

**Microdonnées pour la recherche statistique et scientifique,
les possibilités pour les chercheurs d'obtenir des données individuelles auprès de Statbel**

Erik Meersseman et Patrick Lusyne

INTRODUCTION

Statbel, l'office belge de statistique, collecte, produit et diffuse des chiffres fiables et pertinents sur l'économie, la société et le territoire belges. Afin de produire ces statistiques, Statbel collecte des données. Pour ce faire, Statbel demande des données existantes à des institutions publiques ou privées et organise des enquêtes auprès des citoyens ou des entreprises.

La diffusion de chiffres se fait principalement via des statistiques globales et anonymes. Il s'agit de résultats où un chiffre se rapporte toujours à plusieurs citoyens ou entreprises. Aucune information confidentielle n'est donc divulguée. Cependant, pour certains chercheurs, ces statistiques globales et anonymes ne suffisent pas pour réaliser leur recherche. La législation prévoit la possibilité que Statbel communique dans certains cas des données individuelles (microdonnées), où chaque ligne du fichier contient les données d'une personne, d'un ménage ou d'une entreprise.

Cet article décrit les possibilités pour les chercheurs d'obtenir des données individuelles dont dispose Statbel à des fins statistiques et scientifiques.

CADRE LEGAL

Le traitement des données par Statbel ainsi que le partage de ces données avec des tiers sont régis légalement par la *Loi du 4 juillet 1962 relative à la statistique publique* (ci-après la 'loi statistique'). La loi statistique part du principe que toutes les données obtenues et traitées par Statbel pour la production de statistiques sont protégées par le secret statistique.

Par ailleurs, en ce qui concerne les données à caractère personnel, le *Règlement général sur la protection des données* (Règlement (UE) n° 2016/679 du Parlement européen et du Conseil du 27 avril 2016 - ci-après le 'RGPD') et la *Loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel* (ci-après la 'loi-cadre RGPD') sont également applicables.

Toute communication de données individuelles se fera dans les limites autorisées par la législation visée ci-dessus. Il convient de noter que la loi statistique protège également des données portant sur des entreprises ou des personnes décédées, ce qui n'entre pas dans le champ d'application du RGPD.

À QUELLES FINS DES MICRODONNEES PEUVENT-ELLES ETRE DEMANDEES?

Les chercheurs peuvent demander des microdonnées si celles-ci sont nécessaires pour leur recherche statistique ou scientifique.

On entend par 'nécessaire' que la recherche n'est pas possible sur la base de données globales et anonymes que Statbel diffuse sans limitations via le site web www.statbel.fgov.be:

- **Be.STAT**¹, la base de données électronique de Statbel, permet aux chercheurs de créer des tableaux sur mesure;
- Via les **Open data**², Statbel met à disposition des données agrégées à un niveau assez détaillé.

La finalité pour laquelle les données sont demandées se limite à la recherche statistique et scientifique. Toutefois, cet objectif peut être interprété au sens large. Les recherches destinées à soutenir la prise de décisions politiques ou les recherches effectuées par des institutions privées sont également prises en considération à condition que le traitement se fasse de manière indépendante, transparente et repose sur des méthodes scientifiques. Les résultats de la recherche devront également être rendus publics.

¹ <https://bestat.statbel.fgov.be/bestat/index.xhtml>

² <https://statbel.fgov.be/fr/open-data>

Le résultat du traitement doit consister en des statistiques globales et anonymes et en des rapports d'enquête qui ne peuvent avoir aucune conséquence individuelle pour les citoyens ou les entreprises. Naturellement, ces résultats globaux et anonymes publiés peuvent ensuite être librement utilisés à des fins non statistiques ou non scientifiques.

QUI PEUT DEMANDER DES DONNEES?

La loi statistique énumère les destinataires autorisés à recevoir des microdonnées de Statbel. Il s'agit tout d'abord des administrations fédérales, régionales, provinciales et locales et des organismes publics. Appartiennent également à cette catégorie les autorités statistiques régionales, le Bureau fédéral du Plan et le service statistique de la Banque nationale de Belgique.

