

## ARTIKELEN

# Legal big data en wet- en regelgeving: perspectieven en uitdagingen

F.L. Leeuw

### 1. Achtergrond en vraagstelling

De idee dat big data voor het recht en de rechtswetenschap belangrijk zijn, is al meermalen geadresseerd.<sup>1</sup> Toch is het niet a priori evident dat juristen, onder wie wet- en regelgevers, ‘iets’ met big data willen of kunnen. Zij staan daarin niet alleen; ook andere beroepsgroepen hebben vraagtekens bij deze ontwikkeling. ‘Het zijn niet door ons zelf verzamelde data’, ‘De verhouding tussen steekproef en populatie is ongewis’, ‘Wie zegt dat het mensen zijn die de data “maken” (en niet machines)?’, ‘Welke rol spelen internetbedrijven?’ en ‘Van wie zijn die data eigenlijk?’ zijn het soort kanttekeningen en vragen. Dat Susskind<sup>2</sup> over ‘disruptieve technologieën’ voor juristen rept, waaronder big data, is niet verwonderlijk.

Uiteraard waren er ook takken van sport die wel al snel in zee gingen met big data, zoals de marketingwereld en onderzoekers die maatschappelijke sentimenten meten en daarbij niet primair op enquêtes willen afgaan. Voor genetici en astronomen waren big data ook al langere tijd niet uitzonderlijk. Tegelijkertijd is de zorg over privacy, transparantie en eigendom van data door de jaren heen eerder groter dan kleiner geworden, dit ondanks de onlangs ingevoerde Algemene verordening gegevensbescherming (AVG). Breed uitgemeten affaires inzake datalekken en de Facebook/Cambridge Analytica-kwestie,<sup>3</sup> maar ook (ethische) discussies over wat – bijvoorbeeld – ‘predictive policing’ met behulp van big data over gezichtsherkenning vermag, spelen eveneens een rol.

De wereld van big data is dynamisch, snel en vol verwachtingen, maar ook omgeven met zorgpunten. In dit artikel benoem ik eerst enkele kenmerken van (legal) big data en ga daarna in op deze vier vragen:

- Wat betekent het werken met legal big data voor het maken, invoeren en evalueren van wet- en regelgeving?
- Zal de manier van wetgeven onder invloed van (legal) big data (*analytics*) (kunnen) veranderen en wat zou dat betekenen voor het werk van wetgevers?
- Welke uitdagingen zijn er voor de (nabije) toekomst?

1 B. Custers & F.L. Leeuw, ‘Legal big data. Toepassingen voor de rechtspraak en juridisch onderzoek’, *NJB* 2017/34, p. 2449-2456; C. Prins, ‘Big Data en de rechterlijke macht’, *NJB* 2015/30, p. 1538; R. Susskind, *Tomorrows Lawyers*, Oxford: Oxford University Press 2013.

2 Susskind 2013.

3 Zie [https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge\\_Analytica\\_data\\_scandal](https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal) (geraadpleegd op 16 oktober 2018).

F.L. Leeuw

- Zou er (meer/specifieke) wetgeving moeten komen met het oog op big data, de analyses en het gebruik?

De term 'big data' verwijst naar de enorme omvang van de hoeveelheden gegevens en daarbij passende 'analytics'.<sup>4</sup> Er is geen minimale hoeveelheid vereist om van big data te kunnen spreken, maar het is wel duidelijk dat het gaat om vele terabytes (TB), petabytes (PB) en exabytes (EB) met gegevens. Stephens e.a.<sup>5</sup> vergeleken vier big-dataterreinen en stelden de vraag wie over een paar jaar de 'grootste' zal zijn: astronomie (opslag 1 EB jaarlijks), Twitter (1-17 PB), YouTube (1-2 EB) of genomics (2-40 EB).

Naast aandacht voor deze (eerste) V ('volume') wordt over vier andere V's gesproken. V # 2 heet 'velocity' en betreft de snelheid waarmee data gegeneerd worden. Het viraal gaan van socialemediaberichten, het mondiaal realtime checken van creditcardbetalingen, het meten van vervoersstromen via smart apparaten en tal van andere voorbeelden laten zien waar het bij deze V om gaat. V # 3 heet 'variety' en betreft de diverse typen data die onderdeel zijn van big data. Documenten, audio, video, foto's, (f)mri-metingen, QS-gegevens,<sup>6</sup> met andere woorden gestructureerde en ongestructureerde data. V # 4 staat voor 'veracity': het verschijnsel dat er van alles en nog wat aan data verzameld wordt dan wel zich zelf verzamelt, maar er vragen te stellen zijn bij de betrouwbaarheid ervan. V # 5 heet 'value' en gaat om wat de waarde (vooral in economische zin) van big data is.

In een recent rapport van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW)<sup>7</sup> schrijft de commissie:

'belangrijkste onderscheid tussen "gewone" data en big data in hoeveelheden en grote verscheidenheid aan soorten (hoge dimensionaliteit) van gegevens, die op talloze terreinen kunnen worden vergaard, opgeslagen, gekoppeld en geanalyseerd. De hoeveelheden en hoge dimensionaliteit blijven explosief toenemen.'

- 4 KNAW, *Big data in wetenschappelijk onderzoek met gegevens over personen*, Amsterdam 2018. In dit rapport wordt het op p. 8 zo verwoord: 'Big data is als begrip moeilijk onduidelijk vast te leggen. Hoewel big data in vrijwel alle wetenschapsgebieden wordt gebruikt, is "big" relatief en afhankelijk van de discipline. In bijvoorbeeld de natuurkunde geldt vaak een andere opvatting over "veel" dan in de sociologie.'
- 5 Z. Stephens e.a., 'Big Data: Astronomical or Genomical?', *PLoS Biology* (13) 2015, e1002195, doi: 10.1371/journal.pbio.1002195.
- 6 QS staat voor Quantified Self, de zelfmetende mens, die dat doet met behulp van slimme apparaatjes en apps, waarbij vaak continu en realtime grote hoeveelheden gegevens opgeslagen worden (en ook weer door de zelfmetende mens en andere zelfmeters gebruikt worden). Zie L. Cornet e.a., *De 'zelfmetende' justitiabele. Een verkennend onderzoek naar technologische zelfmeetmethoden binnen justitiële context*, Den Haag: Boom juridisch 2018.
- 7 KNAW 2018, p. 13.

