wetenschap zoveel data verzameld als nu, en ook de hoeveelheid onderzoeksdata die wordt ingewonnen, opgeslagen, verwerkt en geanalyseerd groeit exponentieel. De Nationale Wetenschapsagenda onderscheidt daarom expliciet een Big Data route.³

Bèta-wetenschappen als astronomie en meteorologie maken al decennia lang gebruik van grote databestanden, en hier ontstaat een stroomversnelling. Hierbij gaat het meestal niet om persoonsgegevens. Door de groeiende beschikbaarheid van data gaan ook sociale wetenschappers meer data gebruiken. Steeds meer historische bronnen komen digitaal beschikbaar, mensen laten steeds meer digitale broodkruimels achter op sociale media, mensen dragen steeds meer sensoren bij zich, menselijk gedrag wordt geregistreerd in belcentra en in de logs van webserver, overheidsinformatie komt steeds vaker beschikbaar als open data en in virtuele omgevingen als online games wordt menselijk gedrag gesimuleerd. De sociale wetenschappen krijgen daarmee de beschikking over een steeds rijkere variatie aan data voor datagedreven onderzoek.

In onderwijsinstellingen worden gegevens over het onderwijsproces en de onderwijsorganisatie in toenemende mate opgeslagen en geanalyseerd. Deze instellingen registreren steeds meer gegevens in administraties en beschikken ook vaker over analytische computerprogramma's voor gepersonaliseerd leren. «Learning analytics» is hiervan een voorbeeld waarbij leerstof op maat wordt aangeboden doordat digitale leermiddelen herkennen waar een leerling in zijn ontwikkeling staat en

² Wetenschappelijke Raad voor het Regeringsbeleid. 2016. Big Data in een vrije en veilige samenleving. WRR-Rapport 95.

³ Kamerstuk 29 338, nr. 149

waarbij een leerling gerichte feedback krijgt op basis van data uit het onderwijsproces. Behalve voor het monitoren van de ontwikkeling van leerlingen en studenten gebruiken onderwijsinstellingen data bijvoorbeeld ook voor het optimaal ondersteunen van medewerkers bij hun functies, voor een efficiënte bedrijfsvoering, voor een heldere verantwoording en voor effectieve samenwerking met partijen buiten de school.

Ook informatie uit de culturele sector komt steeds meer digitaal beschikbaar. Zo werken diverse sectoren uit de erfgoedwereld samen in het Netwerk Digitaal Erfgoed om dit erfgoed voor het publiek digitaal toegankelijk te maken. Een belangrijke activiteit in dit netwerk is het samenbrengen van data uit verschillende erfgoedcollecties.

Vijf thema's

De ontwikkelingen rondom Big Data gaan snel, en nieuwe inzichten ontstaan voortdurend. Om goed en verantwoord met Big Data om te kunnen gaan, is het belangrijk te zorgen dat er voldoende ruimte is om het thema verder te kunnen verkennen. Tegelijkertijd is het cruciaal om hierbij publieke waarden als privacy te borgen.

We zijn al langer met onze bestuurlijke partners in gesprek om in het openbaar bestuur de datahuishouding op orde te brengen. We maken afspraken over open standaarden en informatiearchitecturen, we stellen data-catalogi op en we maken afspraken over werkprocessen. Met deze partijen zijn we het gesprek gestart over welke randvoorwaarden op orde moeten zijn om te zorgen voor goed en verantwoord gebruik van Big Data. Uiteindelijk heeft dit geleid tot het identificeren van de volgende vijf thema's:

1. Infrastructuur op orde

Datagedreven onderzoek voor wetenschap en onderwijs stelt hoge eisen aan hard- en software en aan data-infrastructuren. Het gaat in tijden van Big Data immers niet alleen om grote volumes aan data, maar ook om data die met grote snelheid worden ingewonnen, die steeds complexer worden en die vanuit verschillende bronnen en van vele locaties komen.

