Vrije Universiteit Brussel

Groupe de recherche en Sociologie – Interface Démographie

Groupe de recherche en Santé publique

Vrije Universiteit Brussel

Boulevard de la Plaine 2

1050 Bruxelles

Belgique

RÉSUMÉ

DEVELOPPEMENT D'UNE DATAWAREHOUSE REPRENANT LES DONNEES BELGES EN MATIERE DE CONDITIONS DE VIE AU TRAVAIL

CAHIER DES CHARGES HUT/DIRACT/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

Juin 2013

Contexte et Objectif

La « **qualité du travail** » en tant que concept peut être définie d'un point de vue analytique et scientifique, et décrite à l'aide des 4 dimensions: (1) conditions de travail – (2) contenu du travail – (3) conditions d'emploi – (4) relations de travail. Dans la pratique, leur intégration n'est toutefois pas évidente. La qualité du travail comprend un ensemble de « déterminants » et de « conséquences » possibles dans un rapport d'accroissement des risques ou de protection, tant pour le travailleur individuel que pour l'organisation.

Les **risques de santé liés au travail** à propos desquels une base de données doit être constituée sont liés à l'individu (sa santé et son bien-être général), à l'organisation et par extension au contexte social plus large dans lequel l'individu et l'organisation fonctionnent. Les caractéristiques de l'individu et de l'organisation même jouent aussi un rôle.

La promotion de bonnes pratiques de prévention et le monitoring d'information des pouvoirs publics nécessitent d'abord et impérativement **une bonne définition et une bonne connaissance de ces « déterminants »**, de leur répartition dans l'organisation et les catégories professionnelles et de leur rapport aux conséquences en matière de bien-être et de santé. Une fois cette connaissance acquise, il est important de **disposer de données pertinentes** sur les différents déterminants.

Beaucoup de (bases de) données, d'études et d'enregistrements existants contiennent des données sur les risques liés au travail et leurs conséquences possibles. Ces données peuvent être importantes pour le monitoring d'information des pouvoirs publics et/ou peuvent servir à soutenir les pratiques de prévention et la connaissance scientifique.

L'objectif central de ce projet est de donner des pistes pour améliorer l'ouverture à la recherche des informations existantes sur les « Conditions de vie au travail – Qualité du travail » et favoriser leur intégration. Les **objectifs spécifiques** de cette recherche exploratoire sont plus **précisément** les suivants:

1. l'identification et la description de bases de données utiles;
2. l'identification d'acteurs importants, à savoir des personnes et des institutions qui sont en possession de données ou disposent d'une expertise pertinente;
3. la création d'un cadre pour une liaison utile des données disponibles dans une datawarehouse et/ou un système de gestion de données;

4. l'établissement d'un plan pour l'analyse future de données dans le cadre (1) d'une vérification de la qualité, (2) d'une validation croisée entre bases de données et (3) du développement d'indicateurs de base (référenciation) et de leur description.

Procédure et Méthodologie

Compte tenu des objectifs détaillés ci-dessus en tant que fil conducteur, nous avons procédé à l'exécution par phases des différentes parties de la recherche.

Une première phase a abouti à une **identification et à l'établissement d'une liste des sources de données utiles**, aussi bien de données d'enregistrement que de données d'enquête. Nous avons d'abord examiné les caractéristiques technico-administratives générales de chacune des sources de données listées. Nous avons également procédé à l'identification des partenaires « utiles ».

Dans une deuxième phase, nous avons **largement analysé le contenu des bases de données**. Cette analyse a été effectuée à l'aide d'un **modèle conceptuel préalablement explicité** de classement des informations en question. Trois catégories principales ont été utilisées ici:

1. les déterminants de la qualité du travail ou les risques,
2. les caractéristiques (socio-démographiques) générales de l'individu et de l'organisation,
3. les conséquences pour le bien-être et la santé du travailleur individuel.

Dans une troisième phase, nous avons **élaboré un projet de structure pour la datawarehouse**. Nous sommes autant que possible partis des possibilités des bases de données et des structures existantes avec leurs liens actuels. Nous avons évalué la possibilité de créer de nouveaux liens avec des bases de données non encore reliées.

