

Décembre 2024

Courrier des statistiques 12



Rédaction en chef

Catherine Fresson-Martinez

Contribution

Insee : Odran Bonnet, Mylène Chaleix,
Jean-Michel Durr (retraité), Boris Guannel,
Tristan Loisel, Sylvain Moreau
CépiDc : Élise Coudin, Aude Robert
DEPS : Amandine Schreiber, Laure Turner
DSECE : Ketty Attal-Toubert
Sies : Karl Even

Directeur de la publication

Jean-Luc Tavernier

Directeur de la collection

Pascal Rivière

Rédaction

Catherine Fresson-Martinez, Solenn Ily,
Marine Le Roux, 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 2024 « Reproduction partielle
autorisée sous réserve de la mention
de la source et de l'auteur ».



Courrier des statistiques N12

SOMMAIRE

Présentation du numéro <i>Catherine Fresson-Martinez</i>	4
Les statistiques du commerce extérieur - Spécificités, défis et dimension européenne <i>Ketty Attal-Toubert et Boris Guannel</i>	7
Les statistiques sur les causes de décès - Classer et coder... dans la classification internationale des maladies <i>Élise Coudin et Aude Robert</i>	27
Mesurer et qualifier l'insertion professionnelle des étudiants quittant l'enseignement supérieur - Le dispositif InserSup <i>Karl Even</i>	51
Des chiffres pour la culture - 60 ans d'objectivation scientifique au service des politiques culturelles <i>Amandine Schreiber et Laure Turner</i>	71
La statistique annuelle d'entreprises : sa nature, son histoire, ses enjeux <i>Sylvain Moreau</i>	93
L'économie racontée par les données bancaires - Ce que nos relevés de comptes disent de nous <i>Odran Bonnet et Tristan Loisel</i>	115
Les revues par les pairs : un levier au service de la statistique publique <i>Mylène Chaleix et Jean-Michel Durr</i>	137

PRÉSENTATION DU NUMÉRO

L'histoire de la statistique, les outils liés aux enquêtes, les règlements européens, les sources administratives... Depuis sa renaissance en 2018, le Courrier des statistiques vous invite à découvrir le système statistique national au travers d'articles explorant des sujets très divers, faisant ainsi intervenir des auteurs issus des différentes structures de la statistique publique. Ce numéro ne fait pas exception à cette règle, avec quatre articles issus de services statistiques ministériels (SSM) et d'une autre autorité nationale (en anglais, *other national authority* : ONA) qui participent pour la première fois au Courrier des statistiques. Trois articles complètent ce numéro : un article socle sur la statistique structurelle d'entreprises, un article plus institutionnel sur le processus de la revue par les pairs et un dernier article informant sur l'utilisation de données privées, prélude au prochain numéro qui y consacrera un dossier.

La production de toute statistique provient d'un besoin, besoin datant parfois de plusieurs siècles. Il en est ainsi des statistiques du commerce extérieur. L'article de Ketty Attal-Toubert et Boris Guannel revient sur leur genèse et témoigne de leur importance pour la France et à l'échelon européen, hier comme aujourd'hui. En effet, mesurer les importations, les exportations et le solde commercial reste au cœur des débats, et l'intérêt qu'on leur prête a été amplifié par la mondialisation. Récemment, la coopération entre les Autorités nationales statistiques de l'Union européenne en charge du commerce extérieur s'est développée afin de faire face aux enjeux émergeant de la numérisation de l'économie. Des règlements européens se sont imposés pour permettre l'élaboration et la diffusion de statistiques cohérentes entre les États membres.

Dans un autre registre, celui des statistiques de la culture, Amandine Schreiber et Laure Turner nous font découvrir comment depuis plus de 60 ans, les enquêtes comme celle sur les pratiques culturelles des Français permettent d'établir des constats objectifs utiles aux politiques, dans un environnement en perpétuel mouvement du fait des innovations technologiques successives. Par ailleurs, l'enjeu économique est fort et les acteurs sont très divers. Pour suivre ces professionnels et mesurer leur activité, un cadre européen harmonisé a été établi. Mais les particularités du secteur culturel imposent d'avoir recours à des nomenclatures spécifiques.

