

Datagedreven werken – zegen of vloek?

Wat is eigenlijk datagedreven werken, wat kun je er mee? Datagedreven werken is, zo lijkt het op het eerste gezicht, op basis van (onbewerkte) gegevens gaan werken en (bij)sturen. En dit gaat ver. Het is namelijk vaak een excuus om héél veel data te gaan verzamelen of samenvoegen.



Afb. CC0 Public Domain - Free for personal and commercial use

Als ICT'er heb ik altijd geleerd: gegevens (data) zijn **géén** informatie. Verzamelde data wordt omgezet in informatie. En met die informatie wordt iets gedaan.

Als het gaat om bedrijfsmatige data (verkoopcijfers, productiecijfers, afzetgebieden, resultaten advertencampagnes, sturing koopgedrag, enzovoorts) is er geen enkel bezwaar zolang er geen persoonsgegevens worden verwerkt en/of persoonsgegevens met expliciete toestemming zijn verzameld voor dit doel.

Echter, datagedreven werken wordt ook op ander gebied ingezet. Er wordt allerlei data over mensen verzameld, data van burgers. En daar worden analyses op gedaan en op basis van die analyses beslissingen genomen.

Data waar je, vanuit **privacy-oogpunt**, recht op inzage en correctie op moet hebben. Alleen.. **niemand** vertelt je dát ze die data over je verzamelen, daarmee aan het werk zijn en soms verregaande conclusies trekken op basis van die data! Laat staan dat je toestemming wordt gevraagd hiervoor.

Daarnaast worden er in de uitgangspunten, mijns inziens, mogelijk onjuiste beslissingen genomen of aannames gedaan. Waardoor de uitkomsten van de analyses ‘*null and void*’ zijn, in het beste geval, of, in het ergste geval, tot verkeerde conclusies en beslissingen leiden die verstrekkend kunnen zijn.

Gedragsverandering

Het ministerie van Justitie en Veiligheid (JenV) heeft in 2019 [een uitgebreid rapport](#) gepubliceerd waarin zij uit de doeken doet hoe zij middels datagedreven werken onder andere analyses (willen) doen op personen en gezinnen (families) en vervolgens willen gaan sturen op gedragsverandering.

Na het doornemen van het document van het Ministerie van JenV zie ik toch wel wat haken en ogen voor wat betreft het inzetten van datagedreven werken op die manier. Al was het maar wegens de enorme *verzamelwoede* van data (en bewaren er van) die daardoor ontstaat.

Een recent voorbeeld op dat gebied is de veiligheidsdienst. Bits of Freedom heeft hier [een behoorlijk dossier](#) over opgebouwd. De geheime diensten bewaren (nog steeds) onrechtmatig de gegevens van burgers, [zo bleek recent](#).

*De toezichthouder heeft al aangegeven dat dit **onrechtmatig** is, maar **de minister schuift dit oordeel terzijde** en laat deze praktijk in stand. ([BOF](#))*

Dat is allemaal “big data”. Wat doen ze daar mee?

Wat is datagedreven werken

*Datagedreven werken is het **sturen** op basis van onbewerkte feiten en ruwe **data**. Met behulp van **data** is het mogelijk om de processen zo efficiënt mogelijk in te richten, waardoor beslissingen sneller, beter én accurater genomen kunnen worden.*

Datagedreven werken is wezenlijk anders, zo is de claim, dan informatiegestuurd werken:

Informatie is data die verwerkt is zodat er een betekenis aan gegeven kan worden.

Informatiegestuurd werken kan gezien worden als werken op basis van verschillende KPI's (key performance indicators) welke gemonitord worden en waar op gestuurd kan worden.

Informatiegestuurd werken is een onderdeel van datagedreven werken. Echter is datagedreven werken zo ingericht dat data als het startpunt dient om veranderingen te brengen. ([datakingdom](#))

Datagedreven werken: waarom eigenlijk?

Een datagedreven werkwijze zorgt ervoor dat je **sneller** weet wanneer iets goed of juist niet goed gaat. Hierdoor is het mogelijk om sneller aan de knoppen te draaien om **veranderingen** teweeg te brengen. En niet alleen dat maar je weet ook nog eens aan welke knoppen je het beste kunt draaien voor het gewenste resultaat. ([E-Mergo](#))

In dat geval zul je dus de data bewerkt moeten hebben of bepaalde **indicatoren** moeten hebben en is er geen sprake van onbewerkte, ruwe, data maar sprake van **interpretatie** van de data middels bijvoorbeeld een ‘dashboard’, gevoed door kunstmatige intelligentie (Artificial Intelligence, vaak afgekort tot “AI”).

