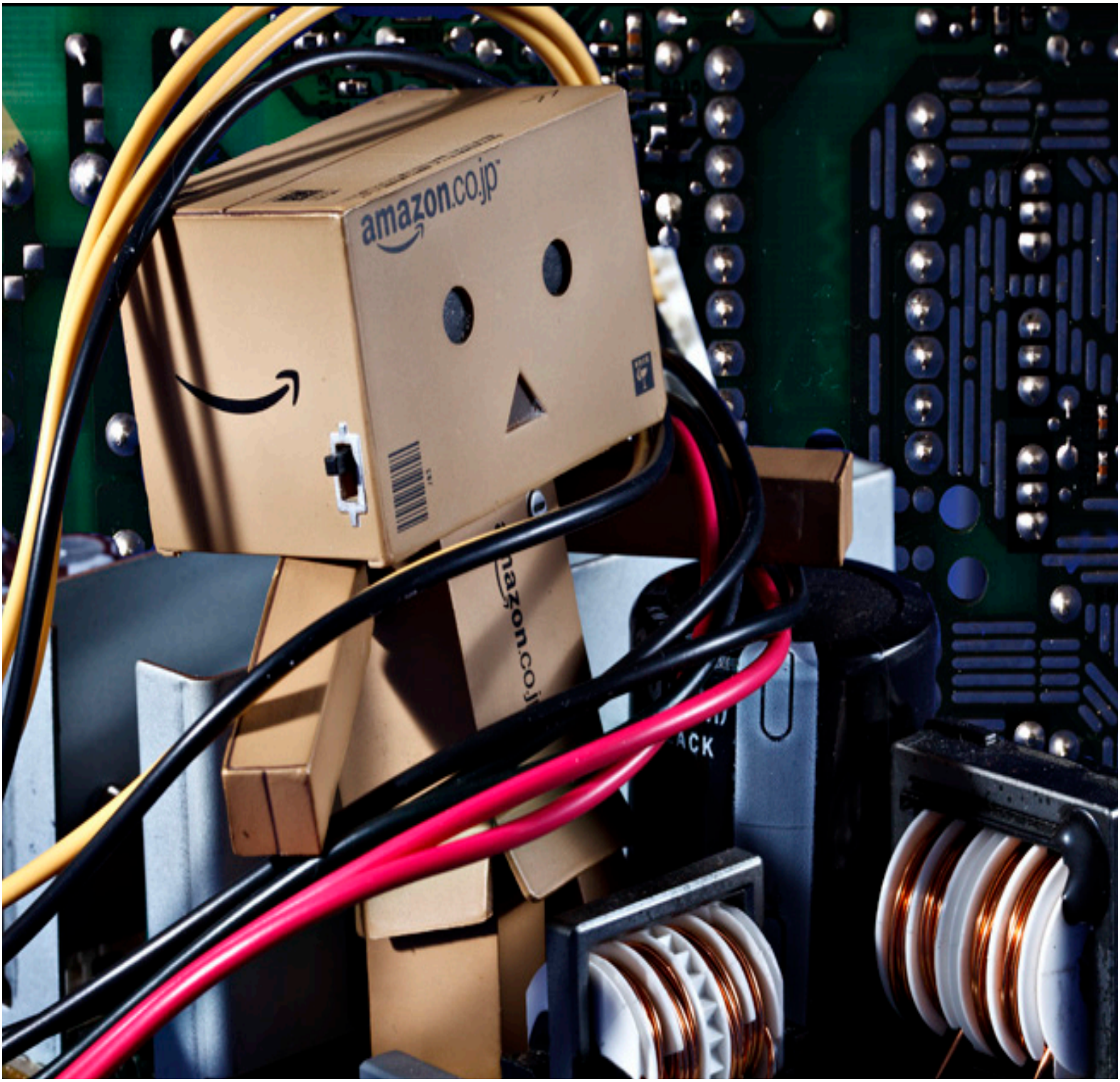
Small Data: The Next Big Thing?

Een discoursanalyse naar de belofte van ‘small data’



Anne Last

Masterscriptie: Small Data: The Next Big Thing?

Begeleid door Marianne van den Boomen
Student Anne Last (3727823)

Master Nieuwe media en digitale cultuur
Faculteit Geesteswetenschappen
Universiteit Utrecht, Augustus, 2013
Coverfoto: Christine Rademaker, 2012

Inhoud

Introductie - De 'nieuwe' claim **5**

Hoofdstuk 1 - Voorbij de hype: van big naar small data **9**

1.1 Big data *10*

1.2 Big data: too big to control *12*

1.3 Big data: too big to handle *14*

1.4 Small data als alternatief discours *17*

Hoofdstuk 2 - Methodologie Foucauldiaanse discoursanalyse **21**

2.1 Michel Foucault *21*

2.2 Het discours *23*

Hoofdstuk 3 - Discoursanalyse small data **25**

3.1 Externe uitsluitingsprocedures *26*

3.2 Interne uitsluitingsprocedures *31*

3.3 Regels voor deelnemers aan het discours *35*

Conclusie **43**

Literatuur **47**

Introductie

De ‘nieuwe’ claim

Sinds het internet bijna twee decennia geleden algemeen beschikbaar werd voor het grote publiek, hebben journalisten en wetenschappers zich met name gericht op het beschrijven en duiden van specifieke nieuwe toepassingen en de maatschappelijke veranderingen die deze toepassingen teweeg hebben gebracht. Zo konden we dankzij het internet opeens informatie delen met mensen aan de andere kant van de wereld, zelf ons eigen medium creëren en onze gedachten en meningen publiceren, ons verenigen in sociale netwerken en muziek, films en foto’s – al dan niet illegaal – verspreiden. Al deze toepassingen, en meer, hebben verstrekkende gevolgen gehad voor de samenleving waarin we leven.

Sinds 2011 begint de discussie zich echter te verschuiven van de toepassingen van het internet naar de elementaire bouwstenen ervan: data. We worden ons steeds meer bewust van het feit dat het internet in zijn totaliteit inmiddels een haast oneindige hoeveelheid gegevens bevat die, wanneer zij op de juiste manier gebruikt en gecombineerd worden, een schat aan waardevolle informatie op kunnen leveren (Manovich, 2011). Deze data worden door alle gebruikers van het internet al dan niet bewust gecreëerd bij iedere handeling die zij online doen, zelfs wanneer zij zich niet letterlijk achter hun computer bevinden, bijvoorbeeld wanneer zij betalen met hun credit card, boodschappen doen met hun bonuskaart of reizen met hun OV chipcard. En dus gaat de discussie niet langer alleen over huidige toepassingen, maar ook over de potentie van alle onontdekte verbanden die deze tsunami aan data met zich meebrengen.

Wanneer de discussie over data gaat is ‘big data’ op dit moment een van de dominante concepten. Big data draaien om de verzameling, analyse en consumptie van, zoals de naam impliceert, enorme hoeveelheden data (Bloem *et al.,* 2012). De term ‘big data’ is zonder twijfel één van dé technologie modewoorden van 2012. Sommige technische en business weblogs schrijven artikelen over het onderwerp met titels als: “Het nieuwe goud” of “Unlocking the Big Data goldmine” (Beemster, 2013). Anderen schrijven bijvoorbeeld over slimme data-toepassingen in de medische wereld, onder de titel “How Big Data will save your life” (Wired, april 2013). Stuk voor stuk grote beloftes, gerelateerd aan bedrijfseconomische of maatschappelijke vooruitgang.

Wanneer we deze ‘veelbelovende goudmijn’ aan data van dichtbij bekijken, blijkt al snel dat er ook een keerzijde is. Een groeiende groep critici wijst erop dat big data enkel in het voordeel werken van grote bedrijven en overheden, en niet in het voordeel van de mensen die de data produceren (Morozov, 2013). Sommigen menen zelfs dat consumenten en burgers alleen gebruikt

worden door bedrijven als bron voor het genereren van data, vaak zonder medeweten van de eigenaar hiervan (Bits of Freedom, 2013). Een geheel andere groep critici wijst erop dat in de praktijk veel bedrijven of individuele consumenten niet over de expertise of technologie beschikken om grote hoeveelheden data te verzamelen of te verwerken tot bruikbare informatie. Hoe veelbelovend de enorme berg data die 'big data' belooft ook lijkt, het daadwerkelijk goed inrichten van de enorme databases die nodig zijn voor het verwerken ervan is voor velen simpelweg te complex (Stillman, 2013; Ho, 2013).

Als antwoord op de kritiek op het gebruik van big data lijkt echter ook alweer een alternatief op te komen: small data. Small data staan, volgens de voorvechters van het begrip, voor kleine, gedetailleerde hoeveelheden data die gemakkelijk opgeslagen en verwerkt kunnen worden op een 'normale' computer, in tegenstelling tot big data waar de rekenkracht van gespecialiseerde systemen of meerdere machines voor nodig is. Een focus op small data in plaats van op big data, zou ertoe leiden dat consumenten en burgers beter in staat zouden zijn om hun eigen data te beheren en er de voordelen van te plukken (Pollock, 2013b; Searls, 2012a).

Het probleem met de term small data is echter dat er nog geen consensus lijkt te bestaan over wat het begrip precies inhoudt, anders dan dat het over kleinere hoeveelheden data gaat dan dat bij de term 'big data' het geval is. Sinds 2012 wordt er op verschillende manieren over small data geschreven en gesproken en er zijn inmiddels meerdere interpretaties ontstaan over de beoogde voordelen van kleine specifieke datasets ten opzichte van grote datasets. Het is hierbij opvallend dat small data bijna nooit op zichzelf worden gepropageerd, maar altijd in samenhang met kritiek op big data. Het lijkt wel alsof het begrip alleen maar vorm krijgt als tegenmacht tegenover big data. Een groot deel van de kritiek op big data is ook juist afkomstig van juist diegenen die er een alternatief voor willen uitdragen.

Zo zijn het juist die mensen die bezwaar maken tegen het feit dat burgers de controle over hun data verliezen bij big data-analyse, die small data positioneren als middel voor sociale innovatie in het publieke domein. Dit standpunt komt met name terug in teksten van de Open Knowledge Foundation. Adviseurs die zich richten op ondernemers, start-ups en lokale bedrijven wijzen op het probleem dat big data alleen door grote bedrijven opgeslagen en verwerkt kunnen worden en bieden small data als handzamer alternatief aan, zoals aan de volgers van de Small Data Group. Ook zijn er mensen, zoals de bekende auteur Doc Searls, schrijver van onder andere *The Cluetrain Manifesto* (2001), die menen dat big data de verhouding tussen consument en aanbieder verstoren. Hij omschrijft vervolgens small data-analyse als manier om de consument weer de macht in het aankoopproces terug te geven. Op het eerste gezicht lijkt het alsof er simpelweg onduidelijkheid is over de term en het een kwestie van tijd is voordat er consensus bereikt is. Maar wanneer we beter naar de standpunten van de diverse voorstanders van small data kijken, zien we dat zij allemaal

vanuit hun eigen perspectief small data als oplossing voor een specifieke beperking van big data neerzetten.

In deze scriptie probeer ik te begrijpen waarom de diverse voorstanders van small data het begrip op zulke verschillende manieren duiden. Het concept small data is geen autonoom begrip, maar wordt neergezet als tegenmacht tegenover big data. Om die redenen is het niet per se interessant om te analyseren wat het concept precies betekent, maar om te begrijpen wat het belang is dat de voorstanders van small data te verdedigen hebben.

Daarom is mijn hoofdvraag: welke belangen spelen er bij het duiden van de term small data?". Ondersteunende deelvragen hierbij zijn ten eerste: "wat behelst het begrip big data precies, wat zijn de problemen die volgens sommigen rondom dit begrip spelen en hoe kunnen small data daar volgens hen een alternatief hiervoor vormen?, en ten tweede "welke verschillende posities zijn er te onderscheiden in de betekenisvorming rondom small data en hoe krijgen small data vorm als tegenmacht tegenover big data?" De hypothese is dat de belangrijkste spelers in het debat over small data allen een ander nadeel van big data aangrijpen om small data vanuit hun eigen perspectief als alternatief te presenteren. En dat er om die reden geen gedeeld begrippenkader ontstaat waarmee het begrip 'small data' volwaardig en volledig tegenover het begrip 'big data' kan worden geplaatst.

Om deze vragen te beantwoorden gebruik ik de filosofie van de Franse denker Michel Foucault als uitgangspunt. Foucault beargumenteert dat kennis en macht aan elkaar gerelateerde begrippen zijn en dat via analyse van de taal – het domein waar kennis gedeeld wordt – onderliggende machtsstructuren zichtbaar gemaakt kunnen worden (Foucault, 1975). Een kritische analyse van het discours (of: vertoog), richt zich op het blootleggen van vele, soms tegenstrijdige betekenissen in de teksten door de juiste vragen te stellen. Zo kunnen geïmpliceerde verbanden zichtbaar gemaakt en opnieuw geïnterpreteerd worden (White, 1995: 173). Om de machtsstructuren en belangen achter de betekenisvorming van small data bloot te leggen, analyseer ik de manier waarop het begrip 'small data' in het discours betekenis krijgt en de manier waarop dit discours als contradiscours ten opzichte van big data functioneert.

In het eerste hoofdstuk probeer ik de eerste deelvragen te beantwoorden, door een zo feitelijk mogelijke beschrijving te geven van het ecosysteem rondom big data, waarna ik vervolgens de mogelijke problemen en beperkingen van het verzamelen, verwerken en analyseren van big data bespreek. Daarna licht ik in ditzelfde hoofdstuk toe, hoe volgens sommigen small data een oplossing zouden kunnen bieden. Hier zullen zich de eerste contouren van de belangen die small data met zich meebrengen zichtbaar worden. In het daaropvolgende hoofdstuk bespreek ik de

methodologische verantwoording door het denken van Michel Foucault over het discours uiteen te zetten en beargumenteer ik waarom dit een geschikte analysemethode is om de hoofdvraag te beantwoorden. In het derde en laatste hoofdstuk voer ik vervolgens de discoursanalyse uit. In deze discoursanalyse analyseer ik op basis van de ideeën uit *De orde van het vertoog* (Foucault, 1982) diverse recente (populair-)wetenschappelijke teksten en opiniestukken, zoals boeken, artikelen en blogposts van auteurs over small data. Ik baseer mij hierbij op de verschillende stromingen van kritiek op big data zoals deze in het eerste hoofdstuk naar voren komen. Om antwoord te kunnen geven op de tweede deelvraag, richt ik mij in deze analyse op de claims en beloftes die de verschillende deelnemers in het discours naar voren brengen. Ik beschrijf op welke manier de betekenis van het begrip small data ontstaat als beloofd alternatief voor de problemen rondom het verzamelen, verweken en beheren van big data. Ten slotte geef ik in de conclusie antwoord op de hoofdvraag, door de belangen die spelen bij de betekenisvorming van het begrip in kaart te brengen.

Rondom het begrip small data spelen verschillende claims en beloftes een rol. Dit is echter niet vreemd, de (wetenschappelijke) geschiedenis leert ons dat innovatie in technologie altijd diverse beloftes en verwachtingen met zich meebrengt. Telkens wanneer er 'nieuwe' technologie geïntroduceerd wordt ontstaat er een diversiteit aan verhalen, door de manier waarop mensen denken en spreken over deze ontwikkelingen (De Vries, 2012). En hoewel we niet in de toekomst kunnen kijken voorbij de reeds bestaande structuren, kunnen we wel van de geschiedenis leren in soortgelijke gebeurtenissen. Dit motiveert mij om bij te dragen aan het onderzoeken van de manier waarop er gesproken en geschreven wordt over small data en de manier waarop het discours vorm krijgt als contradiscours op big data. Op deze manier probeer ik via deze scriptie een bijdrage te leveren aan theorieën die zich toeleggen op de introductie van nieuwe technologieën en op de mogelijke reacties die hierop ontstaan. Zo helpt dit onderzoek dergelijke reacties te nuanceren.

Tevens probeer ik via dit onderzoek een bijdrage te leveren aan de wetenschap van nieuwe media binnen het vakgebied van ‘digital humanities’. Er zijn een aantal dominante debatten gaande binnen dit vakgebied als het om (big) data gaat en deze scriptie raakt aan drie daarvan. Ten eerste het debat rondom persoonlijke data en privacy van internetgebruikers, de vraag in hoeverre de mensen van wie data verzameld worden nog wel controle hebben over hun data en de vraag in hoeverre dit zo zou moeten zijn. Ten tweede spelen er kwesties rondom de commodificatie van data. Wat gebeurt er als data handelswaar worden en het niet duidelijk of het economisch voordeel dat hiermee behaald kan worden terug vloeit naar de originele producten van de data? En ten slotte is er de discussie of data ‘open’ toegankelijk zouden moeten zijn en om welke data dit dan zou gaan. De discussie over small data en de manier waarop het begrip betekenis krijgt als reactie op big data raakt aan alle drie deze debatten.

