

Comité de sécurité de l'information Chambre sécurité sociale et santé
--

CSI/CSSS/24/404

DÉLIBÉRATION N° 24/196 DU 5 NOVEMBRE 2024 RELATIVE À LA COMMUNICATION DE DONNÉES À CARACTÈRE PERSONNEL PSEUDONYMISÉES DU DATAWAREHOUSE MARCHÉ DU TRAVAIL ET PROTECTION SOCIALE PAR LA BANQUE CARREFOUR DE LA SÉCURITÉ SOCIALE À L'UNIVERSITAT POMPEU FABRA DANS LE CADRE D'UNE ÉTUDE SUR LES EFFETS INTERGÉNÉRATIONNELS DU CONGÉ DE PATERNITÉ (PROJET « LEAVE4NEXTGEN »)

Vu la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, en particulier les articles 5 et 15 ;

Vu la demande de l'*Universitat Pompeu Fabra* ;

Vu le rapport de la Banque Carrefour de la sécurité sociale ;

Vu le rapport du président.

A. OBJET

1. L'*Universitat Pompeu Fabra*, basée à Barcelone, réalise actuellement une étude sur les effets intergénérationnels du congé de paternité (dans le cadre du projet « *Leave4NextGen* ») et elle souhaite à cet effet avoir recours à des données à caractère personnel pseudonymisées issues du datawarehouse marché du travail et protection sociale, géré par la Banque Carrefour de la sécurité sociale, afin d'analyser les effets de l'introduction du congé de paternité sur la prochaine génération. L'étude propose d'utiliser des données administratives pour observer si le congé de paternité a un impact sur le long terme sur les décisions en matière d'éducation, de choix de carrière et de fécondité (probabilité d'avoir des enfants à un jeune âge) des jeunes adultes dont le père a bénéficié du congé de paternité. L'objectif poursuivi vise à faire avancer la connaissance sur ce sujet en contribuant à la publication d'articles dans des revues scientifiques.
2. Alors que la Directive (UE) 2019/1158 du Parlement européen et du Conseil du 20 juin 2019 *concernant l'équilibre entre vie professionnelle et vie privée des parents et des aidants et abrogeant la directive 2010/18/UE du Conseil* a incité de nombreux Etats membres de l'Union européenne à adopter des politiques en matière de congé de paternité, la Belgique figure parmi les pionniers de ce mouvement, en ayant introduit un congé de paternité de deux semaines par la loi du 10 août 2001 *relative à la conciliation entre l'emploi et la qualité de*

vie, entrée en vigueur le 1^{er} juillet 2002, ce qui offre un cadre attrayant à la conduite d'une étude relative aux effets intergénérationnels de ce congé.

3. Afin d'analyser les effets de l'introduction du congé de paternité en juillet 2002 sur la prochaine génération, trois échantillons de données à caractère personnel pseudonymisées relatives à trois groupes cibles déterminés seront communiqués.
 - Le premier échantillon concerne les personnes nées entre les années 1999 et 2002¹ (ci-après les personnes de référence) qui sont potentiellement directement impactées par la prise d'un congé de paternité par leur père ;
 - Le second échantillon concerne les partenaires des personnes de référence et vise à observer des effets dits « d'accouplement assortatif », c'est-à-dire des phénomènes selon lesquels des individus qui présentent des caractéristiques similaires (par exemple, le même niveau d'étude) ont tendance à se mettre plus souvent en couple ;
 - Le troisième échantillon concerne les parents des personnes de référence et permet d'effectuer des analyses d'effets hétérogènes selon les caractéristiques des parents (différence de revenu ou d'âge).

La stratégie d'identification de l'effet causal repose sur une comparaison entre les personnes nées avant et après la mise en place du congé de paternité. Cette méthode repose sur l'hypothèse que les personnes nées autour de la réforme sont similaires en tout point, excepté la possibilité que leur père avait de prendre un congé de paternité. Plus les personnes sont nées à une date proche de la réforme et plus il est facile d'assumer que la seule différence entre elles est le fait que leur père ait eu accès à un congé de paternité à partir du 1^{er} juillet 2002. Il est également primordial d'observer les données des personnes nées durant les années qui précèdent la réforme. La raison est que les personnes nées à différents moments de l'année ont des caractéristiques différentes, il convient ainsi d'estimer ces différences durant des années sans réforme et retirer cette « saisonnalité ».