Durant l'exercice, nous nous sommes également intéressés aux **atouts, lacunes et points noirs actuels**, ce qui correspond au quatrième objectif de l'étude.

Nous avons enfin **élaboré** dans une cinquième phase un **plan de première analyse** sur la base de la structure développée pour le réseau intégré de données. Nous avons ici aussi consacré du temps à l'identification des difficultés possibles et des problèmes de qualité à prévoir.

Aperçu des résultats de recherche

Une **liste de sources de données pertinentes** a été établie, validée et hiérarchisée. Ont tenu lieu de critères: « l'exhaustivité de l'information », « l'actualité des données » (les plus récentes), « le mode de collecte des données » (plutôt données d'enregistrement que données d'enquête, eu égard à la possibilité de liaison des données), « le niveau d'agrégation des informations stockées dans la base de données ».

Par base de données retenue, nous avons établi **une fiche de synthèse** qui reprend de manière structurée: (1) les métadonnées sur la base de données même, (2) les variables qui ont trait aux déterminants de la qualité du travail (facteurs de risque), (3) les indicateurs généraux de nature socio-démographique et organisationnelle et (4) les conséquences pour le bien-être et la santé du travailleur individuel. La méthode de travail par fiche de synthèse a permis de présenter les caractéristiques et le contenu des bases de données sur une base comparable. Les fiches ont été mises à la disposition du donneur d'ordre.

Certaines parties prenantes et organisations de même que certains instituts et leurs collaborateurs ont intérêt, pour diverses raisons, à aménager une datawarehouse « Conditions de vie au travail ». Nous avons plus particulièrement dressé la liste des organisations et des instituts intéressés qui subissent l'impact de ou exercent une influence sur l'ouverture à la recherche des informations visées. En fonction de leur expertise dans la branche (technique ou spécifique), mais aussi de leur utilité en tant qu'utilisateur (potentiel) en quelque qualité que ce soit. Ce sont des « **partenaires utiles** » pour cet exercice et son développement ultérieur.

Nous avons finalement été confrontés à une **structure de bases de données très complexe**. L'approche ou la perspective, le mode de conservation et le niveau d'abstraction diffèrent souvent selon le but effectif de la base de données. La préservation de la vie privée et la prévention de l'identifiabilité, notamment dans le cadre de la position concurrentielle d'acteurs privés, sont d'autres défis à relever. Les sensibilités autour du double rôle des autorités – de contrôle et de responsabilité politique – constituent un obstacle supplémentaire. Il est clair que dans l'ébauche d'une proposition de structure de données centralisée, il faut tenir compte de ce cadre général très complexe.

Du point de vue de la recherche, il était tout à fait approprié de **proposer une approche intégrée** et d'accorder les différents flux d'information. Il faut trouver un juste milieu entre les liaisons « scientifiquement supérieures » à un niveau d'abstraction aussi bas que possible et une référenciation plus « réaliste sur le plan opérationnel » à un niveau d'abstraction plus élevé. Il est clair que seule une approche intégrée offre un gain d'information. Elle constitue aussi un premier pas vers la formulation de propositions pour une meilleure coordination par l'harmonisation et la standardisation. La stratégie d'intégration optimale est en d'autres termes dictée par le contenu et la structure des bases de données existantes et par les missions des différents acteurs qui les contrôlent aujourd'hui.

Il était donc extrêmement important de disposer d'un aperçu clair et précis des bases de données existantes. Des **tableaux récapitulatifs** ont été établis à partir des fiches de bases de données individuelles et ces tableaux indiquent pour chacun des éléments du modèle conceptuel si cet aspect a été repris ou non dans la base de données et si oui, de quelle façon. Le tableau montre très clairement comment l'information – du moins théorique – peut être rassemblée.

