

Synthèse de l'assemblée plénière du Cnis du 23 janvier 2024

L'assemblée plénière annuelle du Cnis s'est déroulée en présentiel à Bercy sous la présidence de Patrice Duran. Elle a réuni une soixantaine de participants, membres du Conseil et représentants du service statistique public.

La réunion s'est déroulée en trois temps. D'abord une présentation des principales innovations au service de la statistique publique depuis 20 ans. Puis une séquence consacrée à l'élaboration d'un nouveau programme de moyen terme du Cnis pour les années 2024-2028, et au vote des avis. Enfin, une table ronde sur les réponses apportées aux besoins statistiques.

Jean-Luc Tavernier, DG de l'Insee, a brossé 20 ans d'innovation, en choisissant d'évoquer successivement les enquêtes, les données administratives, les données privées et la diffusion. S'agissant des enquêtes, il a souligné pour les entreprises la généralisation des enquêtes sur internet et l'utilisation de données fiscales pour alléger la charge pour le répondant, et pour les ménages le développement d'enquêtes multimodes pour tenter d'enrayer l'érosion du taux de réponse : l'enquêté peut choisir le mode par lequel il répond à l'enquête, et de fait les taux de réponse se maintiennent en France alors qu'ils baissent ailleurs. Les départements d'outre-mer font l'objet d'un programme spécifique et d'un suréchantillonnage (par rapport aux régions de métropole) systématique pour fournir des résultats à l'échelle des Outre-mer. De nouvelles enquêtes ont vu le jour et ont souvent été rééditées dans de nombreux domaines : PIAAC sur les compétences des adultes, PISA sur le suivi des acquis des élèves, Sans domicile, Bénéficiaires de l'aide alimentaire, Trajectoires et origines, Vécu et ressenti en matière de sécurité, Modes de garde, TRELO sur les travaux de rénovation énergétique dans les logements, etc. Parmi les enquêtes auprès des entreprises, deux innovations récentes : Acemo-Covid et l'enquête désormais annuelle sur les prix prévisionnels de l'électricité et du gaz.

Au chapitre des données administratives, la Déclaration sociale nominative a été conçue en intégrant la finalité statistique, les appariements de données individuelles se sont développés et ont permis la construction de bases statistiques exhaustives telles que FILOSOFI (revenus localisés sociaux et fiscaux) ou FIDELI (fichier démographique sur les logements et les individus), bientôt RESIL (répertoire d'individus et de logements), mais aussi des dispositifs sur la formation professionnelle, un suivi des trajectoires des jeunes, de la formation initiale à l'emploi, un panel démographique incluant des données de santé... Par ailleurs, le recours aux données privées est encadré depuis 2016 par la loi pour une République numérique : les données de caisse entrent maintenant pour plus de 10 % dans le calcul de l'indice des prix à la consommation, les transactions par cartes de crédit sont exploitées depuis la crise Covid, de même que des comptes bancaires, contrairement aux données de téléphonie mobile dont le partenariat a du mal à s'inscrire dans la durée.

Concernant la diffusion, les progrès ont notamment concerné le raccourcissement des délais de publication de tous les indicateurs, la mise à disposition de données individuelles selon des procédures adaptées aux différents publics et des modalités affichées, ainsi que la documentation des bases.

Lors de la discussion, une question a porté sur les possibles biais que pourraient introduire dans le calcul de l'IPC les enseignes qui ne transmettent pas leurs données de caisse. Les redressements mis

en œuvre permettent d'éviter ces biais. L'exploitation des traces numériques a fait l'objet d'une autre question : l'Insee en a utilisé lors de la crise du Covid pour estimer la part de la population qui avait quitté l'agglomération parisienne, mais cela pose des problèmes d'accès d'une part (les réseaux de téléphonie mobile vendent leurs données) et des questions de conservation des données, le temps de l'étalonnage nécessaire à l'établissement de statistiques, d'autre part.

