

De toepassing van datamining ter bestrijding van discriminatie op de arbeidsmarkt

Samenvatting

Op basis van de ordonnantie van 16 november 2017 werd de Gewestelijke Werkgelegenheidsinspectie (GWI/IRE) belast met het uitvoeren van praktijktesten in het kader van de inbreuken op de antidiscriminatie-ordonnanties van 4 september 2008. Dit was een antwoord op de bewijsproblemen bij discriminerende praktijken op de arbeidsmarkt. Deze tests zijn strikt omkaderd in de regelgeving: ze kunnen enkel worden uitgevoerd als er ernstige aanwijzingen van discriminatie zijn en nádat er een klacht of melding werd ingediend.

In deze aanbeveling gaan we in op de mogelijkheid om dataminingstechnieken te gebruiken als ernstige aanwijzing voor discriminerende praktijken. Datamining leidt tot knipperlichten die een objectieve aanwijzing vormen voor mogelijke discriminatie, en op basis waarvan gerichte controles kunnen worden uitgevoerd (zie punten 1 en 2).

We stellen verschillende manieren voor om datamining uit te voeren. Enerzijds datamining in de strikte zin van het woord, waarbij algoritmen moeten worden gecreëerd die tot doel hebben de databanken systematisch te doorzoeken op gedragingen die mogelijk overkomen met discriminerende praktijken van ondernemingen (zie punten 3 en 4). Anderzijds bestaan er ook vereenvoudigde dataminingstechnieken. Zo zijn er bijvoorbeeld analyses die het mogelijk maken om de personeelssamenstelling van een bedrijf te vergelijken met de gemiddelde samenstelling van werknemers in die sector. Zo kunnen ondernemingen geïdentificeerd worden die significant verschillen van dat gemiddelde. Een andere mogelijkheid van vereenvoudigde datamining is een beter gebruik van de databanken van reeds uitgevoerde controles en van meldingen (zie punt 5).

Unia pleit ervoor om voldoende financiële en personele middelen in te zetten om dataminingstechnieken te ontwikkelen. Dat zal het mogelijk maken om te evolueren van een reactieve controles op basis van klachten - die momenteel op zijn grenzen botst zoals aangetoond in het recente Unia-rapport over praktijktesten in het Brussels Hoofdstedelijk Gewest - naar een strategie van gerichte discriminatiecontroles op basis van knipperlichten. Die knipperlichten vormen objectieve aanwijzingen voor mogelijke discriminatiepraktijken. De werkzaamheden op het gebied van datamining die door zowel de federale als de gewestelijke inspectiedienst worden uitgevoerd, zijn complementair. Unia beveelt de oprichting aan van een partnerschap tussen de federale inspectie Toezicht op de Sociale Wetten en de Brusselse Gewestelijke Werkgelegenheidsinspectie. Zij kunnen een gezamenlijk datamining-project uitvoeren, waardoor het mogelijk wordt om federale en regionale databanken te kruisen, technische vooruitgang op het gebied van datamininganalyse uit te wisselen en inspectie-expertise te bundelen. Andere actoren die over gegevens of expertise beschikken, moeten bij dit gezamenlijke project worden betrokken, zoals bijvoorbeeld de RSZ, Unia, het Instituut voor Gelijkheid van Vrouwen en Mannen, Actiris, enz. Dit innovatieve project zal verschillende actoren in staat stellen om elkaar te versterken en een reëel antwoord te bieden op de strijd tegen discriminatie als politieke prioriteit in het Brussels Hoofdstedelijk Gewest.

Inhoud

1	Wat is datamining?	1
2	Datamining en sociale inspectiediensten: van sociale en fiscale fraude tot discriminatie	1
3	Het koppelen van databanken met expertise op het gebied van discriminatie	1
4	Concrete discriminatiescenario's	2
5	Vereenvoudigde datamining	4
6	Aanbevelingen	5

1 Wat is datamining?

Datamining verwijst naar verschillende statistische hulpmiddelen en informaticatoepassingen die de analyse van een grote hoeveelheid gegevens mogelijk maken. Het doel van datamining is verbanden in de data blootleggen en terugkerende dynamieken die niet direct zichtbaar zijn bij het beschouwen van de massa aan data als geheel. Het is dus een kwestie van het onthullen van informatie, vervat in de massa van gegevens. Deze informatie kan dan worden gebruikt om beslissingen te nemen.

