



# Data science maturity van gemeenten

## White paper

datum	4 april 2017
auteur(s)	dr. Patricia Prüfer dr. Natalia Kieruj
Met bijdragen van	prof. dr. Arjan van den Born prof. dr. Geert Duijsters prof. dr. Ronald Leenes dr. Johan Wolswinkel dr. Wesley Kaufmann Richard Haans, Msc.
versie	1.0

© CentERdata & JADS, Tilburg & Den Bosch, 2017

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.



## Inhoudsopgave

1	Inleiding.....	2
	Box 1: Arjan van den Born (hoogleraar Creative Entrepreneurship, TiU): Big data, big wisdom & de opkomst van policy science .....	5
2	Context en achtergrond .....	6
	2.1 Verkenning Data Lab voor gemeenten.....	6
	Box 2: Wesley Kaufmann (universitair hoofddocent Public Governance, TiU): Verdiepen en verbreden van data science in de context van de Omgevingsvisie ....	7
	Box 3: Richard Haans (AiO Public Governance, TiU): Het potentieel van big data voor nieuwe inzichten over 'regionale bedrijvigheid' .....	8
	2.2 Data maturity modellen.....	9
3	Data maturity scan gemeenten .....	11
	3.1 Opzet vragenlijst .....	11
	3.2 De vragenlijst in het veld.....	12
	3.3 Diepte-interviews met gemeenten .....	14
	Box 4: Ronald Leenes (hoogleraar Regulering door technologie, TiU): Big data, big difference? .....	16
4	Uitkomsten maturity scan .....	18
	4.1 Een eerste algemene indruk.....	18
	4.2 Indeling in clusters .....	19
	4.3 De invloed van de clusters op data maturity van gemeenten.....	23
	4.4 De invloed van de clusters op data maturity van gemeentemedewerkers.....	28
	4.5 De invloed van achtergrondkenmerken op data maturity van gemeentemedewerkers .....	30
5	Discussie en aanbevelingen.....	31
	Box 5: Geert Duijsters (hoogleraar Entrepreneurship, TiU): Technologische disruptie - kansen voor samenwerking .....	33
	Box 6: Johan Wolswinkel (universitair hoofddocent Bestuursrecht, TiU): Data en Openbaar Bestuur – zijn gemeenten 'too small for big data'?.....	35
6	Conclusies .....	36
7	Literatuur .....	37
A	Scoringsniveaus maturity scan .....	38
B	Diepte-interviews .....	39
C	Clusters deelnemende gemeenten.....	40



# 1 Inleiding

De meeste organisaties maken wel gebruik van data, maar zijn toch niet in staat om alle voordelen die het gebruik van data kan opleveren volledig uit te buiten. Vaak worden data gebruikt om de resultaten van bepaalde acties of interventies uit het verleden weer te geven, maar worden ze niet gebruikt om beslissingen en strategieën voor de toekomst richting te geven. Het juiste gebruik van data kan een groot strategisch voordeel opleveren en het is dus zeker de moeite waard voor organisaties om zich meer toe te leggen op efficiënt en creatief gebruik van data. Een organisatie moet de unieke mogelijkheden van data begrijpen, en moet ook inzien hoe de volwassenheid (*maturity*) van de organisatie het mogelijk of juist onmogelijk maakt om op strategische wijze gebruik te maken van data analyse om op die manier zijn doelen te behalen.

*Data science* is een relatief nieuwe discipline die tot nieuwe innovaties moet gaan leiden. Met data science bedoelt men het analyseren en interpreteren van grote hoeveelheden data die, vaak complex en ongestructureerd, tegenwoordig dagelijks gegenereerd worden (*big data*). De kunst van data science is het transformeren van data in waardevolle acties.

Dit vakgebied beslaat meer dan het conventionele gebruik van data en statistiek. Wat data science onderscheidt van statistiek is de holistische benadering. Er is steeds meer data 'in het wild' te vinden en data scientists verzamelen deze data, gieten de data in een bruikbare vorm en presenteren het verhaal dat de data vertellen.

Conventioneel statistisch onderzoek is over het algemeen veel meer gericht op het verleden. Er worden hypothesen opgesteld en met behulp van data worden deze getoetst om aan te tonen dat bepaalde processen en verbanden in het verleden zo hebben plaatsgevonden. Data science daarentegen is er veel meer op gericht om ook voorspellingen te doen voor de toekomst en op die manier aanbevelingen te doen voor belangrijke beslissingen die genomen moeten worden.

Maar wat betekent de *datafication*, het steeds meer toenemende belang van data, die de maatschappij en de manier hoe overheidsinstellingen werken volkomen zal veranderen nu eigenlijk precies? Wat gaat er veranderen en hoe moeten overheidsinstellingen daarmee omgaan en op inspelen? Wat zou de meerwaarde kunnen zijn van beleidsvorming die op meer data gebaseerd is?

Laten we beginnen met een voorbeeld op gemeenteniveau. New York City is onder de voormalige burgermeester Michael Bloomberg begonnen alle beschikbare big data te koppelen en te gebruiken om de gemeentelijke dienstverlening te verbeteren en de kosten daarvan tegelijkertijd te verlagen. Zo werd er bijvoorbeeld een nieuwe strategie bedacht om huisbranden te voorkomen.

Het was al bekend dat huizen die illegaal onderverdeeld zijn om meer bewoners te huisvesten met een veel hogere kans in vlammen opgaan dan andere huizen. In New York worden dan ook 25.000 meldingen per jaar gemaakt over overvolle huizen maar de gemeente heeft maar 209 inspecteurs die zich daarmee bezig kunnen houden.



Daarom heeft een team van data scientists zich hiermee bemoeid en dit team heeft een database gemaakt waarin de informatie van alle 900.000 gebouwen in de stad gekoppeld is aan informatie van 19 andere instellingen; belastingen, (fraude bij) energie- en waterverbruik, onbetaalde rekeningen en afsluitingen, criminaliteitsstatistieken en meldingen over overlast gevend gedrag en spoedeisende hulp.

Vervolgens zijn deze gegevens vergeleken met brandweerinformatie over alle huisbranden van de afgelopen vijf jaren, rekening houdend met de zwaarte van elke brand. De hoop van de data scientists was om verbanden tussen deze data te herkennen en op die manier patronen op te sporen welke huizen een groot risico lopen om in vlammen op te gaan. En het bleek inderdaad dat het type gebouw uitmaakt net als het bouwjaar. Onverwachts bleek ook dat gebouwen met uitwendig metselwerk veel minder risico hadden op brand dan huizen met standaard houten muren.

Op grond van al deze informatie is een waarschuwingssysteem ontwikkeld dat aangeeft of een melding over een overvol huis meteen in behandeling moet worden genomen of niet. Interessant hierbij is dat geen van de verzamelde indicatoren op zichzelf brand kan veroorzaken. Maar ze zijn wel allemaal sterk gecorreleerd met huisbrand en vooral in combinatie krachtige voorspellers. Dit systeem leidde er toe dat de 209 inspecteurs gericht aan de slag konden gaan, er veel vaker overvolle huizen met een groot risico onmiddellijk ontruimd werden en het aantal huisbranden duidelijk daalde.

Naast dit voorbeeld zijn er ook voor Nederlandse gemeenten tal van toepassingen te bedenken waar data science van belang kan zijn. Sommige gemeenten zijn daarmee ook al bezig. Zo gaat de gemeente Alkmaar bijvoorbeeld van start met een pilotstudie waarbij data verkregen met behulp van sensoren een rol kunnen spelen in de verbetering van de leefomgeving. De luchtkwaliteit en geluidsniveaus in verschillende gebieden zal gemeten worden en met deze informatie kan vervolgens gericht actie worden ondernomen. De gemeente Utrecht is momenteel bezig met een project in het kader van inbraakpreventie. Het doel is om bepaalde risicogebieden te identificeren op zogeheten *heat maps* en om tegelijkertijd ook in kaart te brengen welke omgevingsfactoren ervoor zorgen dat bepaalde gebieden een hoger risico met zich mee brengen.

Daarnaast loopt in Amsterdam een project waarbij de intensiteit van potentiële overlast wordt voorspeld op basis van portofoon- en digitale bonddata, omgevingskenmerken en weerdata. Deze informatie kan vervolgens worden ingezet voor een efficiëntere handhaving.<sup>1</sup>

Enerzijds is er dus sprake van de mogelijkheden en kansen van data science waardoor nog waardevollere informatie en inzichten uit data gehaald kunnen worden voor datagedreven en *fact-based* beleid. Anderzijds zijn er ook uitdagingen verbonden aan data science. Naast de waarborging van privacy en de technische kennis over hoe men met deze grote hoeveelheden data om moet gaan in technische zin, is het van belang om ook te weten hoe de juiste vragen gesteld moeten worden. Zonder deze kennis en inzicht is niet duidelijk wat belangrijk is om te weten en hoe waarde toegevoegd kan worden. Wat is de waarde

---

<sup>1</sup> Zie bijvoorbeeld de website voor dit project: [www.bigdatagemeenten.nl](http://www.bigdatagemeenten.nl).



van data en de daaruit te genereren informatie en wat het is belang daarvan voor de samenleving en/of organisaties?

Om het bovenstaande te bewerkstelligen zijn er mensen nodig die niet alleen technisch capabel zijn, maar ook gedegen kennis op het gebied van statistiek en wiskunde hebben én daarnaast nieuwsgierig zijn, nieuwe dingen willen ontdekken, willen puzzelen of door nieuwe combinaties te maken iets 'nieuws' opleveren. Naast inventief, creatief en onderzoekend zijn is het ook belangrijk om het grote geheel in het oog houden.

Hoe staan de Nederlandse gemeenten er wat dat betreft voor? Wat moet een gemeente leren over data science en wat betekent data science voor gemeentelijke besturen en beleidsmedewerkers? Zijn gemeenten al (enigszins) in staat om data te gebruiken om beleid beter te maken en/of daadwerkelijk datagedreven beleid te maken?

Dit white paper levert eerste antwoorden op deze vragen en geeft aanbevelingen voor gemeenten op verschillende niveaus van datavolwassenheid (*data maturity*). In het volgende hoofdstuk schetsen we de achtergrond van data maturity en beschrijven de pilot die de Vereniging Nederlandse Gemeenten en het Kwaliteitsinstituut Nederlandse Gemeenten (verder: VNG/KING) hebben uitgevoerd met de Jheronimus Academy of Data Science (verder: JADS).

In diverse hoofdstukken bevinden zich in boxen korte standpunten en ideeën van bij deze pilot betrokken onderzoekers van JADS. Zo leest u onderaan in deze inleiding in box 1 de bijdrage van Arjan van den Born die als academisch directeur van JADS van begin af aan betrokken was bij dit initiatief.<sup>2</sup>

De onderzoekscomponent van deze pilot, het uitvoeren van een data maturity scan in de vorm van een vragenlijst onder gemeenten, wordt toegelicht in hoofdstuk 3. Hoofdstuk 4 toont de uitkomsten uit deze scan en een indeling in drie groepen van gemeenten op basis van de mate van data maturity. De inzichten uit de maturity scan brengen we in hoofdstuk 5 in de discussie samen met de *lessons-learned* uit de pilot en komen uiteindelijk tot aanbevelingen voor Nederlandse gemeenten. We ronden dit rapport met een concluderend hoofdstuk af.

---

<sup>2</sup> Naast de onderzoekers van JADS en CentERdata hebben ook medewerkers van KING een bijdrage geleverd aan dit white paper. Vooral Eric de Kruik en Mark Gremmen leverden waardevolle commentaren en hebben er aan bijgedragen dat de aanbevelingen nog rijker en nuttiger werden.



## Box 1: Arjan van den Born (hoogleraar Creative Entrepreneurship, TiU): Big data, big wisdom & de opkomst van policy science

De opkomst van big data zal de gemeente veranderen. De hoeveelheid data in steden en dorpen neemt exponentieel toe: data over de verkeer- en vervoerstromen, over *sustainability*, economische activiteit, gezondheid en sociaal welzijn. Wat betekent dit voor de gemeente en hoe zij functioneert? Om dat in kaart te brengen, kunnen we misschien leren van de trends die we nu zien in het bedrijfsleven. Wat kan de gemeente leren van bedrijven als booking.com, Facebook, Coolblue en Brand Loyalty?

Als je naar deze toonaangevende bedrijven kijkt zie je ten eerste dat kennis van statistiek en programmeren belangrijker wordt. Bij bedrijven als Facebook, Microsoft en Amazon komen statistici in de raad van bestuur. De programmeur is voor de 21<sup>e</sup> eeuw wat de fabrieksarbeider voor de 20<sup>e</sup> eeuw was. Kennis van statistiek en IT zal op alle niveaus bij de gemeente nodig zijn. Ook de gemeentesecretaris wordt een beetje een *nerd*.

Ten tweede zie je dat datagedreven organisaties continu terugkoppeling ontvangen over de effectiviteit van hun acties. Wat werkt en wat werkt niet? Dat zal ook in toenemende mate gebeuren bij gemeenten. Deze continue terugkoppeling over de effectiviteit van beleid zal leiden tot een verandering in de discussie tussen politiek en beleidsmedewerkers. Beleid wordt veel sneller geëvalueerd en aangepast. Niet pas na enkele jaren zoals nu, maar in de toekomst soms al na enkele dagen en/of weken. Doen en denken gaat veel meer samen.

Ten derde zie je dat datagedreven organisaties uitblinken in het opzetten van experimenten om zo spelenderwijs te leren. Zij voeren duizenden experimenten per maand uit. Ook gemeenten zullen in toenemende mate experimenten opzetten om zo te leren wat werkt en wat niet in het sociale en economische beleid. De gemeente wordt zo een lerende, wetenschappelijke organisatie. Beleid en wetenschap komen dicht bij elkaar; *policy science* ontstaat.

Ten vierde zie je nieuwe samenwerkingsmodellen opkomen, soms gebaseerd op Blockchain-achtige technologie. Om te kunnen leren van de data is het soms nodig om data te delen. Denk aan verschillende gemeenten die willen leren of een bepaalde aanpak op het gebied van jeugdzorg werkt. Dan zal je samen moeten werken met andere gemeenten en instituties om samen voldoende rijke data te hebben om echt te kunnen zeggen of een bepaalde aanpak werkt.

Tot slot. Big data is geen *big wisdom*. Met de opkomst van big data ontstaan ook nieuwe gevaren. Natuurlijk *privacy* en *security*. Maar ook dreigt er *information overload* waarbij de lange termijn vergeten wordt en er hijgerig achter hypes wordt gelopen. Om deze en andere valkuilen te vermijden zal de gemeente moeten investeren in de kennis over big data; wat het wel kan, maar ook wat het niet kan.



## 2 Context en achtergrond

### 2.1 Verkenning Data Lab voor gemeenten

In oktober 2016 hebben VNG/KING en JADS de intentie uitgesproken om met elkaar te gaan samenwerken. Het achterliggende doel van deze samenwerking is de data science competenties van gemeenten te verhogen. Beide partijen delen de droom om *fact-based policy making* naar een hoger plan te trekken en zijn daarom de “Verkenning Data Lab voor gemeenten” begonnen.

JADS is in 2016 opgericht als een samenwerking op het gebied van data science door Tilburg University (TiU), Technische Universiteit Eindhoven (TU/e), de gemeente Den Bosch en de provincie Noord-Brabant. Het doel is om onderwijs, onderzoek en samenwerking met overheidsinstellingen en bedrijven rondom data science te bevorderen. Ook het onderzoeksinstituut CentERdata, een stichting gevestigd op de campus van TiU, is als partner betrokken en levert zowel technische ondersteuning als inhoudelijke bijdragen aan JADS. Tevens leidde CentERdata het onderzoeksgedeelte van de verkenning.

VNG/KING en JADS zijn een pilot begonnen om gezamenlijk ervaringen op het gebied van ‘data science voor gemeenten’ op te doen. Afhankelijk van de ervaringen in deze pilot en de (potentiële) meerwaarde is het de bedoeling om een langdurige samenwerking aan te gaan. Het idee is om een ecosysteem op te richten waarin verschillende partijen een bijdrage leveren en waar samenwerking en alle stappen van het data-waarde proces (o.a. verzamelen, integreren, analyseren, presenteren) worden ondersteund.

De bedoeling van deze pilot was om in een periode van vijf maanden te starten met twee afgebakende thema’s: regionale bedrijvigheid en omgevingswet. Deze vraagstukken zijn in projectvorm opgepakt, waarbij gebruik is gemaakt van de support en software van Synerscope en van CentERdata. Naast het opleveren van directe resultaten van deze twee projecten waarbij met verschillende gemeenten naar verschillende data gekeken werd, zijn deze projecten ook gebruikt om te verkennen hoe gemeenten data science kunnen gaan gebruiken, waar de waarde ligt, en waar de uitdagingen. Daarnaast werd er gekeken hoe JADS en VNG/KING in samenwerking met anderen een platform en ecosysteem zou kunnen bieden om gemeenten te helpen met deze uitdaging en wat dit zou betekenen voor een samenwerking.