Ook voor legal big data geldt dat er geen definitie is. Goanta<sup>8</sup> formuleert het zo:

‘Taken in a more technical meaning, Big Data is simply a mass of data that is not structured like traditional datasets and needs more real-time analysis (...). *Such analysis can include coding the law or legal observations.* The benefits of investigating such data masses lie in the fact that new patterns can be drawn from the information and new lessons can be learned in terms of management and visualization. Perceiving the law as a source of information that can be systematically collected for further analysis can change the way in which the success or the failure of a certain policy is perceived and visualized.’ [cursivering: FL] (l.i., fl)

De aandacht voor cijfers en statistieken in de wereld van het recht is niet nieuw en kent een lange traditie (denk aan Quetelets studie van de relatie tussen leeftijd en criminaliteit; de jurimetrie, het werk van (Amerikaanse) ‘legal realists’ uit de vorige eeuw). Tegelijkertijd zijn er belangrijke verschillen tussen ‘toen’ en ‘nu’. Een wezenlijke is: digitalisering van de inhoud van juridische documenten (jurisprudentie, wetboeken, memories van toelichting bij wet- en regelgeving, contracten, testamenten, convenanten en mediation, maar ook rapporten van waarheidscommissies, toezicht- en handhavingsrapportages, statistieken, video- en audio-opnamen van rechtszaken, mailverkeer, verwijspatronen in zaaksbehandelingen enzovoort) maakt toepassing van analysetechnieken van kwantitatieve aard mogelijk. Bijvoorbeeld om verwijspatronen in de rechtspraak te analyseren, consistentie tussen uitspraken te meten, de toegankelijkheid van de gebruikte juridische taal te meten, aan ‘legal prediction’ te doen of, zoals in de Verenigde Staten, om de doelmatigheid van advocatenkantoren te onderzoeken. Methode- en analyseontwikkelingen zijn een tweede wezenlijk ander punt van verschil tussen ‘toen’ en ‘nu’.

## 2. Wat betekent het werken met legal big data voor wetgeving en regelgeving?

Het belang van legal big data voor wet- en regelgeving beschrijf ik in eerste instantie per (een wat gestileerde) fase in het wetgevingsproces (ontwikkelen, invoeren en evalueren).

### 2.1 Ontwikkelen van wet- en regelgeving

Een belangrijk item bij het denken over nieuwe wetgeving is welk (maatschappelijk) vraagstuk via wetgeving aan te pakken is. Het Integraal Afwegingskader voor beleid en regelgeving (IAK)<sup>9</sup> geeft aan dat het dienstig is probleemanalyses uit te voeren, krachtenveldanalyses te doen, oog te hebben voor gedragsaspecten van wat met de wet- en regelgeving beoogd wordt, zakelijke doelstellingen te formule-

8 C. Goanta, ‘Big Law, Big Data’, *Law and Method* (3) 2017, p. 5.

9 Zie [https://www.kcwj.nl/kennisbank/integraal-afwegingskader-beleid-en-regelgeving/\(geraadpleegd op 12 oktober 2018\)](https://www.kcwj.nl/kennisbank/integraal-afwegingskader-beleid-en-regelgeving/(geraadpleegd op 12 oktober 2018)).

F.L. Leeuw

ren en expliciet te beargumenteren waarom wet- en regelgeving aangewezen instrumenten zijn om het betreffende vraagstuk aan te pakken.

Daar kunnen (legal) big data bij helpen. Denk aan sentimentanalyses van wat in de sociale media en elders op internet te vinden is over de te ‘behandelen’ maatschappelijke problemen. In plaats van mensen en organisaties vragen te laten beantwoorden, kunnen inzichten met grotere snelheid en omvang via sentimentanalyses worden ingewonnen. Aan de hand van taalanalyse kan worden vastgesteld hoe er over een onderwerp van (toekomstige) wetgeving wordt geschreven. Sentimentanalyse omvat niet alleen maatschappelijke sentimenten, maar kan ook het ‘klantsentiment’ over producten, in ons geval relevante wetgeving(scomplexen), betreffen. Informatie uit diverse bronnen, zoals Twitter, online communityforums en e-mails, kan op die manier worden omgezet in voor probleemanalyses en krachtenveldbeelden bruikbare inzichten.<sup>10</sup>

Uiteraard zijn sentimenten slechts sentimenten, oftewel ‘attitudes’ en ‘percepties’. Hoe belangrijk ook, zij alleen vullen geen gaatjes. Daarom is het belangrijk dat wetgevers deze ‘sentimentanalyses’ aanvullen met wat begin jaren zeventig *Lexis legal research services* heette. Inmiddels is deze voorziening doorontwikkeld tot brede platforms.

Ten tweede en belangrijker is dat ook gebruik gemaakt wordt van ‘repositoria’ van resultaten uit evaluatieonderzoek naar wat werkt, wat niet, voor wie en voor wie niet. Begonnen in de medische wetenschappen (‘Cochrane Library’) bestaan er thans enigszins vergelijkbare kennisschatkammers op tal van andere maatschappelijke velden, zoals criminaliteit en veiligheid, onderwijs, ontwikkelingssamenwerking, voedsel, milieu, arbeid, economie enzovoort. Gebruikmaking van deze bronnen bevordert het *evidence based* worden van wetgeving (zie verderop).