2 Datamining en sociale inspectiediensten: van sociale en fiscale fraude tot discriminatie

In België wordt datamining op federaal niveau al op bevredigende wijze gebruikt om sociale en fiscale fraude te bestrijden. Door de gegevens van de duizenden controles die de Sociale Inspectie al heeft uitgevoerd te koppelen aan "sociale" gegevens over bedrijven en hun werknemers, heeft datamining het mogelijk gemaakt om bedrijven of sectoren te identificeren die een risico op sociale fraude lopen. Hierdoor is het mogelijk om gerichte controles uit te voeren en meer overtredingen op te sporen. Op het gebied van belastingfraude was de Bijzondere Belastinginspectie een van de pioniers in Europa met betrekking tot het gebruik van datamining voor het analyseren van frauduleuze netwerken, waardoor in 2014 bijna een miljard euro aan overheidsbijdragen is binnengehaald¹. Zo heeft datamining de belastingdienst in staat gesteld om miljoenen gegevens uit Offshore Leaks te verwerken om de omvang van ontdoken activa op te sporen en te documenteren.

Dezelfde werkwijze kan worden gebruikt bij het zoeken naar discriminatiemisdrijven. De datamining legt de nadruk op aanwijzingen van potentieel discriminerende praktijken binnen specifieke sectoren, privébedrijven, overheidsinstellingen of arbeidsbemiddelaars. Die aanwijzingen zullen dus dienen als knipperlichten voor de federale inspectie Toezicht op de Sociale Wetten (TSW/CLS) en de Gewestelijke Werkgelegenheidsinspectie (GWI/IRE) waardoor meer gerichte onderzoeken en controles mogelijk worden. Die strategie laat dus toe om doelgericht op zoek te gaan naar situaties van discriminatie, en zo de efficiëntie van de controles te vergroten. De datamining maakt het mogelijk om alleen die bedrijven of overheidsinstellingen te controleren waar de analyse van geaggregeerde gegevens aanwijzingen van mogelijke discriminatie aan het licht brengen. Datamining alleen zal echter niet volstaan om te concluderen dat er sprake is van discriminatie, maar het zou wel een van de aanwijzingen kunnen vormen voor het bestaan van mogelijks discriminerende praktijken.

3 Het koppelen van databanken met expertise op het gebied van discriminatie

De inspectiediensten beschikken over omvangrijke administratieve databanken die door de werkgevers worden gevoed in het kader van hun administratieve taken. Die databanken bevatten informatie over bedrijven (grootte, sector, enz.) en over de personen die er werken, zoals hun statuut en de duur van hun tewerkstelling, alsook enkele van hun persoonlijke kenmerken (geslacht, leeftijd, nationaliteit, woonplaats, enz.). Het is technisch mogelijk om ze ook te koppelen aan andere administratieve databanken, zoals de Kruispuntbank van de Sociale Zekerheid (KSZ). Dit zou het mogelijk maken om informatie over persoonlijke kenmerken toe te voegen, met name de origine van de persoon (waardoor het mogelijk wordt om verder te gaan dan de nationaliteit). De RSZ heeft al toegang tot dit soort

