



Vrije Universiteit Brussel

Vakgroep Sociologie - Interface Demography

Vakgroep Maatschappelijke Gezondheidszorg

Vrije Universiteit Brussel

Pleinlaan 2

1050 Brussel

België

EXECUTIVE SUMMARY

ONTWIKKELING VAN EEN DATAWAREHOUSE DAT DE BELGISCHE GEGEVENS BEVAT INZAKE ARBEIDSOMSTANDIGHEDEN

LASTENBOEK HUT/DIRECT/2011/AP/3

Prof. Dr. Christophe Vanroelen

Prof. Dr. Patrick Deboosere

Prof. Dr. Pascale De Quint

Prof. Dr. Anne-Marie Depoorter

Anne Van Outryve

Karen Van Aerden

Juni 2013

Context en Doelstelling

Het begrip '**Kwaliteit van de arbeid**' kan analytisch-wetenschappelijk worden gedefinieerd en beschreven aan de hand van de 4 A-dimensies: Arbeidsomstandigheden – Arbeidsinhoud – Arbeidsvoorwaarden – Arbeidsverhoudingen. In de praktijk echter is hun integratie niet evident. Kwaliteit van de arbeid omvat een set van 'determinanten' in een risicoverhogende of beschermende relatie tot een set van mogelijke 'gevolgen', voor zowel de individuele werknemer als voor de organisatie.

De **arbeidsgebonden gezondheidsrisico's** waarover een database moet worden samengesteld, hebben raakvlakken met het individu (zijn/haar gezondheid en algemeen welzijn), de tewerkstellende organisatie en bij uitbreiding de bredere maatschappelijke context waarbinnen individu en organisatie functioneren. Ook de kenmerken van het individu en de organisatie zelf zijn medebepalend.

Voor de ondersteuning van een goede preventiepraktijk en voor de ontwikkeling van een beleidsinformerende monitoring is een goede **definitie en kennis van deze 'determinanten'**, hun spreiding over organisaties en beroepsgroepen en hun relaties met gevolgen in termen van welzijn en gezondheid een eerste absolute vereiste. Eens deze verworven zijn, is **het beschikken over relevante gegevens** betreffende de afzonderlijke determinanten van belang.

Heel wat bestaande data(banken), onderzoeken en registraties bevatten gegevens over arbeidsgebonden risico's en hun mogelijke gevolgen. Deze verschillende databronnen kunnen relevant zijn voor beleidsinformerende monitoring en kunnen ter ondersteuning voor de preventiepraktijk en de wetenschappelijke kennis worden gebruikt.

De **centrale doelstelling van dit project** is pistes aan te reiken die kunnen leiden tot een betere ontsluiting en integratie van de bestaande gegevens met betrekking tot 'Arbeidsomstandigheden – Kwaliteit van de Arbeid'. Meer **in detail zijn de specifieke objectieven** voor dit exploratief onderzoek de volgende:

1. het identificeren en beschrijven van relevante databanken
2. het identificeren van relevante actoren, met name personen en instellingen die in het bezit zijn van gegevens of over relevante expertise beschikken
3. het opzetten van een kader voor een zinvolle verbinding van de beschikbare gegevens in een datawarehouse en/of datamanagement systeem

4. het opzetten van een plan voor toekomstige gegevensanalyse in het kader van (1) kwaliteitsverificatie; (2) crossvalidering tussen databanken en (3) het ontwikkelen van basisindicatoren (benchmarking) en hun beschrijving.

Procedure en Methodologie

Met bovengenoemde, gedetailleerde doelstellingen als leidraad werd overgegaan tot het in fasen uitvoeren van de verschillende onderdelen van het onderzoek.

Een eerste fase leidde tot de **identificatie en oplisting van relevante databronnen** (zowel registratiegegevens als surveydata). Iedere opgelijste gegevensbron werd in eerste instantie gescreend op de algemene technisch-administratieve karakteristieken van die databank. Er werd tevens overgegaan tot het identificeren van de 'relevante' partners.

In een tweede fase werd ook de **inhoud van de databanken uitgebreid gescreend**. Dit gebeurde aan de hand van een eigen **vooraf geëxpliciteerd conceptueel model** voor ordening van de betrokken informatie. Hierbij werden drie hoofdcategorieën (in subcategorieën opgedeeld) gehanteerd:

1. determinanten van de kwaliteit van de arbeid of arbeidsgebonden risico's
2. (sociaal-demografische) achtergrondkenmerken van individu en organisatie
3. gevolgen voor het welzijn en de gezondheid van de individuele werknemer.