Le CépiDc (Centre d'épidémiologie sur les causes médicales de décès) de l'Inserm est, avec le CASD (Centre d'accès sécurisé aux données), un des deux ONA à fournir des statistiques officielles, et nous entraîne dans le domaine médical. Pouvoir mesurer précisément les causes de décès est un impératif pour comprendre la situation sanitaire de la France. Cela requiert une codification pertinente de ces causes à partir du certificat de décès renseigné par les médecins. En préambule, Élise Coudin et Aude Robert racontent l'histoire des statistiques sur les causes de décès depuis le XIX^e siècle puis elles expliquent les méthodes de codage utilisées, de la codification manuelle au système de règles puis à l'utilisation de l'Intelligence artificielle avec le codage par prédictions de réseaux de neurones profonds.

Comment mesurer l'insertion professionnelle des étudiants quittant l'enseignement supérieur ? Karl Even nous présente le dispositif InserSup déployé par la sous-direction des systèmes d'information et des études statistiques (Sies) du ministère de l'Enseignement supérieur et de la Recherche fin 2023. Ce dispositif rassemble de nombreuses données administratives exhaustives, émanant de différents ministères. Ces données sont appariées afin de produire et mettre à disposition, à un niveau très fin, une multitude d'indicateurs

pour qualifier l'insertion des jeunes, selon leur formation par exemple. Ce système permet, entre autres, aux étudiants de choisir une formation en pleine connaissance des possibilités d'emploi à l'issue de ce cursus, en se connectant aux différentes plateformes de poursuite d'études.

Après la découverte de quatre milieux très différents les uns des autres, un retour à l'Insee, avec un article sur la statistique structurelle d'entreprises. Dans les années 70, l'Insee a conçu un système statistique très élaboré pour l'observation sur les entreprises. Sylvain Moreau relate cette construction, bâtie sur des données administratives complétées par des enquêtes permettant de construire des bases de données micro-économiques sur les entreprises. Tout en conservant les définitions du début, ce système a cependant considérablement évolué et doit continuer de s'adapter à la complexité des organisations. C'est tout l'enjeu des prochaines années.

Dans ce numéro également, une fois n'est pas coutume, un article institutionnel. Mylène Chaleix et Jean-Michel Durr décrivent le processus de la revue par les pairs. Afin de s'assurer que le Code des bonnes pratiques en matière de statistiques est respecté par les instituts nationaux de statistiques (INS), une équipe constituée de quatre experts va auditionner un système statistique national (INS et ONA). Dans cet article, les trois cycles des revues sont présentés avec leurs recommandations et leurs conséquences pour la France mais aussi dans d'autres pays. La confiance dans les statistiques officielles est un enjeu important pour le système européen. Si les revues par les pairs existent depuis presque 20 ans, leur méthodologie est susceptible d'évoluer lors des prochains cycles.

Enfin, un article sur l'utilisation de données privées à des fins statistiques nous décrit les données bancaires exploitées par l'Insee pour documenter l'évolution de la situation financière des Français lors de l'épisode de la Covid-19 ou encore les impacts d'une remise à la pompe sur la consommation de carburant des ménages. Odran Bonnet et Tristan Loisel expliquent comment l'Insee a accès depuis 2020 à des échantillons anonymisés de comptes bancaires de deux grandes banques : La Banque Postale et Crédit Mutuel Alliance Fédérale. De nombreux défis ont été relevés pour exploiter ces données, qui apportent un regard complémentaire sur l'analyse conjoncturelle économique. Comment résoudre les difficultés inhérentes à l'utilisation de données conçues à des fins de gestion bancaire ? Des premières réponses sont apportées.

L'Insee s'intéresse à d'autres données privées à des fins d'usage statistique : par exemple, les données de carte bancaire ou les données de téléphonie mobile. Mais cela est une autre histoire à découvrir dans un dossier du prochain numéro du Courrier des statistiques. À suivre...