Een andere met (legal) big data gepaard gaande manier van werken in de gevallen dat ‘klassieke’ statistieken en onderzoeken niet beschikbaar zijn om de ernst van een bepaald vraagstuk te kennen, is beschreven door Willemsen<sup>11</sup> (in 2018). Het ging om de vraag hoe goed zicht te krijgen op cybercriminaliteit en in het bijzonder ransomware. Hij werkt aan een manier om via Google-zoekgedrag tot een schatting te komen van het aantal keren dat ransomware-aanvallen zich voordoen.

Een derde bijdrage van legal big data betreft het langs deze weg beantwoorden van de vraag hoe toegankelijk, begrijpelijk en leesbaar juridische documenten eigenlijk zijn. In de Verenigde Staten deden Katz en Bommarito<sup>12</sup> dat voor de US Code. Zij ontwikkelden een meetmethode om de complexiteit ervan zo objectief mogelijk vast te stellen. Ze deden dit mede vanuit het gezichtspunt van de maat-

10 Rekening wordt gehouden met echokamers en machines die antwoord geven (in plaats van mensen).

11 Het gaat om de bijdrage van Frank Willemsen (WODC), waar in het WODC rapport ‘Tasten in het duister’ (2018) in hoofdstuk 3.3. iets over gezegd wordt.

12 D. Katz & M. Bommarito, *Measuring the Complexity of the Law: The United States Code*, 2013, <http://ssrn.com/abstract=2307352> (geraadpleegd op 4 juli 2015).

schappelijke relevantie: juridische teksten moeten zo toegankelijk mogelijk zijn. Dat dit ook geldt voor wetsteksten spreekt voor zich, alsook voor andere juridische documenten zoals hypotheekakten. Utrechtse onderzoekers<sup>13</sup> publiceerden daarover en ongeveer gelijktijdig riep de Hoge Raad voor de Justitie in België alle juridische professionals op om bewust om te gaan met hun taalgebruik voor de rechtszoekende.<sup>14</sup>

De bovenstaande ontwikkelingen zijn in belangrijke mate slechts mogelijk vanwege (data)technologie. Er is echter ook een meer *rechtstreekse invloed* op wetgeving mogelijk, waarbij de modus operandi van wetten *zelf* verandert. In de literatuur zijn daar voorbeelden van te vinden. Een daarvan is in box 1 aan te treffen.

### **Box 1. Voorbeeld van hoe onder invloed van technologie en big data de aard van wet- en regelgeving kán veranderen**

Het voorbeeld is afkomstig van Casey en Niblett.<sup>15</sup> Hun vertrekpunt is dat waar wetgevers thans (moeten) kiezen uit het werken met ‘regels’ (‘rules’) of ‘standaarden’ (‘standards’) om te realiseren dat doelen van wetten behaald worden, die manier van werken vervangen kan worden door wat zij ‘microdirectives’ noemen. ‘Rules’ zijn de voorschriften en aanwijzingen hoe een individu of organisatie moet opereren om compliant aan wetgeving te zijn. Of dat ook gebeurt, is niet zeker en vereist (kostbaar) toezicht op de naleving van gedrag van burgers en organisaties. Waar ‘rules’ een beoogde ex-antewerking hebben en vaak precies en afgebakend zijn, zijn ‘standards’ juist de uitspraken in wetgevingsteksten waarin termen zoals ‘redelijk’, ‘excessief’ en ‘materieel’ voorkomen. ‘Rules indicate to an individual whether certain behavior will violate or comply with the law. When a rule is enacted, effort must be undertaken by lawmakers to give full and precise content to the law before the individuals act. Standards, on the other hand, are imprecise when they are enacted.’ Casey en Niblett<sup>16</sup> zien voor de toekomst dat ‘microdirectives’ daarvoor deels in de plaats komen. ‘To see how the mechanism might work, consider the regulation of traffic speed. In a world of rules and standards, a legislature hoping to optimize safety and travel time could enact a rule (a sixty miles-per-hour speed limit) or a standard (“drive reasonably”). With microdirectives, however, the law looks quite different. The legislature merely states its goal. Machines then design the law as a vast catalog of context-specific rules to optimize that goal. From this catalog, a specific microdirective is selected and communicated to a particular driver (perhaps on a dashboard

13 Zie <https://www.uu.nl/nieuws/hoe-kunnen-we-onbegrijpelijke-hypotheekaktes-leesbaar-maken> (geraadpleegd op 15 oktober 2018).

14 Zie [www.csj.be/nl/content/persbericht-heel-justitie-moet-begrijpelijker-spreken-en-schrijven](http://www.csj.be/nl/content/persbericht-heel-justitie-moet-begrijpelijker-spreken-en-schrijven) (geraadpleegd op 16 oktober 2018).

15 A. Casey & A. Niblett, ‘The Death of Rules and Standards’, *Indiana Law Journal* (92) 2017, afl. 4, Article 3, p. 1401-1447.

16 Casey & Niblett 2017, p. 1404.

display) as a precise speed for the specific conditions [he or] she faces. For example, a microdirective might provide a speed limit of 51.2 miles per hour for a particular driver with twelve years of experience on a rainy Tuesday at 3:27 p.m. The legislation remains constant, but the microdirective updates as quickly as conditions change.'

De auteurs komen tot dit idee omdat datatechnologie én communicatietechnologie het werken met realtime, gepersonaliseerde directives steeds meer mogelijk maken, iets dat overigens in het verlengde ligt van de ontwikkeling van 'Quantified Self'(-metingen) en 'personalized medicine'. Cornet e.a.<sup>17</sup> publiceerden een uitgebreide studie over wat op dat gebied binnen het ministerie van Justitie en Veiligheid aan 'research & development' gebeurt.

