

**Sectoraal Comité van de Sociale Zekerheid en van de Gezondheid
Afdeling « Sociale Zekerheid »**

SCSZ/08/155

BERAADSLAGING NR. 08/047 VAN 2 SEPTEMBER 2008 MET BETREKKING TOT DE MEDEDELING VAN GECODEERDE PERSOONSgegevens DOOR DE KRUISPUNTBANK VAN DE SOCIALE ZEKERHEID AAN DE UNIVERSITÉ LIBRE DE BRUXELLES, DE KATHOLIEKE UNIVERSITEIT LEUVEN EN DE UNIVERSITÉ CATHOLIQUE DE LOUVAIN MET HET OOG OP HET OPTIMALISEREN VAN HET PROJECT OASIS DOOR MIDDEL VAN DE INTEGRATIE VAN DATAMINING EN HET BETER AFBAKENEN VAN HET FENOMEEN SOCIALE FRAUDE

Gelet op de wet van 15 januari 1990 *houdende oprichting en organisatie van een Kruispuntbank van de Sociale Zekerheid*, inzonderheid op de artikelen 5 en 15, § 1;

Gelet op de aanvraag van de Université Libre de Bruxelles, de Katholieke Universiteit Leuven en de Université Catholique de Louvain van 11 augustus 2008;

Gelet op het auditoraatsrapport van de Kruispuntbank van de Sociale Zekerheid van 20 augustus 2008;

Gelet op het verslag van de voorzitter.

A. ONDERWERP VAN DE AANVRAAG

- 1.1.** De Université Libre de Bruxelles, de Katholieke Universiteit Leuven en de Université Catholique de Louvain verrichten momenteel een onderzoek met een dubbel opzet: enerzijds het optimaliseren van het project OASIS door er datamining in te integreren en anderzijds het beter afbakenen van het fenomeen sociale fraude.

Op dit ogenblik bestaat er in het kader van de samenwerking tussen sociale inspectiediensten een systeem van alarmen voor frauderisico's, het project OASIS. De sociale inspecteurs baseren zich op OASIS om te selecteren welke ondernemingen gecontroleerd zullen worden. Het systeem verschaft eveneens samenvattende sociale informatie per onderneming. Een belangrijke doelstelling van OASIS bestaat erin om zo

snel mogelijk te kunnen reageren op eventuele fraude. De selectie van ondernemingen gebeurt op basis van knipperlichten, die op basis van de expertise van de sociale inspecteurs in het systeem werden ingebouwd. Tot nu toe bestaat er echter geen enkele toepassing (datamining) die deze knipperlichten kan controleren, valideren of verbeteren. Bovendien werden de gegevens tot hiertoe uitsluitend gebruikt door de sociale inspectiediensten.

De dynamische modellering, die mogelijk gemaakt wordt door de datamining, zal een fijnere beschrijving mogelijk maken van (nieuwe) mechanismen van fraude: de inspectiediensten, de politieke verantwoordelijken en de geïnteresseerde onderzoekers zullen het fenomeen, zoals het zich in België voordoet, beter leren begrijpen.

Met het oog op deze systematische datamining dient de huidige selectie van gegevens in het OASIS-datawarehouse geconfronteerd te worden met het “kadaster van onderzoeken” van de sociale inspectie. Op die manier kunnen bestaande alarmen en fraudescenario’s geoptimaliseerd worden voor de betrokken controlediensten. Daarenboven dient te worden nagegaan hoe op basis van het bestaande datawarehouse een interface kan worden ontworpen ten dienste van verschillende gebruikersgroepen, met name overheidsdiensten, politieke verantwoordelijken, wetenschappers en het geïnteresseerde publiek. Dankzij een gedifferentieerde toegang tot het OASIS-datawarehouse en de interface kunnen de verschillende gebruikersgroepen (sociale inspectie, federale overheidsdiensten, universiteiten) de onderzoeksresultaten, indicatoren en gegevens die voor hen relevant zijn raadplegen en gebruiken.

- 1.2. Het voorliggend onderzoeksproject richt zich uitsluitend op het optimaliseren van het opsporen van potentiële “frauderende” ondernemingen, waarbij gegevens op werknemersniveau worden geaggregeerd tot ondernemingsniveau, volgens de berekeningswijze van de alarmen die thans gebruikt worden in OASIS.