Daarnaast worden data niet altijd meer opgeslagen in een vooraf bepaalde structuur, maar steeds vaker worden ze zo ruw (oorspronkelijk) mogelijk opgeslagen, waarbij het structureren onderdeel is van de analyses die ermee worden uitgevoerd. Ook deze werkwijze leidt tot nieuwe eisen aan de omvang en functionaliteit van de e-Infrastructuur. Een groot deel van de data, met name in de geesteswetenschappen en Life Sciences, is herleidbaar tot personen en daarmee privacygevoelig. Toegang tot dergelijke data dient goed te worden beschermd.

Grootschalige wetenschappelijke instrumenten als de Square Kilometre Array en de Large Hadron Collider genereren zulke grote hoeveelheden data dat de juiste middelen nodig zijn om hiermee om te gaan. Data moeten over datacenters in de hele wereld worden verspreid en door wetenschappers uit de hele wereld worden geanalyseerd. Ook in andere wetenschapsgebieden spelen vergelijkbare trends.

Wetenschappelijke onderzoekers moeten kunnen beschikken over een hoogwaardige e-infrastructuur die de data kan transporteren, analyseren, bewerken, opslaan en duurzaam archiveren, en die eenvoudige toegang mogelijk maakt. Diezelfde databenadering wordt steeds vaker ingezet voor het onderwijs.

Het kabinet erkent dat deze ontwikkelingen hoge eisen stellen aan de e-Infrastructuur. Daarom stelt OCW sinds 2014 structureel € 18,5 miljoen aan SURF ter beschikking voor de innovatie van de e-Infrastructuur. In aanvulling daarop heeft EZ voor de periode 2011–2019 € 19 miljoen beschikbaar gesteld en heeft OCW € 6 miljoen aanvullend beschikbaar gesteld voor de periode 2016–2019.⁴ Daarnaast heeft SURF een voorstel ingediend in het Toekomstfonds en ruim € 11 miljoen gekregen voor de e-Infrastructuur van het wetenschappelijk onderzoek en onderwijs.

Mede door de infrastructuur van SURF heeft de Nederlandse wetenschap een goede uitgangspositie om al die data te transporteren, analyseren en bewerken, delen en op te slaan. Het gaat hierbij om een flexibele e-infrastructuur, bestaande uit high performance computing, geavanceerde netwerken en samenwerkingsinfrastructuren. Deze geavanceerde e-Infrastructuur biedt het Nederlandse wetenschap en onderwijs de kans om baanbrekend te blijven met datagedreven onderzoek, en vormt een goed uitgangspunt om bij te dragen aan de European Open Science Cloud, die door de Europese Commissie op 19 april 2016 is aangekondigd, en die zal bestaan uit een virtuele bundeling van bestaande wetenschappelijke datafaciliteiten uit verschillende wetenschappelijke disciplines en EU-lidstaten.⁵

Het is van belang dat deze e-infrastructuur ook in de toekomst van hoge kwaliteit blijft. Toekomstige investeringen in de e-Infrastructuur en investeringen in grootschalige wetenschappelijke onderzoeksfaciliteiten via de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) of door universiteiten vragen om regie en afstemming. Zoals aangegeven in de kabinetsreactie op het AWTI-rapport «Klaar voor de toekomst» heeft NWO binnen de permanente commissie voor Grootschalige Onderzoeksfaciliteiten een werkgroep ICT opgericht om zorg te dragen voor deze afstemming.⁶ Deze werkgroep ICT is begin 2016 begonnen met een inventarisatie van de behoeften van gebruikers van de e-Infrastructuur. Deze werkgroep ICT bouwt voort op het werk van ePlan dat in 2016 een enquête onder meer dan 1.000 onderzoekers uitvoerde en waar de behoefte aan een goede e-Infrastructuur uit naar voren kwam.⁷ Op basis van de inventarisatie van de werkgroep ICT kan worden bezien of er voldoende middelen beschikbaar zijn om ervoor te zorgen dat de e-Infrastructuur tot de wereldtop blijft behoren.

