

Juillet 2026

Courrier des statistiques 15



Rédaction en chef

Emmanuelle Nauze-Fichet

Contribution

Insee : Lino Galiana, Olivier Lefebvre,
Thomas Lellouch, David Levy,
Clotilde Masson, Hervé Piffeteau
Cour des comptes : Benjamin Belrhomari,
Lauriane Martzel, Lisa Menez,
Léo Zabrocki-Hallak
Drees : Pierre-Yves Cabannes
Eurostat : Pierre Bischoff
Université de Lille : Béatrice Touchelay

Directeur de la publication

Fabrice Lenglard

Directeur de la collection

Pascal Rivière

Rédaction

Solenn Ily, Emmanuelle Nauze-Fichet,
Yann Quélénnec, Pascal Rivière

Composition

Agence Efil
90 boulevard Heurteloup
37 000 Tours
02 47 47 03 20
www.efil.fr

Photo de couverture

Getty Images

Éditeur

Institut national de la statistique
et des études économiques
88, avenue Verdier
92541 MONTROUGE CEDEX

www.insee.fr

© Insee 2026 « Reproduction partielle
autorisée sous réserve de la mention
de la source et de l'auteur ».



Courrier des statistiques N15

SOMMAIRE

Présentation du numéro <i>Pascal Rivière</i>	4
Enquêter les personnes sans domicile – Quand la statistique publique rencontre un terrain inhabituel <i>Pierre-Yves Cabannes et Thomas Lellouch</i>	9
De nouvelles nomenclatures d'activités – Les classements en théorie et en pratique <i>Clotilde Masson</i>	31
Le système statistique européen – 2004-2015 : une période de crises et de consolidations <i>Pierre Bischoff et Hervé Piffeteau</i>	53
Une source harmonisée pour mesurer le tourisme européen – Les données de plateformes collectées par Eurostat <i>David Levy</i>	73
Qu'apportent les métiers de la <i>data</i> à la statistique publique ? <i>Lino Galiana et Olivier Lefebvre</i>	95
Extension du domaine de la statistique : les évaluations de politiques publiques à la Cour des comptes <i>Benjamin Belrhomari, Lauriane Martzel, Lisa Menez et Léo Zabrocki-Hallak</i>	117
Les statistiques coloniales et leurs archives <i>Béatrice Touchelay</i>	139

PRÉSENTATION DU NUMÉRO

À quoi sert la statistique publique ? Fondamentalement, elle sert à poser les bases d'une information de qualité à même d'éclairer de manière objective et rigoureuse le débat public.

Comment procéder pour y parvenir ?

Il faut déjà collecter les données élémentaires. À cette fin, le système statistique public réalise régulièrement des enquêtes sur le terrain. Certaines ont fait l'objet d'articles dans les précédents numéros du Courrier des statistiques, comme l'enquête Trajectoires et Origines [N8] ou l'enquête Emploi [N6]. De telles opérations sont nécessaires : elles permettent d'obtenir, parfois dans des conditions difficiles, des données qui ne sont présentes dans aucune base de données administrative, ni même privée.

Dans ce numéro, Pierre-Yves Cabannes et Thomas Lellouch nous présentent une enquête particulièrement emblématique, terriblement exigeante, qui fait l'honneur de la profession : l'enquête Sans domicile, menée au printemps 2025. Une telle opération, hors normes, ne peut en aucun cas reprendre une démarche standard et il a fallu concevoir pour elle une méthodologie sur mesure, tenant compte de la diversité des situations. L'originalité concerne de multiples aspects de l'enquête : le fait de passer par l'intermédiaire de services d'aide, la grande variabilité des configurations de terrain (hébergements, distributions de repas, accueils de jour et services itinérants), les « questionnaires qui parlent » en 21 langues, etc. La collecte a mobilisé près de 1 000 agents, pour un total de 17 000 entretiens. La complexité de sa mise en œuvre rend compte de la tension permanente entre exigence méthodologique et contraintes opérationnelles. Elle explique la rareté du dispositif, la dernière enquête équivalente datant de 2012. Dire que les résultats sont très attendus relève de l'euphémisme !

Pour nourrir le débat public, il faut par ailleurs s'appuyer sur une architecture solide permettant la comparabilité dans l'espace et le temps. Cela veut dire par exemple des concepts partagés, stables, éprouvés. Un autre pilier majeur d'un fonctionnement professionnel et maîtrisé, ce sont les nomenclatures. Discrètes mais puissantes, elles irriguent les statistiques publiques. On peut citer les nomenclatures socioprofessionnelles [N4], d'infractions [N7], mais aussi de catégories juridiques, de maladies...

Clotilde Masson nous parle ici d'une nomenclature jouant un rôle tout à fait majeur : la nomenclature d'activités. Sans elle, pas de secteurs ni de branches d'activité, qui sont centraux dans la grammaire de la statistique d'entreprise et de la comptabilité nationale. L'auteure décrit d'abord ce qu'est une nomenclature en général : c'est un objet riche, qui ne se limite pas à une liste de catégories, fût-elle structurée de façon arborescente. Celle des activités existe depuis fort longtemps, et elle a évolué au gré des transformations du monde qu'elle est censée représenter. Une nouvelle version de la nomenclature d'activités, c'est un événement très rare, qui a lieu tous les 10-15 ans, et qui s'appuie sur une préparation considérable. En particulier, depuis 1993, la nomenclature française est imbriquée dans celle de l'Europe, elle-même articulée avec celle des Nations Unies. Inévitablement, le résultat est fait de multiples compromis.

Si l'on veut que les statistiques publiques soient considérées comme une référence, il est nécessaire qu'elles inspirent confiance, qu'elles offrent des garanties de neutralité, de

transparence, de qualité. Cela requiert une capacité à se remettre en cause, et un cadre institutionnel solide : le Conseil national de l'information statistique (Cnis) [N6], le Comité du label [N5], l'Autorité de la statistique publique [N5], les revues par les pairs [N12].

Pierre Bischoff et Hervé Piffeteau montrent que cette confiance n'est pas un acquis immuable : elle se conquiert par le professionnalisme, mais un grain de sable peut la remettre en cause, et la rétablir est alors une conquête. Leur article est le deuxième volet d'une série chronologique de trois papiers sur le système statistique européen, dont le premier volet s'achevait en 2003. Notre grain de sable fait irruption en 2004 et c'est un gros grain : c'est la crise grecque, avec la falsification des comptes publics. Toute la statistique européenne s'en retrouve fragilisée. En raison de cette crise, mais pas seulement, l'Europe met en place une liste de principes (neutralité, transparence...) qui composent un « code de bonnes pratiques ». Celui-ci donne un cadre à l'ensemble des instituts nationaux de statistique (INS), qui demeure important aujourd'hui. Mais nous ne sommes pas au bout de nos peines, car un deuxième grain de sable, en l'occurrence une deuxième crise grecque, secoue l'édifice en 2009. Il conduit à consolider le code de bonnes pratiques, et notamment à s'assurer de sa bonne utilisation. Difficiles à vivre, ces crises n'en ont pas moins été fécondes pour consolider l'architecture institutionnelle des statistiques, et notamment sa législation par l'élaboration d'une nouvelle loi statistique européenne.

Produire des statistiques utiles au débat public, c'est aussi innover, en permanence, pour tenir compte des évolutions de l'environnement et du champ des possibles. L'innovation peut passer par de nouvelles sources de données [N13] : les données de caisse [N3], les comptes bancaires [N12], la téléphonie mobile [N13], les transactions bancaires [N13].

L'article de David Levy porte ainsi sur une source de données privées essentielle pour compléter les statistiques du tourisme : les données de plateformes (Airbnb, Booking, Expedia, Tripadvisor). Ici, l'innovation est aussi institutionnelle : après quelques expérimentations disparates, aux succès variables, c'est Eurostat qui a pris la main, jouant de fait un rôle de producteur de données utilisées par les INS, à front renversé. C'est d'autant plus légitime que ces plateformes de réservation ne se prêtent pas facilement à un découpage en pays : l'Europe semble la bonne maille. Envisager d'utiliser ces données nécessite toute une phase de préparation avec chacun des opérateurs, qui se matérialise in fine par des conventions. Les données de plateforme complètent très utilement le dispositif d'observation du tourisme sur un domaine difficile à appréhender : les locations de courte durée entre particuliers. Néanmoins, ce partenariat public-privé soulève des défis structurels : dépendance aux accords volontaires, opacité relative des algorithmes de gestion des plateformes et absence d'identifiant unique pour traiter les doublons de capacité d'hébergement. Pour l'avenir, un travail de consolidation s'impose.

L'innovation, nécessaire, ne se limite pas aux sources d'information. Elle renvoie aussi aux méthodes et environnements de travail des professionnels de la donnée. Dans le contexte de développement massif des traces numériques liée à l'activité, on assiste depuis quelques années à une évolution profonde, inspirée par les apports de la *data science* : nouvelles infrastructures comme le SSPCloud [N7], nouveaux formats [N9], processus plus standardisés d'intégration de données administratives [N9].

Lino Galiana et Olivier Lefebvre s'intéressent aux implications sur nos manières de travailler. Cette nouvelle discipline engendre progressivement une transformation en profondeur des métiers et l'émergence de nouveaux profils spécialisés : *data analysts*, *data scientists*, *data engineers*. La statistique publique ne peut ignorer ces transformations : elle s'attache à les mettre au service de ses missions, dans le respect de ses valeurs. La révolution de la *data science* favorise un rapprochement des compétences entre directions métiers et informatiques, au profit par exemple de la capacité à associer des données de formes et de natures diverses pour répondre à de nouveaux besoins, avec une exigence renforcée de transparence, un nouvel impératif de répliquabilité... Mais cela représente aussi un défi d'adaptation continue, dans un environnement où les besoins, les outils et les méthodes évoluent rapidement. L'Insee et les services statistiques ministériels se sont dotés d'infrastructures et d'organisations pour explorer, expérimenter, industrialiser et maintenir ces nouveaux processus.

Tout en adoptant ces nouvelles manières de travailler et en mobilisant ces nouvelles sources, on peut également innover institutionnellement, et aller au-delà du champ bien balisé de la statistique publique. Ainsi, depuis plusieurs années, des pôles de science des données, appelées aussi pôles *data*, se sont développés dans plusieurs inspections, par exemple à l'inspection générale des finances (IGF) [N13] et à l'inspection générale des affaires sociales (IGAS) [N13]. Leur fonction est différente, puisque leur rôle est avant tout d'aider à l'élaboration, au pilotage et à l'évaluation de politiques publiques sur des sujets souvent très ciblés.

Benjamin Belrhomari, Lauriane Martzel, Lisa Menez et Léo Zabrocki-Hallak nous parlent dans ce numéro de la *data science* à la Cour des comptes. Les juridictions financières – Cour des comptes et Chambres régionales et territoriales des comptes – ont depuis 2008, en plus de leurs missions historiques (contrôler, juger, certifier), une mission d'évaluation des politiques publiques. À cette fin, elles ont de plus en plus recours à l'analyse statistique comme élément de démonstration dans leurs rapports. Celle-ci leur permet d'aller plus loin dans leur appréciation du bon usage de l'argent public. Le département d'analyse et de science des données de la Cour des comptes, de par le positionnement de celle-ci, possède la particularité de pouvoir accéder à de très nombreuses sources administratives. C'est une équipe conséquente de *data scientists* et *data analysts*, ayant déjà plusieurs années d'expérience. Elle emprunte son cadre conceptuel et ses méthodes à la recherche tout en conservant l'essentiel de ses propres normes et procédures. Comme c'est le cas pour les pôles *data* des inspections, elle est soumise à une exigence de rapidité, qui modifie la manière d'appréhender les sujets par rapport aux démarches de la statistique publique.

Enfin, nourrir le débat public, c'est aussi s'inscrire dans le temps long, comme on a pu le voir avec l'histoire du lien entre statistique publique et débat démocratique depuis la naissance de l'Insee [N9 et N11], celle de la statistique européenne depuis l'origine [N14], ou celle des échanges de données et publications statistiques depuis le XIX^e siècle [N14].

Béatrice Touchelay remonte ici au XIX^e siècle pour s'intéresser à un sujet parfois oublié, celui de la statistique dite « coloniale ». Les archives, et en particulier certaines déposées très récemment, apportent un éclairage original sur la manière de réaliser des statistiques au temps de l'empire colonial : on y trouve des lettres de responsables mais aussi

d'enquêteurs, des tableaux de suivi, des parcours de collecte... On constate, avant la Seconde Guerre mondiale, la difficulté à mettre en place une organisation adéquate dans les territoires pour produire des statistiques harmonisées, faute de moyens et de compétences, voire de sensibilité aux enjeux. On voit aussi à quel point les concepts, les nomenclatures, se fondent implicitement sur une vision du pays colonisateur pouvant se révéler très décalée par rapport à la réalité du terrain.

Ce numéro N15 met ainsi en lumière de multiples facettes d'une profession qui a vocation à rendre compte de façon objective et aussi rigoureuse que possible du monde qui l'entoure : grandes enquêtes, grandes nomenclatures, contexte institutionnel européen, nouvelles sources de données, nouvelles démarches et environnements de travail, entités voisines de la statistique publique, et pour finir une description historique nous apportant en creux un recul utile sur les exigences et limites du métier.

Bonne lecture !

Pascal Rivière
Directeur de la collection, Insee

Enquêter les personnes sans domicile

Quand la statistique publique rencontre un terrain inhabituel



Pierre-Yves Cabannes* et Thomas Lellouch**

Comment dénombrer et décrire une population qui, par définition, échappe aux cadres habituels de la statistique publique ? L'enquête Sans Domicile, menée en 2025 par l'Insee et la Drees, relève ce défi en interrogeant directement les personnes sans domicile à partir des services d'aide qu'elles fréquentent. Faute de base de sondage, le dispositif repose sur un échantillonnage à plusieurs degrés : agglomérations, services d'aide et individus. Il s'appuie également sur la mise en place de protocoles adaptés à des configurations de terrain très variées (hébergements, distributions de repas, accueils de jour et services itinérants dits « d'aller-vers »), ainsi que sur des outils spécifiques pour enquêter les personnes non francophones. La collecte, qui a mobilisé au printemps 2025 près de 1 000 agents et permis de réaliser environ 17 000 entretiens, illustre la tension permanente entre exigence méthodologique et contraintes opérationnelles. Au-delà des résultats très attendus sur les personnes sans domicile (nombre, profils, conditions de vie, parcours), l'édition 2025 apporte un retour d'expérience inédit sur les conditions concrètes nécessaires pour produire des données fiables sur les personnes vivant hors des logements ordinaires.

 How to count and describe a population that, by definition, falls outside the usual boundaries of official statistics? The national Homeless Survey (*Enquête Sans Domicile*), jointly conducted in 2025 by INSEE and the Directorate of Research, Studies, Evaluation and Statistics (DREES), addresses this challenge by reaching homeless people directly through the support services that they use. As there is no conventional sampling frame, the study relies on a multi-stage sampling design: urban areas, support services and individuals. It also uses protocols adapted to a wide range of field settings (shelters, meal distribution services, day centers and outreach services) as well as tools specifically designed to survey non French-speaking individuals. The data collection involved nearly 1,000 staff and resulted in around 17,000 interviews in the spring of 2025. It highlights the constant tension between methodological requirements and field constraints. Beyond the long-awaited findings on the homeless population — including numbers, profiles, living conditions and trajectories — the 2025 edition provides unprecedented insight on the practical requirements for producing high-quality data on individuals who do not live in ordinary dwellings.

* Chef du bureau Lutte contre l'exclusion, Drees.
pierre-yves.cabannes@sante.gouv.fr

** Directeur de projet Statistiques de la grande pauvreté, DSDS, Insee.
thomas.lellouch@insee.fr

Qui sont les personnes sans logement en France, quelles sont leurs conditions de vie, et combien sont-elles ? Ces questions reviennent régulièrement dans le débat public, au détour d'un reportage lors d'un hiver rigoureux, ou dans les débats politiques avant les élections. Elles paraissent à la fois simples à poser, et essentielles pour mieux connaître la situation des personnes les plus exclues de notre société. Y répondre de manière rigoureuse constitue néanmoins l'un des défis les plus complexes auxquels la statistique publique est confrontée.

Les personnes sans domicile échappent aux dispositifs classiques de connaissance statistique, en particulier les enquêtes en population générale. L'absence de logement stable et la forte mobilité des personnes entre les situations supposent donc d'adapter



Les personnes sans domicile échappent aux dispositifs classiques de connaissance statistique, en particulier les enquêtes en population générale.



profondément les méthodes statistiques habituelles, pour produire une connaissance approfondie de ces publics.

Face à ce défi, la France a fait un choix original : interroger directement les personnes sans domicile à travers une enquête statistique nationale dédiée : l'enquête Sans Domicile¹. Réalisée en 2001 et en 2012 par l'Insee et l'Institut national des études démographiques (Ined)², puis en 2025 par l'Insee et la Direction de la recherche,

des études, de l'évaluation et des statistiques (Drees)³, elle constitue un dispositif singulier, exigeant et ambitieux. Cet article revient sur les principes, les enjeux et les choix méthodologiques de cette enquête, en montrant comment elle tente de concilier rigueur statistique et contraintes concrètes de terrain.

► **Personnes sans abri, personnes sans domicile : de qui parle-t-on ?**

Avant de parler de méthodologie d'enquête, la première question est celle de la définition du public cible. D'un point de vue sémantique, la catégorie des « sans domicile » est définie en creux, par l'absence de domicile. C'est donc une catégorie très hétérogène, puisqu'on part de situations vécues par les personnes à un moment donné, et non de caractéristiques communes des personnes elles-mêmes.

Plusieurs questions se posent pour préciser de manière statistique les contours de cette catégorie. Comment définir l'absence de domicile ? Doit-on privilégier une approche du domicile plutôt « physique » (un toit sur la tête et des équipements de base), « légale » (en lien avec le statut d'occupation) ou « sociale » (qui renvoie à la notion de se sentir « chez soi ») ? Quelle référence temporelle choisir : une photographie à un instant donné ou une observation sur une période de référence ?

¹ Le nom officiel de l'enquête, tel que validé par le Comité du label de la statistique publique (Christine et Roth, 2020), est « enquête auprès des personnes fréquentant des services d'aide », compte tenu de la méthodologie d'enquête et du champ atteint par cette dernière (voir infra).

² Sur les résultats des éditions de 2001 et 2012, voir notamment Brousse (2006a), Join-Lambert (2006), Yaouancq et al. (2013), Yaouancq et Duée (2014).

³ La Drees est le service statistique ministériel dans les domaines de la santé et des solidarités. La description de l'enquête de 2025 a fait l'objet d'un article dans le blog de l'Insee (Lellouch et al., 2025). Voir également le lien suivant : <https://www.insee.fr/fr/information/7634303>.

Entre 1993 et 1996, le groupe de travail du Conseil national de l'information statistique (Cnis)⁴ sur les personnes sans abri s'est penché sur ces questions (Cnis, 1996). Il a établi une classification générale des situations de précarité de logement reposant sur quatre dimensions : l'aspect physique de l'habitat, le statut d'occupation, la qualité du logement et la stabilité/précarité (au sens temporel, c'est-à-dire la garantie ou non de pouvoir demeurer dans son logement au-delà d'une certaine durée). En 2011, le groupe de travail du Cnis sur le mal-logement a mis à jour cette classification et l'a enrichie avec deux autres dimensions : l'environnement du logement et l'adéquation entre le logement et le ménage qui l'occupe (Join-Lambert et al., 2011).

Pour définir le champ des personnes sans domicile, la statistique publique s'est concentrée uniquement sur les deux premières dimensions (type d'habitat et statut d'occupation), notamment pour des raisons opérationnelles⁵, et a stabilisé dans les années 2000 la définition statistique de deux grandes catégories de situations « à un moment donné » (**figure 1**) : celle des **personnes sans abri**, dormant dans la rue ou dans un autre lieu non prévu pour l'habitation (voiture, jardin public, gare, parking, etc.), et celle des **personnes sans domicile**, qui inclut aussi les personnes mises à l'abri dans le cadre d'un dispositif d'hébergement (en centre collectif, à l'hôtel ou dans un logement ordinaire)⁶. Une catégorie plus large de **personnes sans logement personnel** a également

► **Figure 1 - Les différentes situations des personnes sans domicile**



⁴ La mission principale du Cnis est d'assurer la concertation entre les producteurs et les utilisateurs de la statistique publique. Anxionnaz et Maurel (2021) décrivent le rôle et le fonctionnement de cette instance.
⁵ Marpsat (2007) détaille les choix ayant conduit à retenir cette définition ; elle évoque une définition « au sens restreint ».
⁶ La définition des populations « sans domicile » et « sans abri » fait l'objet d'un article du blog de l'Insee (Albouy et al., 2021). Elle est également explicitée dans le document méthodologique de l'enquête de 2001 (Brousse, 2006b).

été définie et permet d'inclure d'autres situations d'exclusion vis-à-vis du logement⁷. Cette classification⁸ n'épuise pas la question des situations frontalières entre les catégories, qui évolue avec le temps, et sera éclairée par la publication des résultats de l'enquête Sans Domicile 2025.

D'autres pays ont élaboré leurs propres définitions. L'harmonisation internationale est complexe, car les contours des catégories sont nécessairement articulés avec les dispositifs nationaux de prise en charge des personnes à la rue. Des avancées significatives ont néanmoins eu lieu au niveau européen. Ainsi, en 2007, la Fédération européenne d'associations nationales travaillant avec les sans-abri (Feantsa) a mis au point une « typologie de l'exclusion liée au logement » appelée ETHOS (pour *European Typology on Homelessness and housing exclusion*). Cette typologie définit une gradation des situations d'absence de logement ou de mal-logement (sans abri, sans logement, logement précaire, logement inadéquat), illustrant par là même le continuum des situations vis-à-vis du logement⁹.

► Connaître les personnes sans domicile : une difficulté inhérente au public cible

Une fois précisée la définition du public cible, comment recueillir de l'information le concernant ?

La difficulté à connaître statistiquement les personnes sans domicile tient d'abord à la nature même de leur situation résidentielle : il n'existe pas de liste exhaustive de lieux d'habitation à partir de laquelle tirer un échantillon représentatif. Autrement dit, il n'y a pas de « base de sondage ». Les méthodes classiques d'enquêtes en population générale, qui sélectionnent les individus à partir d'un répertoire de logements, apparaissent donc inopérantes. Par ailleurs, l'absence de logement stable implique une forte mobilité des personnes et des situations : les personnes peuvent alterner plusieurs fois entre rue, hébergement d'urgence ou hébergement chez un tiers. Cette instabilité rend également très complexe le dénombrement précis de cette population.

Ces difficultés ne sont pas propres à la France. Partout, les instituts statistiques, les pouvoirs publics et les chercheurs se heurtent à des obstacles similaires pour dénombrer et/ou décrire la population sans domicile. Des approches variées ont ainsi été développées pour tenter de contourner ces obstacles¹⁰ (**encadré 1**).

Le **recensement de la population** réalisé dans chaque pays a vocation à couvrir l'ensemble de la population résidant sur le territoire, y compris donc les personnes sans domicile. Mais la méthode de collecte sépare les personnes sans abri du reste des

⁷ Cette catégorie inclut notamment la situation de personnes contraintes de dormir à l'hôtel à leurs frais, d'être hébergées chez des particuliers ou d'occuper un logement sans titre (Rougerie, 2020).

⁸ Pour une discussion sur les enjeux de cette catégorisation, voir par exemple Loison (2023), et Mercier et Montgaud (2025).

⁹ Voir le lien <https://www.feantsa.org/ethos>. Cette approche globale du mal-logement est également portée par la Fondation pour le Logement des Défavorisés (anciennement Fondation Abbé Pierre), qui publie chaque année un rapport sur l'état du mal-logement en France.

¹⁰ Voir notamment Marpsat et Yaouanck (2016), O'Sullivan et al. (2023), Boisseuil et Hinnekint (2025) et Organisation de coopération et de développement économiques [OCDE] (2025) pour les comparaisons internationales.

personnes sans domicile, et les informations collectées sont généralement limitées. Il s'agit donc de mettre en place une approche statistique dédiée pour mesurer plus précisément le phénomène. Voici les principales méthodes utilisées en France et ailleurs :

- **Comptages de rue** : Il s'agit d'une opération à l'échelle locale, à un moment donné. Des équipes parcourent différents secteurs d'un territoire sur une nuit ou une journée pour y dénombrer les personnes à la rue, et le cas échéant leur proposer un court questionnaire. C'est l'objet des Nuits de la Solidarité en France, qui ont lieu dans plusieurs

► **Encadré 1. Outils de mesure et comparaisons internationales**

Les pays ont recours à des méthodes diverses pour appréhender les situations des personnes sans domicile, avec pour chacune des atouts et des limites.

Approche	Exemples en France	Exemples à l'étranger	Ce que la méthode permet	Principales limites
Comptage de rue ponctuel	Nuits de la solidarité	Belgique, Canada, Colombie, États-Unis, Italie, Luxembourg, Royaume-Uni	Estimer le nombre de personnes sans abri sur un territoire à un instant donné	Couverture partielle du champ des personnes sans domicile, non-représentativité au niveau national, peu d'informations individuelles collectées
Enquêtes auprès des structures	Enquête auprès des établissements et services en faveur des adultes et familles en difficulté sociale (ES-DS)	Allemagne, Danemark, Espagne, Finlande, Lituanie	Décrire l'offre d'aide et les grands profils de publics accueillis au niveau national	Couverture partielle du champ des personnes sans domicile, pas d'interrogation des personnes elles-mêmes, peu d'informations individuelles collectées
Données administratives	SI SIAO, logement social	Allemagne, Chili, Danemark, Hongrie, Norvège, Royaume-Uni	Suivre les profils et les parcours au sein de différents dispositifs publics	Couverture partielle du champ des personnes sans domicile, pas d'interrogation des personnes elles-mêmes, peu d'informations individuelles disponibles
Données d'activité des associations	Secours Catholique, Restos du Cœur, Secours Populaire, Banques Alimentaires, etc.	Nombreuses associations dans de nombreux pays	Décrire les volumes et profils des personnes fréquentant différentes associations	Collecte décentralisée, et systèmes d'information différents selon les organismes, statistiquement non représentatif
Enquête par sondage auprès des personnes	Enquête Sans Domicile (2001, 2012, 2025)	Espagne (2005, 2012, 2022), États-Unis (1987, 1996), Italie (2011, 2014)	Décrire finement les profils, parcours et conditions de vie des personnes sans domicile	Coût élevé, complexité

villes, souvent un soir d'hiver¹¹. À Paris, la dernière opération menée en janvier 2026 a conduit au décompte de 3 857 personnes sans abri¹². Des décomptes de ce type ont déjà eu lieu dans de nombreuses villes dans le monde.

- **Enquêtes auprès des responsables de structures** : Les personnes sans domicile fréquentent des services d'aide, par exemple les services d'hébergement ou les accueils de jour. Il est donc possible de collecter des informations sur les effectifs et les profils en interrogeant les responsables de ces structures. Par exemple, en France, la Drees réalise tous les quatre ans une enquête sur le champ des centres d'hébergement : l'enquête auprès des établissements et services en faveur des adultes et familles en difficulté sociale (ES-DS)¹³.

- **Données administratives** : Il existe un certain nombre de bases de données centralisées au niveau des différentes administrations, qui contiennent des informations sur les personnes sans domicile. En France, par exemple, le système d'information des services intégrés d'accueil et d'orientation (SI SIAO) centralise l'ensemble des offres et des demandes d'hébergement des personnes (en urgence via le numéro 115 ou dans le cadre d'une procédure d'insertion avec accompagnement d'un travailleur social).

- **Données d'activité des associations** : Elles permettent de capter le « pouls » de la situation réelle sur le terrain. Cependant, elles sont difficilement exploitables d'un point de vue statistique au niveau national, car le monde associatif est multiple et complexe et les personnes ont souvent recours à différentes associations.

Certaines méthodes visent avant tout à produire un ordre de grandeur du nombre de personnes sans abri, tandis que d'autres privilégient la description des profils des personnes. Aucune n'offre, à elle seule, une estimation rigoureuse des effectifs, des profils et des conditions de vie, qui soit représentative au niveau national et qui couvre l'ensemble du champ des personnes sans domicile.

C'est dans ce contexte que s'inscrit le choix français d'une enquête statistique, directement auprès des personnes sans domicile. Un choix finalement naturel : pour connaître une population, quoi de mieux que de l'interroger directement ?

► Les objectifs de l'enquête Sans Domicile : dénombrer mais surtout mieux connaître les personnes

L'enquête a pour objectif principal de décrire les profils, les parcours et les conditions de vie des personnes sans domicile. Il s'agit notamment de mieux connaître leurs caractéristiques socio-démographiques, leur état de santé, leur relation à l'emploi, leurs revenus, les violences dont elles peuvent être victimes, sujets pour lesquels une comparaison avec la population générale peut être menée. Il s'agit aussi d'aborder des difficultés spécifiques aux personnes sans domicile, comme l'accès aux services de base (eau, alimentation, hygiène...), l'accès au logement, le recours à certaines prestations

¹¹ Voir le guide méthodologique : Délégation interministérielle à l'hébergement et à l'accès au logement [Dihal] et Insee (2021).

¹² Voir le lien : <https://www.paris.fr/pages/nuit-de-la-solidarite-2026-33001>.

¹³ Voir la page de la Drees dédiée à l'enquête : <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/06-lenquete-aupres-des-etablissements-et-services-en-faveur-des-adultes>.

et services et les relations avec les associations et les organismes. Une attention toute particulière est portée aux trajectoires des personnes (logement, emploi ou migrations le cas échéant) et à l'identification des facteurs ayant conduit au processus d'exclusion (par exemple, en abordant les difficultés dans l'enfance)¹⁴.

Moins centrale dans la conception de l'enquête mais très attendue, l'estimation du nombre de personnes sans domicile en France métropolitaine (et, en leur sein, du nombre de personnes sans abri) est un autre objectif important de l'enquête.

► Les principaux enjeux et difficultés d'un projet « hors normes »

La mise en œuvre d'une enquête statistique auprès des personnes sans domicile soulève de très nombreux enjeux, qui s'ajoutent aux difficultés rencontrées dans les enquêtes classiques. Le principal d'entre eux tient à l'absence de « base de sondage », ce qui oblige à recourir à **une méthodologie indirecte pour atteindre les personnes, en passant par les lieux et les services qu'elles fréquentent**.

La diversité des situations observées au sein des différents lieux d'enquête (hébergements, distributions de repas, accueils de jour, services d'aller-vers¹⁵, etc.) complexifie la manière dont on va sélectionner et interroger les personnes sur le terrain. Chaque configuration impose des modalités spécifiques de contact et de tirage au sort et complique l'application de protocoles standardisés.

La proportion importante de personnes non francophones parmi la population cible constitue une difficulté supplémentaire, notamment du fait de l'impossibilité de connaître à l'avance les langues parlées par les personnes à interroger. Cela suppose de disposer d'outils de collecte adaptés.

À ces contraintes méthodologiques et techniques s'ajoutent des enjeux humains et éthiques. Les personnes interrogées se trouvent souvent dans des situations de grande vulnérabilité, avec parfois une forte méfiance vis-à-vis des institutions. La relation d'enquête doit donc être construite avec soin, dans le respect du consentement des personnes et de la confidentialité des informations recueillies. La qualité des données dépend donc étroitement de la capacité des enquêteurs et enquêtrices à instaurer un climat de confiance, et à adopter une posture adaptée vis-à-vis de ce public fragile. La relation de confiance doit également être construite avec les associations.

¹⁴ Cette volonté de mieux comprendre les trajectoires se manifeste également par des travaux prévus après la collecte, via des post-enquêtes et des appariements avec des données administratives, qui permettront d'avoir des éléments de parcours après l'enquête.

¹⁵ Les services d'aller-vers sont des services itinérants, dont le but est d'atteindre des personnes qui ne se rendent pas forcément elles-mêmes au sein des dispositifs. Il peut s'agir de maraudes parcourant l'espace public pour venir en aide aux personnes sans abri, mais aussi de services intervenant dans des lieux de vie informels, de type squats ou bidonvilles, pour améliorer les équipements dans ces lieux ou proposer des accompagnements.

► Un partenariat indispensable avec les acteurs de la lutte contre l'exclusion



L'enquête Sans Domicile repose sur une relation partenariale étroite avec le monde associatif et institutionnel intervenant auprès des personnes sans domicile.



L'enquête Sans Domicile repose sur une relation partenariale étroite avec le monde associatif et institutionnel intervenant auprès des personnes sans domicile. Cette relation est absolument fondamentale pour la réalisation de l'enquête.

Les associations, en particulier, ont joué un rôle central à plusieurs étapes du dispositif. En amont, elles ont été associées au recensement des services d'aide et à la constitution du questionnaire. Leur connaissance fine du terrain a été essentielle pour construire un dispositif de collecte adapté

à la particularité du public cible et à l'activité des services d'aide. Pendant la collecte, le rôle des associations s'est révélé déterminant pour faciliter l'accès aux lieux d'enquête et l'acceptation de l'enquête. À maintes reprises, la présence des équipes associatives a contribué à expliquer la démarche et à lever certaines réticences.

De son côté, la délégation interministérielle à l'hébergement et à l'accès au logement (Dihal) a soutenu fortement le projet dans le cadre d'un partenariat dédié. De nombreux autres acteurs publics (Santé publique France, Samusocial de Paris, collectivités locales, etc.), ainsi que le monde de la recherche, ont également contribué au projet et ont été associés aux travaux.

Cette relation partenariale a toutefois rendu plus difficile le processus de décision. Les attentes des acteurs de terrain peuvent différer de celles de la statistique publique, mais aussi de celles des chercheurs ou des administrations. Par exemple, les associations peuvent être réticentes devant un questionnaire trop détaillé et perçu comme intrusif. Côté recherche, cependant, cela peut être vu comme une source d'information très utile. Quant au statisticien public, il doit tenir compte de considérations essentielles pour la collecte, comme la longueur du questionnaire ou la bonne compréhension des questions. La gouvernance du projet a permis de donner un cadre de décision collectif, notamment à travers le comité scientifique de l'enquête et grâce à de nombreuses consultations bilatérales. Finalement, la co-construction du dispositif apparaît comme un équilibre difficile, mais indispensable, entre exigences statistiques et attentes des partenaires.

► Atteindre les personnes : une couverture largement renforcée en 2025

L'enquête Sans Domicile repose sur un principe central : atteindre les individus à travers les services d'aide qu'ils fréquentent. De ce fait, le champ réel de l'enquête est plus large que celui des personnes sans domicile. En effet, certains utilisateurs des services peuvent par exemple avoir un logement ou être hébergés chez un tiers. Ces utilisateurs sont aussi dans le champ de l'enquête. À l'inverse, l'enquête ne couvre pas toutes les personnes sans domicile : certaines personnes sans abri, marginalisées et/ou dans des stratégies

d'invisibilisation, ne seront pas couvertes par l'enquête puisqu'elles ne fréquentent aucun service.

Le choix des types de service s'est fait en arbitrant entre différents objectifs et contraintes :

- cibler des services fréquentés dans une proportion importante par des personnes sans domicile ;
- choisir des services permettant d'accroître la diversité des profils ;
- pouvoir disposer d'un répertoire exhaustif des établissements proposant les types de service sélectionnés ;
- pouvoir établir un protocole d'enquête robuste compatible avec le fonctionnement normal des services ;
- être en mesure d'assurer un bon taux de réponse pour les questionnaires individuels ;
- limiter le nombre de protocoles et la lourdeur du questionnaire.

En 2012, le champ des services couverts par l'enquête comprenait essentiellement l'hébergement dans les centres et à l'hôtel¹⁶, hors du dispositif national d'accueil des demandeurs d'asile et réfugiés (DNA), et les distributions de repas.

En 2025, la couverture des personnes sans domicile s'est améliorée fortement en intégrant de nouveaux types de service.

En 2025, la couverture des personnes sans domicile s'est améliorée fortement en intégrant de nouveaux types de service. **Deux axes d'amélioration étaient souhaités : mieux couvrir les personnes sans abri et inclure les personnes hébergées au titre du DNA.**

En effet, en 2012, les personnes sans abri n'étaient atteintes que via les distributions de repas et les haltes de nuit, alors qu'une part importante des

personnes sans abri ne fréquentent pas ces services. Pour l'édition 2025, les accueils de jour, les services d'aller-vers, ainsi que des points fixes d'accueil en rue (sans distribution de repas) ont été ajoutés au champ de l'enquête. L'intégration des services d'aller-vers a aussi permis d'intégrer un certain nombre de situations frontalières, en interrogeant directement des personnes vivant dans des squats, campements ou bidonvilles, puisque certains services sont spécialisés sur ces lieux. Les accueils de jour permettent, eux, de passer dans de meilleures conditions des questionnaires aux personnes sans abri, et avec un bon taux de réponse.

Quant aux établissements du DNA, les personnes qui y sont hébergées étaient prises en compte dans le dénombrement des personnes sans domicile résultant de l'enquête de 2012, mais elles n'avaient pas été interrogées à l'époque. Tous ces nouveaux champs ont été testés sur le terrain en amont de la collecte.

¹⁶ Il s'agit notamment de l'hébergement dit « généraliste », par exemple en centre d'hébergement et de réinsertion sociale (CHRS) ou en centre d'hébergement d'urgence (CHU), de l'hébergement médico-social ou de l'hébergement de familles dans le cadre de l'aide sociale à l'enfance (ASE).

La couverture de l'enquête s'est aussi améliorée à d'autres niveaux. Ainsi, alors que le champ portait en 2012 sur les personnes de 18 ans ou plus dans les agglomérations de 20 000 habitants ou plus, celui de 2025 porte sur les personnes de 16 ans ou plus dans les agglomérations de 5 000 habitants ou plus. L'extension aux personnes de 16 et 17 ans permet notamment d'intégrer une partie des mineurs non accompagnés (MNA).

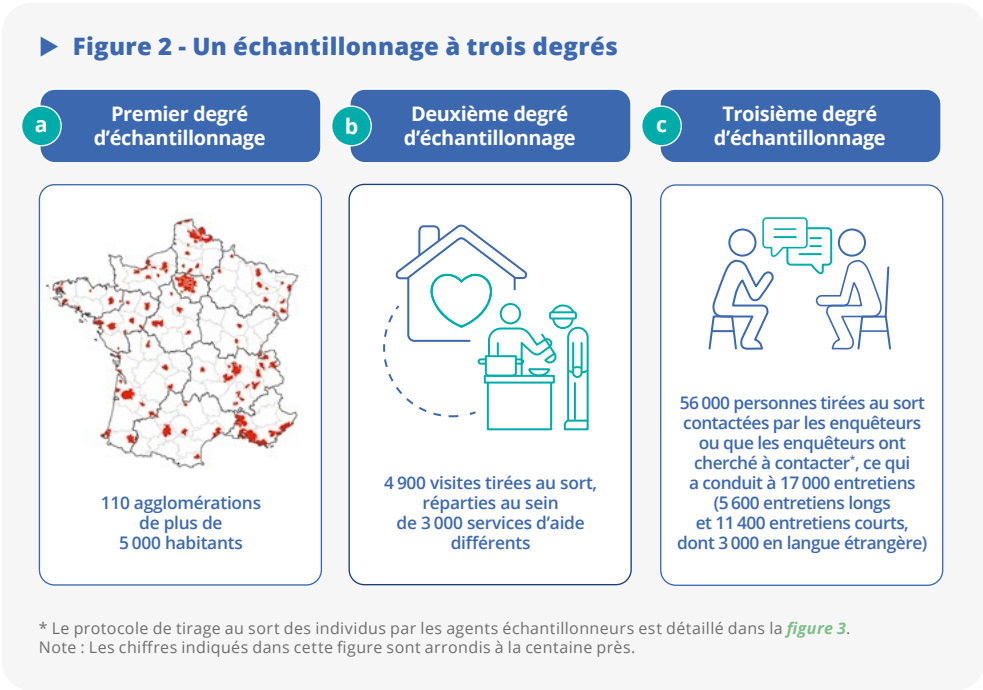
Au total, le champ de l'enquête Sans Domicile de 2025 est l'ensemble des personnes de 16 ans ou plus, vivant dans une agglomération de 5 000 habitants ou plus, et qui sont utilisatrices d'un des services d'aide répertoriés.

► **Un échantillon construit en trois temps**

La constitution de l'échantillon de l'enquête Sans Domicile est très particulière, à la fois car les personnes sont approchées via la fréquentation de certains services d'aide et car la dernière étape de l'échantillonnage est effectuée sur le terrain par des enquêteurs ou des agents de bureau.

L'échantillon de l'enquête Sans Domicile est constitué en trois temps (*figure 2*).

Tout d'abord, 110 agglomérations (« unités urbaines »)¹⁷ de France métropolitaine d'au moins 5 000 habitants ont été tirées au sort en octobre 2023. Ce tirage est à probabilités proportionnelles à la population municipale et équilibré spatialement et selon des variables d'effectifs estimés de certaines catégories de personnes sans



¹⁷ Voir la définition au lien suivant : <https://www.insee.fr/fr/metadonnees/definition/c1501>.

domicile. Les agglomérations de plus de 200 000 habitants ont été sélectionnées d'office, de même que les agglomérations de Calais et Dunkerque. Pour ces dernières, l'enjeu était de pouvoir intégrer dans les statistiques nationales les personnes sans domicile en transit vers l'Angleterre, dont le profil est spécifique.

Un répertoire exhaustif des structures portant des services d'aide dans le champ de l'enquête a ensuite été constitué au sein de ces 110 agglomérations. Ces structures ont été interrogées en 2024 dans le cadre du premier volet de l'enquête, dit volet « Structures ». Réalisé un an avant la collecte auprès des personnes, par Internet ou téléphone, ce volet consiste à interroger les responsables de structures sur le type de services qu'elles proposent, les jours et horaires d'ouverture, ainsi que sur les effectifs et les grands profils des publics qu'elles accueillent.

Cette opération soulève des difficultés spécifiques, car les services d'aide forment un paysage institutionnel hétérogène et évolutif. Cela rend très difficile en particulier la définition d'une typologie des services d'aide adaptée à la complexité des différentes situations. Prenons une maraude qui fait plusieurs arrêts au cours de son parcours pour distribuer de la nourriture : doit-elle être comptabilisée en tant que distribution de repas ou en tant que service itinérant ? Il n'était ainsi pas toujours aisé de faire entrer toutes les situations dans les cases théoriquement étanches de notre typologie. Cette complexité a été in fine source de difficultés lors de la collecte sur le terrain en 2025, car le type de certains services n'avait pas toujours été correctement identifié.

Grâce aux réponses au volet « Structures », une base d'environ 6 000 services d'aide a été constituée. **À partir de cette base, 4 919 « visites »¹⁸ réparties dans 3 028 services ont été tirées au sort en janvier 2025**, pour y mener le volet « Individus ». Une visite correspond à un jour d'ouverture d'un service donné. Le tirage au sort s'est fait selon une probabilité proportionnelle à l'effectif fréquentant le service estimé d'après le volet « Structures ».

Ce tirage au sort constitue une étape clef du dispositif d'enquête, située à l'interface entre la conception méthodologique et la mise en œuvre concrète de la collecte. Certains types de service, accueillant des profils spécifiques, ont ainsi été sur-représentés : c'est le cas par exemple des accueils de jour ou des maraudes, qui sont fréquentés par plus de personnes sans abri. La sélection des services doit également composer avec des contraintes opérationnelles. Par exemple, les maraudes après minuit ou les services ouverts uniquement le dimanche ont été exclus d'emblée de la base pour des raisons de faisabilité. Cette étape met une nouvelle fois en lumière la tension permanente entre rigueur statistique et faisabilité de terrain.

Enfin, **le tirage au sort des individus a été effectué sur le terrain au printemps 2025** par un agent échantillonneur, pendant la phase de collecte du volet Individus. Un outil (« table de tirage ») a été fourni aux échantillonneurs afin qu'ils puissent réaliser ce tirage au sort le plus facilement possible (voir infra), même si, en pratique, la réalité du terrain s'est révélée parfois complexe.

18 Les visites (autrement appelées « service-jour ») correspondent à un jour d'ouverture d'un service donné. La base des services a été convertie en base des services-jours grâce aux informations sur les jours d'ouverture données dans le volet « Structures ». Le tirage a été ensuite effectué au niveau « visite » (et non « service »).

► Dans chaque service, définir un protocole adapté

Les types de service dans le champ de l'enquête sont très divers. Pour chacun d'eux, un protocole d'enquête a été élaboré, visant à la fois à répondre aux besoins méthodologiques (garantir un tirage aléatoire, récolter les informations nécessaires à l'organisation de la collecte et à la pondération des questionnaires – voir infra) et aux

“

En hébergement, le tirage s'est fait sur une liste des personnes de 16 ans ou plus hébergées dans le service. Hors hébergement, le tirage s'est fait en constituant une « file d'attente », réelle ou théorique.

”

objectifs de rendement de l'enquête, mais aussi à s'adapter aux contraintes opérationnelles du terrain. En effet, il était primordial de ne pas perturber le fonctionnement des services et de mettre les équipes de l'Insee dans les meilleures conditions possibles sur place.

De grandes familles de protocole ont pu être constituées. Tout d'abord, les services se distinguent selon la méthode de tirage au sort des personnes (*figure 3*). En hébergement, le tirage s'est fait sur une liste des personnes de 16 ans ou plus hébergées dans le service. Hors hébergement, il n'y a pas de liste et le tirage s'est fait en constituant une « file d'attente », réelle (par exemple, les personnes qui

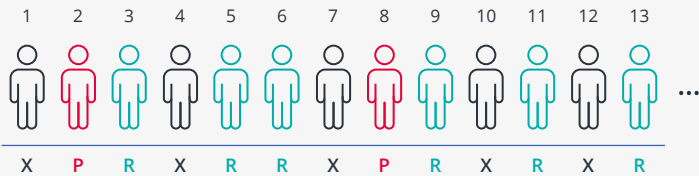
attendent dans une distribution de repas) ou bien théorique (par exemple, les personnes rencontrées au fur et à mesure du parcours d'une maraude). Dans tous les cas, la table de tirage fournie aux agents échantillonneurs leur a permis de réaliser le tirage au sort.

Au sein de l'hébergement, une distinction importante se fait entre l'hébergement dans un même lieu de type collectif (« hébergement regroupé ») et l'hébergement éclaté en plusieurs appartements dans le parc de logements ordinaires (« hébergement diffus »). L'impossibilité d'envoyer en même temps une équipe Insee dans tous les appartements d'un service d'hébergement diffus a ainsi conduit à effectuer un tirage au sort des personnes en amont de la visite, à prendre des rendez-vous avec les personnes sélectionnées et à autoriser les équipes d'enquêteurs à se séparer.

Hors hébergement, les services se distinguent notamment selon le fait d'être en intérieur (par exemple, les accueils de jour) ou en extérieur (par exemple, les points fixes d'accueil en rue), et d'être fixes (par exemple, les distributions de repas) ou itinérants (les maraudes). Ces différences impliquent des contraintes logistiques liées aux conditions de collecte sur le terrain. En intérieur, il est assez souvent possible d'avoir un endroit pour s'asseoir et passer les questionnaires, voire pour s'isoler. Ce peut être nettement plus compliqué en extérieur, où le service et les enquêteurs sont aussi soumis aux aléas climatiques. Dans les services itinérants, il faut s'assurer de pouvoir suivre les maraudes (pour certaines, motorisées) et de les retrouver en cas de séparation. Il faut aussi ne pas être trop nombreux, pour faciliter le contact avec des personnes sans abri isolées. Cela a ainsi conduit à n'envoyer qu'un seul enquêteur dans les maraudes, en plus d'un agent échantillonneur.

► **Figure 3 - Échantillonnage des individus dans les structures d'accueil**

Hors hébergement



Selon la table de tirage suivante :

P	2				8			
R		3	5	6		9	11	13

Hébergement

Identifiant	N° de chambre	Prénom	Nom	Tirage
1	101	Gaston	Lagaffe	P
2	101	Lucky	Luke	X
3	101	Capitaine	Haddock	X
4	102	Tryphon	Tournesol	R
5	102	Yoko	Tsuno	R
6	201	Roberto	Rastapopoulos	X
7	201	Dupond	Dupont	R
8	201	Bianca	Castafiore	X
9	202	Spirou	Fantasio	P
10	202	Benoît	Brisefer	R
11	202	Michel	Tanguy	X
12	301	Adèle	Blanc-sec	R
...	...	...	...	...

Selon la table de tirage suivante :

P	1				9		
R		4	5	7		10	12

Note : Pour effectuer le tirage, l'agent échantillonneur s'appuie sur une table de tirage constituée de numéros tirés aléatoirement.
Les notations s'expliquent de la manière suivante :
• X : L'individu n'est pas contacté par l'enquêteur.
• P : Rang principal. Un questionnaire est systématiquement proposé à l'individu situé à ce rang.
• R : Rang de remplacement. Un questionnaire est proposé à l'individu situé à ce rang si le rang principal précédent et tous les rangs de remplacement intercalés entre lui-même et le rang principal n'ont pas répondu à l'enquête (refus de répondre, impossibilité de contact, langues de l'enquête non parlées...).

Hors hébergement, il y a parfois plusieurs « sous-services » pour le service et le jour sélectionnés. Par exemple, plusieurs sorties de maraude peuvent avoir lieu le même jour, les distributions de repas du soir d'un service peuvent avoir lieu dans plusieurs endroits en même temps. Il n'était matériellement pas possible de suivre tous ces sous-services le même jour. Aussi, un tirage aléatoire du sous-service à enquêter le jour de la visite a été effectué en amont de la visite.

Au bout du compte, malgré tous les aspects auxquels l'équipe de conception a pu penser, se basant notamment sur les tests menés, les protocoles se sont parfois heurtés lors de la collecte à des difficultés du terrain. Par exemple, en accueil de jour, la constitution d'une file théorique pour le tirage au sort s'est parfois révélée particulièrement complexe lorsque les personnes arrivaient en très grand nombre au même moment, lorsqu'elles



La collecte a confirmé que certains protocoles conçus comme robustes sur le papier étaient beaucoup plus difficiles à appliquer dans les configurations de terrain inhabituelles.



revenaient plusieurs fois dans l'accueil lors d'une même visite ou lorsque l'équipe de collecte arrivait dans un lieu alors que des personnes étaient déjà présentes. Les protocoles se sont heurtés aussi à d'autres imprévus, par exemple la difficulté de repérer les personnes à interroger lorsque celles-ci changent d'habits dans l'accueil de jour (si celui-ci propose des douches par exemple).

Dans tous les cas, il a fallu que les échantillonneurs et les enquêteurs témoignent d'une certaine capacité d'adaptation et la coordination avec le service a été primordiale sur cet aspect aussi. La collecte a confirmé que certains protocoles conçus comme robustes sur le papier étaient beaucoup plus difficiles à appliquer dans les configurations de terrain inhabituelles.

► L'interrogation des individus

L'interrogation des individus constitue le cœur de l'enquête Sans Domicile. C'est à cette étape que sont recueillies les informations sur les personnes sans domicile. Elle repose sur une interrogation en face-à-face par un enquêteur de l'Insee d'une personne fréquentant un service d'aide, sélectionnée au hasard. L'entretien se base soit sur un questionnaire long, d'une durée d'une heure environ, disponible uniquement en français, soit sur un questionnaire court d'une trentaine de minutes, disponible en français et dans 21 langues étrangères (**encadré 2**).

Les deux questionnaires sont généralistes et traitent des mêmes thématiques, le questionnaire long étant plus détaillé que le court sur chaque thème. Situation familiale, santé, revenus, emploi, éducation, ou encore famille et relations sociales : les questionnaires abordent de nombreux domaines. Des modules spécifiques sur les parcours (parcours résidentiels, parcours migratoires), ainsi que sur la santé mentale et les violences, ont été développés par rapport aux éditions précédentes. D'autres modules portent enfin sur la fréquentation des services, les conditions de vie et l'accès aux services de base : se nourrir, se laver, charger son téléphone, recevoir du courrier, etc. L'enquête vise ainsi à mettre en lumière les difficultés de la vie quotidienne.

La conception des questionnaires a fait l'objet d'un travail approfondi, associant statisticiens publics, administrations, chercheurs et associations. Les formulations ont été testées sur le terrain plusieurs fois et ajustées afin d'être compréhensibles par des publics très variés. Une attention particulière a été portée à la clarté des questions, à l'ordre des modules et à la progressivité des thèmes abordés, afin de faciliter le déroulement de l'entretien.

► Encadré 2. Des « questionnaires qui parlent » pour enquêter les personnes non francophones

Une part importante des personnes sans domicile ne maîtrise pas suffisamment le français pour répondre à un questionnaire. Sans dispositif spécifique, ces personnes seraient mécaniquement exclues de l'enquête, introduisant un biais important dans les résultats. L'édition de 2012 avait introduit des questionnaires dans 14 langues étrangères : le questionnaire (sous format papier) était traduit et distribué aux personnes non francophones échantillonnées, qui devaient le remplir elles-mêmes puis le renvoyer à l'Insee ou le remettre à l'enquêteur. À l'époque, près de 1 200 questionnaires en langue étrangère avaient été collectés.

L'édition 2025 a permis des innovations techniques importantes sur ce point. Les questionnaires ont été informatisés et des outils ont été développés pour que les tablettes des enquêteurs puissent lire les fichiers audio dans la langue de l'enquêté. Les questions ont été pré-enregistrées dans 21 langues étrangères et diffusées oralement aux enquêtés. Ces derniers pouvaient écouter les questions dans leur langue sous casque ou grâce au haut-parleur de la tablette.

Concrètement, lorsqu'une personne tirée au sort était identifiée comme non francophone, l'enquêteur déterminait dans un premier temps sa langue grâce à un document papier visuel, le cahier multi-langues, permettant à l'enquêté d'indiquer la langue qu'il parlait le mieux. Puis, l'enquêteur renseignait sur la tablette cette langue

et présentait la tablette à l'enquêté, qui répondait lui-même en mode auto-administré sur l'outil. L'enquêteur restait bien sûr à côté le temps de l'entretien pour pouvoir assister techniquement l'enquêté, en cas de difficulté sur l'outil. Au global, environ 3 000 questionnaires en langue étrangère ont été collectés.