In een derde fase werd een **blauwdruk voor de structuur van het datawarehouse ontwikkeld**. Hierbij werd zoveel mogelijk uitgegaan van de mogelijkheden van de bestaande databanken en structuren en hun huidige links. De mogelijkheid tot het creëren van nieuwe links met nog niet verbonden databanken werd geëvalueerd.

Tijdens de oefening werd ook aandacht besteed aan de **huidige troeven, lacunes en knelpunten**, wat overeen komt met de vierde doelstelling van de studie.

Tenslotte werd in de vijfde fase een **plan ontwikkeld voor eerste analyses** op basis van de structuur voor het geïntegreerde datakader. Ook hier werd tijd genomen voor de identificatie van de mogelijke te verwachten knelpunten en kwaliteitsproblemen.

Overzicht van de onderzoeksresultaten

Een **lijst van relevante gegevensbronnen** werd opgemaakt en gevalideerd. In deze lijst werd ook een prioritaire orde aangelegd. Als criteria golden hierbij 'omvattendheid van de aanwezige informatie', 'ouderdom van de gegevens' (meest recent), 'wijze van gegevensverzameling' (registratie boven survey, wegens linkbaarheid van de data), 'aggregatieniveau van de informatie in de databank'.

Per weerhouden databank werd **een synthesefiche** opgesteld waarin de volgende zaken op gestructureerde wijze werden opgenomen: (1) metadata over de databank zelf, (2) variabelen die betrekking hebben op de determinanten van de kwaliteit van de arbeid (risicofactoren), (3) achtergrondkenmerken van individu en organisatie en (4) gevolgen voor het welzijn en de gezondheid van individuele werknemers. De werkwijze via synthesefiches liet toe zowel eigenschappen als inhoud van de databanken onderling vergelijkbaar weer te geven. De fiches werden aan de opdrachtgever ter beschikking gesteld.

Bepaalde stakeholders, organisaties en hun medewerkers hebben om uiteenlopende redenen belang bij het inrichten van een datawarehouse 'Arbeidsomstandigheden'. Specifiek werden de belanghebbende organisaties en instituten opgelijst die invloed ondervinden van of invloed (kunnen) uitoefenen op het ontsluiten van de beoogde informatie. Dit kan zijn op basis van hun expertise in het vakgebied (technisch of inhoudelijk), maar ook met betrekking tot het nut ervan als (potentiële) gebruiker in welke hoedanigheid dan ook. Zij zijn '**relevante partners**' voor deze oefening en de verdere uitwerking ervan.

We bleken uiteindelijk te worden geconfronteerd met een **erg complexe structuur van databanken**. Benadering of perspectief, wijze van opslag en niveau van abstractie verschillen vaak naargelang het actuele doel van de database. Het vrijwaren van de privacy en het voorkómen van identificeerbaarheid (onder meer in het kader van de concurrentiële positie van private actoren) komen hier nog als uitdagingen bovenop. De gevoeligheden rond de dubbele rol van de overheid – controlerend en beleidsmakend – vormen nog een bijkomende hindernis. Het is duidelijk dat bij het ontwerpen van een voorstel voor een gecentraliseerde datastructuur met dit erg complexe algemene kader dient rekening te worden gehouden.

Vanuit onderzoeksoogpunt was het absoluut relevant om een **geïntegreerde aanpak voor te stellen** en de verschillende informatiestromen op elkaar af te stemmen. Hierbij dient een juiste middenweg te worden gevonden tussen 'wetenschappelijk superieure' verbindingen op een zo laag mogelijk abstractieniveau en een meer 'operationeel realistische' benchmarking op een hoger abstractieniveau. Het is duidelijk dat enkel een integrerende aanpak informatiewinst oplevert. Ze vormt bovendien een aanzet tot het formuleren van voorstellen voor een betere afstemming door harmonisatie en door standaardisatie. De optimale strategie voor integratie wordt met andere woorden gedicteerd door de inhoud en de structuur van de bestaande databanken en door de opdrachten van de verschillende actoren die ze vandaag controleren.

Het was dus erg belangrijk te beschikken over een duidelijk overzicht van de bestaande databanken. Uitgaand van de uitgewerkte databankfiches werden **overzichtstabellen** opgesteld waarin voor elk van de elementen van het conceptueel model aangegeven werd of dit aspect al dan niet in de database werd opgenomen en zo ja, op welke wijze. Deze tabellen maken onmiddellijk duidelijk hoe de informatie - minstens theoretisch - kan worden samengebracht.

