Gebruikte afkortingen

|  |  |
| --- | --- |
| ADR | Auditdienst Rijk |
| AGL | Analyseteam Gegevenskwaliteit Loonaangifte |
| AVG | Algemene Verordening Gegevensbescherming |
| CAO  | Centrum voor Applicatie Ontwikkeling |
| CAP | Centrale Administratieve Processen |
| CIE | Centrum voor Infrastructuur en Exploitatie |
| D&A | Data & Analytics |
| DG | Directeur Generaal |
| DMF | Data Management Forum |
| EHi | Expertisecentrum Handhaving en intelligence |
| FIOD | Fiscale Inlichtingen en Opsporingsdienst |
| GO | Grote Ondernemingen |
| IA | Investeringsagenda |
| ICOV | Infobox Crimineel en Onverklaarbaar Vermogen |
| IO | Implementatieondersteuning |
| ISC | Internet Service Centre |
| IV | Informatievoorziening |
| Ki&S | Klantinteractie & -services |
| MKB | Midden- en Kleinbedrijf |
| MT | Managementteam |
| MTHV | Methoden, technieken, voorschriften en hulpmiddelen |
| OC | Onderdeelscommissie |
| PDB | Particulieren Dienstverlening en Bezwaren |
| SMART | Specifiek, Meetbaar, Acceptabel, Realistisch en Tijdsgebonden |
| TSL | Toeslagen |
| UCA | Unit Concernadministratie |

**Inhoudsopgave**

[1 Inleiding 4](#_Toc486344014)

[2 Definitie data-analyse 5](#_Toc486344015)

[3 Huidige situatie data-analyse 6](#_Toc486344016)

[Capaciteit data-analyse 6](#_Toc486344017)

[Typen data-analyse 7](#_Toc486344018)

[Opdrachtverstrekking 7](#_Toc486344019)

[Profielen en kruisbestuiving 7](#_Toc486344020)

# Inleiding

In de afgelopen jaren is **het gebied van data-analyse ingrijpend veranderd**. Enerzijds als gevolg door het beschikbaar komen van technologie, die nieuwe vormen van ontsluiting van data en analyses mogelijk maakt. De Belastingdienst heeft hierin fors geïnvesteerd. Dat heeft ertoe geleid dat er inmiddels op een aantal plekken intensief gewerkt wordt met een gemeenschappelijke data-laag, waarin met minimale investeringen al meerdere systemen gekoppeld zijn. Daarnaast is ervaring opgedaan met een nieuwe manier van werken om data-analyse gericht in te zetten om waarde toe te voegen in het primaire proces. Dit is mede mogelijk gemaakt door nieuw jong talent aan de Belastingdienst te binden dat specifieke expertise bezit op het gebied van geavanceerde data-analyse. En tegelijkertijd vindt op veel plaatsen binnen de Belastingdienst nog data-analyse op een meer traditionele manier plaats.

Een majeure verandering is gelegen in het maatschappelijk besef en de daarmee gepaard gaande eisen, toezicht en controle op privacy en beveiliging. De verplichting die de AVG (Algemene Verordening Gegevensbescherming) oplegt versterkt het interne besef dat de Belastingdienst aan de lat staat om zich goed te kunnen verantwoorden hoe met data(-analyses) van burgers en bedrijven wordt omgegaan.

In dit rapport wordt het **belang van data-analyse** voor de Belastingdienst als een reeds besloten vaststaand leidend principe beschouwd. Net als ten tijde van de initiële formulering van de Investeringsagenda wordt data-analyse gezien als een belangrijk middel om de primaire doelstellingen van de Belastingdienst beter, goedkoper en meer in control uit te kunnen voeren. Data-analyse maakt het bijvoorbeeld mogelijk om onderscheid te maken tussen burgers en bedrijven voor wat betreft hun bereidheid en hun vermogen hun verplichtingen na te komen. Daarnaast stelt data-analyse de Belastingdienst in staat om de interactie met burgers en bedrijven te moderniseren en om in het toezicht en de inning kwantitatief onderbouwde en controleerbare keuzes te maken en effecten te meten. Data-analyse is op deze manier een vereiste om inhoudelijke ambities van de veranderingen in de Belastingdienst te realiseren.