La collecte sur le terrain a déployé ce dispositif à grande échelle : si les outils ont globalement bien fonctionné, les difficultés n'ont pas manqué. En particulier, la durée d'interview pour les non-francophones a été largement sous-estimée (de l'ordre du double de celle anticipée). La longueur et la complexité de certaines questions, tout comme la difficulté de traduire certains termes très spécifiques (par exemple, ceux liés à des dispositifs institutionnels propres à la France), obligeaient parfois l'enquêté à devoir passer beaucoup de temps sur une question. Il était par ailleurs difficile pour l'enquêteur de pouvoir aider l'enquêté en cas de blocage, faute d'une langue commune.

Malgré ces difficultés, le dispositif a permis d'améliorer la couverture de l'enquête, notamment en intégrant les personnes ne sachant pas lire ou écrire, et devrait renforcer significativement la qualité des données collectées pour les personnes non francophones. Il illustre la capacité de la statistique publique à innover pour répondre à des contraintes de terrain fortes, mais au prix d'une complexité accrue du dispositif et de coûts supplémentaires.

La relation enquêteur-enquêté constitue un enjeu central de cette phase. Les personnes interrogées peuvent se trouver dans des situations de grande vulnérabilité, marquées par des expériences de vie très difficiles, voire traumatisantes. En cas de forte vulnérabilité identifiée, les enquêteurs ont pu distribuer aux enquêtés un flyer avec des numéros d'aide d'urgence. Ces entretiens ont parfois généré une charge émotionnelle forte pour des enquêteurs, amenés souvent à enchaîner plusieurs entretiens dans la même journée, pendant une période de collecte qui a duré trois mois. Un dispositif de soutien psychologique a donc été prévu pour accompagner les agents de l'Insee qui en ressentaient le besoin.

► Une collecte exceptionnelle par son ampleur et par l'engagement des équipes

La phase de collecte de l'enquête Sans Domicile constitue l'aboutissement opérationnel de plusieurs années de préparation, avec une charge de travail très importante. Elle s'est déroulée entre le 31 mars et le 12 juillet 2025 et a mobilisé près de 1 000 agents de la statistique publique, dont environ 500 enquêteurs et enquêtrices de l'Insee et autant d'agents de bureau échantillonneurs¹⁹. Cette phase, particulièrement intense sur le plan organisationnel et humain, a été marquée par une très forte dimension collective, aussi bien entre agents sur le terrain qu'au niveau du pilotage articulé entre la direction générale de l'Insee et les directions régionales.

La collecte repose en effet sur une articulation étroite entre différentes catégories d'agents :

- les enquêteurs de l'Insee qui interrogent les personnes ;
- les agents accompagnateurs qui les échantillonnent²⁰ ;
- les gestionnaires en bureau chargés de la formation des enquêteurs et de l'organisation logistique de la collecte ;
- les équipes de conception responsables du suivi national de la collecte et de l'assistance aux équipes de terrain sur les sujets méthodologiques.

La coordination entre tous ces acteurs a été essentielle pour garantir le respect des protocoles, mais aussi l'efficacité de la collecte sur le terrain.

Pour les gestionnaires en bureau des directions régionales et les équipes de conception, les conditions d'organisation de la collecte ont été difficiles, notamment juste avant et au début de la collecte, avec une charge de travail importante. Il y a eu de nombreux imprévus sur la phase de premier contact avec les services, par rapport aux informations connues avant

“

Pour les enquêteurs, les conditions de collecte ont souvent été exigeantes. Ils ont dû faire preuve d'une forte capacité d'adaptation.

”

la collecte : fermetures de services, changements de type de service, doublons ou fusions de service, changements de jours d'ouverture, indisponibilité des responsables. Ceci a conduit à un travail significatif d'assistance pour répondre aux multiples questions des enquêteurs.

Pour les enquêteurs, les conditions de collecte ont souvent été exigeantes. Enquêter debout dans la rue, parfois en période de canicule, dans des lieux très variés, y compris tôt le matin ou tard le soir : ces situations ne sont pas aisées pour les enquêteurs qui ont dû faire preuve d'une forte capacité d'adaptation. Il convient aussi de mentionner les enjeux de sécurité, notamment en cas de situations de tensions sur le terrain. Par exemple, la collecte a dû être abandonnée à mi-période près de Grande-Synthe, dans le département du Nord, car les conditions de sécurité des agents n'étaient pas réunies.

¹⁹ Dans le cadre d'un partenariat avec l'École nationale de la statistique et de l'administration économique (Ensaie), des étudiants ont également fait partie des échantillonneurs sur le terrain.

²⁰ Dans les services hors hébergement.

Faire face à de nombreuses situations imprévues, tout en maintenant un niveau élevé d'exigence méthodologique : c'était l'un des défis de l'enquête Sans Domicile. En dépit de toutes ces difficultés, la collecte a pu bénéficier d'un très fort engagement des agents de la statistique publique. Celui-ci tient à la conscience des enjeux sociaux du projet et à la volonté de contribuer à une meilleure connaissance des situations de grande pauvreté. Cet engagement constitue un facteur clé de la réussite de l'opération.

Les résultats de collecte sont à la hauteur de la mobilisation des équipes : sur environ 3 500 visites, environ 17 000 entretiens ont été finalement réalisés avec des personnes fréquentant des services d'aide, dont plus de 5 600 questionnaires longs. Les objectifs ont ainsi été dépassés.

► Un aval statistique particulièrement complexe

Comme après la collecte de chaque enquête, les travaux de redressement des données ont débuté, mêlant des phases d'apurement des questionnaires individuels, de traitement de la non-réponse et de calcul de pondérations. Pour l'enquête Sans Domicile, la mise en pratique de ces phases se révèle particulièrement complexe, d'autant que le traitement statistique comporte aussi des phases propres à cette enquête. En particulier, il faut également apurer la sur-enquête (**encadré 3**), dont les résultats servent au calcul des pondérations, mais aussi, toujours dans l'optique de ce calcul, traiter les « anomalies de collecte » concernant les services enquêtés et réaliser un « partage des poids » (voir infra).

Les « anomalies de collecte » correspondent à des situations relatives aux services d'aide non prévues avant la collecte. Par exemple, des services considérés initialement comme distincts n'étaient en fait qu'un seul et même service, ou bien ont fusionné entre-temps. Des services ont aussi changé de jours d'ouverture. De manière plus classique, il faut également traiter l'absence de réponse de certains services au moment du volet « Individus », que ce soit parce qu'ils se sont révélés hors champ (par exemple, en raison d'un changement d'activité) ou en tant que non répondants (par exemple, s'ils ont refusé de participer à l'enquête).

► Encadré 3. Une sur-enquête pour pouvoir pondérer les questionnaires

Pour pouvoir exploiter l'enquête et s'assurer que l'ensemble de la population est bien représenté, il faut attribuer un poids à chaque répondant. Contrairement aux enquêtes traditionnelles, l'absence de base de sondage implique que la probabilité réelle de tirage au sort des personnes n'est pas connue avant la visite sur le terrain. Dans le cas de l'enquête Sans Domicile, elle dépend directement de la fréquentation réelle et du nombre de personnes contactées le jour de la visite, qui sont inconnus avant de réaliser la visite.

Pour cette raison, comme pour l'enquête de 2012, une « sur-enquête » a été menée par les enquêteurs, via des outils ad hoc. Il s'agissait notamment de collecter des informations sur la fréquentation

le jour de la visite et sur le sexe et le caractère francophone (ou non) des personnes contactées pour se voir proposer un questionnaire (y compris celles qui n'ont pas répondu). Ces informations ont été recueillies auprès des responsables de services d'aide et des personnes contactées pour passer un questionnaire individuel.

Entre le tirage au sort des personnes à interroger dans les services d'hébergement, l'organisation de la collecte, la coordination avec les différents acteurs (services, autres enquêteurs, direction régionale de l'Insee) et cette sur-enquête, le rôle des enquêteurs dans le cadre de l'enquête Sans Domicile dépasse donc largement la seule passation de questionnaires auprès des individus !

Lors de l'apurement des questionnaires individuels, une attention toute particulière est portée aux questions relatives au lieu où la personne a passé la nuit précédant l'enquête et au statut d'occupation de ce lieu. En effet, ces questions permettent de déterminer qui est sans abri, qui est sans domicile et qui ne l'est pas. Une attention importante est par ailleurs portée aux questions sur la fréquentation des services d'aide durant la semaine précédant l'enquête. Celles-ci servent à estimer le nombre de « liens », i.e. le nombre de fois où la personne a fréquenté un service dans le champ de l'enquête au cours de la collecte, élément entrant en compte dans le calcul des pondérations.

Grâce à la sur-enquête, les poids de sondage des questionnaires individuels peuvent ensuite être calculés, la non-réponse des individus traitée et des marges de calage au niveau des « prestations »²¹ estimées. L'étape de « partage des poids », qui consiste à diviser les poids au niveau des prestations par le nombre de « liens », permet d'aboutir à des pondérations en termes de personnes. Cette phase permet aussi de corriger les doubles comptes, des personnes pouvant par exemple fréquenter à la fois un centre d'hébergement et un accueil de jour. En effet, plus une personne fréquente des services dans le champ de l'enquête, plus sa probabilité d'être sélectionnée est importante et donc plus le poids de son questionnaire doit être diminué.

Une fois l'aval statistique finalisé, les premières études seront lancées et les données diffusées, avec notamment la mise en place d'un comité d'exploitation de l'enquête.

► Un projet emblématique pour la statistique publique —

L'enquête Sans Domicile illustre les défis auxquels la statistique publique est confrontée lorsqu'elle cherche à aller au-devant de tous les publics, y compris les personnes les plus exclues de notre société. Derrière une question en apparence



L'enquête Sans Domicile illustre les défis auxquels la statistique publique est confrontée lorsqu'elle cherche à aller au-devant de tous les publics, y compris les personnes les plus exclues de notre société.



simple – combien de personnes vivent sans logement et dans quelles conditions – se cache un dispositif méthodologique très sophistiqué, mobilisant des choix de conception difficiles et un investissement humain considérable. Le coût financier de l'enquête est lui aussi important, estimé à environ 11 millions d'euros, à la hauteur de l'ambition du projet.

En choisissant d'interroger directement les personnes sans domicile, la statistique publique accepte de s'éloigner de son cadre habituel d'enquête, et de confronter ses outils à des réalités de terrain exigeantes. Ce choix impose

également d'accepter des compromis entre pureté théorique et faisabilité concrète, sans renoncer à l'exigence de rigueur. La collecte de 2025 a en particulier permis de mieux mesurer l'ampleur des adaptations nécessaires sur le terrain.

²¹ Une prestation correspond à l'usage d'un service donné un jour donné. Par exemple, si un centre est ouvert 90 nuits et héberge chaque nuit 100 personnes, alors il y a 9 000 prestations.

L'exemple de l'enquête Sans Domicile révèle l'ampleur du travail nécessaire pour produire des données fiables sur des populations très spécifiques. Il montre que la connaissance statistique ne va pas de soi et qu'elle repose sur une succession de décisions méthodologiques, souvent invisibles, mais déterminantes pour la qualité des données produites. Les premiers résultats de l'enquête sont prévus pour la fin de l'année 2026.

Remerciements

Cette enquête n'aurait pu aboutir sans l'engagement exceptionnel de l'ensemble des agents et partenaires qui y ont contribué. Nous souhaitons tout d'abord saluer le travail remarquable des centaines d'agents mobilisés pendant la phase de terrain – enquêtrices et enquêteurs, échantillonneuses et échantillonneurs, ainsi que gestionnaires en bureau –, dont l'implication a été déterminante pour la réussite du projet. Nous remercions aussi vivement l'ensemble des équipes informatiques, méthodologiques et fonctions support de l'Insee ayant permis la mise en place du dispositif de collecte, ainsi que le soutien des directions de l'Insee et de la Drees tout au long du projet.

Une mention particulière revient aux équipes de conception Insee/Drees et du pôle des enquêtes nationales ménages, et en particulier à Opale Echegu, Thierry Mainaud, Léa Mauro, Isabelle Massini et Christèle Woeber qui ont travaillé plusieurs années sur l'enquête. La contribution des équipes, essentielle à chaque étape du projet, a été marquée par un engagement et une motivation absolument exceptionnels, ainsi que par une capacité constante à innover et à s'adapter pour relever les nombreux défis méthodologiques et logistiques.

Nous adressons également notre reconnaissance aux personnels des services d'aide, qui ont accueilli et facilité le travail des équipes de la statistique publique pendant la collecte, ainsi qu'à l'ensemble des institutions et associations partenaires, pour leur soutien, leur disponibilité et leurs conseils tout au long de ce projet.

Enfin, nous souhaitons remercier chaleureusement les personnes usagères des services d'aide qui ont accepté de participer à l'enquête. Leur contribution a été évidemment indispensable pour produire des données permettant de mieux comprendre leurs conditions de vie.

► Bibliographie

- ALBOUY, Valérie, LEGLEYE, Stéphane et LELLOUCH, Thomas, 2021. Connaître les personnes sans domicile est encore plus important que les dénombrier. In : *Le blog de l'Insee*. [en ligne]. 13 avril 2021. Insee. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://blog.insee.fr/connaître-les-personnes-sans-domicile-est-encore-plus-important-que-les-denombrier/>.
- ANXIONNAZ, Isabelle et MAUREL, Françoise, 2021. Le Conseil national de l'information statistique – La qualité des statistiques publiques passe aussi par la concertation. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2021. Insee. N° N6, pp. 123-142. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5398693?sommaire=5398695>.
- BOISSEUIL, Clément et HINNEKINT, Chloé, 2025. Errance et marginalités dans huit métropoles internationales. In : *Site de l'Apur*. [en ligne]. Novembre 2025. N° 279. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.apur.org/fr/population-societe/social-inclusion/errance-marginalites-huit-metropoles-internationales>.
- BROUSSE, Cécile, 2006a. Devenir sans-domicile, le rester : un problème lié à l'accès au logement ou à la rupture des liens sociaux et familiaux ? In : *Économie et Statistique*. [en ligne]. 1^{er} octobre 2006. Insee. N° 391-392. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/1376415?sommaire=1376427>.
- BROUSSE, Cécile, 2006b. L'enquête sans-domicile 2001. Première partie : Définition de la population sans-domicile et choix de la méthode d'enquête. In : *Insee Méthodes*. [en ligne]. 1^{er} août 2006. Insee. N° 116. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/2579262>.
- CHRISTINE, Marc et ROTH, Nicole, 2020. Le Comité du label – Un acteur de la gouvernance au service de la qualité des statistiques publiques. In : *Courrier des statistiques*. [en ligne]. 31 décembre 2020. Insee. N° N5, pp. 39-52. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5008698?sommaire=5008710>.
- CNIS, 1996. *Pour une meilleure connaissance des sans-abri et de l'exclusion du logement, rapport final du groupe de travail sur les sans-abri*. ISBN 2-11-066393-6.
- DIHAL et INSEE, 2021. Nuit nationale de la solidarité – Compter pour agir – Guide méthodologique. In : *Insee Méthodes*. [en ligne]. 20 mai 2021. N° 139. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5370337>.
- JOIN-LAMBERT, Marie-Thérèse, 2006. Une enquête d'exception – Sans-abri, sans-domicile : des interrogations renouvelées. In : *Économie et Statistique*. [en ligne]. 1^{er} octobre 2006. Insee. N° 391-392. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/1376412?sommaire=1376427>.
- JOIN-LAMBERT, Marie-Thérèse, LABARTHE, Julie, MARPSAT, Maryse et ROUGERIE, Catherine, 2011. Le mal-logement. In : *Rapport d'un groupe de travail du Cnis*. [en ligne]. Juillet 2011. Cnis. N° 126. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.bns.insee.fr/ark%3A/12148/bc6p070pd98.texteImage>.

- LELLOUCH, Thomas, MAINAUD, Thierry, CABANNES, Pierre-Yves et ECHEGU, Opale, 2025. Trouver un lieu pour dormir, se nourrir, se laver, maintenir des liens : une enquête pour connaître les difficultés des personnes sans domicile. In : *Le blog de l'Insee*. [en ligne]. 31 mars 2025. Insee. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://blog.insee.fr/enquete-sans-domicile-2025/>.
- LOISON, Marie, 2023. Le sans-domicilisme. Réflexion sur les catégories de l'exclusion du logement. In : *Sans-abri : expériences et politiques*. 24 avril 2023. Drees. Revue française des affaires sociales. N° 2023/1, pp. 29-50.
- MARPSAT, Maryse, 2007. Explorer les frontières. Recherches sur des catégories « en marge ». In : *Documents de travail*. [en ligne]. 7 mai 2007. Ined. N° 145. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.ined.fr/fr/publications/editions/document-travail/explorer-les-frontieres-recherches-sur-des-categories-en-marge>.
- MARPSAT, Maryse et YAOUANCQ, Françoise, 2016. Avant-propos – L'enquête Sans-Domicile 2012 : histoire et place en Europe. In : *Économie et Statistique*. [en ligne]. Insee. N° 488-489, pp. 7-23. [Consulté le 11 mai 2026]. Disponible à l'adresse : https://www.persee.fr/doc/estat_0336-1454_2016_num_488_1_10708.
- MERCIER, Alice et MONTIGAUD, Lucile, 2025. Le sans-domicilisme à Marseille – Classifications, chiffres et perspectives. In : *Rapport de l'Observatoire des Pauvretés*. [en ligne]. Octobre 2025. [Consulté le 11 mai 2026]. Disponible à l'adresse : <https://www.uriopss-pacac.fr/actualites/rapport-de-lobservatoire-des-pauvretes-le-sans-domicilisme-marseille-classifications>.
- OCDE, 2025. OECD Monitoring Framework to Measure Homelessness. In : *OECD publishing*. [en ligne]. 28 janvier 2025. [Consulté le 23 avril 2026]. Disponible à l'adresse : https://www.oecd.org/en/publications/oecd-monitoring-framework-to-measure-homelessness_3e98455b-en.html.
- O'SULLIVAN, Eoin et al., 2023. Homelessness in the European Union. In : *Study requested by the PETI committee*. [en ligne]. 22 novembre 2023. Parlement européen. [Consulté le 23 avril 2026]. Disponible à l'adresse : [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2023\)755915](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2023)755915).
- ROUGERIE, Catherine, 2020. Difficultés de logement ; les situations de logement appréciées selon les dimensions proposées par le Conseil national de l'information statistique. In : *Documents de travail*. [en ligne]. 9 novembre 2020. Insee. N° F2020-02. [Consulté le 11 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4770825>.
- YAOUANCQ, Françoise, LEBRÈRE, Alexandre, MARPSAT, Maryse, RÉGNIER, Virginie, LEGLEYE, Stéphane et QUAGLIA, Martine, 2013. L'hébergement des sans-domicile en 2012 – Des modes d'hébergement différents selon les situations familiales. In : *Insee Première*. [en ligne]. 2 juillet 2026. Insee. N° 1455. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/1281324>.

- YAOUANCQ, Françoise et DUÉE, Michel, 2014. Les sans-domicile en 2012 : une grande diversité de situations. In : *Insee Références*. [en ligne]. 19 novembre 2014. Insee. [Consulté le 23 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/1288519?sommaire=1288529>.

De nouvelles nomenclatures d'activités

Les classements en théorie et en pratique



Clotilde Masson*

À la veille de l'entrée en vigueur de la nouvelle version de la nomenclature d'activités française (NAF), l'article rappelle d'abord les propriétés et les contraintes des nomenclatures, supports indispensables de l'élaboration de données statistiques. Les nomenclatures structurent un domaine d'étude en formant des sous-ensembles cohérents, groupant par similitude des objets qui sont par nature tous différents. Les critères sur lesquels s'appuient les similarités procèdent du regard porté sur l'univers étudié.

À la fin du XVIII^e siècle, le classement des activités économiques s'appuie d'abord sur l'origine des matières premières mobilisées (inputs) : produits minéraux, produits végétaux, produits animaux. Puis, au cours du XIX^e siècle, les industriels se regroupent plutôt par familles de produits, donc de débouchés (outputs). La prise en compte des processus de transformation dans la logique de classement intervient plus tard, sans toutefois supplanter les deux précédents critères.

À partir de 1993, la France adopte la NAF, dont la structure s'inscrit dans celles des nomenclatures d'activités européenne et onusienne afin de faciliter les comparaisons internationales. Dès lors, l'histoire de ces classifications devient commune, et leurs évolutions, nécessairement concertées et coordonnées, mobilisent les statisticiens et de nombreux autres utilisateurs sur toute la planète.

 With the new version of the French nomenclature of activities (NAF) about to be implemented, this paper first describes the characteristics and constraints of classifications, which are essential tools for the production of statistical data. A classification provides structure for a field of study by forming coherent subsets, grouping together objects that are all different by nature but share similar traits. The criteria on which these similarities are based are determined by the approach taken on the universe being studied.

At the end of the 18th century, the classification of economic activities depends on the source of the raw materials used (inputs): mineral products, plant products, animal products. But during the 19th century, industries are most often grouped by families of products, which means by market outlets (outputs). Later on, the classification logic takes the transformation process into account; it does not, however, replace the two preceding criteria.

From 1993, France implements the NAF, whose structure is in line with that of the European and UN's classifications to facilitate international comparisons. From that point onwards, these classifications share a common history and their developments, necessarily collaborative and coordinated, involve statisticians and many users worldwide.

* Cheffe de la division Nomenclatures économiques, DSE, Insee.
clotilde.masson@insee.fr



La NAF 2025 devient la nouvelle référence pour le classement des activités économiques et pour la détermination du code APE des entreprises.



La **NAF 2025**, nouvelle version de la **nomenclature d'activités française (NAF)**¹, entrera en vigueur en janvier 2027². Dès lors, c'est selon cette nouvelle référence que sera déterminé le **code d'activité principale exercée (code APE)**³ de chaque entreprise, qui matérialise son classement sectoriel, variable clé de la statistique d'entreprises.

Ce code d'activité jouit d'une certaine notoriété car il a aussi de nombreux usages hors de la sphère statistique. En revanche, les coulisses des

classifications statistiques, les contraintes de leur élaboration et les limites de leurs usages sont plus méconnues.

Les **nomenclatures statistiques** structurent un domaine d'étude en formant des sous-ensembles cohérents, groupant par similitude des objets qui sont par nature tous différents.

« ... or les choses ne se présentent pas d'elles-mêmes ainsi groupées à l'observation. Nous pouvons bien apercevoir plus ou moins vaguement leurs ressemblances. Mais le seul fait de ces similitudes ne suffit pas à expliquer comment nous sommes amenés à assembler les êtres qui se ressemblent ainsi, à les réunir en une sorte de milieu idéal, enfermé dans des limites déterminées et que nous appelons un genre, une espèce, etc. » (Durkheim et Mauss, 1903).

Les choix de regroupement opérés, les critères sur lesquels s'apprécient les similarités, intègrent nécessairement des éléments de connaissance ou de conjecture sur le domaine étudié. C'est même une des conditions de la pertinence d'une nomenclature. Pourrait-on établir une classification des maladies sans avoir aucune connaissance médicale ?

En conséquence, lorsque le domaine étudié évolue, il faut réinterroger les nomenclatures qui s'y rapportent. C'est ainsi que les transformations du paysage économique intervenues au cours des quinze dernières années amènent à réviser les nomenclatures d'activités.

► **Nomen calare : « appeler par son nom »**

Le mot « nomenclature » vient du latin *nomen calare*, « appeler par son nom ». **Une nomenclature subdivise l'univers étudié en sous-ensembles identifiables (caractérisables, reconnaissables) et « nommés »**. La population française, par exemple, peut se décomposer en femmes et hommes, mais aussi en enfants, adolescents et adultes, ou encore en habitants de zones rurales, périurbaines et urbaines. Ces derniers peuvent être des agriculteurs, des artisans ou commerçants, des ouvriers, des cadres, etc⁴.

¹ Voir le lien : <https://www.insee.fr/fr/information/2406147>.

² Décret n° 2025-736 du 31 juillet 2025 portant approbation des nomenclatures d'activités et de produits françaises. Voir les références juridiques en fin d'article.

³ Voir le lien : <https://www.insee.fr/fr/metadonnees/definition/c1888>.

⁴ Ce qui relève pour ce dernier point de la nomenclature des professions et catégories professionnelles (PCS), exemple important de nomenclature statistique (Amossé, 2019 ; Amossé, 2020).

Les nomenclatures définissent ainsi des catégories qui permettent de décrire une population, après avoir classé les individus qui la composent, i.e. après avoir attribué à chacun une catégorie.

Les nomenclatures statistiques présentent généralement une structure hiérarchique de partitions emboîtées qui décomposent le domaine d'étude en sous-parties de plus en plus restreintes (Guibert et al., 1971). Cette structure peut être obtenue et interprétée de deux manières :

- « en descendant », c'est-à-dire en décomposant le domaine d'abord en grands ensembles, qui sont ensuite décomposés en sous-ensembles de plus en plus petits ;
- « en montant », c'est-à-dire en réalisant d'abord des petits regroupements d'items considérés comme très semblables, puis en agrégeant progressivement ces regroupements, toujours par similarité, pour constituer des ensembles de plus en plus grands.

Mais selon quels critères évaluer la similarité ? C'est là le cœur de la classification !⁵ Ces critères ne sont pas universels, ne vont pas de soi (*figure 1*). Ils doivent être définis en lien étroit avec les objectifs de l'étude.

« Former une classe de certains objets, ce n'est autre chose que donner un même nom à tous ceux que nous jugeons semblables ; et, quand de cette classe nous en formons deux ou davantage, nous ne faisons encore autre chose que choisir de nouveaux noms pour distinguer des objets que nous jugeons différents. C'est uniquement par cet artifice que nous mettons de l'ordre dans nos idées ; mais cet artifice ne fait que cela ; et il faut bien remarquer qu'il ne peut rien faire de plus. En effet, nous nous tromperions grossièrement, si nous nous imaginions qu'il y a dans la nature des espèces et des genres, parce qu'il y a des espèces et des genres dans notre manière de concevoir » (de Condillac, 1780).

Les unités observées sont généralement complexes, et les observateurs, statisticiens ou non, n'en appréhendent pas tous les mêmes facettes. Ainsi, deux unités pourront être considérées comme très semblables voire identiques par un acteur, quand un autre, qui observe d'autres caractéristiques, notera au contraire de grandes différences. **La nomenclature doit donc expliciter les critères selon lesquels sont regroupées les unités pour former les catégories qui la constituent.**

« ... rien de plus tâtonnant, rien de plus empirique (au moins en apparence) que l'instauration d'un ordre parmi les choses ; rien qui n'exige un œil plus ouvert, un langage plus fidèle et mieux modulé ; rien qui ne demande avec plus d'insistance qu'on se laisse porter par la prolifération des qualités et des formes. Et pourtant un regard qui ne serait pas armé pourrait bien rapprocher quelques figures semblables et en distinguer d'autres à raison de telle ou telle différence : en fait, il n'y a, même pour l'expérience la plus naïve, aucune similitude, aucune distinction qui ne résulte d'une opération précise et de l'application d'un critère préalable » (Foucault, 1966).

⁵ Sur l'importance de la classification en tant qu'activité humaine, voir Bowker et Star (2000).

► Figure 1 - Bécassine, et l'importance du point de vue de l'observateur

10

BÉCASSINE A DE L'ORDRE



Source : Album « L'enfance de Bécassine », écrit par Caumery et illustré par Joseph Porphyre Pinchon, publié en 1913 chez Henri Gautier.



Les catégories doivent être « parlantes ».



Par ailleurs, les différentes catégories de la nomenclature doivent correspondre à des situations identifiables du domaine étudié. Elles doivent à la fois présenter un intérêt statistique et être comprises et reconnues. Autrement dit, les catégories doivent être « parlantes » pour

l'ensemble des acteurs concernés par la nomenclature :

- les statisticiens, qui recueillent les données utiles aux classements et produisent des indicateurs et des études ;
- les entités enquêtées, en vue du repérage de leur catégorie ;
- les utilisateurs qui exploitent les études et les indicateurs, ou qui simplement mobilisent la nomenclature à des fins de gestion.

En effet, si ce n'est pas le cas, ces différents acteurs risquent de ne pas interpréter de la même manière les rubriques de la nomenclature et les classements associés. Car une nomenclature statistique définit des catégories par la spécification des propriétés communes aux unités regroupées, et non par l'énumération de ces unités (non nécessairement connues a priori). Ces propriétés communes peuvent être, par exemple : être une femme de 18 ans ou plus, un élève de CE2, un mammifère, une boisson non alcoolisée, etc.

En outre, il est préférable que les catégories soient simples à énoncer. Par exemple, si l'on définit des tranches d'âge, on ne retiendra pas des seuils tels que « 17 ans et trois mois », « 31 ans et demi », mais plutôt « 18 ans », « 30 ans ». Toutefois, il est clair que le choix de catégories pertinentes dépend des usages auxquels est destinée la nomenclature.

Enfin, le nombre de catégories de la nomenclature dépend du nombre d'unités à classer (chaque groupe détaillé devant contenir un nombre significatif d'unités), mais aussi de la stabilité au cours du temps des groupes d'unités que constituent les catégories. La granularité de la nomenclature résulte également d'un arbitrage entre le niveau de détail souhaité et le niveau de détail praticable (auquel le recueil d'information est possible ou soutenable).

En effet, on peut penser qu'une nomenclature très détaillée, comportant un grand nombre de rubriques élémentaires, permettra une description fine de la population étudiée. Mais cela n'est vrai que si toutes les unités de la population peuvent être classées précisément. Or, collecter l'information requise pour un classement très détaillé peut s'avérer très coûteux, voire impraticable pour d'autres motifs (par exemple de confidentialité, d'accès à l'information). Et si les classements ne sont pas précis, l'intérêt du détail de la nomenclature est perdu !

Classer selon les catégories de la nomenclature

Une fois établies les différentes rubriques de la nomenclature, il reste à fournir le mode d'emploi de cet outil de classement. C'est-à-dire qu'il faut préciser comment déterminer, pour chaque unité de la population étudiée, de quelle catégorie de la nomenclature elle relève. Car, aussi bien choisis et parlants que puissent être les noms des catégories,

ils ne suffisent pas dans tous les cas à assurer une compréhension partagée de leur contenu exact.

Ainsi, pour que la nomenclature offre le langage commun dans lequel les différents utilisateurs auront la même compréhension des catégories, il est utile de fournir en sus de la structure les éléments suivants⁶ :

- **des notes explicatives**, qui décrivent le contenu de chaque catégorie, et en précisent les frontières ;
- **un système de codification**, idéalement reflétant la structure de la nomenclature ;
- **un corps de règles pour classer**, qui explicitent les critères de répartition, et précisent aussi certaines conventions de classement par défaut ;
- le cas échéant, des **tables de correspondances** avec des versions antérieures de la nomenclature, ou avec d'autres nomenclatures.

Toutefois, même avec une nomenclature très précisément explicitée, **le classement de certaines unités complexes ou atypiques peut s'avérer épineux**. Cela peut être ainsi le cas pour une unité qui combine des caractéristiques renvoyant à des catégories différentes de la nomenclature. Les règles de classement peuvent alors définir des règles de priorité entre les caractéristiques, ou des seuils. Par exemple, la nomenclature douanière définit la catégorie suivante de biens : « Tissus de coton, contenant moins de 85 % en poids de coton, mélangés principalement ou uniquement avec des fibres synthétiques ou artificielles, d'un poids n'excédant pas 200 g/m² ». Cependant, comme on l'a évoqué, plus les catégories sont définies avec précision, plus le recueil d'informations permettant de classer les unités dans la bonne catégorie risque d'être difficile ou coûteux.

Comment ces éléments s'illustrent-ils dans la nomenclature d'activités française ?

► La NAF, nomenclature d'activités française



Le classement des activités économiques dans la NAF s'appuie sur leurs trois composantes que sont les inputs, le processus et les outputs.



Le premier objectif de la NAF est de classer les activités économiques françaises en un nombre « raisonnable » d'ensembles cohérents, afin de fournir une grille d'observation et de mesure de la production nationale de biens et services. Le second objectif est de permettre la comparaison avec d'autres pays, ce qui implique, comme on le verra, d'harmoniser sa structure avec celles de nomenclatures internationales.

Le classement des activités économiques dans la NAF s'appuie sur leurs trois composantes que sont les inputs, le processus et les outputs. En effet, exercer une activité économique, c'est :

⁶ Voir à ce sujet Bruneau (1992).

- mobiliser des intrants (inputs) : des matières premières, des ressources humaines, des machines, des brevets, etc. ;
- dans un processus : assemblage, extraction, forgeage, conception, transport, commerce, etc. ;
- dont résultent des biens ou des services (outputs) : des articles fabriqués, des prestations de soins, d'enseignement, de conseil, de réparation, etc.

La NAF regroupe les activités qui sont similaires du point de vue de ces trois composantes. Les critères d'ordre juridique ou organisationnel sont écartés en tant que tels (par exemple, le fait que l'activité soit réalisée par une personne physique ou morale, une entreprise publique ou privée, avec un seul ou plusieurs établissements). Seules sont prises en compte les caractéristiques intrinsèques de l'activité.

Un peu d'histoire

Structurer la nomenclature d'activités à partir des inputs, des processus et des outputs laisse toutefois la place à de nombreuses possibilités de regroupements. De fait, les choix retenus ont évolué au cours du temps.

À la fin du XVIII^e siècle, une représentation « naturaliste » de l'économie prédomine, selon laquelle l'activité humaine « fait fructifier » les produits de la nature. Dans cet esprit, les industries françaises sont classées en trois grandes rubriques relatives à l'origine des matières premières : les produits minéraux, les produits végétaux et les produits animaux. C'est alors la logique des inputs qui prévaut.

À mesure que l'industrie se développe, la variété et la diversité des activités et produits de ce secteur rendent l'approche par les inputs de plus en plus inadaptée. Au cours du XIX^e siècle, les industriels se regroupent plus naturellement par famille de produits, donc d'outputs. Dans ce contexte, la nomenclature du recensement industriel de 1861 marque un tournant en classant les industries principalement en fonction de leurs produits.

Trente ans plus tard, lors du dépouillement du recensement professionnel de 1896, Lucien March⁷ introduit la notion « d'activité de l'entreprise », au sens du processus de transformation qui s'y opère. Soixante ans après, la nomenclature des activités économiques (NAE), créée par le décret du 9 avril 1959⁸, reste orientée principalement par le critère de finalité, donc d'outputs. Cependant, la composante « processus » fait son entrée. Ainsi, les auteurs de la NAE précisent dès la préface que lorsque la technique paraît primer nettement sur la finalité, c'est ce critère qui est retenu. Par exemple, la fabrication de chaussures en caoutchouc est classée avec les industries du caoutchouc et non avec la fabrication de chaussures. De même, les carrières de matériaux de construction sont rapprochées des carrières de matériaux divers. La NAE est remplacée en 1973 par la nomenclature d'activités et de produits (NAP), élaborée dans le même esprit (Rousseau, 1975)⁹.

⁷ Lucien March (1859-1933) était un démographe et ingénieur statisticien d'État français. Il a dirigé la Statistique générale de la France, un des ancêtres de l'Insee, de 1896 à 1920.

⁸ Décret n° 59-534 du 9 avril 1959 relatif à la nomenclature des activités économiques. Voir les références juridiques en fin d'article.

⁹ La NAP est créée par le décret n° 73-1036 du 9 novembre 1973 portant approbation des nomenclatures d'activités et de produits. Voir les références juridiques en fin d'article.

En un siècle et demi, chacune des trois composantes inputs / processus / outputs aura donc servi de critère d'agrégation d'activités économiques. Mais au-delà, se dessine le « critère d'association », qui vise à classer ensemble des activités fréquemment associées au sein des entreprises. C'est ce qui fait de la nomenclature un reflet des structures industrielles existantes, et l'ancre ainsi dans son époque.

Une harmonisation européenne

En Europe, c'est en 1970 que les États membres définissent une nomenclature commune pour classer les activités économiques : la nomenclature statistique des activités économiques dans la Communauté européenne (NACE).

La première version de la NACE précède seulement de quelques années la NAP, mais il n'y a pas de correspondance simple entre elles. L'enjeu d'articulation entre la nomenclature européenne et les nomenclatures des États membres va par la suite gagner en importance. En effet, la perspective d'un marché unique à l'horizon 1993 nécessite une bonne comparabilité des statistiques nationales et communautaires sur le secteur productif. À cette fin, un règlement européen¹⁰ établit en 1990 une NACE révisée, la NACE Rev. 1, dans laquelle doivent obligatoirement s'emboîter les nomenclatures nationales des États membres.



***La NAF s'inscrit
dans un réseau international
de nomenclatures d'activités.***



C'est dans ce contexte qu'est créée la NAF, en 1993¹¹. La nomenclature d'activités française s'inscrit dès lors dans un réseau international de nomenclatures d'activités et de produits. La structure de la NAF est rigoureusement imbriquée dans celle de la NACE, elle-même étroitement liée à celle de

la nomenclature d'activités des Nations Unies, la Classification internationale type, par industrie, de toutes les branches d'activité économique (CITI). La définition des rubriques de la NAF est donc contrainte par ce cadre, de même que les règles de classement statistique qui s'y rapportent (Boeda, 1993).

Un réseau international de nomenclatures d'activités

Pour rester harmonisées entre elles et continuer de permettre les comparaisons internationales de statistiques, les nomenclatures des Nations Unies, de l'Union européenne et de ses États membres doivent évoluer en même temps. Ainsi, depuis les années quatre-vingt-dix, leurs révisions sont conduites aux mêmes périodes et de manière coordonnée : en 1990-1993, en 2002-2003, en 2007-2008¹² et en 2024-2025.

¹⁰ Règlement (CEE) n° 3037/90 du Conseil, du 9 octobre 1990, relatif à la nomenclature statistique des activités économiques dans la Communauté européenne. Voir les références juridiques en fin d'article.

¹¹ La NAF est créée par le décret n° 92-1129 du 2 octobre 1992 portant approbation des nomenclatures d'activités et de produits, pour une entrée en vigueur au 1^{er} janvier 1993. Voir les références juridiques en fin d'article.

¹² Sur cette phase de révisions, voir Lacroix (2009) et Roussel (2008).

Notons qu'en 1997, la création de la nomenclature nord-américaine NAICS (*North American industry classification system*) a failli bousculer la définition d'une activité économique, entendue alors comme « combinaison de ressources pour produire des biens ou des services spécifiques » (Lacroix et Fuger, 2008). En effet, la NAICS ne se base que sur le processus de production, n'étant pas contrainte par une articulation activités-produits. Parallèlement, le système nord-américain de classement des produits (NAPCS pour *North American product classification system*) est totalement orienté vers l'utilisation des produits et les marchés.

Finalement, la définition principale de l'activité économique est conservée. Mais la note introductive de la CITI, reprise dans la NACE et la NAF, indique que le poids attribué à chacun des trois critères inputs, processus et outputs varie d'une division à l'autre, sans qu'il soit en général possible de déterminer le critère prépondérant.

► Une structure, des codes, des libellés... et des notes explicatives

La structure de la NAF

La structure de la NAF 2025 compte cinq niveaux emboîtés :

- **22 sections ;**
- **87 divisions ;**
- **287 groupes ;**
- **651 classes ;**
- **747 sous-classes.**

Les quatre premiers niveaux correspondent exactement à la structure de la NACE Rev. 2.1¹³. Dans la nomenclature française :

- 75 classes sont décomposées en plusieurs sous-classes. Ces décompositions créent au total 171 sous-classes spécifiques à la France.
- 576 classes ne sont pas décomposées. Dans chacune, il n'existe qu'une seule sous-classe, identique à la classe.

La **figure 2**, qui présente un extrait de la nomenclature, illustre l'emboîtement des cinq niveaux. De la section à la sous-classe, chaque rubrique de la nomenclature est identifiée par un **code** et un **libellé** (exemple : « 42.1 » – « Construction de routes et de voies ferrées »). En outre, des **notes explicatives** précisent son contenu (**figure 3**). En effet, le libellé, volontairement court, ne suffit pas toujours à spécifier quelles activités sont incluses dans la rubrique ou au contraire en sont exclues. Les notes fournissent généralement des exemples et contre-exemples pour aider à bien cerner les frontières de la rubrique.

¹³ Règlement délégué (UE) n° 2023/137 de la Commission du 10 octobre 2022 modifiant le règlement (CE) n° 1893/2006 du Parlement européen et du Conseil établissant la nomenclature statistique des activités économiques NACE Rev. 2. Voir les références juridiques en fin d'article.

► **Figure 2 - NAF 2025 : Extrait des rubriques de la section F**

NACE Rev. 2.1	NAF 2025 Sous-classes	Intitulés
F	SECTION F — CONSTRUCTION	
41	Construction de bâtiments résidentiels et non résidentiels	
41.0	Construction de bâtiments résidentiels et non résidentiels	
41.00	Construction de bâtiments résidentiels et non résidentiels	
	41.00G	Construction de maisons individuelles
	41.00H	Construction d'autres bâtiments
42	Génie civil	
42.1	Construction de routes et de voies ferrées	
42.11	42.11Y	Construction de routes et autoroutes
42.12	42.12Y	Construction de voies ferrées de surface et souterraines
42.13	Construction de ponts et tunnels	
	42.13G	Construction de ponts et de viaducs
	42.13H	Construction de tunnels
42.2	Construction de réseaux et de lignes	
42.21	42.21Y	Construction de réseaux pour fluides
42.22	42.22Y	Construction de réseaux électriques et de télécommunications
42.9	Construction d'autres ouvrages de génie civil	
42.91	42.91Y	Construction d'ouvrages maritimes et fluviaux
42.99	42.99Y	Construction d'autres ouvrages de génie civil n.c.a.
43	Travaux de construction spécialisés	

n.c.a. : Non classé ailleurs.
NACE : Nomenclature statistique des activités économiques dans la Communauté européenne.
NAF : Nomenclature d'activités française.
Source : Page de la NAF 2025 sur le site de l'Insee.
Voir le lien : <https://www.insee.fr/fr/information/8617910>.

Les codes de la NAF

Les codes de la NAF 2025 matérialisent à la fois la structure hiérarchique de la nomenclature française et son emboîtement dans les nomenclatures internationales (*figure 4*).

Les sections et divisions sont identiques (même code, même libellé, même contenu) à celles de la CITI Rev. 5 et de la NACE Rev. 2.1 :

- la **section** a un code alphabétique : une lettre (de A à V) ;
- la **division** a un code à deux chiffres (parmi 01 à 99¹⁴).

¹⁴ Par exemple, la section A comporte trois divisions 01 à 03, puis la section B comporte cinq divisions 05 à 09. Il n'y a pas de division 04, ni 34, ni 44, ni 45, par exemple : cela laisse quelques « trous », permettant de prévoir des aménagements de la structure sans tout bouleverser.

► Figure 3 - NAF 2025 : Extrait de la note explicative des rubriques de la section F

41.00G

Construction de maisons individuelles

Cette sous-classe comprend :

- les activités générales ou « tous corps d'état » de construction de maisons individuelles, portant la responsabilité globale de la construction ;
- la construction de maisons individuelles préfabriquées nécessitant un assemblage sur place.

Cette sous-classe comprend également :

- l'aménagement ou la rénovation complète de maisons individuelles existantes, impliquant plusieurs activités de constructions spécialisées.

Cette sous-classe ne comprend pas :

- la promotion mobilière de logements, voir 68.12Y ;
- les activités d'architecture et d'ingénierie, voir groupe 71.1 ;
- la gestion de projets de construction pour compte de tiers, voir groupe 71.1 ;
- les activités de maîtrise d'œuvre, voir groupe 71.1.

NAF : Nomenclature d'activités française.

Source : Page de la NAF 2025 sur le site de l'Insee.

Voir le lien : <https://www.insee.fr/fr/information/8617910>.

Les groupes et classes sont identiques à ceux de la NACE, mais ils peuvent différer de ceux de la CITI. Pour marquer cela, dans les codes de la NACE, comme dans ceux de la NAF, un point a été introduit après les deux chiffres des divisions :

- Le **groupe** a un code à 3 chiffres avec un point après les deux premiers (ex : 42.1).
- La **classe** a un code à 4 chiffres avec un point après les deux premiers (ex : 42.12).

Enfin, le code de chaque sous-classe de la NAF est composé du code de sa classe d'appartenance et d'une lettre :

- Y lorsque c'est l'unique sous-classe (par exemple 01.11Y pour l'unique sous-classe de la classe 01.11) ;
- G, H, J, K ou L, pour distinguer les différentes sous-classes d'une classe, lorsqu'il y en a plusieurs, aucune classe n'ayant été décomposée en plus de cinq sous-classes (par exemple 41.00G et 41.00H pour les deux sous-classes de la classe 41.00).

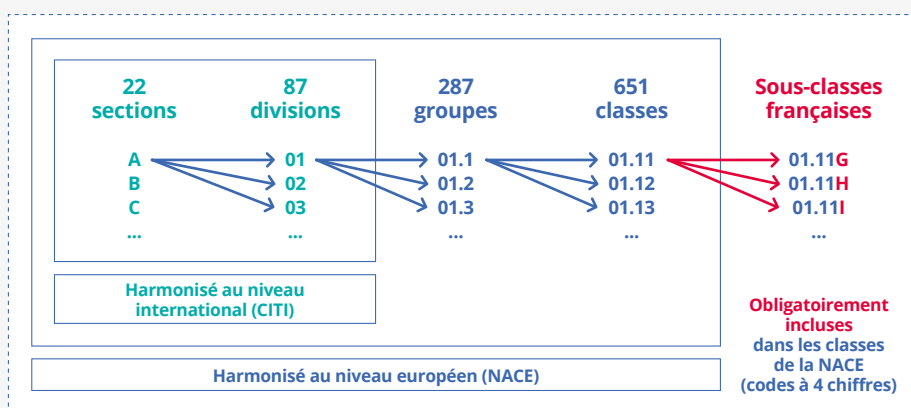
► Les usages de la NAF

Permettre des regroupements ad hoc pour différents usages

Ayant été définies précisément, les différentes catégories de la nomenclature peuvent être combinées différemment pour mieux répondre à certains besoins.

Ainsi, les comptables nationaux utilisent des regroupements de sections et divisions pour communiquer les données de comptabilité nationale d'un grand nombre de pays. Au niveau le plus agrégé, ils définissent dix grands ensembles d'activités issus de regroupements de sections où, par exemple, toutes les industries (sections B à E) sont regroupées en un seul ensemble.

► **Figure 4 - Les codes de la NAF matérialisent l'emboîtement des nomenclatures à différentes échelles**



NAF : Nomenclature d'activités française.

CITI : Classification internationale type, par industrie, de toutes les branches d'activité économique.

NACE : Nomenclature statistique des activités économiques dans la Communauté européenne.

Certaines analyses économiques s'appuient sur des regroupements spécifiques¹⁵ de catégories de la nomenclature. Par exemple, la filière bois réunit des activités relevant du domaine agricole, de l'industrie, du commerce, des transports et du conseil.

Par ailleurs, les regroupements d'activités dans une même catégorie de la nomenclature n'ont pas toujours un caractère d'évidence.

Par exemple :

- Où classer l'activité des pharmacies ? Avec les activités pour la santé humaine, où sont classées celles des hôpitaux, des laboratoires d'analyses médicales, des médecins et d'autres professionnels de santé ? Ou avec les activités de commerce, puisque les pharmacies achètent des biens pour les revendre sans transformation ?¹⁶
- De même, le transport par ambulance doit-il être classé avec les activités de transport ou avec les activités liées à la santé ?¹⁷

Les nomenclatures résultent souvent de choix conventionnels et de compromis : « ...une nomenclature se négocie ; elle ne s'impose pas » (Boeda, 2008).



La NAF, nomenclature de statisticiens, a de nombreux usages non statistiques.



La NAF procède donc à des classements, dont la logique est explicitée dans les notes explicatives, tout en laissant la possibilité d'alternatives. Elle est définie par les statisticiens en vue d'opérer des regroupements économiquement pertinents des

¹⁵ On retrouve cette idée de « regroupements spécifiques » dans la PCS (Amossé, 2020).

¹⁶ C'est ce deuxième choix qui a été retenu dans la NACE, et donc dans la NAF.

¹⁷ C'est aussi ce deuxième choix qui a été retenu dans la NACE, et donc dans la NAF.

entreprises et des établissements, selon la nature de leur activité (Volle, 1980). Cet objectif étant commun à de nombreux acteurs économiques et administratifs, la NAF constitue une référence largement partagée, y compris au-delà des utilisations statistiques. En particulier, elle permet de déterminer le code APE des entreprises, des associations et des organismes publics, ainsi que de leurs établissements. Ce code est utilisé par de multiples acteurs à des fins d'études ou de gestion, comme le recouvrement des impôts, l'octroi de licences commerciales, etc.

Code NAF, code APE

Le code APE d'une unité (entreprise, établissement, etc.) caractérise son « activité principale exercée », définie comme l'activité qui contribue le plus à la valeur ajoutée produite. Ce code joue un rôle important pour l'entreprise (Roussel, 2008).

Par construction, chaque unité a un seul code APE, même si elle exerce plusieurs activités, et même si ces dernières sont très diverses. Résumer l'ensemble des activités d'une entreprise dans un unique code est, par nature, fatalement réducteur. Les utilisateurs du code APE doivent toujours se le rappeler. Le code APE indique **à quel secteur d'activité une unité est rattachée pour la production d'indicateurs statistiques sectoriels**.

L'Insee attribue un code APE aux unités à partir de leurs déclarations, directes (formalités administratives) ou indirectes (via les réponses à certaines enquêtes). Cette opération ne crée par elle-même ni droits ni obligations pour les entreprises, selon l'article R123-231 du code de commerce, relatif au système national d'identification et au répertoire des entreprises et de leurs établissements (Sirene)¹⁸.

Ainsi, dans l'application d'un texte réglementaire ou d'un contrat, le code APE constitue une présomption, mais pas une preuve d'appartenance à un secteur d'activité « visé par ce texte ou ce contrat ». Car le code APE est attribué à des fins statistiques, selon un ensemble de règles harmonisées au niveau européen, lesquelles ne rencontrent pas nécessairement les critères particuliers d'application de ce texte ou de ce contrat.

► La révision mondiale des nomenclatures d'activités

Pourquoi réviser les nomenclatures économiques ?

Au cours du temps, les changements économiques, structurels et organisationnels, ainsi que le développement de nouvelles technologies, entraînent l'apparition d'activités et de

produits nouveaux. Certains d'entre eux supplantent peu à peu une partie des activités et produits existants (*figure 5*). Par exemple, les quinze dernières années sont marquées par l'essor du commerce en ligne, de la distribution de contenus audio et vidéo, de l'économie du partage, ou encore des activités d'intermédiation.

Les nomenclatures changent parce que le monde change.

¹⁸ Voir les références juridiques en fin d'article. Voir aussi Lestang et Rousseau (1982).

► **Figure 5 - Extrait de la nomenclature des entreprises, établissements et toutes activités collectives (1949)**

78-2	78-20	78-200	DÉBIT DE BOISSONS ANNEXÉ À UNE AUTRE ACTIVITÉ.
	78-21	78-210	Buvette dans un établissement de spectacle, de sports : théâtre cinéma, piscine, stade, plage...
	78-22	78-220	Café-bal.
	78-23	78-230	Café-chantant, café-concert, café-spectacle.
	78-24	78-240	Café-épicerie.
	78-25	78-250	Café associé à une autre activité n. d. a.
78-3	78-30	78-300	DÉBIT DE TABAC AVEC DÉBIT DE BOISSONS : café-tabac, débit de tabac avec débit de vin.

Ainsi, la pertinence des nomenclatures statistiques diminue à mesure que le temps passe. Elles doivent être périodiquement révisées, pour permettre de produire une information économique et sociale en phase avec un monde en constante évolution. En particulier, l'introduction de nouvelles rubriques ou de nouvelles notes explicatives permet d'harmoniser les classements d'activités émergentes¹⁹.

Ces changements constituent un défi permanent pour les nomenclatures statistiques (Grouthier et Lachize, 2008). D'un côté, les révisions doivent permettre d'appréhender les nouvelles réalités dans les statistiques produites. De l'autre, elles ne doivent pas être trop fréquentes, car la comparabilité des données dans le temps en souffrirait. Il faut trouver le juste équilibre entre la pertinence exigée des nomenclatures économiques et leur continuité.

Un chantier titanesque

Compte tenu de la coordination mondiale des nomenclatures d'activités, leur révision constitue un chantier colossal, qui implique un nombre considérable d'acteurs, de toute la planète.

La révision 2025 des nomenclatures d'activités a mobilisé pendant six ans non seulement des instituts de statistique de plus de 100 pays, mais aussi des représentants d'organismes internationaux²⁰, de banques centrales, de fédérations professionnelles, d'administrations, ainsi que des chercheurs.