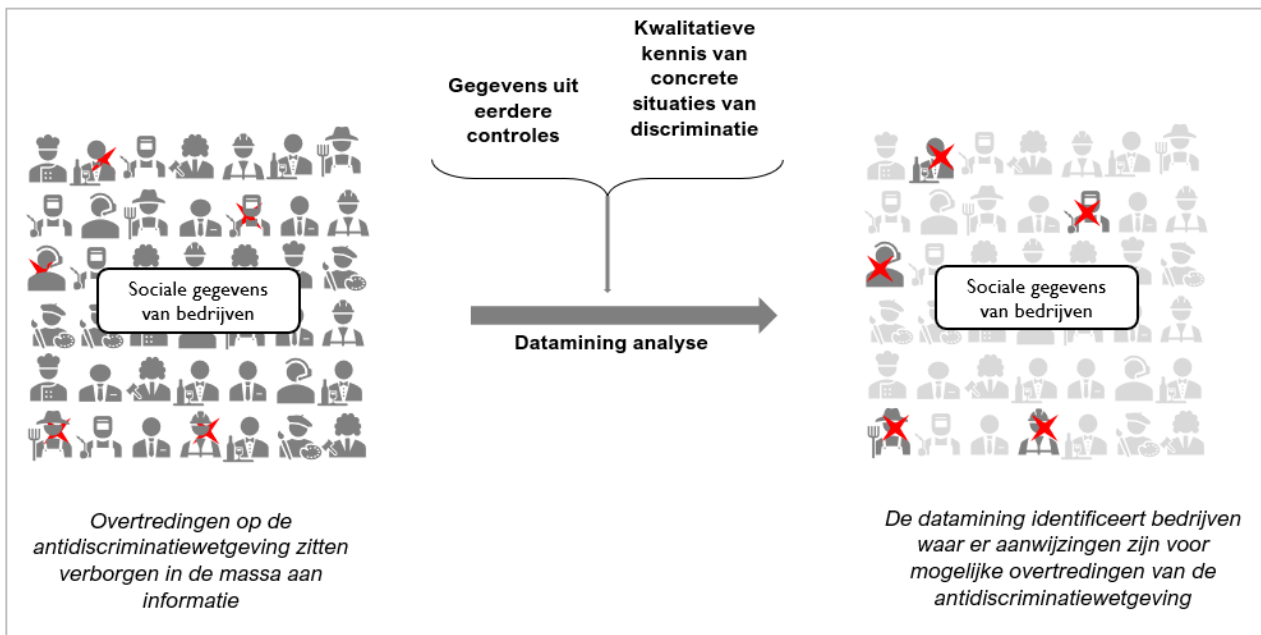
¹ Zie hier : <https://www.agoria.be/nl/Big-data-of-hoe-de-fiscus-ruim-1-miljard-euro-fraude-wist-terug-te-halen>.

gegevens en de inspectiedienst TSW/CLS heeft recent een verzoek ingediend bij de Gegevensbeschermingsautoriteit om zelf ook rechtstreeks toegang te krijgen tot dit soort gegevens.

Het toepassen van datamining bij de bestrijding van discriminatie zou betekenen dat deze databanken worden geanalyseerd om typische discriminatiescenario's ("patronen") te identificeren met behulp van adequate algoritmen. Deze typische discriminatiescenario's moeten ontwikkeld worden op basis van enerzijds de informatie die is verzameld tijdens de talrijke discriminatiecontroles die de federale en gewestelijke inspectiediensten reeds hebben uitgevoerd en anderzijds de kwalitatieve kennis van concrete situaties van discriminatie op de arbeidsmarkt. Die kennis van concrete situaties kan aangeleverd worden door deskundigen op het gebied van discriminatie, in het bijzonder door de verschillende diensten die meldingen en klachten over discriminatie op het gebied van werkgelegenheid verzamelen en behandelen (in het bijzonder Unia, het IGVM, de antidiscriminatie dienst van Actiris Inclusive).

De analyse van de gegevens volgens deze modelscenario's zal een lijst opleveren van bedrijven of overheidsinstellingen waarvoor aanwijzingen van discriminerende praktijken zijn gevonden. De datamining resulteert dus in knipperlichten die aanleiding kunnen geven tot gerichte antidiscriminatiecontroles door de Gewestelijke Werkgelegenheidsinspectie (GWI/IRE).

Figuur 1: Schematische werking van datamining



4 Concrete discriminatiescenario's

Er kunnen veel scenario's gedefinieerd worden om patronen van risicogedrag te ontdekken. Hieronder volgen voorbeelden van discriminatiescenario's die gezocht kunnen worden in de gegevens:

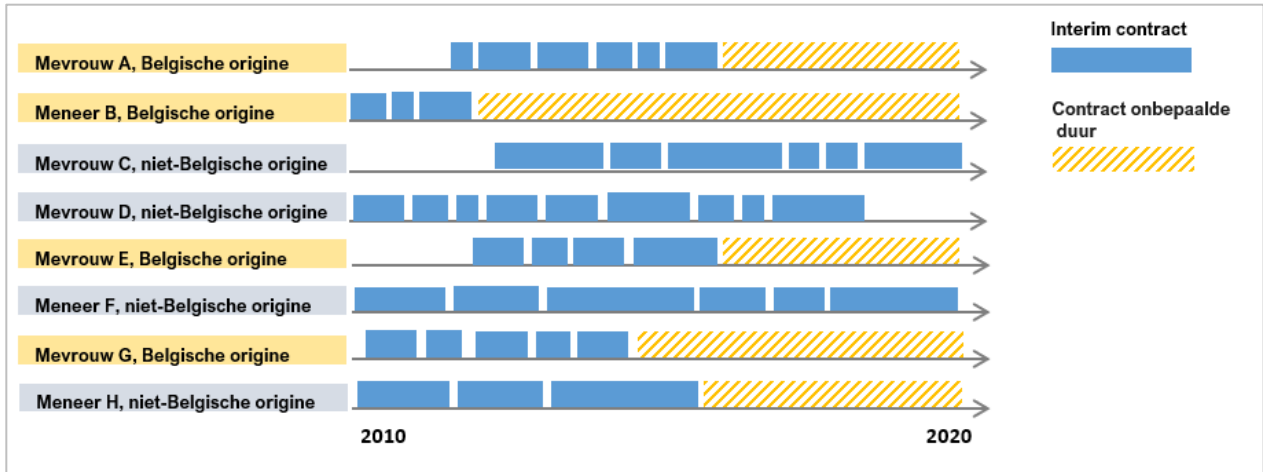
- 1) een bedrijf dat geen oudere werknemers in dienst neemt;
- 2) een onderneming die mensen van Belgische en buitenlandse origine aanwerft via een uitzendbureau, maar nadien enkel mensen van Belgische origine aanwerft in vaste contracten;
- 3) een door het BHG gesubsidieerde organisatie (> 50%) die systematisch mensen ontslaat na een periode van arbeidsongeschiktheid;
- 4) een bedrijf dat geen vrouwen onder de 30 jaar in dienst neemt;
- 5) een onderneming die lagere lonen toekent op basis van de origine of het geslacht van de werknemers;

6) ...

In figuur 2 hieronder is de gegevensanalyse weergegeven voor voorbeeld n°2. Hiervoor is een database nodig met:

- ✓ informatie over de in- en uitstroom van werknemers in het bedrijf (Dimona-in en -out) en het type contract;
- ✓ informatie over de origine van personen (informatie beschikbaar in de KSZ);
- ✓ voor alle Brusselse bedrijven in een sector tussen 2010 en 2020.

Figuur 2: Schematische weergave van een scenario van discriminatie op basis van herkomst van de loontrekkenden (voorbeeld 2).



In deze database kan gezocht worden naar de scenario's "interim + onbepaalde duur" en "enkel interim". In ons voorbeeld bevindt 88% van de werknemers van Belgische origine van het bedrijf zich in een "scenario interim + onbepaalde duur". Terwijl slechts 12% van de werknemers van buitenlandse origine (= meneer H) zich in datzelfde scenario bevindt. Dit verschil in proportie zorgt voor een knipperlicht, wat wijst op een potentieel discriminerende praktijk in dit bedrijf. Deze aanwijzing moet natuurlijk geïnterpreteerd worden; het is geen feitelijke vaststelling van discriminatie. In dit voorbeeld hebben we de profielen vereenvoudigd. Er is dus in werkelijkheid meer informatie nodig alvorens groepen van werknemers te kunnen vergelijken (niveau van de diploma's, anciënniteit, enz.).

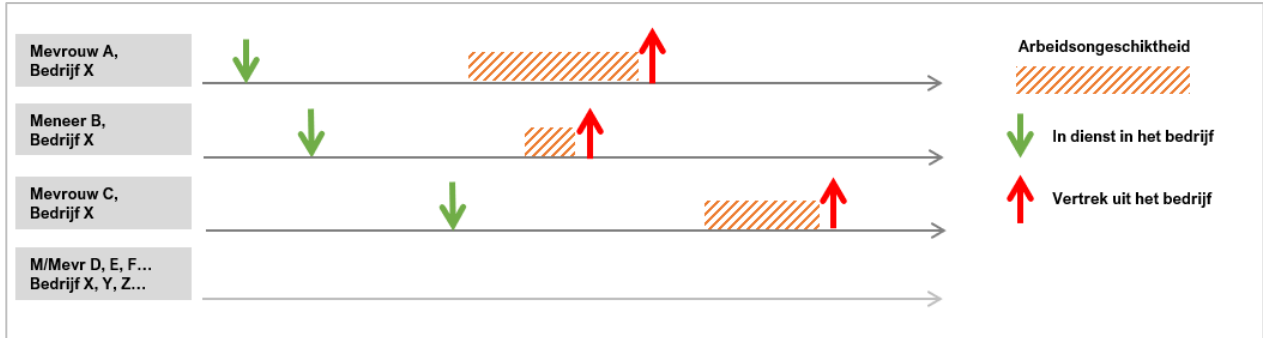
Het onderstaande schema maakt het mogelijk om het mechanisme van datamining visueel weer te geven met betrekking tot voorbeeld n°3. In dit voorbeeld wordt een database gebruikt die de volgende informatie bevat:

- ✓ het binnenkomen en verlaten van het bedrijf (Dimona-in en -uit)
- ✓ perioden van arbeidsongeschiktheid,
- ✓ voor alle Brusselse bedrijven in een bepaalde sector van 2010 tot 2020

In deze massa van gegevens is de bedoeling om de volgende individuele gedragingen te identificeren: werknemers die een bedrijf snel verlaten na een periode van arbeidsongeschiktheid. Dit gedrag zal worden aangeduid als het "arbeidsongeschiktheid + vertrek" scenario. Datamining stelt ons dus in staat om het carrièreparcours van alle werknemers in bedrijven in een sector in één keer te analyseren en om te identificeren of dit typische "arbeidsongeschiktheid + vertrek" scenario herhaaldelijk voorkomt voor meerdere werknemers binnen een bedrijf.

Dit kan door een algoritme te creëren om dit "arbeidsongeschiktheid + vertrek" scenario in de database te identificeren. Het algoritme zal de bedrijven aangeven waarin dit scenario op een herhaalde of systematische manier geïdentificeerd wordt. Dit zal een "knipperlicht" vormen voor organisaties die een risicoprofiel vertonen. De gewestelijke inspectiedienst zal deze signalen moeten interpreteren en kan op basis van dit knipperlicht beslissen om al dan niet een gerichte controle te organiseren bij het bedrijf.

Figuur 3: Schematische weergave van een scenario van discriminatie op basis van gezondheidstoestand (voorbeeld 3).



De interpretatie van de knipperlichten door de gewestelijke inspectiedienst is een belangrijke tussenstap, omdat het slechts een aanwijzing is van een risicoprofiel en niet een bevinding van daadwerkelijke discriminatie. In dit concrete voorbeeld moet het verdachte patroon van de onderneming bij wie een knipperlicht afgaat nog vergeleken worden met het gemiddelde gedrag van vergelijkbare ondernemingen (sector, omvang, enz.).

5 Vereenvoudigde datamining

Die laatste werkwijze kan ook gezien worden als een "vereenvoudigde" manier om bepaalde aspecten van datamining te gebruiken. De databanken geven immers informatie over het sociaal-demografisch profiel van de werknemerspopulatie van elk bedrijf, per sector. Het is dus mogelijk om een "gemiddeld profiel" van werknemers in een sector te definiëren en eenvoudigweg de populatie van werknemers in bedrijven te vergelijken om te bepalen welke bedrijven significant verschillen van het "gemiddelde profiel" van bedrijven in deze sector.

Hieronder wordt een concreet voorbeeld beschreven met betrekking tot discriminatie op basis van nationale herkomst. Voor een bepaalde sector kan men het gemiddelde aandeel werknemers van Belgische en niet-Belgische origine berekenen (waarbij eventueel een onderscheid wordt gemaakt tussen EU en niet-EU origine). Deze gemiddelde verhoudingen worden dan vergeleken met de verhoudingen in elk bedrijf in die sector. Het doel van deze vergelijking is het identificeren van bedrijven waar het aandeel van de werknemers van Belgische origine zeer verschillend is van het gemiddelde aandeel (bijvoorbeeld zeer hoog). Dit is een eerste aanwijzing van een potentieel problematische situatie. Deze vergelijkingen kunnen ook worden gemaakt voor het aandeel van mannen en vrouwen, of verschillende leeftijdsgroepen (bijvoorbeeld werknemers van 50 jaar en ouder).

De relevantie van de resultaten van deze analyse wordt bepaald door de relevantie van de geselecteerde groep van ondernemingen. Met andere woorden, het is belangrijk om vergelijkbare ondernemingen te vergelijken, bijvoorbeeld wat betreft de omvang of het soort activiteit, of de tijdelijkheid van de werkgelegenheid (bijvoorbeeld in het geval van seizoensarbeid). De analyse van de vergelijking is ook belangrijk: in hoeverre kan het verschil tussen het gemiddelde aandeel en het aandeel in de onderneming worden beschouwd als een indicatie van een potentieel probleem? Dit zal besproken en getest moeten worden om tot een degelijk beslissingsondersteunend instrument te komen.