Wat is kunstmatige intelligentie

AI is een toegepaste wetenschap waarbij computers bepaalde menselijke taken, zoals leren, redeneren, analyseren, problemen oplossen en beslissen, overnemen.

De EU beschrijft het als volgt op haar [website](#):

AI is de mogelijkheid van een machine om mensachtige vaardigheden te vertonen - zoals redeneren, leren, plannen en creativiteit. AI maakt het voor technische systemen mogelijk om hun omgeving waar te nemen, om te gaan met deze waarnemingen en problemen op te lossen om een specifiek doel te bereiken. De computer ontvangt data - reeds voorbereid en verzameld via eigen sensoren, zoals een camera - verwerkt deze en reageert erop. AI-systemen zijn in staat om hun gedrag in zekere mate aan te passen, door het effect van vorige acties te analyseren en autonoom te werken.

Uit het rapport van JenV blijkt dat deze waarneming onder andere wordt gedaan door GPS-tracking, cameratoezicht, audio-opnames van gesprekken e.a. (geautomatiseerde) middelen.

Alle data die daarmee verzameld wordt gaat in een grote bak (databases) en middels datagedreven werken wordt op basis van ‘indicatoren’ geprobeerd gedrag van mensen te *sturen* of bepaald (ongewenst) gedrag te voorkomen.

Interpretatie

De dashboards/indicatoren veronderstellen kennis van de data en **hoe die te interpreteren** en te wegen, af te zetten, onderling. Wat weegt zwaarder? Welke handelwijze beïnvloedt de data? Wat voor **vooringenomen** standpunten wegen er in mee?

De weging en verwerking van data (waardoor het *feitelijk* weer omgezet wordt van gegevens naar een beperkte gegevensset i.c. informatie en er geen sprake meer is van werken op basis van ruwe data) zal dus zeer secuur, vrij van aannames en ethisch verantwoord moeten zijn.

Etniciteit

Bepaalde specifieke indicatoren, zo stelt JenV, mogen *niet* meewegen. Waardoor je eerdergenoemde interpretatie mijns inziens juist *niet* waarde vrij maakt. In mijn optiek is vanaf dat punt de data al gemankeerd want: gefilterd.

De *etnische* gegevens bijvoorbeeld. Waarmee een filtering wordt gedaan op de ruwe data en er geen sprake meer is van datagedreven werken omdat relevante gegevens verwijderd zijn.

De aanname is “iedereen is gelijk”. Maar dat is een *vooronderstelling* die kant noch wal raakt. Juist etnische gegevens kunnen een bepaalde invloed hebben op gedrag.

We weten bijvoorbeeld immers allemaal dat in bepaalde bevolkingsgroepen er méér criminaliteit is dan in andere? We weten dat in bepaalde groepen de terreurdreiging hoger is dan in andere?

Een ander bekend fenomeen is dat in bepaalde (arme) wijken de gezondheid van mensen minder goed is en er sprake is van een tot wel zes jaar kortere levensduur..?

Toch klinkt het raar in de oren van mensen als je dat soort feiten (etniciteit, woonomgeving, sociale situatie) juist wél laat meewegen want: “*racisme*”, “*bevooroordeeld*”, “*mensen wegzetten*” en “*kijk maar naar de toeslagenaffaire*”.

Anderzijds is het bewust buiten beschouwing laten van etniciteit, sociale omstandigheden, opleidingsniveau e.d wel degelijk een misser als JenV *gedragsverandering* wil stimuleren met behulp van datagedreven werken. Immers, bepaalde (invloeden op) groepen blijven in je uiteindelijke ‘dashboard’ op die manier buiten het zicht van de mensen of systemen die je wilt beïnvloeden qua gedrag.

In mijn ogen is door deze “woke” politiek correcte houding en filtering van data op grond daarvan, het buiten beschouwing laten van bepaalde data, bij JenV het project op voorhand al gedoemd te mislukken. Want: door bepaalde indicatoren te verwijderen krijg je als resultaat dat andere indicatoren eventueel zwaarder gaan wegen en er een volstrekt vertekend beeld (en daarop te baseren beleid) ontstaat.