2. Voldoende deskundigheid

Een tweede randvoorwaarde om de belofte van Big Data te verzilveren ligt in voldoende deskundigheid. Het toenemende belang van data stelt eisen aan professionals die hier in hun werk mee te maken krijgen. Voor een deel gaat dit om algemene digitale vaardigheden. Het integreren van digitale vaardigheden in het curriculum van primair en voortgezet onderwijs is onderdeel van de discussie met de noemer Onderwijs2032. In het MBO en het HO maken digitale vaardigheden een steeds groter deel uit van het curriculum. Hierbij wordt ook samengewerkt tussen bedrijven en opleidingen, zoals bij het ROC Amsterdam waar met steun van het Regionaal investeringsfonds MBO gewerkt wordt om ICT-opleidingen te laten aansluiten op de snel veranderende ICT-praktijk. Het expertise-

⁴ Kamerstuk 31 288, nr. 431

⁵ <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

⁶ Kamerstuk 33 009, nr. 13

⁷ ePlan. 2016. Infrastructuur duurzaam op maat. Een analyse van de vraag naar ICT-infrastructuur en services in Nederland. Via https://www.esciencecenter.nl/pdf/Infrastructuur_duurzaam_op_maat_versie_1.0B_DEF_03-03-2016.pdf

centrum Mediawijzer.net stelt lesmateriaal beschikbaar over media-wijsheid, programmeren en andere 21^{ste}-eeuwse vaardigheden.⁸

Specifiek gaat het bij Big Data ook om de nieuwe wetenschappelijke discipline «data science» en ontstaat er behoefte aan «data scientists». Een data scientist is iemand die niet alleen data duurzaam en toegankelijk weet op te slaan, maar ook data-analyses kan uitvoeren, bevindingen in context kan plaatsen en deze kan communiceren naar andere wetenschappers en besluitvormers. In de wetenschap is het dus een multidisciplinair geschoolde persoon die de brug kan bouwen tussen enerzijds de wereld van Big Data, en anderzijds de wetenschappelijke methode. In andere sectoren verbindt de data scientist de wereld van Big Data met die van business analytics of learning analytics.

Data science staat volop in de belangstelling en er ontstaan nieuwe opleidingen. Studies met een sterke analytische component, zoals econometrie en bèta-opleidingen, zijn op dit moment in staat om data scientists op te leiden. Eind 2015 waren er in ieder geval 13 van zulke opleidingen. Sindsdien hebben de Technische Universiteit Eindhoven en Tilburg University in Den Bosch de Jheronimus Academy of Data Science (JADS) gesticht die nieuwe opleidingen aanbiedt specifiek gericht op data science, en hierbij wordt ook samenwerkt met overheden en het bedrijfsleven.⁹ ¹⁰ Ook andere universiteiten gaan onderling samenwerkingsverbanden aan. De EU financiert een project genaamd «EDISON» onder leiding van de Universiteit van Amsterdam dat bijdraagt aan de definiëring van het beroep van data scientist door het maken van afspraken over curricula tussen Europese universiteiten, en brengt dit onder de aandacht van werkgevers.¹¹ Dergelijke curricula zijn ook interessant voor partijen die andere wetenschappers omscholen tot data scientist.

Ons blijven tegelijkertijd berichten bereiken over een tekort aan data scientists. Specifiek voor Big Data becijferde de Nationale Denktank bijvoorbeeld een specifiek tekort aan 8.000 data scientists in 2018 bij een verwachte aanwas van 2.000 data scientists.¹² Het World Economic Forum noemde het gebrek aan data scientists één van de belangrijkste obstakels voor bedrijven die Big Data willen toepassen.¹³ Ook de OECD adviseerde landen om meer data scientists te trainen.¹⁴

Binnen het Topsectorenbeleid van de Minister van Economische Zaken is een TeamICT ingesteld. Dit heeft een Human Capital Agenda ICT opgesteld. Doel hiervan is om het aanbod aan goed gekwalificeerde ICT-professionals, waaronder data scientists, te vergroten. In het hoger beroepsonderwijs starten in september 2016 experimenten met vraagfinanciering voor deeltijdopleidingen, ook voor de ICT-sector waar immers een grote behoefte is aan omscholing van werknemers. Volwassenen kunnen met vouchers terecht bij publieke of private aanbieders van opleidingen en er is meer ruimte voor de hogescholen om maatwerk leveren, passend bij de behoefte van werkenden.