Hoofdstuk 1

Voorbij de hype: van big naar small data

Als gevolg van de interactie die mensen, machines en applicaties dagelijks onderling of met elkaar hebben, worden vandaag de dag enorme hoeveelheden data gegenereerd. De bronnen van deze data zijn divers. Zij verschillen van clicks op het world wide web (WWW) en activiteit op sociale media tot het gebruik van mobiele apparaten en sensoren.

Er is tot nu toe veel geschreven over hoe het internet de infrastructuur vormt voor onze moderne netwerksamenleving (Castells, 2001; Van Dijk, 2006; Barabási, 2003; Benkler, 2006). In het big data discours lijkt de discussie echter meer te gaan over de data die over het netwerk verzonden worden, dan over de netwerkstructuur van het internet en de sociale organisatievorm(en) die een netwerksamenwerking mogelijk maakt. In dit discours beschrijft men de manier waarop deze data verzameld en gebruikt worden, de maatschappelijke implicaties hiervan en de voordelen of problemen die dit met zich meebrengt. Anders dan sociologen als Manuel Castells en Jan van Dijk, die de ontwikkeling van de digitale samenleving beschrijven vanuit het begrip van 'het netwerk', focus ik in deze scriptie dan ook op de ontwikkeling van de data die door het netwerk getransporteerd worden en de toepassingen die hiermee gepaard gaan.

Het doel van dit hoofdstuk is een beschrijving te geven van het ontstaan van het concept 'small data' als alternatief voor het begrip 'big data'. Dit doe ik door allereerst dieper in te gaan op wat de term 'big data' precies behelst en welke kansen en mogelijkheden big data-analyse volgens sommigen in het discours biedt. Daarna ga ik in op een gedeelte van de kritische kant van het discours, waarin gewezen wordt op een aantal fundamentele problemen in de verhouding tussen overheden en grote bedrijven enerzijds en burgers, consumenten en kleine bedrijven anderzijds. De kritiek op big data in dit hoofdstuk beperkt zich tot de belangrijkste argumenten die door de voorstanders van het small data discours gebruikt wordt. De kritiek richt zich op een bepaald deel van big data discours en richt zich bijvoorbeeld niet op de methodologische vraag of big data-analyse tot betere inzichten leiden dan andere methoden. Ten slotte zal ik laten zien dat er de afgelopen jaren rondom het concept 'small data' een contra-discours is ontstaan op de voedingsbodem van deze kritiek op big data.

1.1 Big data

Data is een veelgehoorde – en in deze scriptie veelgebruikte – term, maar wat behelst het precies? De definitie van data volgens het woordenboek luidt: "[f]acts and statistics collected together for reference or analysis" (Oxford Dictionaries, 2013). Er wordt hier over gegevens of representaties van feiten gesproken die samen als referentie kunnen dienen of geanalyseerd kunnen worden. Data kunnen van alles zijn: waarnemingen, kleuren, getallen, woorden, etc. Digitale data zijn opgebouwd uit bits. Dit zijn de kleinste eenheden van binaire informatie; zij zijn opgebouwd uit 1-en en 0-en. Deze data kunnen worden vastgelegd op tastbare voorwerpen (digitale datadragers) zoals een DVD, USB-stick of op de harde schijf van een computer. In tegenstelling tot analoge data zoals (geprinte) boeken, kunnen digitale data oneindig gekopieerd en gedistribueerd worden zonder merkbaar verlies van kwaliteit.

Het begrip 'big data' betekent feitelijk niks meer dan een grote hoeveelheid ongestructureerde, geaccumuleerde data. Lev Manovich, professor in digitale Geesteswetenschappen geeft een definitie waarin het begrip meer lading krijgt: "Big Data is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single data set" (Manovich, 2011: 1). Manovich benadrukt dat big data niet op een traditionele computer verzameld, beheerd en verwerkt kunnen worden. Het verzamelen, beheren en verwerken van zulke grote databases vraagt meer computerkracht dan bijvoorbeeld een 'normale' laptop bieden kan.

In plaats van te spreken over 'verzamelen', 'beheren' en 'verwerken', onderscheiden Bloem *et al*. op een andere manier drie fases in het gebruik van big data, namelijk big dataproductie, big datamanagement en big dataconsumptie (Bloem *et al.,* 2012: 14). Zij noemen deze driedeling het ecosysteem van big data, vanwege de onderlinge wisselwerking tussen de fases. Het is opvallend dat zij spreken over 'productie', de fase die logischerwijs aan het verzamelen van data vooraf gaat. Zij laten hierbij in het midden of het verzamelen van data bij de fase van productie of van management hoort. Ook spreken zij over consumptie in plaats van over verwerking, waarbij eenzelfde soort vraag gesteld kan worden: moeten data niet eerst verwerkt worden, voordat zij geconsumeerd kunnen worden?

Een precies afgebakende definitie van het begrip 'big data' is er niet volgens Viktor Mayer-Schönberger en Kenneth Cukier. In hun boek *De big-data revolutie* (2013) stellen zij: "[e]r bestaat geen exacte definitie van big data". Wel gaan zij dieper in op het veronderstelde nut van het gebruik van big data: "de term 'big data' verwijst naar dingen die je op een grote schaal kunt doen en die op een kleinere schaal niet mogelijk zijn, waarmee je nieuwe inzichten verkrijgt of nieuwe vormen van

economische waarde creëert op een manier die invloed heeft op onder andere markten, organisaties en de relatie tussen burgers en overheden" (Mayer-Schönberger & Cukier, 2013: 15).

Mogelijke toepassingen van big data zijn bijvoorbeeld het combineren van diverse informatiestromen over het (online) winkelgedrag van mensen om tot een specifieker productaanbod te kunnen komen, zoals Albert Heijn doet op basis van gegevens van de bonuskaart en webwinkel aankopen. Een ander voorbeeld is het gebruik van digitale accessoires die mensen helpen bij een gezondere levensstijl, zoals stappentellers, specifieke mobiele apps en bijvoorbeeld de Fuelband technologie van Nike. Hierbij worden de data van vele duizenden gebruikers gecombineerd om patronen te herkennen en op basis hiervan adviezen aan gebruikers te kunnen geven (Bonde, 2013). Een derde voorbeeld is de manier waarop bedrijven zoals Google en Twitter in staat menen te zijn om griepepidemieën, beurskoersen of verkiezingsuitslagen te kunnen voorspellen, door het analyseren van het volume, de frequentie of de onderlinge samenhang van bepaalde zoekopdrachten en terugkerende termen op sociale media (Google Grieptrends, 2011; Silver, 2012; Leinweber, 2013). Buiten de commerciële sector maken ten slotte ook overheden gebruik van big datatoepassingen, bijvoorbeeld wanneer zij politie en justitie toegang geven tot meerdere gecombineerde databases om te helpen bij opsporingsactiviteiten.

Zoals uit de voorbeelden hierboven blijkt, hebben big data inderdaad nieuwe mogelijkheden gecreëerd voor de relatie tussen burgers en overheden, en tussen consumenten en bedrijven. Een breed scala aan recente innovaties, zowel in de publieke als de commerciële sector, vindt zijn oorsprong in de analyse van big data. Het is dan ook niet verwonderlijk dat mensen die zich bezig houden met innovatie de afgelopen jaren een toenemende interesse in big data laten zien. Wat wel opvalt, is de optimistische toon van het merendeel van de teksten uit dit discours. Het idee is dat meer kennis (over consumenten en/of burgers) leidt tot betere besluitvorming: "'[j]e kunt niet managen wat je niet meet'. Dankzij big data kunnen managers veel meer meten, en dus weten […] en die informatie kunnen ze vertalen naar betere besluitvorming en performance" (McAfee & Brynjolfsson, 2012: 19). Het centrale idee hierbij is dat ieder onontdekt verband in de data het begin kan zijn van een nieuwe schat aan informatie. Oftewel, iedere ongestructureerde databron is een mogelijke nieuwe goudmijn. Het World Economic Forum, omschrijft big data als "the new oil of the Internet and the new currency of the digital world" (Kuneva, 2011: 5).

Vooral op marketing en business gerelateerde weblogs wordt er veel over big data geschreven. De marketingblog Inc.com vermeldde eerder dit jaar dat uit een enquête van het Amerikaanse onderzoeksbureau Harris Interactive bleek dat 76% van de ondervraagde bedrijven erin geloofde dat het gebruik van big data kansen biedt voor hun organisatie (Stillman, 2013). Dit hoge percentage is terug te zien in een marktprognose die sterk toenemende cijfers laat zien: "[l]arge organizations are in love with big data and big analytics. In fact, back in March, IDC forecast that the

big data technology and services market is expected to grow from $3.2 billion in 2010 to $16.9 billion in 2015" (SDG, 2012). Volgens deze voorspelling lijkt de trend van big data en data-analyse zich voorlopig tot 2015 voort te zetten.

De vraag is echter wat de eenvoudige burger en de gewone consument van deze groeiende waarde van big data meekrijgen. Zij zijn het die iedere dag een groot deel van de bergen data produceren die door grote bedrijven, instellingen en overheden verzameld en geanalyseerd (kunnen) worden. Wat krijgen zij hiervoor terug? Natuurlijk zijn er diverse voorbeelden te noemen van innovatieve, op big data gebaseerde toepassingen, waarvan de burger of de consument direct voordeel heeft. Maar er zijn ook diverse gevallen denkbaar, waarbij het voordeel van het gebruik van big data met name bij overheden of grote bedrijven terecht komt. Denk hierbij bijvoorbeeld aan advertentie-inkomsten van Facebook, waarvan de waarde uiteindelijk gebaseerd is op de persoonlijke data van zijn gebruikers (Tate, 2013). En er zijn zelfs gevallen, waarbij het gebruik van big data nadelig is voor de data producerende internetgebruiker. Bijvoorbeeld wanneer zij zonder het te weten en zonder aanleiding een stuk privacy inleveren, zoals bij het recente schandaal rond Prism het geval is (Clark, 2013).

Het is dan ook niet verwonderlijk dat parallel aan de groeiende interesse in big data, er ook steeds meer kritische geluiden komen. Het zou volgens sommigen immers niet de bedoeling van het internet kunnen zijn dat uiteindelijk met name overheden, grote bedrijven en instellingen voordeel zouden kunnen hebben van de data die dagelijks door miljarden internetgebruikers wereldwijd wordt geproduceerd. Zeker omdat het internet van oorsprong is opgezet als democratiserend medium, waarin gelijke toegang tot informatie centraal staan (Pierce, 1997: xvii).

De kritiek op big data is samen te vatten in grofweg twee argumenten. Enerzijds zijn big data 'too big to control': de schaalgrootte van big data leidt ertoe dat burgers en consumenten de controle kwijtraken over welke data er over hen verzameld worden, hoe deze worden opgeslagen, en wat er vervolgens mee gebeurt. Anderzijds zijn big data 'too big to handle': de infrastructuur die benodigd is om inzichten uit big data te halen is dusdanig complex, dat deze minder goed te gebruiken is voor individuen en kleine bedrijven. Dit zou ertoe kunnen leiden dat in een wereld waarin data steeds belangrijker worden, individuen en kleine bedrijven afhankelijk worden van grote instanties als overheden en multinationals.

1.2 Big data: too big to control

In 2012 creëerden mensen en machines gezamenlijk gemiddeld 2,5 exabytes aan data per dag, deze hoeveelheid verdubbelt naar verwachting iedere 40 maanden (McAfee & Brynjolfsson, 2012: 19). Deze data worden via diverse manieren geproduceerd en verzameld. Een belangrijke bron van

dataproductie op het internet zijn cookies: kleine databestandjes die websites vanaf hun server via de browser op de computer van de sitebezoeker plaatst. Persoonlijke gegevens, voorkeuren en de geschiedenis van het browsegedrag van bezoekers worden op deze manier bijgehouden. De drie meest voorkomende cookies zijn: functionele cookies, cookies ten behoeve van de statistieken en cookies ten behoeve van advertenties. Het zijn met name de data die door het laatste type cookies gegenereerd worden, waar het meeste economische waarde uit te halen is door bedrijven en waar dan ook het meeste kritiek op is (Telecommunicatiewet, 2013).

Om data te produceren hoeven we niet altijd achter onze computers te zitten: we bevinden ons tegenwoordig in een staat van continue connectiviteit. Met een smartphone op zak staan we bijna altijd in verbinding met andere genetwerkte apparaten, online diensten en met elkaar. Dit zal de komende jaren alleen maar toenemen door het gebruik van bijvoorbeeld sensoren en RFID-chips, waarmee apparaten die voorheen op zichzelf stonden ook onderdeel van het netwerk kunnen worden (Bloem *et al*., 2012: 14). Ook wanneer we pinnen, een klantenkaarten gebruiken of met het OV reizen laten we een spoor van digitale data achter. Deze connectiviteit creëert een laag van verbindingen over ons dagelijkse fysieke leven waardoor wij ons in een hybride samenleving bevinden waarin de fysieke en virtuele wereld met elkaar verweven zijn (De Souza e Silva, 2006).

Een nadeel van de schaalgrootte en continue productie van deze hoeveelheid data is dat burgers en consumenten geen inzicht hebben in de data die zij produceren en in de manier waarop deze vervolgens door bedrijven, overheden en organisaties verzameld en opgeslagen worden. Zo kan er op twee manieren een problematische ongelijkheid ontstaan tussen de 'producenten' van data en partijen die hier voordeel uit kunnen halen. Ten eerste verliest men de controle over de data zelf en ten tweede heeft men geen controle over de intenties van de instantie die de data verzamelt (Groot, 2012).

Het eerste probleem dat dit oplevert is dat gewone internetgebruikers het inzicht verliezen in welke data voor welke periode verzameld, opgeslagen en gebruikt worden. Het feit dat wij zoveel data produceren is geen toeval, in veel gevallen is het design van technologie erop gericht dat er data geproduceerd worden, zodat ze vervolgens verzameld kunnen worden. Bedrijven, overheden en instellingen die data verzamelen ten behoeve van big data-analyse streven ernaar grote hoeveelheden en zo divers mogelijke data bijeen te brengen om vervolgens te analyseren (Groot, 2012). Het design van Facebook bijvoorbeeld nodigt de gebruiker uit om zoveel mogelijk gegevens aan te vullen, zoals welke films hij kijkt, welke merken hij leuk vindt en bij plaatsen hij bezoekt. Hoe meer diverse data Facebook van zijn gebruikers verzamelt, des te meer relaties er ontstaan in de informatie die het bedrijf over zijn gebruikers heeft. En deze informatie wordt door Facebook gebruikt om zijn waarde als advertentiemedium te vergroten. Facebook stelt bedrijven in staat om heel

precies (groepen) mensen te bereiken die binnen het klantprofiel van dit bedrijf passen en hen aan te spreken met boodschappen die op hun profiel zijn afgestemd (Tate, 2013).

Er wordt veel kritiek geleverd op deze data-verzameldrang van grote bedrijven en de ongelijkheid die dit oplevert, onder andere door Amerikaanse wetenschapper en schrijver over technologie Doc Searls. Hij stelt zelfs de vergelijking dat klanten gezien worden als vee voor de jacht op persoonlijke data: "[w]e wouldn't need to be tracked if we weren't being cattle. And we won't solve the privacy problem until customers appear to vendors in human form" (Searls, 2012a: 84). De Nederlandse organisatie Bits of Freedom ziet deze verzameldrang als een aantasting van de privacy en vrijheid op het internet. "Die vrijheid is niet vanzelfsprekend. Je gegevens worden opgeslagen en geanalyseerd. Je internetverkeer wordt afgeknepen en geblokkeerd" (Bits of Freedom, 2013). De organisatie richt zich vooral op het ongevraagd verzamelen van persoonlijke data, het langer dan toegestaan opslaan en het verkopen van de data zonder medeweten van de eigenaar. Bits of Freedom streeft voor een open en neutraal internet dat voor iedereen toegankelijk is en waar privacy niet geschonden wordt, en het lijkt alsof de manier waarop big data nu door grote bedrijven en overheden gebruikt worden hier geen plaats in heeft.

Het tweede probleem is dat men geen controle heeft over de intenties van de verzamelende instantie. De gebruiker weet soms niet met welk doel zijn data verzameld worden en kan hier dan ook lastig een mening over hebben. Op het eerste gezicht lijkt het verzamelen van losse databronnen immers vaak geen directe gevolgen te hebben, maar door de databases met elkaar te verbinden en vervolgens te analyseren ontstaan patronen. En het zijn deze patronen op basis waarvan bedrijven en overheden vervolgens uit het zicht van de gewone internetgebruiker hun beslissingen kunnen nemen. Dit kan voor hen onverwachte gevolgen hebben stelt Rein Groot, verdediger van digitale burgerrechten bij Bits of Freedom: "[d]enk aan een hypotheek die je niet krijgt omdat Big Data vertelt dat jouw huis in een gebied met veel wanbetalers staat. Of een zoekmachine die je bepaalde websites niet laat zien omdat Big Data vertelt dat die sites niet passen bij jouw soort profiel" (Groot, 2012).