## 2.2 Uitvoeren van wet- en regelgeving

In Nederland zijn diverse uitvoeringsdiensten actief die enorme aantallen 'mensen en zaken' behandelen en afhandelen, waarbij vigerende wetgeving een van de meest centrale kaders is. Dat gaat gepaard met uitvoeringsproblemen, waaronder vertraging, zoekraken van 'mensen' en 'zaken' en fouten. Big data in de vorm van 'Legal Logistics' kunnen helpen deze processen inzichtelijker en sneller te maken. Dit concept is nog niet erg wijdverbreid. Netten e.a.<sup>18</sup> duiden het als een framework waarin de voor het functioneren van het juridisch systeem relevante data (en daarmee gepaard gaande operaties) geïncorporeerd zijn om op die manier tot nuttige en bruikbare output en outcome-indicatoren voor het functioneren van dit systeem te komen. Doorlooptijden als indicator zijn lang niet altijd bekend of hooguit voor afzonderlijke fasen van en actoren uit bepaalde juridische processen (bijvoorbeeld zaaksbehandelingen in het strafrecht). Wie de cijfers over de aantallen zaken, mensen, diensten, processen, normen en feedbackmechanismen van het 'legal system' kent (vergelijk de overzichten uit publicaties als 'Criminaliteit & Rechtshandhaving' en 'Rechtspleging Civiel en Bestuur'), weet hoe omvangrijk die zijn. Niet voor niets spreken Ruhl e.a.<sup>19</sup> over 'harnessing legal complexity: bringing tools of complexity science to bear on improving law'.

Een perspectief voor de toekomst zijn ook zogenaamde experimenten om implementatieproblemen rondom wet- en regelgeving te monitoren en aan te pakken (vergelijk Jacobs<sup>20</sup>). Als invoering van nieuwe regelgeving in eerste instantie kleinschalig ('experimenteel') gaat gebeuren, kunnen via dit soort experimenten inzichten verkregen worden over hoe de samenleving én uitvoerders daarop reageren, welke beperkingen en mogelijkheden men ziet, en hoe gereageerd wordt als verschillende oplossingen voor implementatieproblemen worden aangedragen

17 Cornet e.a. 2018.

18 N. Netten e.a., 'Legal Logistics – A Framework to Unify Data Centric Services for Smart and Open Justice', *International Journal of E-Planning Research* (7) 2018, afl. 2, p. 51-69.

19 J.B. Ruhl, D.M. Katz & M.J. Bommarito, 'Harnessing Legal Complexity', *Science* (355) 2017/6332, p. 1377-1379.

20 M.J. Jacobs, *Experimentele wetgeving* (oratie Amsterdam VU), 2018.

door de uitvoerders van wet- en regelgeving. Voorbeelden van implementatieproblemen zijn onder andere het gebrek aan een effectieve ‘guidance’ bij invoering, ‘concurrentie’ tussen wie waarvoor verantwoordelijk is, ‘red tape’ en ambtelijke ongeïnteresseerdheid.

### 2.3 *Evalueren van wet- en regelgeving*

In Nederland bestaat een traditie van wetsevaluaties van al enkele decennia oud. In 2008 publiceerden Klein Haarhuis en Niemeijer<sup>21</sup> een meta-evaluatie en -analyse van tientallen wetsevaluaties. (Legal) big data kunnen behulpzaam zijn bij toekomstige evaluaties. Dat moet ook wel, omdat er steeds meer digitaal beleid en wetgeving komen die gedrag van actoren ‘in digitalis’ regarderen (en proberen te beïnvloeden). Zoals Jacobs<sup>22</sup> opmerkt: ‘Niet toevallig koos de Nederlandse Juristen-Vereniging in 2016 het thema “Homo Digitalis” uit voor de jaarlijks op te stellen preadviezen van dit genootschap.’

Wie digitaal beleid op zijn consequenties wil onderzoeken, kan – in principe – gebruik maken van klassieke dataverzamelmethode, zoals surveys, rondetafelgesprekken, focusgroepen enzovoort. Het is verstandiger om, juist in de digitale omgeving, waar veel sporen worden achtergelaten en opgeslagen op pc’s, iPads en (andere) smart devices, dát soort gegevens te gebruiken in plaats van vragenlijsten uit te sturen of per telefoon te ‘behandelen’. Een voorbeeld geeft het proefschrift van Bastiaan Leeuw<sup>23</sup> naar de invloed van anti-piracy beleid (in het bijzonder het Copyright Alert System (CAS) dat in de Verenigde Staten werd ingevoerd) op het gedrag van mensen/piraten. Naast enquêtes onder studenten (bekend om hun illegaal downloaden), waarbij hij vooral op zoek ging naar wat de gedragsmechanismen zijn achter dit gedrag, is gebruik gemaakt van Google-zoekgedrag-data van mensen toen de Amerikaanse jurisdictie CAS afkondigde. De Verenigde Staten werden vergeleken met een relevante en tot op zekere hoogte gelijkende andere jurisdictie (Canada), waar de betreffende interventie níét werd ingevoerd.

Voor evaluaties zijn ook twee andere ontwikkelingen belangrijk: ‘Quantified Self’ (QS), de zelfmetende mens (Cornet e.a. 2018) en Quantified Communities (QC). Bij QS leggen mensen met behulp van apps, sensoren en ‘wearables’ van alles over zichzelf vast en loaden dat soms ook op naar websites waar anderen toegang toe hebben, bijvoorbeeld om aan ‘benchmarking’ te doen. QC is het verschijnsel van smart streets, smart buildings en smart trees, waarbij ook met behulp van GPS-tracking, sensoren en beacons (kleine bluetooth-kastjes die de locatie van personen vastleggen in een openbare ruimte) enorme hoeveelheden data beschikbaar komen. Die kunnen worden gebruikt om vast te stellen hoe bepaalde situaties of

21 C.M. Klein Haarhuis & E. Niemeijer, ‘Synthesizing Legislative Evaluations: Putting the Pieces Together’, *Evaluation* (15) 2008, afl. 4, p. 403-425.