Vanuit haalbaarheidsoogpunt werd een voorstel geformuleerd om '**gelaagd**' tewerk te gaan bij de ontwikkeling van een Datawarehouse Arbeidsomstandigheden. Concreet onderzochten we de haalbaarheid van **twee pistes**, elk met een eigen nut en finaliteit, met name een 'benchmarkingspiste' en een 'individuele verbindingspiste'.

De '**benchmarkingspiste**': Het datawarehouse kan in dit geval worden gezien als een '**Kwaliteit van de arbeid-informatiesysteem**' met als doel het mogelijk maken van relevante uitspraken over het voorkomen van zowel 'indicatoren van de kwaliteit van de arbeid' als potentiële 'gevolgen' op het niveau van een aantal belangrijke 'identifiers' of bijzondere aandachtsgroepen. Dit gebeurt dan op basis van een vooraf opgemaakte inventaris van relevante gegevensbronnen en hun inhoud. Bij dit alles is het van belang de meest gepaste databank(en) voor ieder van de afzonderlijke "queries" te selecteren. Hierbij worden volgende criteria in acht genomen: (1) voorkeur aan registratiegegevens boven surveygegevens; (2) voorkeur aan databanken die de beste garantie geven op representatieve uitspraken voor een zo specifiek mogelijk niveau van aggregatie en (3) de leeftijd van de data.

De '**individuele verbindingspiste**': Hierbij wordt in de letterlijke betekenis van een datawarehouse **tot verbinding van gegevens overgegaan**. Idealiter wordt daarbij een identifier op het niveau van de individuele werknemer gebruikt om uit verschillende databronnen een geheel van onderling verbonden data op te bouwen. De bestaande datastructuur van de Kruispuntbank van de Sociale Zekerheid wordt in dat geval best als 'universum' of 'moederdatabank' aangewend. Op dit moment zijn binnen het universum van de KSZ al zeer relevante dataverbindingen mogelijk in verband met het thema van arbeidsomstandigheden. Er dient te worden gedacht aan uitbreiding van de huidige 'datawarehouse structuur' met andere relevante databanken die vandaag nog niet verbonden zijn met de KSZ. Het opzet van de **linkage op individueel niveau** is methodologisch superieur, daar daadwerkelijke verbanden tussen indicatoren kunnen worden onderzocht aan de hand van statistische analyses op microniveau. Het scenario kan vervolledigd worden door benchmarkinformatie vanuit surveys te verbinden met de registerinformatie. Hierbij moet wel worden opgemerkt dat dergelijke verbindingspistes **potentieel een aantal problemen oproepen**, onder andere in verband met privacy, administratieve belasting, eenvormigheid van gegevensopslag, etc. Het lijkt dan ook aangewezen dit scenario stap voor stap te verwezenlijken in nauw overleg met de databeheerders en in kleinere vervolgprojecten die toegespitst zijn op een specifieke onderzoeksvraag of dataverbinding.

Beide voorstellen (pistes) werden uitgebreid uitgewerkt, met inbegrip van voorbeelden ter verduidelijking. Voor beide pistes werd eveneens overgegaan tot het opmaken van **een aanzet voor een plan voor eerste analyses**.

Conclusie en Aanbevelingen

Het onderzoek naar de mogelijkheden tot de ontwikkeling van een Datwarehouse Arbeidsomstandigheden werd uitgevoerd aan de hand van vijf opeenvolgende fasen. In een eerste fase werden de relevante gegevensbronnen en mogelijke stakeholders geïdentificeerd. In een tweede fase volgde een uitgebreide screening van de inhoud van de verschillende databanken. Vervolgens werd een blauwdruk voor de structuur van het datawarehouse ontwikkeld. Een vierde fase omvatte de inventarisatie van knelpunten en troeven van het in het kader van dit project geformuleerde voorstel. Tenslotte werden suggesties voor verder onderzoek en eerste analyses vanuit het geïntegreerde denkkader gedaan.

Met betrekking tot de integratie van de beschikbare data worden verschillende pistes aangereikt. Deze pistes zijn mogelijk ook combineerbaar. Een eerste scenario bestaat uit het benchmarken van informatie. Hierbij kunnen gegevens uit registratiedata maar ook uit surveys worden aangewend om kengetallen te produceren op het niveau van sub-eenheden binnen de arbeidsmarkt of werknemerspopulatie: sociaal-demografische kenmerken, gegevens over de werksituatie, het soort tewerkstellende organisatie, enz. Een tweede scenario omvat een "rijkere" wijze van data-integratie door het maken van verbindingen op het niveau van de basiseenheden (individuele werknemers) of hogere eenheden (vestigingen of juridische eenheden van tewerkstellende organisaties). In het geval van deze laatste optie kan stapsgewijs vooruitgang worden gemaakt door de verbinding van bijkomende dataregisters bij de bestaande structuur van de KSZ.