1.3 Big data: too big to handle

Niet iedere vorm van big-dataverzameling vindt achter gesloten deuren plaats. Met name in het publieke domein komt er steeds meer data beschikbaar voor burgers, in de vorm van open overheidsdata. In Nederland is de site data.overheid.nl hier een voorbeeld van. Deze site stelt diverse groepen mensen in staat om data die door de Nederlandse overheid is verzameld te hergebruiken voor eigen toepassingen, bijvoorbeeld in het onderwijs of op het gebied van sociale innovatie. Ook in het commerciële domein zijn data steeds vaker vrij beschikbaar. Zo stelt Google

bepaalde informatie over het zoekgedrag van zijn gebruikers beschikbaar. Dit lost natuurlijk het hierboven beschreven probleem niet op: een individuele gebruiker is nog steeds niet in staat om precies bij te houden wat er wel of niet over hem verzameld wordt en wat hiermee door derden gedaan wordt.

Er is echter nog een ander probleem dat veroorzaakt wordt door de schaalgrootte van big data. Zelfs al zouden alle data vrij beschikbaar zijn, dan nog steeds ligt het verzamelen, verwerken en analyseren van databases die de capaciteit van een traditionele computer te boven gaan buiten het bereik van de meeste individuen. Dit is de basis van het tweede grote punt van kritiek: de beoogde voordelen van het gebruik van big data komen alleen ten goede aan grote organisaties die in staat zijn iets met de enorme hoeveelheden data te doen. Het plukken van de vruchten van big data ligt hiermee doorgaans buiten het bereik van burgers en consumenten, en zelfs, binnen de commerciële sector, van kleinere bedrijven. Dit gaat natuurlijk niet op wanneer de grote, data verwerkende bedrijven en overheden besluiten om hun klanten en burgers van deze voordelen te laten genieten, maar ook in dat geval is er sprake van een ongelijke relatie waarbij de macht over de data nog steeds bij de grote instanties ligt.

Om dit punt van kritiek goed te begrijpen, moeten we eerst wat dieper ingaan op wat de factoren zijn die big data 'big' maken. Dit is de hoeveelheid data (volume), de snelheid waarmee deze geproduceerd worden (velocity) en de diversiteit van deze data (variety) (McAfee & Brynjolfsson, 2012).

Het volume en de snelheid van de data maken dat er een gigantische hoeveelheid opslagcapaciteit nodig is om al deze gegevens op te kunnen slaan: "[n]aar schatting verzamelt bijvoorbeeld Wal-Mart per uur 2,5 petabytes aan data van klanttransacties. Eén petabyte is een biljard bytes, het equivalent van wat je in ongeveer 20 miljoen archiefkasten aan tekst zou kunnen opslaan" (McAfee & Brynjolfsson, 2012: 19). Het opslaan van 2,5 petabyte per uur is niet voor iedereen weggelegd. Natuurlijk is dataopslag de laatste jaren een stuk goedkoper is geworden, van $193.000,- per gigabyte in 1980 naar $0,07 per gigabyte in 2009 (Komorowski, 2009). Iemand hoeft geen wiskundige te zijn om te realiseren dat dataopslag uit het bovenstaande voorbeeld een dure aangelegenheid is. Eén van de vaak beschreven voordelen van big data-analyse, namelijk dat je van te voren niet hoeft te weten waar je naar opzoek bent, omdat je toch alles op kan slaan, wordt hiermee een nadeel: "[b]ig data is big. It takes a lot of hardware to store and analyze peta-, yetta-, or (eventually) zettabytes of data. Even using commodity hardware solutions brings a hefty price tag. Public clouds like Amazon Web Services or OpenStack can help mitigate these costs, but there's still a lot of operational support with which to deal. This kind of costs can easily put big data out of the reach of many companies (Proffitt, 2012).

Naast het probleem van opslag, is er ook een probleem met het verwerken van big data. Voorstanders van het gebruik van big data stellen dat de toegenomen rekenkracht van computers één van de belangrijkste drijvende factoren van de populariteit van big data-analyse is. Gordon Moore, visionair van de grote chipsfabrikant Intel, stelde in 1965 de Wet van Moore. Hiermee voorspelde hij dat het aantal transistors op een chip iedere twee jaar verdubbelt (Intel, 2013). Bovendien worden de datasets vaak gedistribueerd opgeslagen, waarbij ze verdeeld worden in stukjes over een gedistribueerd netwerk. Er wordt dan gebruik gemaakt van de rekenkracht van meerdere computers, waardoor er indien nodig relatief gemakkelijk extra capaciteit toegevoegd kan worden (Zwaneveld, 2012).

Dataopslagcapaciteit mag dan wel betaalbaarder zijn geworden en rekenkracht eenvoudig toe te voegen, deze twee samen maken nog geen relationele database waar gemakkelijk gegevens uit onttrokken kunnen worden (Stillman, 2013). Voordat de data bruikbaar zijn om geanalyseerd te worden dienen ze eerst netjes gestructureerd te worden. De Amerikaanse systeemwetenschapper Russell Ackoff beschrijft in 1989 het proces waarbij data tot informatie leidt in de DIKW-hiërarchie (data, kennis, informatie en wijsheid-hiërarchie). Data an sich zijn feitelijk 'symbolen' die op zichzelf geen betekenis hebben, omdat ze geen vormende context hebben. Informatie heeft meer betekenis dan data, omdat het een relatie tussen data afspiegelt. Wanneer verschillende data vergeleken worden leveren de relaties hiertussen informatie op. Als de informatie vervolgens geïnterpreteerd wordt levert dit bepaalde kennis op; er worden conclusies verbonden aan de informatie. Ten slotte geeft wijsheid een vorm van handelingsperspectief voor de toekomst. Iemand met veel wijsheid wordt geacht een goed begrip van de context van de materie te hebben en is in staat principes van een systeem te begrijpen (Ackoff, 1989).

Om big data goed te kunnen structureren, analyseren en vervolgens de verworven kennis te implementeren zijn specialisten nodig die niet standaard in elke organisatie aanwezig zijn. Laat staan dat van individuele burgers en consumenten verwacht mag worden dat zij hiertoe in staat zijn. Zo adviseert het Verkenningsinstituut Nieuwe Technologie dat datawetenschappers idealiter samenwerken met een aantal andere functies binnen bedrijven: "[b]usiness intelligence professionals, marketeers, sales en het it-management moeten met datawetenschappers samenwerken om Big Data in de vingers te krijgen. Behalve werving is dus ook opleiding vereist" (Bloem *et al.*, 2013: 20). En hoewel een groot aantal (business) weblogs, congressen en artikelen in vakbladen gewijd zijn aan instructies over het gebruik van big data, en er zelfs wetenschappelijke opleidingen op het gebied van data-analyse en –interpretatie zijn, is volgens sommigen het gebruik van big data in de praktijk simpelweg te complex voor de gewone leek (Stillman, 2013). De kritiek op big data die te horen is vanuit het perspectief van kleine organisaties kan hierbij net zo goed betrekking hebben op het perspectief van de individuele burger of consument "[a]dding another barrier to small businesses and organizations capitalizing on big data is the fact that big data can be

hard. Hadoop and MapReduce, which are right now topping the chatter charts, are a very good tools to use for commodity storage and analysis. But they can also be very difficult to manage--just ask anyone who had to create a MapReduce batch analysis job" (Proffitt, 2012).

Het opslaan en verwerken van grote hoeveelheden informatie, die in een hoog tempo vanuit verschillende hoeken op je af komen, blijkt in de praktijk dus lang niet voor iedereen uitvoerbaar. Het zijn niet voor niets vaak de succesverhalen van grote bedrijven en overheden, wanneer er positieve verhalen over big data naar buiten worden gebracht. Veel bedrijven en organisaties willen wel, maar slagen er niet in het ecosysteem juist te implementeren. Voor kleine bedrijven, individuen en burgers is het simpelweg niet realistisch om de complexe en kostbare infrastructuur aan te leggen, waarmee van de schaalgrootte van big data geprofiteerd kan worden: "[n]eem retail. Boekverkopers in fysieke winkels konden traceren welke boeken verkochten en welke niet. Als ze een loyaltyprogramma hadden, waren sommige verkopen te relateren aan individuele klanten. Dat was het wel zo'n beetje. Toen het online shoppen doorbrak, groeide het inzicht in klanten enorm. Online retailers konden niet alleen achterhalen wat klanten kochten maar ook waar ze verder nog naar keken: hoe ze door de site navigeerden, in hoeverre ze werden beïnvloed door acties, reviews en de pagina-lay-out en wat overeenkomsten waren tussen individuen en tussen groepen. Al snel kwamen er algoritmen die voorspelden welke boeken individuele klanten graag zouden lezen – algoritmen die steeds beter werden na elke keer dat klanten aanbevelingen opvolgden of juist negeerden. Traditionele retailers hadden eenvoudigweg geen toegang tot dit soort informatie, laat staan dat ze er adequaat op konden reageren" (McAfee & Brynjolfsson, 2012: 19).

Naast de kritiek op de ongelijke verhouding tussen zij die data produceren en zij die deze data verzamelen, is er dus ook kritiek op de ongelijkheid tussen de grote instanties die iets met deze data kunnen en kleinere bedrijven en instellingen, burgers en consumenten voor wie het benutten van deze data niet binnen handbereik ligt.

1.4 Small data als alternatief discours

In de voorgaande twee paragrafen hebben we gezien dat zowel optimistische als kritische geluiden een plek hebben in het big data discours. Wat echter ook duidelijk is geworden is dat veel van de beloofde voordelen met name toekomen aan een relatief kleine groep overheden en grote bedrijven. Ten eerste hebben internetgebruikers vaak geen controle over wat er met hun data gebeurt en ten tweede zijn de benodigde technologie en competenties lang niet voor iedereen beschikbaar, waardoor zij niet de vruchten van de analyse van hun eigen gegevens kunnen plukken. Concreet komen overheden en bedrijven steeds meer te weten over burgers en consumenten, zonder dat zij expliciet toestemming hebben gegeven tot het gebruik van hun data, zonder dat zij op basis van

deze kennis zelf in staat worden gesteld om betere beslissingen te maken en zonder dat de eventuele financiële voordelen naar hen terugvloeien.

Deze problemen vormen de basis van de kritiek op big data. Het is natuurlijk op zichzelf geen probleem dat de voordelen van een bepaalde technologie niet gelijk gedistribueerd zijn. Nergens staat dat dit zo zou moeten zijn: in onze westerse vrijemarkteconomie is het gebruikelijk dat bedrijven concurrentievoordeel hebben van bepaalde innovatie en overheden zijn niet verplicht al hun kennis continu met hun burgers te delen. Maar volgens de critici zorgen de problemen rondom big data voor een groeiende afstand tussen overheid en burger, tussen bedrijf en consument en tussen expert en leek, binnen de context van een medium – het internet – dat groepen juist meer met elkaar in contact zou moeten kunnen brengen.

Een deel van de kritiek op big data komt niet zomaar uit de lucht vallen. Op de voedingsbodem van de kritische teksten in het big data discours ontstaat het laatste jaar een contradiscours, waarin de voor- en nadelen van een alternatief op big data besproken wordt. En dit alternatief is 'small data'. Het verschil tussen de teksten over small data en de kritische teksten uit het big data discours is dat in de teksten over small data niet alleen op de mogelijke problemen rondom big data wordt ingegaan, maar dat deze worden gepresenteerd in context van de mogelijke oplossing die small data kunnen bieden. De manier waarop het begrip 'small data' gepresenteerd wordt geeft de indruk dat de term 'small' bewust gekozen is: als antwoord op de problemen die de schaalgrootte van big data met zich meebrengt. Small data worden als nieuwe belofte als alternatief voor big data gepresenteerd: "Baffled By Big Data? Use 'Small' Data Instead" (Sillman, 2013), "Beyond Big Data: Why Small Data is Taking Off" (SDG, 2013a) en "Big-data benefits with small data: Smaller companies are turning to data, too" (Proffitt, 2012).

Maar wat zijn small data? En wat maakt dat ze zo wezenlijk verschillen van big data? Afhankelijk van wie je deze vraagt stelt zal je een ander antwoord krijgen, maar op het eerste gezicht tekenen zich drie verschillende groepen teksten af in het discours. In deze drie soorten teksten komende steeds andere voordelen van het gebruik van small data naar voren en worden verschillende subjectposities mogelijk gemaakt waar vanuit deze voordelen beoordeeld kunnen worden. Zo zijn er bijvoorbeeld auteurs die de kleine schaal van small data zien als oplossing voor het probleem dat big data alleen door getrainde experts beheerd en geanalyseerd kunnen worden. Hun teksten richten zich met name op de subjectposities van de burger en de overheid en richten zich op het democratiseren van dataverwerking, met als doel om burgers in het publieke domein de voordelen te laten plukken van de data die door henzelf geproduceerd worden. Rufus Pollock, o.a. oprichter van de Open Knowledge Foundation definieert small data als volgt: "[s]mall data is the amount of data you can conveniently store and process on a single machine, and in particular, a high-end laptop or server" (Pollock, 2013b). Kortom, in tegenstelling tot big data kunnen small data

opgeslagen, beheerd en verwerkt worden met gewone computers of laptops. En dit betekent, volgens Pollock, dat geavanceerde data-analyse binnen handbereik komt van de geïnteresseerde leek: [w]hat matters here then is, crudely, the amount of data that an average data geek can handle on their own machine, their own laptop" (Pollock, 2013b).

Andere auteurs richten zich met name op de kansen die small data voor kleinere bedrijven kunnen bieden ten opzichte van de multinationals. Waar de eerste groep auteurs zich met name richt op de democratiserende werking van small data, ten opzichte van big data richt, lijkt deze groep vooral kleine bedrijven aan te willen spreken op het zakelijk voordeel. Het is in de tegenstelling tussen de subjectposities van de kleine, lokale onderneming, versus het grote bedrijf dat hun definitie van small data vorm krijgt: "[m]ake no mistake, small does not mean insignificant. Think of it this way: If we leveraged data in local the same way large e-commerce companies do, we could have a greater impact on commerce than Big Data, since more than 90 percent of retail transactions still occur offline in local businesses. But most of us don't. The reality is that most local businesses haven't figured out how to capture or use this data to grow profits" (Ho, 2013).

Een derde groep andere auteurs ziet het gebruik van small data vooral als oplossing voor het probleem dat consumenten het eigenaarschap en de controle verliezen over de data die zij zelf geproduceerd hebben, wanneer deze voor big data-analyse worden gebruikt. Zo gebruikt Doc Searls de termen 'small data' en 'my data' onderling uitwisselbaar, hiermee implicerend dat de kleinere schaal van small data individuele gebruikers in staat stelt de eigen data te blijven beheren en in hun eigen voordeel te gebruiken. Het lijkt alsof hij wil zeggen dat 'big data' synoniem is aan 'not my data (anymore)'. Dit type teksten hebben betrekking op de subjectposities van de consument en de aanbieder en proberen het voordeel van small data te ontsluiten voor de consument.

Kortom, het lijkt erop dat de manier waarop small data als alternatief worden gepresenteerd voor big data afhangt van het specifieke probleem van big data dat men aan wil pakken. Afhankelijk van de invalshoek die in een bepaalde tekst gekozen wordt, wordt het probleem anders gedefinieerd en wordt er een andere oplossing bij gezocht. Voor nu lijkt de belangrijkste conclusie dan ook dat op het eerste gezicht het begrip 'small data' op zichzelf geen intrinsieke betekenis heeft. Er wordt van buitenaf betekenis aan de term gegeven door hen die kritiek op big data willen uiten, maar deze betekenis richt zich steeds op andere kernmerken van small data. Bovendien spelen er steeds andere subjectposities een rol: waar bijvoorbeeld sommigen spreken over de verhouding tussen consument en aanbieder, spreken anderen over de verhouding tussen burger en overheid.

Om uiteindelijk een uitspraak te kunnen doen over de waarde van small data als alternatief voor big data, moeten we het discours waarin het begrip betekenis krijgt analyseren. Uit verschillende teksten spreken andere belangen: waar sommigen strijden voor een betere wereld en een vrijer internet, zien anderen small data oplossingen simpelweg als commerciële kans. De vraag

is op welke manier deze belangen bepalend zijn voor de betekenis die aan small data gegeven wordt.

De rest van deze scriptie staat in het teken van deze zoektocht. Aan de hand van een analyse van het small data discours, op basis van teksten uit de drie verschillende invalshoeken die ik hierboven heb beschreven, probeer ik de belangen bloot te leggen die ten grondslag liggen aan de argumenten van de verschillende deelnemers aan het discours. Aan de hand van een Foucauldiaanse discoursanalyse kijk ik naar de wijze waarop het begrip small data vorm krijgt en of de verschillende teksten erin slagen de kritiek op big data geloofwaardig over te brengen en small data als alternatief te presenteren. Uiteindelijk leidt dit tot een conclusie over de manier waarop het small data discours als contradiscours vorm krijgt.