22 Jacobs 2018, p. 45.

23 H.B.M. Leeuw, *Punish, Seduce and Persuade: An Empirical Assessment of Anti-Piracy Interventions* (diss. Maastricht), Den Haag: Eleven International Publishing 2017.

F.L. Leeuw

patronen er op bepaalde momenten bij stonden, om langs die weg ontwikkelingen voor en na de invoering van bepaalde wettelijke maatregelen te traceren.

Niet alleen deze data zullen een rol bij wetsevaluaties gaan spelen, maar ook software om vast te stellen of in documenten of in de manier waarop mensen digitaal communiceren veranderingen zijn aan te treffen die op een of andere manier met wet- en regelgeving te maken hebben. Textmining doet dat bij documentenanalyses, maar een ander voorbeeld is te vinden in het werk van Lastdrager<sup>24</sup> naar phishingpraktijken. Bij de Fraudehelpdesk komen per maand gemiddeld 70.000 phishingmails binnen. 64 procent hiervan is afkomstig van een privé-e-mailadres. Lastdrager analyseerde in totaal 700.000 van deze mails en zocht naar patronen in dit enorme databestand. Ook ontwikkelde hij een app.

### 3. Zal de manier van wetgeven onder invloed van (legal) big data (analytics) kunnen veranderen en wat zou dat betekenen voor het werk van wetgevers?

Mijn antwoord op de eerste vraag is bevestigend. Ten eerste zal samenwerking met andere disciplines, waaronder data scientists, belangrijker gaan worden. Wie met big data te maken krijgt en daar niet in thuis is (wat voor redelijk wat wetgevers zal gelden), doet er goed aan *intensieve samenwerking* aan te gaan. Een voorbeeld daarvan betreft de privacywetgeving. Een van de achtergronden van de AVG zijn de grote hoeveelheden data, die ‘realtime’ én koppelbaar zijn en daarmee zorgen inzake de privacy oproepen. Om de uitvoering van deze regelgeving te realiseren zijn Privacy Impact Tools nodig, die op diverse plekken ontwikkeld en uitgetest worden (denk aan meetinstrumenten en softwaretools om inzicht te verkrijgen in (potentiële) privacyproblemen). Dit werk kan worden geduid als ‘operationalisering’ van wat in wet- en regelgeving rondom (big) data is neergelegd.

Ten tweede zal het delen van ‘*far away*’-inzichten belangrijker worden. De snelle technologische ontwikkelingen zullen voor diverse uitdagingen zorgen in de wereld van wetgeving. Denk aan robotisering en aansprakelijkheid, het verlenen van staatsburgerschap aan antropomorfe robots, ‘data justice’ (= ‘fairness in the way people are made visible, represented and treated as a result of their production of digital data’<sup>25</sup>) en de (juridische gevolgen van) empathisch wordende artificiële intelligentie.<sup>26</sup> Om over kansen en (on)mogelijkheden van wetgeving effectief van gedachten te kunnen wisselen, is kennisdeling van wat gedragswetenschappers, data scientists en ontwerpers in huis hebben nodig.

24 E. Lastdrager, *From Fishing to Phishing* (diss. Enschede), 2018.

25 L. Taylor, ‘What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally’, *Big Data & Society* (4) 2017, p. 1-14.

26 Zie <https://computerworld.nl/software/105744-empathische-ai-komt-eraan---en-dat-is-mooi> (ge raadpleegd op 9 december 2018).



Denk ook aan het veranderen van het stapsgewijze en lineaire karakter van het wetgevingsproces. Dat is op zichzelf waardevol, maar vooral begrijpelijk en inzichtelijk voor insiders. De veranderingsrichting die ik voorsta is die van een ‘trial-and-errorbenadering’. Bijvoorbeeld door bij het denken over wetgeving zo snel mogelijk naar de vraag te gaan wat de werkzame (gedrags)mechanismen zijn die (zouden) moeten maken dat de wet ‘aanslaat’, en de kennis die over dergelijke mechanismen in repositoria beschikbaar is te confronteren met de assumpties en beelden van beleidsmakers en wetgevers. Denk aan hoe medicatie tot stand komt: een van de eerste stappen is kennis te verzamelen over de ‘compounds’<sup>27</sup> die maken dat de (nieuwe) medicatie kan aanslaan en geen ongewenste neveneffecten sorteert. De vergelijking is relevant omdat wetgeving, metaforisch beschouwd, (soms) een vorm van *maatschappelijke medicatie/interventie* is. Eerder wees Veerman<sup>28</sup> al op het belang om kennis over gedragsmechanismen te gebruiken bij de ontwikkeling van wetgeving en hij werkt aan een inventarisatie daarvan. Nu ‘kennis-repositoria’ met onderzoeksresultaten naar ‘mechanismen’ sneller en gemakkelijker toegankelijk zijn geworden, zou wat Veerman bepleit meer kans van slagen moeten hebben.<sup>29</sup>

Ook op de wijze waarop de afgelopen tijd aan internetconsultatie vorm is gegeven, kunnen (legal) big data invloed hebben. Alhoewel de ministeriële brief van januari 2017<sup>30</sup> interessante perspectieven schetst, is er, gezien de snelheid van veranderingen in de wereld van het recht, data en de samenleving, meer nodig. Met de brief als vertrekpunt kan worden gedacht aan:

- Het met behulp van big data maken van netwerkkarten van wie als lobbyist waar (al dan niet) bij betrokken is. Via textmining kan in kaart worden gebracht wie wat waar en wanneer te berde bracht, op congressen, seminars, YouTube-filmpjes enzovoort. Dat vergroot de transparantie.
- Alhoewel soms als een vorm van ‘legal gambling’ gezien, zou te overwegen zijn om te experimenteren met ‘legal prediction’, gebaseerd op crowdsourcing. Katz e.a.<sup>31</sup> oefenen daarmee en keken naar de Supreme Court in de Verenigde Staten.