Discussie

Er zijn nog heel wat lacunes aanwezig in de dataverzameling omtrent de kwaliteit van de arbeid in België. Zo is er bijvoorbeeld weinig informatie terug te vinden over die aspecten van de kwaliteit van de arbeid die zich op het niveau van de organisatie bevinden, onder andere rond kenmerken van het human resources beleid. Er bestaat hierover weinig survey informatie in België. De preventieve arbeidsgeneeskundige registratie gaat hier ook grotendeels aan voorbij. Koppeling van gegevens uit de Sociale Balans, geregistreerd door de Nationale Bank van België zou op dit vlak al een aanzet tot verbetering kunnen zijn.

Een knelpunt dat verband houdt met het vorige punt betreft het gebrek aan een vaste definitie voor de onderneming als observatie-eenheid bij het rapporteren over (aspecten van) kwaliteit van de arbeid. De veelheid aan gehanteerde registratieniveaus (werkpost, vestigingsplaats, juridische eenheid) maakt dat gegevens uit verschillende bronnen

soms op een verschillende populatie betrekking hebben en daardoor niet vergelijkbaar zijn (vb. juridische entiteiten, lokale vestigingen, exploitatiezetels,...). Hierrond zouden betere richtlijnen of protocollen ongetwijfeld een stap vooruit zijn: de integratie en verbinding van verschillende niveaus van gegevensopslag zal een globaal inzicht in de beroepsgebonden gezondheidsrisico's ten goede komen.

Ook blijkt uit de inventarisatie van verzamelde informatie dat de psychosociale risico's nog steeds onvoldoende gescreend en geregistreerd worden in de courante preventiepraktijk, hoewel er zowel in de wetgeving (cfr. Welzijnswet) als op het werkveld grote – en steeds meer – erkenning is voor het belang van dit soort arbeidsgebonden risico's en de mogelijke gevolgen ervan. Een systematische registratie van het arbeidsgebonden psychosociale risico tijdens het preventief onderzoek is een piste die verder onderzoek verdient. Onder meer in Canada werd op dit vlak een potentieel inspirerend initiatief ontplooid¹.

De voornaamste vragen die rijzen in verband met de ontwikkeling van een datawarehouse Arbeidsomstandigheden, hebben betrekking op de uniformisering van de gegevensverzameling. Dit is immers een noodzakelijke voorwaarde om tot de koppeling van gegevens te kunnen overgaan, maar het is tegelijk het vlak waarop er nog veel hindernissen te nemen zijn. Voor de overheid is hier een belangrijke, sturende rol weggelegd. Het lijkt hierbij aangewezen om stap voor stap tewerk te gaan. De datastructuur van de bestaande Kruispuntbank Sociale Zekerheid lijkt het logische universum om voor een dergelijke koppeling als basis te dienen. Hieraan kunnen – zoals in dit rapport werd uiteengezet – verschillende externe databanken verbonden worden. Dit kan op verschillende abstractieniveaus: de individuele werknemer en het bedrijf waarin hij/zij is tewerkgesteld.

Een bijzondere uitdaging betreft de "carrièrerelange" opvolging van beroepsgebonden gezondheidsrisico's. Vandaag de dag is de levenslange carrière bij één onderneming voor de meeste werknemers verleden tijd. Een behoorlijk aantal werknemers wisselt regelmatig van werkgever. Een gecoördineerde opvolging en registratie van risico's over de carrière is een hele uitdaging, doch erg nuttig met betrekking tot het in kaart brengen van de levenslange blootstelling – onder andere in het kader van de erkenning van eventuele beroepsziekten.

¹ <http://shop.csa.ca/en/canada/occupational-health-and-safety-management/canca-z1003-13bnq-9700-8032013/inv/z10032013/>

Tenslotte lijken alle stakeholders het erover eens dat moet worden toegezien op het onder controle houden van de financiële en de administratieve belasting voor de verschillende actoren bij ieder nieuw initiatief in het kader van gegevensintegratie. Een datawarehouse moet hier vooral efficiëntie introduceren op de langere termijn, maar vraagt op korte termijn ook inspanningen. Hier dient gestreefd te worden naar een situatie waarbij de kosten voor de verschillende actoren in evenwicht zijn met de baten van de ontwikkeling van een datawarehouse.