Du point de vue de la faisabilité, nous avons choisi de procéder pour la **formulation d'une proposition** sur le développement d'une base de données Conditions de vie au travail d'une manière « **hiérarchisée** ». Nous avons étudié la faisabilité de **deux pistes**, chacune avec sa propre utilité et sa finalité, à savoir une « piste de référenciation » et une « piste de liaison de données individuelle ».

La **« piste de référenciation »**: la datawarehouse se limite dans ce cas-ci à l'option d'un « **système d'information sur la qualité du travail** », dont le but est de permettre des jugements pertinents sur l'existence d'« indicateurs de la qualité du travail » et de « conséquences » potentielles au niveau d'un certain nombre « d' identifiants » importants ou de groupes cibles particuliers. Cela se fait alors sur la base d'un inventaire préalablement établi de sources de données pertinentes et de leur contenu. Dans tout cela, il est important de sélectionner la ou les bases de données les plus appropriées pour chacune des « requêtes ».

Dans le choix de la priorité de reprise des sources de données, les critères suivants ont été pris en considération: (1) préférence donnée à des données d'enregistrement plutôt qu'à des données d'enquête, (2) préférence donnée à des bases de données qui offrent la meilleure garantie d'opinions représentatives à un niveau d'agrégation aussi spécifique que possible et (3) l'âge des données.

La « **piste de liaison de données individuelle** »: on **procède** ici **à la liaison de données** au sens littéral d'une datawarehouse. On utilise idéalement un identifiant au niveau du travailleur individuel pour constituer un ensemble de données reliées à partir de différentes sources de données. La structure de données existante de la Banque-Carrefour de la Sécurité Sociale (BCSS) sera de préférence utilisée dans ce cas comme « univers » ou « base de données mère ». Dans l'univers de la BCSS, des liaisons de données très utiles sont déjà possibles à l'heure actuelle pour ce thème. Il faut penser à un élargissement de la « structure de la datawarehouse» actuelle avec d'autres bases de données pertinentes qui ne sont pas encore reliées aujourd'hui à la BCSS. Le but de la **liaison à un niveau individuel** est méthodologiquement supérieur, puisqu'elle permet d'étudier les liens effectifs entre indicateurs à l'aide d'analyses statistiques au microniveau. Le scénario peut être complété en reliant une information de référenciation provenant d'enquêtes à l'information d'enregistrement.

Il faut cependant signaler ici que ces pistes de liaison de données **soulèvent potentiellement plusieurs problèmes**, notamment en matière de vie privée, de charge administrative, d'uniformité du stockage des données, etc. Il semble approprié de réaliser ce scénario étape par étape, en concertation étroite avec les gestionnaires de données et en développant des projets complémentaires axés sur une question d'enquête ou une liaison de données spécifique.

Les deux propositions (pistes) ont été largement développées et clarifiées avec des exemples. Pour les deux pistes, nous avons également fait l'**ébauche d'un plan d'analyse**.

Conclusions et Recommandations

L'enquête a été effectuée en cinq phases successives. Dans une première phase, nous avons identifié les sources de données pertinentes et dressé une liste de toutes les parties prenantes possibles. Dans une deuxième phase, nous avons procédé à un examen approfondi du contenu des différentes bases de données. Nous avons ensuite élaboré un projet de structure pour la datawarehouse. La quatrième phase a été celle de l'inventaire des problématiques et des atouts de la proposition formulée dans le cadre de ce projet. Nous avons enfin fait des suggestions d'autre recherche et de premières analyses à partir d'un cadre de réflexion intégré.

En ce qui concerne l'intégration des données disponibles, différentes pistes sont indiquées. Ces pistes sont éventuellement combinables. Un premier scénario consiste en la référenciation de l'information. On peut utiliser ici des données d'enregistrement mais aussi des données d'enquête pour produire des indices chiffrés au niveau de sous-unités dans le marché du travail ou la population active: caractéristiques démographiques, données sur la situation d'occupation, le type d'organisation employeuse, etc. Un deuxième scénario comprend un stockage de données « plus riche » du fait de liaisons au niveau des unités de base (travailleurs individuels) ou d'unités supérieures (établissements ou unités légales d'organisations employeuses). Dans cette dernière option, on peut progresser par étapes en reliant des registres de données supplémentaires à la structure existante de la Banque-Carrefour de la Sécurité Sociale.