27 ‘Een aantal grote farmaceutische bedrijven beschikken over grote verzamelingen van verschillende kleinmoleculaire verbindingen (compound libraries genoemd).’ Compounds zijn cruciale ingrediënten van medicatie. Zie RVS, *Ontwikkeling nieuwe geneesmiddelen. Beter, sneller, goedkooper*, Den Haag 2017.

28 G.J. Veerman & R. Mulder, *Wetgeving met Beleid. Bouwstenen voor een bruikbare wetgevingstheorie*, Den Haag, Boom Juridische uitgevers 2010.

29 Denk aan Campbell (includerend: 3ie voorontwikkelingswerk/landen), NJi (Nederlands Jeugdinstituut), Corrections Inc., Cochrane, Evidence Networks enz. Een van de wijzen waarop deze kennis overgedragen wordt, is gelegen in het werken met *knowledge maps*.

30 Deze brief is op 24 februari 2017 door de minister van VenJ mede namens BZK aan de Kamer aangeboden. Onderwerp: transparantie van het wetgevingsproces.

31 D. Katz, M. Bommarito & J. Blackman, *Crowdsourcing Accurately and Robustly Predicts Supreme Court Decisions*, 1 december 2017, <https://ssrn.com/abstract=3085710> of DOI:10.2139/ssrn.3085710. Verscheen ook in: *Physics and Society* 11 december 2017.

‘Over the past several years, participants with diverse backgrounds and experience levels have competed to predict the decisions of the Supreme Court of the United States. Collectively, more than 7,000 participants have cast over 600,000 predictions across nearly 450 cases between 2011 and 2017.’

Ze vergeleken deze *crowd sourcing prediction* met andere voorspellingen. In het geval van wetgeving zouden effecten van wetgeving kunnen worden ‘voorspeld’, die dan vervolgens vergeleken kunnen worden met inzichten uit de wereld van ‘evidence-based beleid’.

- Waar in de brief van januari 2017 is aangegeven dat wetgevers verantwoording moeten afleggen over wat een rol heeft gespeeld bij de nieuwe wetgeving (p. 4), is *belangrijker* verantwoording af te leggen over wat door de wetgever *terzijde is gelegd*, en dus géén rol kon of mocht spelen. Ook hierbij kan textminning behulpzaam zijn.
- Ten slotte het punt van het delen van informatie over uitvoeringstoetsen. De brief zegt het zo:

‘Het kabinet onderschrijft het uitgangspunt van de motie-Oosenbrug c.s. dat het in een vroeg stadium actief openbaar maken van informatie over de uitvoerbaarheid een belangrijke bijdrage kan leveren aan de transparantie van het wetgevingsproces. De uitvoeringstoets is echter een lastig te definiëren informaticategorie. Bovendien is het vanwege de uiteenlopende vormen waarin de bevindingen van de toets neerslaan in schriftelijke stukken, niet mogelijk en zinvol om in algemene zin de actieve openbaarmaking van “de” toets voor te schrijven.’

Zinnen en toonzetting verraden het afwezig zijn van het denken in termen van nieuwe (big data-like) technieken van textminning, respectievelijk het werken met ongestructureerde data. ‘Lastig te definiëren’, ‘uiteenlopende vormen’, ‘niet mogelijk en zinvol’ zijn uitingen uit de tijd dat het de wetgever-schrijver was die chocolade moest maken uit de (vaak zeer kleine aantallen) consultaties en reacties. (Legal) big data-technologie analyseert patronen in deze ongestructureerde data, die daarna openbaar gemaakt kunnen worden, en waarvan vervolgens de wetgever, beargumenteerd, kan aangeven hoe hij ermee wil omgaan.

Het onderdeel van de vraag wat de gevolgen voor *wetgevers zelf* zouden kunnen zijn, lijkt relatief simpel, gezien het voorafgaande. Enerzijds gaat het om het intensiveren van de samenwerking met vakgebieden zoals data science en het werven van juristen met kennis, kunde en gevoel voor de big data-wereld. Anderzijds gaat het om het stimuleren van experimentele wetgeving,<sup>32</sup> maar ook om het volgen en beoordelen van wat big data en artificiële intelligentie nú al te bieden hebben voor de wetgevingsprofessie, én wat nog kan gaan komen. Het combineren van inzichten uit diverse, ook niet-alledaagse bronnen, zoals de wekelijkse

32 Jacobs 2018.

*Singularity Hub* met overzichtswerken in wetenschappelijke tijdschriften, is daar toe nodig.<sup>33</sup>

#### 4. Wat zijn uitdagingen, respectievelijk zorgpunten rondom (legal) big data en wetgeving?

‘Zonder wrijving geen glans’, zegt een oud gezegde. Gaan werken met (legal) big data gaat met wrijving gepaard. Wie de ontwikkelingsgang van het Programma Kwaliteit en Innovatie (KEI) ziet,<sup>34</sup> waarbij digitalisering van de rechtspraak een essentieel onderdeel is, weet met hoeveel problemen en zorgpunten zo’n innovatie gepaard is gegaan. Het zou naïef zijn te denken dat digitale innovaties rondom wetgeving ‘a piece of cake’ zijn. Ook wie zou menen dat met de AVG zorgen rondom gebruik van (legal) big data ‘over en uit’ zijn, vergist zich. De AVG gaat vooral over privacy, maar er zijn tal van *andere topics waar zorgen* over te uiten zijn.

