



(Big)

Data

Gedreven

Gemeente

Rol en positie van gemeenten in een samenleving van Big en Open Data



KWALITEITS
INSTITUUT
NEDERLANDSE
GEMEENTEN

In
samenwerking
met



BIGDATA.COMPANY

Maart 2015, 1ste editie

PDF, Den Haag

Auteurs

Eric de Kruijk, *Kwaliteitsinstituut Nederlandse Gemeenten* - Eric.deKruijk@Kinggemeenten.nl

Floris Schoenmakers, *BigData.Company* – Floris@BigData.Company

Mark Dijksman, *BigData.Company* – Mark@BigData.Company

Mark Gremmen, *Kwaliteitsinstituut Nederlandse Gemeenten* – Mark.Gremmen@Kinggemeenten.nl

Martijn Minderhoud, *BigFellows* – Martijn@BigFellows.io

Kwaliteitsinstituut Nederlandse Gemeenten

Voor gemeenten is het, in onze snel veranderende maatschappij, essentieel te beschikken over een toekomstgerichte visie op informatiemanagement. KING is de partner van gemeenten op het gebied van informatie. Een partner met kennis, die bijdraagt aan visievorming en gemeenten helpt .informatiemanagement op orde te krijgen.

BigData.Company

BigData.Company levert met haar big data platform kant en klare toepassingen voor gemeentelijke vraagstukken zoals zelfredzaamheid en huiselijk geweld. Het platform biedt realtime inzicht in problematieken, effecten van interventies en voorspellingen van ontwikkeling van buurten. De toepassingen van BigData.Company stellen beleidsmakers en professionals in staat hun werk meer efficiënt en effectief te doen

BigFellows

Mark Dijksman en Jaring Hiemstra zijn in 2014 BigFellows begonnen met als doel de publieke sector te ondersteunen met big data vraagstukken. BigFellows biedt adviseurs met domeinexpertise en data analyse vaardigheden. BigFellows helpt de publieke sector om meer data gedreven met complexe vraagstukken om te gaan.

Inhoudsopgave

I Inleiding

II Ontwikkelingen van open en big data

1. Nabije Toekomst

Trends en kansen.

III Data gedreven gemeente

2. Bestuurlijke Uitdagingen

Dilemma's van open & big data.

3. Data Science

Hoe kan ik met data science waarde creëren in een gemeente?

4. Lean Startup

Open en big data initiatieven in de gemeente als een start-up.

5. Impact op Organisatie

Welke impact hebben technologische ontwikkelingen en een data gedreven werkwijze op gemeenten?

6. (Big) Data Gedreven Ambtenaar

Hoe ziet de nieuwe ambtenaar er uit?

7. Security & Privacy

Vertrouwen, veilig en verbonden.

IV Frequently Asked Questions

Appendix I: Open Data Onderzoek Nederlandse Gemeenten

Inleiding

Eén van de grondleggers van het internet, Tim Berners-Lee, kwam er snel achter dat **open data** de drijvende kracht is van het internet. In essentie heeft open data twee kenmerken: het is openbaar beschikbaar voor publiek gebruik en het gebruik kent geen beperkingen. Open data mag gratis gebruikt, bewerkt en gedeeld worden.

Om de mate van openheid te beschrijven, heeft hij de “The Five Stars of Open Data”¹ geïntroduceerd (fig. 1). Waar nul sterren betekent dat geen data gepubliceerd wordt, betekent vijf sterren dat datasets toegankelijk en gekoppeld zijn met andere datasets.

Berners-Lee gelooft, net als steeds meer mensen, bedrijven en organisaties, dat open data een belangrijke basis is voor economische én sociale vooruitgang in de wereld.

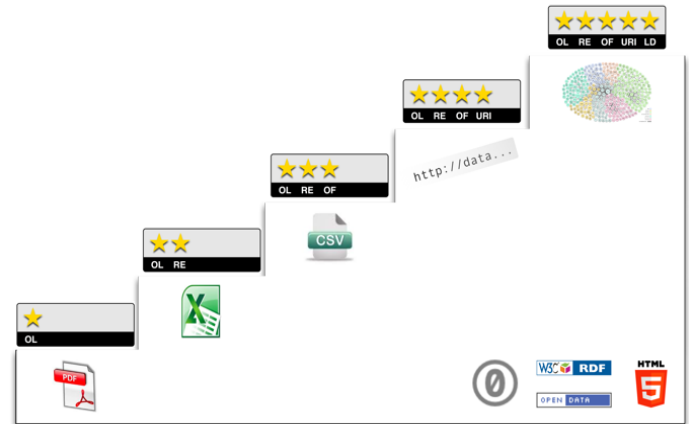


Fig.1: Five stars of open data

- * Bestanden worden beschikbaar gemaakt in een willekeurig format onder een open licentie
- ** Bestanden worden beschikbaar gemaakt als gestructureerde data. Bijvoorbeeld Excel i.p.v. PDF
- *** Bestanden worden beschikbaar gemaakt in software-onafhankelijk format. Bijvoorbeeld CSV i.p.v. Excel
- **** Data is beschreven met metadata (microdata omwille van structuur en uitwisselbaarheid): defined, stadardized and documented
- ***** Relaties beschreven op gestructureerde datasets. Koppeling leiden tot zgn. views (latente dimensie)

¹ <http://5stardata.info/>

Datatechnologie is dusdanig ver ontwikkeld dat het bruikbaar is geworden voor iedere organisatie. Grote hoeveelheden (open) data kunnen op een nagenoeg real-time manier verwerkt, relevant en beschikbaar gemaakt worden. Dit noemen we **big data**. Big data laat zich kenmerken door een verzameling van datasets die zo groot, gevarieerd, en snel veranderend is waardoor gewone database software niet in staat is - op een snelle manier - de data te verwerken en analyseren. Daarbij is de combinatie van verschillende vaardigheden, om zodoende waarde te kunnen creëren met die data, op dit moment schaars.

