



Conseil national
de l'information statistique

Montrouge, le 28 février 2024 – n° 23/H030

ASSEMBLÉE PLÉNIÈRE



Réunion du 23 janvier 2024

COMPTE RENDU DE LA RÉUNION DE

L'ASSEMBLÉE PLÉNIÈRE

- 23 janvier 2024 -

Président : Patrice DURAN, Président du Cnis

RAPPEL DE L'ordre du Jour

I. INTRODUCTION.....	6
II. 20 ANS D'INNOVATIONS AU SERVICE DE LA STATISTIQUE PUBLIQUE.....	6
III. UN NOUVEAU MOYEN TERME POUR 2024-2028.....	9
IV. DISCUSSION ET VOTE DES AVIS GÉNÉRAUX DU MOYEN TERME 2024-2028.....	10
Avis général n°1 : Mesurer la transition écologique et rendre compte de ses enjeux économiques et sociaux.....	10
Avis général n°2 : Mesurer la transformation numérique de l'économie et de la société.....	10
Avis général n°3 : Inscrire la statistique publique dans une perspective internationale.....	10
Avis général n°4 : Rendre compte des enjeux générationnels dans les statistiques.....	11
Avis général n°5 : Enrichir les travaux sur les inégalités et les discriminations.....	11
Avis général n°6 : Mieux éclairer les débats sur les mesures de l'inflation et du niveau de vie.....	12
Avis général n°7 : Mieux prendre en compte les besoins sociaux en matière de statistiques de santé et renforcer la coordination entre producteurs pour y répondre.....	12
Avis général n°8 : Des données territoriales par et pour les territoires.....	12
Avis général n°9 : Donner un cadre à l'utilisation statistique de nouvelles sources de données.....	13
Avis général n°10 : Développer un cadre de référence et une déontologie adaptée aux appariements de données individuelles.....	13
Avis général n°11 : Faciliter l'orientation des utilisateurs dans les statistiques disponibles et poursuivre les efforts de pédagogie.....	13

V. TABLE RONDE : QUELLES RÉPONSES AUX BESOINS STATISTIQUES ?.....	14
Intervention d'Erik Zolotoukhine (Progedo / Adisp).....	14
Intervention de Kamel Gadouche (CASD).....	15
Intervention de Marc Gurgand (EquiPex Idee).....	16
Intervention de Christine Hennion (co-rapporteuse de la mission data et territoires).....	17
Intervention d'Arnaud Degorre (Insee Pays de la Loire).....	18
VI. CONCLUSION.....	20

Liste des participants

Membres du Conseil

Nom	Prénom	Organisme
AUDENIS	Cédric	France Stratégie
BARDAJI	José	France Assureurs
BENKIRANE	Aïda	CGT
BERTHOLON	Raphaëlle	Confédération française de l'encadrement - Confédération générale des cadres (CFE-CGC)
BUSI	Paul	CFDT
CABRILLAC	Bruno	Banque de France
CHARNU	Jean-Michel	CMA France
CHOUGUIAT	Alain	Confédération de l'artisanat et des petites entreprises du bâtiment (CAPEB) U2PL
CLANCHE	François	INED
DEAUVIEAU	Jérôme	Ecole normale supérieure
DOUTEAU	Georges	UNAF
DURAN	Patrice	Ecole normale supérieure
EPAULARD	Anne	Présidente de la commission Système financier et financement de l'économie
LACAILLE	Yves	Union nationale des professions libérales (UNAPL)
OFFNER	Jean-Marc	Président de la commission Territoires
OREFICE	Gianluca	Président de la commission Entreprises et stratégies de marché
POILON	Gwenaëlle	Mouvement des entreprises de France (MEDEF)
POUZIN	Gilles	CFTC
ROBIN	Benoît	Confédération générale du travail - Force ouvrière (CGT-FO)
SALA	Michel	Assemblée nationale
SERIEYX	Yvon	UNAF
TAVERNIER	Jean-Luc	INSEE - DG
TIMBEAU	Xavier	Président de la commission Environnement de développement durable
VALLA	Hugo	Fédération bancaire française
VESSILLIER	Delphine	Fédération française du bâtiment (FFB)

Chefs de SSM

Nom	Prénom	Organisme
ATTAL-TOUBERT	Ketty	Ministère de l'Economie - Service statistique ministériel des Douanes
BRIERE	Luc	Ministère de l'Intérieur - Direction générale des collectivités locales – SSM Collectivités locales
CHEVALIER	Pascal	Ministère de la Justice – SSM Justice
DE PERETTI	Gaël	Direction générale de l'administration et de la fonction publique (DGAFP)
GONZALEZ-DEMICHÉL	Christine	Ministère de l'Intérieur - Service statistique ministériel de la sécurité intérieure
GREFFET	Pierre	Ministère des Armées – SSM Défense
SCHREIBER	Amandine	Ministère de la Culture – SSM Culture
SCHUHL	Pierrette	Ministère de l'Enseignement supérieur – SSM Enseignement supérieur et recherche
SEDILLOT	Béatrice	Ministère de la Transition écologique et de la Cohésion des territoires - Service de la donnée et des études statistiques (SDES)
TOMASINI	Magda	Ministère de l'Education nationale – Direction de l'évaluation, de la prospective et de la performance (Depp)

Autres rapporteurs de commission		
Nom	Prénom	Organisme
BESSONE	Anne-Juliette	Ministère du Travail - Direction de l'animation de la recherche, des études et des statistiques (DARES)
DUEE	Michel	Institut National de la statistique et des études économiques (INSEE) – Département de l'action régionale
GERVAIS	Emmanuel	Banque de France
HOURIEZ	Guillaume	Institut National de la statistique et des études économiques (INSEE) – Département des comptes nationaux
MINODIER	Christelle	Ministère de la Santé et de la Prévention – Direction de la recherche, des études, de l'évaluation et des statistiques (Drees)
MOREAU	Sylvain	Institut National de la statistique et des études économiques (INSEE) – Direction des statistiques d'entreprises (DSE)
PASSERON	Vladimir	Institut National de la statistique et des études économiques (INSEE) - Direction des statistiques démographiques et sociales (DSDS)

Secrétariat général du Cnis		
Nom	Prénom	Organisme
BLANC	Sylvie	Secrétariat général du Cnis
CANCEL	Sébastien	Secrétariat général du Cnis
CRAVATTE	Céline	Secrétariat général du Cnis
DE RAPHELIS SOISSAN	Magali	Secrétariat général du Cnis
FRANCOZ	Dominique	Secrétariat général du Cnis
GUILLAUMAT-TAILLIET	François	Secrétariat général du Cnis
MAUREL	Françoise	Secrétariat général du Cnis
PALA	Frédéric	Secrétariat général du Cnis

Autres participants		
Nom	Prénom	Organisme
BAYET	Alain	Institut National de la statistique et des études économiques (INSEE) - Direction de la diffusion et l'action régionale (DDAR)
BIAU	Olivier	Institut National de la statistique et des études économiques (INSEE) - Direction de la méthodologie et de la coordination statistique et internationale (DMCSI)
BREUIL	Pascale	Comité du label
DEGORRE	Arnaud	Institut National de la statistique et des études économiques (INSEE) - Direction régionale des Pays de la Loire
GADOUCHE	Kamel	Centre d'accès sécurisé à distance (CASD)
GURGAND	Marc	Ecole d'économie de Paris
HENNION	Christine	Rapporteuse de la mission Data et territoires
PROST	Corinne	Institut National de la statistique et des études économiques (INSEE) - Direction de la méthodologie et de la coordination statistique et internationale (DMCSI)
QUELLEC	Jean-Michel	Institut National de la statistique et des études économiques (INSEE) – Inspection générale
REDOR	Patrick	Institut national de la statistique et des études économiques (INSEE)
ZOLOTOUKHINE	Eric	Quetelet-Progedo

La séance est ouverte à 9h35, sous la présidence de Patrice DURAN.

I. INTRODUCTION

Le Président, Patrice DURAN, rappelle que l'année 2024 sera marquée par le démarrage du programme de moyen terme 2024-2028. Ce nouveau programme est le fruit d'un long travail de consultation mené par le Cnis. La demande de statistique publique est croissante et réclame ce type de travail informatif. C'est pourquoi le Cnis mène une réflexion sur la coordination requise par les questions publiques, qui sont largement transversales. L'élaboration du programme de moyen terme par Dominique FRANCOZ a consisté à revenir sur cinq ans de travail des commissions du Cnis et des diverses institutions de la statistique publique. Cette réflexion sur le moyen terme, et sur les avis qui feront l'objet d'un vote, renvoie tant à des logiques de connaissances qu'à des logiques d'action et de contrôle.

Avant de passer à la validation de ces avis, Jean-Luc TAVERNIER reviendra sur l'histoire de la statistique publique depuis une vingtaine d'années. Il évoquera la variété des enjeux ainsi que le travail mené par l'Insee et les services statistiques ministériels. Ensuite, la question des données sera abordée lors d'une table ronde. Cette question est centrale, dans la mesure où la statistique publique pâtit d'une forme de méconnaissance sur ce sujet. Dans un premier temps, le thème des données sera examiné sous l'aspect des besoins des chercheurs en la matière, et du développement des méthodes d'accès. Dans un second temps, c'est la dimension territoriale et la question des données locales qui seront appréhendées. En effet, il est nécessaire de mener une réflexion sur les territoires, eu égard aux difficultés rencontrées par les administrations territoriales et leur faible capacité à récolter et faire usage de données solides et pertinentes.