Les révisions de la CITI et de la NACE ont été conduites de 2019 à 2024 (*figure 6*), selon des processus coordonnés comportant les mêmes principales étapes :

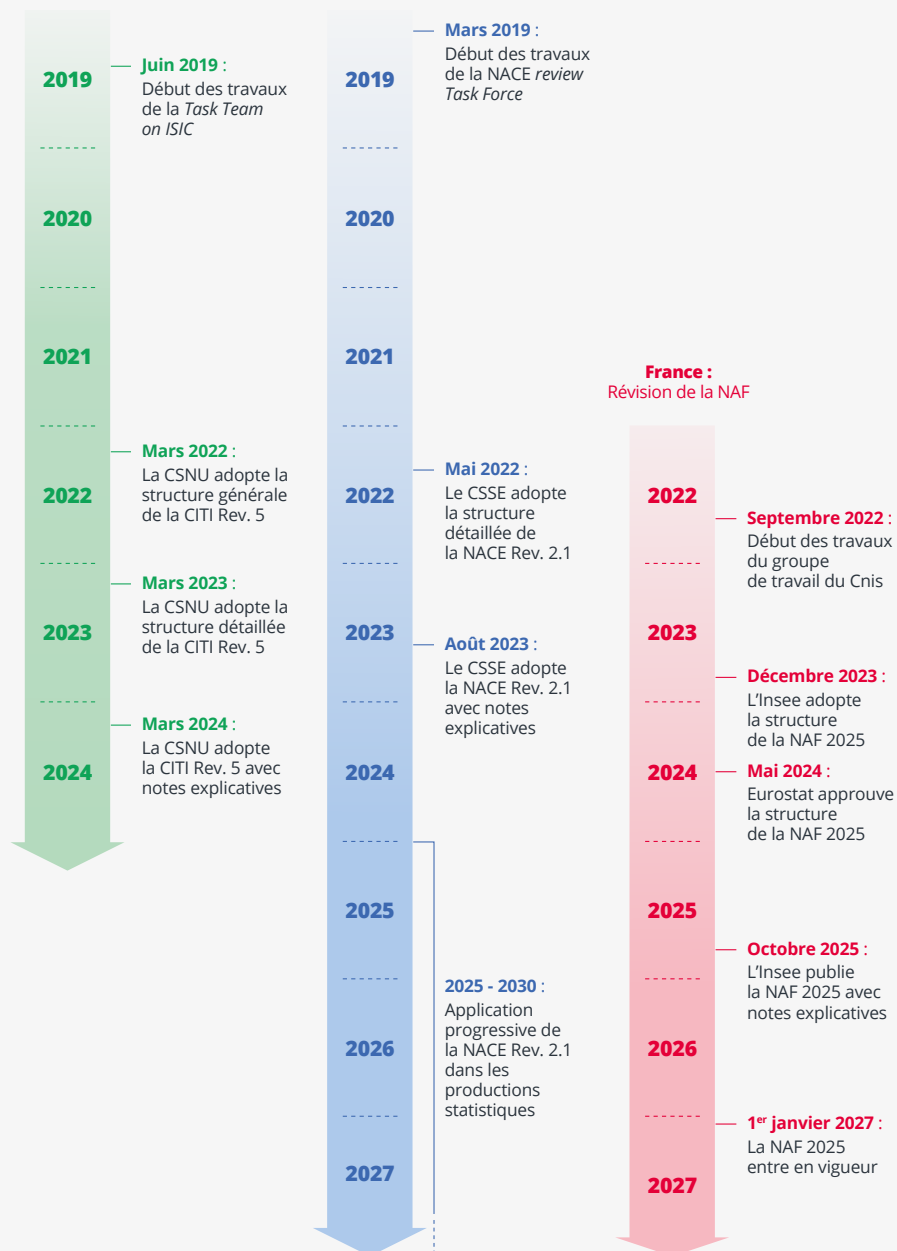
¹⁹ Entre deux révisions, de façon continue, le classement d'activités émergentes fait l'objet de discussions entre experts, et donne lieu à des arbitrages qui font jurisprudence.

²⁰ Fonds monétaire international (FMI), Banque des règlements internationaux (BRI), Organisation mondiale du tourisme (OMT), Organisation internationale du travail (OIT), Organisation de coopération et de développement économiques (OCDE).

► Figure 6 - Chronologie des révisions des nomenclatures

Nations Unies :
Révision de la CITI

Europe :
Révision de la NACE



CITI : Classification internationale type, par industrie, de toutes les branches d'activité économique.
 CSNU : Commission statistique des Nations Unies.
 NACE : Nomenclature statistique des activités économiques dans la Communauté européenne.
 CSSE : Comité du système statistique européen.
 NAF : Nomenclature d'activités française.
 Cnis : Conseil national de l'information statistique.

- consultation des utilisateurs, analyse des éléments recueillis, collecte d'informations complémentaires ;
- élaboration d'un projet de structure de la nomenclature révisée, consultation générale sur ce projet et amendements, adoption de la structure par les instances ad hoc ;
- rédaction des notes explicatives, mise à jour de textes réglementaires, préparation de la mise en œuvre de la nouvelle nomenclature.

La révision de la NACE

Début 2018, Eurostat – l'Office statistique de l'Union européenne – a consulté les instituts nationaux de statistique (INS) des États membres pour évaluer l'opportunité d'engager une mise à jour, voire une révision de la NACE :

« La version actuelle de la nomenclature statistique des activités économiques (NACE Rev. 2) est en vigueur depuis le 1^{er} janvier 2008, en vertu du règlement (CE) n° 1893/2006 du Parlement européen et du Conseil²¹. Plusieurs États membres ont relevé l'inadéquation de cette nomenclature pour la classification de certaines activités économiques, car l'économie a effectivement subi des changements au cours de ces dix dernières années. Nous aimerions recueillir de l'information sur les domaines de la nomenclature que vous considérez désuets et les raisons ; cela nous permettrait d'évaluer si une révision ou une mise à jour serait nécessaire et quelle en serait la portée ».

En retour, près de 1 500 propositions de modifications ont été recueillies, dont la moitié environ touchaient à la structure de la NACE²². Afin de les examiner, Eurostat a créé un groupe de travail européen (NACE *review Task Force*), rassemblant des représentants des INS de dix-huit pays européens (dont l'Insee) et de directions de la Commission européenne.

Les propositions ont été analysées systématiquement, en procédant par section de la NACE. Entre mars 2019 et mai 2022, à intervalle d'un ou deux mois, le groupe de travail européen s'est réuni pour discuter les arguments avancés par les contributeurs, et procéder à des arbitrages.

Sur certains sujets, un consensus a été obtenu assez facilement. C'est le cas par exemple pour la suppression de la classe « Location de vidéocassettes et disques vidéo ». Si cette activité n'a pas entièrement disparu, il n'est plus justifié d'en faire une classe à part entière. Elle a ainsi été intégrée dans la classe « Location et location-bail d'autres biens personnels et domestiques ».

D'autres sujets ont donné lieu à de vifs débats. Par exemple, comment classer la fabrication de vélos électriques ? Il existe en effet un continuum entre le vélo à assistance électrique, encore assez proche d'un vélo musculaire sur lequel un moteur peut être installé a posteriori, et le vélo sur lequel on ne pédale plus, ou à peine, et qui pourrait être considéré comme assez proche d'un motorcycle. La fabrication de cycles équipés d'un moteur électrique auxiliaire reste finalement classée avec la fabrication de bicyclettes.

²¹ Voir les références juridiques en fin d'article.

²² L'Allemagne, la France et l'Italie sont les trois INS européens qui ont formulé le plus de propositions.

Dès le 1^{er} trimestre 2019, la révision de la NACE a induit des travaux préparatoires au sein de chacun des pays participant au groupe de travail européen. En France, sur chaque sujet, les acteurs concernés ont été consultés, non seulement au sein du service statistique public, mais aussi au-delà, pour recueillir en particulier l'avis des organisations professionnelles. Ce sont ces nombreux échanges avec des acteurs, dans tous les secteurs d'activité, qui ont permis de développer les arguments français à porter au niveau européen²³.

La révision de la CITI

Les révisions de la NACE et de la CITI sont menées parallèlement, de manière coordonnée.

Si les statisticiens européens souhaitaient faire évoluer la NACE, ils tenaient toutefois à préserver, voire à raffermir l'emboîtement de la NACE dans la CITI (ou ISIC en anglais²⁴). Les travaux européens devaient donc être conduits en relation étroite avec la division de statistiques des Nations Unies (UNSD²⁵). Celle-ci a mis en place un groupe de travail international (*Task team on ISIC*), mandaté pour préparer la révision de la CITI.

Le groupe de travail international a organisé une consultation mondiale à l'automne 2019, autour de 10 grands thèmes et de 35 sous-thèmes, liés aux évolutions récentes des activités économiques²⁶. Les 110 réponses obtenues proviennent du monde entier : 22 pays d'Afrique, 15 d'Amérique, 38 d'Europe, 23 d'Asie, 6 d'Océanie et 6 organismes internationaux. L'analyse de ces réponses a conduit, à l'issue de plus de trois ans de travaux et consultations, à un projet de nouvelle nomenclature. La CITI Rev.5 a été approuvée par la Commission statistique des Nations Unies (CSNU) en trois temps :

- Mars 2022 : structure générale, jusqu'au niveau des groupes ;
- Mars 2023 : structure détaillée ;
- Mars 2024 : notes explicatives, introduction et plan de mise en œuvre.

La *Task team on ISIC* comptait des représentants d'Eurostat et de pays européens, dont la France, de sorte que les travaux onusiens et européens s'irriguaient mutuellement, et que la cohérence des nouvelles CITI et NACE était une préoccupation permanente²⁷. La structure détaillée de la nouvelle NACE (NACE Rev. 2.1) a pu ainsi être adoptée en mai 2022 par le Comité du système statistique européen (CSSE). Cela a permis à l'Insee d'entamer dès la mi-2022 l'élaboration de la nouvelle NAF, qui devait strictement s'inscrire dans cette structure.

²³ Par exemple, la France a porté la proposition formulée par le Service des données et études statistiques (Sdes) – service statistique ministériel en charge de l'énergie, du transport, du logement et de l'environnement – de classer séparément la fabrication de véhicules électriques et celle de véhicules thermiques. Mais cela n'a pas été retenu, car de nombreux constructeurs produisent les deux types de véhicules, et il aurait alors aussi fallu distinguer la fabrication de véhicules hybrides. En revanche, il a été jugé utile de distinguer ces différents outputs dans la classification des produits.

²⁴ Pour *International Standard Industrial Classification of All Economic Activities*.

²⁵ Pour *United Nations Statistics Division*.

²⁶ La consultation portait notamment sur l'adaptation de la CITI au développement de certaines activités (e-commerce, activités d'intermédiation, nouveaux modes de distribution de contenus audio et vidéo), à l'impact des nouvelles technologies, à la distinction entre commerce de gros et de détail, aux différentes modalités d'externalisation, ou encore aux codes « surchargés » et au cycle de révision des nomenclatures d'activité.

²⁷ Par exemple, le commerce et la réparation d'automobiles étaient classés dans une même division, au sein des activités de commerce. Les avis étaient très partagés en Europe pour modifier ce classement « historique », mais la CITI Rev.5 a déplacé les activités de réparation, désormais classées avec d'autres services de réparation, et ce choix s'est imposé à l'Europe.

La révision de la NAF, contrainte par la nouvelle NACE

L'élaboration de la nouvelle NAF devait se faire en associant le plus possible ses utilisateurs au sein et hors de la sphère statistique, dont notamment les organisations professionnelles. L'enjeu était bien sûr de s'appuyer sur leur connaissance fine des activités. Mais il s'agissait aussi de leur permettre de s'approprier le plus en amont possible la nouvelle nomenclature, d'en être partenaires, et d'en comprendre les contraintes. Il fallait en effet éviter d'élaborer « hors sol » une nouvelle nomenclature qui risquerait de s'avérer ensuite inappropriée.

C'est tout naturellement au sein du Conseil national de l'information statistique (Cnis)²⁸ qu'a été conduite cette concertation, avec la constitution d'un groupe de travail associant des représentants d'organismes professionnels et d'administrations, des statisticiens et des économistes. Par le relais de ses membres, mais aussi à travers une communication initiée sur le site du Cnis, le groupe de travail s'est efforcé de susciter un grand nombre de propositions argumentées d'évolution de la NAF. Chacune d'elles a ensuite été instruite au sein d'un des 20 sous-groupes spécialisés par secteur d'activité, sur la base de critères généraux communs adoptés par le groupe de travail. Ces travaux ont débouché sur la proposition de structure de la NAF 2025, adoptée par l'Insee en décembre 2023, puis approuvée par la Commission européenne en mai 2024, comme le prévoit le règlement européen.

Car l'emboîtement dans la NACE des nomenclatures des États membres est strictement imposé par ce règlement, ce qui réduit fortement les marges de manœuvre pour les nomenclatures nationales. C'est pourquoi il était important de s'impliquer dans les travaux, les réflexions et les choix faits en amont, aux niveaux européen et international.

Devant strictement s'inscrire dans la structure de la NACE, la NAF pouvait soit reprendre exactement la structure européenne, soit définir une structure plus détaillée en décomposant certaines classes en sous-classes. Comme on l'a déjà vu, c'est cette deuxième option qui a prévalu, ne serait-ce que pour conserver certaines catégories déjà présentes dans la NAF, mais aussi pour répondre à des besoins d'information formulés par différents acteurs. Par exemple, les fédérations de taxi et le service statistique des transports souhaitaient un classement distinguant les activités de taxi des activités de voiture de transport avec chauffeur (VTC). Au total, l'évolution de la NAF est liée à des contraintes internationales et à des choix nationaux (**encadré**).

La mise en œuvre de la nomenclature révisée

Ces travaux considérables ne sont toutefois que le prélude du vaste ensemble d'opérations requises pour mener à bien le changement de nomenclature d'activités en France. En effet, les enjeux de la révision de la NAF ne sont pas tant dans l'élaboration de la NAF 2025, fortement contrainte par celle des nomenclatures CITI et NACE, que dans sa mise en œuvre à venir.

²⁸ La mission principale du Cnis est d'assurer la concertation entre les producteurs et les utilisateurs de la statistique publique. Anxionnaz et Maurel (2021) décrivent le rôle et le fonctionnement de cette instance.

► Encadré. Les plus grands changements de la NAF

Les changements hérités des révisions de la CITI et de la NACE :

C'est le secteur du **commerce de détail** qui est le plus fortement restructuré. D'une part, il n'inclut plus les activités de réparation des automobiles et motocycles, désormais classées dans les services. D'autre part, la nouvelle structure ne distingue plus les modalités de vente (en magasin, sur des marchés, en ligne, etc.). Cela a en effet moins de sens aujourd'hui au regard du fort développement du commerce en ligne. Ce sont les produits vendus qui deviennent le critère principal de classement.

Une autre modification importante concerne la mise en évidence des **activités d'intermédiation**. Il s'agit des activités qui facilitent les transactions entre acheteurs et vendeurs, pour la commande et/ou la livraison de biens ou de services, moyennant une rémunération ou une commission. Elles peuvent être exercées sur des plateformes numériques ou par des canaux non numériques. Ces activités se sont déployées dans de nombreux secteurs, par exemple dans les transports, l'hébergement, l'immobilier ou encore le soutien scolaire. Afin de les identifier, tout en préservant l'approche fondamentalement sectorielle de la CITI, il a été décidé de créer des rubriques dédiées au sein de certains groupes ou divisions de la nomenclature.

Les changements spécifiques à la NAF

La NAF 2025 compte au total 747 sous-classes, dont 171 sous-classes spécifiquement françaises, résultant de la décomposition de 75 classes. La plus grande partie des décompositions de classes (59 sur 75) correspond à la reconduction, exacte ou

réaménagée, de décompositions déjà présentes dans la NAF rév. 2. Les 16 nouvelles décompositions sont principalement liées :

- à des activités dont le classement a évolué dans la NACE : par exemple, introduire la distinction entre aéronefs et engins spatiaux au sein de la réparation du matériel militaire, car la NACE ne distingue plus le type d'appareil, mais le caractère civil ou militaire ;
- à l'émergence ou au développement de nouvelles activités économiques, par exemple l'exploitation d'installations de recharge pour véhicules électriques, la livraison à domicile de repas, l'édition de services de médias à la demande, la réparation et l'entretien de cycles et autres engins de la « mobilité douce » ;
- au souhait d'identifier plus précisément certaines activités, par exemple le commerce de gros d'appareils de génie climatique (radiateurs, climatiseurs, etc.), le commerce de détail de produits laitiers et d'œufs, le commerce de détail d'antiquités et de livres anciens, les activités de garde d'enfants, les services pour animaux de compagnie, l'apiculture.

À l'inverse, une vingtaine de décompositions de la NAF rév. 2 ne sont pas reconduites dans la NAF 2025, soit parce qu'elles avaient un poids économique insuffisant, soit, le plus souvent, parce qu'elles n'ont pas été demandées ou que l'instruction n'a pas démontré leur nécessité. Il s'agit notamment d'une dizaine d'activités de fabrication, mais aussi d'activités de commerce de gros, de restauration collective, de promotion immobilière, ou encore des activités des médecins spécialistes.

En particulier, la révision implique de mettre en place un nouveau processus de codification de l'APE (Randriamanana et Faria, 2025) et d'attribuer un nouveau code à l'ensemble des entreprises, associations, organismes publics, établissements.

Cette nouvelle codification induit par ailleurs des modifications des dispositifs d'enquêtes et de production d'indicateurs statistiques.

Enfin, au-delà de la sphère statistique, de nombreux organismes et administrations pour qui le code APE est une variable clé du système d'information doivent revoir leurs programmes et leurs processus de gestion. C'est le cas, par exemple, du Guichet unique des formalités d'entreprises, des organismes de sécurité sociale (Urssaf²⁹), ou de fédérations professionnelles.

Puis la NAF 2025 s'installera dans notre quotidien...

²⁹ Union de recouvrement des cotisations de sécurité sociale et d'allocations familiales (Urssaf).

► Fondements juridiques

- Règlement (CEE) n° 3037/90 du Conseil, du 9 octobre 1990, relatif à la nomenclature statistique des activités économiques dans la Communauté européenne. In : *site de l'Union européenne*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/fr/ALL/?uri=CELEX%3A31990R3037>.
- Règlement (CE) n° 1893/2006 du Parlement européen et du Conseil du 20 décembre 2006 établissant la nomenclature statistique des activités économiques NACE Rév. 2 et modifiant le règlement (CEE) n° 3037/90 du Conseil ainsi que certains règlements (CE) relatifs à des domaines statistiques spécifiques. In : *site de l'Union européenne*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=celex:32006R1893>.
- Règlement délégué (UE) n° 2023/137 de la Commission du 10 octobre 2022 modifiant le règlement (CE) n° 1893/2006 du Parlement européen et du Conseil établissant la nomenclature statistique des activités économiques NACE Rév. 2. In : *site de l'Union européenne*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : https://eur-lex.europa.eu/eli/reg_del/2023/137/oj?locale=fr.
- Code de commerce, article R123-231 relatif au système national d'identification et au répertoire des entreprises et de leurs établissements. In : *site de Légifrance*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000046073507.
- Décret n° 59-534 du 9 avril 1959 relatif à la nomenclature des activités économiques. In : *site de Légifrance*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000492756>.
- Décret n° 73-1036 du 9 novembre 1973 portant approbation des nomenclatures d'activités et de produits. In : *site de Légifrance*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000514722>.
- Décret n° 92-1129 du 2 octobre 1992 portant approbation des nomenclatures d'activités et de produits, pour une entrée en vigueur au 1^{er} janvier 1993. In : *site de Légifrance*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000527438>.
- Décret n° 2025-736 du 31 juillet 2025 portant approbation des nomenclatures d'activités et de produits françaises. In : *site de Légifrance*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000052019213>.

► Bibliographie

- AMOSSÉ, Thomas, 2019. La rénovation de la nomenclature socioprofessionnelle (2018-2019). In : *Rapport du groupe de travail du Cnis*. [en ligne]. Décembre 2019. Cnis. N° 156. [Consulté le 10 avril 2026]. Disponible à l'adresse : https://www.cnis.fr/publications/?_collection=rapports.
- AMOSSÉ, Thomas, 2020. La nomenclature socioprofessionnelle 2020 – Continuité et innovation, pour des usages renforcés. In : *Courrier des statistiques*. [en ligne]. 29 juin 2020. Insee. N° N4, pp. 62-80. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4497076?sommaire=4497095>.
- ANXIONNAZ, Isabelle et MAUREL, Françoise, 2021. Le Conseil national de l'information statistique – La qualité des statistiques passe aussi par la concertation. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2021. Insee. N° N6, pp. 123-142. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5398693?sommaire=5398695>.
- BOEDA, Michel, 1993. NACE et NAF : un regard neuf sur les activités des entreprises. In : *Courrier des statistiques*. [en ligne]. Décembre 1993. Insee. N° 67-68, pp. 11-16. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06z98z3/f1.pdf>.
- BOEDA, Michel, 2008. Les nomenclatures statistiques : pourquoi et comment. In : *Courrier des statistiques*. [en ligne]. Novembre-décembre 2008. Insee. N° 125, pp. 5-11. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05401087v1>.
- BOWKER, Geoffrey C. et STAR, Susan Leigh, 2000. *Sorting things out: Classification and Its Consequences*. Janvier 2000. The MIT Press. ISBN : 9780262269070.
- BRUNEAU, Émile, 1992. SYNAPSE, serveur de nomenclatures. In : *Courrier des Statistiques*. [en ligne]. Juin 1992. Insee. N° 61-62, pp. 67-71. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05429241v1>.
- DE CONDILLAC, Étienne Bonnot, 1780. La logique, ou les premiers développements de l'art de penser. In : *site de la bibliothèque numérique de la BnF*. [en ligne]. L'Esprit et de Bure l'aîné (Paris). [Consulté le 1^{er} mai 2026]. Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k80138k/f12.item>.
- DURKHEIM, Émile et MAUSS, Marcel, 1903. De quelques formes de classification – Contribution à l'étude des représentations collectives. In : *site de la bibliothèque numérique francophone « Les Classiques des sciences sociales »*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <http://dx.doi.org/doi:10.1522/cla.due.deq>.
- FOUCAULT, Michel, 1966. Les mots et les choses – Une archéologie des sciences humaines. In : *site Internet Archive*. [en ligne]. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://archive.org/details/les-mots-et-les-choses-foucault>.

- GROUTHIER, Jean-Philippe et LACHIZE, Jean-Paul, 2008. Le changement de nomenclatures d'activités et de produits dans l'industrie. In : *Courrier des Statistiques*. [en ligne]. Décembre 2008. Insee. N° 125, pp. 67-70. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05413608v1>.
- GUIBERT, Bernard, LAGANIER, Jean et VOLLE, Michel, 1971. Essai sur les nomenclatures industrielles. In : *Économie et statistique*. [en ligne]. Février 1971. Insee. N° 20, pp. 23-36. [Consulté le 10 avril 2026]. Disponible à l'adresse : https://www.persee.fr/doc/estat_0336-1454_1971_num_20_1_6122.
- LACROIX, Michel et FUGER, Marie-Madeleine, 2008. Révision des concepts et principaux changements de structure. In : *Courrier des statistiques*. [en ligne]. Novembre-décembre 2008. Insee. N° 125, pp. 45-54. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05413442v1>.
- LACROIX, Thierry, 2009. The network of international, EU, and French classifications of activities and products: a coordinated revision in 2008. In : *Courrier des statistiques, English series*. [en ligne]. Insee. N° 15, pp. 37-44. [Consulté le 10 avril 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05413389v1>.
- LESTANG, Philippe et ROUSSEAU, Roland, 1982. Les classements « activité principale exercée » de l'INSEE ont-ils des conséquences juridiques ? In : *Courrier des statistiques*. [en ligne]. Octobre 1982. Insee. N° 24, pp. 33-34. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06z97d8.r=courrier%20des%20statistiques%201982?rk=21459;2>.
- RANDRIAMANANA, Nathan et FARIA, Thomas, 2025. Changement de nomenclature statistique à l'ère du machine learning : réentraînement basé sur les LLM (Large Language Models) pour la révision de la NAF. *15^e Journées de méthodologie statistique de l'Insee*. [en ligne]. Novembre 2025, Paris, France. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05613478v1>.
- ROUSSEAU, Roland, 1975. Pourquoi change-t-on de nomenclature ? In : *Économie et Statistique*. [en ligne]. Septembre 1975. Insee. N° 70, pp. 63-67. [Consulté le 19 mai 2026]. Disponible à l'adresse : https://www.persee.fr/doc/estat_0336-1454_1975_num_70_1_1773.
- ROUSSEL, Patrice, 2008. Le code APE, carte vitale des entreprises ou les impacts non statistiques d'un changement de nomenclature. In : *Courrier des statistiques*. [en ligne]. Novembre-décembre 2008. Insee. N° 125, pp. 85-88. [Consulté le 1^{er} mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05413701v1>.
- VOLLE, Michel, 1980. *Le métier de statisticien – Chapitre VII L'évolution des découpages*. Janvier 1980. Hachette. Collection L'Esprit critique. ISBN : 2-01-004529-7.

Le système statistique européen

2004-2015 : une période de crises et de consolidations



Pierre Bischoff* et Hervé Piffeteau**

Ce deuxième volet de l'histoire du système statistique européen (SSE) couvre les années 2004-2015, au cours desquelles les fondements juridiques du SSE vont être considérablement renforcés.

Dans un contexte marqué par l'élargissement de l'Union européenne (UE) à de nouveaux États membres et la réforme des institutions communautaires, le SSE cherche à renforcer ses bases juridiques. L'enjeu est de garantir la fiabilité et l'harmonisation des statistiques européennes nécessaires au pilotage des politiques de l'UE, dont le spectre s'agrandit. Cette consolidation juridique est accélérée à compter de 2004, à la suite de dérapages comptables qui nuisent à la crédibilité du SSE. Le code de bonnes pratiques de la statistique européenne est la première réponse apportée à cette perte de confiance. La Commission européenne souhaite néanmoins aller plus loin en proposant en 2007 une nouvelle « loi statistique » qui deviendra le règlement n° 223/2009, lequel intègre certains principes du code, donne une assise juridique au SSE et définit son schéma de gouvernance. L'organisation du SSE devient plus pyramidale, avec à son sommet le comité du SSE, qui regroupe l'ensemble des directeurs généraux des instituts nationaux de statistique. Cependant, une nouvelle crise des finances publiques à la fin des années 2000 pousse la Commission à réviser cette loi fraîchement établie pour renforcer l'indépendance professionnelle des statisticiens.

 This second chapter of the history of the European Statistical System (ESS) covers the years 2004 to 2015, when its legal grounds were considerably strengthened.

At a time marked by the enlargement of the European Union (EU) to new members and also by the reform of the institutions of the Community, the ESS aimed at consolidating its legal grounds. The goal was to guarantee the reliability and the harmonisation of European statistics, necessary to the management of EU policies—whose spectrum is broadening. The strengthening of the legal basis accelerated starting from 2004, following several accounting misconducts that tarnished the reputation of the ESS. The European Statistics Code of Practice was the first answer brought to this loss in trust. Nonetheless, the European Commission wished to go further by proposing a new "statistical law" in 2007—becoming Regulation n° 223/2009—which encompasses some principles of the Code, provides a legal foundation to the ESS and defines its governance framework. The organisation of the ESS became more pyramidal, with the ESS committee at its top, including all directors general of national statistical institutes. However, a new public finances crisis at the end of the 2000s incited the Commission to revise this newly established law to reinforce professional independency among statisticians.

* Conseiller auprès de la directrice générale d'Eurostat.

pierre.bischoff@ec.europa.eu

Note : Les opinions exprimées dans le présent article sont celles de l'auteur et ne reflètent pas nécessairement la position de la Commission européenne.

** Membre de l'Inspection générale de l'Insee.

herve.piffeteau@insee.fr

À l'aube des années 2000, le **système statistique européen (SSE)**¹ semble en position favorable pour aborder sereinement le nouveau millénaire. Fort des avancées réalisées au cours des cinquante premières années (Bischoff et Piffeteau, 2025), il s'est structuré



Le SSE se prépare maintenant à s'élargir à d'autres pays, tout en essayant de répondre aux nouveaux besoins d'indicateurs.



autour d'un ensemble d'instruments – par exemple, le programme statistique européen – et d'organes – au centre desquels se trouve le comité du programme statistique (CPS). Ces instruments et organes façonnent les orientations stratégiques et opérationnelles sous-jacentes à l'élaboration des statistiques officielles européennes. Le SSE est par ailleurs doté d'un triptyque juridique : la décision du Conseil créant le CPS en 1989, le règlement du Conseil de 1997 relatif à la statistique communautaire (ou « loi statistique de

1997 ») et le règlement du Conseil de 1990 relatif au secret statistique². Le SSE se prépare maintenant à s'élargir à d'autres pays, tout en essayant de répondre aux nouveaux besoins d'indicateurs que l'approfondissement continu des politiques européennes entraîne sans cesse.

Au début des années 2000, l'Union européenne (UE) à quinze a subi une récession économique touchant plusieurs de ses membres. Dans la perspective de son élargissement à vingt-cinq, avec l'adhésion de dix nouveaux pays³, elle a introduit le traité de Nice (2001)⁴ pour réformer ses institutions. L'euro est par ailleurs devenu la monnaie officielle dans douze États membres en 2002.

L'UE engage à présent une réforme supplémentaire de ses institutions (traité de Lisbonne adopté en 2007⁵). L'enjeu est de surmonter les défis démocratiques et institutionnels posés par son expansion, et de renforcer sa gouvernance dans une UE à vingt-cinq entérinée le 1^{er} mai 2004. L'idée est également de pousser davantage l'intégration politique et d'approfondir le projet européen à la suite des grands succès de la « période Delors »⁶, comme la réalisation du marché unique. Enfin, c'est aussi l'époque où émerge la société de l'information à travers le développement des nouvelles technologies de l'information et de la communication, en particulier les réseaux de données⁷, la téléphonie mobile et plus généralement Internet.

Dans ce contexte historique favorable, le SSE réfléchit à son propre avenir et à la façon d'affronter ces nouveaux défis. En septembre 2002, lors de la conférence des directeurs généraux des instituts nationaux de statistique (DGINS) à Palerme, les discussions sur l'avenir du SSE témoignent à la fois de l'optimisme et de l'ambition des responsables statistiques (Eurostat, 2003). En novembre 2002, dans la perspective d'une constitution européenne⁸, le CPS décide de confier à une task force la charge d'établir un plan d'action

¹ Voir le lien : <https://ec.europa.eu/eurostat/fr/web/main/about-us/european-statistical-system>.

² Voir les références juridiques en fin d'article.

³ En 2004, l'UE intègre dix nouveaux pays : Chypre, Estonie, Hongrie, Lettonie, Lituanie, Malte, Pologne, Slovaquie, Slovaquie, Slovaquie, Slovaquie.

⁴ Voir les références juridiques en fin d'article.

⁵ Voir les références juridiques en fin d'article.

⁶ Jacques Delors a présidé la Commission européenne de 1985 à 1994.

⁷ Systèmes grâce auxquels les informations sont transmises sous forme numérique d'un point à un autre.

⁸ Le traité établissant une constitution pour l'Europe (TECE) est signé en octobre 2004. Mais le projet est abandonné en 2005, la France et les Pays-Bas ayant refusé, par référendum, de ratifier le texte.

portant spécifiquement sur les questions institutionnelles et juridiques concernant le futur du SSE. Cette task force recommande notamment l'établissement d'une base solide du SSE dans la constitution, la révision du cadre juridique régissant les statistiques européennes, et l'élaboration d'un code de bonnes pratiques.

L'époque est donc marquée par de nouvelles ambitions et un optimisme quant à la capacité du système à se renforcer toujours plus. La réalité des faits viendra néanmoins contrarier ces plans et va contraindre le SSE à réagir davantage aux événements plutôt qu'à les anticiper, ouvrant ainsi une période où se succéderont crises et consolidations (*figure 1*).

► Les dérapages comptables nuisent à la crédibilité du SSE —

Au début des années 2000, l'harmonisation des comptes nationaux était relativement bien encadrée avec la mise en œuvre, depuis la fin des années 1990, du *système européen des comptes* (SEC 95)⁹. Néanmoins, des « dérapages » comptables pouvaient se produire à deux niveaux.

D'une part, le comptable public pouvait parfois, sous la pression de l'autorité politique, être amené à interpréter de manière contestable les règles de la comptabilité nationale. Dans un rapport de 2010, la commission des finances du Sénat a révélé que la plupart des États membres se sont rendus coupables, à des degrés variables, de ce genre de pratique (Sénat, 2010). Jusqu'au milieu des années 2000, les États membres ont fréquemment recouru à certaines techniques d'optimisation comptable pour réduire leur dette ou leur déficit : partenariats public-privé, titrisation de revenus futurs, recettes exceptionnelles¹⁰. Les deux premières pratiques ont été rendues impossibles par deux décisions d'Eurostat¹¹, datant respectivement de 2004 et 2007. La troisième a, quant à elle, perdu l'essentiel de son intérêt depuis la réforme du *Pacte de stabilité et de croissance*¹² en 2005, puisque la trajectoire des finances publiques est désormais évaluée en fonction du déficit public structurel, hors mesures temporaires.

D'autre part, même si cela était plus rare, les dérives les plus graves concernaient le cas où les données de comptabilité publique utilisées par le comptable national étaient délibérément faussées. Ce phénomène n'aurait cependant concerné qu'un seul État de l'Union européenne : la Grèce (*encadré 1*).

Ainsi, les techniques d'optimisation comptable utilisées par les États membres pour réduire leur déficit public jusqu'au milieu des années 2000 ont pu nuire à la sincérité des finances publiques. La révision des comptes publics grecs en 2004 a également contribué à détériorer la confiance des utilisateurs dans les statistiques européennes. Elle a été non seulement dommageable pour ce pays, mais aussi pour l'ensemble des États membres de l'UE, soit par un effet de contagion, soit par une perte de confiance dans la fiabilité des statistiques des finances publiques.

⁹ Voir le lien : <https://ec.europa.eu/eurostat/fr/web/products-manuals-and-guidelines/-/ca-15-96-001>.

¹⁰ Le rapport du Sénat explicite de manière détaillée le type de pratiques comptables contestables qui a pu être mis en œuvre dans ces domaines (voir pp. 23-26).

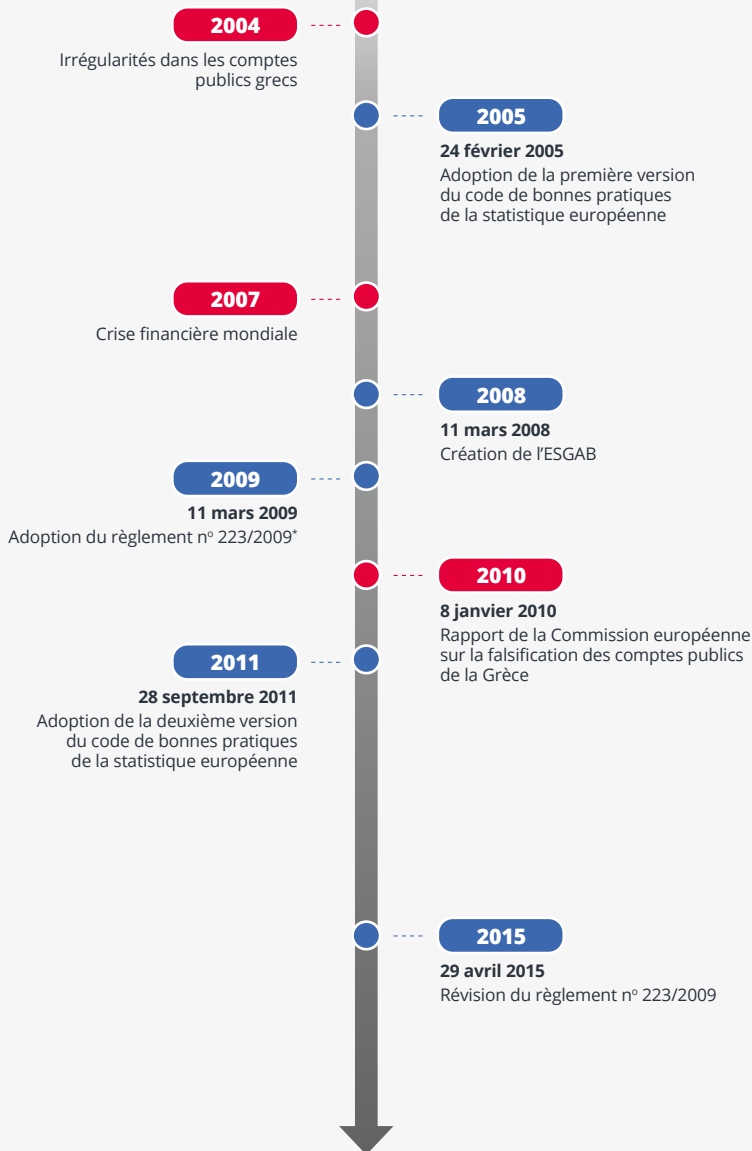
¹¹ Eurostat est l'institut statistique communautaire, direction générale de la Commission européenne.

¹² Voir le lien : <https://www.touteurope.eu/le-glossaire-de-l-europe/pacte-de-stabilite-et-de-croissance/>.

► Figure 1 - L'élaboration du cadre juridique du système statistique européen au gré des crises internationales

● Avancées et crises européennes

● Construction du système statistique européen



* Règlement (CEE) n° 223/2009 du Parlement européen et du Conseil du 11 mars 2009 relatif aux statistiques européennes.

ESGAB : *European Statistical Governance Advisory Board* ou Conseil consultatif européen pour la gouvernance statistique.

► Encadré 1. La révision des comptes publics grecs en 2004

La révision des comptes publics grecs en 2004, à la suite du changement de gouvernement, a révélé d'importantes manipulations des données financières par les autorités grecques en vue de l'entrée de leur pays dans la zone euro*. En septembre 2004, l'institut statistique grec (ELSTAT) notifie à Eurostat des chiffres révisés à la hausse par rapport à ceux notifiés en mars de la même année. Ces derniers concernent le déficit et la dette publics pour les années 2000 à 2003. Les nouveaux chiffres de déficit dépassent largement la limite de 3 % du produit intérieur brut (PIB) fixée par le Pacte de stabilité et de croissance : 4,1 % en 2000, 3,7 % en 2001 et 2002, et 4,6 % en 2003, contre des estimations initiales bien inférieures.

Bien plus grave, les manipulations remontaient en réalité à 1997, lorsque la Grèce a falsifié ses comptes pour intégrer la zone euro. Les déficits réels étaient de 6,6 % en 1997, 4,3 % en 1998 et 3,4 % en 1999, bien au-dessus des critères européens.

Dans un rapport datant de fin 2004, Eurostat identifie les principaux éléments expliquant la révision des chiffres (Eurostat, 2004) : la sous-estimation des dépenses militaires, la surestimation de l'excédent des fonds de sécurité sociale et la révision à la baisse des estimations des recettes fiscales, principalement celles de la taxe sur la valeur ajoutée (TVA). Les divergences dues à l'enregistrement des équipements militaires ont été particulièrement importantes, représentant 90 % du total des révisions en 2000, 50 % en 2001, 75 % en 2002 et 25 % en 2003.

* La Grèce a présenté officiellement sa candidature pour intégrer la zone euro le 9 mars 2000. Cette candidature a été validée par le Conseil européen le 19 juin 2000 et la Grèce a rejoint la zone euro le 1^{er} janvier 2001.

Le Conseil ECOFIN¹³ avait déjà reconnu en juin 2004 l'existence de déficiences dans la compilation et la présentation des statistiques budgétaires à la Commission. Il avait relevé aussi la sensibilité de celles-ci aux alternances politiques et électorales. La Commission était en conséquence invitée à renforcer le contrôle de la qualité de ces données et à proposer des normes européennes minimales pour l'organisation institutionnelle des autorités compétentes en matière de statistiques.

Les événements de 2004 ne font donc en réalité que confirmer les craintes déjà présentes chez bien des acteurs politiques, qui réalisent que l'assise statistique de l'Union économique et monétaire (UEM) est beaucoup plus fragile que ce que les responsables statistiques semblaient indiquer. Le langage feutré des documents officiels ne rend pas compte du caractère brutal avec lequel certaines opinions ont pu s'exprimer à cette époque. De plus, chacun essaye de trouver le bon coupable et le doute s'installe sur la capacité du SSE à délivrer des statistiques suffisamment robustes, pourtant cruciales pour certaines politiques européennes.

En décembre 2004, sous la pression du Conseil ECOFIN, la Commission adopte deux communications successives au Parlement européen et au Conseil.

Dans la première communication, la Commission souligne qu'Eurostat n'est pas habilité à contrôler directement les comptes publics des États membres. Elle met en avant le fait que les échanges entre Eurostat et les services nationaux de statistique reposent sur la bonne foi et la coopération des autorités nationales, mais qu'Eurostat ne peut ni établir directement les données à partir de sources de base, ni contraindre les autorités nationales à fournir des informations, ni procéder à des vérifications sur place.

¹³ Le Conseil pour les Affaires économiques et financières, aussi appelé « Conseil ECOFIN », est la formation du Conseil rassemblant les ministres de l'Économie et des Finances des États membres, ainsi que les ministres compétents en matière de budget lorsque des questions budgétaires sont à l'ordre du jour.

Dans la seconde communication, la Commission expose une stratégie européenne de gouvernance des statistiques budgétaires. Elle propose trois lignes d'action :

1. Elle pointe la nécessité de clarifier les règles portant sur la qualité des données statistiques utilisées dans le cadre de la procédure concernant les déficits excessifs (PDE). Elle annonce une proposition visant à habilitier Eurostat à contrôler directement les comptes gouvernementaux et à procéder à une vérification effective des données notifiées par les États membres.
2. Elle annonce un renforcement des capacités opérationnelles d'Eurostat, avec la possibilité d'effectuer des missions de vérification plus longues et plus approfondies.
3. Elle préconise de mettre au point des normes européennes pour l'indépendance, l'intégrité et la responsabilité des instituts nationaux de statistique.

► Le code de bonnes pratiques est une première réponse à la perte de crédibilité des statistiques européennes —

La solidité du SSE repose en grande partie sur le partage de valeurs communes autour d'une statistique impartiale, fondée sur des approches scientifiques, voire sur une même éthique professionnelle. Dans un système désormais élargi à vingt-cinq États membres, ces valeurs ont besoin d'être renforcées et inscrites dans un « code », garantissant la qualité des statistiques européennes. C'est d'autant plus important que les statistiques européennes servent non seulement au suivi des politiques européennes, mais aussi, désormais, à la gouvernance de l'UEM. Cette recherche d'une démarche qualité systématique avait déjà été initiée par les pays nordiques dans les années 1990 (Franchet, 2004). Elle reflète dans une large mesure les normes internationales déjà adoptées par la Commission statistique des Nations Unies dès avril 1994¹⁴.

La demande politique pour la mise en place d'un tel « code » se fait par ailleurs pressante, compte tenu de la perte de confiance induite par la falsification des comptes publics. Dans ce contexte, le CPS décide d'accélérer les travaux initiés quelques années auparavant. Un projet est élaboré fin 2004-début 2005 sous l'égide d'une task force présidée par le directeur général de l'Insee de l'époque, Jean-Michel Charpin. En février 2005, le CPS adopte formellement le **code de bonnes pratiques de la statistique européenne**¹⁵ (CoP dans son acronyme anglais, pour *Code of Practice*). Comme le soulignent Madinier et al. (2007), c'est une façon de redorer l'image des statisticiens dans ce contexte de crise et de répondre à l'invitation du Conseil à « mettre au point des normes européennes minimales pour l'organisation institutionnelle des autorités compétentes en matière de statistiques ».

Le CoP recouvre toute la gamme des statistiques produites au niveau communautaire, ce qui dépasse de beaucoup le domaine des statistiques économiques et budgétaires. Il vise à définir des normes communes qui couvrent l'environnement institutionnel, les procédures statistiques et les résultats statistiques. Ce sont des principes qui allient déontologie du système statistique européen (valeurs) et qualité des résultats et des processus de production mis en œuvre¹⁶.

¹⁴ Voir le lien : <https://unstats.un.org/unsd/statcom/doc94/f1994.htm>.

¹⁵ Voir le lien : <https://www.insee.fr/fr/information/4140105>.

¹⁶ Voir Briant et Kranklader (2025) à propos des différentes dimensions de la qualité des statistiques.



Le CoP vise à définir des normes communes qui couvrent l'environnement institutionnel, les procédures statistiques et les résultats statistiques.



Le CoP vient ainsi préciser ce que représentent concrètement **les principes d'indépendance et d'intégrité pour les statisticiens**. En réponse à la crise de défiance qui a affecté le SSE, il s'agit de mettre l'accent sur l'indépendance professionnelle des autorités statistiques, afin de renforcer leur intégrité et la crédibilité des données produites. Celle-ci était déjà partiellement mentionnée dans la loi statistique de 1997 au titre du principe d'impartialité, notamment dans le choix des techniques et des méthodologies mises en œuvre pour produire les statistiques officielles¹⁷. Cette indépendance professionnelle va

bien au-delà de l'indépendance scientifique comme le souligne Rouault (2007). Avec le CoP, le CPS a souhaité souligner, d'un point de vue plus institutionnel, l'indépendance des autorités statistiques « à l'égard des interventions politiques et autres interférences externes dans l'élaboration, la production et la diffusion des statistiques »¹⁸.

Il s'agit aussi pour le CPS de mettre en avant **la qualité des résultats et des processus statistiques sous-jacents**. Le CoP définit ainsi un socle de qualité commun garantissant la cohérence et la fiabilité des statistiques produites par les autorités statistiques nationales et européennes, mais aussi la méthodologie et les procédures statistiques à mettre en œuvre pour y parvenir. Il fixe des normes relatives à la création, la production et la communication des statistiques européennes, assurant leur pertinence, leur exactitude et leur transparence. Il vise notamment à renforcer la confiance des utilisateurs, en posant des principes clairs pour des indicateurs de qualité (**encadré 2**).

Dans sa communication de décembre 2004 relative à la stratégie européenne de gouvernance des statistiques budgétaires (voir supra), la Commission avait annoncé que divers instruments juridiques possibles seraient examinés lors de l'étude de la question des normes minimales pour l'édifice institutionnel des autorités statistiques. Cet examen a révélé une grande diversité dans l'organisation des autorités statistiques des États membres. La Commission considère au final qu'il n'est pas approprié de proposer un instrument juridiquement contraignant à la suite de l'adoption du CoP. En mai 2005, elle recommande simplement aux États membres de prendre les mesures requises pour garantir la mise en œuvre de ce dernier. Par ailleurs, afin de surveiller le respect du code au sein du SSE, une évaluation par les pairs sera progressivement mise en place (Chaleix et Durr, 2024). Comme le soulignent Madinier et al. (2007), le CoP n'est pas un règlement, ni une directive, mais une procédure de déclaration et de publicité définissant les meilleures pratiques en matière statistique que les instituts nationaux de statistique (INS) de l'UE doivent viser à mettre en œuvre.

La Commission, si elle salue l'avancée que constitue le CoP, souhaite néanmoins aller plus loin. Elle n'exclut pas à terme la possibilité d'envisager d'autres instruments plus contraignants juridiquement que sa recommandation de mai 2005. En amont de cette

¹⁷ Selon l'article 10 de la loi statistique de 1997, « l'impartialité est une manière objective et indépendante de produire des statistiques communautaires, à l'abri de toute pression émanant de groupes politiques ou d'autres groupes d'intérêt, notamment en ce qui concerne le choix des techniques, des définitions et des méthodologies les mieux adaptées à la poursuite des objectifs définis ».

¹⁸ Cf. indicateur 1.1 du principe 1 du CoP.

future possible transcription juridique, elle complète le dispositif autour du CoP avec la création en 2008 du **conseil consultatif européen pour la gouvernance statistique**¹⁹ (aussi connu sous son acronyme anglais ESGAB, pour *European Statistical Governance Advisory Board*). Ce dernier est spécialement chargé d'exercer un contrôle indépendant sur Eurostat, ainsi que sur le SSE dans son ensemble, en ce qui concerne la mise en œuvre du CoP. L'ESGAB devient ainsi le pendant européen de l'Autorité de la statistique publique (ASP) qui sera créée peu après en France (Bureau, 2020).

► **Encadré 2. Le code de bonnes pratiques de la statistique européenne**

Le **code de bonnes pratiques de la statistique européenne** est une autorégulation fondée initialement sur 15 principes majeurs*, organisés autour de trois piliers : environnement institutionnel, producteurs et utilisateurs (**figure encadré**). Des indicateurs sont par ailleurs associés à chacun de ces principes afin de faciliter l'évaluation de la conformité aux principes du code. Un système de vérification, dans l'esprit d'un audit, complètera ce dispositif : la **revue par les pairs****, dont le premier cycle sera lancé en 2006. Cette dernière a pour objectif de vérifier la conformité aux principes du code et de garantir

par là même le respect des valeurs du SSE et de la qualité de sa production statistique.

Le code et sa revue par les pairs visent donc à renforcer la crédibilité des statistiques européennes et notamment le respect des valeurs déontologiques inhérentes au système statistique. Ainsi, le protocole de revue par les pairs évalue avant tout le système statistique, et non pas l'élaboration d'une statistique donnée. La qualité de cette dernière est quant à elle contrôlée par la fourniture de **rapports sur la qualité***** ou par des mécanismes d'évaluation prévus dans la législation sectorielle correspondante.

* 16 à partir de la 2^e révision intervenue en 2017, avec l'ajout du principe 1bis dédié à la coordination et la coopération.
** Voir le lien : <https://ec.europa.eu/eurostat/fr/web/quality/peer-reviews>.
*** Voir le lien : <https://ec.europa.eu/eurostat/fr/web/quality/quality-monitoring/quality-reporting>.



* Ce principe a été ajouté en 2017, lors de la deuxième révision du code de bonnes pratiques.

¹⁹ Voir le lien : <https://eur-lex.europa.eu/FR/legal-content/summary/european-statistical-governance-advisory-board.html>.

► La nouvelle loi sur la statistique européenne consolide l'assise juridique du SSE

C'est en 2007 que la Commission propose d'aller plus loin en renforçant la législation européenne. Cette proposition débouchera sur l'adoption, le 11 mars 2009, du règlement (CE) n° 223/2009 du Parlement européen et du Conseil relatif aux statistiques européennes²⁰. Cette « loi statistique de 2009 » constitue une avancée majeure vers la mise en place d'un SSE renouvelé et plus efficace, en réponse notamment aux crises récentes. Elle reprend quelques-uns des principes du CoP, en particulier ceux qui ont pour but de renforcer la confiance dans les autorités statistiques, en inscrivant dans la loi des dispositions relatives à l'organisation du SSE.

Cela se traduit d'abord par **la reconnaissance juridique du SSE**. Le SSE est certes établi de longue date, il s'est développé progressivement, mais il n'avait pas de fondement juridique jusqu'à l'instauration de la loi statistique de 2009. L'article 4 de la loi définit ainsi le SSE de la manière suivante :

« Le système statistique européen (SSE) est le partenariat entre l'autorité statistique communautaire, c'est-à-dire la Commission (Eurostat), et les instituts nationaux de statistique (INS) ainsi que les autres autorités nationales responsables dans chaque État membre du développement, de la production et de la diffusion de statistiques européennes. »

Cette reconnaissance juridique est importante à plusieurs titres. En premier lieu, le SSE n'est pas l'unique instance en charge de la production de statistiques européennes officielles. En effet, le **système européen de banques centrales (SEBC)**²¹ produit une vaste gamme de statistiques européennes dans le domaine économique et monétaire. Plus précisément, la Banque centrale européenne (BCE), assistée par les banques centrales nationales, collecte, soit auprès des autorités compétentes nationales, soit directement auprès des agents économiques, l'information statistique qui est nécessaire aux tâches du SEBC. La loi statistique de 2009 permet alors de mieux légitimer le rôle du SSE, chargé pour l'essentiel de la production et de la diffusion des indicateurs de suivi des politiques européennes « autres que celles relevant des banques centrales ».

Cette reconnaissance est d'autant plus nécessaire que les échanges entre les autorités nationales et Eurostat n'étaient pas toujours formalisés juridiquement. Ceux-ci s'appuyaient sur des accords informels entre parties (*gentlemen's agreements*) reposant sur la parole donnée, plutôt que sur des obligations légales, au moins jusqu'à la création du CPS en 1989.

Enfin, dans la loi statistique de 1997, le concept de système statistique européen était soigneusement évité (De Michelis et Chantraine, 2003). Son inscription en 2009 est ainsi primordiale : l'élaboration des statistiques communautaires repose sur une coopération atypique de l'intégration européenne (Bischoff et Piffeteau, 2025). En effet, traditionnellement, lorsque la Commission souhaite proposer de nouveaux règlements,

²⁰ Voir les références juridiques en fin d'article. Voir par ailleurs Mordant (2009) pour une présentation de cette loi.

²¹ Voir le lien : <https://www.touteleurope.eu/le-glossaire-de-l-europe/systeme-europeen-de-banques-centrales/>.

elle les prépare seule et les soumet aux États membres uniquement au stade des négociations politiques. Pour les règlements statistiques, cependant, Eurostat associe très en amont les INS à sa réflexion (Piffeteau, 2019).

Selon Yves Franchet, ancien directeur général d'Eurostat entre 1987 et 2003, « si cette démarche "en réseau" présente de nombreux avantages, elle constituera un obstacle à la reconnaissance du SSE par les autorités de l'UE. Le service juridique de la Commission refusera de reconnaître [jusqu'en 2009] un système qui lui paraît remettre en cause les rôles respectifs de la Commission et des États membres (Conseil). Pour les juristes de la Commission, l'intervention de représentants des pays membres d'une façon officielle avant qu'elle n'ait élaboré sa propre position [était] un risque à ne pas prendre » (Franchet, 2014).

Pour l'élaboration de la législation statistique, le rôle de « guide » du SSE ne sera ainsi jamais reconnu formellement dans les textes juridiques européens. La loi statistique de 2009 se bornera à caractériser ce qu'est le SSE.

La loi statistique de 2009 établit également **un nouveau schéma de gouvernance du SSE**, indispensable pour renforcer la coopération entre les divers acteurs concernés et consolider l'ensemble de ses activités. Dans un contexte où la crise des comptes publics avait pointé un manque de coordination au sein du SSE, le législateur européen



Les INS se voient confier la coordination des activités de développement, production et diffusion des statistiques européennes au niveau national, tandis qu'Eurostat assume ce rôle au niveau communautaire.

a ainsi décidé de clarifier davantage le rôle des différents membres du système. À cette fin, les INS se voient confier une fonction de coordination pour toutes les activités de développement, production et diffusion des statistiques européennes au niveau national, tandis qu'Eurostat assume ce rôle au niveau communautaire. C'est aussi la première fois que le schéma de gouvernance confère à l'INS un rôle central au plan national en tant que point de contact unique pour Eurostat.