Daarnaast maakte een ‘leerproject’ onderdeel uit van de pilot waarbij alle betrokkenen vanuit JADS, CentERdata, Synerscope en VNG/KING die meedraaiden in de twee projecten, ook meedraaiden in het leertraject. Onder dit gedeelte vallen ook de in hoofdstuk 3 beschreven enquête op basis van een maturity model en de diepte-interviews met gemeenten. Uiteindelijk zou de pilot een white paper opleveren over ‘gemeenten en Data Science’; de stand van zaken, uitdagingen en oplossingen.

In boxen 2 en 3 beschrijven twee onderzoekers van de Tilburg School of Governance op Tilburg University (TiU) die betrokken waren bij de twee projecten in de pilot hun ervaringen. Wesley Kaufmann was als wetenschappelijk expert betrokken bij het project dat zich bezig hield met de omgevingswet. Richard Haans was als onderzoeker betrokken bij het project ‘regionale bedrijvigheid’.





## Box 2: Wesley Kaufmann (universitair hoofddocent Public Governance, TiU): Verdiepen en verbreden van data science in de context van de Omgevingsvisie

Het nader vormgeven en ontwikkelen van de Omgevingsvisie staat op dit moment hoog op de agenda van beleidsmakers en andere belanghebbenden die betrokken zijn bij de Nederlandse leefomgeving. En terecht, want de Visie – die concreet tot uiting moet komen met de inwerkingtreding van de Omgevingswet in 2019 – is uitermate ambitieus. In de startnota over de Visie zijn vier strategische doelen beschreven, te weten een concurrerende en duurzame economie, een klimaatbestendige samenleving, een toekomstbestendige woon-werkomgeving en een waardevolle leefomgeving. Er wordt dus nogal wat verwacht van de Omgevingswet, aangezien het succesvol bereiken van één van deze doelen al een hele prestatie zou zijn.

Toch is het niet onlogisch om voor de toekomstige ontwikkeling van ruimtelijke ontwikkeling in Nederland breed in te zetten. De strategische doelen zoals uiteengezet in de startnota zijn namelijk gerelateerd aan *wicked problems*; problemen die zeer complex zijn vanwege een groot aantal belanghebbenden, een hoge mate van verandering van contextuele factoren en een nauwe verwevenheid met andere belangrijke problemen. Hoewel het in beginsel mogelijk is om *wicked problems* te benaderen door op een autoritaire, top-down manier oplossingen te forceren is het – zeker in de Nederlandse maatschappij – veel zinvoller om door middel van breed gedragen samenwerking tot nieuwe inzichten te komen. Het is deze collaboratieve weg die het kabinet heeft ingeslagen; provincies, gemeenten, het bedrijfsleven en maatschappelijke organisaties worden (in elk geval op papier) zoveel mogelijk bij het traject betrokken.

Tegen deze achtergrond rijst de vraag: Hoe kan data science een rol spelen in de (verdere) ontwikkeling van de Omgevingswet? Tijdens de verkenning van het data lab zijn gemeenten in samenwerking met JADS en Synerscope met deze vraag aan de gang gegaan. Een belangrijke eerste stap die in dit kader is gezet is het ontwikkelen van een data science mindset. De razendsnelle ontwikkeling van data science maakt het domein enerzijds zeer interessant, maar zorgt er anderzijds voor dat veel praktische problemen en kinderziektes de kop opsteken. Wat dat betreft heeft data science, zoals onderzoek in het algemeen, nog een hoog *trial and error* gehalte. Het uitwisselen van ervaringen en inzichten in een open omgeving kan de effectieve toepassing van data science sterk bevorderen.

In de tweede plaats is aan het einde van de verkenning ruimte geweest om 'met data te spelen'. Zo is het bijvoorbeeld mogelijk gebleken om met een paar muisklikken 1) voor een bepaalde gemeente overzichtelijk in kaart te brengen waar windmolens geplaatst mogen worden en 2) het gemiddelde energieverbruik per stadsdeel te tonen. Deze verkenningen van data op gemeente- en stadsniveau door middel van visualisatie lichten al een tipje van de sluier op als het gaat om de potentie van data science en ruimtelijke ontwikkeling. Maar daarmee zijn we er natuurlijk nog lang niet.

Om de potentie van data science in de context van de Omgevingswet te ontsluiten is er zowel verdieping als verbreding nodig (waarbij direct opgemerkt dient te worden dat beide elementen nauw verweven zijn). De verdieping zit vooral in het systematisch analyseren van data. Zo kunnen (gekoppelde) datasets niet alleen op verschillende manieren gevisualiseerd worden, maar zijn er ook andere methoden voorhanden (bijvoorbeeld data mining, machine learning, evenals gangbare statistische methoden). In sommige gevallen kan een duidelijke visualisatie van de data al genoeg inzicht verschaffen, maar in veel andere gevallen zullen analyses nodig zijn die 'harde getallen' opleveren. De verbreding van data science zit hem niet alleen in het verkrijgen en koppelen van meer datasets (hoewel dit uiteraard van groot belang is), maar evengoed



in het samenbrengen van zoveel mogelijk belanghebbenden. Het feit dat de ontwikkeling van de Omgevingswet expliciet is ingezet als een samenwerkingsmogelijkheid voor verschillende overheidslagen, het bedrijfsleven en maatschappelijke instanties, zal ertoe moeten leiden dat toekomstige initiatieven verder reiken dan gemeenten alleen.

Uiteraard is het in deze beginfase naïef om te denken dat verschillende belanghebbenden zondermeer bij elkaar kunnen aanschuiven en direct tot nieuwe inzichten zullen komen. Op dit moment is het juist van belang dat gemotiveerde medewerkers aan de slag gaan met het verzamelen, opschonen en analyseren van data binnen de eigen organisatie. Maar op de middellange termijn is het voor een effectieve toepassing van data science in de context van ruimtelijke ontwikkeling cruciaal dat er zo breed mogelijk wordt samengewerkt. Wanneer verschillende belanghebbenden als onderdeel van deze samenwerking hun eigen datasets inbrengen, deze datasets gekoppeld kunnen worden en vervolgens grondig geanalyseerd, dan zijn we een flinke stap dichterbij het realiseren van de ambitieuze doelen van de Omgevingsvisie.

### **Box 3: Richard Haans (AiO Public Governance, TiU): Het potentieel van big data voor nieuwe inzichten over 'regionale bedrijvigheid'**

Tijdens de Mini Hackathon, georganiseerd binnen het kader van de samenwerking tussen het Jheronimus Academy of Data Science, het Kwaliteitsinstituut Nederlandse Gemeenten en de Vereniging van Nederlandse Gemeenten, was het voor mij heel duidelijk wat voor een mogelijkheden big data kan bieden voor gemeenten. Tegelijkertijd kwam uit deze proef ook heel helder naar boven dat big data geen wondermiddel is, en dat menselijke kennis evenals een kritische blik op de data altijd nodig zijn.

Binnen de subgroep Bedrijvigheid was de leidraad de high-tech sector. Het hoofddoel van de Hackathon was om nieuwe manieren te ontdekken om high-tech clusters te omschrijven, waarbij niet uitgegaan werd van externe definities maar in plaats daarvan de data (in dit geval: teksten op organisationele websites) leidend moesten zijn.

Het was indrukwekkend om te zien hoe snel nieuwe ideeën en oplossingen gegenereerd werden door de deelnemers, want waar begin je nu eigenlijk als je niet uit mag gaan van standaard bedrijfsindelingen? Op basis van bestaande kennis van de experts over hoe een typische high-tech organisatie er nou eigenlijk uit ziet werden al snel soortgelijke organisaties geïdentificeerd op basis van het soort woorden dat gebruikt werd door de prototypische organisaties binnen deze clusters. Zo werd helder dat je met de combinatie van menselijke kennis en grootschalige data al snel heel fijnkorrelig onderscheid kunt maken tussen verschillende organisaties, zonder afhankelijk te hoeven zijn van bestaande (rigide) richtlijnen.

Echter werd het ook duidelijk dat je niet blind van de data moet uitgaan. Zo bleken enkele van de grote spelers onverwacht niet door de nieuwe afbakening opgepakt te worden. Hoorden deze organisaties eigenlijk helemaal niet tot deze clusters? Of zaten de data fout? In dit geval leek het een combinatie van verschillende factoren. Zo bleek het dat sommige organisaties helemaal geen website hadden die in de data zaten, waardoor ze per definitie niet geïdentificeerd konden worden. Anderen bleken weer een zeer breed portfolio van activiteiten te hebben, waardoor ze een te grote diversiteit aan woorden gebruikten om netjes binnen de delineatie die uit de eerste analyse naar boven kwam te vallen.

Dit deed me denken aan een uitspraak van Nobelprijswinnaar Ronald Coase: als je de data lang maar genoeg martelt zal deze altijd bekennen. Hier leek echter martelen niet



nodig te zijn: in plaats daarvan kwam de bekentenis snel en bleek deze bovendien incompleet. Het is aan de onderzoeker om er voor te zorgen dat de uitkomsten uit de algoritmes en modellen zo nauw mogelijk overeenkomt met reeds bestaande kennis en expertise. Dit is, gegeven de enorme schaal van de data en de complexiteit van de werkelijkheid, geen makkelijke opgave.

Voor mij liet deze oefening al met al zowel de potentie als de uitdagingen omtrent big data zien. Aan de ene kant leek big data men te bevrijden, creatiever te maken en inzichten te bieden op een schaal die klassieke methodes simpelweg niet aanbieden. Tegelijkertijd was het heel helder hoe belangrijk de menselijke *touch* is voor succesvolle toepassingen van big data analytics. Hiervoor zijn investeringen nodig in niet alleen meer en betere data, maar ook in diepere kennis omtrent nieuwe methodes evenals de praktijk. Geen makkelijke opgave, maar duidelijk de moeite waard!

## 2.2 Data maturity modellen

Het moge duidelijk zijn data science van grote toegevoegde waarde kan zijn voor gemeenten. Daarom is het van belang dat gemeenten wegwijs worden in dit nieuwe vakgebied, waarbij het vanzelf spreekt dat het gemeenten enige tijd en inspanning zal kosten om een kennisniveau te bereiken dat afdoende is. De ene gemeente is waarschijnlijk al verder in deze ontwikkeling dan de andere en het is daarom interessant om eerst in kaart te brengen hoe ver de individuele gemeenten van Nederland zijn in dit proces.

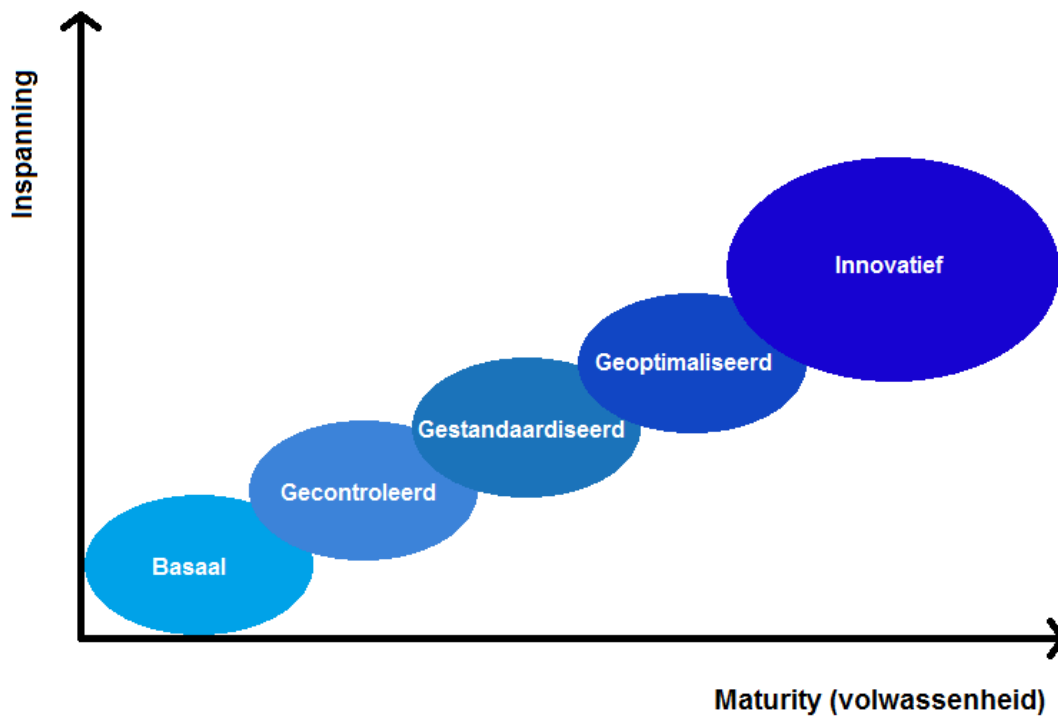
Een model dat hier uitermate geschikt voor is, is het *Maturity Model*. Het concept van dit model werd geïntroduceerd met het Capability Maturity Model (CMM) van de Software Engineering Institute (SEI), en wordt inmiddels in meerdere disciplines toegepast om competenties van organisaties te meten en als basis voor verbetering van processen te fungeren (de Bruin & Freeze 2005; Loshin 2011).

Er wordt uitgegaan van het principe dat een organisatie door bepaalde fasen heen gaat om uiteindelijk volwassenheid te bereiken in een bepaald domein. Deze fasen zijn cumulatief, waarbij een hogere fase voort bouwt op de voorwaarden van de lagere fasen. Het aantal fasen kan variëren, maar het is van belang dat de fasen een onderscheidend vermogen hebben ten opzichte van elkaar, en dat er een logische progressie zit in de opbouw van de fasen. De fasen krijgen allemaal een aparte naam die zo goed mogelijk beschrijft wat elke fase inhoudt. Daarnaast wordt ook per fase een uitgebreide beschrijving gegeven die duidelijk omschrijft wat de voorwaarden van iedere fase zijn en hoe die gemeten worden (vooral de aspecten die nieuw zijn voor die fase) (Klimko 2001).

Een maturity model geeft weer dat er een verband is tussen de mate van volwassenheid (maturity) die een organisatie met betrekking tot een bepaald domein heeft en de inspanningen die de organisatie hiervoor levert. Figuur 2.2 geeft dit verband weer. Zoals de grafiek laat zien gaat men binnen het domein data science door de volgende vijf fasen alvorens de organisatie bij wijze van spreken het eindstation bereikt heeft, en dus volwassen is: basaal, gecontroleerd, gestandaardiseerd, geoptimaliseerd en innovatief.



Figuur 2.2 Maturity Model Data Science



Om een nieuwe fase te bereiken moet de organisatie eerst de eerdere fasen goed onder de knie krijgen, aangezien deze een voorwaarde zijn voor de latere fasen (Booz, Allen, Hamilton, 2015; Dataflux). Wanneer een organisatie zich een bepaalde fase, zoals de geoptimaliseerde fase, eigen probeert te maken worden er extra inspanningen verricht in het kader van die fase. Wanneer de organisatie de geoptimaliseerde fase vervolgens onder de knie heeft, wordt de hoeveelheid inspanningen in die fase aanzienlijk minder aangezien hier een duidelijke efficiëntieslag gemaakt zal zijn.

Dit betekent niet dat er helemaal geen inspanningen meer geleverd worden die raken aan werkzaamheden in de geoptimaliseerde fase, want er zal altijd enige mate van inspanning nodig zijn om de werkzaamheden behorende bij deze fase draaiende te houden, maar het gros van de inspanningen verplaatst zich vervolgens wel naar de volgende fase die eigen gemaakt dient te worden: de innovatieve fase (Booz, Allen, Hamilton 2015).



### 3 Data maturity scan gemeenten

Om een beeld te krijgen van de mate waarin gemeenten in Nederland op dit moment al datagedreven te werk gaan als het om beleidsvorming gaat, is een vragenlijst opgesteld. Deze vragenlijst vormt (in combinatie met een reeks interviews) de *data science maturity scan*. Met behulp van dit instrument meten we in hoeverre gemeenten al volwassenheid hebben bereikt binnen het domein data science.

In dit hoofdstuk lichten we de opzet van de vragenlijst nader toe en beschrijven de veldwerkperiode. Hierbij gaan we ook in op de achtergrondkenmerken van de deelnemende gemeenten en de respondenten. Vervolgens schetsen we de diepte-interviews die gevoerd zijn met negen gemeenten die mee hebben gedaan aan de pilot “Verkenning Data Lab voor gemeenten”.