Catherine Fresson-Martinez
Rédactrice en chef, Insee

Les statistiques du commerce extérieur

Spécificités, défis et dimension européenne



Ketty Attal-Toubert et Boris Guannel***

Les statistiques du commerce extérieur de biens ont été instituées en France il y a environ trois siècles pour répondre au besoin émergent des dirigeants du pays de disposer d'informations précises sur les marchandises entrantes et sortantes du territoire. Depuis plus de deux cents ans, cette mission est confiée aux douanes qui abritent aujourd'hui le service statistique ministériel du commerce extérieur. La mesure des importations, des exportations et du solde commercial alimente toujours les débats et décisions publics, notamment dans le contexte de la mondialisation et des crises récentes (Covid-19, guerre en Ukraine, etc.).

Ces indicateurs, encadrés par des règlements européens exigeants, présentent des spécificités, telles que la finesse de la nomenclature de produits utilisée aussi bien pour la collecte que pour la diffusion des données, la notion de secret statistique « passif », ou encore la possibilité de disposer d'une double vision d'un même flux de part et d'autre de la frontière, en miroir.

La numérisation de l'économie, en particulier l'essor du e-commerce, entraîne depuis quelques années d'importantes mutations des formalités administratives liées au commerce international. Ces changements sont un véritable défi pour ces statistiques d'échanges de biens, dans lequel la coopération européenne joue un rôle important.

 International trade in goods statistics were introduced in France around three centuries ago to meet the emerging need of the country's leaders for accurate information on goods entering and leaving the country. For over two hundred years, this function was assigned to the customs authorities, which today hosts the External Trade Ministerial Statistical Office. The measurement of imports, exports and trade balance continues to contribute to public debate and decisions, particularly in the context of globalisation and recent crises (Covid-19, the Ukraine war, etc.).

These indicators, which are governed by strict European regulations, have a number of specific features, such as the level of detail of the product classification used for both data collection and dissemination, the notion of 'passive' confidentiality of statistics, and the possibility of having a double vision of the same flow on either side of the border, in mirror image.

In recent years, the digitalisation of the economy, in particular the rise of e-commerce, has led to major changes in the administrative formalities involved in international trade. These changes are a real challenge for statistics relating to trade in goods, and one in which European cooperation has an important role to play.

* Cheffe du département des statistiques et des études du commerce extérieur (DSECE), Ministère chargé du Budget et des Comptes publics, direction générale des douanes et droits indirects.
ketty.attal-toubert@douane.finances.gouv.fr

** À la date de la rédaction, adjoint à la cheffe du département des statistiques et des études du commerce extérieur (DSECE), Ministère chargé du Budget et des Comptes publics, direction générale des douanes et droits indirects.
boris.guannel@insee.fr

En France, les statistiques du commerce extérieur, c'est-à-dire la mesure des montants et des quantités de marchandises importées et exportées, existent depuis très longtemps, puisqu'on retrouve les premiers dispositifs administratifs de collecte centralisée de ces flux durant l'Ancien Régime. Connaître les flux d'entrée et de sortie dans le Royaume de France était jugé nécessaire à la mise en œuvre des politiques économiques mercantilistes de l'époque qui visaient à contrôler les importations et à développer les exportations. Aujourd'hui encore, l'intérêt économique et politique pour ces statistiques demeure. Aux niveaux national comme européen, ces statistiques sont considérées comme un indicateur « clé » de l'analyse économique et des politiques publiques, notamment commerciales et monétaires.

► **Les statistiques du commerce extérieur : un besoin existant depuis trois siècles...**