Het gebruik van databestanden met reeds uitgevoerde controles en lijsten van bedrijven die al het onderwerp zijn geweest van klachten over discriminatie is ook een mogelijkheid die onderzocht kan worden.

6 Aanbevelingen

In lijn met haar aanbeveling "Discriminatie op het werk: opsporing en toezicht door de federale arbeidsinspectie"² van maart 2017, pleit Unia ervoor om voldoende financiële en personele middelen in te zetten om dataminingstechnieken te ontwikkelen. Dit zal het mogelijk maken om te evolueren van reactieve controles op basis van klachten - die momenteel op zijn grenzen botst zoals aangetoond in het recente Unia-rapport over praktijktests in het Brusselse Gewest - naar een strategie van gerichte discriminatiecontroles op basis van knipperlichten. Die knipperlichten vormen objectieve aanwijzingen voor mogelijke discriminatiepraktijken. De werkzaamheden op het gebied van datamining die door zowel de federale als de gewestelijke inspectiedienst worden uitgevoerd, zijn complementair. Unia beveelt de oprichting aan van een partnerschap tussen de inspecties TSW /CLS en GIW/IRE om een gezamenlijk datamining-project uit te voeren, waardoor het mogelijk wordt om federale en regionale databanken te kruisen, technische vooruitgang op het gebied van datamininganalyse uit te wisselen en inspectie-expertise te bundelen. Andere actoren, die over gegevens of expertise beschikken, moeten bij dit gezamenlijke project worden betrokken, zoals bijvoorbeeld de RSZ, Unia, het Instituut voor Gelijkheid van Vrouwen en Mannen, Actiris, enz. Dit innovatieve project zal verschillende actoren in staat stellen om elkaar te versterken en een reëel antwoord te bieden op de strijd tegen discriminatie als politieke prioriteit in het Brussels Hoofdstedelijk Gewest.

Vermeldenswaardig is dat de RSZ en de federale Inspectie Toezicht op de Sociale Wetten (TSW/CLS) - op initiatief van de Minister van Tewerkstelling van de vorige regering - al begonnen zijn met een reflectie over en het uitvoeren van technische werkzaamheden met betrekking tot de toepassing van datamining in de zoektocht naar tekenen van discriminatie. Een partnerschap zou dus zeker gunstig zijn voor de werkzaamheden van het Brussels Hoofdstedelijk Gewest op dit gebied.

Concreet bevat deze aanbeveling verschillende voorstellen voor pistes om datamininganalyses uit te voeren. Met betrekking tot de meest geavanceerde datamining technieken zal het nodig zijn om 1) standaard discriminatiescenario's te ontwikkelen die 2) gemodelleerd kunnen worden door algoritmes die 3) in bestaande (of aan te maken) databestanden kunnen worden gebruikt. Het technische karakter van dit programma is belangrijk en zal gespecialiseerde kennis en ongetwijfeld een experimenteerfase vereisen. Daarom beveelt Unia dit partnerschap aan tussen, enerzijds specialisten die al gebruik maken van datamining en, anderzijds actoren in de strijd tegen discriminatie om relevante algoritmen te ontwerpen en te testen.

Bovendien zou de SIOD (Sociale Inlichtingen- en Opsporingsdienst), als coördinerend orgaan dat de sociale inspectiediensten ondersteunt in hun strijd tegen sociale fraude, ook een partner of actor kunnen zijn in het kader van de antidiscriminatie-datamining. Als zodanig is het een eerste aanspreekpunt voor het ontvangen van informatie en goede praktijken. Ten slotte biedt SMALS ook informatie en diensten aan voor het uitvoeren van datamininganalyses, als een non-profitorganisatie die verantwoordelijk is voor het ondersteunen en bijstaan van het ICT-beheer van overheidsorganen.

Unia contacten

Bruno Martens (bruno.martens@unia.be)

Louise Callier (louise.callier@unia.be)

² Te raadplegen via <https://www.unia.be/files/Documenten/Aanbevelingen-advies/FED - 165 - Opsporing en toezicht door de arbeidsinspectie.pdf>