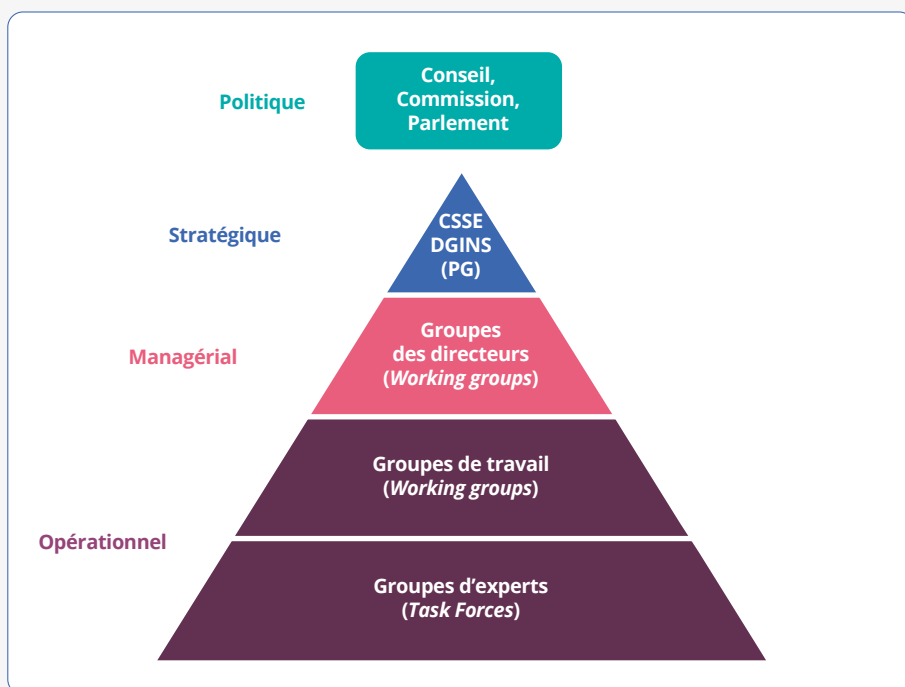
L'organisation du SSE se structure de façon pyramidale (figure 2), avec au sommet

le **Comité du système statistique européen (CSSE)**²² qui se substitue au CPS. La loi statistique précise explicitement que le CSSE est composé des représentants des INS alors que les lois européennes laissent habituellement aux États membres le soin de désigner qui les représentent au sein des comités.

La mission prioritaire du CSSE est de fournir des conseils professionnels au SSE pour tout ce qui concerne les statistiques européennes, mais surtout d'établir des priorités en matière de statistiques. Le CSSE devient ainsi l'organe de pilotage du système. Il hérite également des responsabilités spécifiques de l'ex-comité européen du secret statistique, chargé de la préservation des règles de confidentialité pour les données obtenues à des fins exclusives de production de statistiques européennes (**encadré 3**). Cela signifie qu'il devient la pierre angulaire de l'architecture du SSE, étant chargé de questions telles que

²² Voir le lien : <https://ec.europa.eu/eurostat/fr/web/european-statistical-system/governance-bodies/essc>.

► **Figure 2 - L'organisation du système statistique européen à la suite de la loi statistique de 2009**



CSSE : Comité du système statistique européen.

DGINS : Conférence des directeurs généraux des instituts nationaux de statistique (INS).

PG : *Partnership Group* ou groupe de partenariat. C'est une collaboration entre plusieurs directeurs d'INS et le directeur général d'Eurostat, en général trois fois par an.

la planification et la programmation, la qualité et la méthodologie, la définition des axes prioritaires et la coopération internationale. Progressivement, le CSSE se substituera aux autres comités (Bischoff et Piffeteau, 2025 ; voir figure 3).

La planification et la programmation des travaux sont davantage explicitées.

La production de statistiques d'un haut niveau de qualité et de fiabilité nécessite en effet de faire la part entre, d'un côté, les exigences en matière de données, et de l'autre, les ressources à la disposition des autorités statistiques et le fardeau imposé aux répondants. Or l'extension des politiques communautaires engendre une multiplication des demandes de statistiques européennes. Il importe alors que les autorités statistiques soient en capacité de fixer des objectifs pour réduire ce fardeau au fil du temps, qui pèse à la fois sur les unités enquêtées et sur les autorités statistiques elles-mêmes. Dès juin 2005, le Conseil ECOFIN souligne que « pour permettre aux instituts nationaux de statistique de satisfaire à des normes élevées en matière de statistiques, il est essentiel de pouvoir fixer un ordre de priorités parmi les exigences de l'UE à l'égard des autorités dans ce domaine ». Certes, la mention d'un programme de travail était déjà présente dans la loi statistique de 1997 et celle-ci précisait que des actions prioritaires devaient

être identifiées dans le programme soumis au CPS²³. Toutefois, dans un contexte de défiance à l'égard des statistiques européennes, la nouvelle loi met clairement l'accent sur des aspects tels que :

- l'efficacité au regard du coût ;
- l'établissement des priorités ;
- la réduction de la charge du répondant ;
- l'importance de préciser de quelle façon sera menée telle ou telle action.

À ce titre, la loi statistique de 2009 reprend là encore certaines des dispositions du CoP, notamment le principe 9 intitulé « charge non excessive pour les répondants » et le principe 10 intitulé « rapport coût-efficacité ».

Toutes ces nouvelles dispositions visent à consolider le SSE dans son ensemble et à pallier, en partie, les carences identifiées jusqu'alors. Si elles s'avèrent globalement efficaces pour remédier aux déviations comptables observées dans la très grande majorité des États membres, elles vont se révéler cependant encore insuffisantes pour éviter de nouveaux dérapages comptables.

► Encadré 3. Le concept du secret statistique dans la loi statistique de 2009

La loi statistique de 2009 établit des principes, des règles et des procédures visant à protéger les **données confidentielles**, tout en permettant leur utilisation à des fins statistiques. Elle définit les données confidentielles comme des données permettant l'identification d'unités statistiques, soit directement – par des identifiants formels –, soit indirectement – par le croisement de différentes caractéristiques.

Pour ce qui est de la **protection des données**, la règle fondamentale est que les données confidentielles doivent être protégées de toute divulgation. Il existe cependant deux exceptions à cette règle :

- la première, lorsque l'unité statistique donne son consentement à une telle diffusion ;
- la seconde, lorsque des conditions spécifiques déterminées par un acte arrêté par le Parlement européen et le Conseil autorisent la diffusion, auquel cas les États membres appliquent le principe de « confidentialité passive » : ils ne prennent des mesures de protection qu'à la demande de l'unité

statistique, si celle-ci considère que ses intérêts seraient lésés par la diffusion des données.

Pour ce qui est de la **transmission des données**, le cadre est plus souple lorsqu'il s'agit d'échange de données confidentielles :

- entre deux autorités statistiques du SSE (au sein d'Eurostat et des autorités statistiques des États membres) ;
- entre une autorité statistique du SSE et un membre du SEBC (BCE ou banque centrale d'un État membre).

La loi statistique de 2009 autorise alors un tel échange, et ce indépendamment du lieu d'installation des deux parties, qu'elles soient installées au sein du même État membre ou dans des États membres différents. La condition est que la transmission soit nécessaire pour le développement, la production ou la diffusion de statistiques européennes, ou pour en accroître la qualité. Dans le cadre de toute transmission à un membre du SEBC, cette nécessité doit être justifiée.

²³ Voir chapitre II, article 3, alinéa 3 : « Chaque année, avant la fin du mois de mai, la Commission soumet à l'examen du comité du programme statistique son programme de travail pour l'année suivante. Dans ce programme, elle précise notamment [...] les actions qu'elle considère prioritaires, compte tenu des contraintes financières tant nationales que communautaires. » Comme le souligne Jean-Michel Charpin, « la montée en puissance du débat sur les priorités du programme statistique européen va obliger les pays à progresser dans leur processus de coordination interne des services producteurs de statistiques publiques » (Charpin, 2007).

► Cette consolidation est toutefois mise à mal par une nouvelle crise des statistiques relatives aux finances publiques...

Les mesures décrites précédemment (CoP, ESGAB, loi statistique de 2009) renforcent effectivement la confiance dans les statistiques européennes, notamment à travers davantage de coopération et coordination entre les membres du SSE. Elles s'avèrent toutefois encore insuffisantes dès lors que les problèmes rencontrés vont au-delà d'une simple question de carences statistiques.

La notification par la Grèce des chiffres du déficit et de la dette en octobre 2009 (après un nouveau changement de gouvernement) semble relever du mauvais scénario et présente un air de déjà-vu mais en pire. Une fois encore, les chiffres sont fortement révisés à la hausse par rapport à la notification de mars de la même année. Il s'agit d'un

dysfonctionnement global des institutions publiques, un système où la statistique obéit trop souvent aux ordres du Trésor²⁴.

La fiabilité des statistiques publiques de la zone euro est profondément remise en cause et à un degré encore supérieur à celui de 2004.

Cette fois-ci, les statistiques sont à l'origine d'une crise véritablement existentielle pour l'euro. L'impact est majeur et renvoie à la question des dettes souveraines européennes. La fiabilité des statistiques publiques de la zone euro est profondément remise en cause et à un degré encore supérieur à celui de 2004. De son côté, la Commission européenne déplore que le précédent renforcement des pouvoirs d'Eurostat ne soit pas allé

plus loin et soit resté borné aux questions statistiques (Commission européenne, 2010). En effet, aux termes du règlement (CE) n° 479/2009 du Conseil²⁵, les visites méthodologiques qu'Eurostat peut alors mener pour vérifier la qualité des données sous-tendant les comptes des administrations publiques ne permettent pas d'aller au-delà du domaine purement statistique. Cette limite fournit aux États membres des arguments possibles pour restreindre l'accès aux informations exigées par Eurostat.

Le règlement (CE) n° 479/2009 est ainsi révisé en 2010 par le Conseil, sur proposition de la Commission, pour accorder spécialement à Eurostat un pouvoir étendu d'accès aux comptes de toutes les entités publiques²⁶. Par ailleurs, fin 2011, dans le cadre du renforcement de la gouvernance économique et de la zone euro visé par l'initiative du « six-pack »²⁷, Eurostat devient habilité à ouvrir des enquêtes s'il existe des indices sérieux

²⁴ Auditionné en 2010 par la commission des finances, Paul Champsaur (directeur général de l'Insee de 1992 à 2003) a été très clair à ce sujet : « la statistique grecque était inféodée au Trésor, presque « organiquement », et cette consanguinité a posé problème. [...] Les statisticiens grecs étaient dans une situation qui ne correspondait pas à la norme de déontologie. Le Trésor était beaucoup trop consanguin vis-à-vis de la statistique » (Sénat, 2010).

²⁵ Règlement (CE) n° 479/2009 du Conseil du 25 mai 2009 relatif à l'application du protocole sur la procédure concernant les déficits excessifs annexé au traité instituant la Commission européenne. Voir les références juridiques en fin d'article.

²⁶ Le règlement rectificatif est le règlement (UE) n° 679/2010 du Conseil du 26 juillet 2010. Voir les références juridiques en fin d'article.

²⁷ Le six-pack est un ensemble de six textes législatifs (cinq règlements et une directive), entré en vigueur le 13 décembre 2011, visant à renforcer le Pacte de stabilité et de croissance. En particulier, le premier règlement confère davantage de pouvoir à Eurostat : le règlement (UE) n° 1173/2011 du Parlement européen et du Conseil du 16 novembre 2011 sur la mise en œuvre efficace de la surveillance budgétaire dans la zone euro. Voir les références juridiques en fin d'article.

que des statistiques ont été manipulées intentionnellement ou par grave négligence. Il peut également effectuer des inspections sur place à cette fin, ce qui peut finalement amener le Conseil à infliger une sanction financière à l'État membre concerné. Même tardifs, ces changements sont majeurs, et certains responsables politiques admettent qu'ils auraient dû intervenir plus tôt. Ainsi, une fois devenu Président de la Commission européenne, Jean-Claude Juncker a reconnu publiquement l'erreur de ne pas avoir accepté de donner plus de pouvoir et d'indépendance à Eurostat quand il représentait le Luxembourg au Conseil (**encadré 4**).

La question relative à l'indépendance professionnelle des INS vis-à-vis du pouvoir politique reste néanmoins ouverte. Même si elle apparaît juridiquement pour la première fois dans la loi statistique de 2009, elle reste traitée de manière encore insuffisante et partielle. Elle n'est mentionnée que dans l'article 2 au titre des principes statistiques, tels que décrits dans le CoP.

► **Encadré 4. Extrait d'un discours du président Juncker au forum de la BCE sur les banques centrales de 2019, célébrant les vingt ans de l'Union économique et monétaire**

« Il y a une troisième leçon à tirer : l'euro est une question d'intérêt commun.

Ce troisième moment, cette troisième leçon, nous vient de Grèce lorsque, en octobre 2009, le gouvernement a admis que les statistiques officielles avaient faussé les chiffres réels de la dette et du déficit pendant des années. Cela a alimenté la méfiance et l'incertitude à tous les niveaux. Et cela a révélé l'une des plus grandes contradictions et absurdités de l'Union économique et monétaire européenne.

En réalité, Eurostat avait envoyé dix délégations à Athènes afin d'améliorer la fiabilité des statistiques grecques entre 2004 et 2010. Cependant, les États membres résistaient depuis des années à un renforcement des règles européennes en matière de statistiques. Premiers ministres, ministres des Finances, gouvernements, tous arguaient que la

souveraineté nationale était plus importante que des données fiables et comparables.

Je dois assumer ma part de responsabilité, car j'étais parmi ceux qui ont demandé à la Commission de ne pas poursuivre dans cette voie. J'ai voté contre la proposition de la Commission visant à donner plus de pouvoir et plus d'indépendance à Eurostat. Ce fut une grave erreur. Si nous avions fait le bon choix, nous n'aurions jamais connu la crise grecque telle que nous l'avons vécue. Les actions de ces institutions, Eurostat et d'autres, doivent être responsables et transparentes, ce qui signifie également qu'elles doivent être ancrées dans le processus décisionnel de l'Union européenne. C'est pourquoi j'insiste pour que les solutions intergouvernementales héritées des années de crise soient progressivement intégrées au cadre institutionnel de l'Union européenne. »

► **... suivie d'un nouveau renforcement de l'indépendance professionnelle des statisticiens**

Au-delà des carences institutionnelles qui affectent le fonctionnement des entités responsables de la gestion des recettes et des dépenses publiques dans les États membres, il s'avère que le positionnement des autorités statistiques vis-à-vis du système administratif ne garantit pas pleinement l'indépendance, l'intégrité et la responsabilité des autorités statistiques nationales. L'indépendance professionnelle de certains instituts statistiques à l'égard du ministère des Finances n'est toujours pas assurée. Pour ces derniers, la notification des données relatives à la procédure concernant les déficits excessifs, telle que prévue par la loi européenne, n'a donc pas suivi les principes juridiquement contraignants qu'il convenait d'appliquer « pour la production de statistiques européennes de haute qualité ». Elle a été influencée par d'autres facteurs.

C'est pourquoi le Conseil ECOFIN de novembre 2009 pointe la nécessité, d'une part, de mettre en place des méthodes de travail transparentes et fiables entre les services nationaux concernés, d'autre part, de revoir leur organisation institutionnelle, afin de renforcer cette indépendance professionnelle. C'est notamment à la suite de cette recommandation

Seules des statistiques « neutres » offrent une base solide pour prendre des décisions pour l'avenir.

qu'Eurostat cherchera à renforcer, par de nouvelles dispositions législatives, la gouvernance du SSE et en particulier son indépendance vis-à-vis du pouvoir politique.

L'efficacité des politiques européennes passe par des statistiques publiques crédibles et des décisions fondées sur des éléments probants. La conception, la production et la diffusion de statistiques publiques fiables ne doivent pas être laissées à la charge

des administrations ordinaires, souvent influencées par l'orientation politique des gouvernements. Le risque est que les statistiques soient considérées comme des outils au service d'ambitions politiques diverses. Or, c'est l'inverse qui doit prévaloir. Seules des statistiques « neutres », qui rendent compte des évolutions passées effectives, que la situation soit ou non en accord avec les objectifs politiques, offrent une base solide pour prendre des décisions pour l'avenir.

Dans un rapport publié en 2012, la Cour des comptes européenne précise : « Dans une Union où les décisions politiques doivent se fonder sur des informations probantes et où un nombre croissant de décisions proviennent directement de données statistiques ou d'indicateurs qui en sont dérivés, il est essentiel que le public puisse se fier aux statistiques européennes » (Cour des comptes européenne, 2012).

Dans ses différents rapports de l'époque, l'ESGAB partage ce point de vue et réaffirme également que des efforts restent à faire en vue de la pleine application du CoP et des principes déjà inscrits dans la loi statistique de 2009. Il souligne en particulier que cette loi devrait être révisée pour :

1. renforcer l'indépendance professionnelle des INS ;
2. dissocier la nomination et la révocation de leurs dirigeants de toutes considérations politiques ;
3. clarifier leur rôle de coordination au sein des systèmes statistiques nationaux ;
4. définir des engagements en matière de confiance dans les statistiques.

En avril 2012, la Commission propose de réviser la loi statistique de 2009 en ce sens. Un règlement rectificateur sera ainsi adopté le 29 avril 2015²⁸. Il renforce le principe d'indépendance en précisant que celle-ci se réfère non seulement à la production de statistiques, mais aussi à leur diffusion et à leur disponibilité. Il précise aussi que cette indépendance doit s'affranchir de « pression émanant de groupes politiques, de groupes d'intérêt, d'autorités nationales ou d'autorités de l'Union ». Autrement dit, il reprend une disposition du CoP visant à inscrire des éléments relatifs à l'indépendance dans le

²⁸ Règlement (UE) n° 2015/759 du Parlement européen et du Conseil du 29 avril 2015 modifiant le règlement (CE) n° 223/2009 relatif aux statistiques européennes. Voir les références juridiques en fin d'article.

droit européen, et pas uniquement dans le droit national, ce qui leur donne davantage de poids compte tenu de la hiérarchie des normes juridiques. Des considérations supplémentaires sont introduites, concernant par exemple le statut et les responsabilités fonctionnelles des chefs des autorités statistiques, ou encore les décisions sur le contenu et le calendrier des communiqués statistiques. Elles renforcent encore l'indépendance globale tout en respectant les dispositions juridiques et administratives différentes et parfois divergentes dans les États membres, expression de leurs traditions politiques et culturelles respectives. De l'ensemble de cette révision, la Commission escompte une plus grande confiance du public dans les statistiques officielles.

► **Malgré ces crises, le SSE a su se réformer dans la perspective de nouveaux défis**

L'histoire du SSE aura ainsi été ébranlée fortement par les deux crises grecques de 2004 et 2009. Au début des années 2000, les bases de sa construction semblaient solidement établies. Le système n'en demeurerait pas moins un simple partenariat entre Eurostat, les INS et les autorités nationales adjacentes, si bien que l'hétérogénéité des situations des États membres le rendait fragile.

L'approfondissement du projet politique européen, marqué notamment par une intégration politique plus forte, nécessitait de renforcer ses fondements juridiques. Mais les deux crises grecques marqueront un coup d'arrêt à cette consolidation. Elles entacheront la crédibilité des statistiques européennes, qui par essence doivent être des indicateurs fiables et harmonisés pour la conduite des politiques européennes de plus en plus larges.

Renforcer la confiance dans les statistiques officielles européennes est ainsi devenu au début des années 2000 un enjeu majeur pour pérenniser le SSE. Le code de bonnes pratiques de la statistique européenne, la loi statistique de 2009 et sa révision en 2015 permettront non seulement de préserver le système, mais aussi de le renforcer dans un contexte d'élargissement continu de l'UE. En effet, l'UE comptera 28 États membres fin 2015, à la suite de l'adhésion de la Bulgarie et de la Roumanie en 2007 et celle de la Croatie en 2013.

Fort de cette consolidation juridique, le SSE reste confronté à de nouveaux défis et notamment ceux induits par la nouvelle stratégie européenne de la donnée. Celle-ci vise à positionner l'UE comme un leader dans l'économie fondée sur les données : d'une part, en favorisant la circulation des données, tout en apportant des garanties aux entreprises et aux individus ; d'autre part, en créant des espaces européens communs des données pour stimuler l'innovation et la compétitivité mondiale. Dans ce contexte, le SSE cherche à exploiter de nouvelles sources de données détenues pour l'essentiel par des entités privées, en particulier par des entreprises, afin de produire de nouvelles statistiques à un niveau de détail de plus en plus fin et satisfaire les attentes des utilisateurs dans des délais de plus en plus contraints.

Au milieu des années 2010, alors qu'il vient de traverser une série de crises et de consolidations, le SSE doit ainsi se consacrer à de nouvelles tâches, tout en gardant à l'esprit que l'indépendance professionnelle et la crédibilité des statistiques restent des enjeux permanents, comme l'a appris la période qui s'achève.

► Fondements juridiques

- Traité de Nice modifiant le traité sur l'Union européenne, les traités instituant les Communautés européennes et certains actes connexes. In : *site de l'Union européenne*. [en ligne]. 26 février 2001. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:12001C/TXT>.
- Traité de Lisbonne modifiant le traité sur l'Union européenne et le traité instituant la Communauté européenne. In : *site de l'Union européenne*. [en ligne]. 13 décembre 2007. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:12007L/TXT>.
- Décision (CEE, Euratom) n° 89/382 du Conseil du 19 juin 1989 instituant un comité du programme statistique des Communautés européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:31989D0382>.
- Règlement (Euratom, CEE) n° 1588/90 du Conseil du 11 juin 1990 relatif à la transmission à l'Office statistique des Communautés européennes d'informations statistiques couvertes par le secret. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A31990R1588&qid=1749649474352>.
- Règlement (CE) n° 322/97 du Conseil du 17 février 1997 relatif à la statistique communautaire. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A31997R0322&qid=1760342561620>.
- Règlement (CEE) n° 223/2009 du Parlement européen et du Conseil du 11 mars 2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32009R0223&qid=1760342713619>.
- Règlement (CE) n° 479/2009 du Conseil du 25 mai 2009 relatif à l'application du protocole sur la procédure concernant les déficits excessifs annexé au traité instituant la Communauté européenne. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv%3AQJ.L_.2009.145.01.0001.01.FRA&toc=OJ%3AL%3A2009%3A145%3ATOC.
- Règlement (UE) n° 679/2010 du Conseil du 26 juillet 2010 portant modification du règlement (CE) n° 479/2009 en ce qui concerne la qualité des données statistiques dans le contexte de la procédure concernant les déficits excessifs. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1775553181583&uri=CELEX%3A32010R0679>.
- Règlement (UE) n° 1173/2011 du Parlement européen et du Conseil du 16 novembre 2011 sur la mise en œuvre efficace de la surveillance budgétaire dans la zone euro. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32011R1173&qid=1775553453526>.

- Règlement (UE) n° 2015/759 du Parlement européen et du Conseil du 29 avril 2015 modifiant le règlement (CE) n° 223/2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32015R0759&qid=1776433686939>.

► Bibliographie

- BISCHOFF, Pierre et PIFFETEAU, Hervé, 2025. La construction du système statistique européen – 1952-2003 : un grand élan. In : *Courrier des statistiques*. [en ligne]. 15 décembre 2025. Insee. N° N14, pp. 7-26. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8675600?sommaire=8667496>.
- BRIANT, Pierrette et KRANKLADER, Élodie, 2025. Des démarches qualité pour sécuriser et optimiser la production de statistiques publiques. In : *Courrier des statistiques*. [en ligne]. 15 décembre 2025. Insee. N° N14, pp. 119-142. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8675630?sommaire=8667496>.
- BUREAU, Dominique, 2020. L'Autorité de la statistique publique – Dix ans d'activité, pour une statistique indépendante et de qualité. In : *Courrier des statistiques*. [en ligne]. 31 décembre 2020. Insee. N° N5, pp. 21-38. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5008696?sommaire=5008710>.
- CHALEIX, Mylène et DURR, Jean-Michel, 2024. Les revues par les pairs : un levier au service de la statistique publique. In : *Courrier des statistiques*. [en ligne]. 16 décembre 2024. Insee. N° N12, pp. 137-158. [Consulté 07 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8264560?sommaire=8264562>.
- CHARPIN, Jean-Michel, 2007. L'évolution du rôle de coordination de l'INS au niveau national dans un système statistique européen renforcé. In : *Courrier des statistiques*. [en ligne]. Décembre 2007. Insee. N° 121-122, pp.5-9. [Consulté le 22 mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05461057v1>.
- COMMISSION EUROPÉENNE, 2010. Rapport sur les statistiques du déficit et de la dette publics de la Grèce. In : *Rapport de la Commission européenne*. [en ligne]. 08 janvier 2010. [Consulté le 07 janvier 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A52010DC0001>.
- COUR DES COMPTES EUROPÉENNE, 2012. La Commission et Eurostat ont-ils amélioré le processus de production de statistiques européennes de manière à en renforcer la crédibilité et la fiabilité ? In : *Rapport spécial n° 12*. [en ligne]. 21 septembre 2012. [Consulté le 07 avril 2026]. Disponible à l'adresse : https://www.eca.europa.eu/Lists/ECADocuments/SR12_12/SR12_12_FR.PDF.
- DE MICHELIS, Alberto et CHANTRAINE, Alain, 2003. « Mémoires d'Eurostat – Cinquante ans au service de l'Europe ». In : *site d'Eurostat*. [en ligne]. 15 mai 2003. Eurostat. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-49-02-183>.
- EUROSTAT, 2003. The Future of the European Statistical System – 88th DGINS Conference. In : *site d'Eurostat*. [en ligne]. 19-20 septembre 2002. 88^e conférence des directeurs généraux des instituts nationaux de statistique (DGINS). [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/ks-51-03-293n>.

- EUROSTAT, 2004. Rapport d'Eurostat sur la révision des données grecques relatives au déficit et à la dette publics (en anglais). In : *site d'Eurostat*. [en ligne]. 1^{er} décembre 2004. Emerald Publishing. [Consulté le 08 avril 2026]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/greece>.
- FRANCHET, Yves, 2004. Introduction of TQM in Eurostat, a supranational statistical service. In : *The TQM Magazine*. [en ligne]. 1^{er} octobre 2004. Vol. 16, n° 5, pp. 341-346. [Consulté le 28 mai 2026]. Disponible à l'adresse : <https://doi.org/10.1108/09544780410551278>.
- FRANCHET, Yves, 2014. Développement de la statistique de l'Union européenne : une action coopérative multinationale en réseau. In : *Statistique et société*. [en ligne]. Janvier 2014. Vol. 2, n° 1, pp. 15-19. [Consulté le 08 avril 2026]. Disponible à l'adresse : https://www.persee.fr/doc/staso_2269-0271_2014_num_2_1_889.
- MADINIER, Chantal, LHÉRITIER, Jean-Louis et PUIG, Jean-Pierre, 2007. La mise en place d'un Code de bonnes pratiques, pilier de la nouvelle gouvernance statistique européenne. In : *Courrier des statistiques*. [en ligne]. Mai-décembre 2007. Insee. N° 121-122, pp. 11-17. [Consulté le 08 avril 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05461108v1>.
- MORDANT, Guillaume, 2009. La nouvelle loi statistique européenne. In : *Courrier des Statistiques*. [en ligne]. Décembre 2009. Insee. N° 128, pp. 25-32. [Consulté le 22 mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05393066v1>.
- PIFFETEAU, Hervé, 2019. Un nouveau triptyque juridique pour les statistiques européennes. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 93-109. [Consulté le 08 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254229?sommaire=4254170>.
- ROUAULT, Dominique, 2007. L'indépendance professionnelle dans les textes fondamentaux des systèmes statistiques européens. In : *Courrier des Statistiques*. [en ligne]. Décembre 2007. Insee. N° 121-122, pp. 27-36. [Consulté le 22 mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05462125v1>.
- SÉNAT, 2010. Le système statistique européen à l'épreuve de la crise grecque : renforcer la légitimité d'Eurostat. In : *site du Sénat*. [en ligne]. 31 mars 2010. Rapport d'information n° 374 (2009-2010) de M. Jean Arthuis, fait au nom de la Commission des finances. [Consulté le 07 avril 2026]. Disponible à l'adresse : <https://www.senat.fr/rap/r09-374/r09-3740.html>.

Une source harmonisée pour mesurer le tourisme européen

Les données de plateformes collectées par Eurostat



David Levy*

Depuis dix ans, l'essor des plateformes numériques a profondément transformé le marché de l'hébergement touristique, rendant incomplète la vision offerte par les seules enquêtes traditionnelles. Si la statistique publique mesure de longue date la fréquentation de l'hôtellerie classique, elle peinait jusqu'alors à quantifier les locations de courte durée entre particuliers, un segment pourtant devenu majeur. Pour combler cet angle mort, une coopération inédite a été initiée en 2019 entre Eurostat et quatre acteurs mondiaux (Airbnb, Booking, Expedia et TripAdvisor).

Ce dispositif repose sur une gouvernance inversée : les plateformes transmettent directement à Eurostat des microdonnées standardisées sur les nuitées effectivement réalisées. Contrairement au *web scraping*, cette source mesure l'activité réelle et non l'offre théorique. L'intégration de ces données permet désormais un suivi mensuel et une analyse géographique fine, révélant la dynamique touristique des zones rurales et littorales peu dotées en hôtellerie.

Toutefois, ce partenariat public-privé soulève des défis structurels : dépendance aux accords volontaires, opacité relative des algorithmes de gestion des plateformes et absence d'identifiant unique pour traiter les doublons de capacité d'hébergement. Le passage de ces données au statut de statistique européenne officielle marquera une étape clé dans l'hybridation des sources, entre enquêtes classiques et données massives détenues par le secteur privé.

 For a decade now, the rise of digital platforms has profoundly transformed the tourist housing market, making partial the vision offered by traditional surveys only. Official statistics have long been measuring the number of people staying in conventional hotels, but it has struggled to quantify short-term rental between private individuals, which has become a major segment on the tourist housing market. To make up for this blind spot, an unprecedented collaboration started in 2019 between Eurostat and four global players (Airbnb, Booking, Expedia and TripAdvisor).

This plan is based on an inverted organisation: these platforms directly provide Eurostat with standardised micro-data on nights actually spent. Unlike web scraping, this method measures actual activity and not theoretical supply. The integration of this data now allows for a monthly monitoring and a sharp geographic analysis, disclosing the dynamics of tourism in rural and coastal areas with fewer hotels.

However, this partnership between the public and private sectors raises structural challenges: the dependence on voluntary agreements, the relative opacity of platform management algorithms and the absence of a unique ID to handle accommodation capacity duplicates. The recognition of this data as official European statistics will mark a key moment in the hybridisation of sources, between traditional surveys and data massively owned by the private sector.

* Responsable national Insee pour le tourisme, Insee.
david.levy@insee.fr

Au cours des dix dernières années, les locations touristiques entre particuliers ont connu une croissance soutenue dans la plupart des pays européens. Leur essor s'explique à la fois par celui des plateformes numériques de réservation (mettant en relation offreurs et demandeurs via un site web), par l'attractivité des hébergements (offrant davantage de souplesse que les dispositifs traditionnels), mais aussi par le développement d'un tourisme moins concentré sur les zones d'hôtellerie classique. Par exemple, les touristes venant visiter Paris peuvent séjourner dans des quartiers moins centraux et plus résidentiels comme le 19^e arrondissement, voire en proche



Face à l'essor des locations de courte durée entre particuliers, la statistique publique a rencontré une difficulté croissante pour maintenir une vision complète du secteur de l'hébergement touristique.



banlieue comme à Montreuil. Ou encore, dans des territoires ruraux comme l'Ardèche ou le Massif central, ces locations permettent de séjourner dans des communes peu dotées en hôtels. Face à cette transformation profonde du marché de l'hébergement, la statistique publique a rencontré une difficulté croissante pour maintenir une vision complète du secteur.

Les enquêtes des instituts nationaux de statistique (INS) mesurent de longue date le tourisme sous plusieurs angles : enquêtes auprès des hébergeurs, enquêtes auprès des ménages, suivis administratifs des établissements classés¹.

Si ces outils permettent de suivre les hébergements collectifs (hôtels, campings, villages de vacances, etc.), ils ne captent qu'imparfaitement la location de courte durée entre particuliers. Dans un premier temps, des initiatives nationales ont cherché à compléter le champ, notamment via du *web scraping*² ou des fichiers communaux d'enregistrement (voir supra). Cependant, ces approches fragmentées restaient difficilement comparables entre territoires.

Pour répondre à cette lacune, Eurostat³ a initié dès 2019 une coopération avec plusieurs grandes plateformes de réservation. Celle-ci repose sur la transmission régulière de microdonnées à Eurostat, qui se charge ensuite de les harmoniser, puis de les agréger. Les premières séries ont été publiées en 2020, couvrant l'ensemble des États membres⁴. Cette initiative représente une évolution majeure : pour la première fois, une source statistique européenne provient directement d'acteurs privés opérant sur l'ensemble des États membres.

Cet article présente cette source nouvelle et décrit comment Eurostat en garantit la qualité statistique. Il détaille les informations transmises, les traitements effectués, les apports pour la connaissance du tourisme et les limites liées à la nature même de la donnée privée. Enfin, il discute les enjeux plus larges pour la statistique publique à l'heure où les données détenues par les entreprises prennent une place croissante dans la production d'indicateurs économiques et sociaux.

¹ Le Courrier des statistiques en est le témoin : voir par exemple Viennet (1977). Sur les enjeux sociaux de la mesure, voir Cousin et Réau (2016).

² Le *web scraping*, ou moissonnage en français, est une technique d'extraction automatisée de données de sites web.

³ Eurostat est l'institut statistique communautaire, direction générale de la Commission européenne.

⁴ Voir Ulrich (2021) pour une exploitation de ces données par la France et Lesur (2025) pour un panorama plus large des sources de données privées explorées en France par la statistique publique.

► Les statistiques « traditionnelles » du tourisme

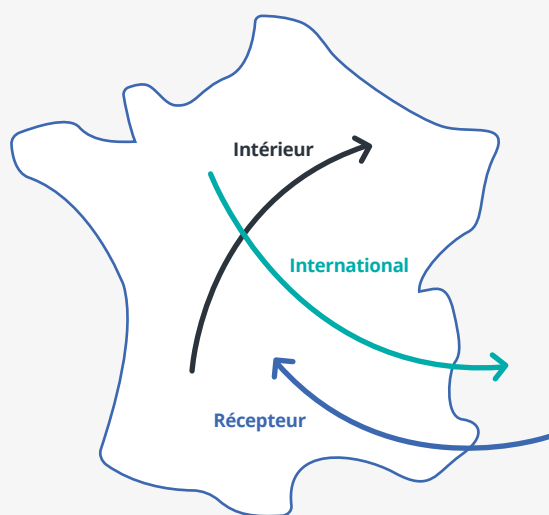
La définition internationale du tourisme, établie par l'ONU Tourisme⁵, est la suivante :

« Le tourisme est un phénomène social, culturel et économique qui suppose des mouvements de personnes vers des pays ou des lieux situés en dehors de leur environnement habituel intervenant pour des motifs personnels ou pour affaires et motifs professionnels. Ces personnes sont appelées visiteurs et peuvent être des touristes ou des excursionnistes ; des résidents ou des non-résidents ; le tourisme englobe leurs activités, dont certaines impliquent des dépenses touristiques »⁶.

L'ONU Tourisme précise par ailleurs que sont exclus du champ, d'une part, les déplacements des professionnels du transport, d'autre part, tous les déplacements dont le motif principal est d'être employé sur le lieu visité (saisonniers, frontaliers, stagiaires, etc.).

Pour avoir une représentation aussi complète que possible du champ du tourisme, les statisticiens publics ont mis en place plusieurs dispositifs de collecte d'information visant à appréhender les trois aspects du tourisme : le tourisme intérieur (au sein d'un pays),

► **Figure 1 - Le champ du tourisme : exemple de la France**



- Tourisme intérieur : résidents français voyageant en France
- Tourisme international : résidents français voyageant à l'étranger
- Tourisme récepteur : non-résidents voyageant en France

⁵ Institution spécialisée de l'Organisation des Nations Unies (ONU), appelée aussi Organisation mondiale du tourisme (OMT). Voir le lien : <https://www.untourism.int/fr/a-propos-onu-tourisme>.

⁶ Voir le glossaire de l'ONU Tourisme : <https://www.untourism.int/fr/glossaire-de-tourisme>. Cette définition est issue du rapport de l'ONU établissant les recommandations internationales pour les statistiques sur le tourisme (ONU, 2008). Les touristes se distinguent des excursionnistes par le fait qu'ils séjournent au moins une nuit sur place. Les résidents sont entendus au sens national (résidents du pays).



Les statisticiens publics ont mis en place plusieurs dispositifs de collecte d'information pour appréhender le tourisme intérieur, le tourisme international et le tourisme récepteur.



le tourisme international (vers un autre pays) et le tourisme récepteur (depuis un autre pays) (**figure 1**). Ainsi, la France gère plusieurs processus de collecte auprès des hébergeurs et des ménages dont l'Insee coordonne l'ensemble afin d'assurer une harmonisation des concepts (**figure 2**).

Les principales enquêtes sont réalisées en respectant un cadre européen harmonisé et servent à répondre aux obligations européennes en matière de statistiques sur le tourisme (**encadré 1**). En particulier, l'Insee réalise chaque mois :

- l'**enquête de fréquentation dans les hébergements collectifs de tourisme**⁷ auprès des hôtels, campings et autres hébergements collectifs de tourisme : il s'agit du dispositif le plus ancien concernant le tourisme dont les premières collectes remontent au début des années 1980 ;
- l'**enquête Voyages professionnels et personnels**⁸ auprès des ménages résidents. Cette enquête remplace l'ancien dispositif de Suivi de la demande touristique (SDT) depuis fin 2025. Elle interroge les personnes sur leurs déplacements touristiques, que ce soit à l'occasion de vacances, week-ends ou séjours professionnels. Elle s'intéresse aux caractéristiques de ces voyages, telles que les moyens de transport utilisés, le type d'hébergement, les dépenses.

De manière symétrique, la Banque de France réalise usuellement une enquête auprès des visiteurs internationaux venant en France : l'enquête auprès des visiteurs étrangers (EVE). Interrompue depuis 2023, cette enquête est en cours de rénovation. Elle sert

► **Encadré 1. Obligations européennes en matière de statistiques du tourisme**

Le cadre juridique européen (notamment le règlement (UE) n° 692/2011 concernant les statistiques européennes sur le tourisme*) impose aux États membres de produire et transmettre à Eurostat des données harmonisées sur le tourisme. Les obligations portent sur deux volets principaux : la **demande touristique** (ménages résidents) et l'**offre d'hébergement collectif touristique** (hôtels, campings, etc.).

Sur le premier volet, le règlement impose une collecte annuelle ou triennale d'informations sur :

- le nombre de voyages des ménages résidents (courts et longs séjours) ;
- leurs destinations (nationales et internationales) ;
- les motifs (personnels, pour affaires) ;

- la durée et le type d'hébergement ;
- les dépenses touristiques.

Sur le second volet, le règlement impose une collecte mensuelle auprès des établissements nationaux sur :

- leur capacité d'hébergement (établissements, chambres, lits) ;
- les nombres d'arrivées et de nuitées ;
- la ventilation de ces dernières par pays de résidence et par région.

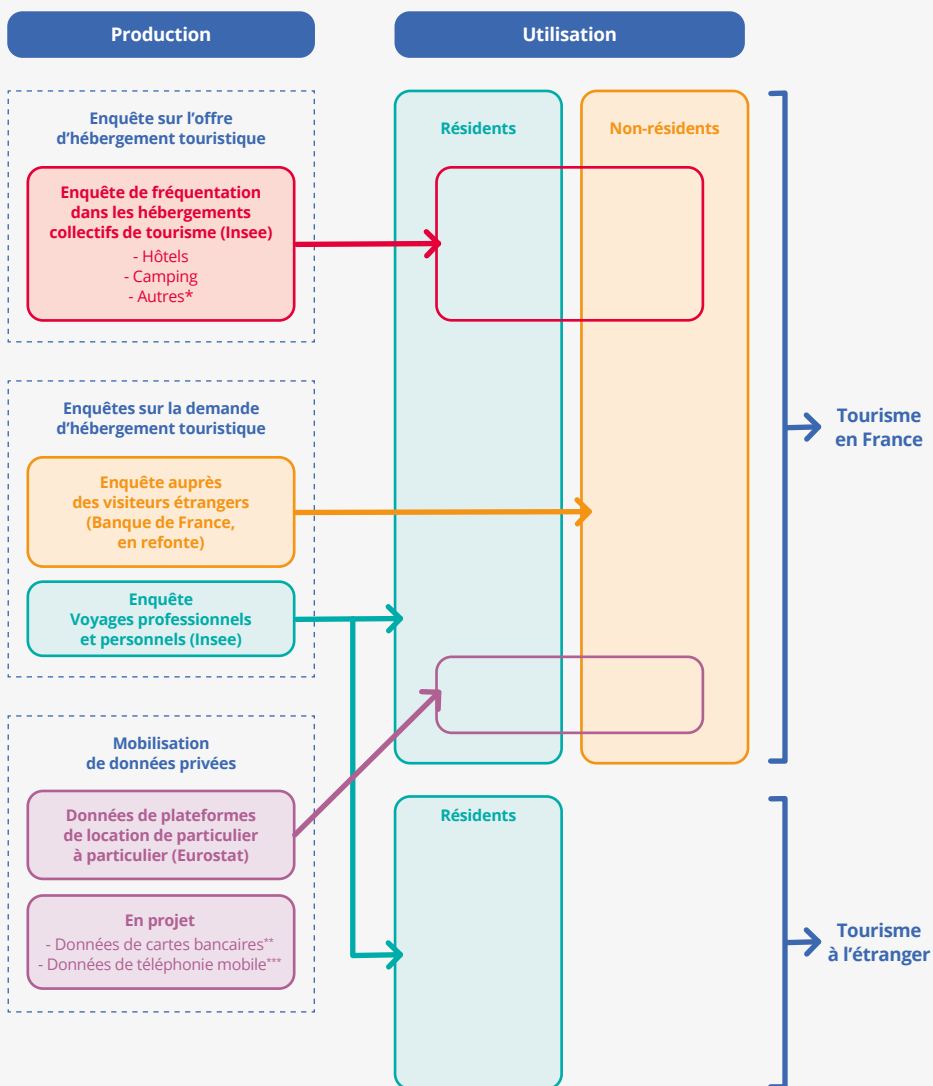
Les États doivent respecter des définitions harmonisées, garantir la qualité statistique (comparabilité, fiabilité, actualité) et transmettre les données dans des délais et formats normalisés. L'ensemble des statistiques produites en France visent à répondre à ce règlement.

* Voir les références juridiques en fin d'article.

⁷ Voir le lien : <https://www.insee.fr/fr/metadonnees/source/serie/s1039>.

⁸ Voir le lien : <https://www.insee.fr/fr/information/6965150>.

► **Figure 2 - Les sources statistiques sur le tourisme et leur utilisation**



* Autres hébergements collectifs de tourisme (auberges collectives, centres internationaux de séjour, centres sportifs, résidences de tourisme et résidences hôtelières, maisons familiales de vacances et villages de vacances).

** Voir Boittelle et al. (2025) : <https://www.insee.fr/fr/information/8546947?sommaire=8546949>.

*** Voir Joubert et al. (2025) : <https://www.insee.fr/fr/information/8546945?sommaire=8546949>.

Note : Les enquêtes de fréquentation auprès de l'offre d'hébergements collectifs et l'enquête Voyages professionnels et personnels répondent au règlement européen sur les statistiques du tourisme (règlement (UE) n° 692/2011, voir les références juridiques en fin d'article).

Les enquêtes auprès des ménages et individus (enquête Voyages professionnels et personnels et enquête auprès des visiteurs étrangers) sont utilisées pour calculer le compte satellite du tourisme, mesure macroéconomique des effets du tourisme.

Lecture : Les enquêtes de fréquentation dans les hébergements collectifs de tourisme renseignent sur une partie du tourisme en France (résidents français et non-résidents), celle dans les hébergements collectifs.

L'enquête Voyages professionnels et personnels renseigne sur l'ensemble du tourisme des résidents français, que ce soit en France ou à l'étranger, dans tous les types d'hébergement.

normalement à estimer les dépenses des touristes étrangers afin d'alimenter la ligne « voyages » de la balance des paiements des comptes nationaux⁹.

Enfin, à côté de ces grandes opérations, sont menées des enquêtes sur des segments particuliers qui ne peuvent être couverts par une enquête nationale : enquêtes dans les aéroports, sur la fréquentation des lieux de culture ou encore sur la fréquentation des lieux de mémoire.

Le socle de la statistique touristique repose sur un dispositif permettant de quantifier le phénomène sous toutes ses dimensions. D'un côté, les enquêtes de fréquentation auprès des hébergements collectifs fournissent une mesure précise de l'offre marchande classique en comptabilisant les nuitées et le taux d'occupation ; elles apportent en outre une dimension conjoncturelle à l'activité du tourisme en France. De l'autre, les enquêtes auprès des ménages captent la demande. Ces outils permettent non seulement de dénombrer les touristes et les excursionnistes, mais aussi d'analyser les comportements : motifs du séjour (agrément, affaires, visites familiales), modes de transport et types d'hébergement, y compris non marchands.

Au-delà de ces indicateurs de flux, ces sources sont les piliers du **compte satellite du tourisme (CST)**¹⁰. Ce cadre macroéconomique, cohérent avec la comptabilité nationale, permet d'évaluer la contribution du secteur au produit intérieur brut (PIB). En croisant les dépenses déclarées par les voyageurs (transport, restauration, activités culturelles) et les dépenses de consommation finale des ménages, le CST isole la « consommation touristique intérieure ». Cette approche monétaire est indispensable pour mesurer les retombées économiques réelles du tourisme intérieur, y compris sur le plan des emplois induits.

► Les locations entre particuliers : un segment en pleine expansion mal couvert par les statistiques standards

Les locations de courte durée entre particuliers constituent l'un des segments les plus dynamiques du tourisme européen. Dans plusieurs États membres, elles représentent entre 15 % et 25 % des nuitées marchandes, et jusqu'à 35 % dans certaines zones très touristiques. C'est le cas notamment de la côte méditerranéenne (Alpes-Maritimes en France, Malaga en Espagne). La diffusion des plateformes a facilité l'accès à une offre très dispersée : logements de résidences secondaires, hébergements occasionnels, locations très saisonnières, habitat diffus dans des communes de petite taille (Devaux, 2024).

Ce marché, par nature éclaté et peu structuré, échappait largement aux dispositifs statistiques classiques. Les enquêtes auprès des hébergeurs se concentrent sur les établissements déclarés et structurés (hôtels, campings, résidences touristiques). Les enquêtes auprès des ménages fournissent des informations essentielles sur les pratiques touristiques, mais ne permettent pas un suivi mensuel ni une analyse géographiquement fine. Quant aux fichiers de régulation locale qui peuvent exister dans certains pays

⁹ Dans le cadre de la comptabilité nationale, la balance des paiements mesure les transactions entre les résidents et les non-résidents. Parmi celles-ci, les transactions liées aux voyages correspondent aux achats de biens et services effectués, d'une part par les non-résidents voyageant en France (qui constituent des « recettes » pour la France), d'autre part par les résidents voyageant à l'étranger (qui constituent des « dépenses » pour la France).

¹⁰ Voir le lien : <https://www.insee.fr/fr/statistiques/7458178>.



Les locations de courte durée entre particuliers constituent l'un des segments les plus dynamiques du tourisme européen.



comme en Espagne (régional), en Italie (régional), en Allemagne (municipal), aux Pays-Bas (municipal) ou au Portugal (registre national avec déclinaison locale), ils reflètent les politiques propres à chaque territoire : ils ne peuvent être agrégés au niveau européen en raison de la diversité des informations et des niveaux géographiques disponibles.

Cette absence de vision harmonisée représentait un enjeu important pour les décideurs publics, les acteurs du tourisme et les collectivités territoriales. Ces derniers peinaient en effet à quantifier l'impact réel des locations entre particuliers sur leurs territoires, tant sur le plan économique que sur l'évolution de l'offre d'hébergement.

► Des expériences locales infructueuses

Plusieurs pays ont tenté d'améliorer leur connaissance du phénomène. Des opérations de *web scraping* ont été menées afin de collecter automatiquement les données visibles sur les plateformes : nombre d'annonces, caractéristiques des logements, prix affichés. D'autres initiatives reposaient sur des fichiers municipaux d'enregistrement des meublés, obligatoires dans certaines grandes villes touristiques comme à Berlin, Munich ou Amsterdam. Ces approches ont produit des éclairages intéressants mais restaient limitées. En particulier, le *web scraping*, tel qu'il était pratiqué, ne pouvait collecter que des informations sur les annonces visibles sur les sites des plateformes de location, mais ni sur leur concrétisation (location effective), ni sur le nombre de nuitées concernées. S'ajoutait à ces difficultés le coût important de maintenance des programmes informatiques pour s'adapter en permanence aux évolutions des sites web.

► Une démarche européenne commune

À la suite de ces opérations, Eurostat a donc entrepris de structurer une démarche commune, permettant de disposer d'une vision cohérente et comparable à l'échelle des États membres. Cette démarche répondait à des demandes des pays membres réunis en groupe de travail.

Après des premières négociations infructueuses avec Airbnb en 2016, le projet a pris un nouveau départ fin 2018 sous la forme d'une initiative conjointe de deux directions générales de la Commission européenne : celle du marché intérieur, de l'industrie, de l'entrepreneuriat et des petites et moyennes entreprises (GROW) et celle des statistiques (Eurostat). Dès le début, l'approche a été multidisciplinaire, réunissant non seulement des *data scientists*¹¹ et des statisticiens, mais aussi (des deux côtés) des experts juridiques et contractuels, des experts en informatique et des experts en politiques publiques. Un autre facteur de réussite essentiel a été l'inclusion de plusieurs entreprises détenant ce type de données.

¹¹ Voir l'article de Galiana et Lefebvre sur l'apport des métiers de la *data* à la statistique publique dans ce même numéro.

L'élargissement à quatre plateformes (Airbnb, Booking, Expedia, Tripadvisor) a non seulement contribué à une meilleure représentativité et réduit la dépendance exclusive à un seul acteur, mais a également créé des conditions de concurrence équitables pour les différents acteurs de ce segment du marché de l'hébergement.

Ces quatre acteurs couvrent une très grande partie du marché européen des locations de courte durée. Leur coopération a permis d'élargir la base de soutien au projet et a finalement conduit à la conclusion d'accords de non-divulgaration signés entre les quatre plateformes et la Commission européenne en mars 2020. Ces accords portent principalement sur les grands principes de l'échange de données et sur les mesures de sécurité à prendre en matière de transmission, de traitement, de stockage, de partage et de diffusion des données.

Eurostat agit comme un hub¹² central pour la collecte des données, inversant la gouvernance traditionnelle du système statistique européen (SSE)¹³. Ce modèle inversé est une petite révolution culturelle pour les INS, habitués à maîtriser leur chaîne de production de A à Z, puis à transmettre les résultats à Eurostat. Pour accompagner

l'institut statistique européen dans le projet, un groupe de pilotage a été constitué, composé d'experts de cinq pays (Espagne, France, Lettonie, Slovaquie et Finlande).

Comme c'est le cas pour les données administratives, les données détenues par les plateformes n'ont pas été conçues à l'origine à des fins statistiques.

Comme c'est le cas pour les données administratives (ou les autres données de source privée mobilisées par la statistique publique), les données détenues par les plateformes n'ont pas été conçues à l'origine à des fins statistiques : elles sont issues des systèmes de gestion des différents opérateurs¹⁴. De ce fait, une partie importante des discussions techniques avec les plateformes a porté sur l'alignement de leurs catégories d'hébergement (non harmonisées) sur

les concepts et définitions existants pour les statistiques du tourisme. En pratique, les données extraites des systèmes de gestion sont transmises directement à Eurostat dans le cadre d'accords de confidentialité, sous forme de données anonymisées. Eurostat se charge ensuite de les harmoniser, puis de les agréger. Il les diffuse ensuite exclusivement sous forme d'agrégats statistiques.

Contrairement aux informations issues du *web scraping* ou de fichiers administratifs locaux, ces données ne décrivent pas l'offre théorique d'hébergement, mais l'activité effectivement réalisée : seules les locations ayant donné lieu à au moins une nuitée sont incluses. Cette convention méthodologique constitue un choix central, visant à mesurer l'occupation réelle plutôt que le stock potentiel de logements proposés.

Le champ sectoriel est par ailleurs strictement limité aux hébergements relevant du groupe 55.2 de la nomenclature européenne d'activité NACE¹⁵ (« hébergement

¹² Ou « concentrateur » : appareil relié à plusieurs machines en réseau, et permettant de concentrer les données pour les transmettre par un unique canal (source : Le Robert).

¹³ Sur le SSE, voir l'article de Bischoff et Piffeteau dans ce même numéro.

¹⁴ Au sujet du traitement plus général de ce type de données par la statistique publique, voir Lamarche et Rivière (2025).

¹⁵ NACE pour nomenclature statistique commune des activités économiques dans la Communauté européenne. Voir à ce sujet l'article de Masson sur la révision des nomenclatures d'activité dans ce même numéro.

touristique et autre hébergement de courte durée »). Il s'agit du groupe dans lequel les plateformes de l'économie collaborative sont principalement actives. Les hôtels et autres hébergements collectifs traditionnels, qui relèvent d'autres groupes de la NACE, sont exclus du dispositif pour éviter les doublons avec les enquêtes de fréquentation réalisées par l'ensemble des pays européens.

Enfin, la couverture géographique concerne l'ensemble des États membres de l'Union européenne ainsi que plusieurs pays de l'Association européenne de libre-échange (AELE)¹⁶. Les données sont transmises mensuellement, avec un délai d'environ 45 jours après la fin de la période de référence. Les premières séries exploitables couvrent les années 2018 et 2019, puis sont mises à jour de façon trimestrielle.

► Les informations transmises par les plateformes

Définir les données nécessaires à la statistique publique a nécessité de nombreuses discussions avec les plateformes. Tout l'enjeu repose sur le passage de la donnée comptable à la donnée statistique. Trois jeux de données distincts ont été définis par Eurostat en concertation avec les plateformes afin de coller au mieux aux fichiers administratifs de l'ensemble des acteurs, représentant en quelque sorte leur dénominateur commun.

Définir les données nécessaires à la statistique publique a nécessité de nombreuses discussions avec les plateformes.

Le premier fichier concerne la capacité d'hébergement. Transmis annuellement, il décrit le nombre d'hôtes, de logements et de lits proposés. Cette information est essentielle pour connaître les tensions qu'il peut y avoir sur le marché de la location. Dans les enquêtes standards, l'indicateur privilégié pour résumer cette information est le taux d'occupation. Il s'agit du rapport entre le nombre de chambres occupées et offertes ou, pour les campings, entre le nombre d'emplacements occupés et offerts. Malheureusement, dans la phase actuelle du projet sur les données de plateformes,

seules les informations relatives à l'occupation sont diffusées. Celles sur l'offre ne le sont pas en raison de problèmes de doublons entre plateformes, certains logements étant proposés à la location sur plusieurs sites.