Les origines des statistiques françaises du commerce extérieur sont lointaines. Ainsi, déjà en 1713, quelques jours avant la signature du traité d'Utrecht qui mit fin à la Guerre de Succession d'Espagne, l'envoyé français en charge de l'économie rapportait à son ministre qu'il fallait « faire un état général de tout ce qui est en France, et de ce qui en sort, pour connaître si nos dettes actives excèdent les passives, les espèces de marchandises qui nous causent le plus de dépenses, et chercher les moyens de les diminuer, pour savoir enfin les États qui apportent du profit à la France pour [sic] leur commerce, afin de cultiver l'alliance de leurs souverains, et regarder avec indifférence celles de ceux dont la correspondance pourrait nous être préjudiciable ». Ainsi fut créé, deux semaines plus tard, le « Bureau de la balance du commerce » (Charles et Daudin, 2011), en réponse au désavantage ressenti par les négociateurs français au congrès d'Utrecht face à leurs homologues anglais qui disposaient depuis 1693 de statistiques détaillées sur les exportations et importations de leur Royaume. La mission de ce bureau était de collecter l'information sur les échanges avec l'étranger, les colonies et comptoirs commerciaux établis outre-mer. Au cours de son existence, de 1713 à 1791, il a ainsi pu produire une série, presque complète, de statistiques sur le commerce extérieur (Beaud, 1964).

► **... dont la collecte a été confiée à la douane dès la Révolution**

Durant la Révolution française, la Régie nationale des douanes¹ hérita naturellement des anciennes attributions du Bureau de la balance du commerce en matière de collecte et de production des statistiques sur les importations et exportations. Cette activité occupait une place importante au sein de cette administration. Ainsi, il était demandé aux directeurs régionaux de la Régie que « les opérations relatives à la balance du commerce soient considérées par tous les préposés [...] comme une partie essentielle de leurs occupations » (Association pour l'histoire de l'administration des douanes, 2024 ; Clinquart, 1978). Ces statistiques étaient produites trimestriellement et annuellement, et leur communication était réservée aux seuls ministres.

¹ Créée le 1^{er} mai 1791, la Régie avait notamment la mission de percevoir les droits de douanes sur les biens importés et ses 15 000 agents étaient décrits comme des « préposés à la police du commerce extérieur ».

À partir de 1825, sous Charles X, apparaît une publication annuelle désormais imprimée, produite par l'administration des douanes : le Tableau général du commerce de la France avec ses colonies et les puissances étrangères (**tableau**).

► **Tableau général du commerce de la France avec ses colonies et les puissances étrangères - 1825**

DÉSIGNATION DES MARCHANDISES.	UNITÉS.	IMPORTATIONS.						
		TOTAL DES QUANTITÉS ENTRÉES, soit pour l'entrepôt, soit pour la consommation, soit pour le transit.		QUANTITÉS ADMISES POUR LA CONSOMMATION avec paiement de droits.				
		Quantités.	Valeurs.	Quantités.	Valeurs.	Droits perçus.		
REPRODUCTION.								
MATIÈRES NÉCESSAIRES À L'INDUSTRIE.								
SUBSTANCES propres à la médecine et à la parfumerie.	Sangues.....	Pièce.	9,041,225.	271,236.	9,041,225.	271,236.	9,945.	
	Cantharides, cloportes.....	Kilog.	9,750.	72,150.	6,715.	49,691.	4,958.	
	Musc.....	Idem.	35.	15,050.	18.	7,740.	2,468.	
	Castoreum.....	Idem.	167.	50,100.	94.	28,200.	194.	
	Antales et coris.....	Idem.	168.	504.	"	"	"	
	Vessies.....	Idem.	43.	258.	43.	258.	6.	
	Corne de cerf en morceaux.....	Idem.	4,872.	3,410.	4,872.	3,410.	271.	
	Râpure de cornes de cerf et d'ivoire.....	Idem.	2,171.	15,19.	2,171.	15,19.	235.	
	Ambre gris.....	Idem.	7.	9,800.	7.	9,800.	467.	
	Éponges.....	} communes.....	Idem.	111,398.	557,990.	61,003.	305,015.	42,504.
			Idem.	7,398.	118,368.	4,852.	77,632.	11,064.
	Yeux d'écrevisse.....	Idem.	168.	504.	168.	504.	34.	