II. 20 ANS D'INNOVATIONS AU SERVICE DE LA STATISTIQUE PUBLIQUE

Jean-Luc TAVERNIER, Directeur général de l'Insee, rappelle que la loi 51-711 du 7 juin 1951 stipule que la statistique publique repose sur deux piliers : les enquêtes statistiques menées par l'Insee, qui s'inscrivent dans le programme du Cnis, et les données administratives. La loi pour une République numérique (LRN) de 2016 a apporté un troisième pilier : les données privées. Cette importante nouveauté ne doit toutefois pas occulter les bouleversements qui se sont produits dans le domaine des enquêtes statistiques et des données administratives. Il convient d'évoquer également les évolutions en matière de diffusion et surtout de mise à disposition des données à des fins de recherche, qui sont considérables.

Jean-Luc TAVERNIER aborde les efforts permettant de réduire la charge pour le répondant lors des enquêtes. Les entreprises ont d'abord obtenu la possibilité de répondre par Internet aux enquêtes. Ensuite, le dispositif d'*Élaboration des statistiques annuelles d'entreprises* (Ésane) a permis de s'appuyer davantage sur les données fiscales et dans une moindre mesure sur les enquêtes. Ces deux initiatives se sont révélées efficaces, ainsi que l'a montré une récente consultation des entreprises destinée à nourrir une loi sur la simplification des entreprises. Des efforts similaires ont été entrepris pour les ménages afin de prévenir l'érosion du taux de réponse aux enquêtes. La polémique à propos de l'enquête sur l'emploi au Royaume-Uni témoigne des effets négatifs de ce phénomène d'érosion. La France a su y échapper par la multiplication des modes de réponses possibles, en face à face, par téléphone ou par Internet. Par ailleurs, l'Insee dispose désormais d'un programme d'enquête cohérent dans les départements et territoires d'Outre-mer, où il suréchantillonne les enquêtes de manière à affiner les résultats.

Beaucoup d'enquêtes sont anciennes, les enquêtes sur l'emploi, les conditions de vie, le budget des familles, l'emploi du temps ou le logement. De nombreuses enquêtes sont cependant plus récentes qu'elles y paraissent. Ainsi, les enquêtes sur les compétences des adultes (Piaac) et compétences des élèves (Pisa), menées dans un cadre international sous l'égide de l'OCDE, datent de moins de vingt ans. De même, des enquêtes importantes sur les évolutions de la société et les conditions de vie ont été initiées au XXI^e siècle. L'enquête *Sans-domicile*, dont la première édition remonte à 2001 et la deuxième à 2012, connaîtra une troisième édition en 2025. Les bénéficiaires de l'aide alimentaire font l'objet d'une enquête lancée en 2021 dans le contexte de la crise sanitaire, alors que ce phénomène n'avait jamais été appréhendé par un travail de statistique publique. Il convient de citer également l'enquête sur les ressources des jeunes, les deux millésimes de l'enquête *Trajectoires et origines* sur les immigrés et les descendants d'immigrés, les enquêtes sur le mode de garde, sur la rénovation énergétique des logements, ou encore sur les questions de victimation. Celles-ci ont été enrichies et ont pris une importance croissante, d'abord sous l'appellation CVS (*Cadre de vie et sécurité*), puis dans le cadre du dispositif VRS (*Vécu et ressenti en matière de sécurité*) du service statistique de la Sécurité intérieure.

La variété des approches, en outre, a mis en avant le ressenti et les questions subjectives dans l'enquête sur les conditions de vie, à travers le baromètre d'opinion de la Direction de la recherche, des études, de

l'évaluation et des statistiques (Drees), et dans les enquêtes spécifiques sur le ressenti ou les Français et la nature.

Il y a dix ans, de nombreux agents de l'Insee s'interrogeaient sur la nécessité de ce type d'enquête à l'avenir, étant donné la multiplication des traces numériques accessibles sur Internet. Or l'expérience a montré que des questionnaires dotés d'un long protocole et conçus par des professionnels restaient les plus à même d'éclairer certains faits sociaux. L'Insee doit cependant procéder chaque année à des arbitrages. En effet, les enquêtes rituelles régies par des règlements européens ou organisées sous l'égide de process internationaux ne laissent guère de place pour des enquêtes spécifiques. De plus, ces enquêtes appellent à être reprises après un certain nombre d'années.

Concernant les entreprises, si les innovations sont moindres, des questionnements spécifiques ont émergé relativement à des crises. Ainsi, l'enquête *Acemo-Covid* a montré comment les entreprises se projetaient lors des confinements. Plus récemment, une enquête a été menée auprès des fournisseurs d'électricité dans le contexte de la crise des prix de l'énergie.

Jean-Luc TAVERNIER évoque deux évolutions notables concernant les données administratives. La première tient à la veille sur les nouveaux processus administratifs. Une vigilance est nécessaire sur ce point, afin que les besoins de la statistique publique soient pris en compte dès le début de la conception de ces nouveaux processus. La *Déclaration sociale nominative* (DSN) illustre ce principe, puisque l'Insee y trouve la source de nombreuses innovations en termes de production statistique sur l'emploi et les salaires. La seconde évolution, plus importante, est le recours désormais banalisé aux appariements entre les bases de données issues de différentes sources administratives. Plusieurs dispositifs illustrent cette méthode. Depuis 2012, le dispositif *Fichier localisé social et fiscal* (Filosofi) permet d'apparier, au niveau des ménages, les revenus fiscaux et les revenus non imposables, de manière à obtenir des indicateurs d'inégalités, de revenus ou de pauvreté selon les territoires, jusqu'au niveau communal, voire infracommunal. Le *Fichier démographique sur les logements et les individus* (Fidéli), sert quant à lui de base de sondage. Une base de données fiscales et du registre du commerce et des sociétés, Fidelimmo, appariée avec des données cadastrales, permet de connaître la concentration de la propriété immobilière en France. Enfin, le projet Resil (*Répertoire statistique d'individus et de logements*) pallie la disparition de la taxe d'habitation.

D'autres appariements portant sur l'évaluation de l'enseignement et de la formation professionnelle ont été effectués, parmi lesquels le dispositif ForCE (*Formation, chômage, emploi*), portant sur les débouchés des bénéficiaires de dispositifs de formation professionnelle, ou le panel Trajam (*Trajectoires des jeunes appariées aux mesures actives du marché du travail*) sur les bénéficiaires d'emplois aidés. Récemment, le service InserJeunes a été créé afin de mesurer l'employabilité des jeunes ayant suivi une formation professionnelle, et s'est généralisé. Quant à InserSup, cet outil appréhende les débouchés d'un point de vue statistique, par appariement des données sur les différentes filières de l'enseignement supérieur.

Dans le domaine de la santé, des appariements entre l'*Echantillon démographique permanent* et le *Système national des données de santé* (SNDS) permettent de suivre les parcours de soins de façon longitudinale. Enfin, des appariements entre enquêtes et données administratives sont pratiqués, à l'exemple de l'*Enquête revenus fiscaux et sociaux* (ERFS) qui rend possibles des mesures sur la distribution des revenus et les inégalités.

Jean-Luc TAVERNIER indique que les évolutions notables dans le domaine des données privées sont plus récentes. La loi pour une République numérique de 2016 a donné un cadre juridique pour l'accès aux données privées. Au niveau européen, un règlement similaire est en construction. Le premier cas permis par cet accès aux données privées est celui des données de caisses de la grande distribution, qui a permis de calculer plus de 10 % de l'indice des prix depuis 2020. La crise s'est avérée le catalyseur d'un certain nombre de coopérations. Ainsi, les données de transactions de cartes de crédit ont contribué à appréhender la chute de la consommation lors du premier confinement, dès mars 2020. Par la suite, deux publications, en coopération avec la Banque postale, ont eu pour objet des indicateurs de précarité financière à partir d'un recueil de données de comptes bancaires. L'Insee a par ailleurs pu mesurer les déplacements des Parisiens lors du premier confinement à partir des données de téléphonie mobile, bien que ce partenariat ait rencontré des difficultés pour s'inscrire dans la durée. D'autres partenariats ont été conclus avec des plateformes de services telles que Airbnb ou Booking, et ont donné lieu à des travaux pionniers.

La méthode de travail sur ces données privées a évolué au fil du temps. L'import des volumineuses données de caisses dans le système d'information de l'Insee a ainsi nécessité une organisation particulière. Par la suite, l'Insee a plutôt travaillé directement dans le système d'information du partenaire privé en s'efforçant d'y appliquer ses algorithmes et ses requêtes.

Jean-Luc TAVERNIER rappelle que si l'exercice de la diffusion a considérablement évolué depuis vingt ans, l'approche de l'Insee n'a pas varié dans ce domaine. L'Institut garde pour exigence d'être le premier utilisateur de ce qu'il produit. Autrement dit, à l'image des services statistiques ministériels, l'Insee ne diffuse pas de chiffres sans les contextualiser et les commenter.

