

Paris, le 25 mai 2023

Synthèse du rapport d'information sur *l'accès et le traitement des données par l'Insee dans le cadre de la politique d'open data*

présenté par **M. Michel Sala**,
rapporteur spécial des crédits des programmes 220 et 305 de la mission *Économie*

en application de l'article 146, alinéa 3, du règlement de l'Assemblée nationale

La crise de la Covid 19 a **exacerbé les limites de l'accès aux données** par l'Insee, coupé de ses ressources traditionnelles, il a su innover en développant de nouveaux partenariats avec des acteurs de la société civile et en diversifiant ses sources.

Vingt années après les prémices de l'ouverture des données de l'Insee, l'extension de l'*open data*, désormais partie intégrante de la culture de l'institution, tend à s'étendre à de nouvelles frontières sans que le cadre normatif ou la doctrine de diffusion des données privées à destination des acteurs publics ne soient établis.

Le rapporteur spécial a décidé de s'intéresser **à la manière dont l'Insee rend disponible les données dans le cadre de la politique d'open data ainsi qu'aux modalités d'accès pour l'Institut aux données de nature privée**. Une partie de l'avenir de la statistique publique se situant, désormais, dans l'exploitation des données privées.

À cette aune, le rapporteur montre que si l'Insee s'impose comme un acteur de référence dans la production statistique (I) il n'en demeure pas moins doublement concurrencé par un cadre normatif complexe et par des acteurs privés disposant de toujours davantage de données (II). Un tel constat impose aujourd'hui, pour l'Insee, de se tourner vers la donnée de nature privée, en vue de fluidifier et d'intensifier les flux de données, au service de la statistique publique (III).

1 | UN ACTEUR DE RÉFÉRENCE DE LA PRODUCTION STATISTIQUE, GARANT DE LA BONNE INFORMATION PUBLIQUE ET RÉCIPIENDAIRE D'UNE ÉTENDUE LARGE DE DONNÉES

➤ L'Insee bénéficie d'un cadre normatif ancien et exigeant entourant son activité statistique.

– Le texte de référence applicable aujourd'hui encore à l'Insee a été promulgué en 1951 et détermine les modalités de réalisation des enquêtes statistiques. Surtout, il assujettit l'Institut à un respect absolu au **principe de « secret statistique »**.

– L'Insee réalise, selon des rythmes déterminés, mensuel pour certains indices et annuel pour des études et des recensements, des travaux dont la qualité assure **la bonne information des décideurs et du public**.

– Les productions de l'Insee sont adossées à **des données particulièrement éclectiques**, que ce soit des données en provenance d'enquêtes par échantillonnage, des données administratives ou encore des données de nature privée dans certains cas.

➤ L'*open data* a eu pour effet d'ouvrir les bases de données au public, poussant l'Insee à devenir pédagogue et accessible.

– L'*open data* a eu pour effet une ouverture large et gratuite des données produites par l'Institut. Pour certaines d'entre elles, auparavant revendues aux diffuseurs et utilisateurs de données, **la fin des redevances perçues a eu pour effet de diminuer les recettes perçues par l'Institut**. C'est près de **11 millions de produits issus de la revente de données qui ont dû être compensés par l'État**.

– Par ailleurs, cette politique d'ouverture des données a fait évoluer les pratiques de l'Insee. Désormais, l'enjeu n'est plus tant à l'ouverture qu'à la recherche d'une qualité optimale de donnée disponible.

– Aujourd'hui, **la France apparaît en tête des classements européens en matière d'ouverture des données publiques**, ce qui constitue une véritable réussite nationale.

2 | L'INSEE EST CORSETÉ PAR UN CADRE JURIDIQUE PEU LISIBLE ET PAR DES ACTEURS PRIVÉS TITULAIRES D'UNE MASSE CROISSANTE DE DATA

➤ L'activité statistique de l'Insee est aujourd'hui limitée par un cadre normatif ancien, peu lisible et mouvant.

– Le rapporteur constate que le cadre applicable à la statistique publique est affecté de lourdeurs ce qui conduit à brider l'utilisation de certaines *data*. D'une part et de manière paradoxale, des données administratives, mais d'autre part et surtout, des données de nature privée. Cela constitue un **obstacle majeur à l'innovation et à la transformation de la production statique**.

Par ailleurs, les mouvements induits par la présentation des textes *Data Act* et *Data Governance Act* en 2022 au niveau européen ainsi que par la révision en cours du Règlement n° 223/2009 sont

sources d'incertitudes pour les services statistiques et l'Insee et ne semblent pas répondre aux évolutions nécessaires à une exploitation pérenne et facilité des données privées à des fins statistiques.

➤ **Au surplus, l'Institut apparaît concurrencé tant du point de vue de l'exploitation que de l'accès aux données par des acteurs de la société civile.**

– Une concurrence qui tient tout d'abord à **la massification de la donnée disponible**, les *Big Tech* apparaissent en effet comme les mieux placées pour exploiter les données dont le taux de croissance annuel moyen est de 26 % environ.

– Une concurrence qui revêt ensuite **une incidence économique dans la mesure où la donnée constitue une manne financière forte sur l'ensemble de sa chaîne d'utilisation** (donnée matière première, donnée comme levier de gain de productivité et donnée comme actif stratégique sur un marché).

– Le rapporteur spécial note enfin que le cadre applicable aux données privées est aujourd'hui particulièrement protecteur pour les entreprises, empêchant un accaparement vertical en l'absence d'une notion juridique claire, fondée par exemple autour de la **notion de « donnée d'intérêt général »**.

3 | AFIN DE CONFORTER LA QUALITÉ DE L'INFORMATION STATISTIQUE L'INSEE S'ORIENTE VERS LA DONNÉE PRIVÉE, À LA FAVEUR D'UNE DÉMARCHÉ PLUS INCLUSIVE ET PARTENARIALE AVEC LES ACTEURS DE LA SOCIÉTÉ CIVILE

➤ **Le rapporteur s'est intéressé aux partenariats tissés entre l'Insee et des acteurs privés ayant permis de mobiliser des données d'une nouvelle nature à des fins de statistique publique.**

– À cet égard, le rapporteur spécial relève que le partenariat établi avec les établissements bancaires, notamment avec La Banque Postale, apparaît comme un **modèle de réussite en matière de mise à disposition des données et appelle à démultiplier cette pratique**.

– En revanche, le partenariat avec les acteurs de la téléphonie mobile appelle un constat plus nuancé. Toutefois, il atteste de **l'intérêt d'avoir recours à des données de nature nouvelle pour la production statistique**.

➤ **Le rapporteur a ainsi cherché à formuler des propositions permettant des flux de données plus intenses et récurrents à destination des services statistiques et de l'Insee.**

Les principales recommandations du rapport :

- Modifier la rédaction actuelle de l'article 3 *bis* de la loi du 7 juin 1951 afin d'élargir le recours aux données privées.
- Cartographier avec les acteurs concernés les données qui pourraient être communiquées dans le cadre de convention de gré à gré.
- Proposer des garanties renforcées en matière de sécurité aux acteurs privés concernés par les échanges de données.
- Prévoir un dispositif de réquisition des données en cas de besoin d'intérêt général.
- Sanctionner les refus systématiques et injustifiés dans l'accès aux données.
- Développer la démarche expérimentale de labellisation de séries statistiques produites hors du champ des SSM et de l'Insee.
- Favoriser des concertations en tables rondes inter-acteurs afin d'anticiper les besoins en matière de protection des données personnelles.