#### 3.1 Opzet vragenlijst

Maturity models, of maturity scans, zijn tot stand gekomen om de competentie van organisaties op een bepaald domein vast te stellen en te waarderen, en om de groei van een organisatie binnen een bepaald domein richting te geven (Huner et al. 2009; Becker et al. 2009; Pöppelbuss & Röglinger 2011). Dit wordt bij voorkeur gedaan op basis van een zo veel mogelijk omvattende set van criteria, meestal met behulp van scores op Likert schalen (bv. een antwoordschaal die loopt van 1 ‘helemaal niet competent’ tot 5 ‘heel erg competent’) op bepaalde criteria (de Bruin & Freeze 2005).

Aangezien er nog geen maturity scan bestond voor het thema *Data Science*, is de enquête tot stand gekomen door te kijken naar maturity models uit andere vakgebieden zoals informatietechnologie. Vervolgens zijn op basis van deze inzichten uit de literatuur de vragen door het *Projectteam Data Science* van TiU (Bosma et al. 2016) aangepast en toegespitst op data science met behulp van experts.

Deze eerste versie is getest op een grote organisatie. Hierbij werd aanvullende input over missende elementen verzameld tijdens diepte-interviews met verschillende medewerkers en managers van deze organisatie. Op basis van deze inzichten heeft CentERdata de enquête vervolgens aangepast om de maturity scan onder de Nederlandse gemeenten af te nemen in het kader van de pilot.

In zekere zin is het uitvoeren van de data maturity scan ook een experiment, waarbij dit nieuwe instrument voor het eerst ingezet is bij gemeenten. In hoofdstuk 4 lichten we toe tot welke inzichten je al dan niet kunt komen op basis van deze ‘pilot maturity scan’.

In de vragenlijst komen verschillende componenten van het werken met data en *fact-based* beleid aan bod. Deze componenten zijn vervolgens weer opgedeeld in meerdere subcomponenten (zie hieronder). Voor iedere component wordt door middel van een vragenset vastgesteld in welke fase men zich bevindt, dus in hoeverre men de bijbehorende werkzaamheden al onder de knie heeft.

Zoals gebruikelijk is er in deze maturity scan gekozen voor vijf fasen. Deze fasen krijgen allemaal een omschrijving die weergeeft waar een organisatie zoal mee te maken heeft



wanneer deze zich in een bepaalde fase bevindt (zie ook figuur 2.2). Een uitgebreide omschrijving van deze fasen, ofwel scoringsniveaus is te vinden in bijlage A. De interne validiteit en betrouwbaarheid van de survey(vragen) werd aangetoond door middel van factoranalyse. Vervolgens zijn op basis daarvan de vragen heringedeeld naar (sub)categorieën.<sup>3</sup>

De vier componenten met telkens de vier subcomponenten zijn als volgt:

- 1. Data en technologie**
  - a) Data
  - b) Data governance
  - c) Infrastructuur
  - d) Alignment IT-business
- 2. Processen**
  - a) Reikwijdte
  - b) Methoden
  - c) Verankering
  - d) Meten en sturen
- 3. Management en organisatie**
  - a) Strategie en doelen
  - b) Organisatie
  - c) Governance
  - d) Waardecreatie
- 4. Cultuur en vaardigheden**
  - a) Mensen
  - b) Competenties
  - c) Communicatie
  - d) Cultuur

Uiteindelijk wordt hieruit duidelijk in welke fase de gemeenten die deel nemen aan het onderzoek zich bevinden en in hoeverre er sprake is van volwassenheid op het gebied van data science. Dit alles leidt tot een beter begrip van de sterke en zwakke punten van gemeenten als het gaat om datagedreven beleidsvoering. Vervolgens kan gericht gekeken worden naar waar eventuele verbetering plaats zouden moeten vinden.

## 3.2 De vragenlijst in het veld

De enquête voor de maturity scan is door CentERdata in november gemaakt door de eerste versie aan te passen voor gemeenten. Vervolgens heeft VNG/ KING de enquête naar de gemeentesecretarissen van alle Nederlandse gemeenten gestuurd met het verzoek om de vragenlijst door de meest geschikte persoon of afdeling in te laten vullen. De doorlooptijd waarin de vragenlijst ingevuld kon worden was twee maanden, 28 november 2016 tot 1 februari 2017.

Na het versturen van twee reminders is de vragenlijst uiteindelijk door 115 personen volledig ingevuld. Aangezien deze personen soms voor meerdere gemeenten werken of de

---

<sup>3</sup> De uitkomsten uit de exploratieve en confirmatieve factoranalyse tonen we verder niet in dit white paper. Op verzoek zijn deze analyses verkrijgbaar bij CentERdata.



enquête voor een samenwerkingsverband hebben ingevuld, betreft deze informatie 127 verschillende gemeenten.<sup>4</sup>

Van de huidige 388 gemeenten in Nederland heeft dus ongeveer een derde deelgenomen aan de maturity scan. Doordat de vragenlijst naar de gemeentesecretarissen is gegaan, weten we niet precies wie de enquête uiteindelijk heeft ingevuld. Daarnaast we weten niet of de gehele vragenlijst door dezelfde persoon is ingevuld of door meerdere personen, noch of de enquête is ingevuld door een specialist of manager. Desalniettemin kunnen we de eerste interessante bevindingen uit deze 'pilot scan' halen en aanbevelingen doen voor gemeenten in alle drie de clusters.

Ter onderbouwing zijn uit CBS data de meest voor de hand liggende achtergrondkenmerken van gemeenten verzameld om te achterhalen of de mate van data-volwassenheid beïnvloed wordt door de grootte van een gemeente of haar stedelijkheid. Door deze achtergrondkenmerken van alle Nederlandse gemeenten te vergelijken met de achtergrondkenmerken van de gemeenten die mee hebben gedaan aan deze scan komt naar voren dat dit een representatieve steekproef is. De patronen en inzichten uit deze analyse leveren dus een representatief beeld voor Nederland.

In tabel 3.2 zijn de beschrijvende statistieken van de achtergrondkenmerken van alle respondenten van de vragenlijst te zien zoals die zelf zijn gerapporteerd. Daarnaast tonen we de beschrijvende statistieken van de achtergrondkenmerken voor de deelnemende gemeenten.

---

<sup>4</sup> De uitkomsten uit de scan zijn dus grotendeels gebaseerd op één waarneming per gemeente. Sommige gemeenten hebben de vragenlijst meer dan één keer ingevuld. Dit zijn de gemeenten Breda (3x), Cuijk (2x), Den Bosch (2x), Grave (2x), Leidschendam-Voorburg (3x), Leudal (2x), Mill en Sint Hubert (2x), Rozendaal (2x), Woensdrecht (2x) en Zeist (2x). In deze gevallen hebben we deze waarnemingen alle meegenomen in de analyse.



Tabel 3.2: Beschrijvende statistieken achtergrondkenmerken

Totaal (N=127)		
<b>Geslacht</b>	<i>Vrouw</i>	25 (19,7%)
	<i>Man</i>	102 (80,3%)
<b>Leeftijdscategorie</b>	<i>18 – 24 jaar</i>	1 (0,8%)
	<i>25 – 30 jaar</i>	0 (0,0%)
	<i>31 – 40 jaar</i>	21 (16,5%)
	<i>41 – 50 jaar</i>	45 (35,4%)
	<i>51 – 60 jaar</i>	53 (41,7%)
	<i>&gt; 60 jaar</i>	7 (5,5%)
<b>Opleidingsniveau</b>	<i>VMBO/LBO/MAVO</i>	1 (0,9%)
	<i>MBO/HAVO</i>	5 (4,4%)
	<i>VWO</i>	1 (0,9%)
	<i>HBO</i>	77 (60,6%)
	<i>WO</i>	43 (33,9%)
<b>Inwoners</b>	<i>&lt; 15.000</i>	30 (23,6%)
	<i>15.001 - 25.000</i>	27 (21,3%)
	<i>25.001 - 35.000</i>	14 (11%)
	<i>35.001 – 60.000</i>	26 (20,5%)
	<i>60.000</i>	30 (23,6%)
<b>Stedelijkheid</b>	<i>&lt; 500 adressen per km<sup>2</sup></i>	21 (16,5%)
	<i>500 – 1.000 adressen per km<sup>2</sup></i>	38 (29,9%)
	<i>1.000 – 1.500 adressen per km<sup>2</sup></i>	25 (19,7%)
	<i>1.500 – 2.500 adressen per km<sup>2</sup></i>	10 (7,9%)
	<i>Meer adressen per km<sup>2</sup></i>	33 (26%)

Bron: Bewerking gegevens data maturity scan en CBS data.

### 3.3 Diepte-interviews met gemeenten

Nadat de vragenlijsten door de respondenten waren ingevuld is er met een aantal gemeenten contact opgenomen of ze bereid zouden zijn om een aanvullend diepte-interview te houden. CentERdata heeft in totaal 9 interviews afgenomen, deels in persoon en deels telefonisch.<sup>5</sup>

Een interview duurde ongeveer 45 minuten en had als doel om nadere inzichten te krijgen in de volwassenheid van gemeenten als het gaat om data science en datagedreven beleid, waar gemeenten tegen aan lopen en waar zij kansen in zien. De achtergrond van onze interviewpartners net als hun functie binnen een gemeente was divers, van informatie-manager, medewerker innovatie, teamleider analyse en intelligentie tot afdelingshoofd automatisering.

<sup>5</sup> De geïnterviewde gemeenten waren: Apeldoorn, Eersel, Enschede, Gorinchem, Helmond, Leidschendam-Voorburg, Schagen, Woerden en de samenwerkingsverband Waddeneilanden. De interviewvragen zijn te vinden in bijlage B.





Over het algemeen sloten de bevindingen van de geïnterviewde personen goed aan bij de bevindingen uit de maturity scan. We spraken gemeenten op alle niveaus van data-volwassenheid, qua kennis en kunde, qua (automatiserings-)processen en tools waarover een gemeente beschikt, maar ook qua mate van datagedreven beleid en bestuurlijke draagvlak van data science.

De geïnterviewde gemeenten zijn het eens dat het thema data en data science op alle aspecten van (de beleidsvoering van) de gemeente invloed zal hebben en dat het vooral negatieve (financiële en sociale) effecten op gemeenten zal hebben als ze niet in staat zijn om adequaat te versnellen op data science ontwikkelingen.

Als kansen worden merendeels mogelijke efficiëntieverbeteringen en meer 'beleid op maat' genoemd, maar ook gerichte decentralisatieprocessen en de bevordering van burgerparticipatie door data science.

Uiteraard zien de geïnterviewde gemeente ook uitdagingen, met name als het gaat om de inbedding van datagedreven beleid in zowel bestuurlijke processen als in beleidsprocessen. De veranderingen die gemeenten door data science en datagedreven beleid te wachten staan moeten breed gedragen worden door het gemeentebestuur en door de burgers. Het is daarom belangrijk dat de goede toepassingen worden gezocht voor data science en dat de inzichten daaruit data-analyse gepast vertaald worden naar gemeentebeleid.

Men benadrukt dat datagedreven beleid wel verbinding moet maken en moet houden met de emoties en de leefwereld van burgers om de maatschappelijke acceptatie te waarborgen. Hierbij is uiteraard ook rekening te houden met juridische aspecten en privacywaarborging.

De bevindingen uit de interviews komen, samen met de uitkomsten uit de maturity scan terug bij de discussiepunten en aanbevelingen.

Een uitgebreide discussie vanuit een ander perspectief vindt u in box 4. Ronald Leenes gaf tijdens de pilot als hoogleraar Regulering door technologie een masterclass aan geïnteresseerde gemeentemedewerkers.



#### **Box 4: Ronald Leenes (hoogleraar Regulering door technologie, TiU): Big data, big difference?**

Gemeenten en anderen worden momenteel geconfronteerd met op zijn minst twee grote omwentelingen: big data (en data science) en de invoering van de Algemene verordening gegevensbescherming (Avg). Voor velen is de eerste nog een beetje een vergezicht. Hoewel gemeenten over veel gegevens beschikken worden deze toch vooral gebruikt voor de klassieke doelen waarvoor ze zijn verzameld. Gegevens over de inwoners voor aanslag gemeentelijke belastingen, stempassen, onderwijs etc., gegevens over objecten voor aanslag OZB etc. Toepassingen waarbij patronen in data worden gezocht of waarbij gegevensbestanden aan elkaar worden gekoppeld zijn zeldzamer. Maar, de andere bijdragen en deze white paper lezend en deelnemers aan de VNG/KING pilot sprekend, is duidelijk dat er hard wordt nagedacht over data science toepassingen in gemeenteland.

De tweede omwenteling raakt velen in gemeenteland. De huidige Wet bescherming persoonsgegevens (Wbp), die de verwerking van persoonsgegevens reguleert wordt op 25 mei 2018 vervangen door de Algemene verordening gegevensbescherming. Dat heeft gevolgen voor het klassieke gebruik van persoonsgegevens door gemeenten, maar zeker voor de nieuwere data science toepassingen. Het is dus van belang aandacht te besteden aan de veranderingen die de Avg met zich meebrengt.

Veel blijft (vrijwel) hetzelfde. Persoonsgegevens zijn nog steeds 'alle informatie betreffende een geïdentificeerde of identificeerbare natuurlijke persoon'. De Avg rekt de grenzen daarvan op een paar punten wat op. Zo vallen bijvoorbeeld IP-adressen vanaf nu met zekerheid onder de categorie identificatoren (vreselijk woord voor datgene dat de koppeling tussen een individu en haar data mogelijk maakt). IP-adressen zijn daarmee vergelijkbaar geworden aan namen en identificerende nummers (denk aan BSN). Ook de beginselen achter de dataprotectieregels zijn gelijk gebleven: de verwerking moet behoorlijk zijn (eerlijk, rechtmatig en transparant), doelbepaald (welbepaalde, gerechtvaardigde doelen) en proportioneel (toereikend, ter zake dienend en noodzakelijk) zijn.

Nieuw zijn de introductie van onder meer de begrippen pseudonimering ((onomkeerbare) loskoppeling van de identificerende gegevens van 'inhoudelijke' gegevens) als middel om risico's voor 'betrokkenen' te verminderen, 'profilering' als risicovolle handeling die extra aandacht vereist, privacy by design, en de gegevensbeschermingseffectbeoordeling (met daarbij behorende gegevensbeschermingseffectbeoordelingsverantwoordelijke).

De algemene vernieuwing van de Avg is erop gericht om minder een 'checkbox' exercitie te zijn met melding aan de Autoriteit Persoonsgegevens voorafgaande aan verwerking, en meer (echte) verantwoordelijkheid bij de verantwoordelijke neer te leggen. Bij verwerkingen die risico's op kunnen leveren voor de rechten en fundamentele vrijheden van betrokkenen (denk burgers in uw gemeente), moeten deze risico's vanaf het begin van een ontwerptraject voor gegevensbescherming serieus genomen worden, in kaart worden gebracht en worden beperkt. Privacy by design (en by default) zijn de verwoording daarvan (vanuit de idee: voorkomen is beter dan genezen). Wanneer de risico's serieus zijn (bijvoorbeeld doordat met gevoelige gegevens wordt gewerkt, of dat de verwerking grote hoeveelheden mensen betreft, of er nieuwe technieken (bijvoorbeeld biometrie) worden gebruikt, zal een gegevensbeschermingseffectbeoordeling plaats moeten vinden die risico's en maatregelen systematisch in kaart brengt in een rapportage. De verantwoordelijke kan vervolgens door getroffen burgers of de Autoriteit persoonsgegevens worden aangesproken (accountability) wanneer er brokken vallen.



Op het vlak van data science heeft de Avg ook allerlei implicaties. Daar waar het bij klassieke gegevensbestanden doorgaans betrekkelijk eenvoudig was te beschrijven wat het doel van de verwerking is, is dat bij meer geavanceerde data science toepassingen mogelijk lastiger. Is van te voren wel duidelijk waarom je in data gegenereerd in het domein van de Omgevingswet gaat zoeken naar verbanden? Wie weet wat er allemaal uit gegevens te halen valt die worden gegenereerd door smartphones (in ieder geval waar het druk is in de stad)? Hoewel doelbepaling en doelbinding wellicht lastiger zijn dan voorheen, vereist de Avg wel dat doelen welbepaald moeten worden beschreven en gerechtvaardigd zijn. Hier ligt dus een uitdaging voor beleidsmakers en een ieder die zich binnen de gemeente met big data bezig houden.