Ook buiten de Investeringsagenda is data-analyse niet weg te denken uit een moderne Belastingdienst. Effectieve sturing op operationele processen en effectieve verantwoording over behaalde resultaten kunnen niet zonder data-analyse. De vraag is daarmee niet óf data-analyse plaats moet vinden, maar hóe dit kan gebeuren op een manier die optimaal is voor de Belastingdienst.

De centrale vraag in deze excempt is

1. Hoe vindt data-analyse binnen de Belastingdienst in de huidige situatie plaats?

# Definitie data-analyse

In dit rapport wordt de volgende definitie gebruikt: het gaat om werkzaamheden die erop gericht zijn om beschikbare data analyses op te leveren die gebruikt worden ter ondersteuning van de managementaccounting (stuurinformatie / managementinformatie), de handhaving en de interactie met burgers en bedrijven (signaleringsanalyse / inzicht-analyse / risicoanalyse).

Bij de inventarisatie van de huidige situatie bleek dat ook met deze definitie in de hand nog veel ruimte voor interpretatie was. Ook doen **veel begrippen** de ronde om het werkveld nader te structureren, die vaak een gedeeltelijke overlappende betekenis hebben.

In dit rapport wordt ervoor gekozen om binnen het werkveld data-analyse **twee hoofdterreinen** te onderscheiden:

1. **Descriptieve data-analyse**: een feitelijke weergave van data ten doel van bijvoorbeeld managementinformatie, inzichten voor besluitvorming of dagelijkse operatie. Vragen die met behulp van deze analyses beantwoord kunnen worden, zijn bijvoorbeeld ‘wat is er gebeurd?’, ‘hoe veel, hoe vaak en waar is dit gebeurd?’ en ‘waar zit het probleem?’.
2. **Geavanceerde data-analyse[[1]](#footnote-1)**:door middel van geavanceerde statistische technieken worden waarde en intelligentie aan data toegevoegd. Typische analyses in deze categorie zijn signaleringsanalyse, statistische analyses, voorspellende analyses en prescriptieve analyses. Vragen die met behulp van deze analyses beantwoord kunnen worden, zijn bijvoorbeeld ‘waar is actie nodig?’, ‘waarom gebeurt dit?’, ‘wat gaat er gebeuren?’ en ‘wat is de beste aanpak?’.

Beide typen analyses dragen vanuit hun eigen invalshoek bij aan het realiseren van de **doelen van de Belastingdienst**. Descriptieve analyses helpen bij het efficiënt en accuraat functioneren van processen en een effectieve control op de organisatie. Geavanceerde analyses richten zich op het direct ondernemen van acties ten behoeven van het minimaliseren van risico’s en nalevingstekorten binnen de handhaving, het optimaliseren van de interactie met burgers en bedrijven en het borgen en/of verhogen van opbrengsten tegen lagere kosten.

Voor deze tweedeling is gekozen omdat de processen en capaciteiten die nodig zijn bij het opzetten, uitvoeren en implementeren van de twee soorten analyses verschillend zijn en bij de verdere uitwerking van de ontwerpprincipes daarmee andere keuzes kunnen worden gemaakt per type analyse.

# Huidige situatie data-analyse

Op basis van de werkwijze zoals beschreven in de inleiding is een inventarisatie gemaakt van de huidige situatie van het werkveld data-analyse. Figuur 1 hieronder geeft een **feitelijke weergave van de reacties** vanuit de betrokken bedrijfsonderdelen.

FIGUUR 1

|  |
| --- |
|  |

Uit deze inventarisatie en aanvullende gesprekken binnen de projectgroep volgen **vier voornaamste inzichten** over de huidige situatie (die hieronder nader toegelicht worden):

* Capaciteit voor data-analyse is omvangrijk en versnipperd georganiseerd,
* Descriptieve analyses vinden in alle bedrijfsonderdelen plaats, terwijl geavanceerde analyses bij minder bedrijfsonderdelen gebeuren,
* De verdeling tussen structurele analyses en ad hoc analyses is nagenoeg gelijk (wijze van opdrachtverstrekking),
* De profielen die werken aan data-analyse zijn divers, met beperkte kruisbestuiving.