Het vrij toegankelijk beschikbaar stellen van (publieke) data, alsmede het benutten van de mogelijkheden van big data, staan hoog op de agenda van Nederland en de Europese Unie². De publieke sector wordt hier specifiek genoemd: *“Big Data will make it possible to increase efficiency in the public sector.”*³

² <http://ec.europa.eu/digital-agenda/en/open-data-0>

³ <http://ec.europa.eu/digital-agenda/en/making-big-data-work-europe>

Het opstellen en toegankelijk maken van data is een essentiële volgende stap in de ontwikkeling van de digitale infrastructuur, en daarmee in de ontwikkeling van Europa: in economische ontwikkeling (commerciële toepassing van data) en in het oplossen van maatschappelijke opgaven.

De overheid heeft een essentiële rol in het vrij beschikbaar stellen van data die door het publiek gegenereerd is. Dit white paper gaat in op deze ontsluiting van data en de waarde van publieke data. Daarbij hebben we speciale aandacht voor de veranderingen die gepresenteerd worden aan de gemeenten door de samenleving van open en big data.

II Ontwikkelingen van open en big data

1. Nabije Toekomst

Trends en kansen.

Trends

De hoeveelheid data neemt exponentieel toe. Alleen al de afgelopen twee jaar is de totale hoeveelheid data wereldwijd verdubbeld.⁴ Data wordt gegenereerd door mensen, apparaten, publieke en private organisaties. Een grote, relatief nieuwe, databron is de mens zelf: je kan vrijwel alles van jezelf meten.⁵ Technische ontwikkelingen zorgen ervoor dat deze data steeds beter en sneller gecombineerd en benut kan en zal worden. Nieuwe toepassingen hebben direct impact op de samenleving. De data gedreven middelen zijn onmisbaar geworden. Er wordt actief gebruik gemaakt van deze middelen door de gehele samenleving heen. Data staat daarbij centraal. *Hoe gaan we deze data*

4 <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

5 <http://www.frankwatching.com/archive/2014/03/27/de-doorbraak-van-quantified-self-meten-weten-leven/>

toegankelijk maken en zo goed mogelijk in dienst stellen van de inwoners?

Kansen voor gemeenten

Door toenemend gebruik van digitale middelen, waarbij het gebruik van IT, apps, wearables en gadgets centraal staat, zullen gemeenten de komende jaren gedwongen worden om na te denken over wat deze ontwikkelingen voor hen betekenen. De 'dwang' komt met name voort uit het feit dat inwoners en organisaties sneller inspringen op de mogelijkheden. Informatiebehoefte van inwoners komt voort uit de ontwikkelingen in social media, apps en geïntegreerde toepassingen. Er zullen nu eisen gesteld gaan worden aan de gemeente op dit punt. Gemeenten kunnen niet achterblijven.

Dataverwerking, -ontsluiting en -analyse speelt een steeds belangrijkere rol om maatschappelijke opgaven op te lossen met maximaal effectieve en efficiënte inzet

van middelen. Beschikbare data en technologische mogelijkheden biedt organisaties kansen om veel gericht en pro-actiever te werken en kennis en ervaringen van klanten en inwoners intensief te benutten. Daarnaast stelt het inwoners in staat om te profiteren van en bij te dragen aan **transparantie, signalering, keuzevrijheid** en **legitimiteit** van het bestuur. Dit is een grote verandering voor gemeenten, waar IT en data nu vooral een ondersteunende rol hebben in administratie en bedrijfsvoering.

Anno 2015 worden de mogelijkheden van big en open data gezien binnen de overheid, maar het aantal projecten waar succesvol toegevoegde waarde is gecreëerd met behulp van big data toepassingen is nog gering. Een significant deel van de Nederlandse gemeenten benutten het potentieel van open en big data nog helemaal niet. Het **‘Open Data Nederlandse Gemeenten’** onderzoek - dat is uitgevoerd in opdracht van KING - toont bijvoorbeeld aan dat weinig gemeenten zelf data ontsluiten via een

open data portal (zie ook Hoofdstuk 4 en Appendix I voor meer informatie).

De meeste gemeenten beginnen net met de verkenning van het open stellen van data en zijn nog niet toe aan big data omdat inzicht, handvatten en vaardigheden ontbreken. Terwijl een gemeente een onmisbare rol speelt in het succes van open data.