Laat ik het maar plat zeggen: er is een wezenlijk, groot, verschil tussen iemand met een vluchtelingenstatus die in een achterstandswijk woont en een geprivilligeerd gezin in een vindex-wijk. Die kun je niet in één bak gooien en vervolgens deze groepen, gefilterd en ontdaan van specieke kenmerken die van invloed zijn, met dezelfde meetwaarden gaan beoordelen.

De gemeente Zaandam (zie verder) kwam bijvoorbeeld *juist* tot de ontdekking dat een groot deel van de mensen met een *vluchtelingenstatus* (75%) afhankelijk bleven van de bijstand. Waardoor zij, door wél “*etnisch te profileren*” op basis van deze gegevens konden gaan inzetten op het toeleiden naar werk voor deze specifieke groep.

Waar en waarom wordt datagedreven werken ingezet?

Datagedreven werken wordt op heel veel terreinen ingezet. In de commerciële sector, het lichtende voorbeeld voor veel mensen, bijvoorbeeld. Zoals BOL.COM, CoolBlue e.v.a. bedrijven.

Maar het wordt *ook* ingezet door de EU, door ministeries, zoals JenV, provincies en gemeenten. Met name voor het Sociaal Domein wil men dit (gaan) gebruiken binnen veel gemeenten. Op de [website van VNG Realisatie](#) is er ook het nodige over te vinden.

In het rapport “[Ontwikkelmodel Datagedreven Gemeenten](#)” lezen we het volgende er over:

Door datagedreven te werken, kunnen gemeenten efficiënter hun doel bereiken. In het sociale domein helpt goed gebruik van informatie bijvoorbeeld om mensen met een mogelijke hulpvraag eerder in beeld te krijgen. En in het fysieke domein om doelmatiger en doeltreffender om te gaan

met onderwerpen als onderhoud, parkeren, of het ophalen van afval. Het Ontwikkelmodel Datagedreven Gemeente helpt je op weg. In deze paper lees je hoe. (p5)

Wanneer een gemeente zich er op richt mensen met een hulpvraag in beeld te krijgen, kan daar moeilijk een bezwaar tegen gemaakt worden. Het doel is nobel (voorkomen dat mensen in de problemen komen) en is ten gunste van de burger.