Hoofdstuk 2

Methodologie
Foucauldiaanse discoursanalyse

Dit hoofdstuk staat in het teken van de Foucauldiaanse discoursanalyse, geïnspireerd op het gedachtengoed van de Franse denker Michel Foucault. Hiermee analyseer ik de betekenisvorming van de term small data. Immers, de manier waarop er geschreven en gesproken wordt – het discours – biedt volgens Foucault inzicht in de maatschappelijke gedeelde betekenis die een bepaald begrip heeft (Foucault, 1982: 10). En deze gedeelde betekenis is belangrijk, want volgens Foucault ontstaat kennis, en daarmee 'waarheid', niet op autonome wijze in een authentiek individu, maar zijn de belichaming van bepaalde machtsstructuren binnen een samenleving. In de moderne samenleving waarin wij leven, zijn de begrippen 'macht' en 'kennis' volgens Foucault wederkerig: de manier waarop kennis wordt geconstrueerd is volgens Foucault onlosmakelijk ingebed in machtsstructuren. Omgekeerd creëert de inzet van kennis een bepaalde macht voor de instituties die hun kennis over een onderwerp onder subjecten verspreidt (Lambrechts, 1981: 545-6). Waarheid is bij Foucault dus datgene wat waar wordt gemaakt, oftewel "de verzameling van regels volgens welke men het ware en het onware scheidt, en men aan het ware specifieke machtseffekten toekent" (Fontana, 1977: 26).

Omdat er rondom het begrip small data zoveel verschillende, soms tegenstrijdige belangen spelen, is het interessant om in dit relatief vroege stadium, waarin small data nog niet veel meer is dan een belofte, naar het ontstaan van discours te kijken. Het gaat er hierbij niet om wat small data precies inhouden. Het gaat erom dat we erachter komen hoe de partijen die belang hebben bij de macht om de term small data te betekenis te geven het discours beïnvloeden. Om deze relevantie te kunnen zien is het belangrijk om goed te begrijpen wat de kern van het gedachtengoed van Foucault is, wat een discours is en hoe het gevormd wordt, en hoe het geanalyseerd kan worden.

2.1 Michel Foucault

Michel Foucault (1926-1984) is een Franse intellectueel die geassocieerd wordt met de stromingen van het poststructuralisme en het postmodernisme in de westerse filosofie. De invloed van het gedachtegoed van Foucault beperkte zich niet alleen tot de filosofie, maar had ook zijn uitwerking in diverse andere disciplines, zoals sociologie, antropologie en geschiedenis (Mills, 2003: 1; Gutting, 2013). De filosofie van Foucault is bovendien instrumenteel van aard. Niet alleen beschrijft hij abstracte concepten, hij geeft concrete aanwijzingen voor hoe men door zijn bril de wereld kan

analyseren. Zo stelt hij: “[a]l mijn boeken zijn kleine gereedschapskisten, zo u wilt. Wanneer men ze wil openmaken, gebruik wil maken van deze of gene zin, van dit of dat idee of van een bepaalde analyse zoals men een schroevendraaier of een sleutel gebruikt, om de machtssystemen, [...] kort te sluiten, te diskwalificeren of kapot te maken” (Lambrechts, 1981: 553-4).

De rode draad in het werk van Foucault is relatie tussen het subject en zijn waarheid: hoe het individu over zichzelf denkt en over zijn vermogen om de wereld om zich heen te kennen. In zijn werk probeert Foucault op een nieuwe manier de mogelijkheidsvoorwaarden van ons “alledaags weten” vast te stellen (Lambrechts, 1981: 554). Volgens hem zijn we door de heersende filosofische opvattingen van de moderniteit gaan geloven in de belofte dat de mens een autonoom rationeel en daardoor vrij wezen is. Een wezen dat de wereld om zich heen steeds beter kan leren begrijpen en zo vooruitgang kan boeken op wetenschappelijk en moreel vlak. Maar uiteindelijk, zo betoogt hij, zijn waarheid en rationaliteit vooral machtsmiddelen van de heersende instituties die onze zoektocht naar de waarheid over de wereld en onszelf vormgeven (Oosterling, 2012).

Dit denken ontstaat bij Foucault als hij in 1961 het boek *Folie et déraison. Histoire de la folie à l'âge classique* schrijft. Hierin beschrijft hij hoe de betekenis van het begrip ‘gekte’ door de geschiedenis heen veranderd is. In de Renaissance werden gekken in de kunst en literatuur afgebeeld als wijzen of als mensen die hun eerlijke, ware aard tonen. Tijdens de Verlichting vervolgens, de tijd van de rede, werden de gekken beschreven als ‘de onredelijken’ en van de samenleving geïsoleerd. In ons moderne tijdvak (‘épistèmè’ in de woorden van Foucault), worden gekken als aparte groep bestempeld die speciale zorg nodig heeft en ‘beter’ gemaakt moet worden. De grote vraag die dit bij Foucault oproept is of we dus wel kunnen spreken van de ware betekenis van het begrip ‘gek’, of dat de huidige medische wetenschap ons simpelweg in staat heeft gesteld om de gekken te kunnen controleren met de belofte naar deze waarheid op zoek te zijn (Foucault, 2006a).

Volgens Foucault focust de moderniteit zich op de autonomie van ‘het subject’, terwijl er feitelijk achterliggende regels zijn. Deze regels worden ons niet van bovenaf opgelegd, volgens de piramidevormige machtsstructuur die de klassieke samenleving kenmerkte. Wij hebben deze regels geïnternaliseerd, simpelweg door in deze tijd te leven en deel te nemen aan het discours: hoe we spreken en hoe we schrijven. We worden in ons denken gecontroleerd, zonder dat we door hebben op welke manier en door wie (Oosterling, 2012). Kennis geeft ons geen vrijheid, zoals we zo graag willen geloven, de wil tot waarheid oefent juist een dwingende macht uit (Lambrechts, 1981: 547). Onze individualiteit is niet de bron van onze kennis, maar het gevolg van het discours waaraan wij deelnemen. Het overbrengen van kennis is feitelijk het onderwerpen van het individu aan wat ‘normaal’ is (Oosterling, 2012).

Vandaar de focus van Foucault op de relatie tussen kennis en macht. Vanuit een destijds nieuwe invalshoek haalt de Fransman fel uit naar het humanisme, de filosofie die als uitgangspunt de

mens als kennend subject centraal stelt. Daarentegen verdedigt Foucault de antisubjectfilosofie, die juist dit idee verwerpt door te stellen dat het subject niet ten grondslag van het filosofisch denken ligt (Lambrechts, 1981: 521-2). Hij eindigt het boek *De woorden en de dingen* (origineel in 1966), dat zijn grote doorbraak betekende, met de woorden: "[d]e archeologie van ons denken toont eenvoudig aan dat de mens een uitvinding van recente datum is. En mogelijk ook dat zijn einde nabij is. Als deze disposities zouden verdwijnen [...], dan zouden we er zeker van kunnen zijn dat de mens zal verdwijnen, als een gezicht in het zand op de vloedlijn van de zee" (Foucault, 2006b: 453). Hij bedoelt hiermee natuurlijk niet dat er feitelijk mensen van de aardbodem zullen verdwijnen, maar dat het conceptuele begrip 'de mens', het subject als autonoom kennend wezen, uit ons denken zal verdwijnen.

2.2 Het discours

Maar hoe kunnen we dan voorbij de schijn van de vanzelfsprekendheid van onze kennis komen? Dit kan volgens Foucault door analyse van het discours. Specifieker spreekt Foucault van het geheel aan uitspraken dat tot dezelfde 'discursieve formatie' behoort: "wanneer men in een groep van uitspraken eenzelfde referentiesysteem kan opsporen en beschrijven, een bepaalde mate waarin uitspraken van elkaar afwijken, eenzelfde theoretisch netwerk en eenzelfde veld van strategische mogelijkheden" (Foucault, 1968: 29).

In *De orde van het vertoog* (origineel in 1971) begint Foucault direct met zijn belangrijkste stelling: "[i]k veronderstel dat in elke maatschappijvorm de productie van het gesproken en geschreven woord gecontroleerd, geselecteerd en georganiseerd wordt, en dan opnieuw onder de mensen gebracht" (Foucault, 1982: 10). Dit gebeurt, zo vervolgt hij, "door middel van een aantal procedures die tot taak hebben het vermogen en de gevaren van het woord te bezweren, het vervreemdend gebeuren dat erin besloten ligt te bedwingen, en het drukkende en angstaanjagende aspect ervan uit de weg te ruimen" (Foucault, 1982: 10). Door analyse van het vertoog (de Nederlandse vertaling van discours) worden die procedures zichtbaar waarmee het discours binnen de grenzen van 'het normale' gehouden wordt. Immers, zo schrijft Foucault verderop, "[het discours is] niet eenvoudigweg datgene waarin strijd of stelsels van overheersing tot uitdrukking worden gebracht, maar ook datgene waarmee wordt gestreden, de macht waarvan men zich meester tracht te maken" (Foucault, 1982: 11). Het discours geeft dus niet alleen een inkijkje in de heersende machtsstrijd, maar is zelf ook de het middel en de inzet van deze strijd.

In het vervolg van *De orde van het vertoog* beschrijft Foucault drie sets van de hierboven genoemde procedures. Ik zal, in navolging van Foucault, in het volgende hoofdstuk deze procedures (ook wel:

uitsluitingsmechanismes) toelichten en op basis hiervan het discours rondom de belofte van small data analyseren.

Hoofdstuk 3

Discoursanalyse small data

Zoals toegelicht in hoofdstuk één is het doel van dit onderzoek om te beschrijven hoe het small data discours als counterdiscours functioneert tegenover het discours over big data. Foucault heeft ons laten zien dat wij, om dit te kunnen begrijpen, moeten kijken naar de manier waarop machtsstructuren binnen het small data discours tot uitdrukking komen en zo de betekenis van het begrip small data vormgeven. Met andere woorden, hoe worden verhalen en kennis over small data bewerkt (gecontroleerd, geselecteerd en georganiseerd) en in circulatie gebracht? In dit hoofdstuk ga ik op zoek naar de impliciete regels in het discours die uitdrukking geven aan deze machtsstructuren. Zij zorgen ervoor dat de deelnemers aan het discours, geen radicale uitingen doen, zich aanpassen aan de algemeen geaccepteerde normen en dat eventuele "angstaanjagende of benauwende aspecten uit de weg worden geruimd" (Foucault, 1982: 10).

Om het small data discours te analyseren, maak ik gebruik van drie sets van procedures, zoals Foucault deze aanreikt in *De orde van het vertoog.* Het zijn deze procedures die er als het ware voor zorgen dat de deelnemers aan het discours binnen de lijntjes blijven en dat zij die dat niet doen automatisch worden uitgesloten. Foucault onderscheidt hierbij drie typen procedures: externe uitsluitingsmechanismes, interne uitsluitingsmechanismes en regels voor deelnemers aan het discours. En ieder type procedure bestaat weer uit een aantal concrete verschijningsvormen (Foucault, 1982). Per set van procedures analyseer ik de verschillen, overeenkomsten, belangen en tegenstrijdigheden in verschillende teksten uit het small data discours.

Ik volg hierbij de globale indeling van het discours, zoals ik deze beschreven heb in hoofdstuk één, waarin drie invalshoeken op de betekenis van big data te onderscheiden zijn. Ten eerste zijn er de teksten waarin het gebruik van small data als democratisch alternatief voor big data-analyse worden beschreven; door de kleinere schaal is het verzamelen en analyseren van small data, in tegenstelling tot big data, wel mogelijk voor de gewone burger. Ten tweede zijn er de teksten waarin small data met name betekenis krijgen als een manier om ook kleine bedrijven de voordelen van data-analyse te laten plukken. Dit in tegenstelling tot big data, waarvan gesteld wordt dat alleen grote bedrijven in staat zijn om deze te analyseren en om te zetten in waardevolle kennis. Ten slotte zijn er teksten waarin small data gezien worden als de hoeveelheid data die door een consument over zichzelf verzameld, beheerd en geanalyseerd kan worden en zo zijn positie ten opzichte van aanbieders versterkt.

3.1 Externe uitsluitingsprocedures

De eerste set van externe procedures bestaat volgens Foucault uit drie methoden van uitsluiting, namelijk het verbod, de tegenstelling tussen rede en waanzin en het onderscheid tussen waar en onwaar.

Het verbod is de meest duidelijke en meest voorkomende vorm van uitsluiting, aldus Foucault. Niet iedereen heeft hetzelfde recht van spreken, sommige dingen 'mogen' nou eenmaal niet. Bijvoorbeeld omdat er een taboe op heerst op het onderwerp of omdat de spreekomstandigheden bepaalde privileges of rituelen vergen (Foucault, 1982: 10). Het is opvallend dat in teksten waarin over de democratiserende kracht van small data in de publieke sfeer gesproken wordt, er met geen enkel woord gerept wordt over de mogelijkheden die small data bieden voor consumenten en voor kleine bedrijven. Zo ziet de Open Knowledge Foundation het als zijn missie om te voorkomen dat de kennis en macht die met het analyseren van grote hoeveelheden data gepaard gaat, in de handen van slechts een paar grote instanties terecht komt: "[…] to prevent the centralisation of data-power in the hands of the few. Redistributing knowledge power is what we're all about at the Open Knowledge Foundation" (OKF, 2013a). Er wordt gestreden voor 'openheid' van kennis, zoals de naam van de stichting ook impliceert. De mogelijkheden die dit op kan leveren worden echter alleen in de publieke sfeer gezocht: "making government more accountable, culture more accessible and science more efficient" (OKF Labs, 2013). Er wordt niet gesproken over mensen in hun rol als consument of kleine ondernemer, maar alleen over hen die sociale verbetering op de agenda hebben: "civic hackers, data wranglers and ordinary citizens intrigued and excited by the possibilities of combining technology and open information for good" (OKF Labs, 2013).

Hier lijkt sprake te zijn van een ongeschreven verbod, zoals Foucault dat bedoelt. Met andere woorden, het is een taboe voor leden van de Open Knowledge Foundation om te schrijven en spreken over hoe hun kennis benut kan worden ten behoeve van winstmaximalisatie of een betere consumentenpositie. Andersom negeren zakelijke blogs de mogelijkheden van small data in de publieke sfeer. Zij richten zich alleen op de mogelijkheden van small data voor kleine bedrijven. Een interessant voorbeeld is de Small Data Group, die in eerste instantie de indrukt wekt small data als algemeen democratiserende technologie te zien: "[small data] is really where the value is created, opinions are formed, insights are shared, and action are made. By non data scientists. Everyday" (SDG, 2013a). Deze algemene quote over small data kan net zo goed vanuit de subjectpositie van de ondernemer of marketeer, als vanuit de subjectpositie van de burger gelezen worden. Uiteindelijk richten zij zich echter op de mogelijkheden voor marketeers: "[t]his shift impacts (and enables) how marketers can create more effective way to reach and engage customers and employees, and how

they can bring their stories to life" (SDG, 2013c). Het lijkt onwaarschijnlijk dat het begrip small data binnen deze invalshoek betekenis krijgt als technologie waarmee burgers de controle over hun eigen data weer terug krijgen en applicaties kunnen maken voor het algemeen nut.

Het tweede middel tot uitsluiting kan in de methode van Foucault blootgelegd worden via de indeling van mensen tussen rede en waanzin. Dit is een scheidslijn die de tegenstelling tussen rede en waanzin gebruik om "hen wiens spreken een zekere geldigheid bezit" af te zetten tegen "hen wiens spreken geen waarheid kan behelzen noch ook van enig belang kan zijn" (Foucault, 1982: 11). Foucault doelt hiermee in op het spreken van de gek, de krakzinnige die in de geschiedenis meerdere betekenissen heeft gekend. Waar de gek in vroeger tijden gezien werd als iemand die de wereld tot nieuwe inzichten kon brengen, wordt de gek in de moderne tijd juist zoveel mogelijk buiten het discours geplaats (Foucault, 2006a). Bij het analyseren van het small data discours als tegenmacht tegenover het big data discours zien we eenzelfde soort verschuiving van het begrip.

Binnen het big data discours is een veelgehoorde kritische mening als het gaat om dataverzameling die van de Russische Evgeny Morozov. Morozov veroordeelt de tirannie van bedrijven die persoonlijke data-verzamelen, zoals dit in Silicon Valley gedaan wordt en de manier waarop dit wordt toegejuicht door de publicaties van bijvoorbeeld O'Reilly Media. De Rus stelt dat in deze Amerikaanse bedrijven een 'geek culture' heerst, een cultuur die beheerst wordt door technische innovators die een sterke invloed hebben op ontwikkelingen van grote commerciële organisaties. Er heerst volgens hem in deze omgeving een vorm van naïef optimisme, waarbij men er vanuit gaat dat ieder maatschappelijk 'probleem' een technische oplossing heeft. Hij noemt dit 'solutionisme' en veroordeelt het Amerikaanse solutionisme dat heerst in Silicon Valley. Techniek kent volgens velen in Silicon Valley alleen maar positieve effecten en daardoor worden mensen blind voor de negatieve effecten, aldus Morozov (Morozov, 2013). Het is opvallend dat hij bij het benoemen van deze groep hamert op de term 'geek'. Het lijkt erop dat Morozov de oorspronkelijke, denigrerende betekenis, van geek in ere probeert te stellen. Het woord stamt af van het Oud Engelse 'geck', waar ook het Nederlandse gek van afgeleid is (Oxford Dictionaries, 2013). Morozov sluit deze geeks uit, omdat hij hen niet staat acht kritisch na te denken over de gevolgen die dataverzameling op grote schaal met zich meebrengt, zij zijn blind voor de negatieve kanten hiervan. Om met de woorden van Foucault te spreken: ze zijn niet in staat tot redelijk nadenken, hun mening doet er niet toe volgens hem.