Tegelijkertijd worden mensen zich steeds bewuster van het feit dat zij data zelf creëren. Met name privacy en eigenaarschap van de data is hier belangrijk. Het eigenaarschap ligt nu veelal nog bij organisaties en instanties. Dat zal in de toekomst mogelijk veranderen. Mensen zullen steeds meer inzien dat zij de data zelf genereren. En willen daarom zelf meer invloed gaan uitoefenen op wie hun data mag gebruiken en wie niet. Voor veel data die in handen is van gemeenten betekent dit impliciet dat publiek gegenereerde data publiek moet zijn. Inwoners verwachten van hun gemeente dat (publieke) data op een toegankelijke manier beschikbaar wordt gemaakt voor verwerking en bewerking.

III Data Gedreven Gemeente

2. Bestuurlijke Uitdagingen

Dilemma's van open & big data.

Regelmatig wordt gesteld dat, wanneer men meer gebruik maakt van informatietechnologie en big data, de ruimte van professionals en burgers beperkt wordt. Wij geloven in het omgekeerde. Goed gebruik van informatietechnologie biedt professionals en inwoners het gereedschap om zelfstandiger, sneller en betere beslissingen te nemen. Sterker nog: het maakt een deel van de staf en interne regels overbodig die professionals nu beperken in hun professionele ruimte. Tegelijkertijd brengen data gestuurde processen ook nieuwe verantwoordelijkheden en noodzakelijke afspraken mee voor gemeenten. Afspraken over, en sturing op, onder andere kwaliteit, technische comptabiliteit, privacy en koppelbaarheid van systemen en databronnen worden steeds belangrijker.

Voordat de verwachte potentie gerealiseerd kan worden, zullen bestuurders en ambtenaren voor verschillende dilemma's komen te staan:

- Privacy vraagstukken (zie ook hoofdstuk 7);
- Juridische vraagstukken met betrekking tot het openbaar maken van data;
- De rol van de lokale overheid;
- Verdienmodel: kosten en baten of maatschappelijke kosten en baten;
- Ethische uitdagingen;
- Data bewerking en verwerking (vaardigheden en middelen);
- Waarborging van consistentie en kwaliteit.

Bovenstaande aspecten zijn een greep uit de dilemma's waar een gemeente te maken heeft of krijgt nog voordat zij de weg van het data gedreven werken in kan slaan. Mogelijk nog interessanter is hoe de organisatie omgaat

met consequenties van deze manier van werken. Wanneer een inwoner of ambtenaar een probleem constateert op basis van geanalyseerde data, leidt dat in veel gevallen tot een verplichting tot handelen: nu heb je het gezien en kan je er niet meer omheen.

Hoe men omgaat met geconstateerde problematiek is essentieel in het succesvol maken van een data gedreven gemeente. Sommige dilemma's zullen van tevoren bediscussieerd moeten worden, de meeste dilemma's zullen ontstaan tijdens de (big) data activiteiten. Om überhaupt tot het dilemma-gedeelte van big data te komen, zal een gemeente verschillende zaken in gang moeten zetten. In de komende paragrafen worden de belangrijkste punten belicht.

3. Data Science

Hoe kan ik met data science waarde creëren in een gemeente?

Data science is de professie van het verzamelen en analyseren van data. In een steeds meer data gedreven wereld wordt data science steeds belangrijker voor innovatie en ontwikkeling van (publieke) organisaties. De rol van data specialisten in gemeenten zal daarbij steeds meer ontwikkelen van ontsluiters naar (ondersteuner in) interpretatie. Het primaire doel voor gemeenten zal vaak zijn om de data betekenis te geven door het in context te plaatsen. Een belangrijke uitdaging voor gemeenten is daarom om deze kwaliteiten te organiseren. Zie ook hoofdstuk 6, '(Big) data gedreven ambtenaar'.

Tegelijkertijd maken analyse en visualisatie toepassingen de drempel lager voor professionals (bestuurders en inhoudelijke specialisten) om informatie te interpreteren en productief te benutten in het dagelijkse werk.

Voordat met data technologie en data science waarde uit data gehaald kan worden, moeten de maatschappelijke opgaven en de mogelijke aanpak helder zijn: wat wil je weten, waarom wil je het weten en hoe ga je het meten, met welke periodiciteit? Wat wil je bereiken, tegen welke kosten? Wat zijn de consequenties? Het gaat om vragen op beleidsniveau: bepaling van prioritering en inzicht in mogelijke preventieve aanpak. Maar ook om handelingsvraagstukken in de uitvoering: concrete dagelijkse keuzes en sturing. Goede vraagarticulatie is daarbij een grote uitdaging, en dit vraagt een andere manier van denken en werken dan nu gangbaar is in veel gemeenten.

Een data gedreven gemeente zou, onder andere, de volgende zaken kunnen nastreven (figuur 2):

- Volgen, (near) real-time monitoren, van ontwikkelingen en eigen prestaties;
- Inzicht in effecten van beleid en uitvoering en daarmee de mogelijkheid om effectief bij te sturen;
- Optimalisatie van werkprocessen door data gedreven ondersteuning in de werkprocessen;
- Nieuwe waarde creëren door ontwikkeling en innovatie van diensten en dienstverlening;
- Organisatie transformatie, door ontwikkeling van compleet nieuwe diensten of door vergaande verandering van de werkwijze.

