

Question Parlementaire n°62 du 26 novembre 2018 de M. le Député Sven Clement

Question

Le STATEC met-il à disposition du public des micro-données anonymisées ? Si oui, où et comment se les procurer ? Si non, comment le STATEC justifie-t-il la rétention de telles données ?

Réponse

Le STATEC met déjà à disposition du grand public des micro-données anonymisées. Celles-ci ont été collectées dans le cadre de deux enquêtes majeures conduites par le STATEC et ses homologues européens : d'une part EU-SILC, une enquête sur les revenus et conditions de vie des ménages, et d'autre part l'enquête « Forces de Travail », consacrée à la mesure de l'activité professionnelle des particuliers. Ces deux enquêtes sont organisées de manière harmonisée dans toute l'Union européenne, et les micro-données sont rassemblées par Eurostat. Sur <https://ec.europa.eu/eurostat/web/microdata/labour-force-survey> et <https://ec.europa.eu/eurostat/web/microdata/statistics-on-income-and-living-conditions>, tout le monde peut télécharger des fichiers de micro-données anonymisées provenant des États membres de l'Union. Le STATEC a livré la collection complète des données luxembourgeoises sur la période de 10 ans courant de 2004 à 2013.

Sur son site, Eurostat informe les utilisateurs que les données contenues dans les fichiers en question, appelés « public use files », ont été dûment anonymisées, notamment pour résister à des tentatives plus ou moins fines de ré-identification des individus. Ces techniques d'anonymisation vont nettement plus loin que l'effacement des données nominatives et des données d'identification « évidentes » comme par exemple un numéro de matricule national à 13 chiffres.

En effet, un jeu de micro-données ne résiste pas de la même façon aux tentatives de ré-identification suivant qu'il est analysé par quelques chercheurs soumis à des engagements stricts de confidentialité et ayant déclaré des finalités explicitement déterminées à l'avance, ou qu'il est analysé par une personne quelconque. En effet, dans ce second cas, on augmente de manière considérable le risque qu'une personne figurant dans le fichier soit identifiée par quelqu'un qui la reconnaîtrait à travers ses seules réponses aux questions socio-économiques du questionnaire du STATEC. L'application de ces techniques fortes d'anonymisation conduit inévitablement à un appauvrissement de l'information contenue dans un fichier de données mis à disposition du grand public. C'est un des prix à payer pour garantir la protection des données. Pour certaines enquêtes, la protection des données doit encore être augmentée. C'est notamment le cas de l'enquête de victimation « Crime » auprès des ménages, ou de certaines enquêtes auprès des entreprises, récoltant des données qui ne sont pas soumises à la publicité selon la loi du 19 décembre 2002 concernant le registre de commerce et des sociétés.

Le STATEC diffuse également des données à travers le portail Open Data (<https://data.public.lu>), à ce jour il en est un contributeur important, avec 160 tableaux de données agrégées. En tant qu'institut national de la statistique et des études économiques, le STATEC salue et soutient l'initiative Open Data. En effet, il a bien conscience que le citoyen luxembourgeois finance les collectes des micro-données afin d'établir des statistiques publiques ; il est donc naturel que l'accès à ces données ne lui soit pas purement et simplement interdit. Ce thème figure également à l'ordre du jour de Conseil supérieur de la statistique, qui représente la société civile et surveille les travaux du STATEC et des

autres membres du système statistique luxembourgeois, notamment en émettant et publiant un avis annuel. Suite à la dernière « Peer Review », une visite formelle des experts d'autres instituts nationaux de statistiques de l'Union Européenne, le Conseil supérieur de la statistique s'est vu confier la mission d'orchestrer l'élaboration de règles et procédures précises pour l'accès aux micro-données au sein du système statistique luxembourgeois.

Enfin, il convient de signaler que le STATEC, qui exécute également une mission de recherche, donne accès aux données individuelles à des chercheurs, et à des étudiants en master ou en doctorat. À ce titre, la loi du 10 juillet 2011 portant organisation du STATEC prévoit et encadre dans son article 16 l'accès aux micro-données *in situ*, à des fins scientifiques. Le législateur précise dans ce contexte que « la recevabilité de la demande et l'autorisation d'accès à des fins scientifiques aux données confidentielles sont subordonnées à l'examen du bien-fondé et de l'intérêt scientifique des projets de recherche pour lesquels l'autorisation est sollicitée, ainsi qu'à l'examen de la qualification scientifique du ou des demandeurs. » La protection des données, qu'elles proviennent de ménages ou d'entreprises, est là encore mise en avant par la disposition explicite « Les renseignements individualisables ne peuvent en aucun cas être divulgués. »

La collecte et le traitement de données à des fins de production de statistiques publiques sont basés sur la confiance du public dans le traitement confidentiel des données. Une remise en cause de cette confiance, par un accès non contrôlé ou illégal aux informations fournies par nos concitoyens et nos entreprises, serait de nature à porter gravement atteinte au STATEC. D'un autre côté, le STATEC a pleinement conscience que les données qui lui sont confiées ont vocation à être mises à disposition du public de la manière la plus diverse et transparente possible. La « data science » de demain se doit aussi d'avoir un volet citoyen, au risque de s'enfermer dans une tour d'ivoire. Il convient donc de trouver un équilibre juste entre protection et diffusion, entre sécurité et ouverture.