Il convient de remarquer que les délais de production et de diffusion se sont raccourcis à la demande des utilisateurs, notamment les utilisateurs en provenance des marchés. Le contrôle qualité, l'organisation des jeux de données et l'accès à ces données se sont nettement améliorés. L'articulation avec les métadonnées, de manière à s'assurer de la qualité et de la cohérence des données, s'est fortement accrue. Les vecteurs de diffusion ont naturellement changé, le site internet ayant progressivement remplacé le papier, désormais résiduel. Des innovations en matière de data visualisation, la conclusion de partenariats avec des sociétés de production de vidéos ou dessins animés, et l'usage des réseaux sociaux permettent aujourd'hui de s'adresser à un public large.

Si les données macros sont naturellement en open data, la mise à disposition des données individuelles est entrée dans une nouvelle ère. Auparavant, pratiquer de la microéconométrie sur des données individuelles était réservé aux cadres de l'Insee et à leurs co-auteurs. Désormais, cette mise à disposition s'est formalisée, à la faveur du recours au comité du secret, avec l'ouverture d'un portail dédié, et grâce à l'assurance que sont respectées les règles de protection des données grâce au Centre d'accès sécurisé aux données (CASD). Ces innovations ont permis d'accroître considérablement le volume de données mises à disposition.

En conclusion, Jean-Luc TAVERNIER indique que ces différentes évolutions ont été conduites afin de remplir six objectifs. Le premier est de limiter la charge d'enquête. Le deuxième est de réduire les délais de production et de diffusion, et de se montrer réactif, par exemple en cas de crise, afin de réaliser des enquêtes ad hoc. Le troisième est de rendre possible l'évaluation des politiques publiques à travers les appariements et la mise à disposition de données pour la recherche. Quant aux trois derniers objectifs, ce sont les objectifs cardinaux soutenus par le Cnis depuis de nombreuses années. Le premier est de sortir de la dictature de la moyenne, ce que l'on peut estimer désormais acquis. Le deuxième est d'améliorer le suivi des trajectoires afin de ne pas se contenter de photographies instantanées. L'Insee s'y efforce grâce à des enquêtes panel ou des infrastructures telles que *l'Echantillon démographique permanent*. En outre, le renforcement de la capacité à détenir un identifiant, à l'image de *l'Identifiant national élève* (INE) commun à l'éducation nationale et à l'enseignement supérieur, ou, prochainement, de la chaîne reliant les délits et les réponses pénales, est primordial. Enfin, le dernier objectif consiste à produire des résultats au niveau des territoires. Or, à l'exception des Outre-mer, l'Insee ne dispose plus des moyens requis par des enquêtes portant sur des échantillons suffisamment larges pour être pertinentes au niveau infranational. C'est la raison pour laquelle le recours aux données administratives et à des méthodes d'estimation sur petits domaines s'avère précieux.

Benoît ROBIN, représentant de FO, fait remarquer que certaines enseignes ne collaborent pas à la mesure de l'évolution des prix à travers l'analyse des tickets de caisse. Bien que les modèles en tiennent compte, ce manque de coopération peut créer des biais. Par ailleurs, à titre d'information, il convient d'observer que le positionnement et l'évolution de la statistique publique française ont contribué à faire évoluer la mentalité dans d'autres enceintes. Ainsi, l'OCDE s'apprête à rendre librement accessible depuis un portail dédié l'ensemble de ses rapports, de ses bases de données et de ses statistiques.

Xavier TIMBEAU, directeur principal de l'OFCE, relève la difficulté d'obtenir et d'utiliser des données numériques de qualité, par exemple les données des opérateurs mobiles. Il demande si un projet d'encadrement des statistiques publiques intégrant des traces numériques est à l'étude.

Jean-Luc TAVERNIER répond à Benoît ROBIN à propos des enseignes de la grande distribution. Il rappelle que la loi prévoyait des sanctions à l'encontre des enseignes qui refuseraient de collaborer. Les deux principales enseignes de *hard discount* répondent désormais aux enquêtes. Dès lors, le risque de biais relatif au manque de collaboration semble désormais négligeable.

En réponse à la question sur les traces numériques, Jean-Luc TAVERNIER indique que l'Insee pratique le *scraping* avec les données qui sont publiques. Ainsi, la Direction de l'animation de la recherche, des études et des statistiques (Dares) mesure les vacances d'emploi à partir d'informations captées sur des sites internet. Dans le contexte de la crise du Covid, travailler à partir des traces numériques, sur Google Maps par exemple, s'avérait stimulant. Néanmoins, pérenniser une collaboration avec des entreprises telles que Google afin de mener un travail en routine semble d'une pertinence douteuse. L'investissement sur les données téléphoniques semble nettement plus prometteur, mais se heurte à de redoutables problèmes d'ordre juridique et financier. Sur le plan financier, la difficulté tient à la nécessité de trouver une entente

financière avec les entreprises valorisant les données téléphoniques, dont l'utilisation à des fins de statistiques est susceptible de réduire la valeur de marché.

III. UN NOUVEAU MOYEN TERME POUR 2024-2028

Dominique FRANCOZ, du Secrétariat général du Cnis, présente le processus ayant abouti à l'élaboration du programme de moyen terme pour 2024-2028. À partir d'octobre 2022, il a été procédé à un bilan du programme de moyen terme précédent. De janvier à septembre 2023, une consultation a été menée auprès de nombreux interlocuteurs, ainsi qu'une consultation publique sur le site du Cnis. Ces consultations ont débouché sur l'organisation de deux *focus groups*. Par la suite, le bilan du moyen terme précédent a été présenté aux commissions thématiques du printemps 2023, et les premières pistes du prochain moyen terme ont été évoquées. Les avis rédigés de chaque commission ont été examinés dans le cadre des commissions d'automne. Enfin, les avis généraux rédigés ont été soumis à validation lors du bureau du 13 décembre. Le programme de moyen terme sera présenté à l'occasion d'un séminaire se tenant le 3 avril 2024. Ce séminaire sera organisé autour de deux tables rondes, l'une portant sur l'inflation et le niveau de vie, l'autre sur l'impact du numérique sur les entreprises.

Les consultations ont représenté 44 entretiens avec 21 producteurs nationaux, une direction d'Eurostat, neuf représentants du bureau ou du conseil, neuf utilisateurs institutionnels et quatre acteurs locaux. La consultation publique a permis de recueillir 150 réponses exploitables et les coordonnées de 62 personnes, qui ont été recontactées en vue de l'organisation des deux *focus groups*, l'un portant sur la transition écologique, l'autre sur les statistiques sociales.

Cette réflexion a été menée dans un contexte marqué par trois éléments. Premièrement, il convenait de prendre en compte les transformations majeures de l'économie et de la société, à savoir la transition écologique, la transformation numérique et le vieillissement de la population. Deuxièmement, des thématiques restaient d'actualité, en l'occurrence les inégalités, les problématiques territoriales et la santé. Troisièmement, la statistique publique est confrontée au triple défi de la multiplication des sources d'information, de la nécessité d'éclairer plus rapidement des phénomènes nouveaux tels que la crise sanitaire, et du besoin accru de coordination des producteurs afin de traiter des questions transverses.

Le choix des thèmes résulte d'une part des besoins exprimés par les utilisateurs, et d'autre part des thèmes du précédent moyen terme s'avérant toujours d'actualité. À l'issue des commissions de printemps, deux réunions avec le Président et les rapporteurs de commissions ont permis d'abord d'identifier les thèmes la rédaction des avis, et ensuite de finaliser la rédaction de ces avis présentés et révisés à l'occasion des commissions d'automne. La consultation publique a également mis en avant des thèmes tels que la santé, les inégalités, la transition écologique et les territoires.

Le programme de moyen terme 2024-2028 se décompose en onze avis généraux et 69 avis de commissions thématiques. Les avis généraux s'inscrivent dans la continuité des thématiques déjà abordées, fédèrent les préoccupations de plusieurs commissions et reflètent des thèmes d'actualité. Ces avis ont été classés en trois grandes thématiques : rendre compte d'un monde qui se transforme, éclairer les débats dans une société complexe, innover pour mieux répondre aux besoins. Treize avis qui n'avaient pas été traités dans le précédent programme de moyen terme ont été repris à l'identique. De nouvelles priorités ont été identifiées, ayant trait à la transition écologique, à l'inflation et au niveau de vie, à la santé et aux enjeux générationnels. Des thèmes communs à plusieurs commissions ont également été relevés, à savoir la transition écologique, la santé, les territoires, les inégalités et les enjeux générationnels.

Nombre d'avis de commission se rattachent aux avis généraux, à l'exception de neuf thèmes : l'économie sociale et solidaire, la maltraitance, les proches aidants, le télétravail, l'économie bleue, la formation, la Nomenclature d'activités française (NAF) 2025 et les services à la personne. Certains avis de commission portent sur des thématiques nouvelles ou qui n'ont pas été traitées dans les dernières années. Ainsi, un avis de la commission démographie et questions sociales et un avis de la commission territoire portent sur l'inflation. La commission emploi a émis un avis sur le télétravail et un avis sur le travail non rémunéré. La commission démographie et questions sociales a émis un avis sur la maltraitance des personnes vulnérables et un avis sur les proches aidants. La commission environnement et développement durable a émis un avis sur la ressource en eau et un avis sur la sobriété. Enfin, la commission entreprise a émis un avis sur l'économie bleue.