Deze innovaties zijn voornamelijk op de eigen organisatie en effectiviteit van gemeenten gericht. In een data gedreven gemeenten, waar transparantie en openheid voorop staan, kunnen ook steeds meer positieve externe effecten ontstaan. Hierbij denken we aan een impact op de publieke participatiegraad, wanneer inwoners in staat gesteld worden om gemeente data te gebruiken. Het is feitelijk een vorm van crowd-sourcing, waarbij inwoners zelf nieuwe inzichten genereren. Inzichten die bruikbaar kunnen zijn voor de ontwikkeling van de stad en de besturing van de gemeente.

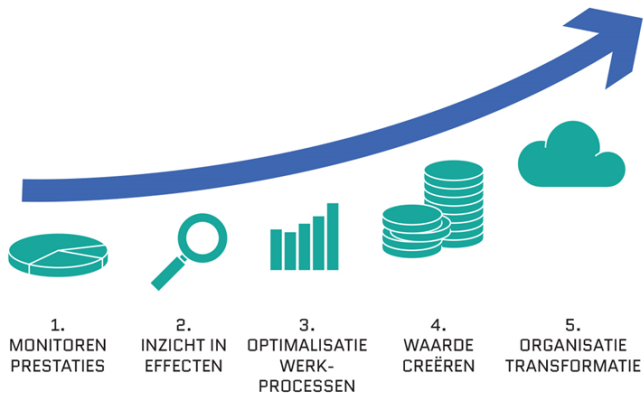


Fig.2: Opeenvolgende aspecten in een data gedreven gemeente

4. Lean Start-up

Open en big data initiatieven in de gemeente als een start-up.

Dat niet alle problemen met techniek, IT en software opgelost kunnen worden, hebben we in de afgelopen 20 jaar voldoende gezien. “Over budget, over time and under performance” zijn helaas bij veel grote ICT projecten, zeker bij de overheid, eerder regel dan uitzondering.

Succesvolle bedrijven realiseren dataprojecten veelal vanuit de **Lean Start-up** principes: ‘**Minimal Viable Product**’ & ‘**Fail Cheap and Fast**’ staan daarbij centraal. Voor tech-startups zijn goedkope mislukkingen en snelle prototypes essentieel. Het leren en daardoor verbeteren staat centraal. Zo wordt tegen minimale kosten snel resultaat neergezet. Ook gemeenten zouden deze weg moeten kunnen volgen. Zet daarom in op projecten, experimenten of pilots om zo de nieuwste en beste technieken te kunnen benutten. Organiseer

doelgerichte projecten, veelal op kleine schaal met beperkte looptijd waar inwoners en andere organisaties op kunnen inhaken. Neem een open data portal als voorbeeld. Een open data portal van een gemeente hoeft niet in een keer goed gevuld te zijn. Door met tien datasets te beginnen van drie verschillende afdelingen en mede daarmee concrete vraagstukken aan te pakken, wordt de kennis en kunde heel snel opgebouwd.

Drie tips om binnen no-time data te kunnen ontsluiten voor het publiek:

1. Maak gebruik van een Open Source platform om data op aan te bieden (zoals CKAN).
2. Maak een werknemer van de gemeente verantwoordelijk en creëer daar een groep van enthousiastelingen om heen.
3. Schrijf online data-contests uit voor het vinden van de meeste innovatieve en/of inventiefste data analyse met behulp van de open gestelde data.

Onderzoek

Open Data Index

Nederlandse Gemeenten.

In Januari 2015 heeft KING in samenwerking met BigData.Company een onderzoek gedaan naar open data portals van Nederlandse gemeenten. De onderzoeksvraag was tweeledig:

Welke Nederlandse gemeenten hebben een eigen open data portal en wat is de kwaliteit van de gepubliceerde datasets op de geïdentificeerde gemeenten portals?

De eerste stap in het onderzoek is het identificeren van portals die door gemeenten gebruikt worden om open data op te publiceren. Ten tijden van het onderzoek hebben 14 gemeenten een eigen open data portal (zie voor meer informatie www.bigdata.company/KING).

Vervolgens hebben we alle gepubliceerde datasets gedownload en via een automatisch proces, met behulp van een SCALA-script (<http://www.scala-lang.org/>), beoordeeld op vier indicatoren. We kunnen concluderen dat een klein gedeelte van de Nederlandse gemeenten een open data portal heeft en dat de kwantiteit en kwaliteit van hetgeen wat op de portals gepubliceerd wordt sterk varieert. Meer informatie over de methodiek en de resultaten is opgenomen in Appendix I.

Het Open Data Nederlandse Gemeenten onderzoek impliceert dat weinig gemeenten in Nederland geslaagd zijn op eigen initiatief een open data portal op te zetten.

5. Impact op Organisatie

Welke impact hebben technologische ontwikkelingen en een data gedreven werkwijze op gemeenten?

Moore's law houdt al sinds 1970 stand: computerkracht verdubbelt iedere twee jaar⁶. Binnen afzienbare tijd zal een computer de intelligentie van mensen kunnen overtreffen. Computer en mensen zijn echter goed in verschillende dingen: mensen kunnen goed beslissingen nemen in complexe situaties. Een computer kan goed met enorme hoeveelheden data om gaan. Succesvolle (big) data oplossingen laten zien dat een computer optimaal benut wordt wanneer het systeem een ondersteunende (en niet autonome) rol speelt.