## **Capaciteit data-analyse**

Binnen vrijwel alle organisatieonderdelen die zijn bevraagd, vinden data-analyses plaats. In totaal zijn volgens een eerste schatting van de organisatieonderdelen ongeveer 400 FTE bezig met data-analyse.

## **Typen data-analyse**

In de huidige situatie vindt **descriptieve data-analyse** bij bijna alle bevraagde bedrijfsonderdelen plaats.Over alle onderdelen heen is rond 25% van de capaciteit (ongeveer 100 FTE) vooral met data-analyses ten behoeve van stuur- en managementinformatie[[2]](#footnote-2) bezig. Hierbij horen bijvoorbeeld data-rapportages op concernniveau die vrij gestructureerd worden uitgevoerd door o.a. bedrijfsvoering.

In aanvulling op de concern-brede stuur- en managementinformatie vinden binnen vrijwel alle organisatie onderdelen aanvullende analyses plaats die deels bedoeld zijn als verdieping op de stuur- en managementinformatie. De manier van werken is divers / weinig gestandaardiseerd. Ook is het beeld dat beperkte transparantie bestaat over welke data bronnen waar en waarvoor worden gebruikt.

**Geavanceerde analyses** vinden op een beperkter aantal plekken plaats, o.a. bij FIOD, D&A, MKB, Toeslagen en Douane. Voorbeelden van analyses, die genoemd zijn in de inventarisatie zijn: analyses voor regie van handhaving op kantoren (bijvoorbeeld binnen MKB), modellen ter signalering van risico’s en fraude (van bijvoorbeeld Toeslagen, D&A, FIOD, Douane), onderzoek over naleveringstekorten, etc. Voor deze analyses wordt gebruikt gemaakt van tools die het analyseren en visualiseren van grotere hoeveelheden data ondersteunen (bijv. SAS, Access, ACL etc.).

## **Opdrachtverstrekking**

Een deel van de data-analyses vindt plaats op basis van een specifieke interne (permanente) opdracht en een deel wordt uitgevoerd vanuit een ad hoc opkomende vraag. De verhouding is, op basis van deze inventarisatie en gecorrigeerd voor de hoeveelheid FTE per onderdeel, globaal als volgt: 58 % structurele opdrachten, 42 % ad hoc werk.

De permanente (meer structurele) opdrachten zijn over het algemeen gebonden aan een voorgeschreven proces (met voorgeschreven procedures, kwaliteitsnormen en proces van managen van opdrachten) en een vast omschreven control-cyclus.

Zodra de opdrachten wat meer ongestructureerd van aard zijn, worden de procedures wat 'losser'. Deze worden uitgevoerd op basis van een beperkt validatieproces en een beperktere control op naleving kwaliteitsnormen (juridisch, procedures, data en methoden), naleving procedures en beperkte evaluatie.

## **Profielen en kruisbestuiving**

De diversiteit van mensen die werken aan data-analyse is groot qua ervaring en opleiding. Zo werken bij sommige onderdelen vooral universitair opgeleide econometristen, gedragswetenschappers en statistici als “*data scientists*”. Bij andere onderdelen (vooral MKB) werken veel medewerkers met een opleiding op gebied van “*electronic data processing*” (EDP’ers). Op weer andere plekken worden data-analyse werkzaamheden vooral uitgevoerd door medewerkers uit de business met affiniteit en enige kennis over data-analyse, maar zonder formele opleiding op dit gebied. De samenwerking en onderliggende uitwisseling van expertise, modellen, etc. tussen deze verschillende groepen is volgens meerdere organisatieonderdelen in de huidige situatie beperkt.

1. Andere begrippen die de ronde doen en binnen dit werkveld vallen zijn, ‘signaleringsanalyses’, ‘statistische analyses’, ‘voorspellende analyses’ en ‘prescriptieve analyses’ [↑](#footnote-ref-1)
2. Het onderscheid tussen stuur- of procesinformatie en managementinformatie of bestuurlijke informatie zit “losjes” in de frequentie (dagelijks vs. wekelijks), mate aggregatie (op een proces of over meerdere processen) en doel (operationaal vs tactisch/strategisch). Een exacte afbakening is een van de vervolgonderzoeken om tot eenduidig taalgebruik te komen. [↑](#footnote-ref-2)