

Une protection renforcée de la vie privée et une plus grande liberté pour les chercheurs



Jean-Pierre LE GLÉAU

Inspecteur général honoraire de l'Insee

Nicolas Belorgey propose une critique du numéro 2014-2 de la revue *Statistique et Société* qui, selon lui, « ne prend pas en compte les besoins d'une démarche rigoureuse aujourd'hui en sciences sociales ».

Il considère en effet que la connaissance en sciences sociales doit s'appuyer sur des outils non seulement quantitatifs, mais aussi qualitatifs. Pour cela, il estime qu'il est nécessaire d'allier deux démarches : le raisonnement sur des cas individuels, afin d'obtenir une bonne compréhension des phénomènes au niveau le plus fin possible, et les tests d'hypothèses statistiques, afin de vérifier le degré de généralité de ces éléments intermédiaires.

L'article intitulé « L'accès aux données confidentielles de la statistique publique – De la sensibilité des données économiques à la sensibilité des données de santé » semble particulièrement visé par ce reproche. Son écriture a sans doute été imprécise, ou mal comprise par le lecteur. En effet, les carences supposées de la démarche qui y est décrite (passage par le comité du secret statistique, puis accès aux données sur un centre sécurisé) constituent justement le point fort de celle-ci, en permettant au chercheur d'effectuer des allers-retours entre les données individuelles et les données agrégées. L'incompréhension provient peut-être du fait qu'il est dit que le centre d'accès sécurisé (CASD) héberge des données de la statistique publique. Il faut bien comprendre qu'il s'agit des données individuelles qui ont permis, après traitement, de produire des statistiques anonymes. Mais le chercheur a accès aux données en amont de ces traitements, les plus détaillées possible qui, bien qu'anonymisées, restent confidentielles à cause des possibilités d'identification indirecte. C'est d'ailleurs pour cela, et afin de protéger la vie privée, qu'un protocole spécifique est nécessaire pour que les chercheurs puissent travailler sur ces données individuelles très détaillées. Ce protocole laisse par contre la plus grande liberté au chercheur dans son travail, et lui permet notamment d'effectuer des allers-retours entre les cas individuels et le cadrage statistique. La seule contrainte qui s'impose à lui est de ne sortir de cette « bulle » que des résultats rendant impossible l'identification d'une personne, contrainte qui semble difficilement contestable.

L'auteur de la critique suggère aussi que les données administratives déjà disponibles doivent pouvoir être enrichies par des informations provenant d'enquêtes réalisées par le chercheur. C'est effectivement une démarche très féconde. Mais celle-ci est tout à fait réalisable dans le cadre du protocole décrit dans l'article. Des appariements de données provenant de diverses sources (administratives ou personnelles) peuvent être réalisés dans l'enceinte du CASD. Ce n'est même que depuis que ce centre d'accès sécurisé a été mis sur pied que ces appariements sont devenus possibles. Certes, les démarches nécessaires pour réaliser de tels appariements sont souvent complexes, lourdes et longues à mettre en œuvre. Mais cette complexité n'est pas à imputer au protocole du CASD, mais tout simplement... au texte de la loi elle-même (notamment la loi informatique et libertés) qui impose des processus très lourds pour les

appariements de fichiers, en particulier ceux contenant le numéro d'inscription au répertoire d'identification des personnes physiques (NIR). Un assouplissement de ces règles figure dans le projet de loi sur la santé qui sera prochainement examiné par le Parlement.

L'auteur propose in fine un autre type de démarche : sélectionner un certain nombre de personnes à partir de données statistiques et compléter les informations sur celles-ci à l'aide d'enquêtes qualitatives. Cette démarche est certainement plus délicate, puisqu'elle suppose de « sortir » de la « bulle » du CASD des données statistiques individuelles pour pouvoir interroger les personnes. Cela est contraire à l'esprit général du processus décrit dans l'article de 2014. On conviendra volontiers que ce plan d'expérimentation, qui consiste à choisir des individus, avec leurs données confidentielles, pour pouvoir les interroger comporte des risques sérieux pour la protection de la vie privée. Il n'est cependant pas formellement impossible, bien que ne s'inscrivant pas dans le processus standard développé pour le CASD. Il suppose l'intervention d'un tiers de confiance et des protocoles complexes, qui ont cependant déjà été mis en œuvre, mais pour l'instant au seul profit de centres de recherche publics.

Comme on le voit, les voies proposées par l'auteur, loin d'être incompatibles avec la démarche « comité du secret – CASD » sont rendues possibles précisément depuis que cette architecture a été mise en place. Il en résulte une protection renforcée de la vie privée et une plus grande liberté pour les chercheurs. Celle-ci est d'ores et déjà appréciée par les chercheurs en sciences économiques ou sociales. Elle pourrait l'être à l'avenir des chercheurs travaillant dans le domaine de la santé.