Les deux autres fichiers concernent l'activité de location. Transmis trimestriellement, ils décrivent cette activité en fonction respectivement des caractéristiques des logements loués ou de celles des voyageurs, selon des critères très peu identifiants (voir infra).

Les variables centrales utilisées dans les analyses (communes aux deux fichiers) sont :

- le nombre de séjours, assimilable à un « proxy »¹⁷ des arrivées touristiques ;
- le nombre de nuits louées ;

¹⁶ L'AELE compte actuellement quatre pays membres : l'Islande, le Liechtenstein, la Norvège et la Suisse.

¹⁷ Variable de substitution pour éclairer le phénomène.

- le nombre de nuitées, calculé comme le nombre de nuits louées pondéré par le nombre de voyageurs hébergés (cette variable est comparable à celle utilisée dans les enquêtes de fréquentation).

Ces indicateurs sont ventilés selon plusieurs dimensions : mois de référence, pays et région (selon la nomenclature européenne NUTS¹⁸), type de logement (entier ou partagé), type d'hôte (particulier ou professionnel), taille du logement, et origine géographique des voyageurs (résidents / non-résidents ou pays de résidence détaillé).

Les plateformes ne transmettent aucune donnée permettant d'identifier individuellement un logement, un hôte ou un voyageur. Les adresses exactes, les prix, les revenus, les motifs de séjour, ainsi que les caractéristiques sociodémographiques des voyageurs sont exclus. Ce périmètre restreint garantit la confidentialité des acteurs et recentre la source sur la mesure de l'activité touristique observée.

► L'insertion des données de plateformes dans le dispositif actuel de production

Il est essentiel d'articuler les données issues des plateformes avec les statistiques traditionnelles produites à partir des enquêtes auprès des ménages et des hébergeurs. Cette mise en cohérence permet notamment de disposer d'indicateurs véritablement comparables, au premier rang desquels la nuitée, unité centrale des statistiques de fréquentation. En mobilisant des concepts harmonisés – séjour, nuit louée, nuitée – (*encadré 2*), il devient possible d'aligner la mesure de l'occupation des locations de courte durée sur celle des hôtels ou campings, et ainsi d'intégrer ce segment dans les agrégats globaux du tourisme. Les données de plateformes apportent une mesure fine et mensuelle



Les données de plateformes apportent une mesure fine et mensuelle de l'activité observée, tandis que les enquêtes fournissent des informations essentielles sur les profils, les motifs et les dépenses.



de l'activité observée, tandis que les enquêtes fournissent des informations essentielles sur les profils, les motifs et les dépenses. Leur articulation garantit la cohérence des niveaux, renforce la comparabilité internationale et permet d'obtenir une vision complète et robuste des dynamiques touristiques.

Dans le cas des données de plateformes, l'unité statistique retenue est le logement effectivement loué au moins une fois durant la période de référence. Un **logement** correspond à une annonce active ayant généré au moins une nuitée. Ainsi, un même hôte – particulier ou professionnel (y compris une conciergerie) – peut apparaître plusieurs fois dans les données s'il loue plusieurs logements distincts. La plateforme transmet pour chaque logement les nuits louées et le nombre de personnes hébergées.

¹⁸ Nomenclature des unités territoriales statistiques. De manière générale et pour la France en particulier, voir les liens : <https://ec.europa.eu/eurostat/fr/web/nuts/overview> et <https://www.insee.fr/fr/metadonnees/definition/c2112>.

La **nuitée** correspond à l’occupation effective d’un logement par un voyageur pour une nuit, conformément aux standards internationaux des statistiques du tourisme. Le nombre de nuitées est calculé à partir des informations de réservation, en tenant compte du nombre de personnes déclarées lors de la réservation. Cet indicateur permet de comparer directement l’intensité de fréquentation entre territoires et périodes, indépendamment de la structure de l’offre. Il permet en outre de positionner cette source à côté des autres statistiques de fréquentation.

Les données permettent de reconstituer les **séjours**, définis comme des suites continues de nuitées dans un même logement. La durée des séjours constitue une information clé pour analyser les usages touristiques : les séjours courts dominent dans les grandes agglomérations, tandis que les séjours plus longs sont plus fréquents dans les zones littorales ou rurales. On retrouve également cette notion dans les enquêtes de fréquentation, il est donc possible de comparer les durées de séjour dans les hôtels ou campings à celles dans les locations entre particuliers.

L’**origine des voyageurs** est mesurée à partir du pays de résidence déclaré par le locataire. Elle permet de distinguer les flux domestiques des flux internationaux, et de décrire finement la structure de la demande touristique. Selon les niveaux de diffusion, cette information est disponible sous une forme détaillée (pays ou grandes zones du monde) ou agrégée (résidents / non-résidents).

Quant à la **localisation des logements**, comme évoqué, les plateformes ne transmettent aucune adresse. Les logements sont géocodés à partir de leurs coordonnées internes, puis rattachés par Eurostat à des unités administratives (NUTS 2 ou NUTS 3). Ce choix assure un compromis entre finesse territoriale et confidentialité.

► **Encadré 2. Définitions clés**

- Logement** : unité ayant eu au moins une nuitée réalisée.
Nuitée : occupation effective du logement par un voyageur pour une nuit.
Séjour : suite de nuitées continues dans le même logement.

Indicateur	Concept Statistique	Unité de mesure	Ce qu’il mesure
Séjour	Flux d’arrivées	Arrivée	Le nombre de réservations
Nuit louée	Occupation du logement	Nuit	La durée d’immobilisation du bien
Nuitée	Fréquentation réelle	Personne-nuit	L’intensité touristique (nuits × voyageurs)

Exemple de calcul : si une famille de trois personnes loue un logement entier pour trois nuits, les statistiques de fréquentation touristique sont : 1 logement loué, 1 séjour, 3 nuits louées et 9 nuitées.

► Le rôle inédit d'Eurostat dans la production de données harmonisées

Contrairement à la majorité des statistiques européennes, les données de plateformes sont envoyées directement à Eurostat qui les diffuse après traitements aux INS. Cette organisation a été retenue afin de garantir l'harmonisation des concepts, des formats et des traitements sur l'ensemble du territoire européen.

Comme évoqué, ce modèle inversé confère un rôle inédit à Eurostat : celui de **producteur de données primaires en lien direct avec des acteurs privés internationaux**. Il confère également un nouveau rôle aux INS, qui maîtrisent habituellement la chaîne de production des statistiques. Cette fois, ils ne sont pas producteurs, mais valideurs. Cette particularité se manifeste par l'enchaînement de deux boucles de validation des données : l'une concerne Eurostat et les plateformes, l'autre Eurostat et les INS (*figure 3*).

Eurostat applique une chaîne complète de contrôles avant toute diffusion :

- des contrôles techniques de format et de complétude ;
- des contrôles de cohérence interne (rapports entre les nombres de logements, de séjours et de nuitées) ;
- des contrôles de continuité temporelle, visant à détecter des ruptures ou des évolutions atypiques (pouvant relever, par exemple, d'un décalage comptable) ;
- des contrôles de cohérence externe reposant sur la comparaison entre les résultats des différentes plateformes.

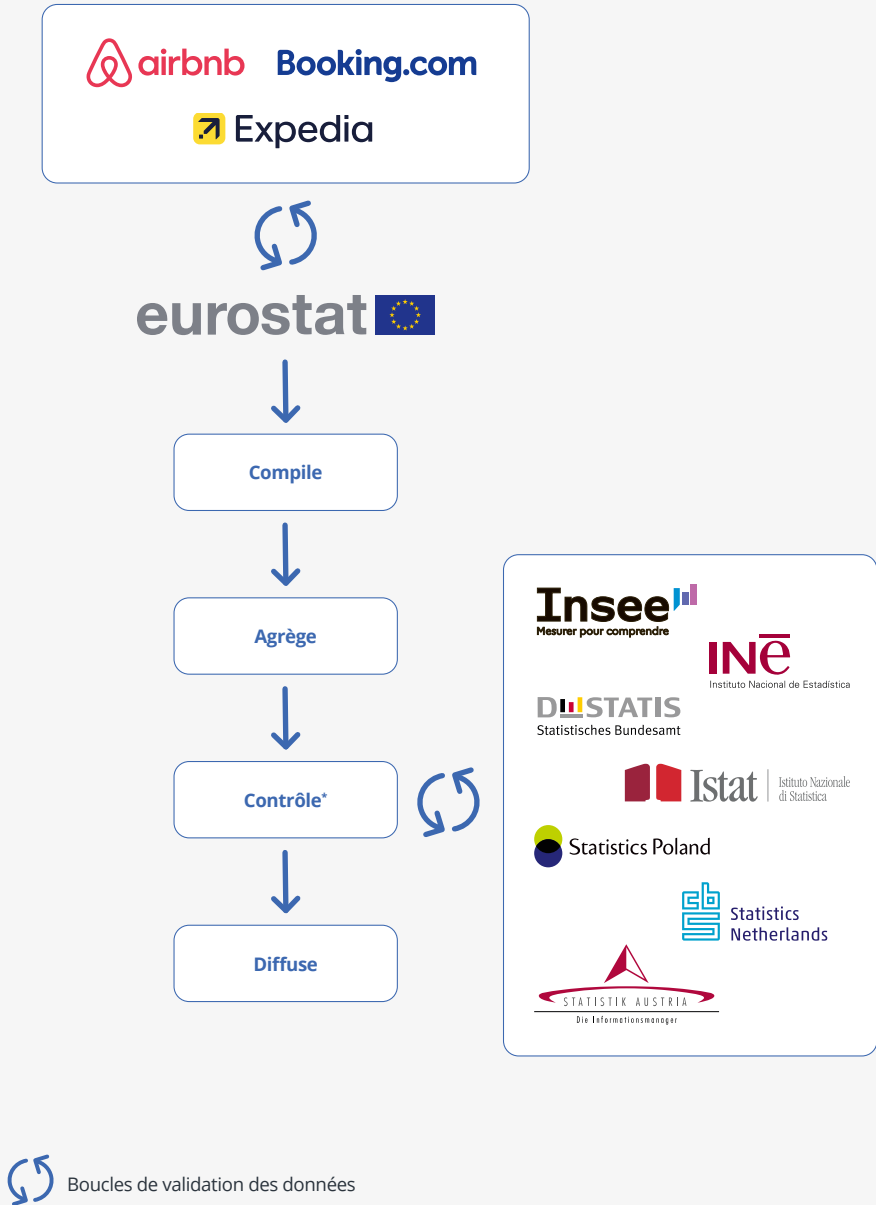
Ces vérifications sont réalisées à l'aide de règles automatisées, complétées par des échanges réguliers avec les plateformes en cas d'anomalie détectée.

La diffusion est encadrée par des règles strictes et des échanges avec les INS. Au niveau le plus agrégé, la diffusion recouvre les pays et les NUTS 2 (qui correspondent en France aux régions). À un niveau plus fin, la diffusion respecte des seuils minimaux d'activité comme le nombre de locations ou le nombre de touristes. Sont concernées les NUTS 3 (qui correspondent en France aux départements) et certaines agglomérations. Plus précisément, certaines données sont diffusées au niveau des *cities*¹⁹, concept d'Eurostat, qui correspond aux grandes communes et à leur banlieue. En France, sur les 110 *cities*, des données issues des plateformes sont disponibles pour 50 à 70 d'entre elles, selon les années et le détail retenu. En dehors de ces niveaux géographiques, aucune diffusion n'est effectuée à un niveau plus fin.

Les INS sont associés à la phase de validation finale, quelques jours avant la diffusion, afin de repérer d'éventuelles anomalies.

¹⁹ Voir le lien : <https://www.insee.fr/fr/information/3642744>.

► **Figure 3 - Le processus de compilation des données de plateformes**



* Cette étape de contrôle en lien avec les instituts nationaux de statistique peut conduire Eurostat à revenir vers les plateformes pour effectuer des révisions.
Lecture : Les données de plateformes sont directement envoyées à Eurostat, qui les traite puis les soumet au contrôle des instituts nationaux de statistique, avant de les diffuser.

► Quels apports pour la statistique publique ?

L'apport principal des données issues des plateformes réside dans leur capacité à éclairer un segment du tourisme jusqu'alors partiellement mesuré. Les enquêtes auprès des ménages résidents (comme l'enquête Voyages professionnels et personnels réalisée par l'Insee) et les enquêtes auprès des visiteurs étrangers (comme l'enquête EVE de la



L'apport principal des données issues des plateformes réside dans leur capacité à éclairer un segment du tourisme jusqu'alors partiellement mesuré.



Banque de France) couvrent en principe l'ensemble des formes d'hébergement, y compris les locations entre particuliers. Toutefois, la taille d'échantillon de ces enquêtes et leur architecture nationale ne permettent ni un suivi infrarégional fin, ni une distinction détaillée des types d'hébergements effectivement utilisés.

Les données de plateformes modifient cette situation à plusieurs égards. Premièrement, elles apportent une mesure directe de l'activité réalisée, et non plus déclarée. Elles reposent sur les transactions effectivement enregistrées, ce qui permet de mesurer les nuitées réellement consommées, avec une fréquence mensuelle et une ventilation territoriale harmonisée à l'échelle européenne. Cette granularité, tant temporelle que géographique, constitue une avancée majeure pour l'analyse conjoncturelle et territoriale²⁰.

Deuxièmement, elles permettent de mieux couvrir les territoires diffus. Les locations entre particuliers sont souvent situées dans des communes rurales, littorales ou de montagne où l'offre hôtelière est limitée. Ces espaces étaient mal captés par les dispositifs classiques centrés sur les établissements structurés. Les données de plateformes rendent visibles ces dynamiques territoriales, y compris dans des zones de petite taille administrative (NUTS 3), sous réserve des règles de confidentialité.

Troisièmement, elles renforcent la comparabilité internationale. Les initiatives nationales fondées sur le *web scraping* ou sur des registres locaux restaient hétérogènes dans leurs méthodes et leurs périmètres. La centralisation opérée par Eurostat assure un cadre conceptuel commun, permettant des comparaisons directes entre États membres. Cela facilite l'analyse des dynamiques touristiques à l'échelle européenne, notamment en période de choc conjoncturel (crise sanitaire, tensions géopolitiques, variations de change).

Quatrièmement, elles contribuent à une meilleure articulation entre offre (l'activité des établissements de tourisme) et demande (la consommation et les pratiques des ménages). Si les enquêtes auprès des ménages décrivent les comportements touristiques (motifs, dépenses, profils sociodémographiques), les données de plateformes décrivent l'occupation effective d'un segment précis de l'offre marchande. Leur croisement analytique permet d'affiner l'estimation de la structure du marché de l'hébergement et d'évaluer la part croissante des locations de courte durée dans les nuitées marchandes.

²⁰ L'utilisation de ces données est également envisagée pour mieux prendre en compte l'hébergement de courte durée via les plateformes dans le futur système de comptabilité nationale (Askenazy et Bourgeois, 2025).

► Les limites de cette nouvelle source

L'apport de cette source est important car elle est exhaustive bien que sur un champ restreint. Cependant, des limites structurelles subsistent. En particulier, la source repose sur des conventions volontaires avec les plateformes qui peuvent être remises en cause si le marché se restructure. La source reste également dépendante des choix techniques opérés par les plateformes et de l'évolution de leurs systèmes d'information, sur lesquels la statistique publique n'a qu'une maîtrise limitée²¹.

Au niveau du champ, la source ne couvre pas, logiquement, certains types de séjours non marchands (famille, amis, résidences secondaires), très présents dans certains pays comme en France. Au niveau des informations recueillies par Eurostat, les prix des séjours enregistrés ne sont pas communiqués. Or c'est une source d'information importante pour mesurer les dépenses des touristes, lesquelles contribuent à l'établissement du compte satellite du tourisme. Les profils sociodémographiques des visiteurs ne sont pas non plus connus : seul le pays de résidence est renseigné, ce qui limite les études sur le tourisme. Enfin, les motifs (personnels, pour affaires) ne sont pas connus également, limitant là encore

L'articulation avec les enquêtes et certains fichiers administratifs demeure indispensable pour couvrir tous les aspects du tourisme.

les usages de cette source. Ainsi, l'articulation avec les enquêtes auprès des ménages et des établissements pour avoir des précisions sur les types de séjours et les dépenses, mais aussi avec des fichiers administratifs comme Sirene²² ou des fichiers fiscaux de taxe sur la valeur ajoutée (TVA), demeure indispensable pour couvrir tous les aspects du tourisme.

Les statistiques de fréquentation s'intéressent à l'occupation mais aussi à l'offre. Le rapport entre les deux fournit le taux d'occupation. Les données de plateformes ne permettent pas d'avoir aujourd'hui

l'offre de logements. Bien que cette information soit transmise à Eurostat, les doublons entre plateformes sont importants et, en l'absence d'identifiant unique par logement, il est très difficile de les identifier.

Ces limites structurelles n'invalident pas l'intérêt de cette source, mais soulignent au contraire la nécessité d'un cadre méthodologique explicite et documenté pour encadrer l'usage de données privées en statistique publique.

L'expérience menée par Eurostat illustre les potentialités mais aussi les défis liés à l'utilisation de données privées. Elle pose plusieurs questions, la première concernant la place des INS dans un tel processus : comment les associer pleinement pour garantir une qualité comparable à celle des statistiques standards qu'ils produisent et diffusent par ailleurs ? D'autres questions se posent qui concernent plus généralement l'ensemble des données privées utilisées par les INS : jusqu'où la statistique publique peut-elle s'appuyer sur des données issues d'entreprises privées, et en particulier comment garantir la pérennité d'un dispositif fondé sur des accords volontaires ?

²¹ Ces problèmes valent pour d'autres sources de données privées sur le tourisme (Cousin et Hillareau, 2018).

²² Système informatique pour le répertoire des entreprises et des établissements. Voir le lien : <https://www.insee.fr/fr/metadonnees/source/serie/s1020>.

► La question de la stabilité du processus

Le dispositif de transmission de données par les plateformes repose aujourd'hui sur des accords volontaires conclus entre la Commission européenne et plusieurs grands opérateurs privés. Ce fondement contractuel constitue à la fois une innovation institutionnelle et une fragilité structurelle.

Contrairement aux statistiques européennes produites dans le cadre du SSE, qui s'appuient sur des obligations juridiques découlant de règlements européens, les données de plateformes ne sont pas issues d'un cadre contraignant pour les entreprises concernées. Leur transmission dépend donc du maintien d'un équilibre entre les intérêts publics et privés²³. En 2024, Tripadvisor a ainsi décidé de se retirer de ce partenariat, qui représentait pour cette entreprise un prix trop élevé. En effet, les coûts techniques et organisationnels pour extraire, standardiser et transmettre des données compatibles avec les standards statistiques européens demandent des ressources importantes. Or, contrairement à des acteurs comme Airbnb ou Booking, Tripadvisor est davantage une plateforme de comparaison et de mise en relation qu'un acteur central de la transaction. Ce retrait illustre la vulnérabilité du dispositif : même si le poids relatif de l'entreprise dans le segment étudié était limité, cet épisode rappelle que la continuité statistique reste dépendante de la volonté des acteurs privés.

Au-delà du risque de retrait, d'autres facteurs peuvent affecter la stabilité du dispositif

1. Les évolutions internes des plateformes :

Les systèmes d'information des entreprises évoluent en permanence : modifications des catégories d'hébergement, changement dans la gestion des annulations, nouvelles modalités de réservation ou d'intermédiation. Ces transformations, motivées par des impératifs commerciaux ou technologiques, peuvent entraîner des ruptures de série ou des variations non directement liées aux dynamiques touristiques. La statistique publique doit alors détecter et documenter ces changements afin d'éviter des interprétations erronées.

2. La dépendance aux formats et aux définitions propres aux plateformes :

Bien qu'un important travail d'alignement conceptuel ait été mené (notamment sur la correspondance avec le groupe 55.2 de la NACE), certaines notions restent ancrées dans la logique opérationnelle des entreprises. La statistique publique ne maîtrise ni l'architecture initiale des bases de données, ni l'ensemble des traitements en amont de la transmission. Cette asymétrie d'information impose un effort permanent de dialogue technique et de documentation méthodologique.

3. Le risque lié à la structure concurrentielle du marché :

Le segment des locations de courte durée est dominé par quelques grands acteurs mondiaux. Une modification stratégique majeure – fusion, retrait d'un marché, évolution

²³ Le règlement européen (CE) n° 223/2009, révisé en 2024, introduit explicitement la possibilité pour Eurostat et les INS de demander l'accès à certaines données détenues par des entreprises privées lorsque cela est nécessaire à la production de statistiques européennes. Avec l'entrée en vigueur fin 2026 du règlement européen (UE) n° 2024/1028 (dit règlement STR pour « *Short-Term Rental Regulation* », i.e. régulation des locations de courte durée), les plateformes devront obligatoirement transmettre des données de fréquentation à Eurostat et aux autorités publiques nationales. La mise en place concrète est prévue pour 2027. Voir les références juridiques en fin d'article.

réglementaire nationale ou européenne – pourrait modifier significativement la couverture de la source. La représentativité statistique dépend ainsi, en partie, de la structure concurrentielle du secteur.

Face à ces risques, plusieurs leviers contribuent à renforcer la pérennité du dispositif

1. La centralisation par Eurostat garantit une harmonisation méthodologique et une capacité de négociation collective au niveau européen, supérieure à celle qu'auraient des initiatives nationales isolées.
2. La reconnaissance progressive de ces données comme statistiques européennes officielles renforce leur ancrage institutionnel. L'intégration future dans un cadre réglementaire explicite – en articulation avec le règlement (UE) n° 692/2011 sur les statistiques du tourisme – offrirait une base juridique plus stable que de simples accords contractuels.
3. La montée en puissance des cadres européens relatifs au partage de données numériques (notamment dans le contexte plus large de la régulation des plateformes) pourrait, à moyen terme, consolider l'accès aux données à des fins statistiques, en clarifiant les obligations des intermédiaires numériques vis-à-vis des autorités publiques.

En définitive, la pérennité du dispositif ne repose pas uniquement sur des considérations techniques. Elle renvoie à une question plus large, sur la place à accorder aux données privées dans la production statistique officielle. L'expérience menée dans le champ du

tourisme montre qu'un partenariat public-privé est possible, mais qu'il nécessite un encadrement juridique, méthodologique et institutionnel rigoureux. La stabilisation du dispositif constitue ainsi un enjeu central pour garantir la continuité des séries, la comparabilité internationale et la confiance des utilisateurs.

L'expérience menée dans le champ du tourisme montre qu'un partenariat public-privé est possible, mais qu'il nécessite un encadrement juridique, méthodologique et institutionnel rigoureux.

Au-delà de ces questions, le passage de ces statistiques d'occupation du statut de « statistiques expérimentales » à celui de « statistiques européennes officielles » ne constitue pas un simple changement de label.

Il suppose un **travail méthodologique et documentaire approfondi**, indispensable pour garantir la reproductibilité, la compréhension et l'usage correct des données par les utilisateurs.

Ce processus implique notamment la formalisation détaillée des concepts statistiques retenus (logement, séjour, nuitée, capacité) et de leurs éventuelles divergences avec les concepts traditionnels du tourisme. Cela implique également la production de métadonnées exhaustives, décrivant les chaînes de traitement, les règles de validation, les choix d'agrégation et leurs justifications. Enfin, pour garantir l'harmonisation des données produites, il est nécessaire d'avoir une documentation des méthodes propres à chaque plateforme (définition des logements actifs, traitement des annulations, estimation du nombre de voyageurs), afin d'explicitier les éventuelles hétérogénéités en amont.

Ce socle documentaire est une condition essentielle pour permettre aux instituts nationaux, aux chercheurs et aux décideurs publics d'interpréter correctement les résultats, d'en assurer la comparabilité dans le temps et de les articuler avec les autres sources statistiques existantes.

En 2025, Eurostat propose de reconnaître la composante « occupation » issue des données de plateformes comme statistique européenne officielle. Cette reconnaissance, si elle est adoptée dans la révision en cours du règlement européen sur les statistiques du tourisme, marquerait une étape importante : elle consacrerait l'aboutissement d'un dispositif reposant sur une base juridique et technique stabilisée, tout en illustrant l'évolution du rôle de la statistique publique face à la montée en puissance des données détenues par le secteur privé.

► Une ouverture de perspectives pour la statistique européenne

La collecte de données de plateformes par Eurostat constitue une avancée majeure pour la mesure du tourisme européen. Cette source comble un angle mort, améliore la couverture du secteur et offre une granularité temporelle et géographique jusque-là impossible à atteindre. Elle renforce également la comparabilité entre pays.

Mais elle repose sur des accords volontaires, ne couvre pas l'ensemble du champ touristique et nécessite un travail conceptuel important. L'articulation avec les enquêtes et les sources administratives reste essentielle pour produire des indicateurs complets et robustes.

En l'état, ces nouvelles données ne peuvent se substituer aux dispositifs d'enquête. Les enquêtes auprès des ménages et des visiteurs étrangers restent indispensables pour appréhender des dimensions fondamentales du tourisme que les données de plateformes ne captent pas : les motifs des déplacements, les profils sociodémographiques des visiteurs, les dépenses, ou encore les séjours non marchands (chez des proches, en résidence secondaire). Elles permettent également de mesurer l'ensemble des formes d'hébergement et d'inscrire les locations de courte durée dans une vision globale des pratiques touristiques. Les données de plateformes doivent donc être envisagées comme une source complémentaire, particulièrement précieuse pour mesurer l'occupation d'un segment spécifique de l'offre, mais qui ne peut à elle seule rendre compte de la complexité du phénomène touristique.

Au-delà du tourisme, cette expérience ouvre des perspectives pour la statistique officielle européenne : l'intégration de données privées semble appelée à se développer, mais elle impose des garde-fous méthodologiques, juridiques et institutionnels pour garantir la qualité, la transparence et la pérennité de la production statistique.

► Fondements juridiques

- Règlement européen (CE) n° 223/2009 du Parlement européen et du Conseil du 11 mars 2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/fr/ALL/?uri=CELEX%3A32009R0223>.
- Règlement (UE) n° 692/2011 du Parlement européen et du Conseil du 6 juillet 2011 concernant les statistiques européennes sur le tourisme. In : *site de l'Union européenne*. [en ligne]. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32011R0692>.
- Règlement (UE) n° 2024/1028 du Parlement européen et du Conseil du 11 avril 2024 concernant la collecte et le partage des données relatives aux services de location de logements de courte durée. In : *site de l'Union européenne*. [en ligne]. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32024R1028>.
- Règlement (UE) n° 2024/3018 du Parlement européen et du Conseil du 27 novembre 2024 modifiant le règlement (CE) n° 223/2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32024R3018>.

► Bibliographie

- ASKENAZY, Philippe et BOURGEOIS, Alexandre, 2025. Vers une meilleure prise en compte de l'hébergement via des plates-formes en ligne au sein des comptes nationaux. In : *Documents de travail*. [en ligne]. 4 mars 2025. Insee. N° 2025-05. [Consulté le 29 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/8376381>.
- BOITTELLE, Mathieu, CUPILLARD, Émilie, JACQUOT, Alain, JOUBERT, Marie-Pierre, LE GOFF, Florian, 2025. Les données de transactions par carte bancaire CB – Quels apports possibles aux analyses conjoncturelles et territoriales ? In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 119-142. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546947?sommaire=8546949>.
- COUSIN, Guillaume et HILLAIREAU, Fabrice, 2018. Les données de téléphonie mobile peuvent-elles améliorer la mesure du tourisme international en France ? In : *Économie et Statistique*. [en ligne]. 11 avril 2018. Insee. N° 505-506, pp. 89-107. [Consulté le 21 mai 2026]. Disponible à l'adresse : <https://insee.hal.science/hal-05308823v1>.
- COUSIN, Saskia et RÉAU, Bertrand, 2016. Chapitre II. Observer, compter et classer. In : *Sociologie du tourisme*. 20 octobre 2016. La Découverte. Collection Repères, n° 535, pp. 19-28. ISBN : 9782707191144.
- DEVAUX, François, 2024. Les habitants temporaires dans l'espace métropolitain : une approche du sujet à travers l'hébergement touristique, la représentation des acteurs publics et la régulation spatiale dans la métropole parisienne. Thèse de géographie. Université Paris I Panthéon-Sorbonne. In : *portail HAL science ouverte*. [en ligne]. 19 décembre 2024. [Consulté le 21 mai 2026]. Disponible à l'adresse : <https://hal.science/tel-05045699v1>.
- JOUBERT, Marie-Pierre, 2025. Les données de téléphonie mobile – Une source de connaissance sur la population et ses déplacements. In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 95-118. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546945?sommaire=8546949>.
- LAMARCHE, Pierre et RIVIÈRE, Pascal, 2025. À propos du temps de production des statistiques publiques. In : *Courrier des statistiques*. [en ligne]. 15 décembre 2025. Insee. N° N14, pp. 95-118. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8675585?sommaire=8667496>.
- LESUR, Romain, 2025. Sources de données privées : panorama et perspectives. In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 73-94. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546943?sommaire=8546949>.
- ONU, 2008. Recommandations internationales 2008 sur les statistiques du tourisme (RIST 2008). In : *site de l'ONU Tourisme*. [en ligne]. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://www.untourism.int/fr/statistiques-du-tourisme/recommandations-internationales-statistiques-du-tourisme>.

- ULRICH, Amandine, 2021. Hébergements proposés par des particuliers via des plateformes – En 2019, Paris et Nice dans le top 10 des villes les plus fréquentées de l'Union européenne. In : *Insee Première*. [en ligne]. 26 novembre 2021. Insee. N° 1879. [Consulté le 13 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5871113>.
- VIENNET, Henri, 1977. La statistique au secrétariat d'état au tourisme. In : *Courrier des statistiques*. [en ligne]. Octobre 1977. Insee. N° 4, pp. 5-6. [Consulté le 29 mai 2026]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06z97p1.r=courrier%20des%20statistiques%201977?rk=429186;4#>.

Qu'apportent les métiers de la *data* à la statistique publique ?



Lino Galiana* et Olivier Lefebvre**

La numérisation progressive des données a profondément transformé la production d'information, dans un contexte où les progrès technologiques et méthodologiques ont amélioré l'efficacité des outils informatiques. Parallèlement, la société s'est accoutumée à disposer en permanence de statistiques dans le débat public et dans le quotidien, et les entreprises utilisent de plus en plus de données dans leurs processus de décision. Face à ces besoins, il existe une forte demande sur le marché du travail pour des profils spécialisés dans la valorisation de données : *data analysts*, *data scientists*, *data engineers*, *data journalists*...

Ces profils favorisent un rapprochement des compétences entre directions métiers et informatiques, ce qui a de nombreuses vertus : possibilité de réduire les délais de livraison, capacité à associer des données de formes et de natures diverses pour répondre à de nouveaux besoins, production plus transparente... Mais cela représente aussi un défi d'adaptation continue, dans un environnement où les besoins, les outils et les méthodes évoluent rapidement.

La statistique publique tire pleinement parti de ces transformations au service de missions inchangées (éclairer le débat public et aider à la décision publique) et de valeurs essentielles (rigueur, transparence, respect de la confidentialité...). L'Insee et les services statistiques ministériels se sont dotés d'infrastructures et d'organisations pour explorer, expérimenter, industrialiser et maintenir ces nouveaux processus.

 The progressive digitisation of data has profoundly transformed the production of information, at a time when technical and methodological advances have improved the efficiency of computing tools. At the same time, society has grown accustomed to constantly having statistics at its disposal in public consultations and in its everyday life. Companies also use more and more data in their decision process. To meet those needs, there is a high demand on the market for profiles specialised in maximising the value of data: *data analysts*, *data scientists*, *data engineers*, *data journalists*...

These profiles foster skill alignment between business-decision actors and IT functions, bringing many benefits: opportunities to reduce delivery time, matching different types and forms of data to meet new needs, producing more transparently... But this also represents a challenge of constant adaptation in an environment where needs, tools and methods change rapidly.

Official statistics make the most of these transformations while serving the same missions (enlightening public debate and supporting public decision-making) and core values such as discipline, transparency, respect of privacy. INSEE and the Ministerial Statistical Offices have developed infrastructures and organisational frameworks to explore, experiment, industrialise and keep those new processes sustainable.

* Chef de la section « Expertise de la fonction publique », Insee.
lino.galiana@insee.fr

** Direction des statistiques démographiques et sociales, Insee.
olivier.lefebvre@insee.fr



**La généralisation
des traces numériques a
profondément transformé
les métiers de la donnée.**



Affirmer que le métier de *data scientist* est « le boulot le plus sexy du XXI^e siècle » (Davenport et Patil, 2012), ou que la donnée est « le pétrole du XXI^e siècle », est quelque peu emphatique et provocateur. Mais cela reflète une réalité. La donnée est d'abord devenue un objet du quotidien. Puis, progressivement et moyennant des investissements conséquents, elle est

devenue un actif central dans les économies numérisées. Son analyse ouvre des champs d'innovation stimulants et l'opportunité d'un enrichissement des connaissances.

La généralisation des traces numériques¹, alliée aux progrès techniques et à un appétit croissant pour les indicateurs chiffrés, a profondément transformé les métiers de la donnée (ou *data*). La donnée est désormais omniprésente et guide des politiques publiques, des stratégies d'entreprises, des choix individuels et des décisions collectives. Elle est davantage qu'un simple intrant informatique : elle constitue un matériau à exploiter, à transformer et à façonner pour créer de la valeur.

Cette évolution a fait émerger de nouveaux profils, comme les *data scientists* et les *data engineers*, qui viennent compléter les compétences des statisticiens (Volle, 1980), des analystes et des développeurs. La coordination de ces multiples profils est désormais indispensable pour traiter des données massives, hétérogènes et souvent non structurées, et pour en faire un levier d'innovation et de connaissance. Ce matériau est protéiforme. Il impose au statisticien d'adapter ses outils en permanence, et d'intervenir plus en amont de la chaîne de fabrication de statistiques.

Dans le champ de la statistique publique, cette hybridation des compétences, jointe à la construction d'environnements de traitement adaptés, a permis d'élargir les méthodes, de mieux exploiter de nouvelles sources de données, sans sacrifier à la transparence. Cette dernière est en effet la garantie, in fine, de la confiance dans les chiffres produits.

► La *data science* pour tirer parti d'un univers de données en expansion et en évolution

Un stock de données, de toutes formes, en explosion

Les baisses continues des coûts de stockage et de traitement ont joué un rôle de premier plan dans la massification de la production de données. Par ailleurs, Internet, les smartphones et les objets connectés créent des traces numériques directes qui peuvent être accumulées dans des entrepôts de données. Ces derniers sont devenus peu coûteux² : un téraoctet de stockage³ coûtait seulement 11 dollars en 2023, contre 6,5 millions de dollars en 1990⁴.

¹ Informations enregistrées par un dispositif numérique sur l'activité ou l'identité de ses utilisateurs (source : Wikipédia).

² La logique d'accumulation à faible coût de données, sans définir en amont les besoins liés aux cas d'usage, a créé un grand engouement pour le big data pendant les années 2010. Ce paradigme est aujourd'hui critiqué par des acteurs comme Jordan Tigrani (créateur de BigQuery, l'offre d'outil big data de Google). Pour en savoir plus, voir Tigrani (2023).

³ C'est l'unité de mesure la plus grande pour un usage courant. Les centres informatiques nécessitent quant à eux des unités de mesure encore plus grandes.

⁴ Voir le lien suivant : <https://ourworldindata.org/grapher/historical-cost-of-computer-memory-and-storage>.

Les outils numériques, devenus omniprésents, sont aussi créateurs de données diversifiées : valeurs quantitatives classiques mais à haute fréquence, textes, images, vidéos, positions GPS, traces informatiques (*log files*⁵)... Cette diversification des types de données et des contenus a permis d'ouvrir l'éventail des thèmes concernés par la *data* et des acteurs économiques impliqués dans sa valorisation. Par ailleurs (opportunité liée à la disponibilité de ce type d'information ou souci d'objectivation ?), nous faisons une consommation effrénée, voire boulimique, de données dans notre vie quotidienne : nous comptons nos pas, choisissons nos restaurants en fonction de la note moyenne donnée par les consommateurs, etc.

Un besoin croissant en ressources de calcul

Cette explosion du volume de données a entraîné un besoin croissant en ressources de calcul. Elle a favorisé des innovations comme, à partir de la fin des années 2000 :

- le calcul distribué, qui permet de résoudre des problèmes complexes en répartissant le travail entre plusieurs ordinateurs ;
- le calcul scientifique sur GPU (pour *Graphics Processing Unit* en anglais, ou processeur graphique en français), qui permet d'accélérer les traitements en parallélisant les tâches nécessitant peu de mémoire.

Les années 2010 ont par ailleurs été marquées par un recours de plus en plus fréquent à des infrastructures aux ressources mutualisées, notamment dans le *cloud*⁶. Ces dernières permettent de bénéficier d'une plus grande puissance de calcul

et d'une sécurité accrue par rapport aux micro-ordinateurs.

La convergence entre accumulation de données, augmentation de la puissance de calcul et progrès des logiciels de programmation a redynamisé des champs de recherche restés à l'état de promesses.

La convergence entre accumulation de données, augmentation de la puissance de calcul et progrès des logiciels de programmation a redynamisé des champs de recherche existant dès les années 1950, mais longtemps restés à l'état de promesses faute de moyens technologiques adéquats. C'est notamment le cas des réseaux de neurones, une des grandes méthodes d'intelligence artificielle⁷. Entre les

premières formalisations académiques des années 1950 et leur adoption industrielle à la fin des années 2000, ils ont eu peu de succès. En effet, peu d'infrastructures étaient suffisamment puissantes pour les mettre en œuvre, mais aussi, rares étaient les cas d'usage justifiant une telle complexité de modélisation.

⁵ Un *log file* ou *log* est un fichier informatique contenant l'historique des événements affectant un processus.

⁶ Ensemble des serveurs distants proposant des services accessibles par le réseau Internet (source : Le Robert).

⁷ Méthode permettant la résolution de problèmes complexes, comme la vision par ordinateur. Voir le glossaire mis en ligne par la Commission nationale de l'informatique et des libertés (Cnil) à l'adresse suivante : <https://www.cnil.fr/fr/glossaire>. Voir Russell et Norvig (2021) pour une histoire détaillée des cas d'usage de l'intelligence artificielle.

Les grands modèles de langage (LLM)⁸ qui prennent aujourd’hui place dans notre quotidien, par exemple pour répondre à des questions administratives ou commerciales à la place de conseillers humains, s’inscrivent dans ce mouvement. Ils sont dans la continuité directe des travaux des années 2010, une décennie riche d’innovations dans le domaine de l’analyse textuelle, elles-mêmes rendues possibles par l’entraînement de modèles sur des corpus textuels de grande ampleur. Certaines tâches, comme l’entraînement de modèles de fondation⁹, nécessitent néanmoins une puissance financière importante. Ainsi, l’entraînement du modèle de langage GPT-4, développé par la société OpenAI en 2023, aurait coûté 80 millions de dollars (Maslej et al., 2025).

Une donnée qui circule beaucoup plus

Les quinze dernières années sont marquées par une ouverture progressive des données (open data), et une multiplication des acteurs associés, en France comme à l’international (Kitchin, 2021). Ce mouvement concerne notamment les données du secteur public, considérées comme un levier pour éclairer le débat citoyen et les décisions de politiques publiques (Bothorel, 2020).



Les quinze dernières années sont marquées par une ouverture progressive des données (open data), et une multiplication des acteurs associés, en France comme à l’international (Kitchin, 2021).



En France, cette dynamique s’appuie à la fois sur :

- des dispositifs institutionnels : création d’Etalab¹⁰ et de la plateforme de partage de données publiques data.gouv.fr en 2011, créations de postes d’administrateurs ministériels des données, des algorithmes et des codes sources à la fin des années 2010 ;
- des outils et cadres techniques, comme les interfaces de programmation applicative (API pour *application programming interface* en anglais)¹¹ ou le Référentiel général d’interopérabilité (RGI)¹² ;
- un cadre législatif (loi du 7 octobre 2016 pour une République numérique¹³) qui facilite l’accès des citoyens, chercheurs et entreprises aux données produites par les administrations, tout en garantissant la protection de la vie privée, en conformité avec la loi Informatique et Libertés¹⁴ et le règlement général pour la protection des données (RGPD)¹⁵.

⁸ LLM pour *Large Language Models* en anglais. Un modèle de langage peut par exemple prédire le mot suivant dans une séquence de mots (voir le glossaire de la Cnil).

⁹ Modèles d’intelligence artificielle « généralistes », qui peuvent être adaptés à une grande variété de tâches (voir le glossaire de la Cnil).

¹⁰ À la suite du décret n° 2011-194 du 21 février 2011 portant création d’une mission « Etalab » chargée de la création d’un portail unique interministériel de données publiques. Voir les références juridiques en fin d’article.

¹¹ Une API est une interface logicielle qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service, afin d’échanger des données et des fonctionnalités.

¹² Le RGI est un cadre de recommandations référençant des normes et standards qui favorisent l’interopérabilité au sein des systèmes d’information de l’administration.

¹³ Voir les références juridiques en fin d’article.

¹⁴ Loi n° 78-17 du 6 janvier 1978 relative à l’informatique, aux fichiers et aux libertés. Voir les références juridiques en fin d’article.

¹⁵ Règlement (UE) n° 2016/679 du Parlement européen et du Conseil du 27 avril 2016. Voir les références juridiques en fin d’article.

La loi pour une République numérique favorise par ailleurs la transmission, pour la production de statistiques publiques, de données détenues par des entreprises privées. Une disposition du même ordre figure depuis fin 2024 dans le règlement européen sur la production de statistiques¹⁶.

Pour les données individuelles, plus sensibles, la diffusion à des fins de recherche a été encouragée, dans un cadre simplifié, tout en conservant un haut niveau de sécurité, grâce notamment au Centre d'accès sécurisé aux données (CASD) (Gadouche, 2019). La diversification des producteurs de données s'est aussi traduite par une ouverture parallèle de données par des entreprises privées, que ce soit gratuitement ou dans une logique de monétisation (Lesur, 2025).

L'effort de normalisation et d'harmonisation constitue une condition préalable pour tirer parti de données de plus en plus nombreuses et hétérogènes. Au-delà de l'interopérabilité technique¹⁷, qui a progressé avec la diffusion des API et de formats de données standardisés (Dondon et Lamarche, 2023), la réutilisation effective suppose aussi une contextualisation suffisante. Il faut en effet pouvoir connaître un minimum l'univers métier dans lequel sont conçues les données pour éviter des incohérences, des interprétations erronées voire des mésusages, ce que la littérature désigne par le terme anglais *data friction* (Edwards et al., 2011). Extraire de l'information pertinente d'une photo, d'un scan de documents, d'un texte ou d'un graphe de relations est en effet plus difficile que si l'on part d'une information déjà organisée et construite pour une finalité bien définie.

Mais au-delà de ces enjeux techniques, l'augmentation et la diversification des producteurs de données transforment également les pratiques : elles ouvrent la voie à des usages plus empiriques et variés.

Des approches plus empiriques, partant davantage de la donnée que de la théorie

Ces données, plus nombreuses, plus accessibles, créent des occasions de valorisation à saisir, parfois imprévues dans leur usage initial.

Le premier usage opportuniste consiste en une mutualisation de données entre plusieurs acteurs lorsque les conditions de partage le permettent. C'est notamment le cas dans le domaine des démarches administratives avec le principe du « Dites-le-nous une fois ». En permettant des transferts de pièces justificatives entre administrations, il simplifie les formalités des citoyens¹⁸ et réduit les risques d'erreurs ainsi que les délais de traitement.

Les données peuvent aussi bénéficier d'une seconde vie en étant réutilisées, souvent par de nouveaux acteurs, à des fins différentes de celles prévues lors de leur conception. Parmi les exemples classiques, on trouve l'utilisation de données de téléphonie mobile pour estimer les flux de population. On détourne alors les traces numériques de leur

¹⁶ Règlement (UE) n° 2024/3018 modifiant le règlement (CE) n° 223/2009 relatif aux statistiques européennes. Voir les références juridiques en fin d'article.

¹⁷ Capacité de matériels, de logiciels ou de protocoles différents à fonctionner ensemble et à partager des informations (source : Larousse).

¹⁸ Par exemple quand une attestation de quotient familial de la caisse d'allocations familiales est nécessaire pour bénéficier de la gratuité d'un service public local.

usage premier – le bornage aux antennes, qui permet d’acheminer les données vers l’utilisateur – vers un autre usage – par exemple l’estimation du nombre de visiteurs d’un événement sportif ou festif (voir aussi Joubert, 2025, pour différents usages intéressant la statistique publique).



Les données peuvent bénéficier d’une seconde vie en étant réutilisées à des fins différentes de celles prévues lors de leur conception.



Cette logique empirique se retrouve aussi dans les méthodes de traitement, notamment à travers le succès des algorithmes prédictifs. En l’occurrence, l’empirisme renvoie au fait que ces méthodes, comme l’apprentissage automatique (machine learning)¹⁹, ne reposent pas sur une théorie préalable mais sur la détection de corrélations utiles à la création de règles de décision. Leur mise en œuvre repose en partie sur de nouveaux métiers, avec une frontière plus ou moins marquée entre

les rôles selon les organisations. Les **data scientists** développent les expérimentations de modèles, les **data engineers** assurent leur intégration dans les systèmes et leur mise à l’échelle, tandis que les **data analysts** font le lien entre les indicateurs découlant des traitements et les recommandations à des fins de décision (*figure 1*).

► Une évolution des manières de travailler

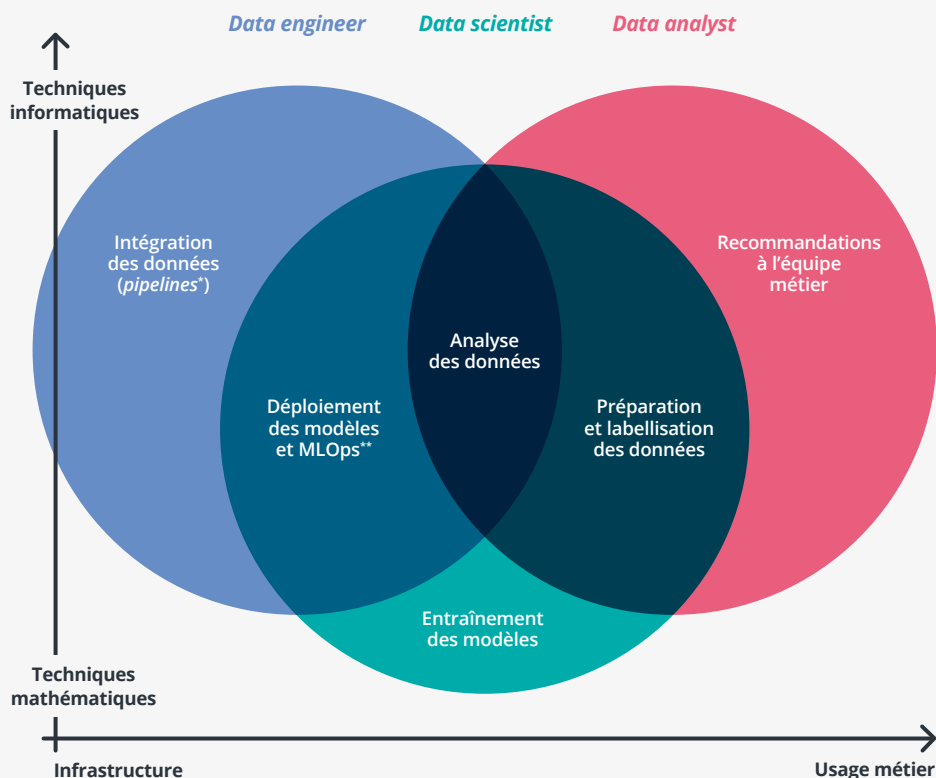
Pour que la statistique publique bénéficie pleinement de ces opportunités, les manières de travailler doivent évoluer. En particulier, deux exigences s’imposent, qui sont étroitement liées : celle de la répliquabilité des traitements et celle de l’ouverture des programmes informatiques (open source). En effet, les données produites doivent être fiables, comparables et suivies dans le temps et les méthodes qu’emploient les statisticiens publics gagnent à être transparentes.

L’exigence de répliquabilité au cœur des chaînes de traitement

Dans un contexte où données, environnements de travail et outils évoluent constamment, pouvoir répliquer un traitement ou reproduire des résultats dans des conditions plus ou moins différentes de leur élaboration d’origine s’avère indispensable pour mettre en production de manière fluide des projets expérimentaux. Les acteurs numériques recourent de plus en plus souvent à des environnements *cloud*, où le principe du microservice prévaut, c’est-à-dire à des environnements logiciels temporaires recréés à la demande. Dans ce contexte, l’adoption de pratiques favorisant la répliquabilité devient indispensable sous peine de ne pas être en mesure de reproduire dans le temps une chaîne de production, ou de simuler l’impact de modifications, en vue d’une optimisation du processus ou d’une analyse de la qualité des données produites.

¹⁹ Champ d’étude de l’intelligence artificielle qui vise à donner aux machines la capacité d’« apprendre » à partir de données, via des modèles mathématiques (voir le glossaire de la Cnil).

► **Figure 1 - Le halo des métiers de la data**



* Traitements de données standardisés.

** MLOps : Voir [encadré 1](#).

Note : Selon les organisations, le positionnement et la taille des cercles peut varier.

Parmi les solutions techniques permettant de répliquer une série de programmes dans son intégralité, la conteneurisation (Comte et al., 2022) occupe une place centrale. Cette approche consiste à créer, pour chaque traitement, un environnement technique maîtrisé de A à Z garantissant sa reproductibilité. Cette approche renforce l'autonomie technique des *data scientists* et *data engineers* ; les premiers sont autonomes pour définir précisément l'environnement utilisé par leurs prototypes que les seconds peuvent ainsi industrialiser plus facilement.

Plutôt que de reconstruire un environnement sur la base de spécifications, les *data engineers* peuvent donc l'adapter directement pour le passage à l'échelle. Le rôle d'un *data engineer* se distingue ainsi de celui d'un développeur applicatif classique ; il ne s'agit pas de développer une chaîne répondant aux besoins du métier (*data scientists* ou *data analysts*), mais d'intégrer et d'organiser les données au sein d'un écosystème plus vaste de services (entrepôts de données consommables par des logiciels statistiques, tableaux de bord, etc.), offrant aux utilisateurs finaux une plus grande autonomie dans l'exploitation de l'information brute disponible.

“

Parmi les solutions techniques permettant de répliquer une série de programmes dans son intégralité, la conteneurisation occupe une place centrale.

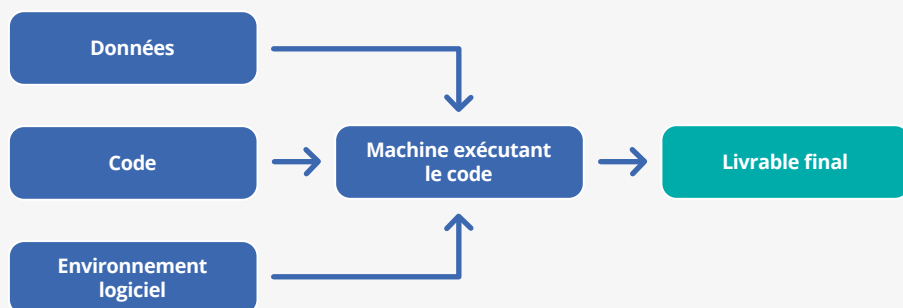
”

Afin d'offrir la flexibilité attendue dans un écosystème mouvant, les pratiques actuelles reposent sur la séparation conceptuelle des trois composantes de l'exécution du traitement : stockage des données, stockage du code et définition de l'environnement logiciel nécessaire au fonctionnement de la chaîne (figure 2). Grâce à cette organisation, il est possible de faire évoluer séparément chaque

composante sans fragiliser l'ensemble. Notamment, des solutions techniques permettent d'expérimenter un traitement nouveau en parallèle de la stabilisation d'une version en production.

L'articulation de ces éléments demande néanmoins un savoir-faire spécifique, ce qui explique le rapprochement entre les compétences d'analyse de la donnée et celles de développement informatique, tout comme la fourniture d'un environnement qui en permette la mise en œuvre.

► **Figure 2 - Architecture d'une infrastructure cloud moderne au service des data scientists et des data engineers**



L'open source, un levier pour des chaînes de traitement flexibles et lisibles

Dans un contexte où les intrants d'une chaîne de production (notamment les données) évoluent fréquemment, il est nécessaire d'être en capacité d'adapter en permanence les outils de traitement. À cet égard, l'open source²⁰ se révèle être un levier stratégique de première importance. Utiliser des solutions construites et éprouvées par une communauté aux besoins similaires permet d'adopter des outils plus agiles, évolutifs et collaboratifs que les solutions propriétaires²¹. Ces dernières sont en effet souvent figées pour garantir leur stabilité (ou la « captivité » de leurs utilisateurs).

²⁰ Un logiciel open source se caractérise par le fait que son code source est libre d'accès, réutilisable et modifiable.

²¹ Un logiciel propriétaire se caractérise par le fait que son code source est fermé, i.e. n'est pas accessible au public.

L'avantage des logiciels open source va bien au-delà de leur gratuité. En premier lieu, ils peuvent s'adapter plus vite aux besoins de leurs utilisateurs, lesquels peuvent eux-mêmes développer des modules, accélérant ainsi le cycle d'intégration de fonctionnalités. Surtout, ces logiciels apportent de la transparence par rapport aux logiciels propriétaires. Contrairement à ces derniers, qui ne communiquent pas forcément sur les choix

méthodologiques, ni sur la manière dont le code les met en œuvre, les outils open source permettent de voir et de contrôler l'ensemble du processus.

Il est nécessaire d'être encapacité d'adapter en permanence les outils de traitement. À cet égard, l'open source se révèle être un levier stratégique de première importance.