Een andere uitdaging ligt op het vlak van geautomatiseerde besluitvorming. Op dat vlak kende de Wbp al regels, bijvoorbeeld dat mensen alleen onder voorwaarden onderworpen mogen worden aan geautomatiseerde besluitvorming. Dat regime is wat aangescherpt. Een wettelijke regeling dat geautomatiseerde besluitvorming toestaat is de koninklijke weg die nog niet altijd wordt genomen. Verder regelt de Avg dat indien geautomatiseerd beslissingen worden genomen over natuurlijke personen, er 'nuttige informatie over de onderliggende logica' moet worden verstrekt.

Vooralsnog lijkt dat voor de meeste toepassingen niet zo lastig. De ontwerpers snappen meestal wel wat de beslisregels zijn en kunnen daar ook uitleg over geven. Dat wordt wellicht lastiger wanneer (unsupervised) machine learning meer ingang gaat vinden. Moderne camerasystemen zijn in staat om te leren allerlei voertuigtypen te herkennen. Op basis daarvan is het denkbaar dat 'verdachte' busjes uit een verkeersstroom worden gepikt. Waarom een bepaald voertuig als verdacht is aangemerkt is bij dat soort (zelf)lerende systemen niet altijd even duidelijk aan te wijzen.

In de context van geautomatiseerde besluitvorming ligt ook een verplichting om onder meer te voorkomen dat de algoritmen discriminerende gevolgen hebben op factoren, zoals ras of etnische afkomst, politieke overtuiging, godsdienst of levensbeschouwelijke overtuigingen, lidmaatschap van een vakbond, genetische of gezondheidsstatus, of seksuele gerichtheid. Voor directe discriminatie (als die bijvoorbeeld zou plaatsvinden op basis van registratie van etniciteit) lijkt dat evident en ook haalbaar (gebruik die gegevens omtrent etniciteit niet). Maar voor indirecte discriminatie is dat lastiger; er zijn concentraties van mensen met bepaalde kenmerken in bepaalde delen van gemeenten waardoor bijvoorbeeld postcode kan leiden tot indirecte discriminatie op etniciteit of godsdienst.

Er ligt een nieuw wetgevingskader voor de verwerking van persoonsgegevens. Dat kader is nog niet helemaal uitgekristalliseerd (er zal nog veel uitleg moeten komen van bijvoorbeeld de Autoriteit Persoonsgegevens, maar ook van VNG en KING) en op punten wellicht ook niet (helemaal) toegesneden op de realiteit van big data, maar we moeten het er wel mee doen. Het is van belang kennis te nemen van de nieuwe regels en daarnaar te handelen, al was het alleen maar omdat er ook serieuze (nieuwe!) sancties staan op overtreding van de regels. Een boete tot maximaal 10 miljoen voor het overtreden van 'verantwoordelijkheidseisen' of zelfs 20 miljoen voor het overtreden van de basisbeginselen zal menig wethouder als onaangenaam nieuws ervaren.



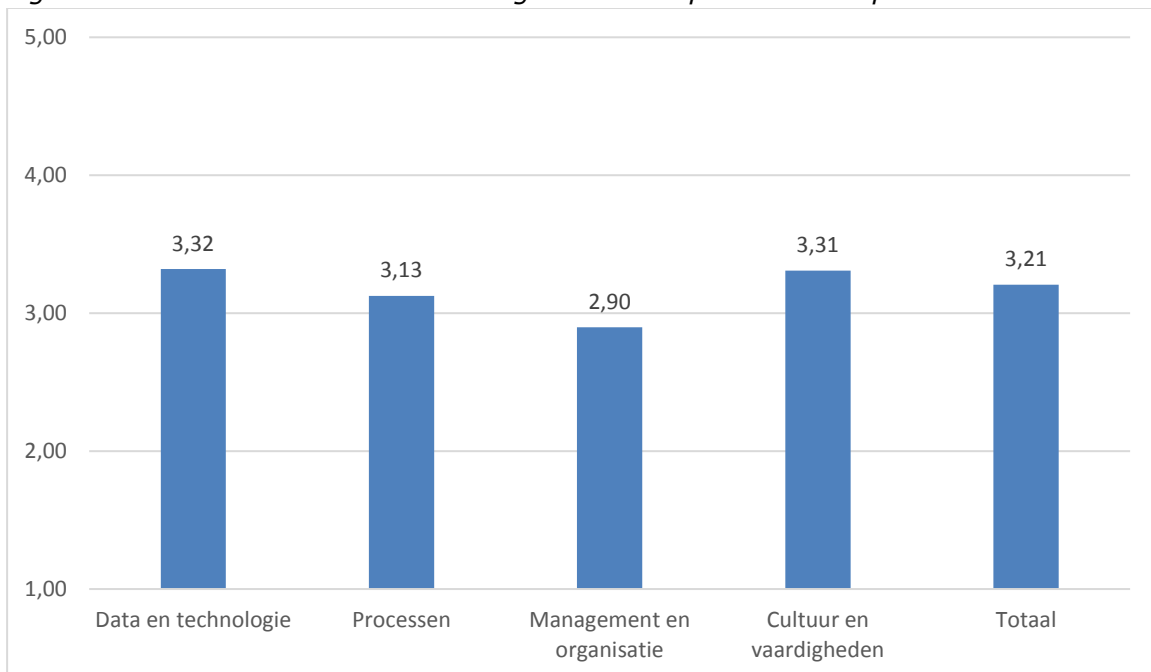
## 4 Uitkomsten maturity scan

Zoals in hoofdstuk 3.1 geschetst was het uitvoeren van de data maturity scan ook een experiment in de zin dat er een nieuw instrument voor het eerst ingezet is bij gemeenten om een poging te doen om de huidige datavolwassenheid in kaart te brengen. In dit hoofdstuk lichten we toe welke inzichten al dan niet vergaarde kunnen worden op basis van deze 'pilot maturity scan'.

### 4.1 Een eerste algemene indruk

In een eerste analyse hebben we gekeken hoe de vragen op de vier componenten gemiddeld zijn beantwoord door de respondenten. Op die manier is te zien in welke van de vijf fasen de gemeenten zich naar eigen inzicht bevinden. Voor de 127 gemeenten komen we op basis van een vijf-punt schaal tot de gemiddelde scores die te zien zijn in figuur 4.1.

*Figuur 4.1: Gemiddelde score van de gemeenten op de vier componenten*



Over het algemeen geven de gemeenten in de maturity scan aan dat ze gemiddeld op een gestandaardiseerd niveau zitten. Op drie van de vier componenten scoren de gemeenten gemiddeld boven de drie, alleen de component 'Management en organisatie' scoort iets lager (op het randje van 'gecontroleerd').

Als we kijken naar de scoringsniveaus van de vijf fasen (zie bijlage A) betekent dit dat de 127 gemeenten al over data van redelijk goede kwaliteit beschikken, dat het verzamelen van cruciale informatie in hoge mate geautomatiseerd is, en dat er data warehouses met actuele gegevens bestaan. De gemeenten rapporteren zelf over een algemeen en gedragen datamodel te beschikken, en over initiatieven voor organisatiebreed technologie-management. Op een gestandaardiseerd niveau hebben gemeentemedewerkers over het



algemeen toegang tot de data die ze nodig hebben en stuurt de organisatie redelijk feitengebaseerd.

## 4.2 Indeling in clusters

Zoals geschetst in 3.1 hebben we door middel van factoranalyse vervolgens gekeken welke componenten en subcomponenten van de vragenlijst het meest informatief zijn voor de maturity scan onder gemeenten. Uit alle vragen zijn acht onderdelen, ook schalen of factoren genoemd, als meest betrouwbaar naar voren gekomen. Uit de vragen behorend tot deze schalen zijn dus de meeste inzichten over de datavolwassenheid van gemeenten af te leiden.<sup>6</sup> De acht schalen, of factoren, die we in de verdere analyse hebben meegenomen en waarvoor we dus resultaten presenteren zijn:

**Schaal 1: Gebruik van data is onderdeel strategie van de gemeente/ datagedrevenheid wordt van bovenaf gefaciliteerd**

**Schaal 2: Tooling en capaciteit voor datagedrevenheid is aanwezig**

Schaal 3: Zelfverzekerd en plezier data-analyse

**Schaal 4: Kennis en kunde aanwezig binnen gemeente voor data science**

**Schaal 5: Procedures en protocollen**

Schaal 6: Werken met data geeft aanzien

Schaal 7: Datagebruik hoort bij werkzaamheden

Schaal 8: Plezier en nut werken met data

De insteek van deze acht schalen is nogal verschillend. Deels worden vragen gesteld die over de gemeente als geheel gingen, deels vragen die over een persoonlijke inschatting of voorkeur gingen. We hebben ons gefocust op de algemene schalen, schalen 1, 2, 4 en 5, deze zijn vetgedrukt in bovenstaande lijst.

Deze keuze is in deze pilot scan bewust gemaakt en afgestemd op de manier van uitzetten van de vragenlijst. Zoals in hoofdstuk 3.2 beschreven hebben we meestal maar één observatie per gemeente, omdat de enquête meestal maar één keer is ingevuld. Daarnaast weten we niet wie de vragenlijst heeft ingevuld; een individu of een team of een afdeling. Verder is niet bekend welke functie de respondent(en) had(den). Daarom kunnen we geen uitspraken doen over de representativiteit van de antwoorden binnen een gemeente, vooral bij de persoonlijke schalen, en wegen deze schalen daarom niet mee bij de inschatting van het volwassenheidsniveau.

De gebruikte schalen leveren alsnog interessante informatie, niet alleen over de randvoorwaarden voor data maturity, maar ook over hoe een gemeente het thema data science het beste aan kan pakken en hoe medewerkers binnen de gemeente denken over data science. De pilot scan is dan ook een succesvolle poging om te komen tot eerste inzichten met betrekking tot data science maturity van gemeenten.

---

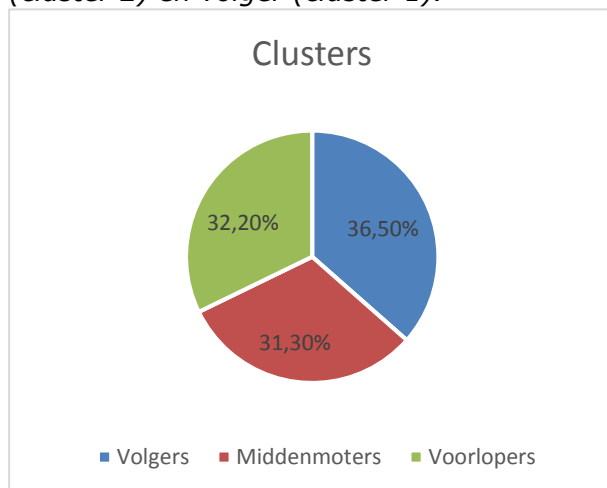
<sup>6</sup> Een aantal van de items (vragen) is niet meegenomen in de factoranalyse, aangezien deze items erg weinig en lage significante correlaties hadden met de andere items. Er waren geen items die te hoog correleren met bepaalde andere items, dus waar sprake was van singularities.



Uit bovenstaande vier schalen zijn vervolgens clusters van gemeenten gedestilleerd om te komen tot een indeling in 3 clusters. Bij de clustering zijn alleen gegevens meegenomen die betrekking hebben op een enkele gemeente. Indien mensen de vragenlijst voor een samenwerkingsverband hebben ingevuld zijn deze antwoorden dus toegekend aan alle gemeenten uit dit samenwerkingsverband. Als meerdere personen daarentegen de vragenlijst hebben ingevuld voor dezelfde gemeente zijn hun antwoorden vervolgens gemiddeld voor de clusteranalyse. Uiteindelijk zijn daarom in de clusteranalyse gegevens meegenomen die betrekking hebben op 115 verschillende gemeenten.<sup>7</sup>

Cluster 3 omvat 37 gemeenten die het meest volwassen zijn als het gaat om data science en fact-based beleid; deze groep noemen we vervolgens de voorlopers. Tot cluster 2 behoren 36 gemeenten die niet meer aan het begin staan op hun weg naar data maturity, maar die wel ook nog een eind te gaan hebben. Dit cluster zijn daarom de middenmoters. Uiteindelijk zitten 42 gemeenten in cluster 1, die nog aan het begin staan van een volledige integratie van data, die informatie nog niet zien als een strategisch succesfactor en die nog geen cultuur van meetbaarheid en toerekenbaarheid hebben waar data innovaties drijft. Deze groep noemen we de volgers.

*Figuur 4.2.1: Procentuele verdeling over de clusters: voorloper (cluster 3), middenmoter (cluster 2) en volger (cluster 1).*



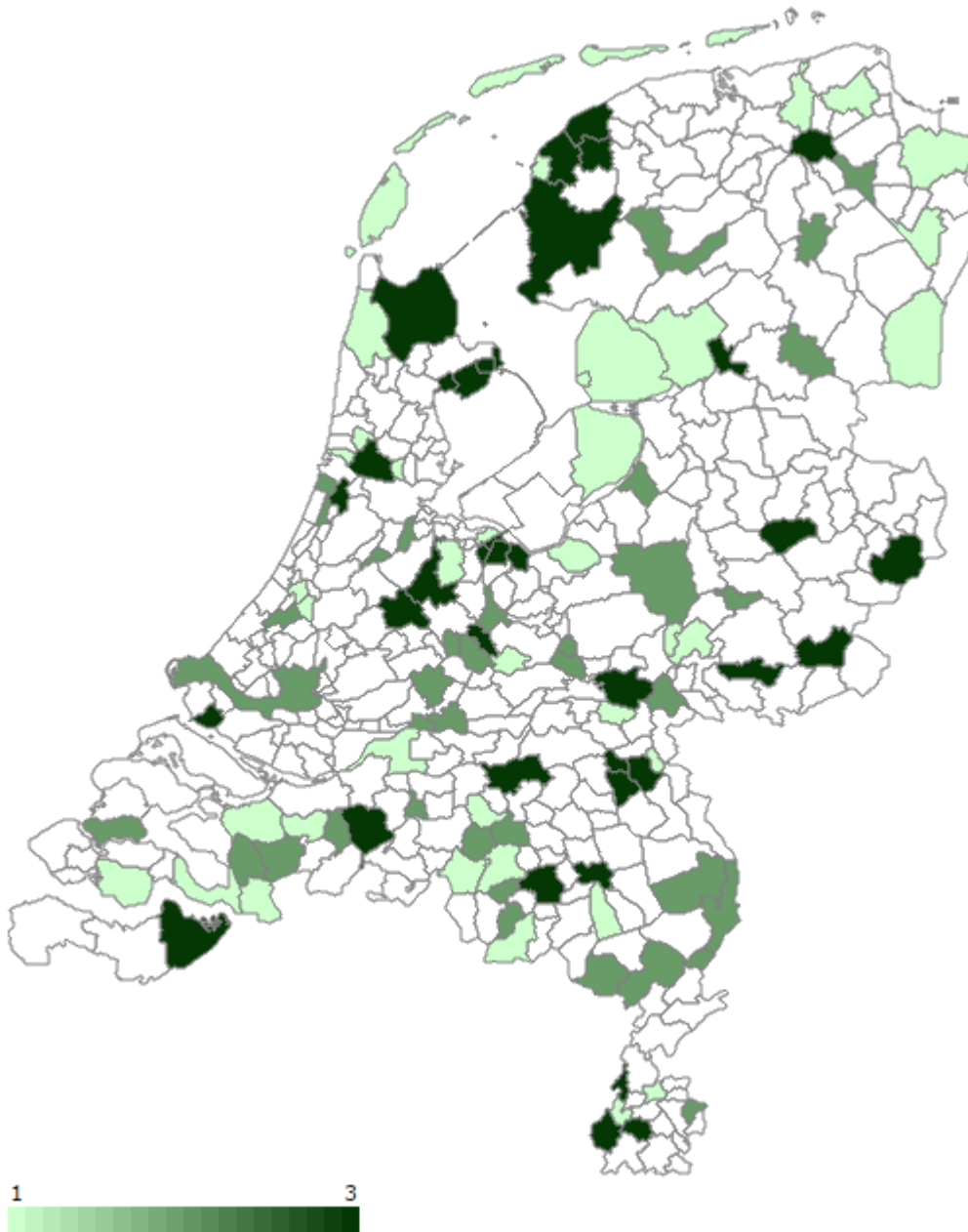
In figuur 4.2.1 ziet u de procentuele verdeling van alle gemeenten over de drie clusters. De groep volgers (blauw) is iets groter dan de middenmoters (rood) en voorlopers (groen), maar zoals te zien zijn de gemeenten redelijk gelijk verdeeld over alle clusters. In bijlage C is de lijst van de 115 gemeenten te zien en hun indeling in een van de drie clusters.