In feite is een grote verzameling data een grote bak 1-en en 0-en. Mensen moeten in staat gesteld worden, door middel van computers, om daar iets mee te doen.

De computer is niet in staat de hele cyclus van opdracht tot uitvoering tot interpretatie uit te voeren. Het gaat om het creëren van een complementaire oplossing waarbij computer en mens beide een belangrijke rol vervullen.

In de toekomst zullen de meeste waardevolle organisaties gebouwd worden door mensen die het best creatieve ideeën van mensen kunnen laten ondersteunen door computers.

Impact op gemeenten

Data is pas nuttig gebruikt als het mensen helpt om meer inzicht te krijgen en iets te verbeteren. Inzichten kunnen mensen helpen productiever te worden en discussies te voeren op basis van objectieve statistieken.

⁶ http://nl.wikipedia.org/wiki/Wet_van_Moore

Een data gedreven manier van werken zal, onder andere, op het volgende impact hebben:

- *Transparantie en verantwoording.* Data krijgt een centrale rol in sturing. Data kan gedeeld worden en bevordert transparantie.
- *Burgerinitiatieven en betrokkenheid.* Op basis van de opengestelde data kunnen burgerinitiatieven groeien.

De toekomstige gemeentelijke organisatie vergt flexibiliteit, reactiesnelheid en de mogelijkheid om te differentiëren en maatwerk te leveren. Door data beschikbaar te stellen, kunnen co-creatie door partners en burgers een integraal onderdeel worden van de organisatie.

Om dit te realiseren is optimalisatie van bestaande processen onvoldoende. Steeds meer publieke managers en bestuurders realiseren zich dat de uitdagingen vragen om een herbezinning op de bestaande organisatie en werkwijze. Deze ontwikkeling is door Jaring Hiemstra⁷ uitgewerkt in zijn visie op **Smart Government**. Kernelementen daaruit zijn op de volgende pagina weergegeven

⁷ <http://www.bigfellows.io/#homepage-team>

Klassieke Overheidsorganisatie	Smart Government
Getrapte verticale communicatie tussen verschillende lagen in de organisatie; veel horizontale overlegvormen en 'project- en vergadertoerisme'.	Horizontale informatie-uitwisseling tussen professionals en tussen professionals en bestuurders; 'Highly aligned, loosely coupled'.
Partners primair selecteren op prijs, burgers als afnemers mogen meepraten en feedback geven. Ze staan 'buiten' de organisatie.	Partners en burgers voegen essentiële waarde toe (citizen generated content) zijn een onderdeel van het primair proces.
Systeemprikkels als driver om doelen te bereiken; flexibele beloning, gedetailleerde jaarplannen. Medewerkers helpen bij personele ontwikkeling.	Intrinsieke motivatie als driver naar excellentie; sturen op waarden en verbinden op ambities (mogelijkheden en vaardigheden). Professionals zélf verantwoordelijk voor ontwikkeling.
Informatietechnologie ondersteunend aan bestaande manier van werken (mail, procesmanagement).	Informatietechnologie gebruiken om te innoveren in het primair proces, efficiënter en sneller intern samen te werken en op nieuw manier met partners en inwoners te werken: inspraak, gedeelde verantwoordelijkheid en successen delen.
Data gebruiken bij beleidsvorming, monitoring en verantwoording. Voor het management en bestuur ('planning en control'). Veel besluitvorming op basis van beelden en meningen: overmatige beleidspoëzie.	Beschikbare interne en externe data gebruiken om dynamisch doelen te definiëren steeds meer te gaan voorspellen en adviseren. Data primair voor professionals (en bestuur, partners en burgers) om keuzes te maken.
Lang cyclisch denken, doen en leren; in cycli van 1-2 jaar of zelfs langer. Nadruk op formeel groots evalueren. Plannen maken op strategisch, tactisch en operationeel niveau.	Kortcyclisch: doen, meten en leren; continu doen en leren, 'fail fast, fail cheap'. Een heldere visie hebben en doen (geen tactisch niveau).
Staf van 22% procent die lijnmanagers 'controlled' en overtuigt van innovatie. Veel logische regels op terrein van HR, Financiën, Juridisch.	Nadruk op innovatie en leren in het primair proces. Beperkte set met afspraken hoe de organisatie wenst te werken. Staf van 5%.

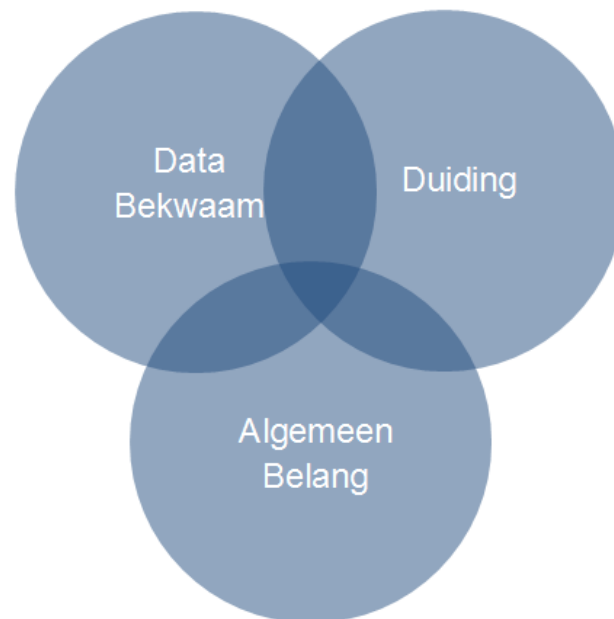
6. (Big) Data Gedreven Ambtenaar

Hoe ziet de nieuwe ambtenaar er uit?