IV. DISCUSSION ET VOTE DES AVIS GÉNÉRAUX DU MOYEN TERME 2024-2028

Françoise MAUREL, Secrétaire générale du Cnis, signale qu'une demande d'amendement, formulée par l'Union nationale des associations familiales (Unaf), a été diffusée en amont de l'assemblée plénière. Elle précise que seuls les avis généraux feront l'objet d'un vote avis par avis. Après ce vote, les avis de commissions, préalablement validés par leur président, feront l'objet d'un vote commun, commission par commission. À cette occasion, les avis spécifiques à 2024 seront soumis à un vote.

Dominique FRANCOZ présente les avis généraux soumis au vote des membres du Conseil.

Avis général n° 1 : Mesurer la transition écologique et rendre compte de ses enjeux économiques et sociaux.

Le Conseil salue les travaux de la statistique publique réalisés pour suivre la transition écologique et les mesures prises pour limiter les effets du réchauffement climatique. Il souligne en outre la nécessité de poursuivre ces travaux et de compléter les statistiques actuelles dans une approche systémique.

Le Conseil demande que les statistiques macroéconomiques prennent davantage en compte les atteintes à l'environnement, au-delà de la performance économique et dans un cadre conceptuel cohérent avec la comptabilité nationale. Il recommande d'enrichir les comptes économiques de l'environnement.

Il souhaite également que les statistiques publiques puissent éclairer les différences de situations et de comportements des agents économiques et des territoires vis-à-vis de la transition écologique (émissions de CO₂, coût et bénéfices de la transition...), et constituent des référentiels pérennes permettant d'éclairer les choix publics et privés.

Avis adopté à l'unanimité des présents.

Avis général n° 2 : Mesurer la transformation numérique de l'économie et de la société.

Les processus numériques modifient en profondeur les échanges économiques et sociaux, rendant certaines métriques obsolètes, telles les approches statistiques sectorielles du numérique.

Le conseil recommande que les statistiques publiques permettent de suivre la numérisation de l'économie et ses impacts, dans un cadre harmonisé au plan international. Il invite notamment à compléter les statistiques macroéconomiques au regard des nouveaux biens et services numériques et à observer la transformation numérique des entreprises. Il encourage également les mesures des compétences numériques des personnes et de l'impact du numérique sur le bien-être et les conditions de vie, ainsi que sur l'accès aux services publics.

Avis adopté à l'unanimité des présents.

Avis général n° 3 : Inscrire la statistique publique dans une perspective internationale.

Le phénomène mondial des économies en crise et les enjeux environnementaux augmentent le besoin de placer les travaux de la statistique publique dans une perspective plus large, européenne voire internationale.

La statistique publique doit continuer de prendre une part active dans les travaux internationaux et européens qui conduisent à la définition de normes statistiques s'imposant aux producteurs. Dans ce cadre, le Conseil souhaite que les besoins des utilisateurs de statistiques nationales et européennes soient davantage pris en compte dans l'élaboration des programmes statistiques européens.

Le Conseil préconise d'approfondir et d'élargir le champ des travaux de comparaisons européennes ou internationales et de veiller, à cet effet, à améliorer et documenter la comparabilité des statistiques à une échelle européenne, voire internationale.

Aïda BENKIRANE, représentante de la CGT, salue cet avis, qui exprime un souhait partagé par de nombreux utilisateurs de statistiques nationales et européennes. Elle remercie le Directeur général de l'Insee, qui présente régulièrement des points précis sur l'actualité statistique nationale et européenne. Consciente des contraintes de calendriers, elle souhaite néanmoins que la préparation des règlements européens fasse l'objet d'un point spécifique à l'ordre du jour du bureau et des commissions compétentes. Ainsi les membres du Cnis pourraient être informés des enjeux et donner leur avis sur ces sujets. De telles

dispositions auraient pu être aménagées au printemps 2023 à propos du règlement 223 de la loi statistique européenne.

Patrice DURAN indique qu'il sera tenu compte de ce souhait. Il ajoute que le Cnis est désormais représenté au niveau européen par la voix de Roxane Silberman, nommée présidente du Comité consultatif européen de la statistique (Esac). La question européenne représente un enjeu central et suscite des interrogations en matière de comparabilité des données et des concepts, et de méthodologie de manière générale.

Yvon SÉRIEYX, représentant de L'UNAF, se félicite également de cet avis.

Jean-Luc TAVERNIER souscrit à ces propos. En réponse à Aïda BENKIRANE, il fait observer qu'il est difficile d'associer de nombreux interlocuteurs aux discussions sur les règlements européens. Cependant, il convient de l'intérêt d'inscrire à l'ordre du jour d'un prochain Bureau un point sur le cas du règlement en cours de négociation sur les statistiques démographiques de population (Esop). Cette négociation mérite d'être analysée afin en particulier d'identifier les biais dans la manière dont les demandes d'utilisateurs sont collectées par la Commission européenne.

Avis adopté à l'unanimité des présents.

Avis général n° 4 : Rendre compte des enjeux générationnels dans les statistiques.

Le vieillissement de la population a des répercussions sur l'ensemble de la société et sur l'économie. A l'autre bout de l'échelle, les jeunes (dont les enfants) ont subi de plein fouet la crise sanitaire avec des répercussions aussi bien sur leurs conditions de vie que sur leur parcours dans le système éducatif et leur insertion dans la vie professionnelle.

Le Conseil demande que les jeunes et les personnes âgées, des seniors au grand âge, fassent l'objet de travaux approfondis et que les statistiques soient mobilisées pour étudier des sujets tels que la pauvreté des jeunes (en particulier des étudiants), les conditions de vie par âge, le travail des seniors, la santé et la dépendance. Il souhaite également que les conséquences économiques, sociales et sociétales du vieillissement de la population soient mieux documentées.

Il recommande que les statistiques couvrent mieux les populations des jeunes et des plus âgés, notamment par l'inclusion plus systématique des personnes vivant en collectivité (établissements médico-sociaux, résidences universitaires, etc.).

Yvon SÉRIEYX souhaite clarifier la situation des foyers d'accueil collectif de l'Aide sociale à l'enfance (ASE), et vérifier qu'ils sont bien compris parmi les établissements médico-sociaux. De même, il se demande si les apprentis sont bien inclus dans la catégorie des étudiants. Il propose le terme de « jeunes en formation initiale » afin d'être certain de n'oublier aucune catégorie de jeunes. De manière générale, il fait part de son souci d'inclure dans le champ d'études des enfants et des jeunes de moins de 18 ans. De très bons exemples d'inclusion des enfants dans les études existent. Ainsi de l'enquête Autonomie, qui inclut les jeunes aidants, ou de l'enquête sur l'alimentation. À l'inverse, certaines enquêtes telles que l'enquête Emploi du temps (11 ans ou plus), l'enquête TIC (15 ans ou plus) sur l'exposition aux écrans ne prennent pas en compte la situation des jeunes enfants. De même, l'enquête globale Transport (5 ans ou plus) fait l'impasse sur les micromobilités des très jeunes enfants. En somme, il convient d'explicitier davantage l'inclusion des enfants dans le champ.

Françoise MAUREL rappelle qu'un avis général est par définition rédigé à un certain niveau de généralité. Cet avis, en outre, a vocation à être décliné dans des avis de commissions thématiques. Il reviendra ensuite aux producteurs statistiques de se saisir de l'esprit de ces avis et de leur donner des champs pertinents.

Avis adopté à l'unanimité des présents.

Avis général n° 5 : Enrichir les travaux sur les inégalités et les discriminations.

Le Conseil se félicite des progrès réalisés en matière d'observation des inégalités, sur les plans macroéconomique et microéconomique.

Le Conseil recommande de poursuivre et développer les mesures des inégalités de toutes natures en mettant l'accent sur les inégalités d'accès (accès aux services publics, aux équipements, aux soins, fracture numérique...) ainsi que sur le cumul de différents types d'inégalités, leur évolution dans le temps et la reproduction des inégalités entre les générations.

Le Conseil sera attentif au suivi des conclusions du groupe de travail du Cnis sur l'observation des discriminations.

Avis adopté à l'unanimité des présents.

Avis général n° 6 : Mieux éclairer les débats sur les mesures de l'inflation et du niveau de vie.

La mesure de l'inflation par l'indicateur statistique principal qu'est l'indice des prix à la consommation donne lieu à des incompréhensions récurrentes du public. Si les conventions internationales qui le fondent ne sont pas contestables, un seul indicateur ne peut à lui seul résumer la diversité des modes de consommation ni leur évolution.

Le Conseil souhaite que la statistique publique puisse apporter des approches complémentaires à la mesure de l'inflation pour éclairer des sujets en débat, notamment la composition du panier de biens et services retenus et les transformations de la consommation (transition écologique, numérisation de l'économie) ou encore les impacts individuels différenciés de l'inflation sur les niveaux de vie (selon la situation financière, la composition du ménage, le type de consommation, la localisation, le type d'habitat...).

Un intervenant se félicite que soient mentionnées dans l'avis la situation financière et la composition du ménage. Il avait suggéré d'ajouter la situation professionnelle. Une récente enquête de la Banque de France indique que, en réaction à l'inflation, 30 % des ménages ont augmenté leurs revenus et 10 % ont entamé une activité professionnelle supplémentaire. Autrement dit, les marges de manœuvre face à l'inflation varient selon les quotités de travail. Cela montre la pertinence de s'interroger sur la situation professionnelle des ménages.

Avis adopté à l'unanimité des présents.

Avis général n° 7 : Mieux prendre en compte les besoins sociaux en matière de statistiques de santé et renforcer la coordination entre producteurs pour y répondre.

