

# BIG DATA ANALYTICS: KANSEN EN RISICO'S

**Auteur:** Drs. Piet Goeyenbier RE RA RO

**Leestijd:** 8 min



**De maatschappij heeft op dit moment te maken met een data-explosie, ook wel aangeduid als Big Data. Welke kansen biedt deze ontwikkeling organisaties en de internal auditor en welke risico's zijn er aan verbonden?**

De laatste jaren is er sprake van een enorme data-explosie, waarvan het einde nog niet in zicht is. Dit fenomeen wordt aangeduid met Big Data vanwege de grote opslag- en verwerkingscapaciteit die nodig is om deze datastroom op te slaan en te verwerken. In dit artikel wordt uiteengezet wat Big Data inhoudt en hoe dit fenomeen het werk van de auditor gaat beïnvloeden. Allereerst worden de relevante begrippen inzichtelijk gemaakt, als aanzet voor een gemeenschappelijk begrippenkader. Vervolgens worden de (toepassings)mogelijkheden behandeld van data-analyse en data mining en de hulpmiddelen die daarbij kunnen worden ingezet. Ten slotte wordt aandacht besteed aan de bijdragen die de internal auditor op dit terrein kan leveren.

## Big Data, wat houdt het in?

Zoals hiervoor aangegeven ontleent Big Data de naam aan de opslagcapaciteit die nodig is om zeer grote hoeveelheden data op te slaan en te verwerken tot informatie. In tabel 1 wordt een overzicht gegeven van de belangrijkste Big Data-bronnen. Waren 25 jaar geleden de interne gegevens de grootste gegevensbron, sinds de komst van internet is dat

volledig veranderd. Op dit moment verdubbelt de hoeveelheid data iedere twee jaar.

| In-/extern                                            | Transactiegegevens                                                                  | Open data<br>(Internet of content)                                | Communicatiegegevens<br>(Internet of people)                | Technische gegevens<br>(Internet of things)         |
|-------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------------------------------|
| Intern (beschikbaar<br>alleen voor de<br>organisatie) | ERP-, workflowmanagement,<br>documentmanagementsyste-<br>men, webportals, et cetera | Office, kennis- en<br>bibliotheeksystemen,<br>intranet, et cetera | E-mail                                                      | Logging-gegevens, TomTom<br>(locatiegegevens), rfid |
| Extern (breed<br>beschikbaar)                         |                                                                                     | Wikipedia, websites,<br>Open Data van overheid<br>en bedrijven    | Sociale media, Twitter,<br>Facebook, LinkedIn,<br>et cetera | Near field communication, Wifi<br>en Bluetooth)     |

Tabel 1. Overzicht elementen Big Data

Big Data wordt gekenmerkt aan de hand van de volgende vier V's: volume, velocity, variety en veracity (vrij vertaald: hoeveelheid, snelheid, diversiteit en waarheidsgetrouwheid). De uitdaging met Big Data is te achterhalen welke data relevant zijn en informatie leveren met toegevoegde waarde voor de organisatie.

Naast het begrip Big Data zijn de begrippen Open Data en Internet of Things relevant.

- Open Data wordt veelal gedefinieerd als de onbewerkte (bron)gegevens die door de overheden (en bedrijfsleven) voor de maatschappij beschikbaar worden gesteld en verwerkbaar zijn voor onder meer analysedoeleinden. Een voorbeeld is de basisregistratie adressen en gebouwen (BAG).<sup>1</sup>
- Internet of Things is een begrip dat vooral wordt gebruikt voor data die worden gegenereerd door apparaten zoals navigatiesystemen, intelligente energiemeters, et cetera. Het is deze groep die verantwoordelijk is voor de data-explosie. In het verlengde daarvan ligt Internet of locations (locatieaanduidingen via postcode, coördinaten, GSM, GPS en 3G/4G).