De gegevens nodig voor het uitvoeren van deze studie zijn terug te vinden in volgende fiscale en sociale databanken: het werkgeversrepertorium, de DMFA-persoonsgegevensbank, de DIMONA-persoonsgegevensbank, het bestand “werven en onderaannemers” (bestand van de werven, opdrachtgevers, hoofdaannemers en onderaannemers), het bestand CDIE (bevat informatie over eventuele schulden bij de RSZ of bij het betrokken fonds voor bestaanszekerheid), de bestanden “rekeningen” en “procedures” (het bestand “rekeningen” bevat het saldo van de bijdragen verschuldigd aan de RSZ en het bestand “procedures” bevat de lopende procedures bij de RSZ), het bestand met betrekking tot de tijdelijke werkloosheid (RVA), de bestanden van de BTW (deze bevatten gegevens met betrekking tot de schulden bij de BTW, de aangiften aan de BTW, de transacties tussen klanten en leveranciers en de signaletiek) en het GENESIS-bestand (het “kadaster van onderzoeken”).

Drie van deze gegevensbanken bevatten gegevens op werknemersniveau (natuurlijke personen), met name de DMFA-persoonsgegevensbank, de DIMONA-persoonsgegevensbank en het werklozenbestand van de RVA. De informatie zal door de leveranciers van de basisgegevens, dat wil zeggen door de beheerders van de betrokken persoonsgegevensbanken, worden geaggregeerd tot op het niveau van de werkgever/onderneming. Vervolgens zullen deze basisgegevens worden ingeladen in het datawarehouse arbeidsmarkt en sociale bescherming en op het niveau van de werkgever/ondernemer worden gekoppeld. Het identificatienummer van de

werkgever/ondernemer zal worden gecodeerd vooraleer de gegevens worden overgemaakt aan de onderzoeker.

Er dient te worden vermeld dat een aantal van de betrokken werkgevers/ondernemingen natuurlijke personen zijn.

1.3. De gegevens afkomstig uit het werkgeversrepertorium zijn:

- de registratie RSZ + controlenummer (gecodeerd);
- de administratieve taalrol;
- de taalrol van de werkgever;
- de datum van aansluiting (maand en jaar);
- de datum van schrapping van een werkgever bij de RSZ (maand en jaar);
- de code van de fusiegemeente (arrondissement);
- de datum van update (maand en jaar);
- het BTW-nummer (gecodeerd);
- de code "juridische vorm" van de onderneming;
- de afkorting van de juridische vorm of zijn beschrijving;
- het type werkgever;
- de datum van toekenning van de categorie (maand en jaar);
- de categorie RSZ - de economische activiteit van de werkgever (drie digits);
- de categorie RSZ - datum schrapping (maand en jaar);
- de categorie RSZ - de oorspronkelijke categorie (drie digits);
- de categorie RSZ - de nieuwe categorie (drie digits);
- de NACE-code (de economische activiteit) (vijf digits);
- de code van de plaats van de exploitatiezetel (arrondissement);
- de code regionalisatie (de aanduiding van aanwezigheid van filialen);
- het taaldecreet;
- de code belangrijkheid (groottecategorieën);
- de code leerling;
- het stamnummer van de behandelde registratie (gecodeerd);
- het stamnummer van bestemming (gecodeerd);
- het stamnummer van bestemming - datum van overdracht (maand en jaar);
- het stamnummer van bestemming - code overname (gecodeerd);
- het stamnummer van oorsprong (gecodeerd);
- het stamnummer van oorsprong - datum van overdracht (maand en jaar);
- het stamnummer van oorsprong - code overname (gecodeerd);
- het nummer van het sociaal secretariaat (gecodeerd);
- de code juridisch repertorium;
- de postcode van de werkgever (arrondissement);
- het al dan niet voorhanden zijn van een gevolmachtigde of curator.

1.4. De gegevens afkomstig uit de DMFA-persoonsgegevensbank zijn:

- het jaar en het kwartaal van de aangifte;
- het gecodeerd RSZ-inschrijvingsnummer;
- de werkgeverscategorie;

Variabelen berekend op ondernemingsniveau:

- het aantal deeltijdse arbeiders;
- het aantal aangegeven werknemers per categorie;
- het aantal bezoldigde dagen per categorie;
- het aantal bezoldigde dagen per categorie (enkel arbeiders);
- het aantal arbeiders met werknemerscode 10, 14, 15, 22, 26, 27, 35;
- het aantal gelijkgestelde dagen per type en per categorie;
- het loon van de arbeiders;
- het aantal arbeiders;
- het gemiddeld aantal uren per week van de werknemer;
- het aantal uren van de prestatie;
- het nummer van het paritair comité.