La crise sanitaire a renforcé le besoin de disposer de données fiables, rapides et cohérentes sur l'état de santé de la population et l'offre de soins. Constatant plus généralement la difficulté des utilisateurs à appréhender l'ensemble des informations produites par les différents acteurs de la statistique publique, de la recherche épidémiologique ou d'autres administrations et leur articulation, le Conseil lancera, à l'issue de l'exercice d'état des lieux mené par le Système Statistique Européen, une concertation spécifique ou un groupe de travail, en vue de préciser le rôle de la statistique publique en matière de coordination des statistiques de santé, d'augmenter la visibilité des données existantes, de cerner les besoins statistiques insuffisamment couverts et de faire des propositions pour y répondre.

Patrice DURAN remarque que la demande en matière de santé est largement transversale. Le rôle de la statistique consiste à préciser la coordination entre des thèmes, des champs, des statistiques, etc. L'ère du fonctionnement des données en silos est terminée.

Avis adopté avec une abstention.

Avis général n° 8 : Des données territoriales par et pour les territoires.

Le Conseil note avec satisfaction les progrès réalisés dans l'élaboration d'outils qui permettent la mise à disposition d'informations plus riches et à des échelles territoriales plus fines avec un niveau de qualité suffisant tout en ayant conscience des enjeux de respect de la confidentialité.

Le Conseil recommande de poursuivre les efforts engagés pour produire plus systématiquement des statistiques à des échelles géographiques fines qui ne suivent pas nécessairement les découpages administratifs, avec un niveau de qualité suffisant.

Afin de pouvoir mieux comparer les territoires, le Conseil invite les producteurs nationaux ou locaux de statistiques à se coordonner et à améliorer la qualité des données produites (en matière notamment de couverture et d'harmonisation), et soutient le partage de bonnes pratiques entre producteurs.

Le Conseil recommande qu'une attention particulière soit apportée au suivi statistique de la transition écologique dans les territoires.

Avis adopté à l'unanimité des présents.

Avis général n° 9 : Donner un cadre à l'utilisation statistique de nouvelles sources de données.

La crise sanitaire a renforcé l'intérêt de la mobilisation de l'ensemble des données disponibles, et notamment les sources privées, pour produire de manière réactive des statistiques adaptées et pertinentes.

Le Cnis demande que les administrations diffusant des statistiques d'intérêt général se coordonnent avec le service statistique public en matière de méthodes et de bonnes pratiques.

Il invite également le service statistique public et les détenteurs de données privées à établir des partenariats pour le partage de ces données en vue de développer de nouvelles statistiques en tant que de besoin. Il recommande de mettre en place un cadre adapté pour pérenniser l'utilisation des sources privées dans le contexte de l'évolution de la réglementation européenne. Le Conseil contribuera à assurer une concertation de qualité en amont sur les besoins et usages pertinents des données.

Avis adopté à l'unanimité des présents.

Avis général n° 10 : Développer un cadre de référence et une déontologie adaptée aux appariements de données individuelles.

Encouragé de longue date par le Cnis, le recours aux informations issues de sources administratives et aux appariements de données individuelles constitue aujourd'hui un moyen de développer et d'enrichir le système d'information statistique avec de multiples bénéfices pour la connaissance (croisement de thématiques, suivi longitudinal, précision augmentée...).

Le Conseil invite à harmoniser et standardiser les méthodes d'appariements dans un objectif de qualité.

Dans un contexte de facilité technique croissante des appariements de sources exhaustives, qui peuvent présenter des risques pour la sécurisation des données personnelles, il recommande de définir et de mettre en place un cadre de référence et une déontologie adaptés à ces traitements de données, pour en assurer un haut niveau de sécurité et une bonne information du public, dans la suite des recommandations du groupe de concertation RESIL.

Avis adopté à l'unanimité des présents.

Avis général n° 11 : Faciliter l'orientation des utilisateurs dans les statistiques disponibles et poursuivre les efforts de pédagogie.

Dans un contexte d'abondance d'informations chiffrées et de concurrence croissante sur le marché de l'information, les statistiques publiques doivent avoir valeur de référence et être aussi suffisamment visibles et accessibles.

Le Conseil souhaite que les utilisateurs puissent être orientés dans l'ensemble des statistiques publiques disponibles, en fonction de leurs besoins et de leurs usages.

Pour augmenter la confiance du public dans les statistiques officielles, il encourage également les démarches de communication proactives à destination des non-experts.

Un intervenant estime que la proactivité devrait concerner des chercheurs qui ne sont pas experts du domaine couvert par telle ou telle banque de données. Ainsi, de nombreux chercheurs sont susceptibles d'être intéressés par l'enquête Emploi sans être eux-mêmes économistes ou experts de cette question.

Patrice DURAN répond que le terme non expert recouvre plusieurs réalités. De manière générale, les thématiques économiques convoquent aujourd'hui des perspectives d'analyse qui relèvent largement de l'ensemble des sciences sociales. Daniel Cohen rappelait en assemblée plénière un an plus tôt qu'il est difficile d'établir un diagnostic général sur des questions sociales sans faire appel à des modes de raisonnement différenciés.

Avis adopté à l'unanimité des présents.

Françoise MAUREL indique que le nombre élevé d'avis des commissions implique de ne pas examiner chacun d'eux en détail. Elle s'enquiert d'éventuelles remarques dans l'assistance et propose de les adopter.

Aïda BENKIRANE, représentante de la CGT, salue la bonne organisation des commissions, mais relève que les documents préparatoires ont été parfois envoyés peu de temps avant les commissions, ce qui a rendu la préparation délicate.

Les avis des commissions sont adoptés.

V. TABLE RONDE : QUELLES RÉPONSES AUX BESOINS STATISTIQUES ?

Patrice DURAN énonce les deux thèmes qui seront développés au cours de la table ronde, d'une part les besoins des chercheurs en microdonnées et le développement des modalités d'accès à ces données, et d'autre part les nouveaux besoins en matière de données locales et leur mise à disposition. Il présente les intervenants : Erik ZOLOTOUKHINE, qui dirige, au sein de l'Infrastructure de recherche (IR) Progedo, les Archives de données issues de la statistique publique (Adisp) ; Kamel GADOUCHE, directeur du Centre d'accès sécurisé aux données (CASD) ; Marc GURGAND, professeur à l'École d'économie de Paris et coordonnateur scientifique de l'EquiPex Innovations, données et expérimentations en éducation (Idee) ; Christine HENNION, ancienne députée et co-rapporteuse de la mission Data et territoires ; Arnaud DEGORRE, directeur de l'Insee Pays de la Loire.

Intervention d'Erik Zolotoukhine (Progedo/Adisp)

Erik ZOLOTOUKHINE rappelle que la création de l'IR Progedo a été validée dans le cadre de la feuille de route du ministère de l'Enseignement supérieur et de la Recherche en 2012. Cependant, Progedo s'inscrit dans une histoire plus longue. Sans remonter aux années 1980 et 1990, période durant laquelle une première convention entre des chercheurs et l'Insee a été signée, il convient de rappeler l'importance du rapport de Roxane Silberman de 1999, établissant les bases de ce qui allait devenir le Réseau Quetelet, dont les missions sont désormais reprises et coordonnées par Progedo. L'activité de Progedo est résumée par sa devise officielle : « développer la culture des données ». Concrètement, cette activité s'articule autour de deux axes principaux. D'une part, Progedo participe et soutient financièrement plusieurs grandes enquêtes internationales telles que Share, *European social survey* (ESS) ou *Generations and gender programme* (GGP). D'autre part, Progedo coordonne les activités de collecte, de documentation et de diffusion de données quantitatives en sciences humaines et sociales, *via* Quetelet-Progedo-Diffusion, anciennement Réseau Quetelet.

En termes de ressources, Quetelet-Progedo-Diffusion s'appuie, au niveau national, sur le service de l'Adisp intégré à l'unité Progedo du CNRS, et sur le service des enquêtes de l'Institut national d'études démographiques (Ined). Au niveau local, Progedo a œuvré à la création d'un maillage territorial constitué par les Plateformes universitaires de données (PUD). 17 PUD, à ce jour, sont implantées dans des universités sur le territoire métropolitain et sont constituées d'un référent scientifique et d'un ou deux ingénieurs. Progedo est par ailleurs partenaire du CASD.

Quetelet-Progedo-Diffusion a pour vocation de diffuser des données quantitatives, en particulier des Fichiers production et recherche (FPR), spécifiquement constitués par les producteurs pour répondre aux besoins des communautés de recherche. Ces FPR représentent une offre complémentaire et intermédiaire entre d'une part, les données anonymisées disponibles en open data, par exemple sur le site de l'Insee et sur les sites des services statistiques ministériels, et d'autre part les données confidentielles accessibles uniquement *via* des accès sécurisés, comme le CASD. Les principales différences entre ces offres tiennent au niveau de pseudonymisation, à la granularité et au détail de certaines variables.

À la suite des récentes évolutions réglementaires, dont le Règlement général pour la protection des données (RGPD) et la loi pour une République numérique (LRN), la diffusion de ces FPR a été redéfinie dans l'avis du 14 décembre 2018 du Comité du secret statistique (CSS). Ces données restent, comme auparavant, réservées à la recherche, mais sont désormais diffusées après habilitation du demandeur par le CSS.