Cette logique fait du code un objet de communication entre des équipes amenées à collaborer sur une chaîne de valeur de données. Ceci explique que les praticiens de la donnée se soient appropriés des outils et des pratiques développés à l'origine dans la communauté du développement logiciel. C'est notamment le cas

du logiciel Git²², qui est progressivement devenu un outil du quotidien. Il permet d'abord de tracer de manière fine les évolutions dans le temps d'un projet et de fiabiliser les échanges de programmes informatiques, du développement au déploiement. Mais aussi, il constitue une porte d'entrée pour les statisticiens et *data scientists* vers l'adoption de pratiques agiles²³ et collaboratives, propices à des projets plus transparents, de meilleure qualité, et plus réactifs. Ce recours à des outils facilitant la répliquabilité fait par ailleurs écho à un débat scientifique plus large sur la « crise de la reproductibilité » (Ioannidis, 2005 ; Kapoor et Narayanan, 2023). Ce débat a amené à l'adoption, par la communauté académique, de règles et pratiques plus exigeantes pour favoriser la transparence et la répliquabilité des publications (Vilhuber, 2020).

La transparence permise par l'open source nécessite cependant une vigilance accrue sur la qualité des composants intermédiaires. Là où un logiciel propriétaire propose, pour un besoin donné, un nombre restreint de manières d'y répondre et des solutions guidées par l'éditeur du logiciel, l'écosystème open source peut mettre à disposition plusieurs outils. Ces derniers n'intègrent pas forcément les mêmes fonctionnalités ou ne sont pas de qualité équivalente. Cette logique responsabilise ainsi les équipes développant des codes. Car la maintenance d'une chaîne de production ne passe pas que par les mises à jour de l'infrastructure, mais aussi par le suivi des différentes briques d'un projet : il convient de s'assurer que les solutions envisagées à un moment restent pertinentes et opérationnelles.

²² Git est un logiciel libre qui permet de gérer les différentes versions d'une application informatique et qui facilite la collaboration entre les développeurs.

²³ La méthode agile est une méthode de gestion de projet favorisant la collaboration, la flexibilité et l'adaptabilité. Elle se concentre sur des cycles de développement courts et itératifs.

► De formidables opportunités à mettre au service de la statistique publique

L'univers de la donnée et de l'intelligence artificielle connaît une expansion accélérée, stimulée par des avancées technologiques qui en facilitent la collecte, la circulation et l'utilisation. Cette profusion de données ou d'outils impose de nouvelles manières de travailler et de s'organiser, et accélère les transformations qui sont en œuvre.

Il y a quelques décennies, le modèle dominant était celui des enquêtes. Dans ce dernier, le statisticien était maître de son « destin statistique », à savoir la conception de l'enquête, l'échantillonnage, le plan d'exploitation, l'analyse et la diffusion. En revanche, il était dépendant de collègues informaticiens pour la réalisation des outils de traitement et des produits de diffusion.

Progressivement, les enquêtes se sont vues complétées par des données administratives, par exemple sur les salaires, les revenus ou les résultats des entreprises. Les prendre en compte demandait un temps important de préparation et de maturation, à la fois politique et juridique (pour accéder aux données), sémantique (pour passer de concepts administratifs à des concepts statistiques) et technique (pour transmettre et traiter ces données, bien plus volumineuses que des données d'enquêtes)²⁴. Parallèlement, les statisticiens ont commencé à s'approprier un outil de traitement statistique des données, qui leur a permis de gagner en autonomie et en flexibilité sur l'analyse des données produites, voire sur la construction de certaines étapes de traitement (par exemple les redressements ou la création de variables supplémentaires par combinaison de variables collectées). L'outil propriétaire en question, bien que riche, était assez peu évolutif, ce qui présentait l'avantage d'un univers bien balisé, relativement stable, mais l'inconvénient d'une dépendance marquée et d'une capacité d'innovation limitée.



Loin d'être en rupture avec les valeurs du statisticien public, les apports de la data science peuvent les renforcer, en réconciliant innovation technique et exigence scientifique et déontologique.



Dans le contexte actuel, les sources de données foisonnent, et la technique, telle que décrite précédemment, permet de les tester assez facilement, et de les intégrer dans les chaînes de production. On gardera en tête que, si la technique permet beaucoup de choses, cela n'exonère pas le statisticien de réflexions et d'échanges sur la légitimité du traitement, la capacité juridique à accéder aux données et la protection de la vie privée.

Le statisticien public se trouve face à une opportunité majeure : enrichir ses outils et méthodes, diversifier ses sources et thématiques d'analyses, tout en restant fidèle à ses principes fondamentaux que sont l'indépendance, la rigueur méthodologique,

²⁴ Plus largement, Lamarche et Rivière (2025) comparent le temps de préparation des processus de production de statistiques publiques selon le type de données sur lesquelles ils sont fondés : données d'enquêtes, données administratives ou données détenues par des opérateurs privés.

le respect de la confidentialité²⁵ et la transparence. Loin d'être en rupture avec ces valeurs, les apports de la *data science* peuvent les renforcer, en réconciliant innovation technique et exigence scientifique et déontologique.

Intégrer la *data science* : une question d'infrastructures et de compétences

Le secteur privé a été précurseur dans l'intégration de profils aux compétences multiples au sein des chaînes de valeur de la donnée. Mais le secteur public – et la statistique publique en particulier – a également pris ce virage depuis plus d'une dizaine d'années²⁶. L'Insee est confronté depuis longtemps aux enjeux du traitement de données produites pour une finalité de gestion plutôt que statistique. Néanmoins, l'exploitation accrue de celles-ci à différents stades du processus de production statistique et les premiers travaux visant à valoriser des données privées (à partir du début des années 2010) ont renforcé les besoins d'infrastructures et de compétences adaptées.

Dans son plan stratégique « Insee 2025 » adopté en 2015, l'Insee retenait parmi ses quatre grandes orientations pour la période 2016-2025 : « Innover et être en première ligne sur les sources de données ». Cela impliquait d'investir dans trois directions indispensables et complémentaires :

- des infrastructures informatiques permettant de traiter ces nouvelles données de manière souple et sécurisée ;
- le développement de compétences pour le traitement de ces données ;
- la mise en place d'un environnement d'innovation de type « *lab* »²⁷, permettant d'explorer de nouvelles sources de données et de conduire des expérimentations en matière de *data science*.

D'un point de vue organisationnel, cela s'est traduit par la mise en place de **deux équipes, l'une positionnée à la direction du système d'information, l'autre à la direction de la méthodologie statistique**. Ces deux équipes interagissent étroitement entre elles et, du fait de leur positionnement transversal, avec les utilisateurs et les porteurs d'innovations. Ce mode de fonctionnement facilite à la fois le travail collectif et la pratique du mentorat propre à développer les compétences et diffuser les bonnes pratiques.

Porté par la **première équipe**, l'Insee a lancé le SSPCloud²⁸ (Comte et al. 2022), une infrastructure de *data science* pensée pour lever les freins à l'innovation dans le secteur public (manque d'autonomie et d'écosystème adapté à l'expérimentation, outils inadaptés, cloisonnements entre directions métiers et direction du système d'information). Conçu comme un bac à sable contrôlé, le SSPCloud permet aux *data scientists* d'expérimenter librement, pour pérenniser ensuite les expérimentations grâce à la conteneurisation. Cette

²⁵ Voir à ce sujet l'article du blog de l'Insee consacré à la manière dont l'institut protège les données qu'il collecte (Lefebvre, 2025).

²⁶ La direction interministérielle du numérique (DINUM) et l'Insee ont consacré un rapport à l'évaluation des besoins de l'État en compétences et expertises en matière de donnée (Bourlange et al., 2021). Y sont évoqués les besoins de recrutement de *data scientists* dans divers domaines d'application pour s'adapter à l'offre de données croissante. Dans une perspective plus centrée sur la statistique publique, voir également Beaurepaire et al. (2022).

²⁷ Abréviation anglaise de *laboratory* (laboratoire).

²⁸ Le SSPCloud est disponible au lien suivant : <https://datalab.sspcloud.fr/>. Le logiciel sous-jacent, Onyxia, évoqué ultérieurement, est rendu public au lien suivant : <https://github.com/InseeFrLab/onyxia>.

infrastructure est devenue un outil central de formation aux écoles ENSAI et ENSAE²⁹ et, dans le cadre de la formation continue, pour les statisticiens publics. Ses outils couvrent l'ensemble des compétences en science et ingénierie des données.

En complément des logiciels d'analyse de données (R et Python), le SSPCloud permet également de disposer d'une large palette d'outils open source couvrant des besoins variés. Ceux-ci sont mis à disposition sous la forme de services préconfigurés et interfaçables (i.e. pouvant être connectés à d'autres systèmes). Ils favorisent la modularité des chaînes de valeur de données en permettant de décomposer celles-ci en maillons, lesquels peuvent évoluer selon les besoins ou l'apparition de nouvelles briques techniques.

Pour favoriser la création de plateformes identiques dans d'autres environnements, l'Insee a développé pour l'occasion le logiciel open source Onyxia (Avouac et al., 2025). Celui-ci permet à d'autres organisations de déployer et de gérer une plateforme équivalente au SSPCloud sur leur propre infrastructure. Onyxia a bénéficié du soutien d'acteurs tels que la direction interministérielle du numérique (Dinum), et d'une communauté active de développement issue, notamment, d'autres instituts nationaux de statistique. Parmi les organisations utilisant le logiciel Onyxia, on trouve des institutions telles que l'institut d'océanographie Mercator Ocean, l'agence publique de coopération technique Expertise France, des opérateurs de l'État, des établissements d'enseignement supérieur et de recherche et des instituts de statistique européens. Leurs missions sont assez diverses, mais ces organisations partagent le besoin commun de mettre à disposition, auprès des utilisateurs de leurs données, des outils flexibles et construits à partir des technologies open source les plus récentes.

Pour renforcer l'autonomie de ses statisticiens dans la production statistique sur données confidentielles, l'Insee a déployé début 2024 une nouvelle infrastructure interne, appelée LS3. Elle offre, dans un cadre conteneurisé, des environnements R/Python et

des outils spécialisés (stockage S3³⁰, MLFlow³¹, etc.), tout en assurant la sécurité des données traitées. L'infrastructure, encore récente, répond déjà aux besoins d'environ 10 % des statisticiens de l'institut.

Conjointement au développement d'infrastructures informatiques, l'Insee a mis en place un *lab* de *data science* (nommé SSP Lab), dédié à l'expérimentation et à l'exploration de nouvelles sources ou de nouveaux usages de la donnée. C'est la **deuxième équipe** évoquée précédemment, dédiée, elle, à l'innovation sur le plan de la méthodologie statistique. Son rôle

est d'accompagner les équipes de production dans leurs démarches d'innovation, de diffuser les pratiques de *data science* dans la communauté des statisticiens publics et de mutualiser les retours d'expérience.



Conjointement au développement d'infrastructures informatiques, l'Insee a mis en place un lab de data science dédié à l'expérimentation.



²⁹ ENSAI : école nationale de la statistique et de l'analyse de l'information ; ENSAE : école nationale de la statistique et de l'administration économique.

³⁰ S3 pour *Simple Storage Service* en anglais (soit « moyen simple de stockage » en français). Il s'agit d'un service de stockage de données sur Internet.

³¹ Plateforme open source au service de la traçabilité des projets utilisant du machine learning.

Cette équipe a notamment piloté les premières expérimentations de modèles de machine learning à l'Insee et a accompagné ensuite leur mise en production. Ce fut le cas par exemple pour le modèle destiné à déterminer le code de l'activité principale exercée (APE) des nouvelles entreprises à partir des libellés d'activité déclarés lors de leur création. Le déploiement rapide du premier prototype, ainsi que l'amélioration continue du modèle, ont été rendus possibles grâce à l'adoption des principes du MLOps (**encadré 1**). Depuis 2022, cette équipe anime également le réseau SSPHub³², ouvert à tous les *data scientists* du secteur public, à travers des ateliers, formations, conférences et échanges de pratiques innovantes.

Contrairement aux apparences, la répartition des missions entre deux équipes ne reproduit pas un cloisonnement entre statisticiens et informaticiens, mais favorise au contraire leur collaboration étroite, indispensable à la diffusion de l'innovation dans un environnement où métiers informatiques et statistiques étaient historiquement assez disjoints. Leur montée en compétences s'est faite de manière coordonnée, en s'enrichissant mutuellement. Par ailleurs, l'appartenance de ces deux équipes à deux directions distinctes permet également de se rapprocher d'utilisateurs de profils différents.

D'autres organisations de la statistique publique ont opté pour une approche différente, en concentrant leurs compétences au sein d'une seule équipe dédiée à la *data science*. C'est le cas de la direction de la recherche, des études, de l'évaluation et des statistiques (Drees)³³, avec son Lab Santé fortement mobilisé pendant la crise sanitaire, ou de la sous-direction des systèmes d'information et études statistiques (Sies)³⁴, dont l'équipe de *data science* développe le « baromètre de la science ouverte » en mobilisant des méthodes d'apprentissage automatique sur des publications de chercheurs. Cette approche de concentration des compétences dans une seule équipe a également été adoptée par les services *data* de l'inspection générale des finances (Bolliet et al. 2025), de l'inspection générale des affaires sociales (Berthe, 2025) ou de la Cour des comptes³⁵.

Une hybridation des compétences nécessaire à l'aboutissement de projets statistiques de grande ampleur

La crise sanitaire a représenté pour plusieurs institutions un accélérateur d'innovation. La production et la diffusion quotidiennes par la Drees de statistiques sur l'épidémie de Covid (hospitalisation, tests positifs...) a représenté un véritable défi. Pour le relever, il est apparu indispensable d'adopter une forme d'agilité et d'arriver à automatiser le processus malgré un environnement instable. La Drees n'a pas désinvesti ce sujet après la crise sanitaire. Pour valoriser le système national des données de santé (SNDS)³⁶, qui permet de produire des indicateurs ou des bases de données telles que celle des restes à charge, plusieurs profils sont mobilisés. Il y a d'abord des ingénieurs capables de traiter les données fournies par la caisse nationale de l'assurance maladie (Cnam). Il y a ensuite des experts issus de différents champs disciplinaires (*data scientists*, statisticiens ou médecins), pour donner un sens socioéconomique aux données traitées.

³² Les ressources communautaires développées dans le cadre de ce réseau sont recensées au lien suivant : <https://ssphub.netlify.app/>.

³³ Service statistique ministériel dans le domaine de la santé et des solidarités.

³⁴ Service statistique ministériel dans le domaine de l'enseignement supérieur et de la recherche.

³⁵ Voir l'article de Belrhomari et al. dans ce même numéro.

³⁶ Voir l'encadré consacré au SNDS dans Berthe (2025).

Intégrer certaines sources de données transactionnelles dans les processus de production représente parfois un véritable défi informatique et méthodologique. C'est le cas de l'intégration de données de caisse³⁷ dans l'indice des prix à la consommation (Leclair, 2019). Le temps important d'expérimentation avant le passage en production s'explique d'abord par le besoin de créer une infrastructure informatique capable de traiter la charge (plus de 1,7 milliard d'enregistrements quotidiens par mois). Il s'explique aussi par la nécessité de s'assurer que l'utilisation de cette source n'introduise pas de

► **Encadré 1. La production à partir de machine learning, un travail d'amélioration continue**

Dans le domaine du développement logiciel, l'approche DevOps vise à favoriser la collaboration entre les équipes de développement (Dev) et celles en charge des opérations (Ops) c'est-à-dire de l'exploitation opérationnelle. Cela passe principalement par l'intégration du cycle de vie d'un projet dans un continuum de phases de travail permettant la simultanéité, et non l'enchaînement séquentiel, de périodes d'expérimentation et de production. L'objectif est de réduire le « mur de la production », c'est-à-dire le risque qu'un projet échoue lors du passage en production par manque d'anticipation, en intégrant dès la conception les enjeux liés au déploiement, à la maintenance et à la remontée de bugs d'un logiciel.

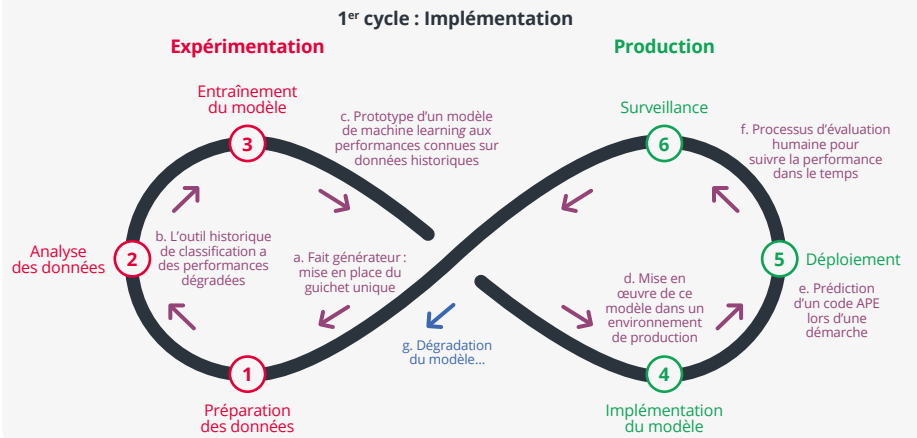
Comme les projets impliquant un modèle de machine learning s'inscrivent dans un travail collaboratif entre différents profils *data* et requièrent des cycles itératifs, l'approche correspondante, dite MLOps (ML pour machine

learning), a adopté les principes DevOps dans ce domaine. La différence de champ d'application (développement logiciel versus machine learning) nécessite cependant des adaptations pour certaines phases. C'est notamment le cas pour l'évaluation, qui mobilise des critères plus larges que la seule satisfaction d'un utilisateur de service.

L'intérêt de l'approche MLOps peut être illustré à partir du projet de codification automatique de l'activité principale exercée (APE) à partir de déclarations textuelles d'activités (en langage naturel) faites dans le cadre des formalités d'entreprise (Avouac et al., 2025) (*figure encadré*).

Dans une première phase d'**expérimentation**, un corpus d'entraînement issu des déclarations passées a permis de tester plusieurs modèles. Cette étape a mobilisé des analystes métiers et gestionnaires d'applications pour définir les besoins, ainsi que des statisticiens et *data scientists* pour concevoir et évaluer le modèle le plus adapté.

Cycles de vie d'un modèle de machine learning : l'exemple de l'activité principale exercée (APE)



³⁷ Informations recueillies par les enseignes du commerce de détail, au moment où les clients passe à la caisse, sur les produits achetés et les prix payés.

rupture de série, dans le contexte d'un indice produit à partir de plusieurs sources (données de caisse, *web scraping*³⁸, relevés d'enquête, données administratives).

L'expérience acquise sur ces projets a permis d'adopter pour les chantiers plus récents une démarche d'ingénierie moderne, tenant compte au mieux de l'état de l'art en la matière. C'est le cas par exemple pour le répertoire statistique des individus et des logements (Résil) de l'Insee (Lefebvre, 2024). Celui-ci mobilise pour sa mise à jour des

La **mise en production** a consisté ensuite à entraîner le modèle retenu, puis à l'utiliser pour prédire le code APE des nouvelles déclarations. Elle s'est déroulée de manière fluide grâce à l'adoption d'une démarche MLOps et à l'utilisation d'une infrastructure conteneurisée. Ce dispositif a permis d'éviter une réécriture du code par les *data engineers* : le métier conserve la maîtrise du traitement des données, tandis que les ingénieurs garantissent l'intégration du modèle dans un environnement de production stable et sécurisé.

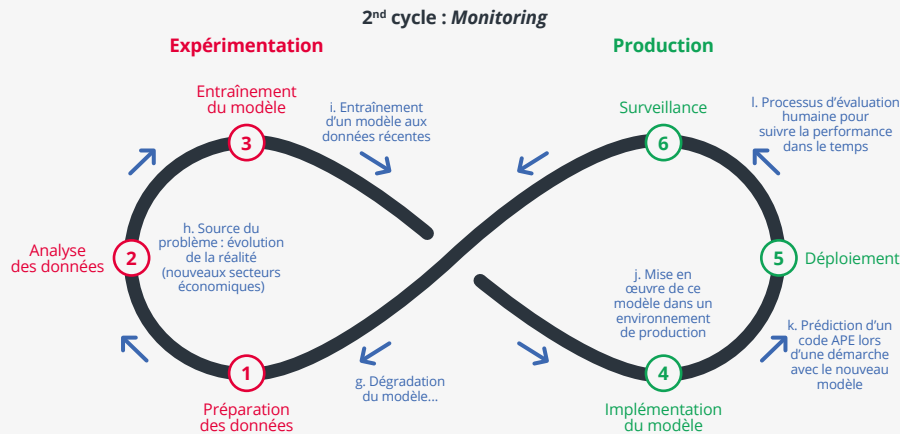
L'un des enjeux majeurs du machine learning en production réside dans la surveillance continue (*monitoring*) du modèle. Un modèle entraîné sur des données historiques peut perdre en pertinence face à des phénomènes nouveaux. Par exemple, l'émergence de nouvelles activités liées à l'économie des plateformes ou à la livraison à domicile change les termes employés dans les déclarations. La surveillance du modèle via la collecte de *logs* et de données annotées* constitue donc un élément essentiel de la démarche MLOps. Elle assure à la fois la qualité des prédictions en

production et l'alimentation des futures phases d'expérimentation, permettant ainsi un cycle d'amélioration continue.

L'adoption de plateformes conteneurisées comme Onyxia contribue par ailleurs à fluidifier ce processus : elles offrent un environnement cohérent entre l'expérimentation (logiciel Python associé à des ressources de calcul adaptées, entrepôt de modèles, stockage de données) et la production, renforçant la reproductibilité et la portabilité (capacité à fonctionner sur un environnement différent) des traitements.

Reste enfin à définir les choix stratégiques : fréquence de réentraînement des modèles, modalités de collecte de nouvelles données annotées, ou seuils déclenchant une nouvelle phase d'expérimentation. Ces décisions, à la fois organisationnelles et techniques, conditionnent la capacité à accélérer les itérations et à améliorer durablement la qualité des productions statistiques fondées sur le machine learning.

* L'annotation est le procédé par lequel les données sont décrites manuellement afin d'être caractérisées (voir le glossaire de la Cnil).



38 Le *web scraping*, ou moissonnage en français, est une technique d'extraction automatisée de données de sites web.

données issues de diverses sources administratives, souvent volumineuses et parfois incomplètes, à harmoniser et à traiter dans un temps contraint. Ainsi, la fabrication de l'univers de référence de Résil implique le traitement de bases de données dont la taille cumulée atteint 12 milliards de lignes, mais la génération d'un univers de référence ne prend que 30 minutes environ. Ceci permet de simuler facilement l'impact de la modification de règles de calcul ou l'ajout de telle ou telle donnée. Les *data scientists* peuvent ainsi concevoir des traitements reproductibles permettant de mettre en relation plusieurs sources dont la mise en production est fluidifiée grâce aux infrastructures conteneurisées de l'Insee.

Un apport dans l'ensemble du cycle de valorisation de la donnée

Au-delà du défi technique, qui peut être relevé grâce à des infrastructures adaptées et à l'acquisition de compétences en ingénierie logicielle, l'adoption pérenne de la science et de l'ingénierie des données pose surtout la question de la répartition des rôles tout au long du cycle de vie d'une production statistique. Les premières étapes d'un projet de valorisation (la phase expérimentale) mobilisent déjà des compétences proches de celles requises pour un développement applicatif, mais la phase de production n'en exige pas moins : elle suppose un suivi continu, garant de la robustesse et de la pertinence des résultats dans le temps.

Ce besoin de suivi, qui mobilise à la fois des compétences techniques et une expertise métier, concerne également les chaînes de production de données. Alors que l'informatique classique repose sur les tests unitaires³⁹, le monde de la donnée exige des validations spécifiques : contrôles de cohérence technique (présence de variables selon la situation), vérifications statistiques (absence de valeurs aberrantes) ou encore comparaisons d'ordres de grandeur avec d'autres sources. Une partie de ces tests, ceux qui s'appuient sur des vérifications statistiques ou des validations métiers, sont généralement conçus par les *data scientists* et *data analysts* et intégrés dans des systèmes à plus grande échelle par les *data engineers*. Disposer des différents profils capables de les mettre en œuvre est ainsi déterminant pour les organisations désirant sécuriser la production et la diffusion de résultats fondés sur les données. Faire en sorte que ces profils partagent une culture et un langage communs est une condition de succès de la spécialisation des rôles. À cet égard, le nouveau nom du corps des administrateurs de l'Insee souligne l'élargissement de la culture scientifique des statisticiens publics aux enjeux propres au traitement de la donnée (**encadré 2**).

► Des progrès indéniables pour la statistique publique dans le respect de ses valeurs essentielles

Les transformations structurelles de l'univers de la donnée et de l'intelligence artificielle permettent de construire des analyses à partir de sources plus nombreuses et plus diverses, d'être en mesure de réagir plus vite aux situations exceptionnelles, d'améliorer la précision de certaines productions. Elles peuvent aussi conduire à quelques gains

³⁹ Méthode consistant, pour vérifier le bon fonctionnement d'un programme complexe, à effectuer des tests séparés sur des parties plus élémentaires (unités).

► Encadré 2. Le nouveau corps des « ingénieurs de la statistique, de l'économie et de la donnée »

Le métier de statisticien s'enrichit de nouvelles facettes. Il s'agit à présent de créer des outils de traitement de la donnée et non plus uniquement de les spécifier. Et il s'agit de veiller à ce que ces outils, et les chaînes de traitement qui découlent de leur assemblage, soient robustes, performantes, suffisamment souples pour s'adapter à des transformations de données ou des nouvelles contraintes, tout en assurant la traçabilité et la reproductibilité des résultats. Ces facettes, ou ces compétences nouvelles ne sont pas sans lien avec des compétences d'ingénieur, dont on attend à la fois des capacités de conception, mais aussi de mise en place d'outils de fabrication.

À cet égard, le changement du nom du corps des administrateurs de l'Insee en ingénieurs de la statistique, de l'économie et de la donnée (Insed)* est révélateur. Cette nouvelle dénomination, sans renier les fondamentaux d'origine (statistique et économie) reconnaît que les compétences des membres de ce corps, au service de la prise de

décision ou de la conception de chaînes de valeur de données, exigent une maîtrise de concepts scientifiques et opérationnels. Elle traduit également le caractère interministériel de ce corps, dont les agents essaient très largement au-delà de l'Insee, et même de la statistique publique. Dans de nombreux champs de l'action publique voire dans le secteur privé, ils apportent leur rigueur scientifique ainsi que leur expertise en analyse économique et en méthodes quantitatives de valorisation de la donnée.

Cette évolution de dénomination est également cohérente avec le fait que l'ENSAI et l'ENSAE, les deux écoles de formation initiale respectivement des attachés de l'Insee et des Insed, soient reconnues par la Commission des titres d'ingénieur comme écoles d'ingénieurs à part entière. Des enseignements consacrés spécifiquement aux enjeux propres à la mise en production et au MLOps sont présents dans le parcours de formation de ces écoles**.

* Ce changement est créé par le chapitre V du décret n° 2025-822 du 12 août 2025. Voir les références juridiques en fin d'article.

** Le contenu de ces enseignements est à l'image des compétences attendues des statisticiens publics : à l'interface de l'informatique et de la statistique. Voir par exemple, au lien suivant, le cours de mise en production de projets de data science construit par Romain Avouac et Lino Galiana pour le cursus d'ingénieurs de la donnée de l'ENSAE : <https://ensae-reproductibilite.github.io/website/>.

sur les délais de production⁴⁰. La statistique publique, même si elle a des exigences spécifiques, est confrontée à des défis proches de ceux du secteur privé. Elle a ainsi connu un mouvement comparable à ce dernier, qu'il s'agisse de l'évolution de ses outils de travail (logiciels et infrastructures) ou des modes d'organisation du travail.

Comme le reste du marché du travail, la statistique publique a commencé à recruter des *data scientists* au début des années 2010 pour en faire un profil très polyvalent, intervenant de l'exploration à la mise en production. Néanmoins, ce profil touche-à-tout montre ses limites dès lors qu'interviennent des enjeux techniques complexes liés à la production. Industrialiser des traitements, fiabiliser l'intégration de données dans un entrepôt ou s'assurer du bon fonctionnement d'une application de production nécessitent des compétences différentes de celles des *data scientists* en charge, par exemple, de la conception d'une méthode d'imputation de valeurs manquantes.

Dans le secteur privé, une réponse à cette complexification a consisté à identifier des rôles distincts (*data engineer*, *data scientist*, *data analyst*), une évolution facilitée par l'adoption progressive de technologies *cloud*. La statistique publique, qui a besoin de concilier innovation méthodologique et exigences de robustesse, traçabilité et pérennité des productions, a tout à gagner à s'inscrire dans ce mouvement. Elle s'est longtemps appuyée, comme de nombreuses organisations, sur une séparation nette entre

⁴⁰ Cela ne peut pas être des gains très significatifs, l'essentiel des facteurs explicatifs du temps de production étant ailleurs (Lamarche et al., 2025).

fonctions informatiques et fonctions statistiques. L'essor du *data scientist* a constitué un premier virage en faisant émerger un profil hybride, à l'interface entre ces deux mondes. Une seconde étape de rapprochement est l'ingénierie des données, qui offre un ensemble d'outils et de pratiques pour fiabiliser l'orchestration et la supervision de chaînes de production, depuis l'acquisition des données brutes à la mise à disposition de données transformées pour les utilisateurs. Pour tirer pleinement profit de cette nouvelle science, une partie des profils d'informaticiens doivent ainsi évoluer vers des profils de *data engineers*.

Dans un contexte où les sources de la statistique publique évoluent fréquemment et où le public demande des statistiques plus fines, plus nombreuses et diffusées plus rapidement, allier flexibilité et traçabilité passe certes par l'adoption de nouveaux outils, mais aussi par l'acculturation des acteurs de la statistique publique aux pratiques de l'ingénierie de données. De la même manière que la statistique publique est nourrie par des savoirs scientifiques en évolution, elle bénéficie de compétences techniques elles aussi diversifiées et en constante adaptation, au service de nos missions et dans le respect de nos valeurs.

► Fondements juridiques

- Règlement (UE) n° 2024/3018 du Parlement européen et du Conseil du 27 novembre 2024 modifiant le règlement (CE) n° 223/2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 05 février 2026]. Disponible à l'adresse : https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ%3AL_202403018.
- Règlement (UE) n° 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données). In : *site de l'Union européenne*. [en ligne]. [Consulté le 20 avril 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/FR/legal-content/summary/general-data-protection-regulation-gdpr.html>.
- Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (loi Informatique et Libertés). In : *site de Légifrance*. [en ligne]. [Consulté le 20 avril 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068624>.
- Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. In : *site de Légifrance*. [en ligne]. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>.
- Décret n° 2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques. In : *site de Légifrance*. [en ligne]. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000023619063/>.
- Décret n° 2025-822 du 12 août 2025 portant dispositions statutaires communes et particulières aux corps interministériels d'ingénieurs de l'État ayant vocation à exercer des fonctions d'encadrement supérieur. In : *site de Légifrance*. [en ligne]. [Consulté le 15 avril 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000052102609>.

► Bibliographie

- AVOUAC, Romain, FARIA, Thomas et COMTE, Frédéric, 2025. L'apport des technologies cloud pour industrialiser le processus d'innovation statistique. In : *Documents de travail*. [en ligne]. Juillet 2025. Insee. N° M2025-05. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/8614378>.
- BEAUREPAIRE, Camille, COQUET, François, LE SAOUT, Ronan, 2022. Qu'est-ce qu'un•e cadre de la statistique publique ? Analyse historique des formations et analyse statistique des fiches de poste. In : *Journées de méthodologie de statistique de l'Insee 2022*. [en ligne]. Mars 2022. Insee. [Consulté le 20 avril 2026]. Disponible à l'adresse : <https://hal.science/hal-05208159v1/document>.
- BERTHE, Juliette, 2025. Le défi des données pour l'inspection générale des affaires sociales. In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 29-47. [Consulté le 20 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546937?sommaire=8546949>.
- BOLLIET, Quentin, FLOYRAC, Aymeric, MAILLARD, Sophie et ROSENZWEIG, Agathe, 2025. L'inspection générale des finances et la science des données – Quelles méthodes pour quels usages ? In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 7-28. [Consulté le 20 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546934?sommaire=8546949>.
- BOTHOREL, Éric, 2020. Pour une politique publique de la donnée. In : *Rapport de la mission sur la politique publique de la donnée, des algorithmes et des codes sources*. [en ligne]. 23 décembre 2020. [Consulté le 05 février 2026]. Disponible à l'adresse : <https://www.info.gouv.fr/rapport/11979-rapport-sur-la-politique-publique-de-la-donnee-des-algorithmes-et-des-codes-sources>.
- BOURLANGE, Danielle, BRUNET, François, CHIGNARD, Simon et EIDELMAN, Alexis, 2021. Évaluation des besoins de l'État en compétences et expertises en matière de donnée. In : *Rapport de la mission auprès de la DINUM et de l'Insee*. [en ligne]. 15 juin 2021. [Consulté le 21 avril 2026]. Disponible à l'adresse : <https://www.vie-publique.fr/rapport/281669-besoins-de-etat-en-competences-expertises-en-matiere-de-donnees>.
- COMTE, Frédéric, DEGORRE, Arnaud et LESUR, Romain, 2022. Le SSPCloud : une fabrique créative pour accompagner les expérimentations des statisticiens publics. In : *Courrier des statistiques*. [en ligne]. 20 janvier 2022. Insee. N° N7, pp. 68-85. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/6035940?sommaire=6035950>.
- DAVENPORT, Thomas H. et Patil, DJ, 2012. Data Scientist: The Sexiest Job of the 21st Century. In : *Harvard Business Review*. [en ligne]. Octobre 2012. Volume 90, pp. 70-76. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>.

- DONDON, Alexis et LAMARCHE, Pierre, 2023. Quels formats pour quelles données ? In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 86-103. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635827?sommaire=7635842>.
- EDWARDS, Paul N., MAYERNIK, Matthew S., BATCHELLER, Archer L., BOWKER, Geoffrey C. et BORGMAN, Christine L., 2011. Science friction: Data, metadata, and collaboration. In : *Social Studies of Science*. [en ligne]. Volume 41, n° 5, pp. 667-690. [Consulté le 21 avril 2026]. Disponible à l'adresse : https://www.researchgate.net/publication/51874125_Science_Friction_Data_Metadata_and_Collaboration.
- GADOUCHE, Kamel, 2019. Le centre d'accès sécurisé aux données (CASD), un service pour la *data science* et la recherche scientifique. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. N° N3, pp. 76-92. [Consulté le 21 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254227?sommaire=4254170>.
- IOANNIDIS, John P. A., 2005. Why Most Published Research Findings Are False. In : *PLOS Medicine*. [en ligne]. 30 août 2005. Volume 2, n° 8. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://doi.org/10.1371/journal.pmed.0020124>.
- JOUBERT, Marie-Pierre, 2025. Les données de téléphonie mobile – Une source de connaissance sur la population et ses déplacements. In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 95-118. [Consulté le 21 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546945?sommaire=8546949>.
- KAPOOR, Sayash et NARAYANAN, Arvind, 2023. Leakage and the reproducibility crisis in machine-learning-based science. In : *Patterns*. [en ligne]. 8 septembre 2023. Volume 4, n° 9. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://doi.org/10.1016/j.patter.2023.100804>.
- KITCHIN, Rob, 2021. *Data Lives : How Data Are Made and Shape Our World*. 3 février 2021. Policy Press. ISBN : 9781529215144.
- LAMARCHE, Pierre et RIVIÈRE, Pascal, 2025. À propos du temps de production des statistiques publiques. In : *Courrier des statistiques*. [en ligne]. 15 décembre 2025. N° N14, pp. 95-118. [Consulté le 21 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8675585?sommaire=8667496>.
- LECLAIR, Marie, 2019. Utiliser les données de caisses pour le calcul de l'indice des prix à la consommation. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 61-75. [Consulté le 21 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254225?sommaire=4254170>.
- LEFEBVRE, Olivier, 2024. Le Répertoire Statistique des Individus et des Logements (Résil) – Un nouvel univers de référence pour les statistiques démographiques et sociales. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2024. Insee. N° N11, pp. 73-94. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8203040?sommaire=8203072>.

- LEFEBVRE, Olivier, 2025. Comment l'Insee protège-t-il les données qu'il collecte ? In : *Le blog de l'Insee*. [en ligne]. 23 avril 2025. [Consulté le 27 avril 2026]. Disponible à l'adresse : <https://blog.insee.fr/comment-l-insee-protege-les-donnees-collectees/>.
- LESUR, Romain, 2025. Sources de données privées : panorama et perspectives. In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. N° N13, pp. 73-94. [Consulté le 27 avril 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546943?sommaire=8546949>.
- MASLEJ, Nestor, FATTORINI, Loredana, PERRAULT, Raymond et al., 2025. In : *Artificial Intelligence Index Report 2025*. [en ligne]. 8 avril 2025. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://doi.org/10.48550/arXiv.2504.07139>.
- RUSSELL, Stuart et NORVIG, Peter, 2021. *Intelligence artificielle – Une approche moderne*. 26 novembre 2021. Pearson. ISBN : 978-2-3260-0221-0.
- TIGANI, Jordan, 2023. Big Data Is Dead. In : *MotherDuck*. [en ligne]. 07 février 2023. [Consulté le 21 janvier 2026]. Disponible à l'adresse : <https://motherduck.com/blog/big-data-is-dead/>.
- VILHUBER, Lars, 2020. Reproducibility and Replicability in Economics. In : *Harvard Data Science Review*. [en ligne]. 21 décembre 2020. Volume 2-4. [Consulté le 21 avril 2026]. Disponible à l'adresse : <https://hdsr.mitpress.mit.edu/pub/fgpmpj1l/release/5>.
- VOLLE, Michel, 1980. *Le métier de statisticien*. Hachette Littérature. L'esprit critique. ISBN 2-01-004529-7.

Extension du domaine de la statistique : les évaluations de politiques publiques à la Cour des comptes



Benjamin Belrhomari*, Lauriane Martzel**, Lisa Menez***
et Léo Zabrocki-Hallak****

Les juridictions financières – Cour des comptes et Chambres régionales et territoriales des comptes – à travers leur nouvelle mission d'évaluation des politiques publiques, font un usage croissant de l'analyse statistique comme élément de démonstration dans leurs rapports. Celle-ci leur permet d'aller plus loin dans leur mission de jugement du bon usage de l'argent public. Ce dispositif emprunte son cadre conceptuel et ses méthodes à la recherche – le recours aux outils quantitatifs de l'évaluation se fait de plus en plus à la façon des articles académiques – en conservant l'essentiel de ses propres normes et procédures. Le délibéré collégial et le contradictoire jouent, dans le cadre institutionnel des juridictions financières, le rôle de validation que la reproductibilité et la revue par les pairs assurent en recherche.

 The French financial courts—the *Cour des comptes*, i.e. France's supreme audit institution, and the Regional and Territorial Chambers of Accounts—through their new mission of evaluating public policies, increasingly rely on statistical analysis as supporting evidence in their reports. It allows them to enhance their capacity to assess the proper use of public funds. This approach borrows its conceptual framework and methods from research—the use of quantitative evaluation tools mirroring more and more the approach of academic papers—while the courts preserve the core of their own norms and procedures. In the institutional framework of financial courts, collegial deliberation and adversarial proceedings serve as the validation mechanism that reproducibility and peer review provide in research.

* Data scientist, Cour des comptes.
benjamin.belrhomari@ccomptes.fr

** Cheffe du département d'analyse et science des données, Cour des comptes.
lauriane.martzel@ccomptes.fr

*** Data scientist, Cour des comptes.
lisa.menez@ccomptes.fr

**** Adjoint à la cheffe du département d'analyse et science des données, Cour des comptes.
leo.zabrockihallak@ccomptes.fr

► Du contrôle des comptes à l'évaluation des politiques publiques

Évaluer une politique publique par la preuve statistique fait désormais partie intégrante du métier de magistrat financier. En 2026, la Cour des comptes a par exemple estimé l'effet des aides à la décarbonation de l'industrie sur les émissions de gaz à effet de serre, ainsi que sur les performances économiques des établissements industriels (Cour



Pendant deux siècles, la Cour des comptes a essentiellement « traduit en nombres » l'activité financière de l'État : dépenses, recettes et effectifs.



des comptes, 2026a). À cette fin, elle a mobilisé des méthodes statistiques d'appariement (Koumarianos et al., 2024) et d'analyse événementielle (Batut et Daviot, 2025), méthodes qui n'auraient sans doute jamais figuré dans un de ses rapports il y a vingt ans.

Cette évolution témoigne d'une transformation profonde de l'institution. Pendant deux siècles¹, la Cour des comptes a essentiellement traduit en nombres – pour reprendre l'expression d'Alain Desrosières (Desrosières, 2008) – l'activité financière de l'État : dépenses, recettes et effectifs.

Ce processus de quantification au service du pilotage de l'État participait d'un mouvement bien plus ancien – dès 1664, Colbert demandait aux maîtres des requêtes² de dresser un état complet des généralités³ (Gille, 1980) – et plus large, du Bureau de statistique de 1800 à l'Insee en 1946.

Cet outil de gouvernement descriptif s'est ensuite progressivement transformé en outil de preuve (Desrosières, 2008), avec le développement de la statistique dite « inférentielle »⁴, puis la dissémination du modèle causal de Rubin⁵ dans les sciences sociales. En fournissant un cadre unifié pour penser l'estimation d'effets causaux, ce modèle a nourri la révolution de la crédibilité en économie⁶ : les résultats fondés sur des expériences quasi-naturelles (Fougère et Jacquemet, 2019 ; Givord, 2010) ont progressivement irrigué l'action publique (Baiz, 2022).

La Cour des comptes n'est pas la seule institution à s'être ainsi transformée. Les institutions supérieures de contrôle britannique et américaine ont emprunté des chemins distincts, mais analogues de recours à la preuve statistique⁷.

¹ La Cour des comptes a été créée en 1807. Voir le lien : <https://www.ccomptes.fr/fr/cour-des-comptes/histoire-et-patrimoine>.

² Titre porté dans la France de l'Ancien Régime par les titulaires de certaines hautes fonctions judiciaires et administratives.

³ Dans la France de l'Ancien Régime, la généralité est une circonscription administrative.

⁴ Desrosières (2008) distingue deux acceptions du terme « statistique » : d'une part, la *Statistik* au sens historique, « science de l'État » descriptive et peu quantitative, visant à rassembler les savoirs utiles au gouvernement ; d'autre part, la statistique *inférentielle*, branche des mathématiques permettant d'induire, de tester et de généraliser des connaissances à partir de données observées, tant dans les sciences de la nature que dans les sciences sociales.

⁵ Voir Rubin (1974), puis sa clarification par Holland (1986).

⁶ Amélioration de la fiabilité des tests empiriques grâce aux progrès des méthodes économétriques et à l'augmentation des données disponibles. Voir Angrist et Pischke (2010).

⁷ Au Royaume-Uni, le *National Audit Office* (NAO) développe une approche de *value for money* (rapport qualité-prix) distincte de l'évaluation des politiques publiques au sens de la Cour. Voir par exemple le lien : <https://www.nao.org.uk/insights/a-planning-and-spending-framework-that-enables-long-term-value-for-money/#background-to-the-report>.

En France, ce mouvement dépasse les seules juridictions financières. En plus des services statistiques ministériels, les inspections générales (Bolliet et al., 2025 ; Berthe, 2025), des opérateurs publics comme la caisse nationale de l'assurance maladie ou la Haute Autorité de santé, et les assemblées parlementaires⁸ investissent à des degrés divers ce même terrain de la preuve quantitative. Elle permet de substituer à l'autorité de statut une démonstration chiffrée dont la force repose sur l'impersonnalité de la procédure (Porter, 1995).

Cette transformation a été initiée à la Cour des comptes par la révision constitutionnelle de 2008⁹ qui lui a confié une mission d'évaluation des politiques publiques. Elle s'accélère avec le projet stratégique Juridictions financières 2025¹⁰ et son objectif de consacrer 20 % des ressources à ces évaluations. La loi dite 3DS de 2022¹¹ a étendu cette mission aux Chambres régionales et territoriales des comptes. **L'évaluation s'ajoute ainsi aux trois missions historiques des juridictions financières : contrôler, juger, certifier.** Le recueil des normes professionnelles des juridictions financières (Cour des comptes et chambres régionales et territoriales des comptes, 2024) dispose que, pour cette

quatrième mission, « le travail d'analyse recherche les liens de causalité entre les actions publiques et leurs effets à court et à moyen terme » (disposition n° IV.30).

La révision constitutionnelle de 2008 a confié à la Cour des comptes une mission d'évaluation des politiques publiques.

Au sein des juridictions financières, cette transformation est présentée non comme une rupture, mais comme le prolongement naturel des missions historiques. Deux Premiers présidents successifs, Denis Migaud et Pierre Moscovici, ont inscrit l'évaluation dans une continuité

revendiquée : objectiver par les nombres est un exercice inscrit depuis bien longtemps dans l'ADN des magistrats financiers (Migaud, 2019) et constitue une exigence renouvelée (Moscovici, 2024). La logique sous-jacente est que le contrôle de la bonne utilisation des deniers publics appellerait un jugement sur les effets des politiques qu'ils financent.

Cette continuité revendiquée masque une transformation plus profonde. La Cour construit un régime de preuve spécifique : elle réutilise des données administratives conçues pour gérer l'action publique afin d'en mesurer les effets, convertit les résultats statistiques qui en sont issus et leurs incertitudes en jugements institutionnels qui n'en portent plus la trace, et substitue au mécanisme de validation de la recherche, i.e. la reproductibilité, celui des juridictions financières : le délibéré collégial et le contradictoire¹².

⁸ Au sein, à l'Assemblée, de la direction du Contrôle et de l'évaluation (cellule LexImpact) et, au Sénat, de la direction de la Législation et du Contrôle.

⁹ Loi constitutionnelle n° 2008-724 du 23 juillet 2008 de modernisation des institutions de la V^e République. Voir les références juridiques en fin d'article.

¹⁰ Voir le lien : <https://www.ccomptes.fr/fr/communiqués-presse/jf2025-le-projet-strategique-des-juridictions-financieres>.

¹¹ Loi n° 2022-217 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale. Voir les références juridiques en fin d'article.

¹² Les juridictions financières fondent leurs décisions sur trois grandes valeurs : l'indépendance de la chambre et des magistrats, la collégialité des décisions et la contradiction avec l'organisme contrôlé.

Des données du contrôle aux données d'évaluation

Dans le cadre des contrôles des comptes et de la gestion, les rapporteurs de la Cour des comptes sont principalement des consommateurs de résultats statistiques. Ils exploitent en premier lieu les productions du service statistique public – celles de l'Insee et des services statistiques ministériels – et les données des administrations contrôlées. Le rapport aux données est alors celui d'un commanditaire, parfois

littéralement : les rapporteurs demandent des chiffres à l'administration ou collectent des chiffres déjà produits, les reçoivent, les vérifient, les confrontent aux pièces justificatives. Le cas échéant, ils les recalculent à partir des pièces brutes pour produire les bons chiffres lorsqu'ils sont incomplets ou manquants.

Les questions posées par une évaluation appellent des données que les productions statistiques standard ne fournissent pas directement.

L'évaluation des politiques publiques modifie ce rapport en profondeur, en étendant largement le champ et la nature des sources mobilisées.

Les questions posées par une évaluation appellent des données que les productions statistiques standard ne fournissent pas directement, des analyses que personne n'a encore conduites, et des indicateurs qu'il faut parfois construire de manière ad hoc. **Le rapporteur passe de principalement consommateur de statistiques à producteur d'analyses quantitatives sur mesure.**

Trois évaluations récentes, exposées ci-dessous, illustrent ce déplacement. Elles soulèvent chacune un problème de données particulier :

- une grandeur d'intérêt inobservée qu'il faut prédire ;
- un phénomène qu'aucune source isolée ne documente directement, et qui impose d'apparier des données hétérogènes ;
- des systèmes d'information de gestion dont la qualité n'est pas garantie.

Pour évaluer la **politique de prévention et prise en charge des accidents vasculaires cérébraux (AVC)**, la Cour des comptes a entre autres étudié l'orientation adéquate des patients et la continuité des soins et de l'accompagnement (Cour des comptes, 2025a). Pour cela, elle a d'abord reconstitué les parcours de soins des patients à partir du système national des données de santé (SNDS), puis elle a imputé statistiquement le niveau de sévérité des séquelles pour le sous-ensemble des patients pour lesquels l'information est manquante (**encadré 1**). Elle a ainsi pu déterminer que, parmi les victimes d'AVC hémorragiques présentant des handicaps lourds [observés ou imputés] à la suite de ce dernier, près d'un tiers n'auraient pas bénéficié d'une hospitalisation dans une unité neurovasculaire. En conséquence, en réponse à la question sur l'efficacité et l'efficience du parcours d'ensemble du patient, la Cour a apporté une appréciation négative.

► Encadré 1. Reconstituer des parcours de soins à partir du SNDS

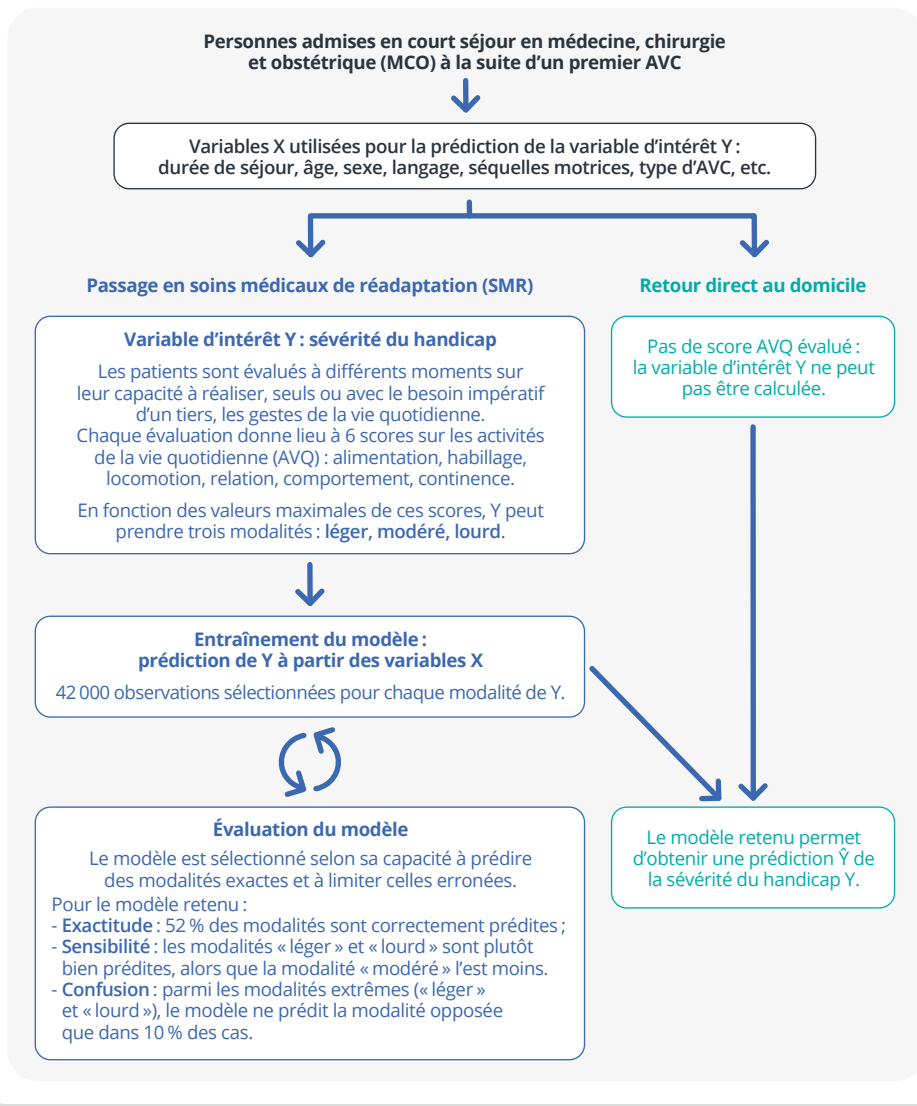
Le système national des données de santé (SNDS) est l'une des plus grandes bases de données médico-administratives au monde : elle relie les remboursements de l'assurance maladie, les séjours hospitaliers et les causes médicales de décès pour la quasi-totalité de la population française, à travers un modèle de données de plusieurs centaines de tables et des nomenclatures médicales spécialisées*. Son exploitation est loin d'être triviale pour produire des indicateurs pertinents à partir des données brutes.

Dans le cas des accidents vasculaires cérébraux (AVC), les parcours de soins à l'hôpital ont été reconstitués plus spécifiquement à partir du

programme de médicalisation des systèmes d'information (PMSI), dans le cadre duquel ont été collectées des informations relatives au patient, à son séjour, ainsi qu'au diagnostic et aux actes dont il a bénéficié. Cependant, les différentes bases du PMSI ne permettent pas d'identifier directement la sévérité des séquelles.

La Cour des comptes a alors estimé un modèle de classification du niveau de sévérité sur la seule sous-population pour laquelle il est observé : les bénéficiaires de soins médicaux de réadaptation (SMR). Elle a ensuite utilisé ce modèle pour imputer un niveau de sévérité aux autres patients pour lesquels l'information était manquante.

* Voir Berthe (2025) pour une présentation détaillée du SNDS.



Pour l'évaluation de la **politique de soutien au développement du biogaz**¹³ (Cour des comptes, 2025b), la Cour des comptes a reconstitué les performances économiques des exploitations et identifié celles qui interagissaient avec un méthaniseur (**encadré 2**). Elle a estimé que les exploitations impliquées dans la méthanisation ont accru leur excédent brut d'exploitation de l'ordre de 20 % par rapport aux exploitations similaires non impliquées. L'effet est significatif pour les exploitations détenant en propre un méthaniseur ; il est beaucoup plus incertain pour celles qui se limitent à l'approvisionnement ou à en récupérer le digestat.

► Encadré 2. Reconstituer les performances économiques des exploitations agricoles

Aucune source ne documente directement les résultats financiers des exploitations impliquées dans la méthanisation. La Cour des comptes a dû appairer quatre sources de données de nature différente :

- les déclarations fiscales des exploitations agricoles, pour reconstituer leurs résultats financiers ;

- le registre parcellaire graphique*, pour les localiser géographiquement ;

- les liens financiers entre unités légales, pour identifier les exploitations (co)propriétaires d'un méthaniseur ;

- le recensement général de l'agriculture**, pour repérer celles qui déclarent interagir avec un méthaniseur.