⁸ <https://www.mediawijzheid.nl/lesmateriaal/>

⁹ Omroep Brabant. 2015. 40 miljoen voor «big data-universiteit» in Den Bosch, <http://www.omroepbrabant.nl/?news/235384992/40+miljoen+voor+big+data-universiteit+in+Den+Bosch+>.aspx

¹⁰ <http://www.jads.nl/>

¹¹ <http://edison-project.eu/>

¹² Nationale DenkTank 2014. Samenvatting Analysefase, www.nationaledenktank.nl.

¹³ World Economic Forum. 2014. The global information technology report 2014. Rewards and risks of Big Data.

¹⁴ OECD. 2015. Data-Driven Innovation. Big Data for Growth and Well-Being.

Deskundigheid gaat ook over bewustwording over de multidisciplinaire rol van data science. Veel van de bovengenoemde opleidingen hebben een monodisciplinair karakter. Wij kijken daarom met interesse naar samenwerkingsverbanden tussen onderwijs- en onderzoeksinstellingen als JADS, het Centre for Bold Cities en Amsterdam Data Science, en naar de curriculumontwikkeling vanuit het EDISON-project.^{15 16} De komende tijd blijven we de ontwikkelingen in het hoger onderwijs op dit terrein volgen en we blijven over dit onderwerp in gesprek met VSNU, Vereniging Hogescholen, SURF, het Netherlands eScience Centre (NLeSC) en NWO. Indien nodig zullen we verdere samenwerking faciliteren zodat instellingen gezamenlijk in staat zijn adequaat op nieuwe ontwikkelingen te reageren.

3. Verbreden open access en open data

Een derde randvoorwaarde voor de goede benutting van Big Data ligt in de beschikbaarheid van data. Voor de publiekgefinancierde wetenschap is open science steeds meer de norm. Hiermee is het besef gegroeid dat kennis alleen waardevol is wanneer het gedeeld wordt. Data zijn de grondstof van informatie, en informatie is de grondstof van kennis. Naarmate meer data worden gedeeld en gekoppeld kunnen bovendien meer verbanden worden gezien. Dat leidt tot waardevolle inzichten. Het doen van datagedreven onderzoek wordt laagdrempeliger naarmate meer bestaande data worden gedeeld.

Met Open Science als prioriteit van het Nederlandse EU-voorzitterschap organiseerde Nederland de conferentie «Open Science – From Vision to Action» op 4 en 5 april jl. in Amsterdam. De deelnemers, die afkomstig waren uit de onderzoekswereld, stelden daar de «Amsterdam Call for Action on Open Science» op, met daarin initiatieven voor de onderzoekswereld zelf en voor overheden. Deze oproep benadrukt onder andere het belang van vindbare, toegankelijke, interoperabele en herbruikbare onderzoeksdata, criteria bij het financieren van onderzoek en de opzet van de benodigde e-Infrastructuur.¹⁷ Mede in reactie op deze duidelijke oproep hebben de EU-onderzoeksm Ministers in de bijeenkomst van de Raad voor Concurrentievermogen van 27 mei jl. ambitieuze doelen gesteld en besluiten genomen over concrete maatregelen. Zij kwamen overeen dat data afkomstig uit publiek-gefinancierd onderzoek een publiek goed zijn en zo open mogelijk moeten zijn: «Zo open als mogelijk, zo gesloten als nodig». Concreet hebben zij besloten dat vanaf 2020 alle wetenschappelijke publicaties via open access beschikbaar moeten zijn. Een reeks maatregelen zal dit bespoedigen. Voor onderzoeksdata wordt in het Europese onderzoeks- en innovatieprogramma Horizon 2020 openheid de norm, tenzij er gegronde redenen zijn om hiervan af te wijken. Dan gaat het om intellectueel eigendom, veiligheid of privacy. Horizon 2020 maakt Data Management Plans een verplicht onderdeel van de procedure om voor onderzoeksfinanciering in aanmerking te komen, en lidstaten worden aangemoedigd om dit ook te laten gelden voor nationale onderzoeksprogramma's. Ook benadrukten de Ministers het belang van een goede infrastructuur voor vindbare, toegankelijke, uitwisselbare en herbruikbare data en riepen zij de Europese Commissie en lidstaten op hiermee aan de slag te gaan. Tenslotte hebben zij afgesproken in de komende jaren gezamenlijk met de Europese Commissie en met de onderzoekswereld op