## Wat betekent Big Data voor de organisatie en auditor?

Bij Big Data gaat het om alle data die voor een organisatie beschikbaar zijn. Dat zijn de gegevens die opgeslagen zijn in de eigen informatiesystemen (zowel de proces- als de beheersystemen), alsook de data die verkrijgbaar zijn via internet (bijvoorbeeld de gegevens van en over concurrenten, klanten, Open Data, et cetera). Big Data biedt kansen voor zowel de organisatie als de auditor. De organisatie krijgt door het combineren van de data uit de verschillende in- en externe bronnen verrijkte gegevens waardoor een gerichtere (markt)benadering mogelijk is. Zo zou de overheid met behulp van de eerdergenoemde BAG na kunnen gaan in welke wijken en in welk type woningen het meest wordt ingebroken. Hierdoor kan gericht actie worden ondernomen om inbraken te voorkomen en de pakkans van inbrekers te vergroten. Ook kan mogelijke fraude met studiefinanciering of huurtoeslag worden opgespoord door na te gaan in welke woningen bovengemiddeld veel mensen wonen.

De informatie die beschikbaar komt uit de interne (transactie)systemen (ERP, WFM, DMS en logfiles) biedt naast de mogelijkheid tot verbetering van de effectiviteit ook mogelijkheden tot efficiencyverbetering door procesoptimalisatie. Daarnaast kan de informatiebeveiliging worden verbeterd door analyse van logginggegevens van toegangspogingen tot websites, infrastructuur, fysieke locaties, et cetera.

De uitdaging voor de organisatie betreft niet alleen de hoeveelheid data, maar vooral het vaststellen van welke data relevant zijn voor de organisatie en wat het opslaan, ontginnen en verwerken waard zijn.

De auditor kan de organisatie op de volgende terreinen ondersteunen:

- vaststellen dat met het beschikbaar stellen van de Open Data en verwerking van Big Data geen wetgeving (bijvoorbeeld privacy) wordt overtreden;

- bijdragen aan de kwaliteitsborging van het proces van data mining;
- uitvoeren van data-analyses om de beveiliging, efficiency, effectiviteit en de aantoonbaarheid/controleerbaarheid van de informatievoorziening en de processen in de organisatie te helpen verbeteren.

## In de jaren tachtig was data-analyse vooral het terrein van de auditor, die met tooling als IDEA en ACL data analyseerde

### Analysemethoden voor data

Door de jaren heen maken de data-analysemethoden dezelfde ontwikkeling door als de hiervoor geschetste ontwikkeling in de data zelf. In de jaren tachtig was data-analyse vooral het terrein van de auditor, die met tooling als IDEA en ACL, ter ondersteuning van de financial auditor, data analyseerde. De toepassingen waren gericht op de data die via de transactiesystemen beschikbaar waren. Er zijn diverse vormen van data-analyse: data-analyse, process mining, data mining en data analytics.

#### Data-analyse

Data-analyse is het proces van inspecteren, opschonen, transformeren en modelleren van data, teneinde nuttige informatie te ontdekken, conclusies te formuleren en besluitvorming te ondersteunen. Data-analyse betreft het gericht zoeken naar verbanden en specifieke gegevens, bijvoorbeeld afwijkingen in gestructureerde gegevensverzamelingen.

Toepassingsgebieden voor de *internal auditor*: voorheen werd data-analyse binnen audit gebruikt om het proces van de jaarrekeningcontrole te vereenvoudigen door het automatiseren van gegevensgerichte controlewerkzaamheden (zoals verbandcontroles, detailcontroles en cijferbeoordelingen). Op dit moment zijn er ontwikkelingen waarbij data-analyse steeds meer gebruikt wordt bij het beoordelen van de effectiviteit van beheersmaatregelen zoals application controls, functiescheiding en autorisaties.

#### Process mining

Process mining is een analysetechniek voor het analyseren van geautomatiseerde bedrijfsprocessen met behulp van event logs (logginggegevens van transactiesystemen ERP, WFM, DMS, et cetera). Bij process mining worden event logs gebruikt om van een proces geautomatiseerd een procesmodel te construeren. Deze procesmodellen geven inzicht in het werkelijke procesverloop (bestaan en werking), dat kan afwijken van de opzet.