In (een deel van) het small data discours krijgt de term juist een geheel andere betekenis. In de teksten waarin small data worden beschreven als democratiserende technologie voor burgers, zijn het de geeks die een tegenmacht kunnen vormen tegen de overheden, grote bedrijven en organisaties die zich met big data-analyse bezig houden. Small data worden hier duidelijk als alternatief voor big data neergezet en zijn volgens bijvoorbeeld Pollock de sleutel tot een

democratischere vorm van data-analyse, waarin geeks en andere geïnteresseerde leken kunnen samenwerken om tot nieuwe inzichten te komen: "[w]hat's interesting (and new) right now is the democratisation of data and the associated possibility of large-scale distributed community of data wranglers working collaboratively. What matters here then is, crudely, the amount of data that an average data geek can handle on their own machine, their own laptop" (Pollock, 2013b). Binnen deze discursieve formatie worden geeks als mogelijke subjectpositie betrokken bij het onderwerp, in hun rol als deelnemers aan een netwerk van geïnteresseerde burgers die samen met de data aan de slag gaan, terwijl we eerder zagen dat Morozov binnen het big data discours de geek buiten het discours probeert te plaatsen.

Als derde externe systeem van uitsluiting bespreekt Foucault het streven naar waarheid via de indeling waar en onwaar. De experts, zij die de zoektocht naar de waarheid vormgeven, zijn bepalend voor wat wij als waar en onwaar zien. Dit houdt verband met de manier van uitsluiting op basis van de tegenstelling tussen rede en waanzin, zoals hiervoor besproken. Foucault merkt op dat de termen rede en waanzin tot een bepaalde hoogte arbitrair zijn. Hij vraagt zich af of het verschil tussen waar en onwaar niet op dezelfde arbitraire manier gezien kan worden, met andere woorden de geproduceerde waarheid wordt gebruikt om andere mensen buiten te sluiten (Foucault, 1982: 13-9). Om volgens Foucault te kunnen analyseren wie vanuit welke positie welke waarheid nastreeft, moeten we niet focussen op het niveau van één propositie of stelling binnen het discours, maar moeten we ons afvragen wat de zoektocht naar waarheid is binnen het gehele discours. Oftewel, we moeten kijken naar welk type uitspraken binnen het discours überhaupt waar kunnen zijn. Hij noemt dat "in de waarheid". Want hierdoor vallen er automatisch ook bepaalde uitspraken "buiten de waarheid:" zij kunnen binnen het discours niet waar of onwaar zijn en worden dus uitgesloten.

In de drie groepen teksten in het small data discours lijken op het eerste gezicht hetzelfde type uitspraken in de waarheid te vallen. Uitspraken over het beoogde voordeel van het gebruik van small data vallen binnen de waarheid en de kritiek op big data waar hun pleidooi voor small data op gebaseerd is ook. Het lijkt daarmee dat het small data discours een coherent en consistent contradiscours is. Uit de verschillende teksten spreken echter verschillende motivaties om uitspraken over het beoogde voordeel van het gebruik van small data als 'waar' te bestempelen, net als uitspraken over het nadeel van het gebruik van 'big data'. Hun zoektocht naar de waarheid is een andere. Des te meer reden om door een Foucauldiaanse bril naar het discours te kijken en het oppervlakkige onderscheid tussen waar en onwaar niet zonder meer aan te nemen.

Uit de teksten waarin over de democratiserende werking van small data in het publieke domein wordt gesproken, blijkt een duidelijke mening over big data. Uitspraken over het beoogde voordeel van big data-analyse voor de burger worden in dit discours als onwaar bestempeld. Uitspraken over de keerzijde van de schaalgrootte van big data voor burgers en over small data als

alternatief zijn waar: “[t]his next decade belongs to distributed models not centralized ones, to collaboration not control, and to small data not big data” (Pollock, 2013a). En: “[t]he discussions around big data miss a much bigger and more important picture: the real opportunity is not big data, but small data. Not centralized ‘big iron’, but decentralized data wrangling. Not ‘one ring to rule them all’ but ‘small pieces loosely joined’” (Pollock, 2013a). De zoektocht naar de waarheid binnen deze teksten beperkt zich echter tot het voordeel van het gebruik van small data tegenover big data voor de burger: “[w]e believe that a vibrant open knowledge commons will empower citizens and enable fair and sustainable societies” (OKF, 2013a). Eventuele uitspraken over het gebruik van small data buiten het publieke domein en de voordelen van small data voor consumenten en/of bedrijven vallen in deze teksten dus buiten de waarheid en lijken zo vanuit het discours uitgesloten te worden.

De zoektocht naar waarheid rondom small data in de teksten van bijvoorbeeld Allen Bonde en de Small Data Group richten zich daarentegen juist op de vraag: “wat is de potentiële waarde van small data voor bedrijven?”. Deelnemers aan het discours die vanuit deze positie redeneren, geven small data betekenis als een aanpak die voordeel op kan leveren voor bedrijven: “brands and investors can benefit from this small data approach, if they focus on creating simple, smart, responsive, socially aware tools and solutions” (SMG, 2012). Vaak worden small data als handzaam formaat van big data beschouwd: “Blaffled By Big Data? Use ‘Small’ Data Instead” (Sillman, 2013). Om die reden is het opslaan, verwerken en analyseren van small data, en de voordelen die hiermee te behalen zijn, voor een grotere groep beschikbaar: “time to bring the power of big data to the masses” (Bonde, 2013). Met ‘masses’ doelt Bonde hier waarschijnlijk op ‘alle bedrijven in plaats van enkel de allergrootsten’ en niet op burgers, consumenten of ‘data geeks’. Zo linkt hij naar een artikel van een collega, waarin het bedrijfsperspectief expliciet gemaakt wordt: “[i]n the world of local businesses, we're not talking about big data. Instead we're dealing with Small Data” (Ho, 2013). Ho, een collega van Bonde en schrijft: “[b]ig data holds lots of promise, sure, but for small businesses, realizing the benefits of this trend may still be pretty far out of reach” (Stillman, 2013). Kortom, deze teksten richten zich met name op uitspraken over de onwaarheid van de stelling: “big data is goed voor bedrijven”. Big data worden namelijk vooral als onhandig en ingewikkeld ervaren (Stillman, 2013; SDG, 2013). Small data worden daar tegenover als simpeler en handzamer alternatief gepositioneerd en dat small data-analyse wél kansen en voordelen kan bieden aan bedrijven is daarmee een ware uitspraak. In deze teksten liggen dus uitspraken die gaan over de waarde van small data voor bedrijven binnen de waarheid.

De derde discursieve formatie die probeert small data als alternatief voor big data te positioneren, beperkt zich tot de specifieke relatie tussen een bedrijf en zijn klanten en tussen aanbieders en consumenten. In eerste instantie lijkt de zoektocht naar de waarheid in deze teksten gelijk aan die in de teksten van bijvoorbeeld de Open Knowledge Foundation, in de zin dat de waarheid rondom small data vooral gezocht wordt in de democratiserende werking voor het

individu: “[t]here is going to be this tectonic shift, as everything sort of re-shapes and re-centers itself around people, around individuals, and around the mountains of data that they have … Everybody talks about ‘big data’. This isn’t big data. This is going to be the era of small data, of my data” (Jeremie Miller geciteerd in Searls: 2012: 196). Deze teksten richten zich echter met name op de subjectpositie van de consument, waar bijvoorbeeld Pollock vooral over de waarde van small data voor burgers spreekt: "[t]his book stands with the customer. This is out of necessity, not sympathy" (Searls, 2012a: 1). Volgens Searls is het verzamelen van consumentendata op zichzelf geen probleem, mits consumenten maar niet de controle verliezen over hun data: "[p]roducing and integrating data sets of all sizes can be a good and useful thing - especially if customers get to do it too, with their own data" (Searls, 2012a: 79). Bovendien schrijft hij expliciet dat dit moet leiden tot een betere positie van de consument ten opzichte van de aanbieder: “[i]n short, we want to see confident, empowered consumers able to make the right choices for themselves – to get the best deals, demand better products or services, and be able to resolve problems when things go wrong” (Searls, 2012a: 195). In deze teksten zijn dus uitspraken over de waarde van small data voor de consument waar en vallen uitspraken over de waarde van small data voor burgers en bedrijven juist weer buiten de waarheid. En hoewel Searls in eerste instantie de positie van de consument probeert te emanciperen ten opzichte van de aanbieder, en meent dat de aandacht de komende tijd naar het door ontwikkelen van small-datatoepassingen uit moet gaan, voorziet hij uiteindelijk een win-winsituatie, waarbij bedrijven nog steeds gebruik kunnen maken van big data: “the end result is that the ‘small data’ that’s yours will be more important than ‘big data’ behind marketing’s guesswork. The two will in the long run dance together. But for now the small data side needs to get its act together. And it will” (Searls, 2012a: 196).

De analyse van de externe uitsluitingsmechanismen doet vermoeden dat er niet alleen drie verschillende invalshoeken in het small data discours zijn, waarin het begrip small data steeds op een andere manier als tegenmacht tegenover big data betekenis krijgt, maar dat er zelfs sprake is van drie gescheiden discursieve formaties. Uit analyse van het verbod kan immers geconcludeerd worden dat zij die zich bezig houden met de democratiserende kracht van small data voor burgers en in het publieke domein op geen enkele manier ingaan op de commerciële waarde ervan voor bedrijven en consumenten. En hoewel alle drie de groepen teksten zich tegen ‘big data’ af lijken te zetten, is hun zoektocht naar de waarheid die small data behelzen en de subjectposities die hierin mogelijk worden gemaakt sterk verschillend.

3.2 Interne uitsluitingsprocedures

Naast de groep procedures die het discours van buitenaf afbakent zijn er ook procedures die als controlestelsel binnen het discours werkzaam zijn. Ze kenmerken zich door classificaties, ordeningen of verdelingen aan te brengen om de dimensie "van het gebeuren en van het toeval" binnen het discours te beheersen (Foucault, 1982: 19). De interne procedures zijn die procedures die zich binnen het discours zelf afspelen en direct gerepresenteerd woorden door tekst: wie reageert op wie? Hoe ontstaat een gedeeld begrippenkader? Hoe wordt dit begrippenkader versterkt door herhaling en variatie? Om de impliciete structuren die in het discours aanwezig zijn bloot te leggen onderscheidt Foucault drie interne uitsluitingsprocedures: het commentaar, de auteur en de disciplines (Foucault, 1982: 19-22).

De eerste procedure is het analyseren van commentaar, het reageren op reeds bestaande tekst. Als voorbeeld noemt Foucault klassieke historische teksten waar in de loop der tijd veel over gediscussieerd is en commentaar op is geleverd. Voor velen zou dit iets zeggen over de kwaliteit van de tekst, maar Foucault kijkt naar de manier waarop commentaar geleverd wordt. Het zou namelijk zo kunnen zijn dat commentaar leveren op een algemeen geaccepteerd historisch werk iets doet met de status van de commentator, de auteur en de tekst. Het begrippenkader dat in de tekst vorm krijgt, vindt door het commentaar zijn vaste plaats in het discours door herhaling van betekenis en identiteit (Foucault, 1982: 20-6).

De meest basale verschijningsvorm van commentaar in het small data discours is het feit dat het discours als geheel commentaar is op het big data discours. Zoals eerder beschreven wordt in de meeste teksten over small data expliciet gereageerd op teksten uit het big data discours: de voordelen van small data vinden hun voedingsbodem in kritiek op big data. Er is echter binnen het small data discours sprake van een interessanter verschijnsel, dat meer zegt over de dynamiek binnen het discours zelf. Het betreft hier het gegeven dat het commentaar zich met name manifesteert binnen de grenzen van de drie verschillende invalshoeken in dit discours. De deelnemers gaan vanuit de verschillende invalshoeken niet met elkaar in discussie over de betekenis van het begrip, maar proberen allen om vanuit hun eigen invalshoek een begrippenkader te bestendigen waarmee small data als alternatief voor big data neergezet kan worden.

De reden hiervoor moet in de teksten zelf gezocht worden. Het is niet zo dat de deelnemers in het discours fysiek niet in staat zijn om op elkaar te reageren. Discussies over small data vinden voornamelijk online plaats, een enkel boek en congres daargelaten, zoals met de meeste onderwerpen die met name de internetgemeenschap aangaan het geval is. In een dergelijke omgeving is reageren de norm, in de vorm van inhoudelijke reacties of door te linken of te retweeten, wat in feite ook vormen van commentaar leveren zijn. Het is ook niet zo dat de verschillende auteurs

niet van het bestaan van de andere teksten op de hoogte zijn. Zo lijkt Bonde op het eerste gezicht te reageren op de teksten van Pollock. Hij plaatste een artikel van Pollock op de hoogste positie in zijn top vijf van aanbevolen artikelen over small data (SDG, 2013b) en reageert één keer direct: "[l]ove the post/topic! The way I've been thinking about Small Data is as the 'last mile' of Big Data - we (often) need Big Data behind the scenes, but the trick is to provide simpler, more consumer-style apps and tools at the front end that work on any device, foster social sharing and help non-technical users turn insights into actions...that are actually helpful in the moment. Readers can see my latest thoughts here: [link]" (Pollock, 2013a). Na een kort compliment, maakt Bonde direct van de gelegenheid gebruik om zijn visie op small data te geven en naar zijn eigen werk te verwijzen. Inhoudelijk blijft echter iedere vorm van reactie uit.

Het lijkt erop dat de werkelijke reden voor het gebrek aan commentaar tussen de verschillende invalshoeken gezocht moet worden in het feit dat de bijbehorende teksten weinig anders met elkaar te maken hebben dan het gebruik van de term 'small data'. Waar bijvoorbeeld Doc Searls spreekt over 'customers' als subject, spreekt Rufus Pollock over 'citizens' en spreken de auteurs rondom de Small Data Group van Allen Bonde over 'local businesses'. Zelfs wanneer Bonde spreekt over 'consumer-style' apps, laat hij de interpretatie open dat het niet gaat om apps voor consumenten, maar om zakelijke apps die net zo makkelijk in het gebruik zijn als consumenten apps. Dus hoewel zij allen het gebruik van small data als een vorm van empowerment zien, worden in de verschillende teksten totaal andere subjectposities mogelijk gemaakt; met name met betrekking tot de vraag wie er baat zou hebben bij het gebruik van small data. Hierdoor lijkt de betekenis van het begrip 'small data' per invalshoek wezenlijk te verschillen, waardoor de drie invalshoeken lastig aan elkaar te relateren zijn. Zoals ook bij het bestuderen van de zoektocht naar de waarheid in de teksten bleek, is de inzet van de teksten uit de drie verschillende invalshoeken anders; er wordt een andere strijd geleverd. Feitelijk maken ze gebruik van een geheel ander begrippenkader en dit maakt het leveren van inhoudelijk commentaar over en weer zeer lastig.

Foucault noemt als tweede uitsluitingsprocedure de auteur. Hiermee bedoelt hij het "beginsel van de groepering van de woorden van het vertoog, als de eenheid en oorsprong van de erin vervatte betekenissen, en als verzamelpunt van de samenhang die ertussen bestaat" (Foucault, 1982: 23). Met andere woorden: het gaat niet er niet om van iedere tekst te weten wie precies de auteur is, maar om te begrijpen op welke manier een tekst op basis van de gebruikte begrippen, redeneringen en stijlmiddelen wordt toegeschreven aan een auteursfiguur, omdat sommige teksten in het vertoog juist hieraan hun betekenis en doeltreffendheid ontlenen (Foucault, 1982: 23-5).

De drie invalshoeken in het small data discours kennen allemaal hun centrale auteur. Niet alleen zijn zij 'auteur' in de traditionele zin van het woord, ook zijn zij auteur in de Foucauldiaanse zin van het woord: hun auteursfiguur wordt zichtbaar in de teksten door het gebruik van specifieke

inhoudelijke kenmerken, begrippen en stijlmiddelen, welke kunnen worden toegeschreven aan het feit dat zij de auteurs zijn (Foucault, 1979). Ook in deze analyse wordt duidelijk hoe de drie invalshoeken zich van elkaar onderscheiden.