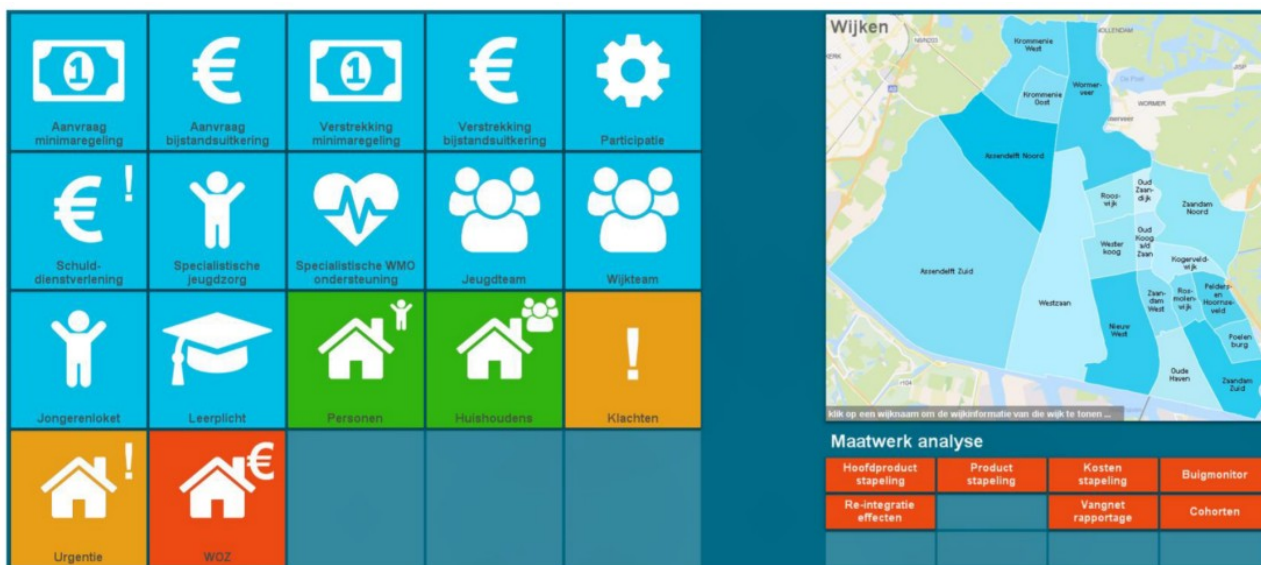
Gemeente Zaandam

Bij de gemeente Zaandam is “big data” ingezet met als doel “om op basis van data maatschappelijke vraagstukken aan te pakken” (“Datagestuurd werken een veranderkundige aanpak” – Tom Pots).

Het initiatief om ‘iets met big data’ te doen kwam vanuit een leidinggevende van de IT-afdeling en Pots kreeg opdracht het verder uit te werken.

De aanleiding in Zaandam was niet de vraag vanuit de werkvloer c.q. het veld, maar omdat big data en datagedreven werken bij steeds meer organisaties op de agenda kwam en komt vond men de tijd rijp er iets mee te gaan doen, om te onderzoeken of het toegevoegde waarde kon hebben.

Interessante conclusies zijn er gekomen uit het dataonderzoek. Zoals bijvoorbeeld het feit dat de meeste overlastmeldingen door slechts 10 personen werden gedaan, dat mensen met verward gedrag maar mondjesmaat de WMO gebruiken (terwijl juist zij daar potentieel voordeel van kunnen hebben waarschijnlijk) en voorspellingen, met 95% zekerheid, over funderingsproblemen.



Dashbord ‘burger’, “Datagestuurd werken een veranderkundige aanpak” – Tom Pots. (afb. overgenomen op basis van citaatrecht)

Zomaar een greep uit de rapportage van Zaandam. Allemaal zonder uitzondering nuttig gebruik van data die je al tot je beschikking hebt en door datagedreven (te) werken leiden tot het in kaart brengen van kansen en problemen.

Hiervoor is een datawarehouse (“datapakhuis”) ingericht welke met QlikView werd ontsloten. De datasets zijn geordend langs twee thema’s: burger en leefomgeving.

De vraag is: mág dat? Kun je gegevens van burgers uit diverse ‘databakken’ halen, samenvoegen in een datawarehouse en vervolgens middels datagedreven werken mee aan de slag gaan? Oók als dat alleen maar nobele, goede, doelen dient?

Profilering

“de inzet van algoritmes haalt regelmatig de media – en meestal niet in positieve zin. Rond de opsporing van fraude zijn verschillende (overheids)organisaties breed uitgemeten onderuitgegaan. Ze koppelden data en voorspellende rekenmodellen aan ‘risicoprofielen’ en trokken onterechte conclusies over mogelijke misstanden.” – [Frankwatching](#).

Er kunnen (dus) ook andere belangen (gaan) spelen bij de overheden. Bijvoorbeeld fraudebestrijding. Of het in beeld brengen van “probleemgezinnen”.

Voorbeeld.

Stel je hebt een dataverzameling (inwoners van een regio) waar je een dashboard op inricht.

Op basis van bepaalde onderzoeken weet je dat “in 9 van de 10 gevallen waarbij er sprake is van (1) laag inkomen, (2) twee of meer kinderen waarvan één of meer met een beperking en (3) woonachtig in bepaalde regio of wijk en (4) er sprake is van bovengemiddeld verzuim op school” er *potentieel* sprake kan zijn van een “probleemgezin” met een meer dan gemiddelde kans op verwaarlozing van de kinderen.

Je stelt vervolgens in het dashboard in dat er bij gezinnen waarbij de 4 indicatoren aanwezig zijn er een ‘attentiesignaal’ komt. Je laat het dashboard zijn werk doen en er komt een lijst uit van gezinnen. Vervolgens ga je ‘pro-actief’ er een ‘gezinscoach’, maatschappelijk werk of jeugdwerk op afsturen voor preventie. Of ‘onderzoek instellen’.

Het gezin is *gemarkeerd* in het systeem, er komt een vlaggetje ‘potentieel probleemgezin’ te staan bij een gezin dat zich totaál niet bewust is van het feit dát ze als zodanig gemarkeerd zijn of dat er überhaupt een (geautomatiseerd) onderzoek naar ze wordt gedaan. Er ontstaat dossiervorming waar verschillende partijen (ongevraagd) inzage in krijgen.