Quetelet-Progedo-Diffusion diffuse exclusivement à la communauté scientifique des chercheurs, des doctorants et des étudiants de master, en France ou à l'international. Cette diffusion s'effectue en accès

restreint, après inscription sur le portail de commande Quetelet-Progedo et signature d'un engagement de confidentialité validé et contresigné par le CSS.

Le catalogue de Quetelet-Progedo-Diffusion réunit les catalogues de l'Adisp et de l'Ined. Le premier est principalement constitué de données de la statistique publique. 75 % de ces données sont produites par l'Insee, 15 % par les services statistiques ministériels, notamment la Dares et la Drees, et 10 % par d'autres producteurs. Le nombre de références actuel s'élève à 1 500. En outre, en cohérence avec la LRN et le mouvement de la science ouverte, l'Adisp est depuis quelques années missionnée pour collecter les données de la recherche académique au sein du CNRS, des universités et des divers instituts de recherche. Quant au catalogue de l'Ined, il est constitué de 70 enquêtes produites par l'Institut.

Au quotidien, le rôle de Quetelet-Progedo-Diffusion est de collecter des données auprès des producteurs, de les faire connaître et de favoriser leur diffusion auprès des chercheurs. Au niveau national, les ingénieurs de l'Adisp et de l'Ined structurent, selon la norme internationale Data documentation initiative (DDI), la documentation proposée, en particulier sur le catalogue en ligne de Progedo. Au niveau local, les ingénieurs des PUD communiquent et valorisent les données auprès des équipes de recherche.

En 2023, 1 159 demandes ont été formulées, et le nombre d'utilisateurs s'élevait à 951. Ces chiffres ont doublé en dix ans, le nombre de références ayant plus que triplé sur la période. Les utilisateurs sont essentiellement des chercheurs et des étudiants, pour la plupart économistes et sociologues, et la moitié appartient à des institutions franciliennes.

Les principales perspectives de Progedo sont l'élargissement du catalogue aux données académiques, ainsi que, avec l'appui du Service statistique public, l'extension des champs thématiques des données FPR, par exemple sur les questions environnementales, encore peu présentes, mais qui intéressent de nombreux chercheurs.

Intervention de Kamel Gadouche (CASD)

Kamel GADOUCHE évoque les environnements de calcul sécurisé du CASD. Les données relèvent de quatre catégories : les données brutes, les données confidentielles, les FPR et les données en open data. Brutes, qu'elles soient des données administratives ou des données collectées par enquête, elles comprennent les noms et prénoms des personnes. Les données confidentielles, soumises à un risque élevé, sont nommées dans la terminologie du règlement européen 557 de 2013 *secure-use files*. Les noms et prénoms sont retirés et l'adresse exacte est remplacée par le nom d'un îlot, d'un quartier ou d'une ville, cependant il reste possible de retrouver des personnes. Dans les FPR, un autre niveau détail est retiré. Enfin, les fichiers open data sont théoriquement anonymisés. Le CASD a pour objet les données du deuxième type, c'est-à-dire les données confidentielles.

La gouvernance du Service statistique public s'appuie sur la loi 51-711 qui, outre la définition des structures de gouvernance, trouve sa traduction concrète dans la production des données et l'organisation de leur diffusion. En effet, ses articles précisent et permettent la collection des données servant à produire des études, les données des citoyens dans l'article 1 bis, et les données des entreprises dans l'article 3 bis. L'article 7 bis permet au service public de récupérer des données administratives, avec comme condition de ne pas les rétrocéder au service qui les ont transmises. Enfin, les articles 6 et 7 bis autorisent la mise à disposition des données ainsi collectées en dehors de la sphère du Service statistique public. Cette permission est effective à un double niveau, au niveau décisionnaire par le Comité du secret statistique, et au niveau pratique par un dispositif d'accès sécurisé tel que le CASD.

Le CASD est un groupement d'intérêt public à but non lucratif qui rassemble l'État, représenté par l'Insee, le Groupe des écoles nationales d'économie et statistique (Genes), la Banque de France depuis 2023, le CNRS, HEC et l'École Polytechnique. Le CASD emploie une trentaine de personnes et dispose de 500 sources de données, représentant environ 600 téraoctets de données. Plus de mille institutions recourent à ses services, en France et à l'étranger.

Le CASD propose des environnements de calcul sécurisés afin d'opérer des traitements sur ces données confidentielles. Il dispose d'une offre logicielle parmi les plus larges au monde, avec plus de 80 logiciels en catalogue. Il apporte un service de support technique, de sauvegarde et de localisation multisites. Enfin, il assure le contrôle de la confidentialité des sorties, la documentation et l'accompagnement pour les démarches et les appariements.

L'utilisateur recourt à un boîtier d'accès dédié et à une carte biométrique afin d'accéder à distance à un environnement cloisonné de calcul, d'où les données ne peuvent sortir sans une intervention manuelle des *data managers* du CASD. En termes de retour des utilisateurs, le niveau de satisfaction mesuré par des enquêtes est très élevé. Le CASD a été le premier organisme français certifié sur la protection des données sur la norme RGPD. Il est également homologué par le Système national des données de santé et dispose d'une autorisation de la Commission nationale de l'informatique et des libertés (CNIL).

Les données des producteurs concernent de nombreux secteurs, tels que la santé, la sécurité, la justice, les revenus, l'éducation, la formation, le logement ou l'économie de manière générale. La gouvernance du CASD permet une procédure d'accès unifiée à ces données *via* le Comité du secret statistique, ce qui représente une spécificité enviable par rapport à des structures homologues à l'étranger. Actuellement, deux tiers des 700 projets menés sur le CASD s'appuient sur des données de plusieurs producteurs. 10 % des projets concernent des études localisées sur un ou plusieurs territoires.

La mise en place du CASD s'est appuyée sur un soutien initial et continu des producteurs, en particulier de l'Insee, du Cnis et du Comité du label de la statistique publique. Les progrès considérables en matière d'environnement de calcul sécurisé ont été permis par la gouvernance, dont découle ce dispositif intersectoriel prévenant tout effet de silo. Ce dispositif est relativement unique au monde à cette échelle.

Intervention de Marc Gurgand (EquiPex Idee)

Marc GURGAND indique que l'EquiPex Idee, financé par l'Agence nationale de la recherche (ANR), a pour vocation de soutenir la recherche en éducation. Une de ses missions est d'améliorer et de renforcer l'accès des chercheurs aux données d'éducation. Cette mission s'exerce dans le cadre d'un partenariat très étroit avec la Direction de l'évaluation, de la prospective et de la performance (Depp). D'une part, l'EquiPex permet, *via* le financement de l'ANR, à des personnes de travailler sur le projet au sein de la Depp. D'autre part, il finance de la prestation de service informatique afin de développer le projet. De plus, un comité des utilisateurs représente la communauté des chercheurs travaillant sur l'éducation et guide l'action de l'EquiPex.

Le projet bénéficie des données exceptionnelles de la Depp. Les importantes opportunités d'appariements internes à ces données placent la recherche française en éducation à un niveau de capacité de production très élevé à l'échelle des standards internationaux. Un long travail d'interaction entre les chercheurs et la Depp a permis notamment la mobilisation de nouvelles données. Ainsi des données sur les mouvements interacadémiques et intra-académiques des enseignants, ou bien des fiches d'orientation des élèves en fin de 3^{ème}, qui représentent une information très riche pour comprendre le processus d'orientation et les biais sociaux.

Ce type de données étaient auparavant mobilisées par un nombre restreint de chercheurs, qui devaient en connaître l'existence, et le cas échéant, y accéder dans les locaux de la Depp à Paris. D'autre part, la Depp était confrontée à un enjeu interne de formalisation et de normalisation de la documentation des données, ainsi que de mise en cohérence des données entre elles. L'EquiPex s'est donc attaché à faciliter l'accès à ces données à des chercheurs issus de nombreuses spécialités. Il s'est agi d'informer sur ces données et de mettre en place un accès à distance. Par ailleurs, un projet interne à la Depp, et articulé à Idee, a créé une dynamique de travail sur ces données.

L'EquiPex Idee mène plusieurs actions. Il a œuvré à la création d'un catalogue en ligne depuis 2022 et alimenté progressivement. Ce catalogue, qui respecte les standards internationaux de présentation des métadonnées, rend transparente la collection des données de la Depp accessibles à la recherche. L'EquiPex développe également un accès à distance, un entrepôt et un espace de travail. Ce développement s'appuie sur l'outil Onyxia développé par l'Insee. Il formalise aussi les opérations d'appariement des données et les protocoles d'enrichissement des données d'enquête par les données administratives. Il construit un outil informatique qui organise le processus interne de création des conventions avec les chercheurs. Enfin, il contribue à la création, à la Depp, du bureau B1, bureau de l'appui à l'évaluation des politiques publiques et de soutien à la recherche, qui renforce les interactions entre la Depp et les chercheurs.

Ces différentes actions ont été initiées en 2022 et ont répondu à une demande très forte. Désormais, des chercheurs issus de différentes disciplines, comme la sociologie, la psychologie et les sciences de l'éducation, accèdent à davantage de données. Cette forte demande, susceptible d'entraîner une saturation des capacités, a conduit à renforcer l'équipe du bureau B1. En outre, l'outil de gestion des conventions est

en phase de recette, et permettra de fluidifier et accélérer le traitement des demandes. Enfin, une version bêta de l'accès à distance est en cours de test.