Les personnes physiques et morales qui poursuivent un objectif scientifique sont également prises en compte. Il s'agit concrètement d'universités nationales et étrangères, de services d'études et d'organisations internationales. Pour les personnes physiques, il est pratiquement impossible de satisfaire aux exigences en matière de sécurité (voir ci-dessous) imposées par Statbel pour l'obtention de microdonnées. Par conséquent, les étudiants (doctorants) ne sont éligibles que si la demande émane de l'université elle-même et que celle-ci se porte garante de la protection des données.

QUELLES SONT LES DONNEES CONCERNEES?

Tant les données (administratives) existantes que Statbel a demandées que les données collectées par Statbel auprès des citoyens ou entreprises au moyen d'enquêtes sont prises en compte.

En ce qui concerne les données (administratives) existantes, Statbel doit respecter les conditions imposées par le Comité de surveillance statistique, le Comité de sécurité de l'information ou le contrat conclu avec le fournisseur des données. Il est généralement stipulé que Statbel ne peut mettre à disposition que les données qu'il a traitées. Les chercheurs à la recherche de données brutes doivent prendre contact avec le fournisseur de données. Par ailleurs, il convient pour chaque variable (ou pour chaque groupe de variables) d'expliquer les raisons pour lesquelles les données sont nécessaires à la recherche. Les données d'enquête que Statbel a collectées sont perçues comme un ensemble cohérent de variables, de sorte que la motivation doit se faire au niveau de l'enquête dans son ensemble.

Statbel ne met à la disposition des chercheurs que des données *pseudonymisées*. Il s'agit de données au niveau individuel (personne, ménage ou entreprise) qui ne permettent pas au chercheur de déduire l'identité de la personne, du ménage ou de l'entreprise. Cela signifie tout d'abord que l'ensemble de données ne peut contenir aucune information permettant une *identification directe*, par exemple le numéro de Registre national, le numéro d'entreprise, le nom, l'adresse, le numéro de téléphone, etc. Cependant, les données doivent également être protégées de l'*identification indirecte*. Cela signifie qu'une combinaison de données ne peut pas mener à une seule personne, un seul ménage ou une seule entreprise. Cet aspect est évalué pour chaque ensemble de données, mais de manière générale:

- la date de naissance est transposée en âge ou tranche d'âge;
- le code postal ou la commune est remplacé par l'arrondissement, la province, la région ou le degré d'urbanisation;
- les montants sont arrondis, écrêtés ou exprimés en classes ou centiles;
- les codes NACE des entreprises sont regroupés.

CHERCHEURS PRIVILEGES

Les services statistiques de Flandre (autorité statistique flamande), de Wallonie (Institut wallon de l'évaluation, de la prospective et de la statistique) et de Bruxelles (Institut Bruxellois de Statistique et d'Analyse), les départements statistiques de la Banque nationale de Belgique, le Bureau fédéral du plan et l'Observatoire des prix du SPF Économie peuvent légalement recevoir des microdonnées *non pseudonymisées* dans le cadre de leur mission au sein de l'Institut des Comptes Nationaux (ICN). Ils constituent une exception, à condition que les microdonnées non pseudonymisées soient nécessaires à l'accomplissement de leur mission

légale de production de statistiques. Il convient de noter que ces institutions et leur personnel sont soumis, dans le cadre de leur mission statistique, au secret statistique, tout comme Statbel (voir ci-dessous).

COUPLAGE DE DONNEES

Il peut y avoir une valeur ajoutée pour les chercheurs si les données de Statbel sont couplées à des données de tiers pour lesquelles le chercheur a obtenu une autorisation dans le cadre de son projet de recherche. Statbel étant légalement reconnu comme *tiers de confiance* (AR du 13 juin 2014), il peut se charger du couplage des deux données et de leur pseudonymisation. Statbel conserve la clé de la pseudonymisation afin que des données supplémentaires puissent être ajoutées ultérieurement.

Dans le cadre d'un couplage, il est important que le chercheur lui-même n'ait pas accès à des données avec identification directe, qui lui permettrait de déduire l'identité de la personne, du ménage ou de l'entreprise.