Ik noem er een paar, zonder volledig te zijn:

- Wie is de eigenaar van de data die burgers zelf ‘aanmaken’ via zoekgedrag, sociale media, rijden in moderne auto’s, gebruik van chipcards en het dragen van fitbits en andere zelfmetingsapparatuur?
- Hoe transparant zijn de algoritmen waarmee gewerkt wordt bij analyses? Wie doorgront wat er gebeurt?
- Worden de wetgever en/of de rechter, de officier van justitie en de politieagent (meer en meer) vervangen door machines en digitale procedures? Hoe zit het dan met de aansprakelijkheid en met de acceptatie vanuit de samenleving van dergelijke ontwikkelingen?
- Hoewel kleiner wordend, omdat de generaties die ‘digital native’ zijn steeds groter worden (in demografisch opzicht), blijft het ‘digital divide’-vraagstuk bestaan: zij die niet of nauwelijks participeren aan en in de internetwereld, zijn vaak afgesloten van ontwikkelingen die het leven veraangenamen, maar ook van ontwikkelingen die alleen ‘in digitalis’ zijn te kennen.

33 Een voorbeeld van dat laatste is het special issue van *PLOS Medicine* (2018) over ‘Machine Learning’. Een voor leken zeer leesbare vooraankondiging en samenvatting verscheen onder de titel: ‘Better Medicine through Machine Learning: What’s Real, and What’s Artificial?’, <https://blogs.plos.org/speakingofmedicine/2018/11/28/better-medicine-through-machine-learning-whats-real-and-whats-artificial/>(geraadpleegd op 9 december 2018).

34 ‘Wij hebben moeten vaststellen dat al vanaf de start van het programma een aantal basiscondities voor een dergelijk grootschalig programma niet voldoende zijn ingevuld c.q. bewaakt, waardoor tijdens de looptijd ongewenste complicaties en tegenslagen zijn ontstaan. De daarop uitgevoerde corrigerende acties hebben naar ons inzicht de afgelopen jaren te weinig effect gesorteerd. Daarmee staat het programma op dit moment ernstig uit balans en zijn directe ingrepen in de vorm van een reset alsnog nodig.’ Dit was een van de hoofdconclusies uit het rapport van TRConsult, *Quick scan Review KEI. Review op risicobeheersing en basis succescondities voor grote ICT-trajecten*, april 2018. Het rapport gaf minister Dekker aanleiding in te grijpen in dit proces.

## 5. Zou er (meer/specifieke) wetgeving moeten komen met het oog op big data analytics én gebruik?

Een ingewikkelde vraag. Een recent rapport van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) houdt zich er (deels) mee bezig,<sup>35</sup> op internet verschijnen er blogs over,<sup>36</sup> ministers van Economische Zaken en Klimaat, Justitie en Veiligheid en Onderwijs, Cultuur en Wetenschap schrijven er brieven over,<sup>37</sup> en instituties zoals de KNAW en NWO wijden er ook aandacht aan,<sup>38</sup> maar het blijft een weerbarstig terrein. Zo schreef Klous<sup>39</sup> niet lang geleden dat

‘wet- en regelgeving achter loopt op de ontwikkelingen van Big Data. In een ecosysteem zijn er altijd drie zaken die geregeld moeten zijn. Er moeten duidelijke onderlinge afspraken zijn (standaarden). Er moet worden bepaald wie toegang krijgt en welke eisen daaraan gesteld worden. Tot slot moet je een oplossing hebben voor conflicten.’

Daarmee is de wetgevingskous echter niet af. Andere auteurs wijzen namelijk op andere zorgpunten en taken voor wet- en regelgevers. Het gaat dan bijvoorbeeld over discriminatie (ten gevolge van gebruik van predictive analytics bij het ‘determineren van individuen’ bij uitkeringen, banen of de schuldenproblematiek), hoe met (‘de schaamte van’) datalekken om te gaan (waarbij persoonlijke informatie publiekelijk beschikbaar komt) en hoe de ‘handel in data’ (en databanken) te reguleren is, bijvoorbeeld omdat er vraagtekens zijn bij de kwaliteit van de (verhandelde) data.

Het AI Now Institute van New York University<sup>40</sup> bracht eind 2018 zijn (derde) jaarrapport uit over de implicaties van AI voor de samenleving. Thema’s zijn wie er eigenlijk verantwoordelijk is als AI schade toebrengt aan mens en maatschappij, de ‘governance’ van de AI-industrie én de regulering van het gebruik van ‘face recognition’-software.

Klous is verder van mening dat

‘het nog zeker tien jaar duurt voor de wetgeving omtrent Big Data op orde komt. (...) Nu zie je bijvoorbeeld dat organisaties verplicht zijn om datalekken

35 WRR, *Big Data in een vrije en veilige samenleving*, Amsterdam: Amsterdam University Press 2016; B. van der Sloot e.a., *Exploring the Boundaries of Big Data*, Amsterdam: Amsterdam University Press, 2016.

36 Zie <https://www.nyenrode.nl/nieuws/n/wet-en-regelgeving-big-data> (geraadpleegd op 19 oktober 2018).

37 Minister van EZK, ‘Nadere informatie over Big data en privacy. Verwerking en bescherming persoonsgegevens’, briefnr. 78, Kamerstuk 32761. Ook: Minister voor Rechtsbescherming, ‘Transparantie van algoritmes in gebruik bij de overhead’, brief van 9 oktober 2018.

38 NWO, *Topwetenschap vereist topinfrastructuur. Adviesrapport nationale digitale infrastructuur voor wetenschappelijk onderzoek*, Den Haag 2017; KNAW 2018, zie voetnoot 4.

39 Zie <https://www.nyenrode.nl/nieuws/n/wet-en-regelgeving-big-data> (geraadpleegd op 14 oktober 2018).

40 Zie <https://ainowinstitute.org/> (geraadpleegd op 9 december 2018).

te melden. De toezichthouder bepaalt de straf. Er is echter een risico dat niet alle organisaties deze verantwoordelijkheid serieus nemen. Een volgende stap zou zijn dat er derde partijen komen, gespecialiseerd in het controleren van rapporteren over datalekken, vergelijkbaar met de accountants van nu die gespecialiseerd zijn in jaarrekeningcontroles.’