Toepassingsgebieden voor de *internal auditor*: het geconstrueerde procesmodel geeft inzicht in de werkelijke uitvoering van de processen (transactiestromen, doorlooptijden, bottlenecks, functiescheidingen en afwijkingen). De internal auditor kan deze procesmodellen gebruiken om de effectiviteit en efficiëntie van de procesuitvoering te beoordelen.

## Data mining in termen van business intelligence is bijvoorbeeld belangrijk voor organisaties in relatie tot hun klantenkring

#### Data mining en data analytics

Data mining is het gericht zoeken naar (statistische) verbanden in gegevensverzamelingen met als doel profielen op te stellen voor wetenschappelijk of commercieel gebruik. Data mining richt zich op het ontginnen van data ten behoeve van voorspellingen. Met data mining kunnen gedrag en trends worden voorspeld en nog onbekende patronen worden ontdekt. Data mining wordt ook gebruikt om onverwachte zeer waardevolle informatie te vinden. De data zitten vaak diep weggestopt in bijvoorbeeld data warehouses. Bij data mining zie je de ontwikkeling dat er naast de gegevens uit de interne bronnen (transactiesystemen) steeds meer data worden gebruikt uit externe bronnen (internet en Open Data).

Toepassingsgebieden voor de *internal auditor*: data mining wordt in verschillende branches zoals overheid, verzekerings- en bankwezen toegepast, voor bijvoorbeeld de opsporing van mogelijk misbruik en/of fraude. Data mining in termen van business intelligence is bijvoorbeeld belangrijk voor organisaties in relatie tot hun klantenkring. Hiermee

kan een beter inzicht in de behoefte van hun klanten worden verkregen en daarmee bijdragen tot het uitbreiden van de klantenkring en omzet. Binnen Internal Audit wordt data mining vooral gehanteerd bij de beoordeling van financiële rapportages door te zoeken naar 'afwijkende' transacties. Data analytics en data science zijn nieuwe begrippen die de huidige en meer wetenschappelijke benadering van data mining duiden.

### Hulpmiddelen voor analyse van data

Zoals process mining een recente ontwikkeling is, zo zijn er ook actuele ontwikkelingen op het gebied van governance risk management en compliance (GRC) en security, information en event management (SIEM). Daarnaast is een ontwikkeling te zien in visualisatietooling. De visualisatietools zijn nuttig om de bomen door het bos weer te zien, zeker bij Big Data: een plaatje zegt nu eenmaal meer dan duizend woorden. Met bijvoorbeeld process mining en geovisualisatie kunnen goede inzichten worden verkregen in de locaties waar bijvoorbeeld pieken en/of knelpunten liggen (zie ook tabel 2).

| Soort analyse         | Data-analyse                                           | Process mining                        | Data mining/<br>Data analytics | Overige tooling (SIEM, techniek, Twitter en visualisatie)                |
|-----------------------|--------------------------------------------------------|---------------------------------------|--------------------------------|--------------------------------------------------------------------------|
| Voorbeelden van tools | IDEA, ACL, Excel, BWISE, GRC-tooling van SAP of Oracle | ProM, Disco, ReflectOne, PMOne, Nitro | SAS, Statistica, SPSS          | Splunk, ArcSight, Qradar, Twitonomy, MicroStrategy Analytics, VOS-viewer |

Tabel 2. Overzicht voorbeelden van tools data-analyse, process en data mining

De data-analysewerkzaamheden die worden uitgevoerd door de internal auditor in het kader van de governance en de jaarrekeningcontrole nemen in de toekomst toe. De nieuwe gegevens die beschikbaar komen, bieden meer mogelijkheden. De nieuwe hulpmiddelen (ook voor de gebieden GRC en SIEM) zijn vooral nuttig in het kader van de governance. Uiteindelijk zullen deze tools naar verwachting worden gebruikt door de eerste en vooral de tweede lijn. Internal Audit kan zich dan meer richten op de opvolging van aanbevelingen en oplossing van problemen (beide door de eerste lijn), die op basis van bevindingen door de tools worden gesignaleerd.