1.5. De variabelen op ondernemingsniveau, afkomstig uit de DIMONA-persoonsgegevensbank, zijn:

- het aantal INSZ (het aantal fysieke personen);
- het aantal inkomende werknemers;
- het aantal uitgaande werknemers.

1.6. De variabelen op ondernemingsniveau, afkomstig van de RVA, zijn:

- het aantal dagen tijdelijke werkloosheid voor type P01 (economische oorzaken);
- het aantal dagen tijdelijke werkloosheid voor type P02 (slecht weer);
- het aantal werknemers in tijdelijke werkloosheid per type.

1.7. De gegevens afkomstig van de BTW zijn, naast het gecodeerd BTW-nummer:

Gegevens met betrekking tot de BTW-aangifte:

- de inschrijvingsdatum (maand en jaar);
- de aard van de aangifte;
- het stelsel;
- de code vraag om terugbetaling;
- het bedrag van de aangifte;
- de periode (betrokken maanden en jaren);
- het nummer van het bevoegd kantoor (gecodeerd);
- de NACE-code (vijf digits);

Gegevens met betrekking tot de schulden bij de BTW:

- het bedrag van de BTW-schulden;
- het nummer RCIV (Regionale Centra voor Informatieverwerking);
- het ontvangstkantoor BTW (gecodeerd);

Gegevens met betrekking tot klanten en leveranciers:

- het BTW-nummer van de klant (gecodeerd);
- het BTW-nummer van de leverancier (gecodeerd);
- het bedrag;

Gegevens met betrekking tot de signaletiek:

- het statuut;
- de datum van gevolg belastingsregime (maand en jaar);
- de datum van beëindiging (maand en jaar);
- de aanduiding van faillissement;
- het bevoegd controlekantoor (gecodeerd).

1.8. De gegevens uit de bestanden “rekeningen” en “procedures” zijn:

- registratie RSZ + controlenummer (gecodeerd);
- het saldo van de rekening in eurocent;
- de code die het saldo van de rekening verheldert;
- het nummer van de procedure;
- de datum van sluiting van de procedure (maand en jaar).

1.9. De gegevens uit de bestanden “werven en onderaannemers” zijn:

Gegevens met betrekking tot de werven:

- het identificatienummer van de werf in 30bis + controlegetal (gecodeerd);
- de datum van handtekening van het contract (maand en jaar);
- de datum van het begin van de werken;
- de datum van het einde van de werken (maand en jaar);
- de datum van het einde van de werf;
- het kredietsaldo van de werf;
- het debetsaldo van de werf;
- het totaalbedrag voorzien voor de werf;

Gegevens met betrekking tot de opdrachtgever:

- het identificatienummer RSZ + controlegetal (gecodeerd);
- het BTW-nummer (gecodeerd);
- het registratienummer (gecodeerd);
- het registratienummer RSZ + controlegetal (gecodeerd);
- het KBO-nummer (gecodeerd);
- het land als de zetel zich in het buitenland bevindt;
- de datum van inschrijving als opdrachtgever (maand en jaar);

Gegevens met betrekking tot de hoofdaannemer:

- het identificatienummer RSZ + controlegetal (gecodeerd);
- het BTW-nummer (gecodeerd);
- het registratienummer (gecodeerd);
- het inschrijvingsnummer RSZ + controlegetal (gecodeerd);
- het KBO-nummer (gecodeerd);
- het land als de zetel zich in het buitenland bevindt;
- de datum van inschrijving als hoofdaannemer (maand en jaar);

Gegevens met betrekking tot de onderaannemer:

- het identificatienummer RSZ + controlegetal (gecodeerd);
- het BTW-nummer (gecodeerd);
- het inschrijvingsnummer (gecodeerd);
- het oud identificatienummer (gecodeerd);
- het nieuw identificatienummer (gecodeerd);
- het KBO-nummer (gecodeerd);
- het land als de zetel zich in het buitenland bevindt.

1.10. De gegevens uit het bestand CDIE zijn:

- het inschrijvingsnummer RSZ + controlegetal van de werkgever (gecodeerd);
- het BTW-nummer van de werkgever (gecodeerd);
- de aanduiding of de werkgever is geregistreerd of niet;
- de geldigheidsdatum van het attest C30bis wat betreft registratie;
- de aanduiding of er afhouding is of niet;
- de geldigheidsdatum van het attest C30bis wat betreft afhouding (maand en jaar);
- al dan niet in orde met verplichtingen t.o.v. RSZ;
- al dan niet in orde met verplichtingen t.o.v. het fonds van bestaanszekerheid;
- de datum waarop een beslissing “geforceerd” werd (maand en jaar).