De teksten die zich richten op de rol van de burger en de mogelijkheden die small data binnen het publieke domein bieden, vinden hun samenhang in de autoriteit van Rufus Pollock. Hij schreef in 2006 het boek *Value of the Public Domain*, waarin hij pleit om intellectuele werken in het publieke domein open te stellen, zodat burgers deze vrij kunnen gebruiken, verspreiden en hergebruiken (Pollock, 2006: 3). Het is dan ook niet verwonderlijk dat de teksten over small data van zijn hand zich allen in ditzelfde betekenisveld bevinden en de mogelijkheden van small data direct koppelen aan open kennis, tegenover het voor de gewone burger gesloten karakter van big data-analyse. Ook is zijn rol als oprichter van de Open Knowlegde Foundation duidelijk terug te zien in de manier waarop deze stichting open data bevordert en open toegankelijke kennis stimuleert. Begrippen als openheid, samenwerking, burgerschap en gedeelde kennis staan centraal in de manifesten en teksten van de stichting.

De teksten die betrekking hebben op de mogelijkheden van small data voor kleine bedrijven kennen een net zo duidelijke auteurstoeschrijving. De meeste teksten binnen deze invalshoek zijn geschreven door Allen Bonde en ook door zijn achtergrond en invalshoek krijgen deze teksten betekenis en winnen zij aan doeltreffendheid. In de teksten draait het om de rol van data in ondernemerschap, concurrentie en 'business intelligence'. Small datatoepassingen zijn in deze teksten een economisch middel, dat aan bedrijven wordt aangeraden door een auteursfiguur die zelf achtergrond heeft in de adviespraktijk. Bonde gebruikt zijn jarenlange ervaring in het bedrijfsleven, onder andere als consultant bij McKinsey, om nieuwe initiatieven voor het bedrijfsleven te ontplooien en te steunen. En small data lijken zo'n initiatief. Ook de autoriteit van de Digital Clarity Group en de Small Data Group wordt met name ontleend aan de regelmatige blogposts van Bonde. De samenhang in de betekenis van small data in deze teksten kan gevonden worden in de mogelijkheden die small data aan bedrijven bieden om concurrerend te zijn.

De teksten die small data benaderen vanuit de positie van de consument en de aanbieder zijn van de hand van Doc Searls c.s. en project VRM, dat Searls leidt bij het aan Harvard gelieerde *Berkman Center for Internet and Society*. Hier is duidelijk het verschil te zien tussen de traditionele omschrijving van het auteursbegrip en de manier waarop Foucault dit duidt. De persoon Searls kent een divers verleden. Zelf spreekt hij over: "the many hats I wear" (Searls, 2012b). De auteur Searls, zoals deze binnen het small data discours betekenis krijgt, heeft echter een consistente inhoudelijke focus. Hij probeert vanuit het perspectief van de consument de infrastructuur van het internet te gebruiken om commerciële transacties op een andere – voor consumenten interessante – manier te organiseren, en small data zijn hier een middel voor. Deze auteurspersoon richt zich duidelijk op de relatie tussen consumenten en bedrijven en probeert hierin de kant van de consument te kiezen: "we

still need the freedom and power to do more with personal data than we can now, when too much of it is in the hands of others who don't care, can't make good use of it, or are likely to misuse it" (Searls, 2012a: 190). Evengoed bedient hij zich van een begrippenkader dat meer bij het bedrijfsleven aansluit: customers, vendors, markets, profits, opportunities etc. Hiermee vervult de auteur een interessante brugfunctie tussen de betekeniswereld van de consument die van de voordelen van small datatoepassingen zou moeten kunnen profiteren en de betekeniswereld van de bedrijven die deze toepassingen mogelijk moeten maken.

De laatste interne uitsluitingsprocedure vindt Foucault in de 'disciplines'. De grenzen van disciplines laten zich goed uitleggen aan de hand van de wetenschappen. Iedere discipline wordt beperkt door de kaders van zijn vakgebied, omdat hier bepaalde regels aan grenzen. Zoals een set van methodes, regels, definities en aannames die geldend zijn binnen een paradigma (Foucault, 1982: 26). Hierbij gaat het er niet om of de uitspraken in teksten binnen deze disciplines waar of onwaar zijn, maar hoe binnen bepaalde disciplines andere meningen buitengesloten worden, door kennis die buiten de marges van de discipline liggen naar de achtergrond te schuiven. Hierdoor ontstaat eenheid binnen de discipline, wat het mogelijk maakt om binnen hetzelfde begrippenkader tot nieuwe inzichten te komen: "[w]il er sprake zijn van een discipline, dan moet er dus de mogelijkheid tot formuleren – en wel tot eindeloos verder formuleren – van nieuwe stellingen bestaan" (Foucault, 1982: 26).

De teksten van Pollock c.s. en de Open Knowledge Foundation vallen binnen de discipline van sociale innovatie. De kern hiervan wordt goed verwoord in de omschrijving van het doel van de Open Knowledge Foundation: "dedicated to promoting open data and open content in all their forms – including government data, publicly funded research and public domain cultural content" (OKF, 2013a). Binnen deze discipline wordt gekeken naar de rol die open data hebben om sociale innovatie mogelijk te maken. Uitspraken over toepassingen van small data binnen aandachtsgebieden als bijvoorbeeld marketing en bedrijfskunde hebben niets te maken met overheidsdata, onderzoek uit de publieke sector en openbare culturele uitingen en vallen duidelijk buiten het domein van deze discipline.

Deze uitspraken horen eerder thuis bij bedrijfskundige en economische disciplines, waar de teksten van bijvoorbeeld Bonde en Pollock zich in begeven. Hun teksten lijken voor eenzelfde type publiek geschreven en deelnemers aan het discours met een zelfde soort achtergrond toe te laten. Bonde stelt dat zijn doel is: "[to] provide insightful perspectives, research and advisory […] for companies looking to leverage the full potential of social analytics and large-scale business intelligence" (SDG, 2013). De titel van Searls zijn boek over small data is niet voor niets *The Intention Economy*, hij geeft hiermee duidelijk aan in welke context men zijn teksten moet lezen.

Toch zit er veel verschil in de teksten van deze twee prominente deelnemers aan het small data discours, met name in de subjectposities die zij mogelijk maken. Bonde schrijft met name over kleine (lokale) bedrijven, hun concurrentiekracht tegenover grote ondernemingen en de rol van small data in de bedrijfsvoering. In feite kijkt hij dus naar small data vanuit de discipline van ondernemerschap, waarin vanuit bedrijven gekeken wordt naar hun positie in de markt en mogelijkheden om te groeien. Searls zijn teksten zijn beter in te delen binnen de marketingdiscipline, juist omdat hij zich richt op de tweezijdige relatie tussen consumenten een aanbieders.

Het lijkt het erop dat bij analyse van de interne uitsluitingsprocedures het patroon versterkt wordt, dat ook bij analyse van de externe uitsluitingsprocedures zichtbaar werd. Namelijk dat er sprake is van niet één, maar drie gescheiden discursieve formaties binnen het small data discours, ieder met hun eigen begrippenkader. In de teksten wordt met name gereageerd op het big data discours en op andere teksten binnen dezelfde invalshoeken, maar zelden op teksten uit een andere invalshoek over small data. Bovendien kennen de drie invalshoeken een duidelijke Foucauldiaanse auteurstoeschrijving en sluiten aan bij drie verschillende disciplines: sociale innovatie, ondernemerschap en marketing.

Twee van deze invalshoeken staan diametraal tegenover elkaar: zij werken vanuit een andere discipline en hun teksten maken geheel andere subjectposities mogelijk. De teksten van Doc Searls nemen echter een middenpositie in. De inzet van de strijd van Searls en zijn betoog voor small data als tegenmacht tegenover big data heeft veel weg heeft van de strijd van Pollock en de Open Knowledge Foundation: beide strijden voor empowerment van het individu, in zijn positie als burger of consument. De teksten van Searls betrekken echter juist wel het commercieel georiënteerde publiek bij het discours, in tegenstelling tot de teksten van Pollock.

3.3 Regels voor deelnemers aan het discours

Niet ieder subject heeft volgens Foucault in gelijke mate toegang tot een discours. De deelnemers zijn gebonden aan een aantal impliciete regels die het discours reguleren. Naast de impliciete en expliciete uitsluitingsmechanismen hebben ook deze regels volgens Foucault een uitsluitende werking. Foucault onderscheidt vier typen gedragsregels: het (woord)ritueel, de vertooggemeenschappen, de doctrines en maatschappelijke toe-eigening (Foucault, 1982: 30-7).

Het ritueel bepaalt volgens Foucault de kwalificaties die deelnemers aan het discours moeten bezitten om te mogen spreken. Deze rituelen zitten verholen in “de gebaren, het gedrag, de omstandigheden en het geheel van tekens die tot vertoog behoren” (Foucault, 1982: 32). Zij die zich niet aan deze codes houden, of deze niet kennen, zijn uitgesloten van deelname. Rituelen in het

small data discours zouden gezocht kunnen worden in het gebruik van jargon. Welk taalgebruik is op welke plaats gepast?

Als we naar het jargon in de discipline van sociale innovatie binnen het discours kijken valt het op dat er gesproken wordt over 'communities', die samenwerken in 'mass-participation activity'. Het gaat hier om de activiteiten van groepen burgers die massaal samen kunnen werken aan het realiseren van open dataprojecten. Het begrippenkader richt zich op samenwerking: 'democratisation of data', 'large-scale distributed community' en 'working collaboratively' (Pollock, 2013b). Dit wordt ook duidelijk in de volgende uitnodiging: "[w]e welcome everyone who shares our vision of promoting open data and open content in all their forms to join our community, whatever your background and skillset" (OKF, 2013a). Iedereen die mee wil helpen open data in het publieke domein te promoten is welkom om mee te werken binnen de samenwerkingsverbanden van de gemeenschap.

De disciplines van ondernemerschap kenmerkt zich door een geheel ander jargon. Een jargon dat vooral denkt in termen van zakelijke transacties. Hier verschijnen individuen niet als burgers, maar als 'customers' en bedrijven als 'sellers' of 'vendors'. Eén van de doelstellingen van SDG illustreert het zakelijk karakter: "[o]ur goal is to provide thoughtful, independent viewpoints on top vendors and early adopters" (SDG, 2013b). Wanneer over small data gesproken wordt, is dat vooral als middel dat zakelijk voordeel kan brengen: "this shift impacts (and enables) how marketers can create more effective way to reach and engage customers" (SDG, 2013c).

Het jargon van de marketingdiscipline verschilt iets van het ondernemerschapsjargon. In principe worden dezelfde zakelijke termen gebruikt, maar het is in de hedendaagse marketingdiscipline niet ongebruikelijk om ook vanuit het voordeel voor de consument te redeneren. Men gaat er vanuit dat het internet een fundamentele verandering betekent voor de machtsrelatie tussen consument en aanbieder: "[b]ig business continues to believe that a free market is one in which customers get to choose their captors. Choosing among AT&T, Sprint, T-Mobile and Verizon for your new smartphone is like choosing where you'd like to live under house arrest. It's why marketers still talk about customers as 'targets' they can 'acquire,' 'control,' 'manage' and 'lock in,' as if they were cattle. And it's why big business thinks that the best way to get personal with customers on the Internet is with 'big data,' gathered by placing tracking files in people's browsers and smartphone apps without their knowledge—so they can be stalked wherever they go, with their 'experiences' on commercial websites 'personalized' for them" (Searls, 2012c). Het is opvallend hoe Searls het jargon van meerdere disciplines tegelijk inzet: "[this] can only be made possible by the full empowerment of individuals inside a new field called VRM, for vendor relationship management. VRM works on the demand side of the marketplace: for you, the customer, rather than for sellers and third parties on the supply side" (Searls, 2012c). Woorden als "full empowerment of individuals" zouden evengoed uit de discipline van de sociale innovatie komen, terwijl "the demand side of the

marketplace" juist weer bij het ondernemerschapsjargon aansluiten. Juist door dit bijna schizofrene gebruik van twee totaal verschillende begrippenkaders, lijkt Searls aan te willen sluiten bij het hedendaagse marketingparadigma dat zakelijke groei juist voortkomt uit het beter behandelen van consumenten. Want, zoals de ondertitel van Searls zijn boek zegt: "customers take charge".

Vervolgens beschrijft Foucault het functioneren van de 'vertoogsgemeenschappen', die er voor zorgdragen dat het discours bewaard blijft en voort blijft bestaan. Een vertooggemeenschap bestaat uit een groep mensen die samengebracht worden door een bepaalde collectieve identiteit of ideologische overtuiging. Binnen deze besloten kringen circuleert het discours volgens strikte regels. Vertooggemeenschappen hebben echter geen duidelijke, maar vaag omlijnde kaders. Het kunnen (in vroegere tijden) geheime gemeenschappen zijn die zich bevinden op het snijvlak van geheimhouding en verspreiding: wat binnen de gemeenschap gebeurt is niet openbaar, maar nieuwe mensen kunnen wel lid worden (Foucault, 1982: 33). Een voorbeeld hiervan is een groep mensen die een bepaalde religie of levensovertuiging aanhangt: de rituelen binnen de religie zijn niet openbaar, maar na lange studie en inwijding kan men lid worden van de gemeenschap en op dat moment worden de rituelen ontsloten. In hoeverre is hier een parallel zichtbaar met de small data-vertooggemeenschap die een alternatief probeert te ontwikkelen naar aanleiding van de big data-analyse hype?

In het geval van small data biedt het veel inzicht om te kijken hoe bepaalde teksten gedistribueerd worden. Deze distributie vindt voornamelijk online plaats. De vertooggemeenschap van de sociale innovatie discipline concentreert zich rondom de site van de Open Knowlegde Foundation van Pollock. Naast de nieuwsfunctie en informatieve functie van de site is het ook een plek waar kennis volgens de strikte regels waar Foucault over spreekt gedistribueerd kan worden. Zo linkt OKF door naar de Shuttleworth Foundation, waar Pollock eveneens bij betrokken is. Deze stichting financiert de OKF Labs dataprojecten. Eenzelfde constructie is er opgezet met OpenGLAM dat ook door OKF geïnitieerd is met als doel het cultureel erfgoed digitaal toegankelijk maken. Om voortzetting van het discours te stimuleren heeft OKF het Panton Fellowship ontwikkeld. Dit is een onderzoekspositie voor wetenschappers die onderzoek doen naar open data. Ten slotte werkt OKF wereldwijd met een groot aantal soortgelijke organisaties samen aan projecten die open data gebruiken ten behoeve van het publieke domein. Voorbeelden hiervan zijn: OpenSpending, The Public Domain Review en School of Data (OKF, 2013b). De OKF laat de distributie van het discours dus via een groot aantal, maar zorgvuldig gecureerde initiatieven verlopen. Hoewel deze organisaties de indruk wekken dat ze veel dataprojecten realiseren zijn er, naast de teksten over wat small data zouden moeten zijn, nagenoeg geen concrete voorbeelden waarbij een project door middel van het gebruik van small data gerealiseerd is.

De vertooggemeenschap van de ondernemerschapsdiscipline is met name georganiseerd via de sites waar Allen Bonde zelf bij betrokken is, namelijk: Small Data Group en Digital Clarity Group. Beide sites worden gekenmerkt door een prominente blogfunctie. Hier verschijnt continu nieuws, zoals uitkomsten van nieuwe onderzoeken, evenementen of nieuwe ontwikkelingen. SDG focust zich zoals de naam al doet vermoeden alleen op small data, waarbij nagenoeg alle blogposts door Bonde zelf geschreven zijn. Hij neemt een actieve functie binnen de vertooggemeenschap, waar de lezers een minder actieve functie innemen. Bonde verzamelt op zijn blog het laatste nieuws over small data op andere voorname websites. Waar de Open Knowledge Foundation zich dus expliciet als groeiende gemeenschap profileert, zijn de SDG en DCG veel meer gesloten van karakter. Ook deze gemeenschap schrijft enkel over de mogelijkheden van small data en heeft zelf geen deel in het realiseren van small datatoepassingen.

Voor de positie binnen het small data discours waar vanuit de marketingdiscipline wordt geredeneerd zijn small data meer een middel dan een doel op zich. Auteurs als Searls lijken uiteindelijk te streven naar een omwenteling in de marketingdiscipline met het Project VRM. De vertooggemeenschap beperkt zich met name tot leden van de *Berkman Center for Internet & Society* van de universiteit van Harvard; een onderzoekscentrum gericht op de impact van het internet op de samenleving. Binnen het Berkman Center leidt Searls het Project VRM, met zijn eigen site en mailinglist (VRM, 2013a). Dus hoewel de vertooggemeenschap met het boek van Searls probeert om marketeers te bereiken, zijn de leden met name academici, wat een mogelijke verklaring kan zijn voor het dubbele begrippenkader dat in hun teksten wordt gebruikt. Vanuit Project VRM wordt er (in sommige gevallen samen met bedrijven) gebouwd aan een aantal small datatoepassingen, zoals Info en Singly. Beide projecten richten zich op het realiseren van een "personal information storing and sharing application" (VRM, 2013b). Door deze tools te gebruiken kunnen consumenten hun online data zelf organiseren, zoals Searls dit in de zeven richtlijnen omschrijft die ik verderop in deze paragraaf noem in de doctrine van de marketingdiscipline.