SECRET STATISTIQUE ET FUITES DE DONNEES

Toutes les données dont Statbel dispose sont protégées par le secret statistique. Cela signifie que les données ne peuvent en aucun cas être utilisées à des fins autres que la recherche statistique et scientifique. C'est pourquoi, dans le cas d'une communication de microdonnées, Statbel impose certaines mesures de sécurité:

- la finalité est définie contractuellement, de même que l'interdiction de diffuser les microdonnées;
- le demandeur doit disposer d'un délégué à la protection des données (Data Protection Officer ou DPO) et d'un conseiller en sécurité (CSI);
- les données doivent être placées sur un serveur sécurisé, qui se situe en Europe et dont l'accès reste limité aux chercheurs impliqués dans le projet;
- le contrat stipule que le résultat ne peut consister qu'en des statistiques globales et anonymes.

COMMENT DEMANDER DES MICRODONNEES?

Seules les données dont Statbel dispose et qui présentent une valeur ajoutée pour la recherche peuvent être demandées. Afin de vérifier si cette condition est remplie, le chercheur doit au préalable se concerter avec un statisticien de Statbel familiarisé avec le thème de la recherche et les données. Pendant cet entretien, il sera déterminé sur la base des besoins du chercheur quelles microdonnées entrent en ligne de compte pour une demande. Un aspect non négligeable est la qualité des données, l'univers, les périodes de référence et le moment où les données sont disponibles. Au cours de cette phase, on vérifiera également si la demande n'implique pas un risque (accru) en matière de sécurité des données, car si tel est le cas, la consultation préalable du DPO est requise.

Si Statbel peut apporter une valeur ajoutée à la recherche en rendant les microdonnées disponibles et qu'il n'y a pas de preuve à première vue d'un risque (accru), le chercheur peut soumettre une demande formelle en utilisant un formulaire standardisé. Le formulaire se trouve sur le site web de Statbel (<https://statbel.fgov.be/fr/propos-de-statbel/que-faisons-nous/microdonnees-pour-la-recherche>) Le formulaire se compose de deux parties. La première partie est le *formulaire de demande de microdonnées* et porte sur les aspects suivants:

- l'identité du responsable du traitement (ou du représentant légal);
- la finalité pour laquelle les données sont demandées (uniquement pour la recherche statistique et scientifique);
- les données demandées (variables, période de référence, source, ...);
- le délai de conservation souhaité avec justification;
- la catégorie de personnes qui auront accès aux microdonnées;
- en quoi consisteront les résultats (ceux-ci doivent être limités à des résultats globaux et anonymes);
- l'identité du DPO.

Le deuxième volet comporte une *déclaration relative aux mesures techniques et organisationnelles* visant à prévenir les violations du secret statistique ou les fuites de données. Concrètement, ce deuxième volet comporte:

- l'identification du conseiller en sécurité;
- la procédure interne relative aux fuites de données;
- la liste des mesures spécifiques visant à lutter contre la violation du secret statistique et les fuites de données;
- la politique de sécurité de l'organisation.

Le formulaire complété constitue la base de la délibération de Statbel sur la demande, et c'est pourquoi la première partie du document doit être signée par le responsable du traitement (ou le représentant légal). La deuxième partie doit être signée par le conseiller en sécurité, le responsable du traitement ou le représentant légal de l'organisation. Le terme '*responsable du traitement*' ne désigne pas la personne à la tête de la recherche concrète pour laquelle les données seront utilisées, mais la personne morale ou physique telle que décrite à l'article 24 du RGPD. Plus précisément, il s'agit de la personne qui fixe les objectifs, fournit les ressources, signe le contrat de confidentialité et est responsable si quelque chose devait mal tourner. Il s'agit par exemple du recteur ou de l'administrateur d'une université, du président d'une administration fédérale, du secrétaire-général d'une administration régionale, etc.

DELIBERATION ET AUTORISATION DE STATBEL

Chaque demande sera évaluée par un comité multidisciplinaire au sein de Statbel. Cette évaluation permet de vérifier non seulement si les conditions légales sont satisfaites, mais également si les données demandées sont proportionnelles par rapport à l'objectif de la recherche et si les mesures techniques et organisationnelles sont proportionnelles aux risques. Le cas échéant, des mesures supplémentaires seront imposées pour lutter contre le risque d'identification indirecte, une réutilisation des données à des fins non statistiques ou non scientifiques ou des fuites de données.