* Voir le lien : <https://artificialisation.developpement-durable.gouv.fr/bases-donnees/registre-parcellaire-graphique>.

** Opération organisée tous les dix ans environ par le Service de la statistique et de la prospective du ministère de l'Agriculture, de l'Agro-alimentaire et de la Souveraineté alimentaire.

Dans le cadre de travaux sur **le temps d'enseignement perdu par les élèves au collège** réalisés à la demande de la Défenseure des droits, la Cour des comptes a exploité les données sur le remplacement de courte durée d'enseignants issues des

systèmes d'information des établissements, dont notamment Pronote (Cour des comptes, 2025c). La qualité et l'homogénéité de ces données varient selon les établissements et les types d'absence et de remplacement. Leur exploitation a ainsi dû être complétée par une enquête nationale auprès des collèges pour documenter les pratiques de saisie dans Pronote et les écarts de mesure entre établissements.

La Cour des comptes doit utiliser et croiser des données particulièrement diverses, et pas nécessairement produites afin d'être exploitées statistiquement.

À l'image de ces trois cas, l'évaluation conduit systématiquement la Cour des comptes à mobiliser des données plus brutes que ce qu'elle aurait obtenu en sollicitant les administrations. Elle doit aussi utiliser et croiser des données particulièrement diverses, et pas nécessairement produites afin d'être exploitées statistiquement.

¹³ Le biogaz est un gaz renouvelable obtenu par la méthanisation, c'est-à-dire par la fermentation de matières organiques. Pour produire du biogaz, des déchets organiques sont rassemblés dans une cuve hermétique, que l'on appelle digesteur ou méthaniseur. Le digestat est le résidu du processus (l'autre produit étant donc le biogaz) : il peut également être valorisé, par exemple comme fertilisant.

Les systèmes d'information produits par les entités notifiées¹⁴ – déclarations fiscales de la Direction générale des Finances publiques (DGFIP), séjours hospitaliers du PMSI, etc. – sont des dispositifs de **lisibilité** pour les pouvoirs publics au sens de Scott (1998) : ils encodent la réalité sociale dans des catégories standardisées pour permettre une gestion efficace par l'État. Ils constituent ainsi un instrument de gouvernement (Desrosières, 2008) pour, dans les exemples évoqués, administrer l'impôt ou gérer l'hôpital. La Cour des comptes réutilise ces données pour évaluer et les détourne ainsi de leur finalité originelle. Elle se heurte systématiquement aux conséquences de la lisibilité créée pour la puissance publique : la simplification qui permet la gestion ne capture qu'imparfaitement ce qu'il faudrait mesurer pour l'évaluation et amène souvent à des biais de couverture (comme dans l'exemple du PMSI pour la mesure de la sévérité des séquelles).

Le service statistique public exploite également directement des données brutes extraites des systèmes d'information administratifs¹⁵. Néanmoins, les juridictions financières s'en distinguent par un recours généralisé à ce type d'exploitation¹⁶, qui est rendu possible par l'utilisation de leur droit de communication (**encadré 3**).

Le fonctionnement des juridictions financières explique cette différence d'usage. **Une équipe de rapporteurs travaille ponctuellement et pour quelques mois sur un sujet sans notion de production récurrente des mêmes indicateurs homogènes au cours du temps.** Cet « hapax statistique »¹⁷ implique de croiser des données disparates qui n'ont pas été collectées pour répondre à l'objet du rapport, et de chaîner des millésimes d'enquêtes qui ont parfois sensiblement évolué au cours du temps. En contrepartie, il n'impose pas les contraintes de production et d'industrialisation des processus propres au système statistique public (SSP)¹⁸. Les procédures des juridictions financières l'empêchent, dans la mesure où l'essentiel des données est obtenu dans le cadre d'une instruction en particulier. Or, les données ainsi recueillies ne peuvent légalement pas être réutilisées pour un autre rapport.

Instruire et estimer : co-construire une analyse quantitative

Le recours accru à la preuve statistique transforme les compétences que la Cour des comptes mobilise. Un effort de formation des rapporteurs qui instruisent les rapports est engagé dès 2014 (Hayez, 2015). En 2016, la Cour crée le **département d'analyse et science des données** (DASD), rattaché à la direction du numérique, pour appuyer les équipes d'instruction sur les travaux comportant une dimension quantitative significative. Le DASD est composé aujourd'hui de 12 *data scientists*, 2 *data analysts*, ainsi que de la cheffe de département et de son adjoint.

¹⁴ La notification désigne ici l'acte formel par lequel la Cour des comptes informe les administrations concernées de l'ouverture d'une procédure et de ses modalités. En matière de contrôle, la notification est adressée au dirigeant de l'organisme contrôlé et, le cas échéant, à ses autorités de tutelle. En matière d'évaluation de politiques publiques, où aucune administration n'est individuellement contrôlée, elle est adressée aux représentants des principales parties prenantes.

¹⁵ Voir de manière générale Koumarianos et Rivière (2025) et, pour un exemple précis, Renne (2018).

¹⁶ C'est le cas plus largement des instances de contrôle et d'évaluation qui se sont dotées d'un pôle de science des données : voir par exemple Berthe et al. (2025) pour l'inspection générale des affaires sociales et Bolliet et al. (2025) pour l'inspection générale des finances.

¹⁷ À l'image d'un mot dont il n'existe qu'une seule occurrence dans un texte.

¹⁸ Si la Cour est une institution publique faisant de la « statistique », elle ne fait pas de la « statistique publique » (Isnard, 2018).

► Encadré 3. Accès aux données

Les juridictions financières disposent de prérogatives spécifiques pour accéder aux données nécessaires à leurs missions de contrôle et d'évaluation. Ces accès sont encadrés par le code des juridictions financières (CJF)* et respectent les principes de protection des données.

Pour la Cour des comptes, l'article L141-5 du CJF permet d'accéder à tous les documents, données et traitements relatifs à la gestion des services et organismes soumis à son contrôle, sans que le secret protégé par la loi puisse être opposé. Les modalités pratiques sont précisées à l'article R141-4 : les rapporteurs sont autorisés à demander la communication des documents sous divers formats, ainsi que l'accès direct aux systèmes d'information. Des conventions peuvent être conclues pour déterminer les modalités techniques d'accès continu aux systèmes d'information.

Les Chambres régionales et territoriales des comptes bénéficient d'un cadre similaire, avec l'article L241-5 du CJF étendant leur droit de

communication aux collectivités publiques, établissements publics et autres organismes soumis à leur contrôle. L'article R241-6 détaille les modalités pratiques, incluant également la possibilité de conclure des conventions pour un accès continu aux systèmes d'information.

Les juridictions financières agissent dans le cadre de leur mission de service public, ce qui fonde la licéité de leurs traitements de données à caractère personnel, conformément à l'article 6, 1^{er} alinéa, point e du règlement général sur la protection des données (RGPD)*. Elles sont considérées comme des « tiers autorisés », disposant d'un droit de communication des documents et informations, y compris des données à caractère personnel, détenus par les organismes contrôlés. Les demandes d'accès aux données se font par écrit et lors des entretiens d'ouverture de contrôle, en respectant le « principe de minimisation » (au regard du but visé). Les données collectées sont conservées de manière sécurisée et sont détruites à l'issue du contrôle.

* Voir les références juridiques en fin d'article.

La Cour intègre de façon croissante parmi les rapporteurs des profils issus de la statistique publique et de la recherche : elle emploie des chercheurs, des ingénieurs de la statistique, de l'économie et de la donnée et recrute directement au grade de conseiller



Les données ne tombent jamais du ciel : elles sont le produit d'un travail d'identification, d'accès, de contrôle, de compréhension et, enfin, de mise en qualité pour les rendre exploitables.



maître des experts en évaluation des politiques publiques. Au sein du DASD, les profils associent méthodes quantitatives et sciences sociales, ce qui permet l'aller-retour entre les enjeux d'un rapport et la conception de l'analyse statistique.

Cet appui couvre l'ensemble de la chaîne, du travail sur les données à la co-construction des questions évaluatives et des choix méthodologiques.

La première étape, et souvent la plus longue, est le travail de préparation sur les données elles-mêmes. Les données ne tombent en effet jamais du ciel (Denis, 2018) : elles sont le produit d'un travail d'identification (pour déterminer lesquelles

sont les plus pertinentes à mobiliser pour un rapport), puis d'accès, de contrôle, de compréhension et, enfin, de mise en qualité pour les rendre exploitables.

Du fait des spécificités évoquées précédemment – croisement de nombreuses sources hétérogènes, données de gestion, décalage entre besoins des juridictions financières et finalités qui ont présidé à la constitution des données – ce processus représente une part importante de l'activité d'un *data scientist*¹⁹. Parfois, les données nécessaires sont

¹⁹ Sur les métiers de la *data*, voir l'article de Galiana et Lefebvre dans ce même numéro.

inexistantes, ou se révèlent après analyse insuffisamment fiables ou pas exploitables. La Cour des comptes est amenée à recommander régulièrement d'améliorer la capacité de l'État à observer des objets pertinents. Par exemple, dans son rapport sur les droits de succession (Cour des comptes, 2024a), elle recommande au ministère de l'Économie de réaliser une étude statistique relative aux droits de mutation à titre gratuit²⁰ avant toute évolution législative.

Contrairement aux étapes ultérieures de production de statistiques descriptives ou d'estimation de modèles statistiques, ce travail est peu visible par les rapporteurs car il produit peu ou pas de traces : tableaux, graphiques, etc. Nous allons voir qu'il peut pourtant constituer un résultat en lui-même.

Pour l'évaluation du pacte Dutreil, dispositif permettant de réduire les droits de succession lors de la transmission d'une entreprise, la Cour des comptes a établi un portrait inédit des entreprises concernées par le pacte (Cour des comptes, 2025d). D'un point de vue statistique, la méthode employée est triviale, puisqu'il s'agit de calculer des moyennes ou d'autres statistiques descriptives sur ces entreprises. La nouveauté réside bien dans le travail de mise en qualité et de croisement des données. Celui-ci a consisté d'abord à identifier les unités légales bénéficiaires du pacte dans les données de gestion brutes de la DGFIP. Puis, en lien avec l'Institut des politiques publiques (IPP), des unités légales transmises en première approche hors pacte Dutreil, au regard des données de la DGFIP, ont été reclassées à partir de leurs caractéristiques observables comme ayant été en réalité probablement transmises via ce pacte (Bach et al., 2025). Ces entités, souvent des holdings, ont ensuite été reliées aux unités légales ayant une activité économique réelle qu'elles détiennent via leurs liens capitalistiques.

Les données fiscales et sociales des holdings et de leurs filiales ont alors permis d'en dresser le portrait économique. Ce travail de réexamen des données brutes est un des facteurs qui ont amené la Cour des comptes à estimer un coût du dispositif de l'ordre de 5,5 milliards d'euros en 2024, contre 0,8 milliard d'euros d'après l'administration.

L'appui des data scientists ne se réduit pas au traitement des données. Ils sont généralement associés à la construction des questions évaluatives.

L'appui du DASD ne se réduit pas au traitement des données. Le *data scientist* est généralement associé à la construction des questions évaluatives. Il détermine avec les rapporteurs si une question de politique

publique est formulable en termes statistiques, si les données disponibles permettent d'y répondre, quelle méthode serait adaptée, et quel degré de certitude on peut en attendre. **Le *data scientist* et les rapporteurs co-construisent l'analyse** : d'abord lors de l'élaboration d'une note de faisabilité préalable à l'évaluation, puis au fil de l'instruction, dans un dialogue continu entre analyse statistique et construction du rapport.

La distinction fonctionnelle entre *data scientist* et rapporteurs pourrait laisser penser que le premier se cantonne aux aspects techniques. Or, la connaissance du fond des sujets et la capacité à traduire des besoins d'instruction en analyses statistiques et,

²⁰ Taxes dues à l'État lors de transferts de biens sans contrepartie financière, comme les donations et les successions, après déduction d'abattements.

réciroquement, des résultats statistiques en informations pertinentes pour le rapport, sont essentielles. Une analyse techniquement irréprochable mais ne répondant pas aux questions évaluatives ne produit pas de résultat utile. De même, considérer la contribution du *data scientist* comme une prestation externalisée qu'il livrerait aux rapporteurs est improductif. Sans intégration dans l'équipe d'instruction, le *data scientist* ne peut saisir les enjeux sous-jacents aux demandes et donc y répondre de manière pertinente. Cette intégration produit une acculturation réciproque : les rapporteurs se familiarisent avec les possibilités et les limites des méthodes quantitatives, les *data scientists* acquièrent les réflexes de l'instruction et la connaissance du cadre institutionnel dans lequel les résultats seront mobilisés. Elle reste hétérogène, selon les sujets et les cultures des Chambres, car la frontière professionnelle (Abbott, 1988) est encore en cours de stabilisation entre rapporteurs et *data scientists*.

Cet environnement de travail distingue les juridictions financières d'autres institutions dans lesquelles exercent des *data scientists*. Dans un laboratoire de recherche, le chercheur maîtrise généralement son sujet en profondeur, sans distinction entre les rôles de rapporteur et de *data scientist*, et dispose d'un temps plus long ; dans le SSP, le statisticien travaille souvent sur des productions récurrentes ou sur des sujets proches sur lesquels il se spécialise pendant l'occupation de son poste. À la Cour des comptes, le *data scientist* est amené à changer régulièrement de sujet, avec à chaque fois un nouveau contexte, des données qu'il découvre potentiellement et dont la documentation est parfois lacunaire, et un calendrier contraint par celui de l'instruction.

► Un régime de preuve hybride

Mesurer, estimer, convaincre

La mobilisation de la preuve statistique dans les rapports d'évaluation d'une politique recouvre **un spectre de méthodes large**. Elle ne se limite pas forcément à la seule identification et estimation d'un paramètre mesurant quantitativement l'effet causal d'une politique publique sur un résultat. Le « Guide des EPP »²¹, interne aux juridictions financières, prévoit notamment qu'il est très fortement recommandé de mobiliser conjointement des méthodes qualitatives et des méthodes quantitatives. De façon schématique, les méthodes quantitatives permettent en effet d'estimer l'amplitude d'un phénomène, là où les méthodes qualitatives aident à en comprendre les ressorts²². Le jugement évaluatif porte sur l'efficacité, l'efficience, la cohérence interne des objectifs et sa cohérence externe avec d'autres interventions, et sur l'utilité de la politique publique. Il amène les juridictions financières à étudier des dimensions peu prises en compte dans la recherche académique, telles que les questions de gouvernance, de formalisation des objectifs ou de coût²³. Si l'ensemble de ces critères se prête à l'emploi de preuves statistiques, l'efficacité et l'efficience en sont les principales concernées, que ce soit pour les mesurer ou pour inférer un effet causal.

²¹ EPP pour évaluation de politique publique.

²² Voir également Cartwright (2010) pour la mobilisation de différentes méthodes.

²³ Le SSP peut aussi être amené à aborder les questions de gouvernance. Voir par exemple Arbeláez Ayala et al. (2025).



Un indicateur descriptif peut parfois suffire en lui-même à étayer un jugement évaluatif, quelle que soit la forme sous laquelle il est présenté (chiffre, tableau, carte, graphique).



Selon l'objet de l'évaluation, cette dernière peut être comprise comme un exercice de « mesure », où, sans passer par un modèle, un indicateur descriptif suffit en lui-même à étayer un jugement évaluatif, quelle que soit la forme sous laquelle il est présenté (chiffre, tableau, carte, graphique)²⁴. La définition de cet indicateur, puis son calcul effectif à partir des données, constituent alors le cœur de l'évaluation. Les juridictions financières recourent pour cela le plus possible aux **données existantes**, principalement produites par le SSP et les administrations notifiées.

Dans l'évaluation de la prévention et de la prise en charge du cancer du sein, réalisée à la demande de la commission des affaires sociales de l'Assemblée nationale, la Cour des comptes a étudié la prise en charge hospitalière des patientes (Cour des comptes, 2026b)²⁵. Il s'agissait d'évaluer si les objectifs fixés par les pouvoirs publics étaient atteints, et notamment celui de l'accessibilité géographique et sociale :

- D'une part, elle a calculé l'accessibilité potentielle localisée à la chirurgie du cancer du sein, un indicateur qui rapporte l'offre (ici d'intervention chirurgicale) à la demande, en intégrant les durées de trajet réelles plutôt que les découpages administratifs (Barlet et al., 2012).
- D'autre part, en se fondant sur les données du SNDS, elle a également estimé le reste à charge moyen des femmes atteintes d'un cancer du sein, après remboursement par l'assurance maladie obligatoire et avant remboursements éventuels par l'assurance maladie complémentaire. Celui-ci est estimé à 1 549 euros en 2026, sachant que les consommations de soins sont analysées sur une période de deux ans à compter de la date de début du cancer du sein.

Pour le cas où les données nécessaires à la mesure recherchée sont inexistantes, la Cour des comptes a également développé un recours croissant aux **sondages** pour recueillir directement la parole des bénéficiaires et des acteurs des politiques publiques évaluées (**encadré 4**)²⁶. Ce recours aux entretiens et aux sondages ad hoc pour compléter les données administratives vise à récupérer la *métis*²⁷ (Scott, 1998), le savoir pratique local que les dispositifs de lisibilité étatiques éliminent par construction.

La « mesure » ne suffit cependant pas toujours à fonder un jugement évaluatif. Lorsque la question porte sur l'efficacité d'un dispositif (a-t-il produit un effet ?), il faut passer de la description à l'inférence causale, c'est-à-dire quantifier la différence entre les résultats observés et ce qu'ils auraient été en l'absence de la politique publique²⁸.

²⁴ Si l'on distingue généralement statistiques descriptives et inférence causale, ici la mesure sert d'abord d'élément de preuve statistique pour évaluer une politique publique, et non pas à décrire en soi.

²⁵ L'analyse des restes à charge a été publiée indépendamment par la suite (Chariot, 2026).

²⁶ Cette pratique fait des juridictions financières un (petit) producteur de données.

²⁷ Terme grec qui désigne cette capacité de l'intelligence qui correspond, non pas à l'abstraction, mais à l'efficacité pratique.

²⁸ Le contrefactuel au sens du modèle de Rubin (Rubin, 1974).

► Encadré 4. Le recours au sondage dans les juridictions financières

Avec la montée en régime des évaluations de politiques publiques et des consultations citoyennes, les juridictions financières déploient de façon croissante des sondages pour recueillir le point de vue des usagers sur la mise en œuvre effective d'un dispositif. Un questionnaire est en effet l'occasion de donner la parole aux acteurs et bénéficiaires de la politique publique pour les entendre sur les freins tout en identifiant des leviers d'action potentiels. Les informations collectées permettent également de mieux appréhender certains mécanismes à l'œuvre pour compléter les analyses quantitatives produites à partir de données administratives.

Le DASD intervient à tous les niveaux de la construction d'un sondage pour améliorer sa qualité scientifique :

- rédaction du questionnaire pour éviter certains biais dans la formulation des questions ;

- conception du plan de sondage pour s'assurer de sa représentativité et de sa précision ;
- prise en compte de la non-réponse à l'aide de méthodes de redressement ;
- analyse et mise en valeur des résultats.

À chaque étape, les *data scientists* œuvrent afin que les résultats obtenus soient généralisables à la population d'intérêt et fournissent des mesures fiables et souvent inexistantes par ailleurs.

Le recueil de la parole s'inscrit également dans une volonté plus large d'ouverture aux citoyens, dont témoigne aussi la plateforme de participation citoyenne lancée en 2022 par la Cour des comptes*. Celle-ci permet à tous de proposer des objets de contrôle ou d'évaluation, comme par exemple la politique en faveur de l'inclusion dans l'emploi des personnes en situation de handicap.

* Voir le lien : <https://participationcitoyenne.ccomptes.fr/>.

Pour l'évaluation du fonds régional de garantie²⁹ de la région Hauts-de-France (Cour des comptes, 2024b), les trajectoires des entreprises bénéficiaires de la garantie de l'opérateur public ont été comparées au cours du temps à celles d'entreprises de mêmes caractéristiques, mais non bénéficiaires (méthode dite d'analyse événementielle). En effet, pour que la différence éventuelle entre les trajectoires des entreprises bénéficiaires et non bénéficiaires puisse être attribuée à l'effet de l'obtention de la garantie, plutôt qu'à d'autres différences entre les deux groupes, il est nécessaire que les caractéristiques des entreprises bénéficiaires et non bénéficiaires soient similaires. Une méthode d'appariement sur score de propension (Givord, 2010) a été employée pour que chaque entreprise bénéficiaire soit comparée à l'entreprise non bénéficiaire la plus ressemblante possible au niveau des caractéristiques observables avant l'obtention de la garantie. Comme pour l'ensemble des évaluations des juridictions financières, le travail préalable est considérable : établir une liste exploitable d'entreprises bénéficiaires à partir des données de gestion de l'opérateur, avec leurs identifiants (numéros Siren³⁰) et leurs dates de traitement, puis les croiser avec des bases de données issues de la statistique publique (FARE³¹). La méthode est adaptée d'un travail académique (Gazaniol et Lê, 2021), ajustée aux contraintes du cas d'espèce – ici, un nombre d'entreprises bénéficiaires très faible.

Les évaluations visent essentiellement à la validité interne, car elles portent sur un dispositif spécifique. Elles éclairent la décision publique et le citoyen sur ce dispositif en particulier, sans nécessairement répondre à une question plus générale. Dans l'exemple précédent, la question est : « le fonds de garantie des Hauts-de-France a-t-il eu un effet ? ». Il ne s'agit pas de la question « les fonds de garantie régionaux sont-ils efficaces en général ? », examinée à partir du cas particulier des Hauts-de-France.

²⁹ La loi permet aux collectivités territoriales d'accorder leur garantie (sous certaines conditions) aux emprunts contractés par des personnes de droit privé. Les fonds régionaux de garantie peuvent intervenir en faveur des entreprises lors de leur création, lors de leur transmission ou lors de la mise en œuvre de programmes spécifiques d'investissements.

³⁰ Voir le lien : <https://www.insee.fr/fr/metadonnees/definition/c2047>.

³¹ Voir le lien : <https://www.casd.eu/source/statistique-structurelle-annuelle-dentreprises-issue-du-dispositif-esane/>.

Ces exemples illustrent un arbitrage récurrent. Les méthodes d'évaluation d'impact mobilisées sont issues de la recherche académique, où elles ne sont utilisées, publiables et publiées que lorsque les meilleures conditions de qualité méthodologique sont réunies. La Cour des comptes doit arbitrer entre robustesse méthodologique et utilité

Les juridictions financières peuvent ainsi publier des résultats non concluants.

d'évaluer une politique publique donnée. Dans l'exemple précédent, le rapport reconnaît les limites méthodologiques inhérentes à l'utilisation d'échantillons si petits.

Contrairement à la recherche, la Cour des comptes n'est pas soumise au biais de publication (Ioannidis et al., 2017). Un résultat à interpréter avec prudence, mais éclairant sur un dispositif que

personne n'a évalué, peut être plus utile au débat public qu'une absence de résultat, au motif que les conditions d'identification ne sont pas pleinement satisfaites. Les juridictions financières peuvent ainsi publier des résultats négatifs (ce dispositif ne produit pas d'effet mesurable, ou les données ne permettent pas de conclure), comme pour l'évaluation du fonds régional de garantie des Hauts-de-France. Le rapport le reconnaît explicitement et conclut que les effets du dispositif sont difficiles à mesurer précisément avec les données disponibles.

Cet arbitrage n'a rien de mécanique : il se négocie au cas par cas entre *data scientists* et rapporteurs, puis est tranché lors de l'examen en délibéré du projet de rapport par la Chambre (*figure*).

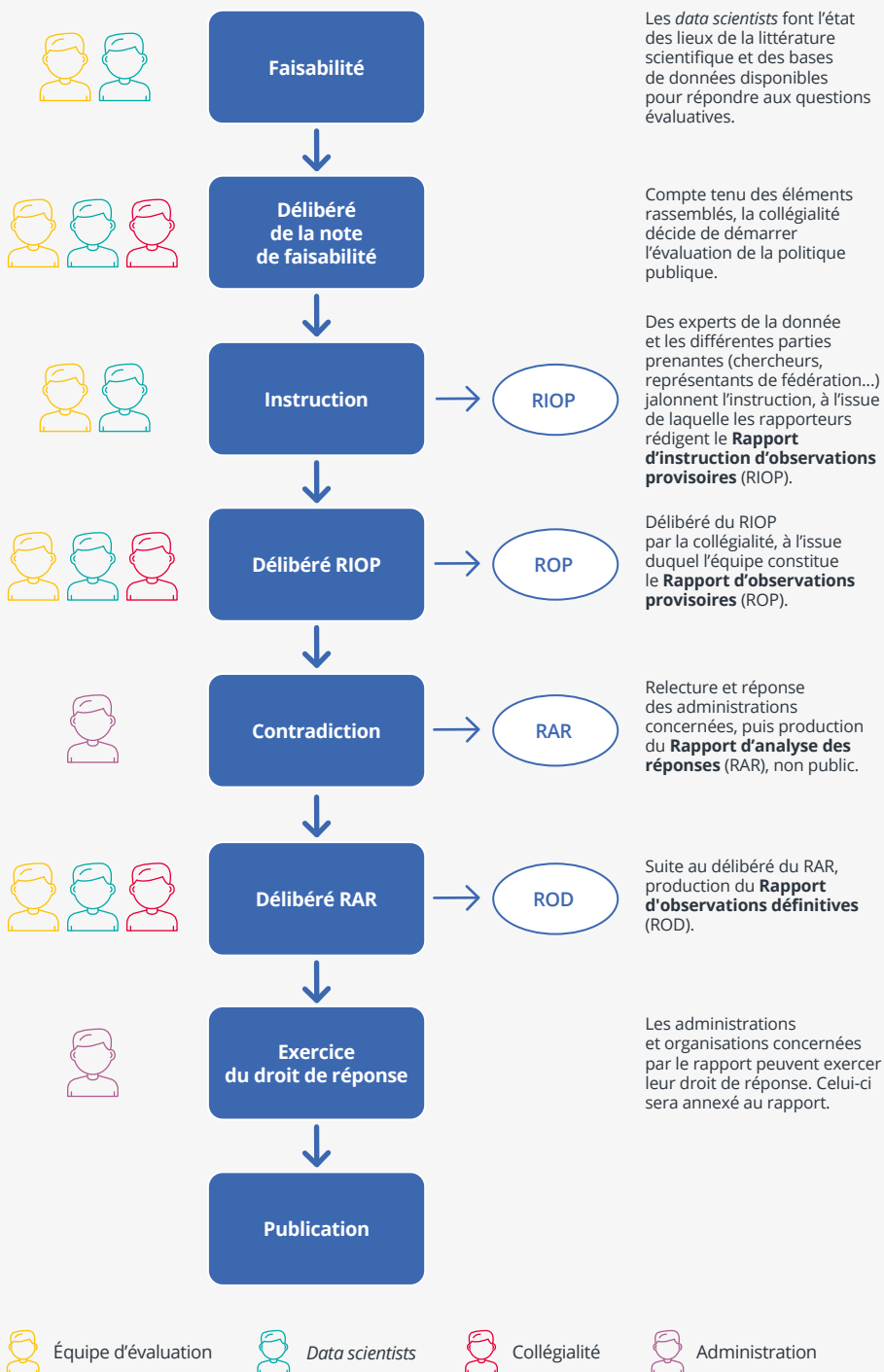
Transparence du résultat, secret de l'instruction

Les *data scientists* arrivent à la Cour des comptes avec une culture professionnelle issue de la recherche ou de la statistique publique, et y rencontrent une culture institutionnelle qui obéit à des logiques différentes. **Dans les juridictions financières, c'est lors du délibéré que la preuve statistique acquiert sa forme finale.** La version définitive d'un rapport est produite par la collégialité³², qui traduit les estimations statistiques et le projet de rapport des rapporteurs en observations et recommandations de la Cour des comptes. Le délibéré, et non la reproductibilité ou la revue par les pairs (Chaleix et Durr, 2024), tient lieu de mécanisme de validation. La crédibilité du résultat émerge de la qualité de l'examen collectif auquel il a été soumis : c'est une forme de rationalité procédurale et communicationnelle³³. Dans ce processus, les inscriptions quantitatives (résultats de modèles, intervalles d'incertitudes), comme les autres éléments produits par l'instruction, sont traduites en énoncés juridiques définitifs (« la Cour observe que »). Ce mécanisme de transformation des inscriptions techniques en droit est analogue à celui que Latour a documenté au Conseil d'État (Latour, 2002).

³² Article L141-1 du CJF. La Cour des comptes considère ce processus délibératif et collégial comme le mécanisme qui garantit l'objectivité et l'impartialité des travaux (Cour des comptes, 2025e).

³³ Voir Baldacchino et Jourdain (2025), qui mobilisent le cadre habermassien de la rationalité communicationnelle.

► **Figure - Processus d'élaboration d'un rapport d'évaluation de la Cour des comptes**



Il peut arriver que l'analyse statistique ne confirme pas l'hypothèse de travail de l'équipe d'instruction, ou qu'elle produise un résultat ambigu là où le rapport appellerait une conclusion tranchée. Le *data scientist* peut aussi estimer qu'un résultat est trop fragile pour figurer dans le rapport quand l'équipe de rapporteurs considère qu'il apporte un éclairage utile malgré ses limites. Ces discussions sont inhérentes à tout travail quantitatif, mais elles prennent une forme particulière dans les juridictions financières : les rapporteurs décident de ce qui figure dans le projet de rapport, et c'est la collégialité, lors du délibéré, qui tranche in fine, parfois à rebours des propositions émises par les rapporteurs. Le rôle du *data scientist* est alors moins de défendre une position méthodologique que de rendre explicites les conséquences des différents choix possibles. Cela suppose à la fois une compréhension suffisante des enjeux du rapport pour formuler ces conséquences en termes pertinents pour les rapporteurs puis pour la collégialité, ainsi qu'une familiarité avec le cadre institutionnel dans lequel le résultat sera discuté puis publié. En amont du délibéré, les analyses quantitatives peuvent être présentées lors des comités d'accompagnement, qui rassemblent des

parties prenantes de la politique publique et des experts, dont certains peuvent être chercheurs ou statisticiens. Ce dispositif permet un examen des choix méthodologiques par des experts extérieurs avant le délibéré.

À la suite du délibéré, les administrations notifiées sont destinataires des observations provisoires et peuvent y répondre. Ce processus de contradiction constitue un test méthodologique exigeant.

À la suite du délibéré, les administrations notifiées sont destinataires des observations provisoires et peuvent y répondre. De plus, leur réponse aux observations définitives est publiée avec le rapport. Ce processus de contradiction constitue un test méthodologique exigeant : les administrations notifiées peuvent contester un indicateur,

un périmètre, un choix de modélisation. Le contradictoire est toutefois une mise à l'épreuve antagoniste : l'administration conteste avant tout les résultats qui la mettent en cause et rarement ceux qui lui sont favorables. Les faiblesses méthodologiques d'un résultat allant dans le sens de l'administration risquent de ne pas être relevées – un angle mort que la relecture interne au DASD et les comités d'accompagnement contribuent à compenser.

Une fois adopté et publié dans un rapport des juridictions financières, le chiffre acquiert un certain statut d'autorité. Ces chiffres sont généralement présentés sans mesure d'incertitude formelle : les rapports communiquent les limites de manière qualitative, dans le texte, sans intervalles d'incertitude. Pourtant tout choix méthodologique est discutable – et discuté lors de l'adoption du projet de rapport en délibéré. Le reste à charge de 1 549 euros estimé pour le cancer du sein dépend, par exemple, de la convention retenue pour la durée du suivi, ici deux ans après la date de début du cancer. Une autre convention produirait un montant différent. Ainsi, une administration peut contester, non pas le calcul lui-même, mais le périmètre ou la définition retenus. La Cour des comptes doit alors défendre non seulement l'exactitude du chiffre, mais la pertinence de la convention qui le sous-tend.

La trajectoire de la Cour des comptes ne se résume pas au passage d'un outil de gouvernement à un outil de preuve. Les recommandations qui concluent chaque rapport reconvertissent la preuve statistique en instrument de gouvernement. Le coût du pacte Dutreil estimé à 5,5 milliards d'euros ne reste pas un résultat : il fonde la recommandation de mieux cibler le dispositif. Le chiffre sert in fine à recommander d'agir, pas à savoir.

Cette production de preuve s'inscrit dans un cadre institutionnel singulier. Elle concilie le secret de l'instruction avec une transparence du résultat : **la préparation des rapports et les échanges avec les organismes notifiés ne sont pas publics mais l'ensemble des rapports sont publiés.**

Cette combinaison crée des tensions spécifiques dès lors que les juridictions financières importent la preuve statistique dans leurs méthodes. En recherche comme en statistique publique, la crédibilité d'un résultat quantitatif repose pour partie sur la possibilité de le reproduire : description des choix méthodologiques, publication des programmes, accès aux données lorsque cela est possible. Le secret de l'instruction ne l'interdit pas en droit, mais les juridictions financières y renoncent de facto.

Cette réticence n'est pas tant juridique qu'organisationnelle et culturelle. La publication de notes méthodologiques et de codes correspond pourtant aux valeurs de transparence de la Cour des comptes ; elle n'est pas encore systématique, faute de procédures arrêtées. Si les rapports sont accompagnés d'un « dossier-liasses-rapport » consignait les pièces justificatives, la définition formelle de ce que constituent ces pièces pour des travaux quantitatifs – méthodes, programmes, environnement, données brutes ? – reste à établir. **La Cour des comptes substitue le délibéré collégial et le contradictoire à la reproductibilité comme mécanisme de validation.**

Les travaux quantitatifs produits par un *data scientist* font l'objet d'une relecture au sein du DASD, qui s'est progressivement structurée pour garantir la solidité technique des analyses. Plus largement, chaque évaluation consolide les compétences et la culture quantitatives au sein de l'institution, et installe des exigences méthodologiques plus explicites.



La preuve statistique sert in fine à la Cour des comptes à recommander aux administrations d'agir : le chiffre produit pour évaluer est reconverti en instrument de gouvernement.



La preuve statistique sert in fine à la Cour des comptes à recommander aux administrations d'agir : le chiffre produit pour évaluer est reconverti en instrument de gouvernement. Ainsi la crédibilité de cette preuve n'est pas une question épistémologique abstraite mais une condition de l'efficacité de ses recommandations. Institution singulière, la Cour des comptes n'a pas à adopter les normes de la recherche académique ou de la statistique publique. Le régime de preuve hybride qu'elle construit ne sera pour autant pleinement convaincant que si la preuve elle-même est ouverte à l'examen.

► Fondements juridiques

- Loi constitutionnelle n° 2008-724 du 23 juillet 2008 de modernisation des institutions de la V^e République. In : *site de Légifrance*. [en ligne]. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000019237256/2025-03-08/>.
- Règlement (UE) n° 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données). In : *site de l'Union européenne*. [en ligne]. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32016R0679>.
- Code des juridictions financières. In : *site de Légifrance*. [en ligne]. Mis à jour le 21 février 2026. [Consulté le 19 mai 2026]. Disponible à l'adresse : https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006070249.
- Loi n° 2022-17 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale. In : *site de Légifrance*. [en ligne]. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFARTI000045197601>.

► Bibliographie

- ABBOTT, Andrew, 1988. *The System of Professions – An Essay on the Division of Expert Labor*. University of Chicago Press. ISBN 0226000680.
- ANGRIST, Joshua et PISCHKE, Jörn-Steffen, 2010. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. In : *Journal of Economic Perspectives*. [en ligne]. Printemps 2010. American Economic Association. Volume 24(2), pp. 3-30. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://www.aeaweb.org/articles?id=10.1257/jep.24.2.3>.
- ARBELÁEZ AYALA, Alejandra, BUCHER, Anne, MOREL, Makiko, SANCHEZ, Pablo, SIBILLE, Léo et TERRAZ, Isabelle, 2025. Rapport final du comité scientifique de l'évaluation du Plan d'investissement dans les compétences. In : *site de la Dares*. [en ligne]. Avril 2025. Dares. [Consulté le 5 juin 2026]. Disponible à l'adresse : <https://dares.travail-emploi.gouv.fr/publication/rapport-final-du-comite-scientifique-evaluation-pic>.
- BACH, Laurent, BOZIO, Antoine, FIZE, Étienne, GUILLOUZOUIC, Arthur et MALGOUYRES, Clément, 2025. L'impact économique des transmissions d'entreprises réalisées avec un pacte Dutreil. In : *site de l'institut des politiques publiques*. [en ligne]. Novembre 2025. N° 62. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://www.ipp.eu/publication/limpact-economique-des-transmissions-dentreprises-realisees-avec-un-pacte-dutreil/>.
- BAÏZ, Adam, 2022. Quelles évaluations des politiques publiques pour quelles utilisations ? In : *site du Haut-commissariat à la Stratégie et au Plan*. [en ligne]. 30 juin 2022. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://www.strategie-plan.gouv.fr/publications/evaluations-politiques-publiques-utilisations>.
- BALDACCHINO, Adeline et JOURDAIN, Édouard, 2024. La raison collective de la Cour des comptes. Examen d'un cas entre démocratie délibérative et libertaire. In : *Revue du Mauss*. [en ligne]. N° 63, pp. 336-351. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://hal.science/hal-05121965v1>.
- BARLET, Muriel, COLDEFY, Magali, COLLIN, Clémentine et LUCAS-GABRIELLI, Véronique, 2012. L'accessibilité aux médecins généralistes libéraux : plus faible en milieu rural. In : *Pour*. [en ligne]. Juillet 2012. N° 214, pp. 29-40. [Consulté le 19 mai 2026]. Disponible à l'adresse : <https://shs.cairn.info/revue-pour-2012-2-page-29?lang=fr&tab=texte-integral>.
- BATUT, Cyprien et DAVIOT, Quentin, 2025. Analyse événementielle. In : *site internet du LIEPP*. [en ligne]. Juillet 2025. LIEPP Fiche méthodologique. N° 53. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://sciencespo.hal.science/hal-05281180v1>.
- BERTHE, Juliette, 2025. Le défi des données pour l'inspection générale des affaires sociales. In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 29-48. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546937?sommaire=8546949>.

- BOLLINET, Quentin, FLOYRAC, Aymeric, MAILLARD, Sophie et ROSENZWEIG, Agathe, 2025. L'inspection générale des finances et la science des données – Quelles méthodes pour quels usages ? In : *Courrier des statistiques*. [en ligne]. 23 juin 2025. Insee. N° N13, pp. 7-28. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8546934?sommaire=8546949>.
- CARTWRIGHT, Nancy, 2010. Hunting Causes and Using Them: Approaches in Philosophy and Economics. In : *Analysis*. [en ligne]. Avril 2010. Oxford University Press. [Consulté le 20 mai 2026]. Disponible à l'adresse : https://www.researchgate.net/publication/48911320_Hunting_Causes_and_Using_Them_Approaches_in_Philosophy_and_EconomicsBy_Nancy_Cartwright_Summary.
- CHALEIX, Mylène et DURR, Jean-Michel, 2024. Les revues par les pairs : un levier au service de la statistique publique. In : *Courrier des statistiques*. [en ligne]. 16 décembre 2024. Insee. N° N12, pp. 137-157. [Consulté le 4 juin 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8264560?sommaire=8264562>.
- CHARIOT, Appoline, 2026. Reste à charge des femmes atteintes d'un cancer du sein après remboursement par l'assurance maladie obligatoire, France. In : *Journal of Epidemiology and Population Health*. [en ligne]. Mars 2026. Elsevier. Volume 74, n° 1. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.sciencedirect.com/science/article/pii/S2950433326001072>.
- COUR DES COMPTES ET CHAMBRES RÉGIONALES ET TERRITORIALES DES COMPTES, 2024. Normes professionnelles – 4^e édition. In : *site de la Cour des comptes*. [en ligne]. 14 mars 2024. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/cour-des-comptes/nous-decouvrir/normes-professionnelles>.
- COUR DES COMPTES, 2024a. Les droits de succession. In : *site de la Cour des comptes*. [en ligne]. Juin 2024. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/publications/les-droits-de-succession>.
- COUR DES COMPTES, 2024b. Évaluation du fonds régional de garantie de la Région Hauts-de-France. In : *site de la Cour des comptes*. [en ligne]. 2 octobre 2024. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/publications/evaluation-du-fonds-regional-de-garantie-de-la-region-hauts-de-france>.
- COUR DES COMPTES, 2025a. Prévention et prise en charge des accidents vasculaires cérébraux. In : *site de la Cour des comptes*. [en ligne]. 29 octobre 2025. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/publications/prevention-et-prise-en-charge-des-accidents-vasculaires-cerebraux>.
- COUR DES COMPTES, 2025b. Le soutien au développement du biogaz. In : *site de la Cour des comptes*. [en ligne]. 6 mars 2025. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/publications/le-soutien-au-developpement-du-biogaz>.
- COUR DES COMPTES, 2025d. Le pacte Dutreil : un dispositif fiscal en forte croissance à mieux cibler. In : *site de la Cour des comptes*. [en ligne]. 11 novembre 2025. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/publications/le-pacte-dutreil-un-dispositif-fiscal-en-forte-croissance-mieux-cibler>.

- COUR DES COMPTES, 2025e. Rapport d'activité 2024 de la Cour des comptes – Pierre Moscovici : des juridictions financières profondément transformées. In : *site de la Cour des comptes*. [en ligne]. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/rapport-dactivite-2024-de-la-cour-des-comptes>.
- COUR DES COMPTES, 2026a. Les aides à la décarbonation de l'industrie du plan de relance et de France 2030. In : *site de la Cour des comptes*. [en ligne]. 11 mars 2026. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/publications/les-aides-la-decarbonation-de-lindustrie-du-plan-de-relance-et-de-france-2030>.
- COUR DES COMPTES, 2026b. La prévention et la prise en charge du cancer du sein. In : *site de la Cour des comptes*. [en ligne]. 28 janvier 2026. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.ccomptes.fr/fr/publications/la-prevention-et-la-prise-en-charge-du-cancer-du-sein>.
- DENIS, Jérôme, 2018. *Le travail invisible des données – Éléments pour une sociologie des infrastructures scripturales*. Avril 2018. Presses des Mines. Sciences sociales. ISBN 9782356715005.
- DESROSIÈRES, Alain, 2008. *Pour une sociologie historique de la quantification – L'argument statistique I*. [en ligne]. Presses des Mines. Sciences sociales. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://books.openedition.org/pressesmines/901#anchor-resume>.
- FOUGÈRE, Denis et JACQUEMET, Nicolas, 2019. Inférence causale et évaluation d'impact. In : *Économie et Statistique*. [en ligne]. Insee. N° 510-511-512, pp. 187-206. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4253057?sommaire=4253159>.
- GAZANIOL, Alexandre et LÊ, Mathilde, 2021. The impact of public loan guarantees on banks' risk taking and firms' growth: Evidence from France. In : *Économie & prévision*. [en ligne]. Direction générale du Trésor. 2021/1 n° 218, pp. 67-104. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://shs.cairn.info/journal-economie-et-prevision-2021-1-page-67?lang=en&tab=texte-integral>.
- GILLE, Bertrand, 1980. *Les sources statistiques de l'histoire de France – Des enquêtes du XVII^e siècle à 1870*. Librairie Droz. ISBN 978-2-600-04091-4.
- GIVORD, Pauline, 2010. Méthodes économétriques pour l'évaluation de politiques publiques. In : *Documents de travail*. [en ligne]. 1^{er} juin 2010. Insee. N° G2010/08. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/1380865>.
- HAYEZ, Philippe, 2015. La Cour des comptes : du contrôle à l'évaluation. In : *Revue française d'administration publique*. [en ligne]. Institut national du service public. N° 155, pp. 707-711. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://droit.cairn.info/revue-francaise-d-administration-publique-2015-3-page-707>.
- HOLLAND, Paul W., 1986. Statistics and Causal Inference. In : *Journal of the American Statistical Association*. [en ligne]. Décembre 1986. American Statistical Association. N° 396, pp. 945-960. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354>.

- IOANNIDIS, John P. A., STANLEY, T. D. et DOUCOULIAGOS, Hristos, 2017. The Power of Bias in Economics Research. In : *The Economic Journal*. 24 octobre 2017. Wiley. Current. Volume 127, n° 605, pp. 263-265. ISSN 0013-0133.
- ISNARD, Michel, 2018. Qu'entend-on par statistique(s) publique(s) ? In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 69-77. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/3646978?sommaire=3647035>.
- KOUMARIANOS, Heidi, LEFEBVRE, Olivier et MALHERBE, Lucas, 2024. Les appariements : finalités, pratiques et enjeux de qualité. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2024. Insee. N° N11, pp. 117-139. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8203044?sommaire=8203072>.
- KOUMARIANOS, Heidi et RIVIÈRE, Pascal, 2025. Statistiques fondées sur des données administratives : esquisse d'un cadre général. In : *Document de travail*. [en ligne]. 23 juin 2025. Insee. N° M2025-03. [Consulté le 9 juin 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/8596376>.
- LATOUR, Bruno, 2002. *La fabrique du droit – Une ethnographie du Conseil d'État*. 10 octobre 2002. La Découverte. ISBN 978-2707135810.
- MIGAUD, Didier, 2019. Entretien avec Didier Migaud, Premier président de la Cour des comptes. In : *Gestion et finances publiques*. [en ligne]. Novembre 2019. JLE. 2019/3 n° 3, pp. 74-81. [Consulté le 20 mai 2026]. Disponible à l'adresse : <https://droit.cairn.info/revue-gestion-et-finances-publiques-2019-3-page-74?lang=fr&tab=texte-integral>.
- MOSCOVICI, Pierre, 2024. La Cour des comptes, entre droit et nombres. In : *Archives de philosophie du droit*. Dalloz. N° 65, pp. 399-407. ISBN 0066-6564.
- PORTER, Theodore M., 1995. *Trust in numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press. ISBN 978-0691029085.

- RENNE, Catherine, 2018. Bien comprendre la déclaration sociale nominative pour mieux mesurer. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 35-44. [Consulté le 9 juin 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/3647029?sommaire=3647035>.
- RUBIN, D. B., 1974. Estimating causal effects of treatment in randomized and nonrandomized studies. In : *Journal of Educational Psychology*. American Psychological Association. Volume 66, n° 5, pp. 688-701. ISSN 0022-0663.
- SCOTT, James C., 1998. *Seeing Like a State – How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press. ISBN 9780300007 0163.

Les statistiques coloniales et leurs archives



Béatrice Touchelay*

Les méthodes statistiques se perfectionnent au XIX^e siècle. La statistique compte parmi les sciences et le statisticien devient le « grand expert social ». La conquête des sociétés européennes par le chiffre accompagne l'élargissement des interventions des États. Ces transformations coïncident avec une nouvelle vague de colonisation française vers les continents africain et asiatique, qui s'appuie sur des cartes et des chiffres et mobilise des techniques et des classifications importées de métropole. La mise en nombres, le classement des populations colonisées et les hiérarchies introduites par les colons ont probablement des conséquences sur l'organisation des sociétés qu'elles sont censées représenter mais que, de fait, elles formatent. Les archives des statistiques coloniales permettent de s'interroger plus largement sur le sens et l'impact des catégories statistiques et sur les motivations de ceux qui les portent. Elles renseignent sur les deux temps des statistiques coloniales : celui de « la statistique coloniale sans statisticien », puis, à partir de 1943, celui du sursaut statistique dans des territoires intégrés à l'Union française qui, au moins dans les mots, ne sont plus complètement considérés comme des territoires asservis. Ce sursaut se manifeste à travers l'organisation de services spécifiques et d'enquêtes sur le terrain, ces dernières étant encadrées par des statisticiens de métropole secondés par des techniciens formés sur place. Le classement récent d'un fonds « statistique » aux Archives nationales d'outre-mer invite à approfondir les réflexions sur le sens des statistiques des territoires anciennement colonisés et sur le métier des statisticiens.

 Statistical methods were refined throughout the 19th century. Statistics became a science, and statisticians became “the great social experts”. The conquest of European societies by numbers accompanied the expansion of state and administrative interventions. These transformations coincided with a new wave of French colonisation of Africa and Asia, which relied on maps and figures and employed techniques and classifications imported from mainland France. The quantification, the classification of colonised populations, and the hierarchies introduced by the colonists probably had consequences on the organisation of the societies they were supposed to represent but which, in fact, they shaped. More broadly, colonial statistical archives allow us to examine the meaning and impact of statistical categories and the motivations of those who created them. They shed light on the two phases of colonial statistics: first, the phase of “colonial statistics without statisticians”, and then, from 1943 onward, the surge in statistical activity in territories integrated into the French Union, which, at least nominally, were no longer considered entirely subjugated territories. This surge manifested itself through the organisation of specific services and field surveys, the latter being overseen by statisticians from mainland France assisted by locally trained technicians. The recent cataloguing of a new series entitled “Statistics” at the National Overseas Archives in Aix-en-Provence invites further reflection on the meaning of statistics in formerly colonised territories and on the profession of statistician.

* Professeure des universités, Université de Lille, Laboratoire Histoire, Histoire de l'art, Archéologie, Textes, Images, Sociétés (HARTIS), UMR CNRS 9028.
beatrice.touchelay@univ-lille.fr



Des administrations coloniales sont mises en place, qui produisent et utilisent des statistiques, et imposent les catégories sociodémographiques adoptées en métropole.



Après une première période de colonisation française aux XVII^e et XVIII^e siècles, une seconde période démarre au XIX^e siècle, tournée vers les continents africain et asiatique. Des administrations coloniales sont mises en place, qui produisent et utilisent des statistiques, et imposent les catégories sociodémographiques adoptées en métropole. Les archives disponibles – en particulier une toute nouvelle série inventoriée aux Archives nationales d’outre-mer avec le soutien de l’Insee (**encadré**) – éclairent les intentions des commanditaires de ces chiffres. Elles montrent comment les statistiques

coloniales sont produites et ce qu’elles reflètent. Elles incitent à réfléchir aux effets des catégories qui les sous-tendent sur les populations concernées, et plus largement au sens et à l’impact des catégories statistiques.

Tout au long du XIX^e siècle, les méthodes statistiques se perfectionnent. Les doutes sur le bien-fondé du calcul de la moyenne s’estompent, les catégories et les nomenclatures que construisent les statisticiens (Guibert et al., 1971) s’affinent. La statistique compte parmi les sciences et « [l]e statisticien, nouveau géomètre devient [...] le grand expert social, capable de prendre la mesure de toute chose » (Perrot, 1987). Les recensements démographiques jouent un rôle important dans la création de la nation et dans la centralisation des pouvoirs politiques. Au même titre que l’harmonisation des poids et des mesures ou que la révolution des transports (canaux, puis chemin de fer), l’essor des statistiques accélère la formation du marché national et la normalisation des pratiques. Ces transformations accompagnent une nouvelle vague de colonisation européenne.

Après s’être entretenus pendant les guerres révolutionnaires et celles du Premier Empire, et convaincus de leur supériorité, plusieurs pays de l’Ouest européen (dont la France) maintiennent leur domination transatlantique sur les « vieilles colonies » et prolongent la conquête du monde. La domination militaire, l’installation des administrations coloniales et l’exploitation des territoires des continents africain et asiatique s’appuient sur des outils, cartes et chiffres, qui mobilisent des techniques et des classifications importées de métropole (Bowker et Star, 1999). Ces outils s’imposent à des populations dont les croyances¹, les structures sociales et politiques, sont différentes et qui, concevant et représentant l’espace autrement, comptent et classent différemment².

Les tableaux préformatés remplis par les administrateurs coloniaux pour les gouverneurs et la métropole enferment les individus qu’ils qualifient « d’indigènes » dans des catégories dont le sens n’est pas partagé. Par exemple, « actif / inactif », dans des économies qui ne sont pas monétarisées³, « jeune / vieux »⁴, alors qu’il n’existe pas d’état civil, ni de

¹ « Ne nous dénombre pas, s’il te plaît, dit Sèni en se levant précipitamment, nous sommes des Bouts-de-bois-de-Dieu, tu nous ferais mourir » (Sembène, 1960, p. 301).

² Cette affirmation prend tout son sens si l’on retient la définition d’Alain Desrosières : « Quantifier, c’est concevoir puis dénombrer – « mettre en nombre » –, classer et compter » (Desrosières, 2008).

³ Cette classification conduit à créer deux secteurs : l’un marchand, tiré par les colons et dans lequel s’insère, ou est insérée de force (travail forcé), une partie des populations colonisées ; l’autre non marchand, artisanal et vivrier, que l’appropriation des ressources naturelles par les colons condamne à la portion congrue.

⁴ Le respect dévolu aux plus âgés, qui est l’un des piliers de la hiérarchie des sociétés colonisées, se trouve en complète contradiction avec les valeurs des colons qui, ne s’intéressant qu’à la force physique utile pour les chantiers, bouleversent l’ordre traditionnel.

► Encadré. De nouvelles archives sur les statistiques coloniales

En 2023, un partenariat scientifique a été mis en place entre les Archives nationales d'outre-mer (ANOM), l'Insee et l'Institut de recherches historiques du Septentrion (IRHiS), unité mixte de recherche sous la tutelle du Centre national de la recherche scientifique (CNRS) et de l'université de Lille. Il consistait à classer le **fonds du Service statistique des Territoires d'outre-mer**, dénommé « fonds STAT »*. Les archives qui ont pu être classées** se répartissent en 226 cartons et regroupent des documents datés des années 1902 à 1960.

Le fonds STAT inclut des documents de natures très diverses : rapports, relevés administratifs, recensements, fiches d'enquête, enquêtes par sondage pour l'Afrique occidentale française. Son périmètre est l'ensemble des territoires d'outre-mer.

Les dossiers relatifs aux statistiques collectées sont plus ou moins complets. Ceux des

enquêtes par sondage menées par les missions socioéconomiques sont généralement très exhaustifs. On peut y retrouver les documents d'organisation et de cadrage de l'enquête (programme, formation des enquêteurs, bases de sondage, documentation...), les données recueillies par les enquêteurs sur le terrain (fiches, questionnaires...), mais aussi des informations sur la préparation des résultats (dépouillement, versions préparatoires des rapports...).

Le plan de classement du fonds suit la logique progressive de l'action statistique :

- 1) la programmation et les moyens ;
- 2) la collecte des sources ;
- 3) la synthèse.