4. En ce qui concerne le premier échantillon (relatif aux personnes de référence), les chercheurs traiteraient par intéressé, les données à caractère personnel pseudonymisées suivantes du datawarehouse marché du travail et protection sociale (généralement la situation par trimestre ou à une date de référence déterminée).

Caractéristiques personnelles (situation au 31 décembre 2023) : le numéro d'ordre unique, l'année et le mois de naissance, le sexe, l'état civil, la cohabitation légale, le nombre d'enfants et l'année de naissance de chaque enfant.

Niveau et domaine d'éducation : le niveau d'études, le domaine d'études, l'année de remise du titre, le titre d'études secondaires supérieures, l'année de délivrance du titre d'études secondaires supérieures, l'année académique, la catégorie d'études, la catégorie d'études (enseignement non-universitaire), le domaine d'études et le diplôme décerné.

¹ D'après les données de Statbel, 454.779 personnes sont nées en Belgique entre 1999 et 2002, dont 111.484 sont nées en 2002.

Position socio-économique et revenus (au dernier jour du quatrième trimestre 2023) : la position socio-économique de la personne au dernier jour du trimestre, le statut du travailleur, le régime de travail, l'équivalent temps plein (ETP) à l'exclusion des journées assimilées, l'ordre d'importance des prestations de travail (pour les personnes qui cumulent plusieurs emplois), le code NACE employeur (2 chiffres), la rémunération ordinaire (en classes), la rémunération imposable brute ONSS et ONSSAPL (en classes), la rémunération brute ONSS et ONSSAPL (en classes) et le revenu INASTI (en classes).

Congé de paternité (de juillet 2002 à décembre 2023) : le congé de paternité, le nombre de jours, les années et mois de début et de fin.

5. En ce qui concerne le second échantillon (relatif aux partenaires des personnes de référence), les chercheurs traiteraient par intéressé, les données à caractère personnel pseudonymisées suivantes du datawarehouse marché du travail et protection sociale (généralement la situation par trimestre ou à une date de référence déterminée).

Caractéristiques personnelles (au 31 décembre 2023) : le numéro d'ordre unique, le sexe et l'année de naissance des partenaires.

Niveau et domaine d'éducation : le niveau d'études, le domaine d'études, l'année de remise du titre, le titre d'études secondaires supérieures, l'année de délivrance du titre d'études secondaires supérieures, l'année académique, la catégorie d'études, la catégorie d'études (enseignement non-universitaire), le domaine d'études et le diplôme décerné.

Position socio-économique et revenus (au dernier jour du quatrième trimestre 2023) : la position socio-économique de la personne au dernier jour du trimestre, le statut du travailleur, le régime de travail, l'équivalent temps plein (ETP) à l'exclusion des journées assimilées, l'ordre d'importance des prestations de travail (pour les personnes qui cumulent plusieurs emplois), le code NACE employeur (2 chiffres), la rémunération ordinaire (en classes), la rémunération imposable brute ONSS et ONSSAPL (en classes), la rémunération brute ONSS et ONSSAPL (en classes) et le revenu INASTI (en classes).

Congé de paternité (de juillet 2002 à décembre 2023) : le congé de paternité, le nombre de jours, les années et mois de début et de fin.

6. En ce qui concerne le troisième échantillon (relatif aux parents des personnes de référence), les chercheurs traiteraient par intéressé, les données à caractère personnel pseudonymisées suivantes du datawarehouse marché du travail et protection sociale (généralement la situation par trimestre ou à une date de référence déterminée).

Caractéristiques personnelles (au 31 décembre 2023) : le numéro d'ordre unique, le sexe et l'année de naissance de chaque parent.

Niveau et domaine d'éducation : le domaine d'étude, le niveau d'instruction et le niveau d'instruction détaillé.