Sur la base de cet avis, le DPO de Statbel rédigera un avis formel. Enfin, c'est le directeur général qui, en tant que représentant légal de Statbel, donne son approbation en signant l'avis du DPO. Un contrat de confidentialité est alors rédigé et soumis à la signature du responsable du traitement du chercheur. La délibération par Statbel et la rédaction du contrat de confidentialité prennent 2 à 3 semaines, à compter de la réception du formulaire de demande dûment complété et signé. Les avis du DPO et les contrats de confidentialité signés sont publiés sur le site web de Statbel: <https://statbel.fgov.be/fr/propos-de-statbel/vie-privee/protection-de-la-vie-privee-rgpd>.

PROCEDURES SIMPLIFIEES

La procédure formelle développée par Statbel pour la communication des microdonnées se limite aux informations nécessaires pour une évaluation correcte de la demande. Si le chercheur souhaite adapter une autorisation existante, Statbel prévoit un *formulaire de demande simplifié*. Cela ne peut concerner qu'une prolongation du délai de conservation, l'ajout de variables supplémentaires, l'ajout de périodes de référence supplémentaires et l'extension de la finalité de la recherche. La simplification par rapport à la procédure normale se situe dans le formulaire de demande qui est moins détaillé.

Pour certaines recherches, des *données agrégées avec un faible risque d'identification indirecte* suffisent pour effectuer la recherche. En raison de ce faible risque, les données ne peuvent être considérées comme globales et anonymes et une délibération de Statbel doit avoir lieu. Ici aussi, un formulaire de demande très simplifié suffit.

DELIBERATIONS ET COMMUNICATIONS DE MICRODONNEES DEPUIS MAI 2018

La procédure décrite ci-dessus pour l'obtention de microdonnées est applicable depuis l'entrée en vigueur du RGPD le 25 mai 2018 et la suppression du Comité de surveillance statistique de l'ancienne Commission de la protection de la vie privée (Commission vie privée). Au cours des années 2018, 2019 et 2020, Statbel a approuvé 209 demandes de chercheurs pour l'obtention de microdonnées, dont 50 via une procédure simplifiée afin d'étendre une autorisation existante. Grâce au contact

préalable entre le chercheur et le statisticien de Statbel, il est rare que des demandes soient refusées. Cependant, dans certains cas, une adaptation de la demande est nécessaire.

Type de demande	2018	2019	2020	Total
<i>Demandes ordinaires</i>	24	39	36	99
<i>Demandes par une autre autorité statistique qui est membre de l'IIS</i>	5	26	16	47
<i>Demandes dans le cadre de l'ICN</i>	4	3	6	13
<i>Extensions d'un contrat existant</i>	12	15	23	50
<i>Total</i>	45	83	81	209

LA PRATIQUE – CAS D'UTILISATEURS

L'évolution des techniques relatives aux données au cours de la dernière décennie a conduit à une vaste réforme de la production régulière de statistiques au sein de Statbel, qui met notamment davantage l'accent sur les sources secondaires, telles que les données administratives. Cela a permis de réaliser d'importants gains d'efficacité, qui ont été mis à profit pour développer des produits de données de plus en plus sophistiqués répondant aux besoins croissants de nos parties prenantes (notamment les départements politiques, le monde académique, etc.).

Les produits de données de Statbel couvrent de nombreux domaines de recherche et de politique (données économiques, sociales et démographiques) et les innovations prennent différentes formes:

- certains nouveaux produits de données sont des extensions directes de la production régulière de données et de statistiques. Cela se fait souvent à la demande d'acteurs institutionnels importants. Les modules ad hoc des enquêtes en sont un exemple.
- dans une certaine mesure, les données administratives sont encore sous-évaluées. Statbel a développé différents produits de données qui dévoilent davantage les sources à sa disposition et conduisent à une valorisation plus riche, permettant de mettre en lumière des phénomènes économiques et sociaux difficiles à cartographier jusqu'à récemment. Le développement récent des données d'origine ainsi que les séries de données longitudinales basées sur le Registre national en sont des exemples.
- en couplant les données administratives, il est possible d'obtenir des informations qui restent totalement inaccessibles lorsque la recherche ou le travail préparatoire à la prise de décisions politiques se fondent sur une seule source. Le census administratif est un exemple de produit standard où ce principe est explicitement mis en pratique. Toutefois, l'écosystème de données très riche de Statbel permet d'intégrer des informations provenant de différentes bases de données de manière beaucoup plus libre. En outre, cela ne concerne pas seulement les données collectées par Statbel, mais ce dernier collabore également avec d'autres détenteurs de données, tels que la Banque-Carrefour de la Sécurité Sociale (BCSS) et l'Agence Inter-Mutualiste (IMA). Tout cela se fait dans le strict respect de la loi.

