



Ministerie van Justitie en Veiligheid

# Big Data

Waarborgen voor datakwaliteit >



# Big Data

## Waarborgen voor datakwaliteit

### Hoe kan zorgvuldig worden omgegaan met data, waarvan de kwaliteit niet vaststaat?



#### Definiëren kwaliteit

Datakwaliteit wordt verschillend gedefinieerd en is contextafhankelijk. Doordat de benodigde kwaliteit van data en het gebruik van bepaalde algoritmen per doel of toepassing kan verschillen, is het definiëren van kwaliteit een belangrijke eerste stap.



#### Kwaliteit bron vaststellen en onderhouden

Een bron is niet altijd betrouwbaar. Daarom dienen er checks and balances te worden ingebouwd in het analyseproces. Valideer de gegevens tegen andere bronnen. Eventuele fouten moeten worden teruggemeld aan de bron.



#### Kennis over de kwaliteit van eigen data

Naast de kwaliteit van (externe) bronnen is kennis en inzicht over kwaliteit van eigen data essentieel: deze kennis moet aanwezig zijn en expliciet gemaakt worden, bijvoorbeeld via het toekennen van een gewicht, voordat er stappen kunnen worden gezet.

### Hoe zorg je ervoor dat data-analysemethoden voldoende kunnen worden gecontroleerd?



#### Openheid en transparantie

Breng de werking van de analyse methode zo in kaart dat de stappen en het proces auditeerbaar zijn. Leg hierbij ook het doel van data-analyse uit. Uitgangspunt moet zijn dat algoritmen en methoden die bij data-analyses worden gebruikt deugdelijk zijn en aan de wetenschappelijke criteria voor goed (statistisch) onderzoek voldoen. Werk met waarschijnlijkheids/ betrouwbaarheidspercentages.



#### Belangenafweging

Het is van belang om zoveel mogelijk openheid te geven over data-analysemethoden. Hierbij moet het belang van transparantie vaak worden afgewogen tegenover andere belangen, waarbij rekening moet worden gehouden met 'gaming the system' waarbij kwaadwillenden de gewenste 'outcome' van data-analyse kunnen manipuleren.



#### Trainen en testen

Het trainen en testen van de data-analysemethode vormen belangrijke en gescheiden onderdelen van het gehele controleproces. Het is van belang ook bij deze onderdelen voldoende transparant te zijn.



# Big Data

## Waarborgen voor datakwaliteit

### Hoe zorg je ervoor dat aanpassingen in de data-analyse zorgvuldig worden gemaakt?



#### Standaardisatie

Wanneer data-analyse wordt toegepast, wordt de gehanteerde methode regelmatig herijkt op basis van resultaten en voortschrijdend inzicht. Daarbij is het van belang zoveel mogelijk gebruik te maken van gestandaardiseerde stappen en processen, zodat de kans op 'vrije' invulling van waarde niet mogelijk is en daarmee de kwaliteit kan worden geborgd.



#### Praktische aanpak

Het analyseren van data is vooral een praktisch en iteratief proces. Belangrijk hierbij is dat de focus gedurende het gehele proces op hetzelfde doel blijft, anders kan de (eigen) bias onbewust worden versterkt.

### Hoe kan controle van data-toepassingen het beste worden ingericht?



#### Interne controle

De onderzoeksresultaten, profielen en correlaties moeten op hun merites kunnen worden ingezien en gecontroleerd, waarbij uitkomsten voldoende worden onderbouwd.



#### Middelen

Als organisatie zou je middelen als AI kunnen inzetten om controlemogelijkheden van big data-toepassingen te verbeteren. Aanvullend is het van belang te realiseren dat burgers zelf niet altijd controlemacht hebben en dat interne controles niet voldoende zijn om vertrouwen te kweken. Door externe organisaties zoals NGO's of een Consumentenbond (externe) audits te laten doen, geef je als organisatie de boodschap het proces zeer serieus te nemen en daarin transparant op te willen treden.





## Colofon

Ministerie van Justitie en Veiligheid  
*Directie Informatisering & Inkoop*  
Postbus 20301 | 2500 EH Den Haag

[informatieplan@minvenj.nl](mailto:informatieplan@minvenj.nl)

Den Haag | september 2018