### De rol van IA bij het data mining proces

Data mining maakt een ontwikkeling door naar business intelligence en data analytics. De internal auditor kan een goede bijdrage leveren aan de kwaliteitsbeheersing van het gehele data mining proces. Bij de grote accountantskantoren zie je de ontwikkeling dat op het gebied van data mining steeds meer wiskundigen, statistici en econometristen worden aangetrokken. Deze functionarissen worden vaak data-analist of zelfs data-scientist genoemd. Mogelijk dat deze ontwikkeling ook plaats gaat vinden in organisaties en binnen Internal Audit.

Bij een data-analyse c.q. data mining- en data analytics-proces zijn de volgende stappen te onderkennen:

- analyse van de informatiebehoefte;
- ontwerp van de mogelijke toepassing, de business case;
- realisatie van de toepassing;
- evaluatie en testen van de gerealiseerde toepassing;
- gebruik en beheer van de gerealiseerde toepassing.

### Door het combineren van verschillende gegevensbronnen kunnen gegevensverzamelingen ontstaan met privacygevoelige, tot individuele personen herleidbare, informatie

De internal auditor en met name de IT-auditor kan op basis van zijn opleiding en ervaring een second opinion geven bij het resultaat van iedere stap. Tevens kan hij aangeven of er sprake is van een goede borging en inbedding van het data analytics-proces binnen de organisatie.

Bij overheidsorganisaties die de hofleveranciers zijn van Open Data, kan de internal auditor een rol spelen bij het borgen van de kwaliteit van de data die als Open Data beschikbaar wordt gesteld. De auditor kan dan onderzoek doen naar het totstandkomings- en beschikbaarstellingsproces van deze Open Data.

Andere aandachtspunten die samenhangen met Big Data en Open Data zijn de privacyproblematiek en de aansprakelijkheid. Door het combineren van verschillende gegevensbronnen kunnen gegevensverzamelingen ontstaan met privacygevoelige, tot individuele personen herleidbare, informatie. Door Open Data op een gedetailleerd niveau beschikbaar te stellen kunnen deze data relatief eenvoudig herleidbaar zijn tot individuele personen. Tevens is de kwaliteit (juistheid, volledigheid en tijdigheid) van de Open Data van belang. De organisatie die de Open Data beschikbaar heeft gesteld kan aansprakelijk worden gesteld, indien die Open Data geen eigendom is van deze organisatie en/of dat met het beschikbaar stellen van deze data belangen van derden (bedrijven of particulieren) worden geschaad. Op dit gebied kan de internal auditor een signaleringsfunctie hebben door tijdig aan te geven of de organisatie hier risico's loopt en of deze risico's in voldoende mate worden beheerst.

De ontwikkeling van data analytics biedt organisaties nieuwe mogelijkheden, zowel voor de beheersing als voor de verdere ontwikkeling van deze organisaties. De organisaties moeten zich bewust zijn van de data-explosie, de bijbehorende aandachtspunten en risico's en nagaan hoe zij daar hun voordeel mee kunnen doen. De internal auditor kan naast het gebruikmaken van deze technieken voor audit, risicobeheersing en governance doeleinden, een rol vervullen bij het toetsen van de kwaliteit van het data analytics proces.

#### Noot

1. In de BAG staan per adres (zie hiervoor [bagviewer.pdok.nl](http://bagviewer.pdok.nl)) de belangrijkste gegevens van een pand geregistreerd (bijvoorbeeld de gebruiksfunctie en oppervlakte).

Dit artikel is geschreven op persoonlijke titel. Collega drs. Hanifi Akcay RE (senior auditor bij ADR) dank ik hartelijk voor zijn bijdrage aan de totstandkoming van dit artikel.

#### Over

Piet Goeyenbier is auditmanager bij de Auditdienst Rijk (ADR) van het ministerie van Financiën. Hij is lid van de Commissie Vaktechniek van IIA Nederland. Verder is hij als extern deskundige betrokken bij de AITAP (post-doctorale IT-audit)opleiding aan de Amsterdam Business School (UvA) en penningmeester van Ngi-NGN Regio Den Haag.

Tags: data analytics

Bron url: <https://auditmagazine.nl/artikelen/big-data-analytics-kansen-en-risicos/>