Als we naar de indeling in de drie clusters kijken voor alle gemeenten op de kaart van Nederland ontstaat het beeld uit figuur 4.2.2.

<sup>7</sup> Zie o.a. voetnoot 4 voor de lijst met gemeenten die de enquête meerdere malen hebben ingevuld.



Figuur 4.2.2: Indeling in clusters: voorlopers (cluster 3), middenmoters (cluster 2) en volgers (cluster 1).



Maar in hoeverre verschillen de gemeenten die tot de drie clusters behoren nu eigenlijk met betrekking tot bepaalde achtergrondkenmerken? We hebben voor deze analyses gegevens van het CBS verzameld over het aantal inwoners van een gemeente en de mate van stedelijkheid van een gemeente (gemeten in adressen per km<sup>2</sup>).<sup>8</sup>

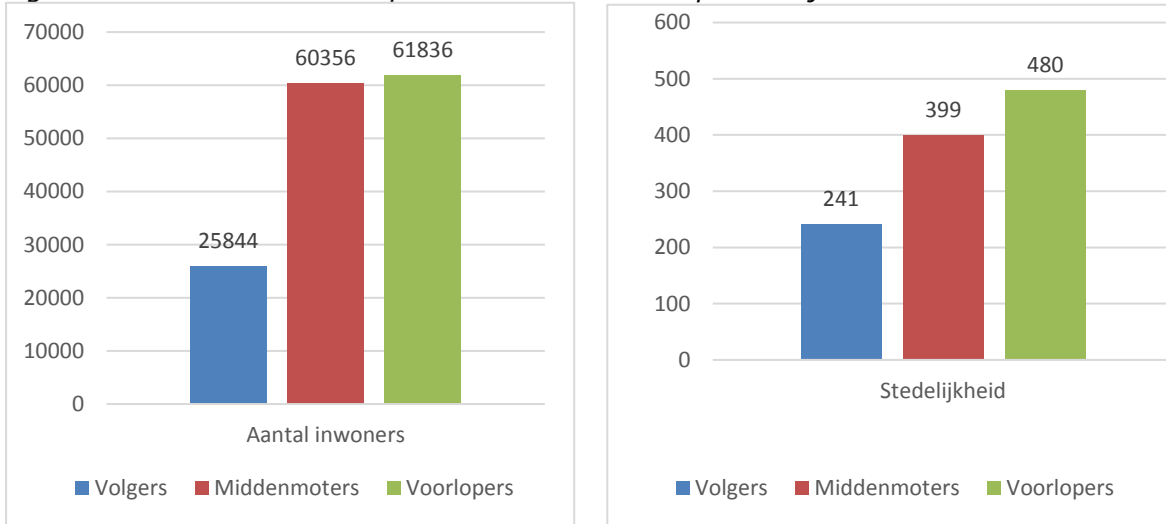
In een eerste stap is gekeken wat het gemiddeld aantal inwoners en het gemiddelde mate van stedelijkheid per cluster is. Zowel met betrekking tot het aantal inwoners als de mate van stedelijkheid zien we zoals te verwachten een stijging over de clusters heen. Onder de

<sup>8</sup> De maatstaven voor stedelijkheid zijn gebaseerd op de CBS-indeling in vijf categorieën, zie ook <https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen?tab=s#id=stedelijkheid--van-een-gebied-->



volgers vinden we duidelijk kleinere gemeenten dan onder de middenmoters of voorlopers. Het verschil tussen cluster 1 en de clusters 2 & 3 is dan ook significant, dat wil zeken aantoonbaar, betekenisvol en betrouwbaar. Ook de mate van stedelijkheid verschilt, waarbij de gemeenten met hogere niveaus van data maturity ook meer stedelijk zijn. Een significant verschil is hierbij alleen te vinden tussen de volgers en de voorlopers, niet tussen de andere clusters. Deze informatie is te zien ins figuur 4.2.3.

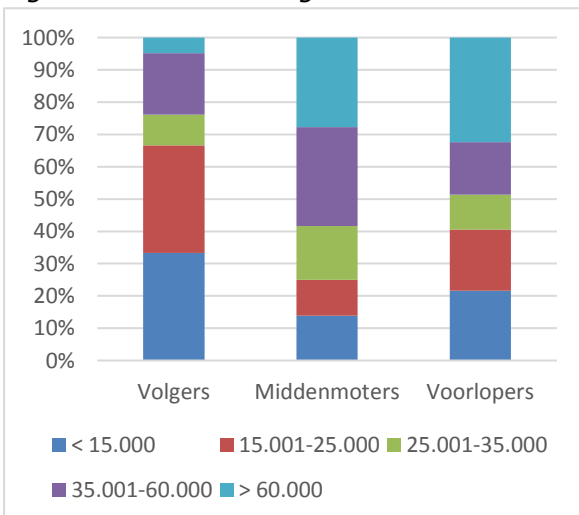
*Figuur 4.2.3: Effect cluster op aantal inwoners en op stedelijkheid*



Uiteraard is er een verband tussen stedelijkheid en gemeentegrootte. Meestal zijn kleinere gemeenten minder stedelijk, bijvoorbeeld zijn in gemeenten met minder dan 10 duizend inwoners geen zeer sterk stedelijke gedeelten te vinden, terwijl in de gemeenten met 250 duizend inwoners of meer (de vier grootste gemeenten van Nederland) juist drie kwart van de mensen in een zeer sterk stedelijk gebied woont. Er zijn echter ook in grote steden gebieden te vinden die als landelijk, ofwel niet-stedelijk gedefinieerd zijn.

Aan de andere kant zijn weinig stedelijk gemeenten vaak plattelandsgemeenten, waarbij juist de grotere plattelandsgemeenten vaker samenwerken met andere gemeenten, bijvoorbeeld in het samenwerkingsverband P10, waardoor er enigszins schaafeffecten te bereiken zijn, maar de mate van stedelijkheid uiteraard niet toeneemt.

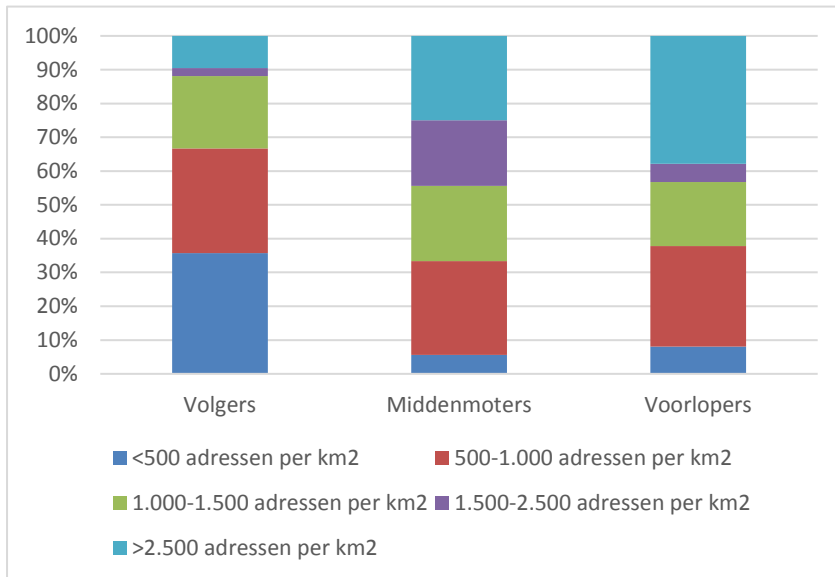
*Figuur 4.2.4: Verdeling aantal inwoners binnen de clusters*



Door een iets andere benadering wordt dit verband nog inzichtelijker (figuur 4.2.4). Als we kijken naar de verdeling van de vijf inwonersklassen binnen de clusters is te zien dat de meerderheid van gemeenten in de groep volgers zeer kleine of kleine gemeenten zijn, terwijl de meerderheid van gemeenten onder de voorlopers juist zeer grote gemeenten zijn.

Dezelfde benadering voor stedelijkheid is te zien in figuur 4.2.5 onderaan. Wederom zijn de minst stedelijke gemeenten vooral te vinden onder de volgers,





terwijl de zeer stedelijke gemeenten vooral in de groep voorlopers te vinden zijn. We zien ook dat deze indeling niet 'perfect' is. Er zijn ook wel zeer stedelijke gemeenten te vinden die tot de volgers behoren, terwijl ook zeer weinig stedelijke gemeenten tot de voorlopers behoren.

#### 4.2.5: Verdeling mate van stedelijkheid binnen de clusters

### 4.3 De invloed van de clusters op data maturity van gemeenten

Het voorgaande hoofdstuk heeft geleid tot een eerste indruk van de verschillende groepen van gemeenten. In het vervolg kijken we in hoeverre de gemeenten uit de drie clusters ook verschillen als het gaat om datavolwassenheid. Hiervoor maken we zoals eerder beschreven onderscheid naar de algemene vragen die betrekking hebben op een gemeente en de persoonlijkere vragen. In deze paragraaf lichten we eerst de relatie van de clusters en achtergrondkenmerken met de algemene vragen toe, voordat we in 4.4 vervolgens ingaan op de relatie met de persoonlijke vragen.

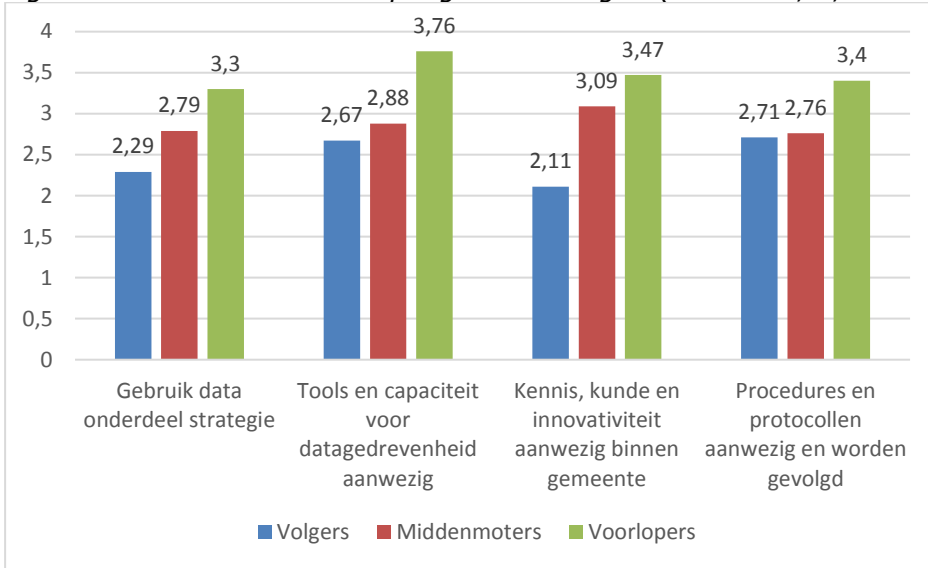
In figuur 4.3.1. is te zien dat er significante verschillen zijn op de scores op de vier algemene schalen. De voorlopers uit cluster 3 scoren over het algemeen significant hoger dan de middenmoters en volgers. De middenmoters uit cluster 2 scoren over het algemeen significant hoger dan de volgers uit cluster 1. Dat betekent dat er daadwerkelijk een grote invloed van de clusters is op de mate van datavolwassenheid, zowel als het gaat om de randvoorwaarden als om het 'doen'.

Alleen voor de schaal 'procedures en protocollen aanwezig en worden gevolgd' vinden we geen significante verschillen tussen de middenmoters en de volgers, hier liggen voor alle twee groepen nog kansen ter verbetering. De voorlopers scoren ook hier significant hoger en zijn er blijkbaar al in staat geweest om de noodzakelijke procedures en protocollen voor datagedreven beleid in te voeren en te handhaven.<sup>9</sup>

<sup>9</sup> Let op: Het is niet vreemd dat de clusters zo duidelijk van elkaar verschillen op deze schalen aangezien deze schalen gebruikt zijn om clusters te vormen die zoveel mogelijk van elkaar verschillen.



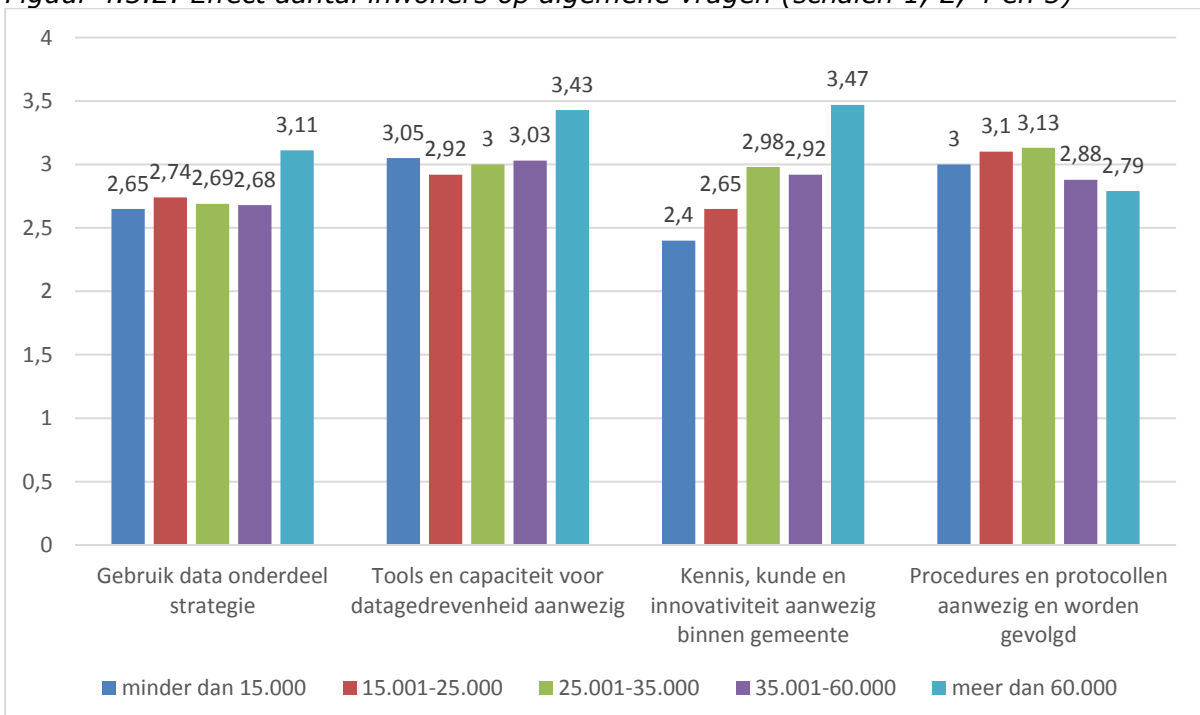
Figuur 4.3.1: Effect cluster op algemene vragen (schalen 1, 2, 4 en 5)



Vervolgens duiken we dieper in de achtergrondkenmerken van de gemeenten om te achterhalen of het aantal inwoners of de mate van stedelijkheid verschil maakt met betrekking tot de inschatting van datavolwassenheid van een gemeente in het algemeen.

In figuur 4.3.2 is een indeling in vijf verschillende categorieën van inwoneraantallen te zien. Deze indeling is gemaakt voor de 127 gemeenten die de maturity scan hebben ingevuld. Er is op drie van de vier schalen verschil te zien, vooral de kleine gemeenten met minder dan 15.000 inwoners geven vaker een lagere score dan de grote gemeenten met meer dan 60.000 inwoners. We zien geen significante verschillen op het onderdeel 'Procedures en protocollen aanwezig en worden gevolgd' tussen de gemeenten.