De gemeente van de toekomst vraagt om een nieuw soort ambtenaar. Niet een specialist die veel weet van data mining, statistiek of programmeren, maar iemand met analytisch vermogen en begrijpt welke rollen op welk moment betrokken moeten worden.

Zelf zal de nieuwe ambtenaar dus vooral **data bekwaam** moeten zijn: de mogelijkheden en beperkingen van data kennen. Daarnaast is het geven van **duiding** een belangrijke factor: causaliteit ontdekken en resultaten van data analyse kunnen interpreteren.

Het derde aspect van de nieuwe ambtenaar is het minst vernieuwend, maar kent wel een nieuwe inslag. Het **algemeen belang** staat, net als vroeger, voorop. Maar door groeiende transparantie bij gemeenten zullen inwoners meer (participatie) invloed krijgen. Daar moet de nieuwe ambtenaar goed mee om weten te gaan.

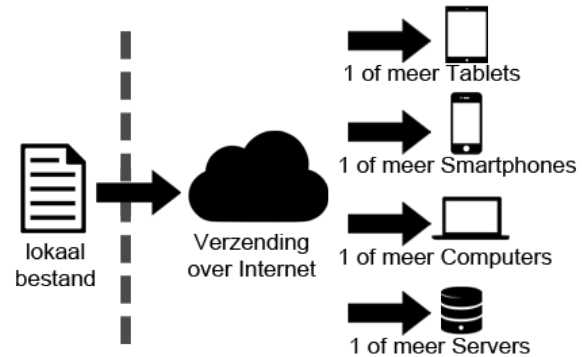


7. Security & Privacy

Vertrouwen, veilig en verbonden.

Wanneer bestanden verzonden worden via het internet komt het op allerlei plaatsen terecht. Op het moment dat een werknemer een bestand van zijn computer naar een externe organisatie stuurt, is het al snel toegankelijk op minstens vier verschillende externe apparaten. Het figuur rechts op de pagina toont hoe het voorheen lokale bestand verspreid wordt naar tablets, smartphones, computers en servers.

In veel gevallen is dit geen probleem en zelfs wenselijk. Vooral bij open data is de replicatie en de beveiliging een minder grote zorg. Echter, voor bewerking van (persoons)gevoelige gegevens in big data toepassingen geldt het tegenovergestelde. Men wil replicatie, verspreiding en ongeautoriseerde toegang zoveel mogelijk tegengaan.



Wenselijk beveiligingsniveau

Een digitale muur bouwen om computersystemen van een organisatie werkt tegenwoordig niet meer. Dat heeft veel te maken met het proces in het bovenstaande figuur. De nieuwe manier van beveiligen zit veel meer in het veilig ontwikkelen, up-to-date houden en in de gaten houden van data en systemen.

Gemeenten zullen voor big data oplossingen veelal gebruik maken van systemen en applicaties van derden. In veel gevallen is dat juist goed, omdat deze organisaties verantwoordelijk gemaakt kunnen worden voor de beveiliging van de systemen. Voor gemeenten zijn de volgende zaken belangrijk:

- Vraag de externe organisatie te laten zien dat zij basis (best practice) beveiligingsmaatregelen in acht nemen.
- Houd als gemeente een goed overzicht van welke data er is en welke datakoppelingen met externe (organisaties) er zijn.
- Gebruik waarde-classificaties op basis van Business Impact Analyse om zo beveiliging en privacy maatregelen af te stemmen op hetgeen wat beveiligd moet worden.

Waarde-classificatie

Het laatste punt behoeft verdere uitleg. De waarde-classificatie bepaalt het gewenste beveiligingsniveau. Maar hoe classificeer je data? Met behulp van standaarden en richtlijnen kun je uitspraken doen over de drie pijlers van beveiliging: vertrouwelijkheid, beschikbaarheid en integriteit van data. De standaarden en richtlijnen betogen in grote lijnen hetzelfde. De **vertrouwelijkheid** van de data kun je opdelen in drie klassen:

1. Vertrouwelijk. Alleen toegankelijk voor een beperkt aantal mensen waarvoor dit strikt noodzakelijk is.
2. Intern gebruik. Toegankelijk voor iedereen binnen een bepaalde afdeling of organisatie.
3. Openbaar. Toegankelijk voor een ieder.

Daarnaast is het in de informatiebeveiligingswereld gebruikelijk om te bepalen in hoeverre het IT proces of het IT systeem **beschikbaar** moet zijn. Bijvoorbeeld: hoe bedrijfskritisch is het HR-systeem van de gemeente? Mag het geen minuut offline zijn, geen uur of geen dag? Als laatste moet men bepalen hoe belangrijk de **integriteit** van de data is waarbij de focus ligt op het kunnen waarborgen van de accuraatheid en consistentie van de data. Voor sommige data is inaccuratie een minder groot probleem dan voor andere datasets. Neem bijvoorbeeld de volgende twee waarden: “Het aantal kilogram gestrooide zout in december” versus “Het rekeningnummer van meneer Jansen”.