Daar kun je *verschrikkelijk* de mist mee in gaan want voldoet iemand aan deze indicatoren dan hoeft er helemaal geen sprake te zijn van een probleemgezin. Het feit namelijk dat deze indicatoren bij probleemgezinnen voorkomen wil niet zeggen dat het indicatoren ván probleemgezinnen zijn. Immers: **correlatie is géén causaliteit!**

Bedenk bijvoorbeeld eens dat bij kinderen met ADHD als diagnose wordt gesteld dat ze “een beperking” hebben én deze kinderen een hoger verzuim hebben. Volslagen ten onrechte kunnen er dan héél veel gezinnen op zo’n (onderzoeks)lijst komen te staan! Terwijl gezinnen in de vinex-wijk waar alle signalen op rood staan en alarmbellen zouden moeten afgaan worden gemist omdat deze in de ‘verkeerde’ postcode-range wonen... of andere indicatoren niet zijn ingesteld die net zo zwaar of zwaarder hadden moeten wegen.

Publieke organisaties, die het leven van mensen proberen te vergemakkelijken en te verbeteren, baseren hun beslissingen steeds vaker op data over de leefwereld, situatie en behoeften van mensen.

Een **klantprofiel**, opgebouwd uit de data die over iemand beschikbaar is, is het hart van zo'n klantgerichte organisatie. Dat is toe te juichen, maar **vereist wel dat de waarde én (privacy)kaders en beperkingen glashelder zijn**. Want keuzes rondom de inzet van persoonsgegevens zijn vaak beslissingen met *vérstreckende* gevolgen. En **deze gevolgen raken die mensen direct in hun leefwereld**. – [Frankwatching](#).

Voorbeeld.

Stel, uw moeder woont in een flatje in een 'achterstandswijk' met veel criminaliteit. U bezoekt haar minstens twee à drie keer per week, want ze heeft enige *mantelzorg* nodig.

Helaas woont u zelf ook in een soortgelijke wijk waar recent nog een paar thuistelers zijn opgepakt. *Boven* uw moeder woont sinds kort één van de opgepakte wietkwekers.

Het ministerie van JenV houdt de wijken in de gaten met camera's. U ziet de camera's wel maar maakt zich geen zorgen. *Integendeel*: veilig gevoel die camera's! Vooral ook in verband met uw moeder immers?

U wordt zelf regelmatig herkenbaar door zo'n camera vastgelegd, uw kenteken komt op de beelden regelmatig voor. Uw telefoon registreert uw aanwezigheid op de zendmasten (driepuntsmeting). Ook deze data wordt door justitie en de veiligheidsdiensten regelmatig verzameld en in hun datawarehouse gestopt.

Op basis van een 'big data' analyse weet de in het systeem ingebouwde AI het 99% zeker: u onderhoudt contact met de veroordeelde crimineel. U bezoekt hem immers meerdere malen per week en verblijft daar gemiddeld ca. één à anderhalf uur per keer. U bent volgens het systeem "waarschijnlijk een handelaar in softdrugs of afnemer van deze veroordeelde wietkweker".

Binnen de kortste keren ziet iemand die het dashboard gebruikt uw gegevens. Constateert vervolgens ook nog dat u recent een bijstandsuitkering hebt aangevraagd. Weliswaar omdat u pas gescheiden bent, maar *dát* ziet het systeem niet.

Er komt een vlaggetje bij uw naam: *schaduw* en bij de bank en andere (semi)overheden een onderzoek starten naar uw handel en wandel. U bent *geprofileerd*, u bent een *potentiele crimineel*.

Bij de Sociale Dienst is inmiddels ook een signaal afgegaan. Die starten ook een onderzoek... hebt u eigenlijk wel recht op uitkering? U schijnt *neveninkomsten* te hebben die niet opgegeven zijn. Wat doet u zo vaak bij die crimineel in die achterstandswijk, hebt u wellicht een *relatie* die u niet gemeld hebt..

Mijn fantasie slaat wat door wellicht. Maar het is zeker niet ondenkbaar dat dit onze toekomst wordt. Profileren op basis van datagedreven werken in welke vorm dan ook is daarom, als je het mij vraagt, een **No Go**.

Wat zijn de **bezwaren**, en de nadelen, die hier aan kleven?