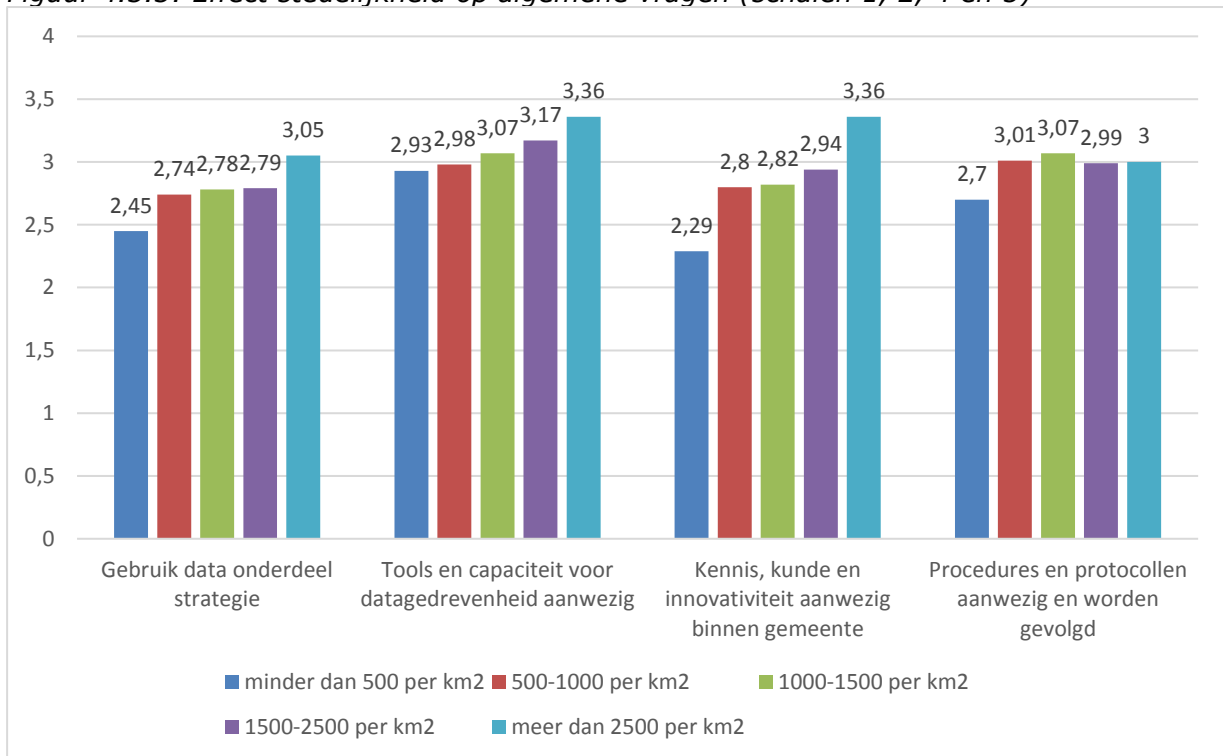
Figuur 4.3.2: Effect aantal inwoners op algemene vragen (schalen 1, 2, 4 en 5)





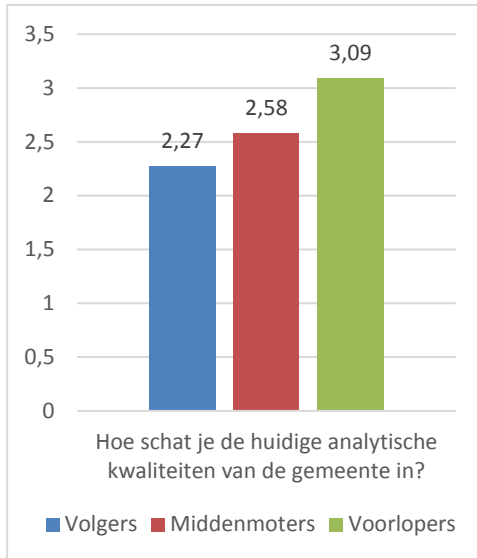
Als we kijken of de mate van stedelijkheid verschil maakt voor de scores op de algemene vragen uit de maturity scan is hier weer meer variatie tussen de gemeenten te zien. Zowel met betrekking tot 'gebruik data onderdeel strategie' als tot 'kennis, kunde en innovativiteit aanwezig binnen gemeente' zijn er significante verschillen tussen de clusters. Vooral niet stedelijke gemeenten (met minder dan 500 inwoners per km<sup>2</sup>) scoren aantoonbaar lager dan meer stedelijke gemeenten, vooral dan de grootste gemeenten (met meer dan 2.500 inwoners per km<sup>2</sup>). Op de andere twee schalen zijn daarentegen geen aantoonbare verschillen tussen de gemeenten te zien, zoals te zien in figuur 4.3.4.

Figuur 4.3.3: Effect stedelijkheid op algemene vragen (schalen 1, 2, 4 en 5)



Vervolgens zijn in de maturity scan vragen aan bod gekomen waarbij de respondenten rechtstreeks een inschatting moesten maken van de data science maturity van hun eigen gemeente. Zo is direct gevraagd hoe men de analytische kwaliteiten van de (medewerkers van de) gemeente inschat. Ook deze vraag was op een vijf-punt schaal te beantwoorden, beginnend bij 'uitstekend'; 'bovengemiddeld'; 'afdoende'; 'onvoldoende' en 'minimaal'.<sup>10</sup>

<sup>10</sup> Vanuit survey-technische overwegingen zijn de antwoordcategorieën op de vragenlijst hierbij omgedraaid. Uiteraard is in de analyses met de laagste score '1' begonnen voor een negatieve inschatting van de analytische kwaliteiten en geeft score '5' uitstekende analytische kwaliteiten weer.



Figuur 4.3.4: Effect clusters op vraag 19

In figuur 4.3.4. is te zien dat bij de beantwoording van deze vraag weer significante verschillen met betrekking tot de clusters naar voren komen. De gemeenten in cluster 1 scoren aantoonbaar lager dan de gemeenten in de cluster 3, maar er zijn geen significante verschillen tussen cluster 1 en 2. De volgers scoren naar eigen inschatting wel 'onvoldoende' op de analytische kwaliteiten, terwijl de gemeenten uit cluster 2 en 3 naar eigen mening 'afdoende' scoren. De voorlopers scoren hierbij nog significant hoger dan de middenmoters.

Net als bij de relatie van de clusters met de scores op alle persoonlijke vragen komt er weer een aantoonbaar verschil tussen de voorlopers en de andere gemeenten naar voren. Deze groep gemeenten heeft blijkbaar ook als het gaat om de (inschatting van de) analytische kwaliteiten van haar medewerkers al meerdere stappen kunnen zetten op de weg naar datavolwassenheid.

Kijken we vervolgens of voor de beantwoording van deze vraag ook de achtergrondkenmerken van de betreffende gemeente van invloed zijn, zo zien we dat een hoger aantal inwoners ertoe leidt dat een beduidend hogere score gegeven wordt. Grote gemeenten met meer dan 60.000 inwoners scoren een 3,11 ('afdoende'), terwijl kleinere gemeenten hooguit een 2,74 scoren en dus ergens tussen 'onvoldoende' en 'afdoende' scoren. Dit is te zien in figuur 4.3.5 op de volgende pagina.

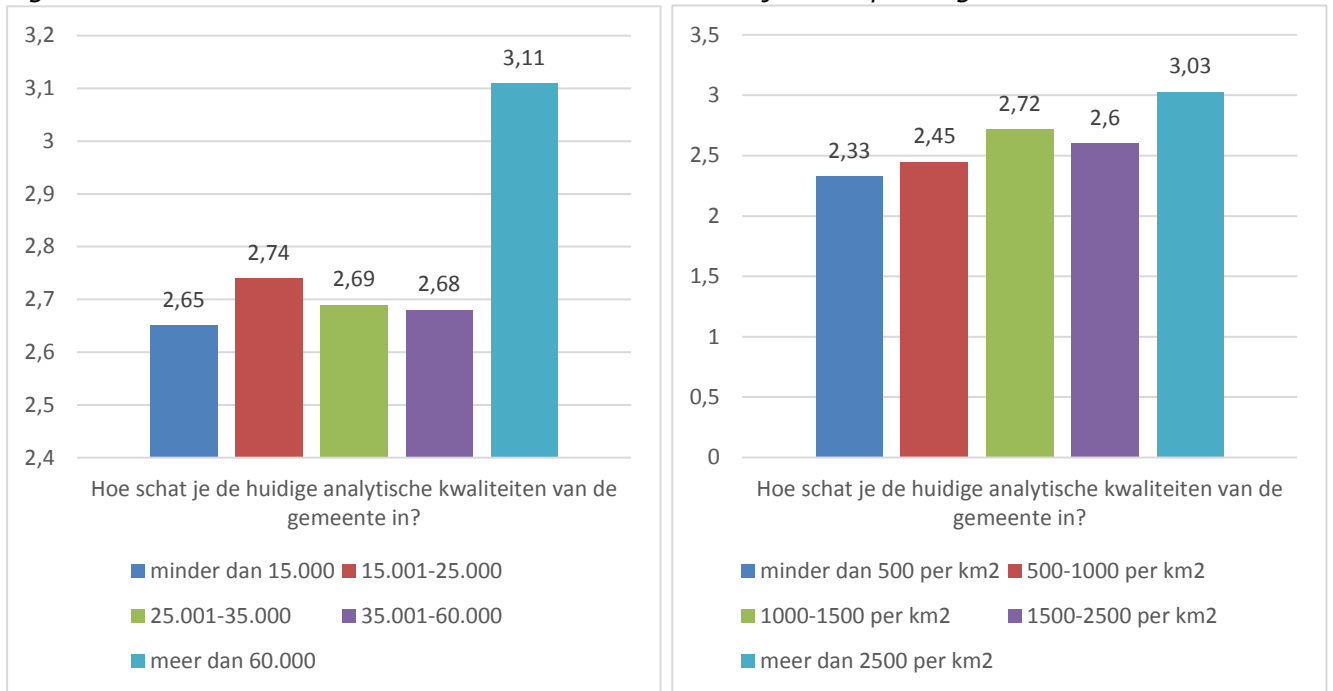
Daarnaast scoren met name de gemeenten met een hoge mate van stedelijkheid (meer dan 2.500 adressen per km<sup>2</sup>) hoger dan de rest en verschillen significant van gemeenten in de laagste twee categorieën van stedelijkheid. Gemeenten uit deze twee minst stedelijke categorieën schatten de analytische kwaliteiten ook duidelijk 'onvoldoende' in.

Hier zou ook sprake kunnen zijn van een relatie tussen een bepaalde schaalgrootte en voldoende analytische kwaliteiten. In dit geval zouden de scores voor een groot deel bepaald worden door gemeenten met O&S afdelingen. Om dit te achterhalen hebben we gekeken of dit te maken kan hebben met het hoeveel gemeenten uit onze steekproef lid zijn van de "Vereniging voor Statistiek en Onderzoek".

Van de 127 gemeenten die aan het onderzoek hebben deelgenomen beschikken 28 gemeenten over een O&S afdeling, waarbij de ruime meerderheid inderdaad tot de zeer grote gemeenten behoort (21 van de 28 hebben meer dan 60.000 inwoners; de resterende 7 gemeenten met O&S afdeling zijn 'grote gemeenten' met 35.001 – 60.000 inwoners). De relatie is iets zwakker voor stedelijkheid, 16 van de 28 gemeenten behoren tot de meest stedelijke, maar de resterende 12 gemeenten met een O&S afdeling zijn verdeeld over de vier andere stedelijkheids categorieën. Over het algemeen lijkt het er op dat gemeenten met een O&S afdeling inderdaad vaak van grotere schaalgrootte zijn. Het zou dus kunnen dat er over het algemeen voor bepaalde analytische kwaliteiten een grotere schaalgrootte nodig is, en dat data science inderdaad voort zou kunnen komen uit de traditionele statistiek.

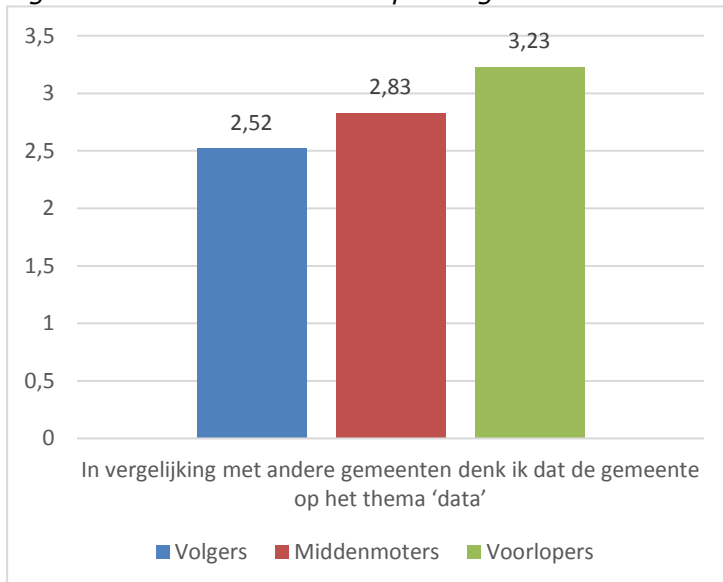


Figuur 4.3.5: Effect aantal inwoners en mate van stedelijkheid op vraag 19



De volgende vraag uit de enquête heeft een vergelijkbare strekking, maar is nog algemener gesteld dan de vorige. Hier zijn de respondenten gevraagd om een inschatting te maken hoe de eigen gemeente scoort 'op het thema data' in vergelijking met andere gemeenten. De mogelijke antwoordcategorieën hierbij waren dat de eigen gemeente 'ver achterloopt', 'achterloopt', 'gelijk opgaat', 'voorloopt' of 'ver voorloopt'.

Figuur 4.3.6: Effect cluster op vraag 20



In zeker zin is dit de vraag naar de data maturity van een gemeente, maar dan rechtstreeks gevraagd aan de medewerkers van deze gemeente. Zoals in figuur 4.3.6 te zien beantwoorden gemeenten in cluster 1 deze vraag opnieuw verschillend van de gemeenten uit de andere twee clusters. Met een score net boven de 2,5 geven de respondenten uit deze groep aan dat de eigen gemeente op het thema data 'achterloopt' in vergelijking met andere gemeenten. De gemeenten in dit cluster kunnen dus vooral nog veel

winst behalen op het gebied van datavolwassenheid, er is ruimte voor een 'herkansing'. Dit is ook een duidelijk lager score dan bij de middenmotors, laat staan bij de voorlopers. Maar ook tussen de laatste twee groepen vinden we nog aantoonbare verschillen en zijn vooral de voorlopers overtuigd dat hun gemeente 'voorloopt'.



Kijken we naar de rol van de achtergrondkenmerken aantal inwoners en stedelijkheid, dan vinden we ook op vraag 20 dezelfde patronen als op de vorige vraag. Grotere gemeenten (met meer dan 60.000 inwoners) geven een score die hoger dan 3,1 is en significant verschilt van alle andere gemeenten. Daarnaast zijn er significante verschillen tussen alle clusters als we kijken naar de invloed van stedelijkheid. De scores van alle vijf groepen liggen weliswaar iets hoger dan bij de vorige vraag, de minst stedelijke groep scoort met een 2,48 net tussen 'achterloopt' en 'gelijk opgaat'. De uitkomsten van deze vraag liggen zo dicht bij de uitkomsten op vraag 19 dat we hier geen figuur voor tonen.

Gerelateerd aan de rechtstreekse vraag naar de datavolwassenheid van de eigen gemeente is in vraag 21 gevraagd of de gemeente over een centraal team beschikt dat organisatiebreed data-analyses uitvoert. Hier was alleen een 'ja' of 'nee' mogelijk en het antwoordpatroon van cluster 1 verschilt significant van dat van clusters 2 en 3. Volgers beschikken dus over aantoonbaar minder teams dan middenmoters of voorlopers, alhoewel er maar een klein (niet significant) verschil te zien is tussen de middenmoters en de voorlopers.

Vervolgens zijn verschillende algemene vragen gesteld naar de inschatting van de respondenten op trends en generieke ontwikkelingen, bijvoorbeeld of men van mening is dat het gebruik van data(-analyse) toegevoegde waarde heeft voor de gemeente, of dat de invloed van sociale media op het beleid van de gemeente de komende jaren verder zal toenemen. Op deze vragen 17, 18, 21 en 25 vinden we geen effecten van de clusters, zowel de volgers, als de middenmoters en de voorlopers zijn het eens en antwoorden (bijna) allemaal met 'ja'. Deze uitkomsten zijn niet getoond, omdat er geen significante bevindingen naar voren kwamen.

#### **4.4 De invloed van de clusters op data maturity van gemeentemedewerkers**

In de eerdere paragrafen zijn de effecten in kaart gebracht van de clusters en achtergrondkenmerken op de algemene vragen die betrekking hebben op de werkwijze van een gemeente en haar datagedreven beleid. In deze paragraaf laten we zien of deze verschillen ook naar voren komen als we naar de schalen met persoonlijkere vragen kijken.

Paragraaf 4.2 lichtte toe dat de vragen in deze vier schalen niet gebruikt zijn om tot de clusterindeling te komen, omdat ze naar persoonlijke inschattingen vragen. Deze keuze is in deze pilot scan bewust gemaakt en afgestemd op de manier van uitzetten van de vragenlijst. Het is echter wel interessant om ook deze uitkomsten te tonen, omdat we deze antwoorden in zekere zin kunnen interpreteren als de 'datavolwassenheid van gemeentemedewerkers'.