Toch is het de vraag of het zo lang gaat duren en zo zal eindigen. Immers, ten eerste gebeurt op regelgevingsgebied al het een en ander:

‘De roep om ethische en juridische kaders voor digitale technologieën klinkt steeds luider. In het voorjaar ondertekenden vierentwintig EU lidstaten de *Declaration of cooperation on Artificial Intelligence*. Daarmee wil Europa zijn concurrentiepositie versterken ten opzichte van Big Tech China (Alibaba, Tencent, Baidu, Didi) en Big Tech Amerika (Google, Facebook, Apple, Microsoft, Amazon). In de verklaring staat dat fundamentele waarden en rechten in de EU het uitgangspunt moeten vormen. Een van die waarden is dat de samenleving als geheel moet kunnen profiteren van de ontwikkelingen en toepassingen van kunstmatige intelligentie.’<sup>41</sup>

Ten tweede is de (centrale) overheid in Nederland al bezig met dit vraagstuk (zie de eerdergenoemde brieven in voetnoot 37). In een van deze brieven wordt beschreven welke wettelijke regimes en organisaties er al zijn om big data analytics en big data-gebruik te reguleren en misbruik aan te pakken. Daarbij wordt een expliciet theoretisch kader gebruikt, namelijk dat van het marktfalen door perverse prikkels. Tegelijkertijd maakt de brief ook duidelijk dat het antwoord op de vraag ‘Volstaat een juridische benadering?’ negatief is. De brief van minister Dekker (9 oktober 2018) beschrijft uitgebreid en toegelicht aan de hand van diverse voorbeelden wat op AI-gebied bij de rechtspleging al gaande is en welke vraagstukken dát oproept.

Dat brengt me tot het derde punt: de ontwikkeling in de richting van *data justice*, een concept dat ik eerder al noemde. Het wordt omschreven als ‘fairness in the way people are made visible, represented and treated as a result of their production of digital data’.<sup>42</sup> Om dit doel te realiseren wordt met methoden gewerkt om ongerechtigheden in data (analyses/rapportages/visualisaties) te voorkomen, respectievelijk aan te pakken. Het zijn geen juridische regels, maar technisch-inhoudelijke regels met een sterke ethische achtergrond. Taylor<sup>43</sup> noemt er enkele, zoals het vaststellen van de mate waarin ‘data used for governance can support power asymmetries, another focusing on the ways in which data technologies can provide greater distributive justice through making the poor visible and another is interested in how practices of dataveillance (“surveillance using digital

41 S. Beerends, *Big data evangelie vergt meer dan wetgeving*, 26 april 2018, <https://www.setup.nl/magazine/2018/04/big-data-evangelie-vergt-meer-dan-wetgeving> (geraadpleegd op 14 oktober 2018).

42 Taylor 2017, p. 6.

43 Taylor 2017, p. 6.

methods”) can impact on the work of social justice organisations’. Via deze *methoden-‘regelgeving’* en *standaarden* wordt misbruik van big data zo veel mogelijk voorkomen dan wel bestraft. Of dat op zichzelf voldoende is, is de vraag, maar in samenhang met andere (juridische) arrangementen zou een eind gekomen moeten kunnen worden. Onlangs publiceerden DeepMind-onderzoekers een studie naar criteria waarmee nagegaan kan worden hoe *veilig* AI is.<sup>44</sup> Wetgevers zouden niet alleen kennis moeten hebben van deze ontwikkelingen, maar ook aandacht moeten besteden aan de vraag hoe zij – als ze in de praktijk functioneren zoals gehoopt – ingebed kunnen worden in juridische kaders.

Ten slotte lijkt het mij voor wetgevers rondom big data belangrijk aansluiting te zoeken bij de beweging die het belang van een *publieke infrastructuur van en voor big data* benadrukt. Het in handen hebben van gigantische volumes aan (ook zeer gespecialiseerde) data door een paar zeer grote bedrijven draagt risico’s bij zich voor mens en maatschappij. Door NWO en de KNAW<sup>45</sup> wordt betoogd dat naast intensieve samenwerking met (private) organisaties, er ook een veilige en bruikbare *publieke infrastructuur* moet komen. NWO zegt het zo:

‘Onvoldoende investeringen in betrouwbare en beveiligde netwerken en data-faciliteiten, zorgt voor afname in vertrouwen in de veiligheid van data en de privacybescherming van persoonsgegevens, bijvoorbeeld bij medisch onderzoek. Omdat er niet genoeg vertrouwen onder gebruikers is dat commerciële cloud computing en storage voldoende garantie kunnen geven over de veiligheid van onderzoeksdata, kan het ontbreken van publieke nationale infrastructuur betekenen dat grote beperkingen optreden in wetenschappelijke innovatie.’

Dit zijn slechts enkele voorbeelden van wat rondom big data op mens en maatschappij af komt. Juridisch dichtschrœien van wat wel en niet kan en kan gaan kunnen, is zinloos en zet een rem op (evolutionaire) vooruitgang. Utopisch geloof in alleen het mooie en prachtige van big data en AI is naïef en kan ongelukken met zich brengen. Wetgeving is geen overbodige luxe, maar zal vermoedelijk laveren tussen wat Thierer e.a.<sup>46</sup> ‘permissionless innovation’ en ‘the precautionary principle’ noemen. Ik wens wetgevers daarbij veel wijsheid.

44 P.A. Ortega, V. Maini & DeepMind safety team, *Building Safe AI: Specification, Robustness and Assurance*, 2018, <https://medium.com/@deepmindsafetyresearch/building-safe-artificial-intelligence-52f5f75058f1> (geraadpleegd op 17 oktober 2018).

45 NWO 2017; KNAW 2018.

46 A. Thierer e.a., *Artificial Intelligence and Public Policy*, Mercator Center, George Mason University 2017, p. 3 en 5. ‘Permissionless innovation refers to the idea that “experimentation with new technologies and business models should generally be permitted by default. The precautionary principle refers to to preemptively limit or even ban certain applications out of fears for worst-case scenarios”.’