Kortom, de vertooggemeenschap rondom de discipline van sociale innovatie lijkt zoveel mogelijk burgers bijeen te willen brengen. De vertooggemeenschap die small data-oplossingen als bedrijfsmiddel voor beter ondernemerschap tracht te verkopen blijft echter beperkt tot één belanghebbende, Bonde, die zichzelf als adviseur in het middelpunt plaatst. De vertooggemeenschap die vanuit de marketingdiscipline redeneert, kijkt vanuit de academische wereld naar concrete toepassingen van small data voor het bedrijfsleven en positioneert zich daaromheen.

Waar de analyse van de vertooggemeenschappen laat zien hoe het discours circuleert, laat de volgende procedure – de doctrine – zien onder welke voorwaarden dit gebeurt en op basis van welke criteria de deelnemers aan een vertooggemeenschap hun positie innemen. Analyse van de

doctrine (of: ideologie) kenmerkt zich door het beschrijven van de voorwaarden voor uitbreiding van het discours, in tegenstelling tot analyse van de vertooggemeenschap die vooral de huidige samenstelling van het discours laat zien. Een doctrine is een leer die door een groep mensen wordt aangehangen: "doordat zij één en dezelfde taal tot gemeengoed maken, bepalen individuen, in willekeurig welk aantal, dat zij er over en weer bijhoren" (Foucault, 1982: 35). De enige voorwaarde voor individuen om 'erbij' te horen is dat zij dezelfde waarheden erkennen en zich conformeren aan het type uitspraken die de leer met zich meebrengt. Op die manier bindt de doctrine individuen aan bepaalde soorten uitspraken. De doctrine bepaalt welke soorten uitspraken binnen het vertoog zijn toegestaan en daarmee ook welke soorten niet. Ook zorgen de uitspraken binnen de doctrine ervoor dat zij die de doctrine aanhangen zich onderling verbonden voelen en zich kunnen onderscheiden van alle anderen (Foucault, 1982: 35-6). Het verschil met het simpelweg identificeren van disciplines, zoals we zagen bij de interne uitsluitingsmechanismen, is dat de doctrine meer ideologisch gedreven is op basis van een inhoudelijke of normatieve verbondenheid.

Wanneer we kijken naar de drie disciplines binnen het small data discours, valt op dat iedere discipline zijn eigen groep aanhangers heeft, die vanuit hun eigen achtergrond redeneren en, gebruikmakend van hun eigen begrippenkader, enkel kunnen discussiëren over datgene wat binnen hun eigen waarheid valt. Zolang ze binnen de kaders van deze waarheid blijven communiceren, ontstaat er een soort wederzijdse versterking van deze waarheid, omdat er zo weinig interactie tussen de disciplines plaatsvindt. Dit maakt het echter nog geen doctrine: dat je in het zelfde gelooft, hoeft niet te betekenen dat je daar ook de zelfde ideologische lading aan geeft en de door Foucault benoemde 'onderlinge verbondenheid' voelt. Deze vorm van verbondenheid is het sterkst te zien bij de discipline van de sociale innovatie. Hier wordt ideologisch gesteld dat small data vrij toegankelijk zouden moeten zijn voor iedereen: "[t]he data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form. […] The data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable. […] Everyone must be able to use, reuse and redistribute – there should be no discrimination against fields of endeavour or against persons or groups" (OKF, 2013b). Het taalgebruik zet bij deze positie ook het hardst de positie van de individuele burger en de machthebbende instituties tegenover elkaar: "[i]n a well-functioning, democratic society citizens need to know what their government is doing. To do that, they must be able freely to access government data and information and to share that information with other citizens" (OKF, 2013b). Daar tegenover wordt bijvoorbeeld over big data gezegd dat het juist op een ideologie van centralisatie is gestoeld: "[b]ig data smacks of the centralization fads we've seen in each computing era" (Pollock, 2013a).

Hoewel ook binnen de discipline van ondernemerschap gepleit wordt voor meer toegankelijkheid van data, is dit minder ideologisch gedreven. Soms lijkt dit wel zo te zijn, bijvoorbeeld wanneer men stelt dat: "[i]t's time to bring the power of big data to the masses" (SDG, 2013d). Maar waarom dit zo zou moeten zijn, daar gaan de teksten verder niet op in. Men blijft een hangen in het argument dat de manier waarop big data opgeslagen, verwerkt en geanalyseerd moeten worden bepaalde praktische bezwaren kent, die deze data voor kleinere bedrijven onhandelbaar maakt: "[a]dding another barrier to small businesses and organizations capitalizing on big data is the fact that big data can be hard. Hadoop and MapReduce, which are right now topping the chatter charts, are a very good tools to use for commodity storage and analysis. But they can also be very difficult to manage--just ask anyone who had to create a MapReduce batch analysis job" (Proffitt, 2012). In feite probeert men simpelweg te profiteren van de hype rondom big data om andere data gedreven oplossingen te verkopen: "[s]mall data is […] taking advantage of the hype and hoopla around big data to remind businesses of the general value of data analytics no matter how big the data set is" (Proffitt, 2012). Wanneer er concreet een voordeel van het gebruik van small data besproken wordt, blijft men ook eerder pragmatisch dan ideologisch. Bijvoorbeeld wanneer beschreven wordt hoe small data, beter dan big data, kunnen helpen bij het opzetten van een e-mail marketingprogramma: "[a] 'small data' set includes a known population that is not expected to have changes to its composition in the short term, which allows marketers to use this data to answer a specific question or need. Conversely, 'big data' would represent your large purchased email list with unknowns, possible duplications, and unsubscribes. This list cannot be used for targeted email marketing sends in its current form" (Kettmann, 2013).

De marketingdiscipline kent een simpele doctrine: small data zorgen voor een betere positie van de consument. Men is niet tegen big data-analyse op zichzelf, maar ziet de hype rondom big data-analyse als een uitvloeisel van een paradigma van controle dat de moderne economie beheerst. Searls stelt in het begin van zijn boek, dat hij verwacht dat dit gaat veranderen: "[o]ver the coming years, customers will be emancipated from systems built to control them. They will become free and independent actors in the marketplace, equipped to tell vendors what they want, how they want it, where, and when – even how much they'd like to pay – outside of any vendor's system of customer control" (Searls, 2012a: 1). Hoewel deze quote als een voorspellende beschrijving is geschreven ('customers will'), wordt de ideologische wens duidelijk uit de beschrijving van de situatie waar Searls zich tegen afzet ('systems built to control them') en wat dit uiteindelijk oplevert ('free and independent actors').

In de '10th aniversary edition' van *The Cluetrain Manifesto* beschrijft Searls zeven doelen voor Vendor Relationship Management (VRM), het technologische framework waar deze nieuwe realiteit volgens Searls aan moet voldoen. Ten eerste moet VRM individuele consumenten voorzien van hulpmiddelen om hun relatie met organisaties te beheren, en niet andersom. Ten tweede

moeten consumenten verzamelaars van hun eigen data worden. Ten derde pleit Searls ervoor om consumenten de mogelijkheid te geven om hun data selectief te delen. Ten vierde moeten consumenten in staat worden gesteld om te controleren hoe hun data worden gebruikt door anderen en voor hoe lang. Ten vijfde zou VRM consumenten de mogelijkheid moeten geven om hun eigen servicevoorwaarden in te dienen en, ten zesde, op deze manier hun wensen kenbaar te maken aan de open markt. En ten slotte vindt Searls dat de technologie voor dergelijke oplossingen gebaseerd moet zijn op open standaarden, open API's en open code (Searls, 2012a: 164). Meer dan een feitelijke beschrijving van wat VRM is, leest dit als een set van 'zeven geboden', waaruit een duidelijke doctrine zichtbaar wordt. De keuze voor de term VRM versterkt de manier waarop Searls zich tegen oude situatie afzet. De term is afgeleid van Customer Relationship Management (CRM), de technologie waarmee bedrijven databases met koopgegevens van klanten aan kunnen leggen en hier aanbiedingen op af kunnen stemmen. De werking van deze systemen wordt zichtbaar wanneer bijvoorbeeld mobiele-telefoonaanbieders ervoor zorgen dat je een telefoontje krijgt vlak voordat je abonnement afloopt, waarin je een aanbieding wordt gedaan. Hiermee verkleint de aanbieder de kans dat zijn de klant besluit klant te worden bij de concurrent, zodra een eventueel interessante mogelijkheid zich voordoet.

Het laatste beheerssysteem volgens Foucault is de sociale toe-eigening van het discours. Als instrument hiervoor wijst hij de opvoeding aan, immers dit is het middel waarmee elk individu toegang krijgt tot welke vertoogsvorm dan ook. De opvoeding is bij uitstek de plek waar kennis en macht samenkomen, omdat kennisoverdracht altijd gepaard gaat met overdracht een stelsel van normen en waarden. Foucault noemt dit daarom "een politiek middel om het systeem van toe-eigening van vertoog [...] te handhaven of te veranderen" (Foucault, 1982: 36).

Dit roept de vraag op wat de plekken zijn waar er in het small data discours wordt opgevoed. Natuurlijk worden binnen de vertooggemeenschappen de leden als het ware opgevoed, maar zij nemen feitelijk al op de 'juiste' manier deel aan het discours. Het lijkt dus interessanter om specifiek te kijken naar welke gezaghebbende instituten de massa 'opvoeden' en zo het discours proberen te verspreiden. Zo blijkt dat nieuwssites The Guardian en CBC News het werk van Pollock plaatsen en citeren. Tevens verschijnt hij op een aantal kleinere datablogs. Zijn teksten, gericht op de democratisering van data voor de burger, vinden logischerwijs hun weg naar de burger, via de grote nieuwsmedia, en via de individuele data 'geek' via de datablogs. Bondes werk, gericht op ondernemerschap, bereikt de massa via zakelijke media als Forbes, Inc, Xconomy en een scala aan kleinere blogs over ondernemerschap. Ook hier sluiten de sociale toe-eigening dus één-op-één aan op de positie die de teksten in het discours innemen. Searls' werk vinden we in Wall Street Journal, TechCrunch en DataRoundtable. Zijn positie in het discours, waarin hij de consument met behulp van de technologie van het internet 'bevrijd' van zijn ondergeschikte rol ten opzichte van aanbieders,

komt goed tot uitdrukking in de manier waarop zijn werk zowel via algemene nieuwsmedia als via zakelijke media en via technologie media naar buiten komen. Waar we eerder konden concludeerden dat er tussen de drie disciplines binnen het discours weinig inhoudelijke interactie is, blijkt dit dus ook te gelden voor verdere verspreiding van het discours. De drie disciplines hebben zo hun eigen kanalen om het grote publiek in aanraking te laten komen met hun eigen waarheid over small data.

Analyse van de externe uitsluitingsmechanismen deed vermoeden dat het discours feitelijk in drieën uiteen valt, waarbij de positionering van small data ten opzichte van big data steeds een andere is. Analyse van de interne uitsluitingsmechanismen versterkte dit vermoeden; door de verschillen in het gebruikte begrippenkader en de mogelijke subjectposities in de teksten, redeneren drie disciplines steeds op een andere manier over small data, waardoor een onderlinge dialoog niet of nauwelijks op gang komt. Het gaat hier om de disciplines van sociale innovatie, ondernemers en marketing. De analyse van de regels voor deelname aan het discours zet deze driedeling op scherp. De vertooggemeenschap rondom de OKF, beschrijft small data vanuit de ideologie van 'open data' als manier om een democratischere verhouding tussen individuele burgers en machthebbers op gang te brengen. Geheel kloppend bij hun filosofie van openheid, betrekken zij zoveel mogelijk mensen bij hun activiteiten. De vertooggemeenschap die zich gevormd heeft rondom de SDG verdedigt de positie van het voordeel van small data-analyse voor kleine bedrijven. Het gaat hier om het streven naar slimme zakelijke vernieuwing, minder dan om een ideologisch gedreven doctrine. Want, vernieuwing is niet per se iets dat zou moeten, maar simpelweg iets dat zou kunnen. De vertooggemeenschap rondom Searls c.s. neemt met name de positie in van pragmatische marketingdenkers, die een gemengde rol spelen tussen het belang voor consumenten en bedrijven in. Zij zijn ideologisch gedreven door het geven van macht aan consumenten, maar vinden medestanders voor dit verhaal in zowel de academische- als in de marketingwereld en lijken de taal van beide werelden te kunnen spreken.

Conclusie

In dit onderzoek heb ik geprobeerd om door middel van het uitvoeren van een Foucauldiaanse discoursanalyse meer inzicht te verkrijgen in de beloftes rondom het begrip 'small data'. Ik heb mij hierbij met name gericht op de manier waarop het small data discours een contradiscours vormt ten opzichte van het big data discours. Immers, er is gebleken dat iedere definitie of uitleg van het begrip 'small data' te definiëren vertrekt vanuit een punt van kritiek op big data. Het begrip lijkt enkel betekenis te krijgen als tegenmacht tegenover de hype rondom big data. Alle voorstanders van het gebruik van small data wijzen op een aantal van de problemen die spelen rondom big data. Deze problemen hebben ik in hoofdstuk één beschreven en de kritiek blijkt uiteen te vallen in twee hoofdpunten.

Ten eerste zouden big data te groot zijn om te controleren, wat zoveel wil zeggen als dat individuele burgers en consumenten het inzicht verliezen in wat er met de data gebeurt, die over hen verzameld is. Zij kunnen er niet meer bij, verliezen de toegang en verliezen daarmee de controle over de bedrijven en overheden die de data verzamelen. Bovendien hebben ze niet de mogelijkheid om de vruchten te kunnen plukken van de inzichten die uit hun eigen data zouden kunnen voortkomen, zo luiden de argumenten van de critici. De voorstanders van small data positioneren vervolgens small data hier als oplossing tegenover: door gebruik te maken van kleinere datasets, kan de individuele consument of burger zelf besluiten welke data hij wel of niet ontsluit aan overheden of bedrijven, en kan men zelf de eigen data analyseren. Mensen als Doc Searls proberen hiermee de machtsverhouding tussen consumenten en bedrijven te verschuiven. Zij beschrijven een mogelijk nieuwe marketingrealiteit, waarin de consument zelf beter in staat is om over zijn eigen voorkeuren en behoeften na te denken en deze behoeften op zijn eigen voorwaarden aan een geselecteerde groep aanbieders te ontsluiten. Aan de andere kant zijn er mensen als Rufus Pollock, die vanuit een bredere filosofie over 'open data' beargumenteren waarom het goed zou zijn dat data niet gecentraliseerd in grote datasets worden opgeslagen, maar gedistribueerd in kleinere data sets worden aangeboden aan individuele burgers. Zij zouden op basis van de gezamenlijke en open analyse van deze data tot inzichten en toepassingen kunnen komen, die een positieve bijdrage leveren aan de samenleving als geheel, los van de invloed van grote instellingen en overheden.

Het tweede argument tegen het gebruik van big data, is dat de grote datasets die hiervoor nodig zijn, te groot zouden zijn om door consumenten, burgers én kleine bedrijven opgeslagen, verwerkt en geanalyseerd te worden. Zelfs al zou men controle over en toegang tot de eigen data hebben, ontbreekt het iedereen behalve zeer grote bedrijven aan de kennis en de financiële middelen om de infrastructuur neer te zetten die hiervoor nodig is, zou luidt het argument. De voorstanders van small data menen vervolgens dat door het gebruik van kleinere data sets dit wel

mogelijk is. Zo benoemd de al eerder genoemde Rufus Pollock de kracht van tientallen, zo niet honderden, 'data geeks' die wereldwijd samen kunnen werken in een decentraal netwerk, waarbij allen een klein stukje van de data-puzzel oplossen. Allen Bonde bekijkt ditzelfde probleem vanuit een geheel andere hoek en stelt dat het gebruik van small data kleinere, lokale ondernemingen in staat stelt om ook de vruchten van data-analyse te kunnen plukken.

Het opvallende aan deze beweging is niet alleen dat de betekenis van het begrip small data enkel en alleen wordt gezocht in de oplossing die small data zouden kunnen vormen voor (één van) de problemen die spelen rondom big data. Het meest opvallende, blijkt na analyse, dat ook de kritiek die door deze auteurs op het gebruik van big data wordt geleverd zich beperkt tot alleen die argumenten die het eigen punt versterken. Men spreekt bijvoorbeeld niet over het fundamentele bezwaar dat je zou kunnen hebben tegen het verzamelen van grote hoeveelheden data an sich. Of over de methodologische vraag of big data-analyse wel tot betere inzichten kan leiden. De reden hiervoor lijkt simpel: voor deze problemen kunnen small data de oplossing niet zijn. Het lijkt er dus op dat het hele discours over small data alleen kan bestaan bij de gratie van zijn rol als tegendiscours. Zonder big data zou het begrip 'small data' niet hebben bestaan. En zonder 'small data' als oplossing, zouden de auteurs in kwestie wellicht niet geïnteresseerd zijn geweest in de problemen rondom big data die ze aankaarten.