1.11. De gegevens afkomstig van het “kadaster van onderzoeken” (GENESIS) zijn:

- het referentienummer van het onderzoek;
- de betrokken inspectiedienst;
- de reden van het onderzoek;
- de instantie die om het onderzoek heeft verzocht;
- de datum van opening van het onderzoek;
- de datum van sluiting van het onderzoek;
- het betrokken gewestelijk kantoor;
- het gecodeerd INSZ van de werkgever (natuurlijke persoon);
- het geboortjaar van de werkgever (natuurlijke persoon);
- het geslacht van de werkgever (natuurlijke persoon);
- de nationaliteit van de werkgever (natuurlijke persoon);
- de bepalingen waarop het onderzoek gegrond is;
- het resultaat van het onderzoek.

1.12. De gevraagde gegevens hebben betrekking op alle werkgevers in alle sectoren. Het betreft ongeveer 250.000 ondernemingen. Alle gegevens worden opgevraagd op basis van maandelijks of trimestriële cijfers met een maximum van vijf jaar, te tellen vanaf de meest recent beschikbare data (dit is ofwel de periode 2002-2006 ofwel de periode 2003-2007). De gegevens van de GENESIS-databank worden opgevraagd sinds de invoering van de databank “kadaster van onderzoeken” tot op de datum van oplevering van de data.

De Kruispuntbank van de Sociale Zekerheid staat in voor het verzamelen en het koppelen van de gegevens en het coderen van de identificatienummers van de ondernemingen vooraleer de gegevens over te maken aan de Université Libre de Bruxelles, de Katholieke Universiteit Leuven en de Université Catholique de Louvain.

- 1.13. De persoonsgegevens zouden worden bijgehouden tot 31 december 2010, waarna zij vernietigd zouden worden.

B. BEHANDELING VAN DE AANVRAAG

- 2.1. Krachtens artikel 5, § 1, van de wet van 15 januari 1990 *houdende oprichting en organisatie van een Kruispuntbank van de Sociale Zekerheid* verzamelt de Kruispuntbank van de Sociale Zekerheid persoonsgegevens bij de instellingen van sociale zekerheid, slaat ze op, voegt ze samen en deelt ze mee aan de personen die ze nodig hebben voor het verrichten van onderzoeken die nuttig zijn voor de kennis, de conceptie en het beheer van de sociale zekerheid.
- 2.2. De mee te delen gegevens hebben enkel betrekking op werkgevers en niet op geïdentificeerde of heridentificeerbare werknemers.

Vermits tussen de betrokken werkgevers zich ook natuurlijke personen bevinden, gaat het (gedeeltelijk) om een mededeling van persoonsgegevens die krachtens artikel 15, § 1, tweede lid, van de wet van 15 januari 1990 *houdende oprichting en organisatie van een Kruispuntbank van de Sociale Zekerheid* een principiële machtiging van de afdeling sociale zekerheid van het sectoraal comité van de sociale zekerheid en van de gezondheid vereist.

- 2.3. Het onderzoek beoogt enerzijds het optimaliseren van het project OASIS door er datamining in te integreren en anderzijds het beter afbakenen van het fenomeen sociale fraude. Het betreft een gerechtvaardigd doeleinde.

De mee te delen persoonsgegevens betreffen enkel werkgevers. Eventuele persoonsgegevens aangaande werknemers van de betrokken werkgevers zullen worden geaggregeerd op werkgeversniveau. In geen geval zal de mededeling aldus betrekking hebben op geïdentificeerde of heridentificeerbare werknemers.

Bij de mededeling van de persoonsgegevens wordt elk werkgeversidentificatienummer vervangen door een uniek betekenisloos volgnummer. De mee te delen persoonsgegevens kunnen slechts door middel van dit uniek betekenisloos volgnummer in verband worden gebracht met de werkgever (natuurlijke persoon dan wel rechtspersoon).

De eigenlijke persoonskenmerken blijven beperkt en worden doorgaans in klassen meegedeeld. Zij lijken niet van die aard dat ze een (her)identificatie van de betrokken werkgever mogelijk maken.