- Profileren op basis van een dataset is profileren op basis van *éénzijdige, onvolledige*, gegevens; immers: **niemand** heeft een **compleet** dataprofiel van een burger. Beslissingen die op basis van een incompleet profiel worden genomen kunnen **verstrekkende gevolgen** hebben voor de betrokkene(n)!
- De **indicatoren** op basis waarvan beslissingen worden genomen: hoe komen die tot stand, op basis van gedegen wetenschappelijk onderzoek en brede, geverifieerde, data-analyse of op basis van ‘ervaring’ en ‘onderbuikgevoel’?
- Alle data die je hebt van een burger in één grote container stoppen, een datawarehouse, **is dat eigenlijk wel legaal?**
 - Dat een burger toestemming heeft gegeven het te gebruiken voor een bepaald doel (bijvoorbeeld vaststellen van recht op uitkering) wil niet zeggen dat je toestemming van dezelfde burger hebt het zomaar te gebruiken om een analyse te doen ten behoeve van jeugdzorg of andere doeleinden – *integendeel!*
 - *De algemene regel is dat verstrekken van persoonsgegevens **alleen** mag als dat **verenigbaar is met het doel** waarvoor de gegevens zijn verzameld. Of dit het geval is, hangt af van de concrete omstandigheden. ([Autoriteit Persoonsgegevens](#)).* Er mag vanaf worden geweken wanneer de verwerking van de persoonsgegevens noodzakelijk zijn om de *publieke taak* goed te kunnen vervullen. Het samenvoegen van data in een datawarehouse om daar middels ‘datagedreven werken’ allerlei leuke ontdekkingen mee te doen valt daar domweg niet onder.
- Heeft **de burger er weet van** dat je dat doet, heeft de burger bewust, impliciet, **toestemming** gegeven én inzage in deze data, hoe zit het met het recht op correctie?
- Wie mag het allemaal raadplegen en waarom? Wat zijn de waarborgen dat de data op de juiste manier wordt gebruikt en door de juiste personen (en waarom zijn zij dat?) wordt geraadpleegd?
- Is er sprake van correcte en herleidbare logging van deze data raadplegingen? Denk aan het debacle met de GGD en inzage in gegevens door niet geautoriseerde personen en de [illegale handel in deze gegevens](#).

*Aan data en daaruit volgende categorieën liggen **altijd** keuzes ten grondslag. Ze reproduceren dus een **perspectief op de werkelijkheid**. Dat effect versterkt bovendien naarmate we die vorm van de werkelijkheid meetbaar maken.*

*Ten tweede: mensen stoppen andere mensen graag in een hokje. Dat helpt ons om de wereld en de mensen die we tegenkomen te begrijpen. En dat conflicteert met de kern van een dataprofiel [...] **dataprofielen zijn niet consistent**: verschillende ordeningsprincipes kunnen door elkaar lopen en de categorieën sluiten elkaar ook niet onderling uit. Je scoort op alles, maar niet in gelijke mate. – [Frankwatching](#).*

Datagedreven werken – Zegen of Vloek?

Er valt natuurlijk nog veel meer over te zeggen, er zijn heel veel voorbeelden te vinden hoe datagedreven werken werkelijk ‘een zegen’ kan zijn maar ook, zoals ik een paar keer aanhaalde, een regelrechte vloek werd of kan zijn.

In heel veel zaken kan datagedreven werken heel nuttig zijn. Maar wanneer we het gaan loslaten op de burger, in handen van amateur-analysten achter een ‘dashboard’ die het verschil tussen gegevens en informatie niet kennen, en privacyregels al helemaal niet, is er mijns inziens een enorm risico dat de gegevens tot volstrekt verkeerde conclusies leiden. Met alle gevolgen van dien.

Belangrijk is, denk ik, dat we niet zomaar de hype moeten volgen. Een systeem of methode gebruiken ‘omdat het er nu eenmaal is’ is nooit de juiste insteek.

Als er vanuit de maatschappij geen vraag naar is, is het domweg niet zinvol hier tijd en energie in te steken. Tenzij het voor de maatschappij een, ook door de burger zo ervaren, toegevoegde waarde heeft én de burger hier – op zijn minst impliciet maar bij voorkeur expliciet – mee instemt.

Dit artikel is op persoonlijke titel geschreven en gebaseerd op openbare bronnen.