

**Commission**  
**« Démographie et questions sociales » du 28 novembre 2019**

**SYNTHÈSE**

La commission « Démographie et questions sociales » du 28 novembre 2019 s'est tenue sous la présidence de Christine D'Autume.

La réunion a débuté par un point d'information sur le règlement européen relatif aux statistiques sociales, publié le 10 octobre et dont l'entrée en application est prévue en janvier 2021. Le processus de préparation de ce règlement a été lancé il y a environ dix ans par un mémorandum des directeurs généraux des instituts nationaux de statistique appelant à la modernisation des enquêtes sociales et à leur plus grande cohérence. Ce règlement vise à structurer et à harmoniser la production des enquêtes sociales auprès des ménages. D'autres règlements-cadres viendront couvrir les questions de population et le recensement. Ce texte marque un grand progrès en matière de cohérence, puisque les enquêtes sociales auprès des ménages étaient auparavant organisées en tuyaux d'orgue, avec un règlement par enquête, conçus au fil du temps, et pas toujours totalement cohérents entre eux dans le contenu et la structure. Le nouveau règlement spécifie des concepts communs et plusieurs noyaux concentriques de variables standardisées, avec des contraintes de qualité renforcées. Ce règlement sera assorti d'une programmation pluriannuelle des enquêtes et comporte sept grandes thématiques : la main d'œuvre, le revenu et les conditions de vie, la santé, l'éducation et la formation, l'utilisation des technologies de l'information et de la communication, l'emploi du temps et la consommation.

Ensuite cinq projets d'enquêtes ont été présentés pour l'obtention d'un avis d'opportunité. Il s'agit de l'enquête ESCAPAD 2021 de l'Observatoire français des drogues et toxicomanies, le dispositif Autonomie 2021-2023 de la Direction de la recherche, des études, de l'évaluation et des statistiques, l'enquête régionale sur le sport à La Réunion réalisée par l'université de la Réunion, l'enquête nationale sur les conditions de vie des étudiants menée par l'Observatoire national de la vie étudiante et enfin l'étude de l'impact sur la santé de l'incendie de l'usine Lubrizol de Rouen par Santé publique France. L'ensemble de ces enquêtes a obtenu un avis d'opportunité favorable.

La commission a ensuite abordé le thème des nouvelles sources de données et de leurs utilisations innovantes dans les champs de la santé et de la consommation à travers trois présentations.

D'abord, la Drees a présenté **un point d'avancement du projet *Health Data Hub***. Il a été rappelé l'essor important des données de santé et les multiples applications dans ce domaine, depuis la recherche jusqu'au pilotage des systèmes de santé en passant par l'appui aux professionnels de santé. Malgré un patrimoine important en France, avoir connaissance des données disponibles et des moyens pour y accéder n'est pas évident. Parfois, les données ne sont pas mises à disposition et il n'est pas possible de leur appliquer des méthodes innovantes.

En 2016, le législateur a créé le Système national des données de santé (SNDS) et mis en place une gouvernance pour accompagner l'accès à ces données, avec l'Institut national des données de santé (INDS). L'article 41 de la loi d'organisation et de transformation du système de santé a ensuite élargi le périmètre du SNDS à l'ensemble des données cliniques recueillies lors d'actes pris en charge par l'assurance maladie. Le *Health Data Hub* est alors créé à partir de l'INDS en élargissant ses missions. Il a en particulier la mission d'organiser et de mettre à disposition les données du SNDS, d'informer les patients et de faciliter leurs droits, notamment le droit d'opposition. Enfin, un comité éthique et scientifique est chargé de se prononcer sur le caractère d'intérêt public des traitements.

L'offre de service du *Health Data Hub* doit accompagner les porteurs de projet dans les différentes étapes, du cadrage du projet et la mise à disposition de documentations collaboratives à l'aide aux traitements via la mise en place d'une plate-forme technologique sécurisée pour l'analyse des données avec les capacités nécessaires aux traitements innovants. Le Hub se constitue peu à peu à partir d'un ensemble de bases de données qui seront référencées dans un catalogue. Ce catalogue listera d'abord les principales bases stratégiques pour la communauté scientifique. Grâce à ce catalogue, l'accès aux données sera facilité.