Position socio-économique et revenus (au dernier jour du quatrième trimestre 2023) : la position socio-économique de la personne au dernier jour du trimestre, le statut du travailleur, le régime de travail, l'équivalent temps plein (ETP) à l'exclusion des journées assimilées, l'ordre d'importance des prestations de travail (pour les personnes qui cumulent plusieurs emplois), le code NACE employeur (2 chiffres), la rémunération ordinaire (en classes), la rémunération imposable brute ONSS et ONSSAPL (en classes), la rémunération brute ONSS et ONSSAPL (en classes) et le revenu INASTI (en classes).

Congé de paternité (de juillet 2002 à décembre 2023) : le congé de paternité, le nombre de jours, les années et mois de début et de fin.

7. L'étude serait réalisée en deux phases. Dans une première phase, la Banque Carrefour de la sécurité sociale communiquerait aux chercheurs des données à caractère personnel pseudonymisées pour environ 23.000 personnes, ainsi que les partenaires et parents (*un échantillon*), représentant cinq pour cent de la population cible, afin qu'ils puissent développer leurs algorithmes à l'aide d'applications spécifiques. Dans une deuxième phase, les chercheurs enverraient leurs codes à la Banque Carrefour de la sécurité sociale afin de les exécuter sur la population cible et auraient ainsi accès aux mêmes types de données à caractère personnel pseudonymisées de l'ensemble de la population (*la totalité de la population*), et ce sur un ordinateur sécurisé dans les locaux de la Banque Carrefour de la sécurité sociale et sous la surveillance permanente d'un collaborateur de cette organisation, pour y appliquer les algorithmes qu'ils ont développés. Les chercheurs pourraient emporter les résultats de leurs analyses, uniquement sous forme de données purement anonymes en dehors du bâtiment de la Banque Carrefour de la sécurité sociale.
8. Les chercheurs conserveraient, en tant que responsables du traitement, les données à caractère personnel pseudonymisées reçues au cours de la première phase jusqu'à la fin du projet « *Leave4NextGen* », c'est-à-dire jusqu'au 31 décembre 2027. Les données à caractère personnel pseudonymisées reçues seront donc détruites au plus tard le 31 décembre 2027. Les résultats de l'étude seront uniquement publiés sous forme anonyme.

B. EXAMEN DE LA DEMANDE

Compétence du Comité de sécurité de l'information

9. En vertu de l'article 5 de la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, la Banque Carrefour de la sécurité sociale recueille des données à caractère personnel auprès des institutions de sécurité sociale, les enregistre, procède à leur agrégation et les communique aux personnes qui en ont besoin pour la réalisation de recherches pouvant être utiles à la connaissance, à la conception et à la gestion de la protection sociale.
10. En vertu de l'article 15 de la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, toute communication de données à caractère personnel par la Banque Carrefour de la sécurité sociale ou une autre institution de sécurité

sociale doit faire l'objet une délibération préalable de la chambre sécurité sociale et santé du Comité de sécurité de l'information².

Licéité du traitement

11. En vertu de l'article 6 du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE*, le traitement n'est licite que si, et dans la mesure où, au moins une des conditions mentionnées à cet article est remplie.
12. La communication des données à caractère personnel pseudonymisées précitées du datawarehouse marché du travail et protection sociale par la Banque Carrefour de la sécurité sociale à l'*Universitat Pompeu Fabra*, pour l'exécution d'une étude sur les effets intergénérationnels du congé de paternité, est licite au sens de l'article 6, 1, alinéa 1^{er}, e), puisqu'elle nécessaire à l'accomplissement d'une tâche d'intérêt public.

Principes relatifs au traitement des données à caractère personnel

13. En vertu du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE*, les données à caractère personnel doivent être collectées pour des finalités déterminées, explicites et légitimes et elles ne peuvent pas être traitées ultérieurement d'une manière incompatible avec ces finalités (limitation des finalités), elles doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées (minimisation des données), elles doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées (limitation de la conservation) et elles doivent être traitées de façon à garantir une sécurité appropriée des données à caractère personnel, y compris la protection contre le traitement non autorisé ou illicite et contre la perte, la destruction ou les dégâts d'origine accidentelle, à l'aide de mesures techniques ou organisationnelles appropriées (intégrité et confidentialité).