Du point de vue des chercheurs, l'intégration très forte entre l'administration et la recherche a suscité une grande satisfaction, ainsi que la capacité d'obtenir un financement du ministère de l'Enseignement supérieur.

Intervention de Christine Hennion (co-rapporteuse de la mission data et territoires)

Christine HENNION rappelle qu'elle est élue locale à Courbevoie et que, avant son mandat de députée, elle avait travaillé dans le secteur privé, sur les sujets de Télécom et de numérique. Appréhender ces deux niveaux d'organisation et d'instance que sont la représentation nationale et la représentation locale, lui a permis d'appréhender l'écart entre les volontés exprimées sur le plan national et leur application dans les territoires.

Le rapport de la mission Data et territoires, corédigé avec Bertrand Monthubert et Magali Altounian, a été remis au ministre de la Transformation et de la Fonction publiques en septembre 2023. Il questionne la capacité des collectivités territoriales, qui produisent une grande quantité de données, à bénéficier elles-mêmes de ces données afin d'évaluer et de piloter leurs politiques publiques. Les auteurs du rapport ont souhaité aller au-delà de la question de l'open data, et interroger l'efficacité de la LRN de 2016, qui est un facteur d'incitation à mettre à disposition des données. Si seuls 16 % des collectivités ont mis des données à disposition du public, cette proportion croît rapidement, les collectivités prenant conscience de l'intérêt de la démarche. Celles qui ne mettent pas des données à disposition sont très souvent contraintes par un manque de moyens humains. En effet, les personnels en capacité d'opérer ces démarches sont insuffisants en nombre et souvent requis par d'autres priorités sur les sujets liés au numérique. Naturellement, les grandes collectivités sont plus à même d'initier la démarche. Cependant, les auditions menées par la mission Data et territoires ont montré que certaines petites collectivités y parviennent également.

Les collectivités qui ne parviennent pas à répondre à la demande formulée dans la LRN ont fait part de leur frustration dans leur relation avec l'État. En effet, leurs représentants signalent une asymétrie entre les demandes et les injonctions de l'État en matière de production de données administratives, auxquelles elles sont tenues de répondre avec diligence, et la lenteur des services étatiques à restituer des données dont elles-mêmes ont besoin pour leur gestion et leurs politiques publiques. Malgré toutes les innovations dans le domaine des données, le manque de coordination avec l'État provoque parfois un surcroît de travail dont les représentants des collectivités peinent à percevoir le sens et l'utilité.

Le rapport de la mission Data et territoires a formulé une série de recommandations. Il met l'accent, d'abord, sur la gouvernance autour des données territoriales afin de pallier le manque de coordination et de dialogue au niveau de la définition même des données à collecter. En effet, certaines données significatives au niveau de la statistique nationale ne permettent pas toujours aux collectivités de tirer des conclusions pour elles-mêmes. Le rapport recommande par conséquent la création d'espaces de collaboration et d'échanges au niveau des territoires. Il suggère de s'appuyer sur les structures préalablement existantes, à l'image du Comité national d'information géolocalisée (Cnig), qui regroupe de nombreux acteurs, y compris l'Insee, et pourrait concerner d'autres instances afin de mieux définir et partager les informations sur les données.

Les collectivités frustrées par rapport aux demandes de l'État réclament une amélioration de cette relation. En effet, elles ont besoin d'un support de proximité, par exemple au niveau des services déconcentrés de l'État à l'échelle départementale. La Direction interministérielle du numérique (Dinum) met en place actuellement des référents sur les données au niveau régional. Toutefois cet échelon est trop vaste et ces référents régionaux sont encore peu identifiés. Dès lors, il convient de descendre au niveau régional le soutien à la démarche de mise à disposition des données par les collectivités.

Menée sur une courte période, la mission Data et territoires n'a pu creuser la question du soutien financier. Sur les sujets liés aux données numériques, le financement s'accomplit le plus souvent dans le cadre d'appels à projets, mais les auditions de la mission ont démontré un réel besoin de mutualisation. En effet, entrer dans une phase d'industrialisation, nécessaire sur ces sujets, demande un financement dans la durée afin d'entretenir les outils, les bases et d'animer des communautés. Or l'État ne met pas suffisamment en œuvre ce type de financement. C'est la raison pour laquelle le rapport de la mission met en avant le préciput utilisé dans les projets ANR afin d'évaluer l'éventuelle pertinence de ce mécanisme de financement.

Parmi les autres recommandations du rapport figure la nécessaire attention à porter à l'acculturation et à la formation à tous les niveaux, y compris au niveau du public. Le rapport a noté que les citoyens font plutôt confiance aux collectivités dans la gestion des données. Il convient dès lors de les inciter à contribuer davantage à la création de données, mais aussi à l'élection de priorités.

Intervention d'Arnaud Degorre (Insee Pays de la Loire)

Arnaud DEGORRE fait part du sentiment, partagé dans l'ensemble des directions régionales de l'Insee, d'un enrichissement du paysage statistique à l'échelle des territoires. Cet enrichissement est lié à l'identification des nouveaux usages par les acteurs locaux, mais aussi aux nombreuses innovations énumérées par Jean-Luc TAVERNIER lors de son intervention. Ces deux éléments bénéficient à l'offre locale et concourent à ce que l'intervenant qualifie de « concordance des données », comme *momentum* dans l'essor des statistiques territoriales.

Le premier facteur de cette concordance des données est la meilleure reconnaissance des données locales. La statistique locale était auparavant vue, parfois par les statisticiens publics eux-mêmes, comme une forme appauvrie de la statistique. En effet, les statisticiens régionaux se heurtaient souvent à l'impossibilité d'atteindre la pureté conceptuelle des enquêtes spécifiques. Par exemple, mesurer le chômage à travers le recensement de la population paraît une démarche pauvre comparée à la pureté de la mesure permise par l'enquête emploi en continu. C'est la raison pour laquelle des extensions régionales d'enquête ont été déployées, de même que l'exploration des méthodes petits domaines afin de tenter de surmonter cette limite de la mesure locale.

Cette vision réductrice de la statistique locale n'a plus cours : au contraire, le contenu informationnel dû à la localisation des données est désormais pris en compte, et en fait une forme augmentée de la statistique. Le Cnig définit ainsi la donnée localisée comme toute donnée dont la composante de localisation apporte de la valeur. La valorisation de cette information géographique s'est substituée, pour ainsi dire, aux limites rencontrées dans les statistiques territoriales comparativement aux sources nationales. L'offre de l'Insee témoigne de ce changement de considération. Un rapport de l'Inspection générale de l'Insee a mis en avant la nécessité de progresser en économie géographique et dans l'analyse spatiale de l'ensemble des phénomènes économiques et sociaux. Un sujet régional n'est donc plus seulement appréhendé sous l'angle local, mais peut l'être aussi sous l'angle national ou par comparaisons entre régions. D'ailleurs, de plus en plus d'études nationales territoriales embrassent l'ensemble des échelles.

L'émergence des sources nouvelles représente la deuxième composante de la concordance des données. À l'échelle des territoires, les sources des opérateurs restent encore à valoriser et le seront sans doute grâce au programme de moyen terme. Ces dernières années, les sources administratives ont fait l'objet d'un regain d'intérêt. La plupart d'entre elles comportent une composante localisée qui leur est native. Les déclarations sociales nominatives, les fichiers localisés sur les rémunérations et l'emploi salarié, le fichier approché des résultats d'Esane (Fare) ou encore les dispositifs Filosofi et Fideli, ont été créés au cours des deux dernières décennies. Le potentiel d'exploitation de ces nouvelles sources administratives est d'autant plus élevé qu'elles peuvent être mobilisées au niveau national et au niveau territorial. Il convient de ne plus opposer ces deux niveaux, mais au contraire de les relier.

Le troisième facteur de la concordance des données est l'accompagnement systématique de la mise à disposition des informations au niveau local selon l'intention d'usage. Cette intention d'usage suppose de procéder à un examen problématisé de la donnée locale. L'Insee propose ainsi trois modalités de réponse à la demande des utilisateurs locaux. La première modalité est la production d'études de cadrage à l'initiative de l'Insee, afin d'inscrire la demande dans un sujet d'actualité relevant du débat public local. Deuxième canal de réponse, les études en partenariat menées avec les collectivités territoriales et les services de l'État se rapprochent de l'idée de la statistique publique comme bien commun. Enfin, la troisième modalité, qui s'est nettement développée récemment, est la notion de conseil-expertise. Il s'agit d'examiner, avec les collectivités territoriales, comment fonctionne la statistique publique et quel apport elle représente pour le système d'information des collectivités et pour le pilotage de leurs politiques publiques.

Il convient de remarquer que, quelle que soit sa modalité, toute réponse est mutualisée dans le réseau de l'Insee, afin qu'elle soit comparable d'une région à une autre, et reproductible. Chaque collectivité territoriale ayant travaillé avec l'Insee doit être en mesure de partager cette expérience avec d'autres collectivités territoriales, dans d'autres régions. La mesure de la consommation foncière en réponse à la loi Climat et résilience de 2021 et à la loi Zéro artificialisation nette (ZAN) de 2023, fournit un exemple de cette mutualisation. Cette mesure a été mise en œuvre par l'Insee en partenariat avec le Service des données et

études statistiques (Sdes) au niveau national, les Directions régionales de l'environnement, de l'aménagement et du logement (Dreal) au niveau régional, et souvent les Conseils régionaux. Une méthodologie commune a été développée afin d'appréhender la logique d'artificialisation observée dans les collectivités territoriales, et surtout en comprendre les vecteurs sous-jacents.