¹⁵ <http://www.leiden-delft-erasmus.nl/nl/centres/centre-for-bold-cities>

¹⁶ <http://amsterdamdatascience.nl/>

¹⁷ Nederlands EU-voorzitterschap 2016. 2016. Amsterdam Call for Action on Open Science, <http://www.eu2016.nl/documenten/rapporten/2016/04/04/amsterdam-call-for-action-on-open-science>

te trekken om Open Science te realiseren, en daarbij op gezette tijden gezamenlijk de voortgang te monitoren.¹⁸

Dat het delen van data en informatie leidt tot nieuwe inzichten, geldt niet alleen voor data die in het wetenschappelijke proces zijn verzameld. Datagedreven onderzoek kan ook gebruik maken van data uit overheidsadministraties of uit andere bronnen. Het beschikbaar stellen van publiekgefinancierde data als open data draagt hiertoe bij. Voor de data van OCW geldt het uitgangspunt van de Wet openbaarheid van bestuur (Wob) «*alles is openbaar, tenzij*» waarbij het beschikbaar stellen voor hergebruik achterwege kan blijven als een weigeringsgrond in de Wob of de Wet hergebruik van overheidsinformatie van toepassing is. Ten gevolge van dit beleid voor meer open data worden bijvoorbeeld steeds meer data van DUO openbaar gemaakt via data.duo.nl. Openbaar betekent: toegang voor de één is ook toegang voor ieder ander, zonder technische of juridische beperkingen. Met dit beleid voorkomen wij dat er rondom publiekgefinancierde data mogelijk zogenoemde «datakartels» ontstaan: afspraken over het delen van data waarbij de toegang tot die gedeelde data beperkt wordt tot de deelnemers van zo'n datakartel.

Maar er zijn grenzen. Data die herleidbaar zijn tot natuurlijke personen worden niet openbaar gemaakt. Voor het onderwijs betekent dit dat open data beperkt blijven tot gemiddelden of totalen. Zo worden de gemiddelde eindtoetsscores alleen per school gepubliceerd wanneer het om minimaal vijf leerlingen gaat. Ook waken we ervoor dat we geen data openbaar maken als een derde partij beschikt over de intellectuele eigendomsrechten op deze data.

Binnen deze grenzen blijven we onderzoeken of we actief meer data beschikbaar kunnen stellen. Ook in de cultuursector ontstaan veel publiekgefinancierde data. Zowel nationaal en Europees werken wij aan duurzame toegang tot deze data en de daarvoor benodigde infrastructuur zoals Europeana, een digitaal platform dat meer dan 53 miljoen kunstwerken, boeken, bladmuziek en historische stukken toegankelijk maakt. Wij zullen laten onderzoeken in hoeverre benaderingen als open access en open data een bijdrage kunnen leveren aan het verbreden van de toegang tot informatie uit de cultuursector.

4. Datagedreven onderwijsonderzoek

In het onderwijs spelen data een steeds grotere rol, zowel in beleid, management als in het onderwijsproces zelf. Ook het wetenschappelijke onderzoek dat onderwijs ondersteunt wordt steeds afhankelijker van data. Data leiden tot inzichten die de kwaliteit van het onderwijs helpen verbeteren.

Veel data in het onderwijs zijn persoonsgegevens. Uiteraard maakt OCW dergelijke privacygevoelige data niet openbaar, tenzij deze door middeling, aggregatie of andere technieken zo worden aangeboden dat deze niet meer herleidbaar zijn tot een individu. Daarmee komt de methode van open data niet in aanmerking als manier om dergelijke data te delen. We willen evenmin dat er een situatie ontstaat waarin commerciële partijen, die data hebben verworven in het kader van een opdracht van een publieke instelling, vervolgens «de grens opzoeken» en commercieel gaan handelen in deze data, ook al is dit ingegeven door het feit dat data zoveel kunnen bijdragen. Behalve dat dit risico's met zich meebrengt voor de privacy van kinderen en studenten, zijn data dan niet meer

¹⁸ Raad van de Europese Unie. 2016. Outcome of Proceedings. 9526/16 RECH 208 TELECOM 100. <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>

publiekelijk beschikbaar voor onderzoek. Het is daarom van belang om het hergebruik van dergelijke data voor wetenschappelijk onderzoek binnen de grenzen van de Wet bescherming persoonsgegevens (Wbp) te organiseren.