Limitation de la finalité

14. La communication de données à caractère personnel pseudonymisées poursuit une finalité déterminée, explicite et légitime, à savoir l'analyse des effets du congé de paternité entre les générations, dans le cadre du projet « *Leave4NextGen* ». Des évaluations récentes montrent que les réformes du congé de paternité ont permis d'accroître la participation des pères à la garde des enfants et ont contribué à modifier les normes en matière de genre. Les chercheurs de l'*Universitat Pompeu Fabra* souhaitent déterminer si cette évolution des normes de genre

² Le Comité de sécurité de l'information constate que les chercheurs souhaitent non seulement obtenir des données à caractère personnel pseudonymisées, mais également des données anonymes (tableaux). Dans la mesure où cette communication est effectuée selon les dispositions de la délibération n° 18/140 du 6 novembre 2018, le Comité de sécurité de l'information ne doit pas se prononcer à cet égard.

peut être transmise à la génération suivante et influencer sur les décisions en matière d'éducation, de choix de carrière et de fécondité des enfants nés après la mise en place du congé paternité.

Minimisation des données

15. Les chercheurs ne sont pas en mesure de réaliser la finalité au moyen de données anonymes, étant donné qu'ils doivent pouvoir suivre la situation de personnes individuelles pendant un certain temps. Ils s'engagent contractuellement à mettre en œuvre tous les moyens possibles pour éviter une identification des personnes concernées. En toute hypothèse, il leur est interdit d'entreprendre toute action visant à convertir les données à caractère personnel pseudonymisées en données à caractère personnel non pseudonymisées.
16. Dans la première phase de l'étude, les données à caractère personnel portent sur environ 23.000 personnes, ainsi que les partenaires et parents (représentant cinq pour cent de la population cible). Le numéro d'identification de la sécurité sociale de chaque intéressé est remplacé par un numéro d'ordre unique sans signification. Les caractéristiques personnelles proprement dites sont limitées aux données strictement nécessaires pour l'étude et sont réparties en classes (regroupement des personnes par mois, des revenus par tranches et des secteurs d'activités par groupes).
17. Les données relatives aux caractéristiques personnelles des intéressés, en particulier les numéros d'ordre unique, permettent d'identifier les mêmes personnes dans les différentes bases de données et ainsi d'effectuer des analyses croisées. De plus, les années et mois de naissance permettent d'identifier les personnes nées avant (groupe de contrôle) et après la réforme de 2002 (groupe de traitement), ainsi que d'estimer les effets saisonniers liés au mois de naissance. Le sexe permet d'observer des effets différents entre les hommes et les femmes, tel que suggéré par de nombreuses études précédentes. L'état civil permet quant à lui d'identifier si les personnes nées après la réforme se sont davantage mises en couple et le nombre d'enfants permet d'identifier si les personnes nées après la réforme ont déjà eu des enfants.
18. Les données relatives aux niveau et domaine d'éducation des personnes de référence et leurs partenaires sont nécessaires premièrement, pour observer le plus haut niveau d'étude atteint par les personnes nées après la réforme et deuxièmement, pour observer si les femmes nées après l'introduction du congé paternité ont une probabilité plus élevée d'étudier dans un domaine STIM (science, technologie, ingénierie et mathématiques). Les données relatives aux niveau et domaine d'éducation des parents de la personne de référence permettent d'effectuer des analyses d'hétérogénéité selon le niveau d'étude des parents et le domaine. Cela permet, par exemple, de déterminer si une fille née après la réforme est davantage susceptible d'étudier dans un domaine STIM (science, technologie, ingénierie et mathématiques) lorsque son père avait également étudié dans ce domaine.
19. Les données relatives à la position socio-économique et les revenus sont nécessaires pour déterminer si les intéressés travaillent, sont au chômage ou en incapacité de travail, ainsi que les différences en termes de temps de travail et de revenus. Ces données permettent également d'observer si le fait de potentiellement étudier dans des domaines différents amène les hommes et femmes concernés à des statuts différents, d'observer si le volume de travail est

similaire, et de mesurer la différence de revenus entre la personne de référence et ses partenaires. Cette différence sera également calculée pour les parents afin de mesurer des effets hétérogènes. Ainsi, les données relatives au revenu, au volume de travail et au secteur d'activités, permettraient d'observer parmi les personnes occupées si les revenus, le volume de travail ou le choix de carrière sont affectés par le congé de leur père, ainsi que ceux de leur partenaire.