De gegevens lijken voorts, uitgaande van hogervermeld doeleinde, ter zake dienend en niet overmatig.

- 2.4. Het gaat om een latere verwerking van gecodeerde persoonsgegevens voor historische, statistische of wetenschappelijke doeleinden, waardoor de bepalingen van afdeling II van hoofdstuk II van het koninklijk besluit van 13 februari 2001 *ter uitvoering van de wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens* dienen te worden nageleefd.

De onderzoekers kunnen hogervermeld doeleinde niet verwezenlijken aan de hand van anonieme gegevens. Zij dienen immers in de mogelijkheid te worden gesteld om de situatie van de betrokkenen op te volgen.

- 2.5.** De Kruispuntbank van de Sociale Zekerheid mag de gecodeerde persoonsgegevens slechts meedelen nadat ze, overeenkomstig artikel 13 van het koninklijk besluit van 13 februari 2001 *ter uitvoering van de wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens*, in het bezit werd gesteld van het door de Commissie voor de Bescherming van de Persoonlijke Levenssfeer uitgereikte ontvangstbewijs van de door de onderzoekers verrichte aangifte van de verwerking voor historische, statistische of wetenschappelijke doeleinden.
- 2.6.** De onderzoekers dienen er zich contractueel toe te verbinden alle mogelijke middelen te zullen inzetten om te vermijden dat de identiteit van de personen op wie de meegedeelde persoonsgegevens betrekking hebben, zou worden achterhaald.

In elk geval is het, overeenkomstig artikel 6 van het koninklijk besluit van 13 februari 2001 *ter uitvoering van de wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens*, verboden om handelingen te stellen die ertoe strekken de meegedeelde gecodeerde persoonsgegevens om te zetten in niet-gecodeerde persoonsgegevens.

Het niet-naleven van dit verbod kan, krachtens artikel 39, 1°, van de wet van 8 december 1992, een strafrechtelijke veroordeling tot een geldboete van honderd tot honderdduizend euro tot gevolg hebben.

- 2.7.** Overeenkomstig artikel 23 van het koninklijk besluit van 13 februari 2001 *ter uitvoering van de wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens* mogen de resultaten van de verwerking voor historische, statistische of wetenschappelijke doeleinden niet worden bekendgemaakt in een vorm die de identificatie van de betrokken persoon mogelijk maakt tenzij deze laatste daartoe zijn toestemming heeft gegeven en de persoonlijke levenssfeer van derden niet wordt geschonden of tenzij de bekendmaking van niet-gecodeerde persoonsgegevens beperkt blijft tot persoonsgegevens die kennelijk door betrokkene zelf publiek zijn gemaakt of die in nauw verband staan met het publiek karakter van betrokkene of van de feiten waarbij deze laatste betrokken is of is geweest. Met voorbehoud van voormelde uitzonderingen, dienen de onderzoeksresultaten aldus op een anonieme wijze te worden gepubliceerd.
- 2.8.** De onderzoekers dienen bij de verwerking van de persoonsgegevens rekening te houden met de wet van 15 januari 1990 *houdende oprichting en organisatie van een Kruispuntbank van de Sociale Zekerheid*, de wet van 8 december 1992 *tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens*, hun uitvoeringsbesluiten en elke andere wettelijke of reglementaire bepaling tot bescherming van de persoonlijke levenssfeer.
- 2.9.** De persoonsgegevens mogen worden bijgehouden zolang zij noodzakelijk zijn voor het verwezenlijken van het onderzoek en ten langste tot 31 december 2010, waarna zij dienen te worden vernietigd.

Om deze redenen, verleent

het sectoraal comité van de sociale zekerheid en van de gezondheid, afdeling sociale zekerheid,

de machtiging aan de Kruispuntbank van de Sociale Zekerheid om de hogervermelde gecodeerde persoonsgegevens onder de hogervermelde voorwaarden mee te delen aan de Universit  Libre de Bruxelles, de Katholieke Universiteit Leuven en de Universit  Catholique de Louvain met het oog op het optimaliseren van het project OASIS door er datamining in te integreren en op het beter afbakenen van het fenomeen sociale fraude.

Yves ROGER
Voorzitter

De zetel van het Sectoraal Comit  van de Sociale Zekerheid en van de Gezondheid is gevestigd in de kantoren van de Kruispuntbank van de Sociale Zekerheid, op volgend adres : Sint-Pieterssteenweg 375 – 1040 Brussel (tel. 32-2-741 83 11)