Les illustrations présentées dans cet article sont toutes extraites de ce nouveau fonds.

* Voir le lien : https://recherche-anom.culture.gouv.fr/archives/fonds/FANOM_00157/view:fonds.

** Le classement a été effectué par Cédric Picard, vacataire, et Amélie Hurel, conservatrice du patrimoine, sous la direction d'Isabelle Dion, directrice des ANOM.

calendrier romain⁵, « marié / célibataire », qui suppose l'homogamie, voire le sacrement religieux, n'ont pas de sens pour tous. Par ailleurs, le recours presque systématique à des catégories ethniques ou religieuses pour classer ces « indigènes » invente des hiérarchies et cimente des oppositions là où il n'existait que des différences. Ainsi, la mise en chiffres, le classement des populations colonisées et les hiérarchies introduites par les colons ont probablement des conséquences sur l'organisation des sociétés qu'elles sont censées représenter mais que, de fait, elles déforment en essayant de les formater.

Les archives coloniales renseignent sur la façon dont ces statistiques et leurs catégories ont été introduites dans les espaces de la seconde colonisation, au moment où l'organisation des services statistiques⁶, les pratiques d'enquêtes et les demandes de chiffres s'affinent dans la métropole. Quelques exemples tirés des archives serviront à montrer comment les catégories statistiques métropolitaines ont été exportées pour servir d'outil de gouvernement des territoires où elles ne correspondaient pas aux réalités locales. Soulignant **l'écart entre les catégories des observés et celles des observateurs**, ces exemples extraits des archives coloniales belges et françaises permettent de réfléchir à **l'impact des catégories statistiques sur les sociétés colonisées**. À rebours, ils permettent aussi de réfléchir aux effets de l'expérience des statistiques coloniales sur les métropoles.

Après les précisions nécessaires sur le sens des statistiques coloniales et de leurs archives, ce texte aborde le développement des statistiques coloniales en deux temps : celui de « la statistique coloniale sans statisticien » (Sanner, 1994), puis, à partir de 1943,

⁵ Ce calendrier va s'imposer progressivement.

⁶ Pour plus de précisions sur les services de statistiques coloniales en France métropolitaine avant 1943, voir Touchelay (2019).

celui du sursaut statistique dans des territoires qui seront intégrés à l'Union française⁷ et qui, au moins dans les mots, ne seront plus complètement considérés comme des territoires asservis. Ce sursaut se manifeste à travers l'organisation de services spécifiques et d'enquêtes sur le terrain, encadrées par des statisticiens de métropole qui s'appuient sur des techniciens formés sur place.

► Le point sur les archives et les statistiques coloniales —

Les archives coloniales

Comme toute archive, les archives coloniales se composent de la documentation, sous toutes ses formes, produite dans les territoires concernés par les administrations et, parfois, par des agents privés : colons, aventuriers, militaires, entrepreneurs ou capitaines d'industrie et leurs épouses, etc. Constituées pour une grande partie par des rapports administratifs adressés par les administrateurs coloniaux à leurs supérieurs, elles sont à interpréter minutieusement, et non pas à prendre pour argent comptant⁸. Ces archives

► **Figure 1 - Armoire contenant les fiches des archives de Hanoï**

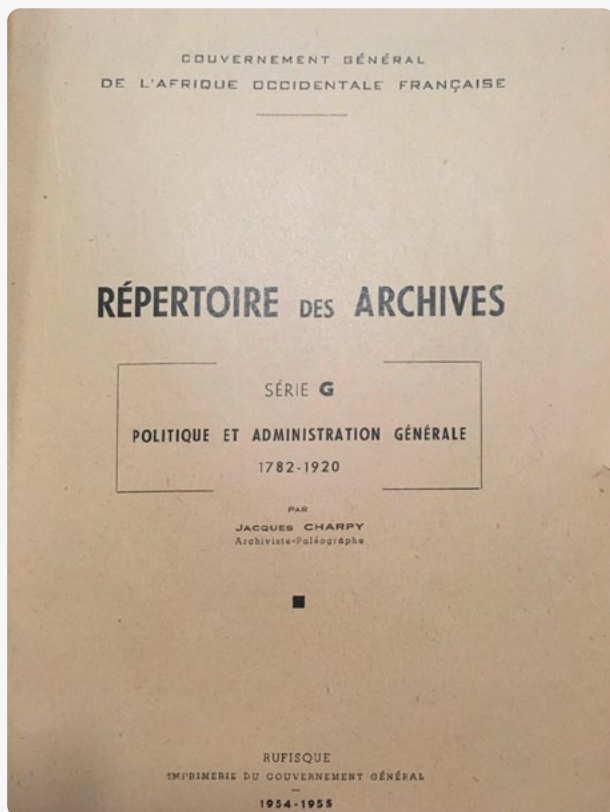


Source : Centre n° 1 des archives nationales du Vietnam, © photo Béatrice Touchelay.

⁷ Structure politique regroupant la France et ses anciennes colonies de 1946 à 1958.

⁸ Voir Tiquet (2017), qui est très éclairant sur la nature particulière des rapports administratifs et sur les précautions nécessaires de l'historien qui cherche à les interpréter.

► **Figure 2 - Inventaire des archives série G**



Source : Archives nationales du Sénégal, Dakar, © photo Béatrice Touchelay.

proviennent rarement des colonisés eux-mêmes. Ces derniers n'y apparaissent que lorsqu'ils sont en contact avec les colons. C'est le cas par exemple dans les archives des tribunaux indigènes ou des hôpitaux, dans celles des syndicats et des partis au moment des luttes pour l'indépendance, ou encore dans celles relatives à des personnalités qui ont acquis des responsabilités politiques à partir de 1945 et de la formation de l'Union française. Cependant, les archives coloniales sont faites avant tout par les colons et pour les colons. Elles ont été conservées et classées par les archivistes des administrations coloniales (*figures 1 et 2*).

Au moment des indépendances, la question du partage et de la restitution des archives provoque des débats. En principe, les règles sont précises : les archives dites « de souveraineté », comme les tracés des frontières ou les cartes des forages et des ressources naturelles, restent sur place ; les archives « de gestion », sur l'organisation administrative et économique des territoires colonisés et leur personnel, sont transférées et conservées en France. On trouve ces dernières :

- aux Archives nationales d'outre-mer (ANOM) à Aix-en-Provence ;

- aux archives diplomatiques de Nantes ou de la Courneuve pour les Affaires étrangères (naturalisation et état civil des Français nés ou décédés dans les anciennes colonies ou dans l'Union française, etc.) ;
- aux Archives nationales du monde du travail (ANMT) à Roubaix pour les fonds d'entreprises ;
- au Service des archives économiques et financières (SAEF) de Savigny-le-Temple, par exemple sur le Fonds d'investissement pour le développement économique et social des territoires d'outre-mer (Fides) ou sur les opérations de coopération réalisées par l'Insee ;
- aux Archives nationales de Pierrefitte-sur-Seine (dossiers de personnalités, etc.), ou dans des centres privés comme les chambres de commerce.

Quelques microfilms et numérisations permettent de conserver ces fonds à la fois là où ils ont été produits (i.e. dans l'ancienne colonie) et en métropole. Ainsi, une partie des archives de l'Afrique occidentale française (AOF) sont disponibles en microfilm aux archives nationales de Paris (site Richelieu) et aux Archives nationales du Sénégal (ANS) à Dakar.

Les statistiques coloniales



Les statistiques coloniales sont le plus souvent produites par les colons et les administrations coloniales.



Les statistiques coloniales sont le plus souvent produites par les colons et les administrations coloniales. Elles peuvent aussi être produites en métropole pour répondre aux besoins des colons. Renseignant sur la situation démographique et géographique du territoire, sur les ressources disponibles et sur leur exploitation, certaines visent à attirer les colons et les capitaux (d'où l'importance des statistiques médicales, pluviométriques et climatiques). D'autres visent à répondre aux

demandes de la métropole, à justifier des crédits attribués à la colonie et à préparer les rares plans d'investissements de la métropole dans les colonies (plan Sarraut de 1921 ou du Front populaire de 1937). À partir de 1919, elles répondent aussi aux demandes de la Société des Nations (SDN), précurseur de l'Organisation des Nations unies (ONU), qui réclame des comptes aux bénéficiaires des territoires sous mandat (comme le Liban ou le Cameroun pour la France⁹).

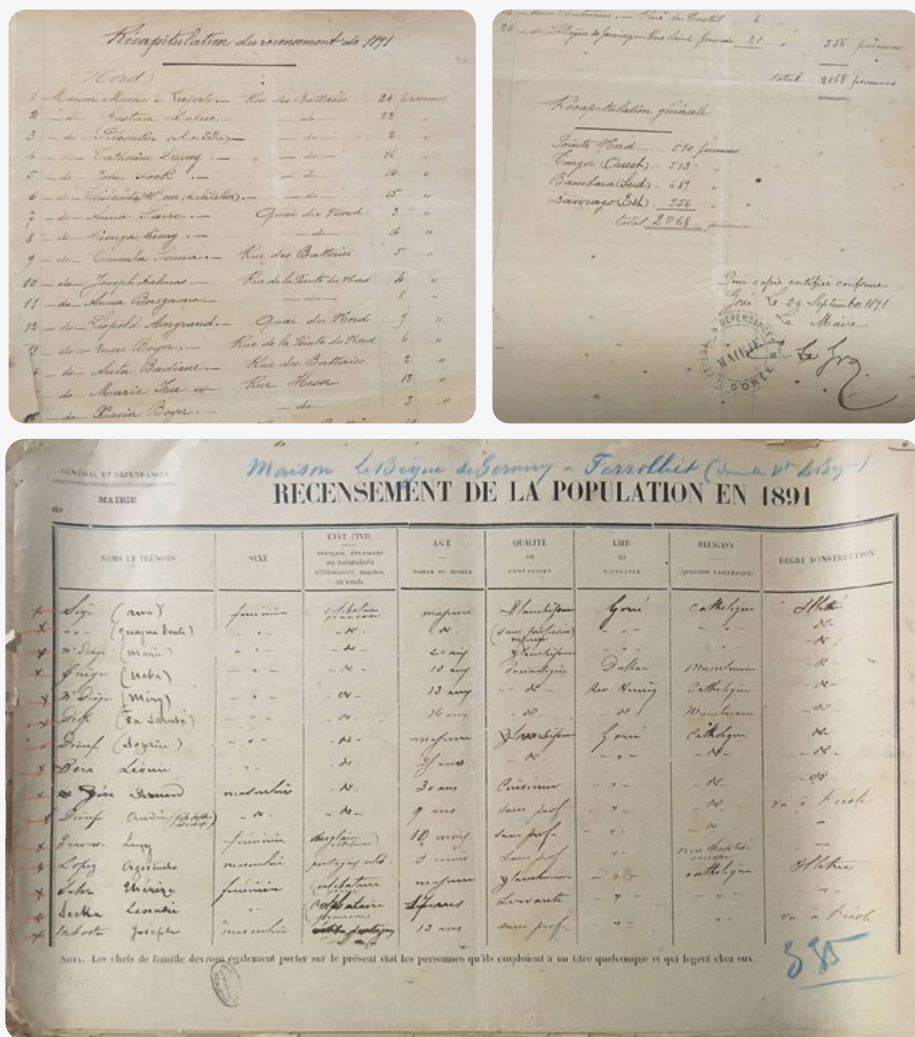
Avant l'installation du service de statistiques du ministère des Colonies en 1943, aucun service de statistiques coloniales générales n'est créé en métropole depuis la reprise de la colonisation française en 1830. Le développement des statistiques coloniales est en partie laissé au bon vouloir et aux capacités des gouverneurs et de leurs subordonnés sur place, qui les voient souvent comme une charge inutile.

Les archives des statistiques coloniales informent sur l'organisation et les pratiques des administrations coloniales, ainsi que sur leurs transformations. Certaines nous apprennent par exemple que, bien que la pratique des recensements démographiques

⁹ Au terme de la Première Guerre mondiale, la France est devenue mandataire du Liban et du Cameroun, anciennes colonies de l'Empire ottoman et de l'Allemagne.

nominatifs¹⁰ fléchisse au XIX^e siècle, comparé à l'Ancien Régime (Malègue, 2018), elle subsiste dans certains territoires comme à Gorée, commune autonome de l'actuel Sénégal¹¹ (*figure 3*).

► **Figure 3 - Recensement de 1891 à Gorée**



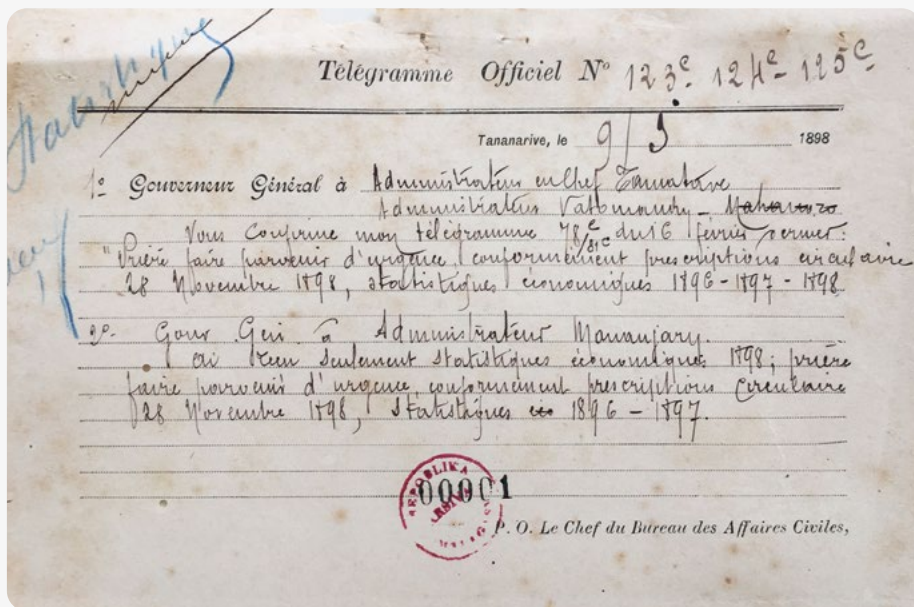
Lecture : Les deux premières photos illustrent le récapitulatif par maison, indiquant pour chacune d'elle le nom du chef de famille et le nombre d'occupants. La troisième photo correspond au recensement détaillé des habitants.
Source : Archives nationales du Sénégal. © photo Béatrice Touchelay.

- 10 « Dans le domaine de la démographie coloniale, les documents d'archives font référence à deux types de recensements, le *recensement numérique* et le *recensement nominatif*. Le premier donne le nombre total d'habitants par localité sans autre renseignement [...]. Le second ressemble au recensement moderne car il enregistre non seulement le nombre mais aussi quelques renseignements supplémentaires, comme la structure des ménages, le sexe et l'âge » (Gervais, 1994).
- 11 Archives nationales du Sénégal (ANS) 3G² 131. Gorée, recensement de la population en 1891.

Un autre enseignement éclaire le motif de la fabrication du chiffre : même s'il est difficile à produire, **le chiffre rassure**. Dès les premiers temps de l'implantation française à Madagascar (Tisseau, 2017) en 1898 et alors que la situation militaire n'est pas stabilisée, le général Gallieni exige des administrateurs qu'ils lui fournissent régulièrement des rapports économiques et un grand nombre d'informations quantifiées (**figure 4**).

Enfin, bien au-delà des nombres, les archives des statistiques coloniales éclairent le processus de domination coloniale et son délitement. Elles informent aussi sur les relations entre la métropole et les colonies, l'organisation du pouvoir dans les colonies elles-mêmes et son échec. Finalement, derrière les fausses certitudes attribuées aux chiffres, elles montrent comment, à travers les biais qu'elles induisent, les statistiques coloniales contribuent à expliquer la faillite de cette domination.

► **Figure 4 - Directive du général Gallieni pour la transmission de statistiques économiques, 1898**



Source : Archives nationales de Madagascar, Antananarivo, © photo Béatrice Touchelay.

► Avant 1943 : des statistiques coloniales sans statisticien —

Les institutions de statistiques coloniales en métropole

En métropole, la production des statistiques coloniales est définie progressivement par des circulaires et des instructions du ministère de la Marine, puis du ministère des Colonies, leurs services se chargeant de la diffusion des résultats. La première édition des « États

de la population, de la culture et du commerce » date de 1832 et d'autres suivront¹². Les dotations budgétaires de ces ministères ne suffisent pas à répondre aux besoins de données. En effet, les statistiques sont rendues particulièrement coûteuses du fait de l'immensité des territoires et de la faiblesse de leur densité démographique, ainsi que de leur sous-encadrement chronique et du manque criant d'infrastructure¹³.

Centraliser et diffuser les chiffres disponibles

À partir de 1855, la production et la diffusion des statistiques coloniales relève d'une institution : l'exposition permanente des colonies, créée à Paris. Les données disponibles sont irrégulières et hétérogènes. Diverses administrations ou structures viennent compléter les services coloniaux métropolitains : le sous-secrétariat d'État aux Colonies en 1881, le Conseil supérieur des colonies en 1883, l'inspection des colonies en 1887 et l'École coloniale (qui ne dispense aucune formation à la statistique) en 1889. Leurs tutelles sont variables avant la création du ministère dédié aux colonies par la loi du 20 mars 1894¹⁴, qui accompagne la stabilisation des frontières de l'empire. La production des statistiques souffre de ces changements. Le service de la documentation officielle, l'exposition permanente, disparaît en 1896 lors de la destruction du Palais de l'Industrie qui l'hébergeait. Sa suppression laisse « les collections à l'abandon dans les caves du ministère au moment où les puissances coloniales étrangères perfectionnent [leur documentation commerciale et coloniale] »¹⁵.

Un début d'harmonisation : l'Office colonial

De graves lacunes subsistent, de nombreux territoires ne retournent pas les tableaux à renseigner et le chiffre de la population des colonies est méconnu.

La situation s'améliore avec la création d'un Office colonial en 1899, chargé de « faire l'éducation coloniale des Français, de centraliser et de mettre à leur disposition les renseignements concernant l'agriculture, le commerce et l'industrie des colonies françaises »¹⁶. Installé au Palais Royal à Paris, il est rattaché au ministère des Colonies et placé sous la surveillance d'un Conseil de perfectionnement. Ce dernier doit « contribuer au développement du commerce entre la métropole et les colonies »¹⁷. L'Office produit des notices

destinées aux émigrants et publie toute une série de statistiques générales à mesure que les informations sont transmises par les colonies au ministère. Néanmoins, de graves lacunes subsistent, de nombreux territoires ne retournent pas les tableaux à renseigner et le chiffre de la population des colonies est méconnu.

¹² Disponible en ligne sur Gallica à partir de 1839 au lien suivant : <https://gallica.bnf.fr/ark:/12148/bpt6k110745v>.

¹³ Archives nationales d'outre-mer (ANOM) Aix-en-Provence (France) 1 AFFECO/101 Conférence africaine française de Brazzaville. Procès-verbal (PV) de la première séance plénière, mardi 1^{er} février 1944, Exemplaire n° 4 de Peter (Pleven) noté « Secret ». Pleven : « en Afrique noire française la démographie commande. La densité générale est inférieure à trois habitants au km², la question de la main-d'œuvre domine tout le problème », p. 9.

¹⁴ Voir les références juridiques en fin d'article.

¹⁵ ANOM AGEFOM//852 Fonds ministériels. Office colonial. Rapport présenté au nom du conseil d'administration de l'Office colonial, dactylographié, non daté, 18 pages.

¹⁶ Archives nationales du Vietnam (ANV) Centre n° 1, Hanoi. Gouvernement général de l'Indochine. 3684 Création, installation et fonctionnement de l'Office colonial à Paris (1895-1915). Rapport du directeur aux conseils d'administration et de perfectionnement sur l'ensemble du service pendant l'exercice 1905, 20 janvier 1906, Melun imprimerie administrative.

¹⁷ ANOM AGEFOM//852 Fonds ministériels. Office colonial. PV des séances du Conseil de perfectionnement, première séance, 7 avril 1900.

À partir de la Grande Guerre, les statistiques commerciales et douanières sont confiées à des agences économiques implantées dans les territoires – qui les financent – et séparées des autres statistiques. En 1919, les travaux des agences territoriales sont centralisés par une Agence générale des colonies, qui remplace l'Office colonial¹⁸. En 1934, l'Agence générale est supprimée par mesure d'économie (mais pas les agences territoriales) et réorganisée l'année suivante¹⁹. En 1937, un service intercolonial d'information et de documentation est chargé de centraliser les attributions des agences territoriales. Il est réorganisé sous Vichy en 1941. Tous ces services centraux ne s'intéressent qu'aux statistiques des échanges extérieurs. À partir de 1920, les statistiques non douanières relèvent d'une direction économique du ministère des Colonies, qui devient en 1935 la direction des affaires économiques²⁰.

À la veille de la Seconde Guerre mondiale, un maigre bilan

À la veille de 1940, l'Algérie dispose d'un service dédié aux travaux statistiques. Ses données sont à peu près analogues à celles qui sont publiées chaque année en métropole, et il en est de même des protectorats de Tunisie et du Maroc. Ailleurs, les lacunes sont béantes.

Productions et usages des chiffres dans les colonies, l'impossible normalisation

Dans les colonies proprement dites, les statistiques doivent être établies suivant un plan uniforme prescrit par le ministère des Colonies. Dans la pratique cependant, les administrations locales ne fournissent pas de données régulières. Seule l'Indochine dispose de son propre service de statistiques. Créé en 1922, celui-ci est pris en charge par des statisticiens détachés de la Statistique générale de la France (SGF)²¹... jusqu'aux restrictions budgétaires de 1933 qui conduisent à leur rapatriement.

Ailleurs, les gouverneurs centralisent sur place les statistiques provenant des différentes administrations et institutions (prisons, hôpitaux, écoles, etc.). Ils disposent aussi des résultats des enquêtes effectuées par les administrateurs à l'occasion de leurs tournées. Ces données sont très hétérogènes et les gouverneurs n'ont aucun moyen de les contrôler. Les exigences de la métropole dépassent le plus souvent leurs capacités d'informateur.

Des exigences démesurées

La nécessité de disposer de statistiques de qualité se renforce à partir de la loi du 13 avril 1900²², qui oblige les colonies à équilibrer leur budget. Dès 1901 à Madagascar,

¹⁸ ANOM AGEFOM//3674 Statut de l'Agence générale des colonies. Rapport sur sa création.

¹⁹ ANOM AGEFOM //902 Création de l'Agence de la France d'Outre-mer (FOM). Rapport de Marius Moutet, ministre des Colonies, pour le président de la République au sujet de la suppression de l'Agence générale des colonies par le décret-loi du 4 avril 1934, 12 mars 1937.

²⁰ ANOM AGEFOM//4101 Ministère des Colonies. Direction des Affaires économiques et du Plan (1835-1964).

²¹ Service statistique rattaché successivement à différentes administrations françaises de 1840 à 1940. C'est un des ancêtres de l'Insee. La bibliothèque de l'Insee a hérité à sa création des collections détenues antérieurement par la bibliothèque de la SGF (Beaurepaire et Boucher, 2025).

²² Loi du 13 avril 1900 portant fixation du budget général des dépenses et recettes de l'exercice 1900. Voir les références juridiques en fin d'article.

► **Figure 5 - Divers tableaux statistiques commandités par le général Gallieni, 1901**

Tableau 1 : État des colons ou sociétés présents ou représentés dans la province

Tableau 5 : État du mouvement commercial provoqué par la vente des produits de la province

Tableau 3 : État des exploitations agricoles, forestières, industrielles entreprises par les Européens et assimilés

23 Joseph Gallieni (1849-1916), général et maréchal de France, fut gouverneur général de Madagascar de 1896 à 1905.

Le premier tableau concerne les colons et les sociétés présents ou représentés dans la province au 1^{er} semestre 1901²⁴. Les colons sont à classer par nationalité, date et lieu de naissance (*figure 5*).

Un second tableau s'intéresse aux colons qui ont quitté la province et pour lesquels il faut renseigner la nationalité, la date et le lieu de naissance, la composition de la famille, la nature de l'entreprise, la cause du départ ou d'abandon de l'entreprise et le lieu éventuel du transfert de l'entreprise. Ces informations sont très lacunaires puisque rien n'oblige les intéressés à déclarer ces changements.

Un troisième tableau renseigne sur les exploitations agricoles, forestières ou industrielles des « Européens et assimilés » (*figure 5*) et un quatrième sur celles des « indigènes et assimilés ». D'autres tableaux, comme le cinquième tableau sur l'état du mouvement commercial (*figure 5*), confirment cette boulimie d'informations.

Plusieurs circulaires des ministres des Colonies (Gaston Doumergue en juillet 1904²⁵, puis Alfred Picard en février 1909) renforcent la pression sur les administrations territoriales, sommées de remplir des tableaux préformatés, de respecter un échéancier rigoureux et de fournir « des statistiques démographiques annuelles ».

Quelques tentatives d'harmonisation

La circulaire Doumergue de 1903 impose « 12 modèles de tableaux »²⁶, celle de 1909 compte 212 pages de prescriptions...²⁷.

Les gouverneurs relaient ces exigences dans les territoires. À Madagascar, par exemple, plusieurs circulaires réclament un état nominatif au 1^{er} janvier 1905 des habitants des principales communes classé par profession (commerçants, employés de commerce, agriculteurs ou ouvriers agricoles, industriels ou ouvriers d'industrie). De nombreuses informations sont demandées : la date et le lieu de naissance, le dernier domicile connu, la composition de la famille, la profession antérieure et les ressources disponibles, la date d'arrivée dans la colonie, la province, le lieu où est situé l'établissement professionnel, ainsi que la nature de l'entreprise ou de la profession. Les personnes sont à classer par nationalité (français, anglais, allemands, russes, hollandais, anjouanais, comoriens, zanzibarites, chinois ou indiens)²⁸.

Réalistes dans certaines régions, ces exigences imposent une charge impossible à accomplir dans les territoires sous-encadrés d'Afrique subsaharienne, au regard des effectifs d'administrateurs territoriaux. Elles poussent ces derniers à privilégier les

²⁴ Archives nationales de Madagascar (ANM), Antananarivo. Province d'Ambatondrazaka. Rapport de l'administrateur Comperat sur la situation économique de la province, statistiques économiques 1^{er} semestre 1901.

²⁵ ANV Gouvernement général de l'Indochine 8848. Note au sujet de la statistique de la population d'Indochine (1904-1912), 4 pages.

²⁶ ANV Gouvernement général de l'Indochine. 138. Statistique de la population de l'Indochine. Note confidentielle du résident supérieur au Tonkin aux résidents chefs de province du Tonkin : « Je ne saurai trop vous recommander d'apporter le plus grand soin à l'établissement de ces statistiques et de procéder avec la plus grande circonspection dans vos enquêtes de recensement de la population annamite, pour éviter de froisser la susceptibilité des indigènes ou d'éveiller leurs craintes. Il importe surtout que les intentions de l'Administration ne soient pas dénaturées et que la population annamite ne puisse être amenée par des renseignements tendancieux à redouter la création d'une nouvelle taxe », Hanoi, 22 septembre 1904, pour ampliation, l'administrateur chef de Cabinet, Signé Bosc.

²⁷ ANOM AGEFOM//852 Fonds ministériels. Office colonial. Pièces de principes (1900-1911), loi de finances et arrêtés.

²⁸ ANM. Statistiques économiques 180/2 Nossi-Bé.



Au terme de la période, les archives des statistiques coloniales et les témoignages des acteurs conduisent à un triste bilan...



statistiques administratives provenant des rapports d'activité des services douaniers, du fisc, des prisons, des écoles ou des hôpitaux, lesquelles ne supposent pas d'enquête spécifique. Les statistiques démographiques qui exigent des enquêtes sont sacrifiées.

Au terme de la période, les archives des statistiques coloniales et les témoignages des acteurs conduisent à un triste bilan : « Il reste beaucoup à faire pour organiser [...] la statistique dans les possessions françaises, éveiller l'intérêt des chefs des administrations locales pour les questions statistiques et leur montrer l'utilité qu'aurait pour eux un service spécialisé » (Ulmer, 1938).

À la veille de la Seconde Guerre mondiale, l'insuffisance des statistiques sur l'Empire français est reconnue. La nécessité d'une réforme s'impose.

► À partir de 1943 : les statistiques coloniales et leurs archives changent de nature

La mise en place du service colonial de statistiques

Pendant l'Occupation, plusieurs éléments changent la donne : la création du Service national de statistiques (SNS)²⁹, l'introduction de la mécanographie³⁰, la forte augmentation des effectifs (6 000 agents au SNS à la Libération, contre 139 à la SGF en 1939), la création d'une école d'application et, enfin, la formation d'un corps d'administrateurs statisticiens recrutés à la sortie de l'école Polytechnique dont certains partiront outre-mer. Dans ce contexte, l'organisation des statistiques coloniales est modifiée par le régime de Vichy. Un **Service statistique colonial** est créé auprès du secrétaire d'État aux colonies en octobre 1943. Il est chargé de toutes les questions statistiques et de leur publication³¹. Il coordonne les activités statistiques des diverses administrations, s'assure de l'uniformisation des méthodes de collecte, d'exploitation et d'interprétation des données et propose toutes mesures utiles au développement et à l'amélioration de la documentation statistique coloniale.

Les lacunes des statistiques disponibles, en particulier des statistiques démographiques, conduisent à approfondir les réformes après la libération³² :

« La France, deuxième pays colonial du monde, est présentement incapable de savoir de façon un peu précise de quelle région proviennent les coloniaux qu'elle envoie dans ses possessions, quelle est leur formation et quel métier ils vont y exercer, entre quels âges

²⁹ Le SNS a été créé en 1941 par la fusion de deux services intégrés alors au ministère de l'Économie et des Finances depuis 1940 : le Service de la Démographie (SD) et la Statistique générale de la France (SGF), déjà évoquée précédemment. Le SNS est l'ancêtre direct de l'Insee, qui lui a succédé en 1946.

³⁰ La mécanographie désigne les techniques mécaniques et électromécaniques de traitement de l'information mobilisant par exemple des cartes perforées.

³¹ Extrait de l'inventaire de la série STAT des ANOM rédigé par Amélie Hurel, conservatrice aux ANOM.

³² ANM Province de Diego Suarez. IV D. 31 Statistique économique par circonscription 1936-1948 Recensement de la population 1946-19 décembre 1945 urgent Copie de la circulaire ministérielle 370/SCS en date du 20 juillet 1945. Le ministre des Colonies aux Gouverneurs généraux et gouverneurs des colonies.

extrêmes se place la vie coloniale, ou dans quelle mesure elle réagit sur la vie familiale (fécondité, dispersion, études, etc.), sur la morbidité, sur la mortalité ».

Les conséquences de cette méconnaissance sont sans appel :

« Pays colonisateur, la France ne peut pas traiter le problème de la colonisation. Or, sa faible population lui interdit de compter sur la pression démographique intérieure pour assurer le recrutement et la sélection spontanée des coloniaux. Aucun pays colonial n'a autant qu'elle besoin d'être dans ce domaine parfaitement renseigné ».

À la Libération, des changements de pratiques

Le « recensement des Européens dans les colonies » est prévu en mars 1946, en même temps que le recensement général de la population de métropole. Outre-mer, les difficultés budgétaires obligent à limiter l'opération « aux Français » présents sur place.

La politique de la France outre-mer est modifiée avec la création de l'Union française, laquelle dote les territoires de représentations parlementaires et met officiellement un terme au système colonial. L'emprise des organisations internationales, l'affirmation des pays non-alignés³³ à partir de 1955 et la poussée des mouvements d'indépendantistes incitent à renforcer la mise en valeur des territoires... ce qui suppose de mieux les connaître.

Les réformes de l'organisation des services centraux du ministère des Colonies, désormais ministère de la France d'outre-mer, transforment le Service statistique colonial en Service statistique des Territoires d'outre-mer (TOM). Ce dernier démarrera son activité en 1948, avec du personnel détaché de l'Insee (créé en avril 1946). Le décret du 29 mai 1947³⁴, qui fixe les modalités de la coordination de la statistique et des études économiques pour la métropole et la France d'outre-mer, définit le cadre des relations du Service statistique des TOM avec l'Insee. L'effectif initial du service est de 15 agents ; il atteindra 22 agents en 1958, sa dernière année d'existence.

La création du Service statistique des TOM introduit une rupture majeure dans la statistique coloniale. Ses archives (**encadré**) attestent de sa professionnalisation au cours des dernières années de la présence française outre-mer³⁵.

La mise en place du Service statistique des Territoires d'outre-mer

Le Service statistique des TOM est chargé de programmer et coordonner l'action statistique menée dans les territoires d'outre-mer, en particulier par les services statistiques locaux qui se développent dans les années 1950. Il est compétent sur l'ensemble des territoires d'outre-mer, à l'exception de ceux de l'Afrique du Nord.

³³ Le mouvement des non-alignés est né de la volonté des pays récemment décolonisés de refuser de s'engager dans la guerre froide, et d'affirmer leur autonomie vis-à-vis des États-Unis et de l'Union soviétique. Il prend forme lors de la conférence de Bandung (Indonésie) en avril 1955, laquelle réunit 29 États africains et asiatiques.

³⁴ Voir les références juridiques en fin d'article.

³⁵ Note établie par Amélie Hurel, conservatrice aux ANOM.

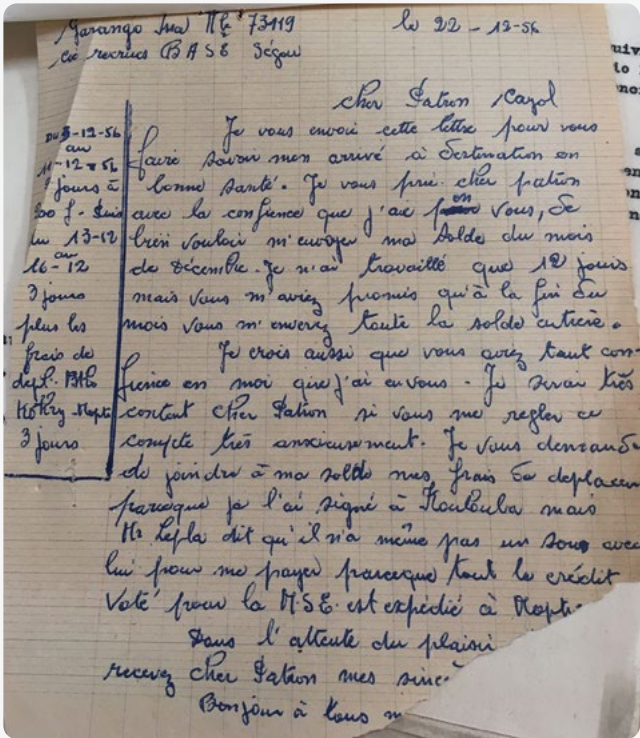
Un rôle qui correspond à de nouvelles préoccupations

Se consacrant d’abord à la conception d’annuaires statistiques d’ensemble, le service décentralise ensuite cette production aux territoires. Il se spécialise alors dans l’organisation d’enquêtes de terrain, menées sur des territoires et auprès de populations précises. Pour analyser plus finement la situation démographique, sociale et économique des TOM, il privilégie les enquêtes par sondage, organisées sous la forme de missions socio-économiques temporaires, effectuées par les statisticiens des services territoriaux.

Le Service statistique des TOM cesse son activité en octobre 1958, lors de la disparition de son ministère de tutelle, à la suite de l’indépendance de la plupart des territoires. Malgré sa courte existence, il a laissé de nombreuses archives (environ 40 mètres linéaires de documents techniques désormais accessibles aux ANOM).

Ces archives confirment le profond renouvellement des objectifs des enquêtes. Les préoccupations deviennent sociales et témoignent d’un intérêt nouveau pour les populations de l’Union française. Elles concernent les conditions de vie, les modes alimentaires, les niveaux de vie ainsi que les résultats économiques des territoires

► **Figure 6 - Enquête Soudan français 1955 : Lettre d'un enquêteur réclamant sa solde**



Source : Archives nationales d'outre-mer, © photo Béatrice Touchelay.

soutenus par le Fides³⁶. Les méthodes sont aussi profondément changées : les coûteux recensements exhaustifs sont remplacés par des enquêtes de terrain confiées aux statisticiens et à des enquêteurs formés sur place (*figures 6 et 7*).

► **Figure 7 - Enquête démographique dans un village de Haute-Guinée, 1955**

Cercle de Kankan - Canton Bati

Village Diousoumana (Foulah)

Village (Diousoumana)	N° lot	N° maison	N° habitant	caté 1	caté 2	caté 3	Total	Alcool 1955	Café 1955	Coton 1955	Cacao 1955	Café 1955	Cacao 1955	Café 1955	Cacao 1955	Tranche de 10 ans et +											
																Tranche de 10 ans et +				Tranche de 10 ans et +				Tranche de 10 ans et +			
																G	P	G	P	G	P	G	P	G	P	G	P
Diousoumana 1	33	51	103	236	6	5	73	360	14	8	109	7	3	10	3	97	96	112	38	67	24	81	23	46	41		
Diousoumana 2	21	36	71	180	7	8	38	248	5	11	107	11	4	6	4	46	49	82	69	12	14	36	58	42	31		
Diousoumana 3	32	54	146	495	19	11	50	375	6	17	151	12	2	8	5	97	101	146	73	73	18	98	65	40	20		
Total	86	141	320	911	32	24	166	983	25	36	367	26	9	26	8	240	246	340	173	93	56	195	206	170	107		
Hameaux																											
Foulah	9	11	30	61	2	6	21	97	4	8	46	7	3	2	2	26	37	46	21	30	2	4	7	7	3		
Diousoumana	1	2	5	10	0	3	9	0	0	3	0	0	0	0	0	3	2	5	2	1	1	2	1	2	1		
Koulikou	2	5	7	26	0	1	3	20	0	0	17	1	0	1	0	7	10	11	3	4	1	4	3	3	2		
Houssou	8	15	39	76	0	2	5	12	2	4	36	5	1	3	1	17	23	28	13	20	7	14	10	16	11		
Pilou	6	8	13	29	0	10	3	44	0	1	25	3	1	4	1	6	8	9	5	6	5	26	17	18	9		
TOTAL	105	187	371	915	37	45	191	1118	33	47	506	53	12	36	12	274	395	437	232	281	72	260	264	152	140		

$$\text{Taux natalité} = \frac{2600}{915} = 2,85\%$$

$$\text{Taux mortalité} = \frac{160}{915} = 16,4\%$$

$$\text{Taux fécondité 15-45 ans} = \frac{500}{110} = 4,55\%$$

$$\text{Taux moyen enfants par femme} = \frac{260}{110} = 2,36$$

$$\text{Taux mortalité infantile} = \frac{160}{56} = 2,86\%$$

$$\text{Age enfant - de 16 ans} = \frac{260}{110} = 2,36\%$$

$$\text{Population} = \frac{915}{110} = 8,32$$

Détail du recensement dans le village de Diousoumana, avec le calcul d'indicateurs démographiques en bas du tableau

Etat modèle 4

ETAT FINAL RECAPITULATIF DE HAMEAU OU DE VILLAGE (ILOP, SECTEUR)

Etabli un état par hameau, quartier, secteur, ilot

1 Cercle : Kankan

Subdivision : _____

Canton : Bati

Village : Diousoumana

Hameau : _____

Quartier : _____

Secteur : Secteur n° 1

Contrôleur : Camara Fawako

Equipe n° : _____

Agents recenseurs : Enquête pilote des agents recenseurs

Superviseur : _____

Date de l'étude : Arrivée : 29-12-54
Départ : 31-12-54

N°	Nom du chef de concession	N° de fiche collective	Nombre de ménages	Nombre d'habitations	Nombre de personnes	Tot
4			5	7		

Bilan du recensement dans le village de Diousoumana

Source : Archives nationales d'outre-mer, © photo Béatrice Touchelay.

³⁶ Quelques exemples disponibles à la bibliothèque des ANOM : Mission socio-économique du Soudan, *La pêche et données économiques diverses, Consommation et production de riz, Le niveau des prix*, Paris : Secrétariat d'état aux relations avec les États de la Communauté, 1960. BIB SOMC BR 3933 ; République du Mali. Mission socio-économique du Soudan, *L'alimentation des populations rurales du delta vif du Niger et de l'Office du Niger*, Paris : Secrétariat d'État aux relations avec les États de la Communauté, 1961. BIB AOMB 13570 ; Mission socio-économique du Soudan, *Étude démographique du Delta vif du Niger. 2^e fascicule, résultats détaillés*, Paris : Imprimerie Technigraphie, [s. d.]. BIB AOM 13222 ; Mission socio-économique du Soudan, *Enquête agricole (rendements) - Rapport provisoire n° 2*, Paris : ministère de la France d'outre-mer, 1957. BIB SOMC BR 3074 ; Mission socio-économique du Soudan, *Enquête démographique 1957-1958. Rapport provisoire n° 3*, [S. l.] : [s. n.], 1958. BIB SOMC BR 3642.

Exemples d'enquêtes approfondies

L'enquête socioéconomique de 1957 au Soudan est un exemple de ces enquêtes approfondies. Elle montre combien les statistiques antérieures étaient incomplètes et approximatives. Les statisticiens découvrent l'importance de la mortalité infantile et la faiblesse de l'espérance de vie dans certaines régions. Les archives permettent de reconstituer les étapes de la préparation et de la réalisation des enquêtes, ce qui est rare. Elles éclairent aussi de façon inédite les relations entre les statisticiens professionnels et les enquêteurs formés sur place, donnant à voir les carnets de note des enquêteurs, les cahiers qu'ils utilisent pendant leur formation, etc.

Dans l'enquête sur les manœuvres, par exemple, les formulaires individuels pré-imprimés sont remplis au stylo par les enquêteurs et vérifiés par les contrôleurs. Ils comportent souvent des commentaires sur la « qualité des réponses » obtenues ou celle de la main-d'œuvre selon les lieux de provenance. Ces éléments disparaissent au moment de la totalisation des bulletins.

Les archives des années 1945-1958 qui ont été consultées donnent une mesure de la faiblesse des connaissances démographiques et des conditions de vie. Ces connaissances sont disparates et discontinues, ce qui les rend difficiles à utiliser. L'autre partie du fonds récemment inventorié aux ANOM, qui est amianté et donc toujours inaccessible, permettra peut-être de combler certains vides.

Les fonds montrent aussi à quel point les pratiques statistiques restent artisanales. Les cartes des villages et de l'emplacement des cases sont tracées au crayon. La notification systématique de l'emplacement de la case du chef de village montre qu'il joue toujours le rôle d'intermédiaire indispensable aux enquêtes. Ces enquêtes pionnières prennent fin en 1958 au moment des indépendances. Elles ont montré que les statistiques sorties du cadre colonial n'étaient pas des instruments au service du pouvoir et de la domination, mais qu'elles servaient à éclairer l'action.

► Des statistiques qui questionnent les statisticiens



La collecte et la restitution des données témoignent d'une situation d'asymétrie typique de la situation coloniale.



L'expérience des statistiques coloniales mise en relief par les archives est riche d'enseignements. Pour n'en retenir que quelques-uns, elle montre d'abord que la production des statistiques ne s'improvise pas, mais qu'elle exige des compétences spécifiques. Dépourvus de formation, les administrateurs coloniaux ne comprennent ni les enjeux, ni la nécessité des enquêtes.

La collecte et la restitution des données témoignent également d'une **situation d'asymétrie** typique de la situation coloniale. Les personnes interrogées, leurs interprètes et tous les intermédiaires impliqués dans les enquêtes sont dans une relation déséquilibrée, qui fausse à la fois la réception et les résultats de la démarche. Ces résultats sont d'autant plus aléatoires que les enquêteurs imposent leurs conceptions – « compter c'est d'abord concevoir », selon Alain Desrosières – et leurs classifications... sans que

les enquêtes ne mesurent ni les motifs, ni les effets de leur classement. Alors que les classifications statistiques résultent de consensus plus ou moins conscients mais toujours négociés dans les sociétés colonisatrices, elles sont imposées aux populations colonisées dont elles ne respectent pas les coutumes. Ces formes de comptage et cette mise en case forcée éloignent les statistiques de la réalité qu'elles sont censées représenter, elles créent des illusions, font écran au lieu de refléter. Au bout du compte, ces instruments fabriquent une réalité plus qu'ils ne la reflètent et les capacités de pilotage de leurs commanditaires sont limitées.

De nombreux administrateurs sont incapables de remplir correctement les tableaux préformatés transmis par leurs supérieurs hiérarchiques³⁷. L'un des quiproquos les plus visibles réside dans la définition de la « famille », qui est pourtant un des critères de classement démographique très fréquent. Le cadre de référence est celui de la « famille » française, c'est-à-dire d'un groupe d'individus du même nom, composé de deux générations, voire trois dans les milieux ruraux : les deux parents et leurs enfants. Dans les territoires colonisés d'Afrique subsaharienne du XIX^e et du XX^e siècle, la « famille » est souvent beaucoup plus large. Elle correspond parfois au « village », autre entité introduite à des fins de contrôle par la colonisation, à la lignée (un ancêtre commun), à

des croyances ou des rites partagés, etc. Il est difficile de savoir dans quelle mesure les rédacteurs des tableaux avaient conscience du décalage entre la réalité et les résultats statistiques qu'ils transmettaient.

**“ En dessinant
une réalité
très partielle,
les statistiques
coloniales donnent
l'illusion d'informer
et de contrôler. ”**

Quelques rapports attestent de lucidité lorsqu'ils dénoncent l'insuffisance des moyens attribués aux enquêtes et aux recensements ; d'autres se permettent des digressions en précisant le nombre d'épouses, alors que cette information n'est pas réclamée ; d'autres encore utilisent les unités de mesure locales, avec ou non conversion en mètres ou en argent... Mais la plupart

contournent les difficultés, en mettant des chiffres faux, en indiquant qu'ils ne peuvent pas répondre ou en laissant la colonne vide. Convaincus de la légitimité de la colonisation, de la supériorité du « modèle » européen et qu'ils participent à une mission civilisatrice légitime, ou bien tout simplement pressés par le temps³⁸, attachés à graver les échelons et à préserver leur carrière, les producteurs de statistiques coloniales ne discutent pas les catégories. Les questions apparaissent progressivement à mesure de la contestation de la supériorité du modèle imposé aux colonisés. Elles sont posées ouvertement à partir de 1945 par les statisticiens professionnels qui ne peuvent que constater l'inadéquation des catégories importées et des sociétés qu'elles sont censées représenter.

³⁷ Voir les exemples des notes 3 et 4.

³⁸ La faiblesse des moyens des administrateurs et l'insuffisance de leur effectif sont dénoncées dans la plupart des mémoires qu'ils ont publiés, voir par exemple Delafosse (1909).

Ainsi, les archives des statistiques coloniales montrent la colonisation en action, elles véhiculent un rapport de domination et illustrent sa violence qui fait fi des classifications antérieures. Ces statistiques posent beaucoup de questions aux statisticiens, sur le fondement et les effets des catégories et des nomenclatures qu'ils définissent, sur les rapports aux territoires et aux populations que leurs statistiques sont censées refléter, sur leur part dans la normalisation des cadres de pensée.

Les statistiques coloniales servent à rendre compte à la métropole, à assurer l'exploitation ponctuelle des ressources et à attirer des colons (Touchelay, 2024). En dessinant une réalité très partielle, elles donnent l'illusion d'informer et de contrôler. Leur héritage est persistant : dès les indépendances, la plupart des États anciennement colonisés se dotent de services de statistiques sans changer radicalement les cadres de pensée introduits pendant la colonisation.

► Fondements juridiques

- Loi du 20 mars 1894 portant création d'un ministère des colonies. In : *site de la Bibliothèque nationale de France*. [en ligne]. [Consulté le 26 janvier 2026]. Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k20162874>.
- Loi du 13 avril 1900 portant fixation du budget général des dépenses et des recettes de l'exercice 1900. In : *site de Légifrance*. [en ligne]. [Consulté le 26 janvier 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000314852>.
- Décret n° 47-963 du 29 mai 1947 fixant les modalités de la coordination de la statistique et des études économiques pour la métropole et la France d'outre-mer. In : *site de Légifrance*. [en ligne]. [Consulté le 26 janvier 2026]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000299679>.

► Bibliographie

- BEAUREPAIRE, Camille et BOUCHER, Thierry, 2025. Des données de papier – Une histoire des échanges de publications statistiques. In : *Courrier des statistiques*. [en ligne]. 15 décembre 2025. Insee. N° N14, pp. 27-50. [Consulté le 6 mars 2026]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8675626?sommaire=8667496>.
- BOWKER, Geoffrey C. et STAR, Susan Leigh, 1999. *Sorting Things Out: Classification and Its Consequences*. The MIT Press. ISBN 9780262269070.
- DELAFOSSE, Maurice, 1909. Les états d'âme d'un colonial. In : *site de la Bibliothèque nationale de France*. [en ligne]. Comité de l'Afrique française. [Consulté le 17 mars 2026]. Disponible à l'adresse : <https://catalogue.bnf.fr/ark:/12148/cb32007188v>.
- DESROSÈRES, Alain, 2008. *Pour une sociologie historique de la quantification. L'argument statistique 1*. Septembre 2009. Presses des Mines. P. 18. ISBN 9782356710024.
- GERVAIS, Raymond, 1994. Recensements en AOF : genèse et signification – Des exemples de la Haute-Volta coloniale. In : *Annales de Démographie Historique*. [en ligne]. Société de Démographie Historique – EHESS. Pp. 339-354. [Consulté le 6 mars 2026]. Disponible à l'adresse : https://www.persee.fr/doc/adh_0066-2062_1994_num_1994_1_1877.
- GUIBERT, Bernard, LAGANIER, Jean et VOLLE, Michel, 1971. Essai sur les nomenclatures industrielles. In : *Économie et Statistique*. [en ligne]. Février 1971. N° 20, pp. 23-36. [Consulté le 9 janvier 2026]. Disponible à l'adresse : https://www.persee.fr/doc/estat_0336-1454_1971_num_20_1_6122.
- MALÈGUE, Fanny, 2018. L'empire en tableaux – Recenser et gouverner les colonies antillaises après la guerre de Sept Ans. In : *Histoire & mesure*. [en ligne]. N° XXXIII-2, pp. 93-114. [Consulté le 9 janvier 2026]. Disponible à l'adresse : <https://doi.org/10.4000/histoiremesure.8038>.
- PERROT, Michelle, 1987. Première mesure des faits sociaux : les débuts de la statistique criminelle en France (1780-1830). In : *Pour une histoire de la statistique*. INSEE Economica. Tome 1, Contributions, 2^e édition. Pp. 125-135. ISBN 2-7178-1260-1.
- SANNER, Pierre, 1994. Contribution à un mémorial du service colonial de statistiques 1923-1958. In : *Journal de la société statistique de Paris*. [en ligne]. Tome 135, n° 1, pp. 73-99. [Consulté le 09 janvier 2026]. Disponible à l'adresse : https://www.numdam.org/item/JFSF_1994__135_1_73_0/.
- SEMBÈNE, Ousmane, 1960. *Les Bouts de bois de Dieu*. Pocket. ISBN 9782266106313.
- TIQUET, Romain, 2017. Rendre compte pour ne pas avoir à rendre des comptes – Pour une réflexion sur l'écrit administratif en situation coloniale (Sénégal, années 1920-1950). In : *Cahiers d'histoire – Revue d'histoire critique*. [en ligne]. N° 137, pp. 123-140. [Consulté le 9 janvier 2026]. Disponible à l'adresse : <https://journals.openedition.org/chrh/6427>.
- TISSEAU, Violaine, 2017. *Être métis en Imerina (Madagascar) aux XIX^e-XX^e siècles*. 22 mai 2017. Karthala. Hommes et sociétés. ISBN 9782811118556.

- TOUCHELAY, Béatrice, 2019. British and French Colonial Statistics: Development by Hybridization from the Nineteenth to the Mid-Twentieth Centuries. In : *British and French Colonialism in Africa, Asia and the Middle East – Connected Empires across the Eighteenth to the Twentieth Centuries*. 14 août 2019. Palgrave Macmillan. Pp. 249-274. ISBN 978-3-319-97963-2.
- TOUCHELAY, Béatrice, 2024. Ce que compter veut dire en situation coloniale et impériale – Introduction. In : *Histoire & mesure*. [en ligne]. N° XXXIX-1, pp. 5-18. [Consulté le 9 janvier 2026]. Disponible à l'adresse : <https://doi.org/10.4000/12ht4>.
- ULMER, Henri, 1938. La statistique dans les pays coloniaux. In : *Journal de la Société de statistique de Paris*. [en ligne]. Tome 79, pp. 231-250. [Consulté le 9 janvier 2026]. Disponible à l'adresse : https://www.numdam.org/item/JSFS_1938__79__231_0/.



PRÉSENTATION DU NUMÉRO N15

Ce numéro N15 du Courrier des statistiques dévoile les coulisses de chantiers d'ampleur de la statistique publique. D'abord l'enquête Sans domicile 2025, une opération hors normes dont l'enjeu est de mieux connaître la situation des personnes dans les situations les plus précaires de notre société. Puis la refonte de la nomenclature d'activités française, en articulation avec celle des nomenclatures européenne et onusienne, pour mieux refléter le monde économique actuel. Le troisième papier poursuit le récit de la construction du système statistique européen, débuté dans le numéro précédent : entre 2004 et 2015, deux grandes crises de confiance vont conduire à consolider fortement le socle juridique de ce « partenariat ».

Les trois articles suivants parlent d'innovation. Le premier décrit l'exploitation, coordonnée par Eurostat, de données de plateformes d'hébergement de courte durée pour compléter les statistiques du tourisme. Le deuxième raconte comment les métiers de la *data* se sont développés, pour traiter des données de plus en plus nombreuses et multiformes, et comment la statistique publique mobilise ces nouvelles façons de travailler. Ces outils servent aussi à certaines instances d'évaluation : le troisième papier présente le pôle *data* de la Cour des comptes et son importante contribution aux évaluations de politiques publiques menées par l'institution.

Le numéro s'achève sur l'histoire des statistiques coloniales en France et les archives qui leur sont associées. Ces dernières soulèvent des questions plus larges sur le sens et l'impact des catégories statistiques et sur les motivations de ceux qui les portent.



ISSN 2107-0903

ISBN 978-2-11-162504-4



9 782111 870055

Insee
Mesurer pour comprendre