Beoordeling van IT-systemen en data die zich daarin bevindt, wordt normaal gesproken gedaan met een

Business Impact Analyse (BIA)⁸. Het resultaat van een BIA is een matrix waarin gemakkelijk zichtbaar is welke systemen en data het meest kritisch zijn.

Privacy

Privacy komt direct terug in de beveiligingspijler vertrouwelijkheid. Daarmee wordt privacy gedeeltelijk ondervangen door beveiliging. Door de technieken op een juiste manier in te zetten kan vertrouwelijkheid gewaarborgd worden. Vanuit technisch perspectief wordt minder aandacht geschonken aan wetgeving en ethiek. Voor big data geldt anno 2015 temeer de vraag **'wat mag ik met data doen en op welke manier mag dat?'**

Niet-herleidbare data mag door een gemeente altijd gebruikt worden voor analyse. Vanuit het wetgevingskader zijn twee zaken belangrijk: doelbinding en bewerkersovereenkomst. Doelbinding is nodig wanneer men persoonsgevoelige data verwerkt. De

relevante vraag is: wordt de data gebruikt voor het doel waarvoor het verzameld is? Wanneer een gemeente er voor kiest om een derde partij in te zetten om analyses uit te voeren, dan is het belangrijk om een bewerkersovereenkomst te hebben met die partij. Daarmee krijgt die partij toestemming om bewerkingen uit te voeren op de aangeleverde data.

In een tijd waar veel data verzameld wordt, maar nog niet zoveel gekoppeld wordt, is het belangrijk om te beseffen dat er grote veranderingen aan komen. Neem **profiling**⁹. Profielen die gemaakt worden van mensen, processen en organisaties worden steeds nauwkeuriger. Privacy is in dat opzicht zeer belangrijk. Maar het moet fungeren als deur en niet als muur. Data gedreven sturing heeft de potentie om veel processen, en daarmee mensenlevens, te verbeteren. Als gemeente is het zaak om altijd nauwgezet met de privacy om te gaan door heldere doelen voor big data projecten te formuleren.

⁸ http://nl.wikipedia.org/wiki/Business_impact_analyse

⁹ Profiling zal in een volgende editie verder belicht worden.

IV Frequently Asked Questions

Ik wil morgen beginnen "iets te doen met big data". Wat zijn de eerste stappen die ik kan zetten?

Big Data moet klein beginnen. Zorg dat je je eigen (interne) databronnen in kaart hebt: relaties tussen datasets, kwaliteit, actualiteit, format, meta-data, eigenaarschap, privacy gevoeligheid, etcetera. En dat (meerdere) data eigenaren binnen de gemeente willen meewerken aan big data experimenten. Daarnaast kent Nederland verschillende Big Data start-ups, win eens (gratis) advies in over de mogelijkheden.

Ik heb een aantal moeilijk te beantwoorden (beleids) vragen. Hoe verken ik of big data daarin wellicht een antwoord kan bieden?

In de eerste plaats is het noodzakelijk om te komen tot een heldere vraagarticulatie. Om zo te beoordelen of de

inzet van big data zinvol is. De vraagstelling moet concreet en behapbaar moeten zijn. Dus niet: "*hoe milieubewust zijn mijn inwoners*", maar: "*zijn bewoners bereid en in staat om tijd en geld te investeren in energiebesparende maatregelen*". Vanuit een heldere vraagstelling kan beoordeeld worden of een big data oplossingsrichting geschikt is. Door vervolgens te richten op een *minimal viable product* kan een gemeenten tegen lage kosten inzicht krijgen in de werkelijke potentie. Verschillende bedrijven bieden Big Data producten die direct inzetbaar zijn. Zie ook hoofdstuk vier.

Hoe kom ik van een prototype of experiment tot een werkende en schaalbare oplossing voor de organisatie? Wanneer een experiment een zichtbare toegevoegde waarde (efficiëntie, kostenreductie, optimalisatie e.d.)

levert, is het zaak om een groep van 'early-adopters' aan te wijzen. Deze groep kan de rest van de organisatie enthousiast maken omtrent de effecten en potentie.

Een gemeente bestaat uit meerdere data eigenaren, hoe krijg ik de gemeente data op een locatie voor analyse?

Iedere afzonderlijke data eigenaar zal toestemming moeten geven. In de toekomst zal die rol waarschijnlijk centraal ingevuld gaan worden door een Chief Information Officer (CIO), welke centraal bepaalt wie tot welke data toegang kan krijgen.

Hoe ga ik als gemeente om met gemeente data die in handen is van derden?

Ontsluiting van data uit 'gesloten' systemen is een van de voornaamste uitdagingen bij een Big Data toepassing. Veelal zullen nieuwe afspraken gemaakt moeten worden met software leveranciers (of de interne IT organisatie), zodat zij op een gemakkelijke manier data kunnen leveren.

Welke privacy overwegingen zijn belangrijk?