Zoals uit de inleidende stukken blijkt is het namelijk juist belangrijk dat de *mindset* verandert van gemeentemedewerkers en alle personen die met data (science) en datagedreven beleid te maken hebben. Bovendien is het wanneer men met data werkt van belang dat men naast de competenties en technische vaardigheden voldoende in staat is om de goede vragen te stellen om daadwerkelijk meerwaarde uit de data te halen en het grote geheel niet uit het oog te verliezen. Daarom zijn er vragen gesteld over de zelfverzekerdheid en het plezier dat men heeft in data-analyse, of datagebruik tot de



werkzaamheden van een gemeentemedewerker hoort, of werken met data aanzien geeft en of men werken met data als leuk en nuttig beschouwt.

De persoonlijke vragen geven daarom een goed beeld van de 'zachte kant' van datavolwassenheid. Vooral de schalen 3 en 8 zijn belangrijke randvoorwaarden om snel tot maturity te komen, want zonder geschikte, geïnteresseerde en creatieve medewerkers is het heel moeilijk voor een gemeente om door te groeien naar het hoogste niveau van datavolwassenheid.

Deze schalen bieden daarnaast enige onderbouwing van de uitkomsten op de algemene schalen. Zowel op de gemeentevragen als op de persoonlijkeren vragen komt hetzelfde beeld naar voren. Ook individuen scoren lager op de vragen indien ze in een 'volgergemeente' werken dan medewerkers van middenmoters en voorlopers. Het enige verschil tussen deze soorten vragen is dat er geen significant verschil tussen middenmoters en voorlopers te vinden is, maar wel zelfde tendens.

Kijken we eerst naar de effecten van de clusters op de persoonlijke vragen dan zien we enig verschil, vooral tussen cluster 1 in vergelijking met de twee andere clusters. Volgers geven dus significant vaker een lagere score op drie van de vier schalen dan de middenmoters of de voorlopers. Alleen op de vraag 'werken met data geeft aanzien' zien we geen verschillen. Zoals te zien in figuur 4.4.1 zijn hier zowel de voorlopers, als de middenmoters en de volgers het eens en geven een gemiddelde score van 3.

Figuur 4.4.1: Effect cluster op 'persoonlijke' vragen (schalen 3, 6, 7 en 8)



Er zijn geen verschillen aantoonbaar met betrekking tot de achtergrondkenmerken grootte van een gemeente (gemeten in aantal inwoners) of stedelijkheid (gemeten in adressen per km<sup>2</sup>).



## 4.5 De invloed van achtergrondkenmerken op data maturity van gemeentemedewerkers

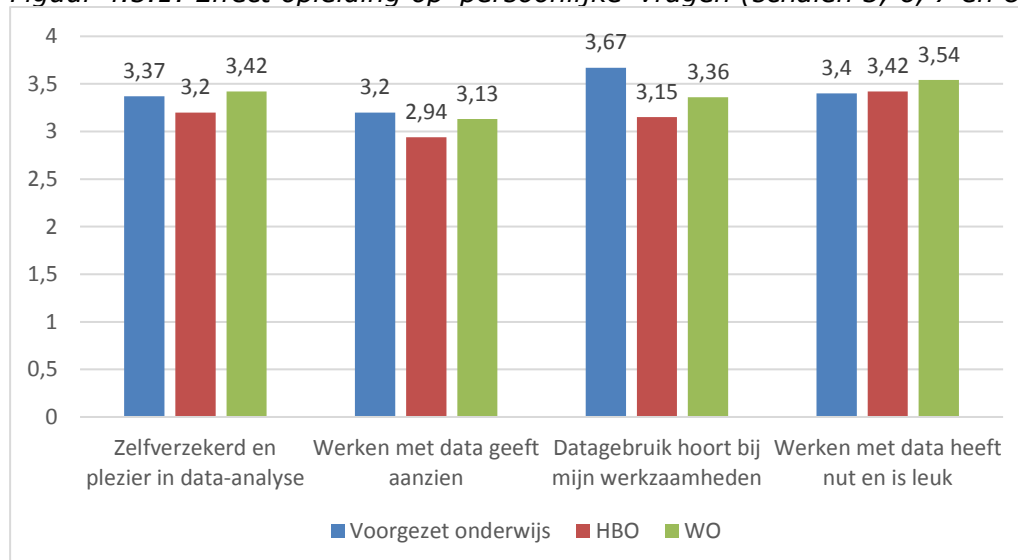
Tenslotte hebben we gekeken of de persoonlijke achtergrondkenmerken van de respondenten invloed hebben op hun antwoordpatronen. Hiervoor gebruiken we de gegevens die aan het begin van de enquête verzameld zijn over de leeftijd, opleiding en geslacht van een respondent.

We vinden dat zowel leeftijd als geslacht geen verschil maken als het gaat om de scores op de persoonlijke schalen. De opleiding van een respondent doet er echter wel toe. Zoals in figuur 4.5.1 te zien geven respondenten met een voortgezet onderwijs diploma altijd hogere scores dan respondenten met een hbo-opleiding en scoren meer in de buurt van respondenten met een wo-opleiding.

Een aantoonbaar verschil zien we alleen als het gaat om een antwoord op de vraag 'Werken met data geeft aanzien'. Hier zien we dat de score die de groep hbo-respondenten geeft significant lager is dan de andere twee groepen respondenten. Mogelijk heeft dit te maken met verschillen in de werkzaamheden en taken van personen met een verschillende opleidingsachtergrond. Wellicht voeren hbo-opgeleide gemeentemedewerkers relatief vaker taken op het gebied van *Business Intelligence* (BI) uit die wel met (veel) data te maken hebben, maar niet leiden tot aanzien binnen de gemeenten. Personen met een wo-opleiding worden mogelijk al vaker ingezet voor 'echte' data science opdrachten wat wel tot meer aanzien onder de collega's leidt.

Hierbij is wel interessant dat juist de groep die geen hbo- of wo-opleiding heeft gevold, maar over een vo-diploma beschikt de hoogste scores geeft. Dit zou er mee te maken kunnen hebben dat deze personen vaker werkzaam zijn bij kleinere en/of minder stedelijke gemeenten. Indien daar alsnog datagedreven gewerkt wordt voren in deze gemeenten juist ook vo-opgeleide gemeentemedewerkers datagerelateerde taken uit en bereiken daardoor aanzien.

Figuur 4.5.1: Effect opleiding op 'persoonlijke' vragen (schalen 3, 6, 7 en 8)







## 5 Discussie en aanbevelingen

De uitkomsten die tot zover zijn beschreven geven een redelijk goed beeld van de stand van data maturity van (een deel van) de gemeenten. Van de huidige 388 gemeenten in Nederland heeft ongeveer een derde (127 gemeenten) deelgenomen aan de maturity scan. Desalniettemin kunnen we interessante bevindingen uit deze scan halen en aanbevelingen doen voor gemeenten in alle drie de clusters.

Uit deze maturity scan alsmede uit de diepte-interviews met gemeenten komt naar voren dat de gemeenten in de scan gemiddeld op een gestandaardiseerd niveau scoren. Dat betekent dat het algemene beeld over de toegevoegde waarde van data (science) positief is. De gemeenten beschikken over het algemeen over data van redelijk goede kwaliteit, verzamelen cruciale informatie redelijk geautomatiseerd, en er bestaan deels data warehouses met actuele gegevens. Daarnaast beschikken veel gemeenten over een algemeen en gedragen datamodel en over initiatieven voor organisatiebreed technologie-management. Gemeentemedewerkers hebben over het algemeen toegang tot de data die ze nodig hebben en de gemeente stuurt redelijk feitengebaseerd.

Doordat de vragenlijst eerst naar de gemeentesecretarissen is gegaan en we niet precies weten wie de enquête uiteindelijk heeft ingevuld, beschikken we niet over uitgebreide achtergrondinformatie met betrekking tot de functie van de betreffende persoon. We weten daarom niet of de enquête is ingevuld door een specialist of manager.<sup>11</sup>

Uit ander onderzoek blijkt echter dat er vaak een groot verschil is tussen de inschatting van managers en die van specialisten. Zo blijkt bijvoorbeeld uit de scan die eerder is gedaan dat managers vaker negatief zijn in hun oordeel over de positieve bijdragen van data science. Managers voelen daardoor bijvoorbeeld minder urgentie voor meer capaciteit op dat gebied. Een oorzaak hiervan kan zijn dat managers over het algemeen aanmerkelijk minder werken met data en daarin, naar eigen zeggen, vaak ook minder bekwaam zijn (Bosma et al. 2016).

Als we de uitkomsten uit de maturity scan van gemeenten relateren aan de bevindingen uit de diepte-interviews wordt duidelijk dat er een zekere overlap bestaat in meningen en inschattingen. In de interviews waren de gemeentemedewerkers echter vaak kritischer dan in de scan. De uitkomsten van de scan zouden dus een overschatting van de werkelijke situatie kunnen zijn, wat te maken kan hebben met het feit dat de enquête gebaseerd is op zelfrapportage.<sup>12</sup>

Gesignaleerde problemen in de interviews zijn vooral dat men te vaak niet afdelings-overstijgend werkt, zowel wat betreft de uitwisseling van data als ook het delen van informatie en rapporten. Vaak beschikken gemeentemedewerkers (ofwel de betreffende

---

<sup>11</sup> Daarnaast weten we niet of de gehele vragenlijst door een en dezelfde persoon is ingevuld of door meerdere personen.

<sup>12</sup> We kwamen ook één keer de situatie tegen waarbij iemand in een interview heel positief sprak over de gemeente waarvoor de betreffende persoon werkte. Uit de scan bleek later dat deze gemeente in vergelijking met andere echter niet bovengemiddeld maar gemiddeld scoorde. Daarom behoort deze gemeente niet tot de voorlopers.



specialisten van een afdeling) maar over een beperkte set aan data die nodig is voor de eigen, lokale taken en opdrachten en kunnen uitkomsten van analyses niet breed gedeeld en/of begrepen worden. Dit wordt in de literatuur ook weleens functie- en informatiesilo genoemd en is een van de meest voorkomende beperkingen voor data science binnen een organisatie en voor datagedreven beleid (Strategy& 2014; Booz, Allen, Hamilton 2015, Bosma et al. 2016).

Om een organisatie of gemeente in staat te stellen data science te beoefenen en niet alleen met een traditionele manier van data analyse bezig te zijn is het juist belangrijk om over dergelijke beperkende factoren heen te stappen. Dit kan in eerste instantie met behulp van data warehouses. Op een hoger niveau is aan te bevelen om geavanceerde *cloud*-technologieën in te zetten. Een cloud geeft de mogelijkheid om beheertaken te automatiseren waardoor nieuwe diensten sneller kunnen worden uitgerold. Daarnaast zijn in een cloud juist alle beschikbare data verzameld, zowel van interne als externe bronnen, en kunnen deze bovendien met elkaar gekoppeld en geanalyseerd worden. Daardoor worden de informatiesilo's open gebroken en zijn gemeenten in staat om met grote en diverse hoeveelheden data te werken. De hogere niveaus van data maturity zijn uiteindelijk alleen door de inzet van dit soort geavanceerde databeheer –en analyse technologieën te bereiken.

Verder komt in zowel de scan als de interviews naar voren dat er vrijwel overal te weinig ruimte is voor samenwerking buiten de standaardkaders of –teams, en om te experimenteren. Daardoor worden de voordelen van verschillende expertises en ideeën die ontstaan bij een samenwerking in een nieuw en divers team vaak niet gerealiseerd. Juist meer afdelingsoverstijgende samenwerking om te komen tot nieuwe inzichten en innovatieve manieren om problemen op te lossen is cruciaal voor data science. Door een dergelijke samenwerking zal in de loop van de tijd ook meer wederzijds begrip ontstaan en een nieuwe cultuur van experimenteren en datagedreven beleid.

Om deze processen te bevorderen zijn vooral de managers of bestuurders gevraagd. Hun visie op data science en hun draagvlak voor datagedreven beleid zijn uitermate belangrijk om een gemeente naar een hoger niveau van datavolwassenheid te brengen. Een gebrek aan die visie c.q. draagvlak kwam zowel in de scan als in de interviews vaak als knelpunt naar voren. Onderbouwing en inbedding van data science in zowel bestuurlijke als beleidsmatige processen binnen een gemeente is cruciaal. 'Draagvlak van boven' draagt uiteraard ook bij aan meer capaciteiten, zowel op het gebied van financiën als personeel.

Aan de financiële kant zijn naast investeringen in menselijk kapitaal ook belangrijke investeringen in datainfrastructuur, databeheer en datamanagement te plegen, bijvoorbeeld in boven genoemde cloud technologieën. Allen als gemeenten over voldoende expertise én technologieën beschikken zijn ze in staat om uiteindelijk naar een hoger niveau van maturity te komen. Als een gemeentebestuur hier zorgvuldige afwegingen maakt en toekomstgerichte, breed gedragen beslissingen maakt ondervindt men minder obstakels op de weg naar datavolwassenheid.

Uit de scan blijkt dat de kennis en kunde van gemeentemedewerkers onderwerpen zijn waarop vaak nog onvoldoende gescoord wordt door de gemeenten. Ontevredenheid hierover kwam ook in de interviews naar voren. Juist medewerkers met meer kennis op het gebied van data science zijn belangrijk om van een gestandaardiseerd



naar een hoger maturity niveau te komen. Dit kan door bijscholing van medewerkers, met het voordeel dat zij al beschikken over gemeente-specifieke kennis. Of door het aannemen van data scientists die hun brede technische expertise weliswaar op een gepaste manier moeten afstemmen op gemeentelijke processen en beleidsvormingen.

Naast meer interne samenwerking is het uiteraard ook aan te bevelen om meer externe samenwerking te zoeken. Dit kan bijvoorbeeld in samenwerkingsverbanden met andere gemeenten of overheidsinstanties, maar ook met andere partners.

Vanuit deze gedachte is ook de pilot "Verkenning Data Lab voor gemeenten" ontstaan. Hier werkten VNG, KING en JADS samen om de data science competenties van gemeenten te bevorderen en door de brede samenwerking *fact-based policy making* naar een hoger plan te trekken. Meer over de 'mitsen en maren' van samenwerken in een fase van diepgrijpende veranderingen is te lezen in box 5. Geert Duijsters gaf tijdens de pilot als hoogleraar Entrepreneurship een masterclass aan geïnteresseerde gemeente-medewerkers.

### **Box 5: Geert Duijsters (hoogleraar Entrepreneurship, TiU): Technologische disruptie - kansen voor samenwerking**

Disruptie lijkt een modewoord, een hype, een gril. Niets is echter minder waar. Disruptie raakt in toenemende mate alles en iedereen. Disruptie slaat als een sluipmoordenaar toe: Overall en altijd. Het ontwricht in toenemende mate onze samenleving, onze manier van werken en onze economie. Disruptie is wellicht de meest verwoestende kracht sinds de industriële revolutie.

Maar wat is nu disruptie? De meest letterlijke vertaling van disruptie is ontwrichting. Het is een kracht die het jarenlange evenwicht in een economische sector kan doen omslaan in een grote mate van instabiliteit waarbij een groot gedeelte van de traditioneel heersende organisaties ten onder zal gaan en andere, tot dan toe onbekende, bedrijven de nieuwe leiders zullen zijn. Digitale disruptie zal leiden tot ontwrichting van onze samenleving waarbij grote verliezers te noteren zijn en onverwachte winnaars. De winnaars zijn niet alleen de als disruptors bekend staande start-ups zoals Airbnb, Uber en Netflix, maar ook grote organisaties die in staat zijn geweest om hun structuren om te vormen naar flexibele, lerende en zelf organiserende instellingen.

Omdat disrupties vaak ontstaan in niches en voor gemeenten relatief onbekende sectoren (bijvoorbeeld data science), is het belangrijk om buiten het eigen klassieke domein te manoeuvreren. In toenemende mate is dan ook externe samenwerking nodig om alle vereiste (data-gerelateerde) competenties te verkrijgen, om krachten te bundelen en om synergie te creëren.

Strategische samenwerking is echter niet makkelijk. Ongeveer de helft van alle samenwerkingsverbanden faalt. Het opbouwen van de juiste alliantievaardigheden kan dit percentage echter tot boven 90% omhoog stuwten. Onderzoek laat zien dat het vooral van belang is te investeren in regelmatige evaluatie met de partner, het opstellen van de juiste evaluatiematen, het maken van gezamenlijke afspraken op basis van een Memorandum of Understanding and Principles (MOUP) en het opbouwen van de juiste kennis rondom het management van samenwerkingsprocessen. Tezamen vormen deze elementen de bouwstenen van een succesvol samenwerkingsmodel.