Le projet a démarré en mars 2018. Un appel à projets a été lancé en janvier 2019. Dix projets pilotes ont été sélectionnés. Ils seront utilisés pour dimensionner la plate-forme, la tester et initier l'accompagnement des porteurs de projet. Ces projets ont été proposés par divers acteurs, institutions ou startups. Tous utilisent les données du SNDS qui sont, la plupart du temps, appariés à d'autres bases de données. La plate-forme devrait constituer un outil facilitant les interactions entre les différents acteurs, producteurs des données et utilisateurs. Ce hub doit garantir l'accès aisé, unifié, transparent et sécurisé aux données de santé.

Une deuxième présentation de la Drees a exposé **les utilisations innovantes du SNDS**. Le Système national de données de santé couvre des volumes considérables de données entraînant de nombreux changements en termes de méthodologie. Pour couvrir dix ans de SNDS, il faudrait traiter 50 milliards de lignes. Plusieurs organisations ont l'habitude de travailler avec ces données, mais la documentation n'est pas toujours la même. Rassembler celle-ci dans un seul endroit permettrait de garantir la qualité en termes d'utilisation des données.

En termes d'analyse, il s'agit de mutualiser les langages. Il existe depuis quelques années des langages de programmation en open source permettant l'utilisation des outils par des personnes différentes. La mutualisation des ressources permet de développer de grands projets open source comme Apache, un projet développé par plus de 7 000 collaborateurs à destination de centaines de milliers de personnes, ou comme le projet de documentation du SNDS. En termes de restitution, certains traitements peuvent être automatisés avec une mise à disposition des utilisateurs finaux d'interfaces graphiques intuitives.

Ces nouvelles sources de données présentent des champs d'application considérables et impliquent des vitesses de calcul impressionnantes. Elles exigent le développement d'une plus grande technicité.

Une troisième présentation de l'Insee s'est intéressée aux **résultats du calcul de l'indice des prix à la consommation (IPC), réalisé à partir des données de caisse**. Jusqu'à présent, pour produire cet indice, l'Insee se fonde sur 200 000 relevés réalisés chaque mois par des enquêteurs dans les points de vente, mais aussi sur un même volume de prix collectés sur Internet ou via des bases de données, comme celle de la Caisse nationale d'assurance maladie sur les services de santé.

A compter de janvier 2020, il est prévu d'intégrer un changement majeur dans les sources utilisées pour le calcul de l'indice en remplaçant une partie des relevés physiques par des données de caisse enregistrées par les enseignes. Lorsqu'un produit est scanné, différentes informations à un niveau très détaillé sont collectées en termes de prix et de quantités consommées. Ces données de caisse seraient utilisées pour l'ensemble des super et hypermarchés sur un champ de produits restreint aux projets alimentaires industriels, d'entretien et d'hygiène-beauté et uniquement en France métropolitaine.

Grâce aux données de caisse l'Insee collecte ainsi, chaque mois, 1,7 milliard d'enregistrements. L'exhaustivité permise par les données de caisse apportera une meilleure précision de calcul et une meilleure information sur les quantités. Par la suite, il est prévu de poursuivre des travaux méthodologiques en vue de produire de nouvelles statistiques, notamment des prix moyens, des comparaisons spatiales de prix et à plus long terme, des IPC régionaux.

Au terme des discussions autour de ces trois présentations, **la commission a voté un avis** qui salue les différentes initiatives et encourage leur poursuite. Elle demande à être tenue informée des avancées des travaux méthodologiques et recommande au service statistique public d'exploiter les potentialités offertes par le *Health Data Hub* pour explorer des thématiques nouvelles et améliorer les statistiques existantes.