DÉSIGNATION DES MARCHANDISES.	UNITÉS.	EXPORTATIONS.						
		TOTAL DES QUANTITÉS SORTIES, en payant qu'en soit l'origine.		QUANTITÉS EXPORTÉES DE L'INTÉRIEUR DU ROYAUME, en payant des droits en franchise ou avec prime.				
		Quantités.	Valeurs.	Quantités.	Valeurs.	Droits perçus.		
PRODUITS NATURELS.								
SUBSTANCES propres à la médecine et à la parfumerie.	Sangues.....	Pièce.	688,460.	20,653.	688,460.	20,653.	370.	
	Cantharides, cloportes.....	Kilogr.	5,987.	44,303.	3,624.	26,818.	9.	
	Musc.....	Idem.	26.	11,180.	20.	8,600.	"	
	Castoreum et civette.....	Idem.	26.	7,800.	9.	2,700.	"	
	Antales.....	Idem.	28,666.	85,998.	18,948.	56,844.	42.	
	Vessies.....	Idem.	15,005.	90,030.	13,085.	78,510.	33.	
	Sang de bouc.....	Idem.	1,293.	5,172.	"	"	"	
	Râpure de cornes de cerf et d'ivoire.....	Idem.	1,820.	1,274.	1,267.	887.	4.	
	Ambre gris.....	Idem.	34.	47,600.	24.	33,600.	"	
	Éponges.....	} communes.....	Idem.	11,232.	67,392.	1,162.	6,972.	3.
			Idem.	1,972.	39,440.	1,369.	27,380.	4.
	MATIÈRES dures à tailler.	} brutes.....	Idem.	66,325.	59,692.	3,138.	3,184.	778.
Idem.			21,925.	4,385.	21,925.	4,385.	68.	
FRUITS.....	} à distiller.....	Kilogr.	148,194.	88,916.	124,286.	74,572.	338.	
		Idem.	1,377,243.	2,065,864.	1,363,186.	2,044,779.	4,190.	
TABAC en feuilles et en côtes.....	Idem.	2,140,049.	428,010.	589,265.	117,853.	6,431.		

Source : Douanes royales de France, 1825 ; base de données « Montesquieu » (Becuwe et alii., 2019).

De nos jours encore, la Direction générale des douanes et droits indirects est l'institution qui collecte les informations sur les flux d'importation et d'exportation de biens, dans le cadre de ses missions d'administration de la frontière, de contrôle des marchandises mais aussi d'information statistique². Au sein de cette institution, la production et la diffusion des statistiques du commerce extérieur sont assurées par le département des statistiques et des études du commerce extérieur (DSECE), l'un des seize services statistiques ministériels (SSM), en collaboration avec trois services douaniers (situés à Toulouse, Lille et Garges-Lès-Gonesse) chargés de la collecte et du traitement des données.

► Une statistique toujours plus importante dans le contexte actuel de fort essor des échanges internationaux de biens



Les statistiques du commerce extérieur, bien que parmi les plus anciennes de la statistique publique, sont aujourd'hui encore au cœur des décisions économiques et politiques.



Les statistiques du commerce extérieur, bien que parmi les plus anciennes de la statistique publique, sont aujourd'hui encore au cœur des décisions économiques et politiques. En effet, malgré la tertiarisation grandissante de l'économie et, plus récemment, le développement des échanges dématérialisés (par exemple, le recours aux plateformes d'achat de livres numériques, de VOD ou de musique plutôt que l'acquisition de leur équivalent physique), les échanges de biens

continuent de représenter une part prépondérante du commerce international de la France. Ainsi, en 2023, le montant des exportations de biens était 1,8 fois plus important que celui des exportations de services ; celui des importations de biens était 2,3 fois plus important que celui des importations de services (Banque de France, 2024).

Ces statistiques répondent, comme à leur origine, au besoin de contrôler les flux entrants, d'évaluer la compétitivité et de définir la politique commerciale du pays. Suivre et analyser les échanges entre les pays est d'autant plus crucial dans le contexte de la mondialisation économique. S'il a existé plusieurs phases de mondialisation (du milieu du 19^e siècle à la Première Guerre mondiale, puis dans les années 1970), celle-ci s'intensifie à partir du début des années 1990. Ainsi, d'importants accords de libre-échange sont mis en place (Mercosur³, Alena⁴, etc.), l'Organisation mondiale du commerce succède au *General Agreement on Tariffs and Trade* (GATT)⁵ en 1995 et le nombre de multinationales s'accroît fortement.