Op dit moment gebeurt het verzamelen van data vaak per onderzoeksproject of door iedere afzonderlijke onderzoeksorganisatie. Vanuit de wetenschap wordt «overbegrazing» gesignaleerd: scholen worden zo vaak benaderd om mee te werken aan onderzoek dat de respons afneemt. Tegelijkertijd is geen enkele partij sterk genoeg om zelfstandig goed datagedreven onderzoek uit te voeren. Samenwerken, het combineren van data, infrastructuur en expertise, binnen de kaders van de Wbp, zijn sleutels tot succes. Wat privacy betreft zijn waarborgen zoals die zijn vastgelegd in de Wbp en de nieuwe Europese algemene verordening gegevensbescherming een belangrijke randvoorwaarde voor datagebruik in onderwijs, cultuur en wetenschap. Hoewel de technologie zich snel ontwikkelt, verwachten wij dat op OCW-terrein de huidige privacywetgeving in beginsel volstaat voor statistische of wetenschappelijke doeleinden.

Snelle ontwikkelingen maken het wel noodzakelijk dat deze kaders voortdurend geborgd worden, bijvoorbeeld omdat zij consequenties hebben voor het gemak waarmee verschillende datacollecties met elkaar verbonden kunnen worden. Dit zou ertoe kunnen leiden dat informatie alsnog tot de natuurlijke persoon herleidbaar kan worden. Wij houden de ontwikkelingen op dit vlak in de gaten en zullen waar nodig passende activiteiten ontplooiën.

Het CBS is een voorbeeld van een veilige omgeving waarbinnen meerdere onderzoekers eens ingewonnen data kunnen hergebruiken terwijl de privacy is geborgd. Uit deze omgeving kunnen geen data naar buiten gaan die herleidbaar zijn tot individuele personen of organisaties. En ook intern is het CBS zo georganiseerd dat voldaan wordt aan de vereisten van de Wbp.

Samen met de onderwijsraden, SURF, NRO en het CBS willen we de komende tijd op dit punt verkenningen uitvoeren. Gezamenlijk hebben we een expertgroep geformeerd waarbinnen we over Big Data spreken. Daarnaast spreekt OCW op bestuurlijk niveau met de onderwijsraden in een gremium genaamd Informatiekamer. Samen met deze partijen laten we een verkennend onderzoek uitvoeren naar wensen, mogelijkheden en beperkingen van een veilige omgeving waarbinnen Big Data kan worden gecombineerd (zoals het CBS dat nu doet voor gestructureerde data).

Vanuit onderzoek is er daarnaast belangstelling voor het gebruik van data uit leerlingadministratiesystemen die scholen via het CBS vrijwillig voor onderzoek beschikbaar stellen binnen de kaders van de Wbp als alternatief voor grootschalige enquêtes. Het gaat hierbij nog niet om Big Data, maar het leert ons mogelijk wel iets over deze manier van samenwerken binnen de verschillende onderwijssectoren. Het NRO is de goede plek om voorstellen hierover nader uit te werken.

5. Ethiek

Ethiek is nadenken over goed handelen en is daarmee een thema dat de vier andere thema's verbindt. Bij de ethische aspecten van Big Data bij OCW springen privacy en discriminatie in het oog. Uiteraard staan de bescherming van persoonsgegevens en het tegengaan van discriminatie voor ons voorop. Maar het gesprek over ethiek is ook een gesprek dat we met elkaar moeten voeren. Daarbij gaat het in eerste instantie over een

gesprek tussen gebruikers van Big Data, zoals bijvoorbeeld in onderwijsinstellingen. Wij gaan daar echter ook graag over in gesprek met uw Kamer, en natuurlijk met onder bestuurlijke partners.