Voor het meer-en-deel van de doelen waarvoor big data wordt ingezet, is het helemaal niet nodig om individua-data te gebruiken. Als gegevens worden niet herleidbaar zijn, zijn deze in veel gevallen vrij bruikbaar. Zie ook hoofdstuk 7, security & privacy.

Hoe krijg ik mensen/medewerkers mee in de data gedreven organisatie?

Wanneer een gemeente eenmaal heeft gekozen voor een Big Data platform (waar schoonmaak, opslag, analyse en visualisatie van data gedaan wordt), kunnen allerlei toepassingen daarop gezet worden. Zo heeft bijvoorbeeld iedere werknemer bij de organisatie Etsy – een beursgenoteerde internet onderneming gefocused op vintage producten - toegang tot hun eigen big data toepassingen¹⁰.

¹⁰ <http://www.networkworld.com/article/2163643/big-data-business-intelligence/etsy-gets-crafty-with-big-data.html>

Hoe zorg ik dat een big data project niet financieel uit de hand loopt?

Bijvoorbeeld door in te zetten op prototypes, waarbij samenwerking op basis van een resultaat gedreven afspraak met een Big Data organisatie aan ten grondslag ligt.

Welke open source tools zijn nu beschikbaar voor het beschikbaar maken van grote hoeveelheden data?

Gemeenten kunnen data publiceren en beschikbaar maken voor het publiek. Bijvoorbeeld via een webapplicatie als CKAN. Meer informatie: opendatahandbook.org.

Hoe combineer en analyseer ik de grote hoeveelheden data die wij hebben?

De uitdaging bij Big Data zit in het combineren van data uit verschillende bronnen in verschillende formats. Een eerste keer koppelen van een nieuwe databron met een big data platform behoeft altijd een handeling van een data scientist. Eenmaal gekoppeld, dan is het gemakkelijk om nieuwe data in te laden en te analyseren.

Appendix I.

Open Data onderzoek Nederlandse gemeenten

Live op www.bigdata.company/KING

Doel

Open data bevordert transparantie en creëert economische waarde. De Open Data Index beoogt het indexeren van de gepubliceerde open data door gemeenten en het beoordelen van de kwaliteit van de gepubliceerde datasets.

Methodiek

De datasets die gemeenten hebben gepubliceerd op hun eigen open data portal zijn gescraped (geautomatiseerd gedownload) en beoordeeld op verschillende aspecten. Wij hebben gekozen om de datasets te beoordelen op de volgende indicatoren:

- Compleetheid: hoeveel missing values heeft het databestanden?



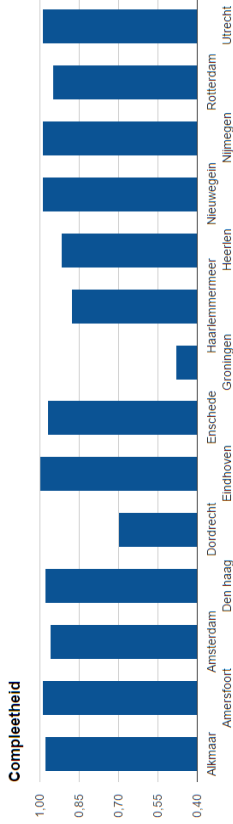
- Bestandsgrootte: hoe groot zijn de bestanden?
- Aantal bestanden: hoeveel bestanden zijn gepubliceerd op de open data portal?
- Toegankelijkheid: in welk bestandsformaat worden de datasets aangeboden?

De scores op de indicatoren opgedeeld in 5 klassen, waar A de hoogst haalbare klasse is en waar E de laagst haalbare klasse is.

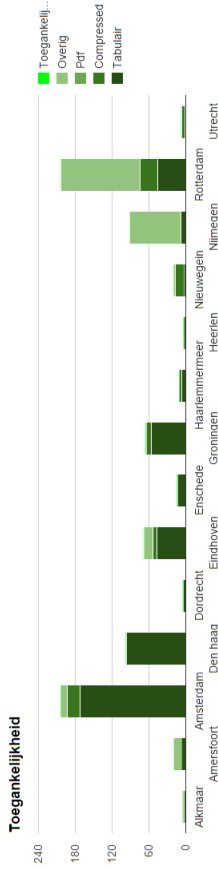
Resultaten

In totaal hebben 14 Nederlandse gemeenten een portal waar open data op gepubliceerd wordt. De kwaliteit en kwaliteit van de open data varieert sterk.

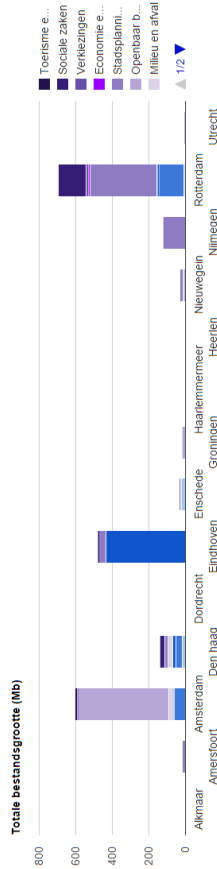
Indicator Compleetheid Datasets



Indicator Toegankelijkheid Datasets



Indicator Bestandsgrootte Datasets



Indicator Aantal Datasets