Vandaar dat ik in hoofdstuk drie dieper ben ingegaan op de wijze waarop het small data discours als contradiscours functioneert. Ik heb de vraag proberen te beantwoorden welke verschillende posities er te onderscheiden zijn in de betekenisvorming rondom small data en hoe small data vorm krijgen als tegenmacht tegenover big data. Hierbij heb ik gebruik gemaakt van de handvatten die Foucault hiervoor biedt; het analyseren van de procedures die het discours reguleren. Uit de analyse van de externe uitsluitingsmechanismes is gebleken dat de drie verschillende soorten teksten over small data in feite tot drie gescheiden discursieve formaties behoren. Binnen deze drie discursieve formaties worden small data telkens op een andere manier als tegenmacht tegenover big data gesitueerd. De zoektocht naar de waarheid is in alle drie de discursieve formaties een andere, omdat zij zich richten op andere problemen rondom big data en small data op een andere manier als oplossing voorstellen. De analyse van de interne uitsluitingsmechanismes liet zien wat de oorzaak hiervan was, dat de teksten vanuit verschillende disciplines zijn geschreven, allen met hun eigen begrippenkader en jargon. Dat maakt het zoeken naar een gemeenschappelijke waarheid lastig. Bovendien is gebleken dat in de teksten uit de drie verschillende discursieve formaties steeds andere subjectposities mogelijk worden gemaakt. Er wordt tussen de drie groepen zelden op elkaar gereageerd, waardoor er geen kruisbestuiving of dialoog plaatsvindt in de verschillende teksten waarin het begrip small data betekenis krijgt. De analyse van de regels voor deelnemers aan het vertoog versterkte vervolgens het idee van drie gescheiden vertooggemeenschappen, waarvan de

grenzen redelijk hard en absoluut zijn. Omdat deze daarnaast tevens vanuit een specifieke doctrine of ideologie redeneren, lijken de verschillende deelnemers ook weinig belang te hebben bij enige vorm van kruisbestuiving.

De enige overeenkomst tussen de drie discursieve formaties is eigenlijk dat ze de term 'small data' gebruiken. De betekenis van de term ontstaat slechts in de vorm van een belofte over de voordelen die small data ten opzichte van big data zouden hebben. Men probeert met argumenten aan te tonen dat small data een redelijk alternatief zijn voor big data. Uit het feit dat ze vervolgens vooral met de voorstanders van big data in debat gaan, en niet met elkaar, kunnen we afleiden dat het belang er niet uit bestaat om een specifieke betekenis van small data te claimen, maar om vanuit de eigen discipline een alternatief voor big data te verspreiden.

Er is dus in feite geen sprake van een contradiscours, maar van drie gescheiden contradiscoursen. Er is sprake van drie sets van argument waarom small data voor een specifieke groep subjecten een zeer specifiek voordeel kent, ten opzichte van een zeer specifiek probleem rondom big data. Hierdoor krijgen diverse problematische aspecten van big data verzameling, opslag en analyse niet de aandacht die ze wellicht zouden verdienen. En blijft het moeilijk om dé betekenis van het begrip 'small data' te duiden, anders dan dat het om minder data gaat dan bij big data het geval is. Op basis van deze conclusies lijkt het lastig vol te houden dat small data een volwaardige tegenmacht vormt tegenover big data.

Maar er is nog een probleem: het gebrek aan feitelijke voorbeelden van hoe small data zouden moeten functioneren, laat staan van gerealiseerde projecten. De teksten in alle drie de discursieve formaties blijven hangen in beloftes, behalve bepaalde projecten die door het Project VRM van Doc Searls gerealiseerd worden. Maar die hebben vervolgens alleen betrekking op de relatie tussen consument en aanbieder en zijn daarmee te beperkt om een geheel contradiscours te kunnen dragen. Sowieso wordt er in alle drie de formaties met een ideologische klinkend begrippenkader gewerkt. In alle teksten wordt beschreven op welke wijze small data een verschil zouden kunnen of zouden moeten betekenen in de machtsrelatie tussen de subjectposities die in de tekst mogelijk worden gemaakt: burger versus overheid, consument versus aanbieder en klein bedrijf versus groot bedrijf. In de teksten waar vanuit de filosofie van open data wordt geredeneerd, beschrijft men dat er projecten gerealiseerd zouden kunnen worden op basis van samenwerking en het gebruik van small data. Maar wat dit oplevert, en wat dit onderscheidt van big data, blijft onuitgewerkt. Daar waar nagedacht wordt over de positie van kleine bedrijven, blijft het tevens bij een beschrijving van de situatie zoals die zou moeten zijn. De teksten staan bol van kritiek op big data en lof over de mogelijkheden die small data in potentie te weeg zouden kunnen brengen. En de enige voorbeelden die genoemd worden, betreffen juist grote bedrijven als Google en Nike.

Samenvattend kunnen we dus stellen dat het small data discours er maar zeer ten dele in slaagt om een volwaardige tegenmacht te vormen tegenover big data. Ten eerste, omdat er maar een klein deel van de mogelijke kritiek op big data binnen de scope van het discours valt. Ten tweede, omdat het contra-discours sterkt gefragmenteerd is. Drie verschillende discursieve formaties geven drie verschillende betekenissen aan de term ‘small data’, maken verschillende subjectposities mogelijk en richten zich steeds op een ander punt van kritiek op big data. Ten derde omdat het discours meer gaat over de manier waarop het gebruik van small data een bepaalde verbetering op zou moeten leveren voor burgers, consumenten en kleine bedrijven, dan over de manier waarop dit vervolgens gerealiseerd zou kunnen worden en wat voor toepassingen hier feitelijk uit voort zouden kunnen komen. Het probleem is dus dat het small data discours niet zo diep gaat, zowel niet in de kritiek op big data, als niet in de beschrijving hoe small data daar vervolgens een antwoord op zou kunnen geven. Small data lijken met name iets is dat een bepaalde groep mensen over de rug van de populariteit van big data aan het publiek probeert te 'verkopen'.

Meer dan een tegenmacht tegenover big data, zou het small data discours ook als parasiet kunnen worden omschreven: het kan alleen bestaan dankzij big data en heeft big data juist daarom heel hard nodig. Het belang dat met name lijkt te spelen binnen de drie discursieve formaties is dan ook het belang om de discussie zo één-dimensionaal mogelijk te houden; om op basis van een ‘verkoopargument’ small data aan de man te brengen. De logische structuur van het small data discours is simpel: “het oude is verkeerd en wel hierom, en onze oplossing is beter en wel daarom”.

Kortom, small data zijn een belofte. Na diepere analyse blijkt het inderdaad niet veel meer te zijn dan dat. En om kans te maken als contradiscours zal er eerst een bredere, gezamenlijke visie op de problemen rond big data geformuleerd moeten worden, waarop small data een krachtig antwoord kunnen zijn dat gebruikt maakt van één overkoepelend en gedeeld begrippenkader. En vervolgens moeten er eerst dan nog een groot aantal small-datatoepassingen gerealiseerd worden.

Literatuur

Ackoff, Russel. 1989. From Data to Wisdom. *Journal of Applied Systems Analysis.* Volume 16, p. 3-9.

Barabási, Albert-László en Eric Bonabeau. 2003. Scale-Free Networks. *Scientific American*. Nr. 288, p. 60-69.

Beemster, Rob. 14 maart 2013. Het Nieuwe Goud. In *Adformatie*, vakblad over reclame, marketing en media. Jaargang 41, Nr 6.

Benkler, Yochai. 2006. Political Freedom Part 2: Emergence of the Networked Public Sphere. In *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* New Haven, CY: Yale University Press.

Bonde, Allen. 2013. Small Data Goes Big Time. *Digital Clarity Group (DCG).* 23 april. http://www.digitalclaritygroup.com/blog/small-data-goes-big-time/.

Bloem, Jaap, Menno van Doorn, Sander Duivestein, Thomas van Manen, Erik van Ommeren. 2012. *Helderheid Creëren met Big Data*. Sogeti VerkenningsInstituut Nieuwe Technologie VINT. Groningen: LINE UP.

Bloem, Jaap, Menno van Doorn, Sander Duivestein, Thomas van Manen, Erik van Ommeren. 2013. *Uw Big Data-potentieel.* Sogeti VerkenningsInstituut Nieuwe Technologie VINT. Groningen: LINE UP.

Castells, Manuel. 2002. *The Internet Galaxy.* Oxford: University Press.

Clark, Liat. 2013. NSA broke Privacy rules Thousands of Times and self-selects which blunders to report. *Wired*. 16 augustus. http://www.wired.co.uk/news/archive/2013-08/16/nsa-failures-and-leaks.

De Souza e Silva, Adriana. 2006. From Cyber to Hybrid: Mobile Technologies as Interfaces of Hybrid Spaces. In *Space and culture 9* (3) 261-278.

Dijk, Jan van. 2006. *The Network Society*. *Social Aspects of New Media.* Londen: Sage Publications Ltd.

Fontana, Alessandro. 1977. Vérité et pouvoir. Entretien avec Michel Foucault. In *L'Arc: La crise dans la tête*, No. 70.

Foucault, Michel. 1968. Réponse au Cercle d'épistémologie. In *Specificaties voor analyse*. N.p.

Foucault, Michel. 1975. *Discipline and Punish: The Birth of the Prison*. New York: Random House.

Foucault, Michel. 1979 (eerste druk 1969). ‘What is an Author’ in *Textual Strategies: Perspectives in Post-Structuralist Criticism*. New York: Cornell University Press.

Foucault, Michel. 1982 (eerste druk 1971). *De orde van het vertoog*. Vertaling: C.P. Heering-Moorman. Meppel: Boom.

Foucault, Michel. 2006a (eerste druk 1961). *History of Madness*. Vertaling: Murphy Jonathan. Oxfordshire: Routledge.

Foucault, Michel. 2006b (eerste druk 1966). *De woorden en de dingen. Een archeologie van de menswetenschappen*. Vertaling: Walter van der Star. Amsterdam: Boom.

Google Grieptrends, 2011. Grieptrends. *Google.org.* N.d. http://www.google.org/flutrends/about/how.html.

Groot, Reinier. 2012. Big Data = Big Deal? *Bits of Freedom.* 21 juni. https://www.bof.nl/2012/06/21/big-data-big-deal/.

Ho, Victor. 2013. Why Small Data May Be Bigger Than Big Data. *Inc.com*. 1 april. http://www.inc.com/victor-ho/why-small-data-may-be-bigger-than-big-data.html.

Intel. 2013. Moore's Law Inspires Intel Innovation. *Intel*. N.d. http://www.intel.com/content/www/us/en/silicon-innovations/moores-law-technology.html.

Kettmann, Courtney. 2013. Big v. Small Data… Why Marketers Today Need to Know the Difference. *Viralheat Social Media Strategy Blog*. 14 januari. http://blog.viralheat.com/2013/01/14/big-v-small-data-why-marketers-today-need-to-know-the-difference/.

Komorowski, Matthew. 2009. A History of Storage Cost. Mkomo.com. 24 juli. http://www.mkomo.com/cost-per-gigabyte.

Kuneva, Meglena. 2011. *Personal Data: The Emergence of a New Asset Class*. World Economic Forum.

Lambrechts, Mark. 1981. De archeologies-genealogiese methode van Michel Foucault. In *Te Elfder Ure.* Nijmegen: Socialistiese Uitgeverij Nijmegen.

Leinweber, David. 2013. Big Data Gets Bigger: Now Google Trends Can Predict The Market. *Forbes.* 26 april. http://www.forbes.com/sites/davidleinweber/2013/04/26/big-data-gets-bigger-now-google-trends-can-predict-the-market/.

Manovich, Lev. 2011. Trending: The Promises and the Challenges of Big Social Data. In *Debates in the Digital Humanities*, The University of Minnesota Press, Minneapolis, MN.

Mayer-Schönberger, Viktor en Kenneth Cukier. 2013. *De big-data revolutie. Hoe de data-explosie al*

onze vragen gaat beantwoorden. Amsterdam: Maven Publishing.

McAfee, Andrew en Erik Brynjolfsson. 2012. Big Data: The Management Revolution. Boston: Harvard Business Publishing.

Mills, Sara. 2003. Michel Foucault. In *Routlegde Critical Thinkers*. Londen: Routlegde.

Morozov, Evgeny. 2013. The Meme Hustler. *The Baffler Nr. 22.* 2 mei. http://thebaffler.com/past/the_meme_hustler.

Novum. 2013. Consumer Insights. In *Adformatie*. 6 februari. http://www.adformatie.nl/nieuws/bericht/tweederde-van-online-consumenten-tegen-tracking/.

OKF. 2013a. The Open Knowledge Foundation Newsletter. *Open Knowledge Foundation*. 7 mei. http://blog.okfn.org/2013/05/07/the-open-knowledge-foundation-newsletter-may-2013/.

OKF. 2013b. What is Open Data? *Open Knowledge Foundation.* N.d. http://okfn.org/opendata/.

Oosterling, Henk. 2012, 16 november. Michel Foucault (1926-1984). In *Durf te denken*. [TV-uitzending]. Amsterdam: Human.

Pierce, Cecilia. 1997. *The Interactive Book: A guide to the interactive revolution*. Indianapolis: Macmillan.

Pollock, Rufus. 2006. *The Value of the Public Domain.* Londen: Institute for Public Policy Research.

Pollock, Rufus. 2013a. Forget Big Data, Small Data is the Real Revolution. *Open Knowledge Foundation Blog*. 22 april. http://blog.okfn.org/2013/04/22/forget-big-data-small-data-is-the-real-revolution/.

Pollock, Rufus. 2013b. What Do We Mean By Small Data. *Open Knowledge Foundation Blog*. 26 april. http://blog.okfn.org/2013/04/26/what-do-we-mean-by-small-data/.

Proffitt, Brian. 2012. Big-data benefits with small data. *ITworld*. 18 april. http://www.itworld.com/big-datahadoop/269166/big-data-benefits-small-data.

SDG. 2012. A Small View of Big Data. *Small Data Group*. 22 september. http://smalldatagroup.com/2012/09/22/a-small-idea/.

SDG. 2013a. Beyond Big Data: Why Small Data is Taking Off. *Small Data Group.* 10 mei. http://smalldatagroup.com/2013/05/10/beyond-big-data-why-small-data-is-taking-off/.

SDG. 2013b. 5 Small Data Articles You Need to Read. *Small Data Group.* 6 juli. http://smalldatagroup.com/2013/07/06/5-small-data-articles-you-need-to-read/.

SDG. 2013c. Content, Scale and Simplicity. *Small Data Group.* 10 augustus. http://smalldatagroup.com/2013/08/10/content-scale-and-simplicity-highlights-from-my-ims13-keynote/.

SDG. 2013d. Small Data Moves into the Spotlight - 10 Vendors Worth Watching. *Small Data Group.* 14 februari. http://smalldatagroup.com/2013/02/14/small-data-moves-into-the-spotlight-10-vendors-worth-watching/.

Searls, Doc. 2012a. *The Intention Economy.* Boston: Harvard Business Review Press.

Searls, Doc. 2012b. Doc Searls Weblog. *Weblogs at Harvard Law School.* 20 februari. http://blogs.law.harvard.edu/doc/.

Searls, Doc. 2012c. The Customer as a God: The Future of Shopping. *Wall Street Journal*. 20 juli. http://online.wsj.com/article/SB10000872396390444873204577535352521092154.html.

Silver, Nate. 2012. *Signal and the Noise*. New York: The Penguin Press.

Stillman, Jessica. 2013. Baffled By Big Data? Use 'Small' Data Instead. *Inc.com*. 4 januari. http://www.inc.com/jessica-stillman/baffled-by-big-data-just-use-data-instead.html.

Tate, Ryan. 2013. Facebook Will Peer Into Your Grocery Bag to Sell an Ad. *Wired*. 11 april. http://www.wired.com/business/2013/04/facebook-groceries/.

Telecommunicatiewet, 2013. Wijziging van de Telecommunicatiewet (wijziging artikel 11.7a), Concept wetsvoorstel ten behoeve van internetconsultatie. *Economische Zaken*. 20 mei.

Vries, Imar de. 2012. *Tantalisingly Close. An Archaeology of Communication Desires in Discourses of Mobile Wireless Media.* Amsterdam: Amsterdam University Press.

VRM, 2013a. Project VRM. *Berkman Center*. 27 maart. http://cyber.law.harvard.edu/projectvrm/Main_Page.

VRM, 2013b. VRM Development Work. *Berkman Center*. 29 juli. http://cyber.law.harvard.edu/projectvrm/VRM_Development_Work.

White, Mimi. 1995. Ideological Analysis and Television. *Channels of Discourse*, *Reassembled*. New York: Routledge.

Zwaneveld, Arend. 2012. Big data: wat is dat eigenlijk? *Webanalisten.nl*. 9 mei. http://www.webanalisten.nl/big-data-wat-is-dat-eigenlijk#_edn2.