Plus particulièrement en ce qui concerne ce dernier point, mais en fait globalement, Statbel offre la personnalisation, bien que de grandes consolidations de données apparaissent, car diverses bases de données créées sur mesure sont réparties entre divers utilisateurs. Il s'agit alors d'un effet d'harmonisation et conduit même les consortiums de recherche à s'organiser quelque peu autour de nos données.

La liste non exhaustive suivante comprend un bref aperçu de plusieurs projets de recherche plutôt récents qui utilisent les microdonnées mises à disposition par Statbel:

- **Immilab** est un projet sponsorisé par la Politique scientifique fédérale belge (BELSPO) qui examine la position des personnes issues de l'immigration sur le marché du travail belge. Des équipes de recherche de l'UA, de l'ULB et de l'UMons participent à ce projet. Statbel y contribue via un couplage innovant entre l'Enquête sur la structure et la répartition des salaires (SES) et les données démographiques;
- **Re-invest** concerne un projet qui met l'accent sur le lien entre la protection sociale, les logements et la santé. Ce projet utilise les données longitudinales 2004-2018 de la statistique sur les revenus et les conditions de vie (Statistics on Income en Living Conditions - SILC) comme base, enrichies des données du Censur (2011) et des informations démographiques (notamment sur les déménagements). Les partenaires de ce projet sont la KU Leuven - HIVA, le Service de lutte contre la pauvreté et l'UCL.
- **SUSPENS** est un projet financé par BELSPO, réalisé par des chercheurs de l'Université d'Anvers, l'ULB et le Bureau fédéral du Plan. Les sources utilisées sont l'enquête SILC et l'enquête sur le budget des ménages de Statbel, afin d'établir un lien entre les revenus, le comportement des consommateurs et les émissions de gaz à effet de serre au niveau du ménage.
- La **VMM** (Vlaamse Milieumaatschappij) calcule un indicateur pour le test d'accessibilité financière de la facture d'eau intégrale par an. Les dépenses pour la facture d'eau intégrale sont comparées aux revenus du ménage, afin de fournir des conseils politiques sur le caractère abordable de l'eau potable. À cet effet, Statbel a couplé les informations de SILC aux données de la facture d'eau de la VMM. Afin de protéger les données contre l'identification par la VMM, des mesures techniques et organisationnelles supplémentaires ont été imposées.
- **Causineq** et les **bases de données de la mortalité** pendant l'épidémie de **COVID-19**. Causineq est une étude à grande échelle de la mortalité différentielle selon le gradient social à laquelle l'UCL et la VUB participent. Dès le début de la crise du COVID-19, Statbel a produit des données hebdomadaires de mortalité similaires utilisées par divers organismes et universités pour cartographier la surmortalité (Sciensano, UCL, VUB, UHASSELT).

Ce qui est peut-être encore plus frappant, c'est que, parallèlement aux progrès technico-méthodologiques réalisés ces dernières années, un changement culturel a également eu lieu au sein de Statbel. Le passage d'une organisation purement 'productrice de statistiques' à un important producteur de données est allé de pair avec l'adoption d'une méthode de travail plus ouverte et, surtout, avec une augmentation significative des services offerts aux parties prenantes.

CONCLUSION

Statbel dispose de données de grande qualité et du mandat légal pour mettre ces données à la disposition de chercheurs dans le cadre de recherches statistiques et scientifiques. En outre, Statbel dispose de l'expertise nécessaire pour créer des produits de données personnalisés de haute qualité. Ces dernières années, Statbel s'est transformé en une organisation qui aide activement les chercheurs en leur fournissant des données qui répondent à leurs besoins spécifiques.