20. Les données relatives au congé de paternité permettent d'identifier les pères qui ont effectivement eu recours au congé de paternité, et d'observer la discontinuité en matière de prise du congé de paternité pour les pères qui ont eu un enfant avant/après la réforme de 2002. Par ailleurs, ces données sont nécessaires pour mesurer les effets intergénérationnels et la probabilité qu'ont les enfants de pères ayant bénéficié d'un congé de paternité de prendre à leur tour un congé de paternité.
21. Les données à caractère personnel pseudonymisées à communiquer au cours de la première phase semblent pertinentes et non excessives par rapport à la finalité précitée. Au cours de la deuxième phase, les chercheurs appliquent les algorithmes qu'ils ont développés à l'aide des données à caractère personnel pseudonymisées reçues précédemment à l'ensemble de la population, dans un environnement sécurisé auprès de la Banque Carrefour de la sécurité sociale et sous la surveillance d'un de ses collaborateurs. Seuls ces résultats, sous forme de données purement anonymes, peuvent quitter les locaux de la Banque Carrefour de la sécurité sociale.

Limitation de la conservation

22. Les données à caractère personnel pseudonymisées seront détruites par les chercheurs dès qu'elles ne sont plus nécessaires à la réalisation de la finalité précitée et ce au plus tard le 31 décembre 2027, date à laquelle la recherche devrait être terminée. Les données seront conservées par la Banque-Carrefour de la sécurité sociale durant deux années supplémentaires, soit jusqu'au 31 décembre 2029. Cette durée de conservation peut, le cas échéant, uniquement être prolongée par une décision explicite en la matière de la chambre sécurité sociale et santé du Comité de sécurité de l'information.

Intégrité et confidentialité

23. Le demandeur met tout en œuvre pour éviter une identification des personnes concernées et s'abstient, à tous les égards, de toute tentative visant à convertir les données à caractère personnel pseudonymisées reçues de la part de la Banque Carrefour de la sécurité sociale au cours de la première phase de l'étude en données à caractère personnel non-pseudonymisées. Par ailleurs, il ne communique, en aucun cas, ces données à caractère personnel pseudonymisées à des tiers. Il publie, en outre, les résultats du traitement qu'il a réalisé uniquement sous une forme qui ne permet d'aucune façon d'identifier les personnes concernées.
24. Pour le surplus, le demandeur tient compte de la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale* et de toute autre disposition réglementaire relative à la protection de la vie privée, en particulier le Règlement (UE)

2016/679 du Parlement européen et du Conseil du 27 avril 2016 *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE* et de la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel*.

25. Le demandeur reconnaît l'application exclusive du droit belge et de l'Union européenne, ainsi que la compétence exclusive de l'Autorité de Protection des Données (APD) et des juridictions belges en cas de litige.

Par ces motifs,

la chambre sécurité sociale et santé du comité de sécurité de l'information

conclut que la communication de données à caractère personnel pseudonymisées du datawarehouse marché du travail et protection sociale par la Banque Carrefour de la sécurité sociale à l'*Universitat Pompeu Fabra*, dans le cadre d'une étude sur les effets intergénérationnels du congé de paternité (projet « Leave4NextGen »), telle que décrite dans la présente délibération, est autorisée moyennant le respect des mesures de protection des données qui ont été définies.

La présente délibération entre en vigueur le 21 novembre 2024.

Michel DENEYER
Président

Le siège de la chambre sécurité sociale et santé du Comité de sécurité de l'information est établi dans les bureaux de la Banque Carrefour de la sécurité sociale, à l'adresse suivante: Quai de Willebroeck 38 - 1000 Bruxelles (tél. 32-2-741 83 11).