Pour le moyen terme, Dominique Francoz a rappelé l'ensemble des consultations qui se sont déroulées tout au long de l'année 2023 pour faire émerger les besoins statistiques et ont conduit les commissions à produire des avis de moyen terme sur leurs thématiques et le bureau à proposer des avis généraux, par nature transverses. Il a ensuite été procédé à la discussion des avis généraux, en pointant les quelques amendements de forme introduits suite aux dernières suggestions des membres du Conseil. Le besoin d'une meilleure prise en compte des utilisateurs nationaux dans les projets statistiques européens a été souligné par un membre du conseil, qui a regretté que le règlement 223 (loi statistique européenne) n'ait pas pu faire l'objet d'une discussion au Cnis au printemps 2023. Le Directeur Général de l'Insee a proposé qu'une étude de cas soit réalisée en bureau du Cnis sur un prochain règlement en cours d'adoption. Après le vote des 12 avis généraux de moyen terme, les avis des commissions, déjà validés par ces dernières, ont été adoptés lors d'un vote unique. Patrice Duran a remercié tous les contributeurs, utilisateurs ou producteurs de statistique, soulignant leur forte implication dans la préparation de ce programme, en particulier les présidents et rapporteurs des commissions. Un séminaire du Cnis, le 3 avril, sera l'occasion d'aborder plusieurs sujets du nouveau moyen terme dans deux tables rondes.

Une table ronde intitulée « Quelles réponses aux besoins statistiques ? » a ensuite cherché à illustrer deux grandes évolutions depuis le [rapport Bozio-Geoffard](#) de 2017 sur l'accès des chercheurs aux données administratives et la loi pour une République numérique (2016) qui a porté l'open data : d'une part le besoin des chercheurs en microdonnées et le développement des modalités d'accès, d'autre part les nouveaux besoins de données locales et leur mise à disposition. [Quételet-Progedo-Diffusion](#) propose à la communauté scientifique des fichiers production et recherche (FPR) d'un niveau intermédiaire d'information entre les données standards anonymisées (et plus agrégées) et les données confidentielles plus détaillées, accessibles sous conditions, en particulier au CASD. Erik Zolotoukhine, son responsable, a précisé que son catalogue contient plus de 1500 références et que le nombre de demandeurs a doublé depuis 2012. Kamel Gadouche, directeur du Centre d'accès sécurisé aux données ([CASD](#)), considère qu'avec plus de 500 sources et environ 80 outils de traitement bénéficiant à plus de mille utilisateurs, l'offre multisectorielle du CASD est assez unique au monde. La Depp est en partenariat étroit avec [l'Equipex IDEE](#) depuis 2 ans, a expliqué Marc Gurgand, coordonnateur scientifique du programme, ce qui facilite l'accès de ses données à des chercheurs de disciplines variées (statisticiens, sciences de l'éducation, psychologues...), propose un catalogue en ligne au standard international (DDI) permettant de formaliser les opérations d'appariement entre enquêtes et données administratives. Christine Hennion a quant à elle présenté les conclusions du rapport [Data et territoires](#) qu'elle a co-écrit, soulignant l'asymétrie d'information entre l'Etat et les collectivités, le manque de coordination entre ministères qui les conduisent parfois à adresser aux collectivités les mêmes demandes. Son rapport recommande d'organiser une gouvernance des données à collecter, en partant de ce qui existe déjà, via le CNIG (Conseil national de l'information géolocalisée) notamment, avec des référents installés par la Dinum dans chaque région. Arnaud Degorre a fait état des progrès dans l'offre de données locales de l'Insee et dans sa capacité à mener des analyses spatiales, à différentes échelles, de façon homogène d'un territoire à l'autre, ce que permettent le renouveau des sources administratives exhaustives et l'organisation mutualisée du réseau de l'Insee. Les services de la donnée locale proposent des API dédiées, des cartographies dynamiques, des comparateurs de territoires, un catalogue de jeux de données interopérable avec d'autres systèmes de données.

Patrice Duran a conclu la séquence en rappelant que la statistique publique est une affaire d'action collective et que les interventions ont bien montré le besoin de coordination.