Discussion

Il subsiste beaucoup de lacunes dans la collecte de données sur la qualité du travail en Belgique. On retrouve par exemple peu d'informations sur les aspects de la qualité du travail qui se situent au niveau de l'organisation, notamment concernant les caractéristiques de la politique des ressources humaines. Il existe peu de données d'enquête sur ce thème en Belgique. Les enregistrements sur la médecine du travail préventive n'y font généralement pas non plus référence. La liaison de données provenant des bilans sociaux enregistrés par la Banque Nationale de Belgique pourrait déjà être une amorce d'amélioration sur ce plan.

Une problématique liée au point précédent concerne l'absence d'une définition bien établie pour « l'entreprise » en tant qu'unité d'observation dans les rapports sur (des aspects partiels) de la qualité du travail. La multitude de niveaux d'enregistrement

utilisés (poste de travail, lieu d'implantation, unité légale) fait que les données provenant de différentes sources ont parfois trait à une population différente et ne sont de ce fait pas comparables (par ex. entités juridiques, établissements locaux, sièges d'exploitation...). De meilleures directives ou de meilleurs protocoles seraient sans aucun doute un pas en avant en la matière: l'intégration et la liaison de différents niveaux de stockage de données contribueront à une vision globale des risques de santé liés au travail.

Il ressort également de l'inventaire des informations collectées que les risques psychosociaux sont toujours insuffisamment évalués et enregistrés dans les pratiques de prévention courantes. Et ce, malgré une – toujours plus – large reconnaissance de l'importance de ce type de risques liés au travail et de leurs conséquences possibles, aussi bien dans la législation (cf. Loi sur le bien-être au travail) que sur le terrain. Un enregistrement systématique du risque psychosocial lié au travail durant l'examen préventif est une piste qui mérite d'être approfondie. Une initiative potentiellement inspirante a notamment été développée au Canada sur ce plan[1].

Les principales questions qui se posent par rapport au développement d'une datawarehouse sur les Conditions de vie au travail ont trait à l'uniformisation de la collecte des données. C'est en effet une condition nécessaire pour pouvoir relier des données, mais c'est en même temps le plan sur lequel il y a encore beaucoup d'obstacles à surmonter. Les pouvoirs publics ont un rôle actif important à jouer ici. Il semble également indiqué de travailler étape par étape. La structure de données de la Banque-Carrefour de la Sécurité Sociale semble être l'univers logique pour servir de base. Différentes bases de données externes peuvent y être reliées – comme exposé dans ce rapport. Et cela peut se faire à différents niveaux d'abstraction: le travailleur individuel et l'entreprise où il/elle est employé(e).

Un défi particulier concerne le suivi « tout au long de la carrière » des risques de santé liés au travail. Au jour d'aujourd'hui, la carrière à vie dans une seule entreprise appartient au passé pour la plupart des travailleurs. Beaucoup de travailleurs changent régulièrement d'employeur. Un suivi et un enregistrement coordonnés du risque professionnel encouru tout au long de la carrière sont un réel défi, qui est toutefois très utile pour brosser un tableau de l'exposition perpétuelle – notamment dans le cadre de la reconnaissance de maladies professionnelles éventuelles.

[1] http://shop.csa.ca/en/canada/occupational-health-and-safety-management/cancsa-z1003-13bnq-9700-8032013/invt/z10032013/

Toutes les parties prenantes semblent enfin d'accord sur le fait qu'il faut veiller à garder le contrôle de la charge financière et administrative pour les différents acteurs à chaque nouvelle initiative lancée dans le cadre de l'intégration de données. Une datawarehouse doit surtout introduire le rendement à plus long terme, mais demande aussi des efforts à court terme. Il faut tendre vers une situation où les coûts pour les différents acteurs sont compensés par l'utilité du développement d'une datawarehouse.