Le suivi des échanges est essentiel dans le contexte de crise. En effet, les crises renforcent la prise de conscience des dépendances vis-à-vis des autres pays pour certains produits stratégiques. Dans l'actualité récente, en l'espace de deux ans entre 2020 et 2022,

² Voir les références juridiques en fin d'article.

³ Le Marché commun du Sud, couramment abrégé Mercosur est une zone de libre-échange qui regroupe plusieurs pays de l'Amérique du Sud. Il a été créé en 1991.

⁴ Alena : l'Accord de libre-échange nord-américain, entré en vigueur en 1994, est un accord instituant une zone de libre-échange entre le Canada, les États-Unis et le Mexique.

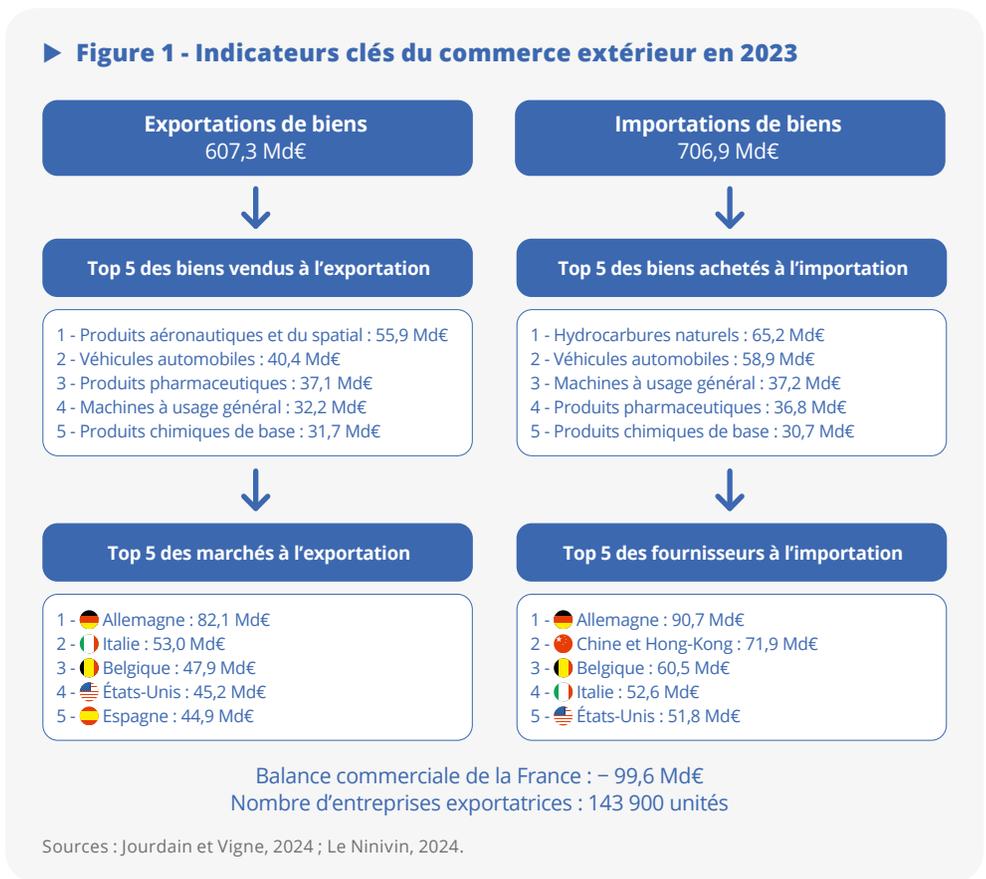
⁵ Le GATT (Accord général sur les tarifs douaniers et le commerce, 1947-1995) est un cadre juridique et financier mis en place à l'échelle internationale au sortir de la Seconde Guerre mondiale, et qui a existé jusqu'à la création de l'Organisation mondiale du commerce en 1995.

la succession de crises rapprochées (la pandémie de Covid-19, les tensions sur les matières premières et le conflit russo-ukrainien) a mis en lumière les conséquences de la mondialisation pour l'ensemble des agents économiques de la nation. Les statistiques du commerce extérieur ont contribué à suivre et analyser ces phénomènes.