Maar wat zijn nu concrete aanbevelingen voor gemeenten in de drie clusters? Om te beginnen bij de **volgers** die nog de meeste stappen moeten zetten op de weg naar data-volwassenheid: 'begin klein, begin bij jezelf', dus bij de eigen (interne) databronnen. Zorg dat je weet wat je in huis hebt qua datakwaliteit en -format, maar ook qua mensen die met data om kunnen gaan. Als een klein groepje mensen dieper in de data duikt en begint te experimenteren, is dat al een grote stap. Maak vervolgens het succes van experimenten en projecten - en uiteindelijk beleid - steeds minder afhankelijk van de kennis en ervaring van deze kleine groep medewerkers. Zorg dat de eerste kleine experimenten bij succes opgeschaald en gestructureerd herhaald kunnen worden. Spreek algemene regels af voor datakwaliteit en -management, voor privacywaarborging en voor gemeentebrede data-analyses. Ga op zoek naar een trekker, een boegbeeld voor data science binnen je gemeente die zowel mensen als middelen en tijd kan regelen voor steeds meer datagedreven beleid.

De **middenmoters** moeten blijven investeren in kennis en capaciteit voor data science en het verbreden van draagvlak binnen het bestuur. Committeer medewerkers formeel aan de datastrategie en laat data governance processen ontstaan. Maak de beschikbare data steeds meer deelbaar en bruikbaar, bij voorkeur in gecentraliseerde organisatiebrede data warehouses. Maak afdelingsoverstijgende samenwerking een *must* en zorg voor een algemeen gedragen datamodel en technologiemanagement. Betrek zowel gebruikers als specialisten in toenemende mate bij het garanderen van kwaliteit van datagedreven beleid en blijf de toegevoegde waarde voor beleid en burgers toetsen. Laat de aandacht steeds meer verschuiven van probleemoplossing naar preventie.

En wat zijn de aanbevelingen voor de **voorlopers**? Aangezien geen enkele gemeente op het hoogste niveau van data maturity scoort in de scan, zijn er nog zaken die veranderd kunnen worden. Zorg ervoor dat data governance in een framework geïmplementeerd is en op de agenda van de hoogste bestuurlijke laag staat. Nieuwe activiteiten moeten aan dit framework getoetst worden en vastgelegde regels moeten geautomatiseerd gewaarborgd worden. Hiermee creëer je breed vertrouwen in datakwaliteit. Maak data volledig geïntegreerd, bijvoorbeeld door in te zetten op geavanceerde cloud technologieën. Ontplooï een cultuur waar analyse, meetbaarheid en toerekenbaarheid belangrijk zijn en waar informatie wordt gezien als een strategische succesfactor zodat er breed vertrouwen is in het gebruik van data science in de dagelijks bedrijfsvoering en beleidsvorming in alle lagen van de organisatie.

Als afsluiting van deze discussiesectie zijn de gedachten van Johan Wolswinkel te lezen over de afwegingen die gemeenten als overheidsinstellingen moeten maken als het gaat om data science. Als universitair hoofddocent Bestuursrecht gaf ook Johan Wolswinkel een masterclass tijdens de pilot.



## Box 6: Johan Wolswinkel (universitair hoofddocent Bestuursrecht, TiU): Data en Openbaar Bestuur – zijn gemeenten ‘too small for big data’?

De ‘datarevolutie’ heeft inmiddels ook het openbaar bestuur bereikt. Overheden zien nieuwe mogelijkheden in het gebruik van (big) data, maar zijn tegelijkertijd nog zoekende naar manieren waarop zij verantwoord kunnen omgaan met nieuwe datatoepassingen in relatie tot hun burgers.

In die ontmoeting tussen data en overheden nemen gemeenten een bijzondere plek in. Waar de kleine schaal van gemeenten ogenschijnlijk in contrast staat met de grote schaal die big data veronderstellen, blijken er juist vele gemeentelijke beleidsterreinen te zijn waarop datatoepassingen hun waarde kunnen bewijzen en reeds hebben bewezen. Voorbeelden variëren van de openbare orde via de sociale zekerheid tot de fysieke leefomgeving.

Een sterkere oriëntatie op ‘datagedreven bestuur’ stelt echter de kaders waarmee gemeenten vertrouwd zijn op de proef. Zo richt openbaarheid van bestuur zich niet meer primair op (gestructureerde) informatie die aan burgers op verzoek ter beschikking wordt gesteld, maar op (ruwe) data waarmee burgers zelf informatie over de gemeente kunnen genereren. Kunnen alle burgers wel op dezelfde manier worden voorzien van data nu niet elke burger een data scientist is? Ook datatoepassingen door de overheid zelf moeten anders worden benaderd. Wat betekent een behoorlijke motivering van besluitvorming als data-analyses de grondslag vormen voor deze besluitvorming? Welke mate van transparantie kan worden betracht over achterliggende algoritmes die de uitkomst van de besluitvorming hebben bepaald? En hoe blijft de bescherming van persoonsgegevens (zichtbaar) gewaarborgd?

Kortom: datagedreven bestuur vraagt om doordenking van klassieke vragen die gemeenten steeds op het netvlies moeten hebben staan in hun omgang met burgers. Nu de oude antwoorden niet altijd meer volstaan, is de uitdaging om nieuwe antwoorden te formuleren die passen in het datatijdperk, maar tegelijk recht blijven doen aan de bescherming van burgers door en tegenover hun gemeenten.



## 6 Conclusies

Dit white paper is een resultaat van de pilot “Verkenning Data Lab voor gemeenten” die VNG/KING en JADS in oktober 2016 hebben opgestart. Het achterliggende doel van deze samenwerking was de data science competenties van gemeenten te verhogen om gezamenlijk *fact-based policy making* naar een hoger plan te trekken.

Deze pilot leverde ervaringen op over ‘data science voor gemeenten’ en over de meest gepaste inrichting van een langdurige samenwerking waarin JADS en VNG/KING met anderen een platform en ecosysteem kunnen bieden om gemeenten te helpen met deze uitdaging.

Naast twee projecten omtrent de thema’s ‘regionale bedrijvigheid’ en ‘omgevingswet’ en het uitvoeren van hackathons met gemeentemedewerkers onder begeleiding van Syner-scope en CentERdata maakte een ‘leerproject’ onderdeel uit van de pilot. Daaronder valt de uitgevoerde *data science maturity scan* voor gemeenten in de vorm van een enquête met aanvullende diepte-interviews.

De bevindingen van dit onderzoek staan in dit white paper. We kunnen concluderen dat de gemeenten in de scan gemiddeld scoren op een gestandaardiseerd niveau, en dat er op basis van de vragenlijst een indeling in voorlopers, middenmoters en volgers gemaakt kan worden.

Zowel voor datagedreven beleid in het algemeen als voor gemeenten op de verschillende niveaus van data maturity zijn aanbevelingen gedaan hoe zij verdere stappen kunnen zetten die kunnen leiden tot het hoogste niveau van datavolwassenheid.

Uiteraard is er geen algemene unieke regel die voor alle gemeenten van toepassing is. Elke gemeente, elke organisatie moet voor zichzelf keuzes maken baserend op de individuele situatie, achtergrondkenmerken en omstandigheden. Bij de eigen organisatie en klein beginnen, verstandig gebruik maken van wat een gemeente, haar bestuur en haar medewerkers qua data en kennis al in huis hebben en weten is altijd een goede startpunt.

De hoofdprioriteiten voor gemeentebesturen op een rij:

- Ontwikkel een duidelijke (big) data strategie
- Toon de toegevoegde waarde van datagedreven beleid aan in experimenten/pilots
- Leg de verantwoordelijkheid voor datakwaliteit en –beheer in handen van een aangewezen persoon of groep, bijvoorbeeld bij een “Chief Data Scientist”
- Verhoog de kennis van medewerkers door training en/of werving zodat men niet alleen over (meer) technische kennis beschikt, maar ook de juiste vragen kan stellen; zorg dat deze vragen door gepaste tools beantwoord kunnen worden
- Maak data science een integraal element van je bestuurs- en beleidsprocessen zodat er een cultuur van datagedreven en feitengebaserd beleid ontstaat waarin de toegevoegde waarde voor de burgers en maatschappelijke acceptatie centraal staan.



## 7 Literatuur

Becker, J., Knackstedt, R., & Pöppelbuss, J. (2009). Exploring the SME Quandary: data governance in practise in the small to medium-sized enterprise sector. *Electronic Journal of Information Systems Evaluation*, 15, 2-13.

Booz, Allen, Hamilton (2015). The field guide to data science.

Bosma, B., Kolkman, D. & Van de Born, A. (2016). Data science maturity. Projectteam Data Science, Tilburg University.

De Bruin, T., & Freeze, R. (2005). Understanding the main phases of developing a maturity assessment model. 16th Australian Conference on Information Systems.

Dataflux. The data governance maturity model. White paper.

Huner, K., Ofner, K., & Otto, B. (2009). Towards a maturity model for corporate data quality management. Proceeding of the 2009 ACM symposium on Applied Computing, 231-238.

Klimko, G. (2001). Knowledge management and maturity models: Building common understanding. Second European Conference on Knowledge Management, 279-288.

Loshin, D. (2011). *The practitioner's guide to data quality improvement*. Chapter 3, 35-51.

Mills, S. (2014). Data science maturity. Booz Allen Hamilton, 1-3.

Pöppelbuss, J., & Röglinger, M. (2011). What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. *Proceedings of ECIS 2011*, paper 28.

Strategy& (2014). Big data maturity – An action plan for policy makers and executives, eerder gepubliceerd in *The Global Information Technology Report 2014: Rewards and Risks of Big Data*, World Economic Forum.



## A Scoringsniveaus maturity scan<sup>13</sup>

Scoringsniveau	Beschrijving
<b>Basaal</b>	Oplossingen en initiatieven op het thema 'data' zijn lokaal en sporadisch; er is weerstand tegen een data-gedreven manier van werken binnen de organisatie; er zijn geen algemeen geldende standaarden en geen organisatiebrede gedeelde resources of management. Er is geen overkoepelende architectuur; veel ad hoc 'oplossingen'; primitieve technologie; rapportages komen direct vanuit lokale en operationele systemen en applicaties (denk aan: Microsoft Excel); IT wordt gezien als cost center. De directie van de organisatie heeft geen transparant en consistent overkoepelend beeld en kan niet feitengebaseerd sturen.
<b>Gecontroleerd</b>	Organisatie is zich bewust van de waarde van data, maar er is weinig vertrouwen in datakwaliteit en -consistentie. Er is veel redundantie en inconsistentie in data en beperkte steun voor het thema 'data' bij de directie. Qua tooling wordt er gebruik gemaakt van 'off the shelf' softwaresuites; geen gedegen informatiesysteem waarin data wordt verzameld.
<b>Gestandaardiseerd</b>	Verbeterde datakwaliteit; verzamelen van cruciale informatie in hoge mate geautomatiseerd; data warehouses met actuele gegevens; algemeen en gedragen datamodel; initiatieven voor een organisatiebreed technologiemanagement; mensen hebben over het algemeen toegang tot de data die ze nodig hebben. De organisatie stuurt meer feitengebaseerd.
<b>Geoptimaliseerd</b>	Hoge datakwaliteit; organisatie benoemt data als kritische succesfactor en heeft een datastrategie; alle relevante data is beschikbaar in gecentraliseerde organisatiebrede data warehouses; constante consolidatie van analytische structuren; flexibele en gelaagde architectuur; oplossingen drijven functionaliteit binnen de organisatie; integratie tussen analytische applicaties. Mensen zijn formeel gecommitteerd aan de datastrategie en worden getraind om data te analyseren; presteren en (bij)sturing vinden plaats op basis van data. De datastrategie is sterk verweven met de algehele strategie.
<b>Innovatief</b>	Breed vertrouwen in datakwaliteit; data is volledig geïntegreerd, informatie wordt gezien als een strategische succesfactor; er heerst een cultuur waar analyse, meetbaarheid en toerekenbaarheid hoog in het vaandel staan en data drijft de innovatie.

<sup>13</sup> Deze scoringsniveaus en de omschrijvingen zijn overgenomen uit de eerste versie van de maturity scan die door Bosma et al. (2016) is afgenomen.





## B Diepte-interviews

De 9 interviews zijn deels in persoon en deels telefonisch afgenomen, en duurden ongeveer 45 minuten.

Vragen:<sup>14</sup>

### **Wat is jouw rol en functie binnen de gemeente?**

- Hoe ontwikkelt de thematiek rond data zich gezien vanuit jouw rol en functie?
- Hoe heeft dit thema zich de afgelopen jaren ontwikkeld?
- Wat heb je hier in jouw werk concreet van gemerkt?

### **Welke kansen en/of bedreigingen zie je op het thema 'data' voor de gemeente?**

- Welke projecten worden er binnen op het thema 'data' uitgevoerd binnen jouw afdeling of werkgebied?
- Kan je een voorbeeld geven van een succesvol project binnen dit thema?

### **Welke tools zijn er binnen de gemeente beschikbaar om te kunnen werken met data?**

- Is er sprake van een algemene manier van werken?
- Kunnen data-analyses tussen afdelingen worden gedeeld? Hoe?

### **Wat zie je als de belangrijkste uitdagingen om het thema 'data' in te bedden in de beleidsprocessen van de gemeente?**

- Waarom?
- Welke aspecten van de beleidsvoering zijn veranderd onder invloed van dit thema? Op welke wijze?
- Heeft u voorbeelden van beleidsprocessen die heel anders verlopen dan enkele jaren geleden?

### **Welke ontwikkelingen op het thema 'data' zou je het liefst versnellen? Waarom?**

- Wat is de (financiële/sociale) impact als de gemeente niet kan versnellen op deze thema's?

### **Welke kennis en kunde zie je als essentieel in de verdere ontwikkelingen van de gemeente op dit thema?**

- Welke kennis en kunde is al aanwezig?
- Hoe breed is deze kennis en kunde beschikbaar?
- Op welke wijze ontwikkelt de kennis en kunde zich binnen de gemeente?

### **Als we 5 – 10 jaar vooruit zouden kijken, op welke aspecten van (de beleidsvoering van) de gemeente heeft dit thema dan de grootste invloed gehad?**

- Waarom deze aspecten?
- Op welke aspecten heeft dit thema weinig of geen invloed gehad?
- Waarom denk je dit?

### **Welke adviezen zou je gemeenten geven die dit thema aan het verkennen zijn?**

- Welke lessen kunnen andere gemeenten trekken uit de ervaringen van uw gemeente?
- Waar zie je de ontwikkelingen toe leiden op de langere termijn?

---

<sup>14</sup> Deze vragen zijn overgenomen uit de eerste versie van de maturity scan die door Bosma et al. (2016) is afgenomen.



## C Clusters deelnemende gemeenten

Cluster 1: 42 gemeenten "Volgers"	Cluster 2: 36 gemeenten "Middenmoters"	Cluster 3: 37 gemeenten "Voorlopers"
Oirschot	Lingewaal	Bunnik
Mook en Middelaar	Noord-Beveland	Drechterland
Beverwijk	Gorinchem (Gorcum of Gorkum)	Oost Gelre
Haaren	Houten	Doetinchem
Schinnen	Hoogeveen	Súdwest-Fryslân
Schagen	Etten-Leur	Meppel
Winsum	Apeldoorn	Maastricht
Zoeterwoude	Rhenen	Hellevoetsluis
Rheden	Zutphen	Valkenburg aan de Geul
Borsele	Hoogezand-Sappemeer	Stichtse Vecht
Emmen	Bloemendaal	Rijssen-Holtén
Werkendam	Venlo	Overbetuwe
Putten	Bergen op Zoom	Woerden
Urk	Uithoorn	Enschede
Harlingen	Ouder-Amstel	Haarlem
Steenwijkerland	Dongen	Hulst
Wijk bij Duurstede	Boxtel	Zaanstad
Bergeijk	Roosendaal	Stein
Reimerswaal	Veenendaal	Hoorn
Stadskanaal	Horst aan de Maas	Helmond
Someren	Beesel	Hollands Kroon
Dronten	Oisterwijk	Eindhoven
Huizen	Elburg	Groningen
Uitgeest	Landgraaf	Bunschoten
Leiden	Rotterdam	Blaricum
Steenbergen	Lingewaard	Franekeradeel
Halderberge	Assen	Het Bildt
Noordoostpolder	Zederik	Menaldumadeel
Meerssen	Nieuwegein	Eemnes
Wijdmeren	Weert	Laren
Oldambt	Berg en Dal	Stede Broec
Hilvarenbeek	Heerenveen	Enkhuizen
Beuningen	Eersel	Breda
Loppersum	Leidschendam-Voorburg	Cuijk
Oostzaan	Leudal	Grave
Ameland	Zeist	's-Hertogenbosch (Den Bosch)
Schiermonnikoog		Mill en Sint Hubert
Terschelling		
Texel		
Vlieland		
Rozendaal		
Woensdrecht		