De kaders ten aanzien van privacy voor wetenschappelijk onderzoek zijn helder en bieden houvast. Voor andere beleidsterreinen is dat mogelijk anders. Zo beraadt het kabinet zich op het rapport «Big Data in een vrije en veilige samenleving» van de Wetenschappelijke Raad voor het Regeringsbeleid.¹⁹ De Minister van Veiligheid en Justitie zal hier namens het Kabinet op terugkomen.

Wanneer in wetenschappelijk onderzoek patronen worden gevonden, kan kennis daaruit vervolgens worden toegepast in praktijksituaties. Ook dat kunnen situaties zijn waarin persoonsgegevens worden verwerkt. Wanneer dergelijke inzichten bijvoorbeeld worden toegepast in moderne leermiddelen die leraren helpen om beter te differentiëren tussen leerlingen, dan kan dit helpen om talenten op elk niveau maximaal ontplooiën.²⁰ Het gebruik van zulke digitale leermiddelen is een nieuwe verwerking van persoonsgegevens, uiteraard is ook hierop de privacywetgeving van toepassing.

Het toepassen van dergelijke kennis op individuen mag niet leiden tot discriminatie. Wanneer een patroon wordt gevonden dat voortijdig schoolverlaten of een lage eindscore voorspelt, dan is het zaak dat zulke kennis wordt gebruikt om leerlingen die dat nodig hebben extra begeleiding te bieden. Niet om leerlingen te weigeren aan de poort. En bij het toepassen van kennis op basis van Big Data, is het wenselijk dat er altijd goed wordt gekeken naar de persoon zelf.

Er zijn in theorie legio meer en minder extreme casussen denkbaar. Het is niet mogelijk deze van te voren allemaal uitvoerig te doordenken op elke mogelijke consequentie. Tegelijkertijd is niets doen geen optie, en alleen een «bedoeling» uitspreken is niet goed genoeg wanneer het om grondrechten als privacy of het discriminatieverbod gaat.

Met onze bestuurlijke partners is afgesproken om ethische vraagstukken die samenhangen met datagebruik vroegtijdig te bespreken wanneer die zich voordoen. Bij elke ontwikkelingsstap die we zetten, zoals gezamenlijke «hackatons», zullen we ethische aspecten expliciet in de evaluatie meenemen. Daarnaast blijven wij actief onderzoeken waar dilemma's ontstaan en waar afspraken met onze bestuurlijke partners of andere maatregelen nodig zijn. Om zulke dilemma's te doordenken blijven we in gesprek met wetenschappers, bestuurders en mensen in de praktijk. Zo maken we de ethische aspecten van Big Data bespreekbaar binnen de sectoren, en zo hopen we samen een kader te ontwikkelen dat door de betrokken partijen wordt gedragen en gebruikt zodat we de beloftes van Big Data verantwoord kunnen verzilveren. Dit kader zal bestaan uit afspraken die kunnen worden aangepast als de snelle ontwikkelingen rondom digitalisering daarom vragen.

Tot slot

Big Data vraagt om ruimte en vertrouwen zodat de kansen die deze ontwikkeling biedt gepakt kunnen worden. Dat kan alleen maar als ook de risico's ten aanzien van publieke waarden als privacy worden geborgd. Dat is een gemeenschappelijke opgave. Met de hierboven beschreven vijf

¹⁹ Wetenschappelijke Raad voor het Regeringsbeleid. 2016. Big Data in een vrije en veilige samenleving. WRR-Rapport 95.

²⁰ Kamerstuk 32 034, nr. 3.

thema's en onderliggende acties hebben wij er vertrouwen in dat we hiervoor een voldoende krachtig fundament leggen.

Ook in de nabije toekomst zullen ontwikkelingen rondom Big Data en digitalisering snel blijven gaan. Om adequaat te reageren op de nieuwe kansen en risico's die ontstaan blijven wij in gesprek met belanghebbenden, experts en bestuurders. Samen blijven wij deze ontwikkelingen volgen en wanneer er aanleiding toe is zullen wij uw Kamer over relevante ontwikkelingen informeren.

De Minister van Onderwijs, Cultuur en Wetenschap,
M. Bussemaker

De Staatssecretaris van Onderwijs, Cultuur en Wetenschap,
S. Dekker