Cet exemple de réussite illustre le quatrième vecteur de concordance des données, à savoir la conjonction des sources. En effet, pour parvenir à cette mesure, différentes sources statistiques ont été fédérées, telles que les Fidéli, les données des fichiers fonciers, les données du recensement, etc. Des rapprochements intersectoriels ne sont plus une impossibilité au niveau territorial.

Un autre exemple illustre ces évolutions. L'Insee travaille sur le risque environnemental et la connaissance des populations exposées au risque de pollution en confrontant diverses sources. Il mobilise à la fois la statistique publique, à travers le recensement de la population et la localisation des habitants, et des sources externes de données, ici des données collectées par les associations agréées de surveillance de la qualité de l'air. Ainsi, déterminer quels sont les groupes sociaux concernés permet de procéder à des analyses territorialisées des risques de pollution.

Au registre des améliorations en cours ou à apporter au niveau de l'offre de données locales, l'effort consiste à se porter au-delà de l'open data, ainsi que le recommandait le rapport de la mission Data et territoires. La priorité de l'Insee en région est de déterminer les moyens d'accès aux données et surtout les modalités de traitement des données proposées aux utilisateurs. Il s'agit là du cœur de la notion de « service de la donnée ». Dans cette perspective, l'Insee offre désormais un accès normalisé aux données localisées avec une interface de programmation d'application (API) dédiée, l'API données locales. Des outils de cartographie dynamique permettent non seulement d'utiliser les sources de la statistique publique, mais aussi d'injecter ses propres sources. Des comparateurs de territoire sont également disponibles sur le site de l'Insee. À l'avenir, l'Insee mettra à disposition des outils multiscalaires afin de réconcilier les échelles d'observation selon les données de référence. En outre, une forme de révolution silencieuse se prépare avec l'apparition, à la suite du projet Mélodi, d'un catalogue de jeux de données, qui transformera l'usage des données pour les spécialistes comme pour le grand public. En effet, ce dispositif permettra une pleine exploration et une pleine connaissance des jeux de données locaux ou nationaux, ainsi que de la filiation de ces données selon les standards internationaux. De plus, une approche d'interopérabilité concourra au partage de ce catalogue avec les autres instances ouvrant leurs données.

Sur tous ces sujets, il convient avant tout d'écouter les besoins des utilisateurs, tant au niveau local que national. Réciproquement, il est nécessaire d'écouter l'offre des utilisateurs, puisqu'ils détiennent eux-mêmes une partie des données en mesure de répondre à leur demande. L'Insee mène plusieurs actions afin de conduire cette écoute au niveau local. Un groupe de travail national mené de façon concertée entre l'Insee et le secrétariat général du Cnis se penche sur les canaux de captation des besoins des utilisateurs locaux. Par ailleurs, des *focus groups* sur les utilisateurs avancés des données locales ont été initiés en interne à l'Insee. Enfin, l'Insee s'efforce d'être présent dans les salons et les forums dédiés à la donnée géolocalisée, comme le salon GéoDataDays ou le salon de la data.

Patrice DURAN remercie les intervenants et note la primauté du lien entre producteur et utilisateur de données, relevé par chacun d'eux. De ce lien de dépendance émergent les capacités de coordination. Il convient également de souligner la réussite que représente l'administration déconcentrée de l'Insee, qui rend perceptible un besoin de contextualisation et facilite grandement les échanges, malgré la relation complexe entre l'État et les collectivités territoriales.

Patrice DURAN s'enquiert de la coordination des directions régionales de l'Insee avec les Chambres régionales des comptes.

Arnaud DEGORE indique que les directions régionales de l'Insee se sont mises en relation avec leurs réseaux respectifs en 2023. Cette mise en relation relève d'une logique d'ouverture des données pour les exercices d'évaluation des politiques publiques dont les Chambres régionales des comptes pourraient être saisies. Cette évaluation n'est pas conduite en partenariat entre les Chambres régionales et l'Insee, pour des raisons déontologiques. Néanmoins, la démarche consiste à faciliter l'identification des matériaux statistiques permettant d'obtenir des éléments de contexte pour des exercices d'évaluation.

Béatrice SÉDILLOT, responsable du Sdes, SSM du ministère de la Transition écologique, évoque les données environnementales. Le Sdes a mis à disposition de Progedo récemment l'enquête portant sur la rénovation énergétique des maisons individuelles et l'enquête logement 2020. Les données environnementales ne sont quant à elles pas collectées par des enquêtes. En revanche, un travail important a été mené au niveau du ministère *via* le portail *Green data for health* (GD4H), qui permet d'établir un

catalogue de données environnementales en lien avec les enjeux de santé notamment. Il s'agit de données ouvertes, en application de la convention d'Aarhus. Il est important que les chercheurs connaissent l'existence de cette plateforme contenant une centaine de bases de données.

Yvon SÉRIEYX souhaite des éclaircissements sur les méthodes de transformation de fichiers bruts en fichiers faiblement anonymisés, puis en fichiers FPR. Il demande quel type d'accompagnement pourrait permettre aux équipes de produire systématiquement des fichiers FPR qui seraient stimulants pour la recherche et utiles aux actions de la société civile.

Erik ZOLOTOUKHINE répond qu'il s'agit d'une question complexe que Progedo et l'Adisp ne traitent pas directement. Elle relève d'abord des producteurs, et en particulier les nouveaux producteurs, qui constituent les données FPR. Le fichier FPR est dépourvu de définition précise. La difficulté ne réside pas dans le fait de retirer des valeurs identifiantes, mais dans la granularité et le croisement de fichiers qui potentiellement comportent 2 000 variables. En effet, il n'est pas garanti que le croisement de dix ou vingt de ces variables ne puisse permettre une identification. Cependant, après vingt ans d'usage, il n'est pas fait état de quelconque accident sur ces fichiers diffusés sous contraintes et dont l'accès est restreint.

Jean-Luc TAVERNIER demande à Christine HENNION de quelle manière l'exécutif compte se saisir de la première recommandation du rapport de la mission Data et territoire, à savoir la mise en place d'une gouvernance.

Christine HENNION indique que le rapport a été commandé par le ministère de la Transformation et de la Fonction publiques, mais concerne plusieurs ministères ainsi que les collectivités territoriales, en particulier les régions, parmi lesquelles certaines ont étoffé des structures déjà existantes. Certaines recommandations du rapport réclament des décisions de l'État. Par exemple, il est question d'organiser des échanges d'agents entre collectivités afin d'améliorer la compréhension des fonctionnements et des besoins respectifs. Par ailleurs, la Dinum s'organise au niveau des territoires et au niveau des ministères, et chaque ministère a dû donner sa feuille de route sur les questions numériques.

Jérôme DEAUVIEAU, représentant des chercheurs en sciences économiques et sociales au Cnis, estime que les différents dispositifs sont en mesure de répondre aux demandes des chercheurs. Toutefois, le prix d'accès au CASD a nettement augmenté depuis quelques années. Ce coût, qui est justifié, ne pose pas de problème en soi. En revanche le mode de financement de ce type de coût n'est pas toujours adapté, en particulier pour les doctorants en sciences sociales.

Kamel GADOUCHE concède que cette question du coût de l'accès au CASD lui est souvent adressée. Avant 2016, la gratuité de l'accès entraînait des effets pervers, de nombreuses demandes ne se traduisant pas par de l'activité. Par ailleurs, le CASD avait obtenu en 2011 un financement de plusieurs millions d'euros à travers un EquiPex qui s'est achevé en 2019 et qui exigeait un autofinancement à terme afin de pérenniser l'équipement. En effet, le CASD ne bénéficie pas de dotation de l'État. La contribution des membres du consortium constituant le groupement d'intérêt public permet de maintenir un coût modeste puisque les membres du consortium ne tirent aucun bénéfice direct de leur contribution.

Kamel GADOUCHE insiste sur une distinction : ce n'est pas l'accès au CASD qui représente un coût, ni les données, mais l'environnement de calcul. Il se dit favorable à des subventions afin de réduire le coût de ce service. Cependant, il convient de rappeler que les exigences en matière de traitement, de calcul et surtout de sécurité augmentent et représentent un coût.

VI. CONCLUSION

Patrice DURAN remercie les intervenants, ainsi que Dominique FRANCOZ pour son travail sur le programme moyen terme. Il conclut l'assemblée plénière en soulignant que la statistique publique est bien une affaire d'action collective. À ce titre, les besoins de coordination sont particulièrement évidents. En ce qui concerne les collectivités territoriales, la formation est primordiale dans la mesure où nombre d'entre elles sont insuffisamment équipées pour produire des statistiques de qualité comme pour en faire un usage approprié. Il souligne également le caractère central et décisif de la question des appariements. Ce sujet détient un important potentiel d'interaction entre des administrations ou des acteurs insuffisamment reliés aujourd'hui.

Enfin, Patrice DURAN salue Françoise MAUREL, qui s'apprête à quitter le Secrétariat général du Cnis, et la remercie très chaleureusement pour la qualité de son travail et de son implication.

La séance est levée à 12h45.