Par exemple, la crise de la Covid-19 a révélé la vulnérabilité liée à cette dépendance, notamment pour les produits de santé. Les statistiques du commerce extérieur ont permis de suivre leurs approvisionnements, aux niveaux national et européen grâce à une coordination des États membres.

Concernant la vulnérabilité énergétique de la France et l'impact de l'embargo sur le pétrole russe dans le contexte de la guerre en Ukraine et de la crise énergétique (Vigne, 2023), la mesure des flux et du poids des partenaires dans les échanges est également indispensable. Il en est de même pour d'autres thématiques telles que l'autonomie alimentaire des départements et régions d'outre-mer (DROM), la dépendance aux importations de produits liés à la transition écologique (voitures électriques, batteries, métaux stratégiques, panneaux solaires, etc.) ou le développement de l'intelligence artificielle (composants électroniques).

Ainsi, les statistiques du commerce extérieur ont peut-être encore davantage qu'auparavant une place prépondérante au sein de la statistique publique (*figure 1*).



► Un indicateur « clé », au niveau national comme au niveau européen...

Au même titre que les comptes nationaux, le chômage ou les prix à la consommation, les statistiques du commerce extérieur jouent un rôle central pour la statistique publique française. En effet, au niveau national, elles permettent de répondre à divers besoins :

- alimenter des productions phares du service statistique public. Au sein de celui-ci, ces statistiques constituent un intrant pour des productions majeures : les comptes nationaux établis par l'Insee⁶ et la balance de paiements établie par la Banque de France ;
- éclairer les débats et décisions publics. D'une part, les échanges extérieurs de biens composent, via le solde extérieur, une sous-composante du PIB. De ce fait, ils représentent un facteur incontournable d'analyse conjoncturelle de la croissance du PIB. D'autre part, ces statistiques sont également utilisées, souvent au niveau le plus fin de la nomenclature des produits, afin de produire des indicateurs de compétitivité extérieure de la France et éclairer la décision publique ;
- réaliser des études économiques dans le domaine du commerce international. Les données individuelles d'échanges de biens sont exploitées par les chargés d'études et chercheurs, pour répondre à de nombreuses questions sur l'économie internationale.

Au total, les statistiques du commerce extérieur constituent une véritable « consommation intermédiaire » pour le système statistique public français.



Ces statistiques font partie des principaux indicateurs économiques européens (PIEE), ensemble de 26 indicateurs essentiels pour la surveillance de la zone euro.



Par ailleurs, du point de vue des instances publiques européennes, ces statistiques font partie des principaux indicateurs économiques européens (PIEE), ensemble de 26 indicateurs essentiels pour la surveillance de la zone euro⁷ (*Economic and financial committee*, 2008).

La mise en place du marché unique européen le 1^{er} janvier 1993 a garanti la libre circulation des biens au sein de l'Union européenne (UE), avec d'une part l'application d'une législation douanière uniformisée pour les échanges de biens entre les pays de l'UE et ceux en dehors de l'UE, et d'autre part la suppression des formalités douanières au sein de l'UE. Pour autant, le suivi des flux entre ces pays pour établir les statistiques du commerce extérieur a été maintenu par le Conseil européen.

⁶ Les exportations de biens figurent dans le « Tableau des ressources en produits » et les importations de biens dans le « Tableau des emplois finals ».

⁷ [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Principal_European_economic_indicators_\(PEEI\)/fr](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Principal_European_economic_indicators_(PEEI)/fr). Voir les références juridiques en fin d'article.

► ... encadré depuis longtemps par une réglementation européenne

Le règlement européen « Intrastat »⁸ de 1991 a instauré, en vue de l'établissement des statistiques des échanges de biens entre États membres, un système permanent de collecte statistique, appelé « système Intrastat ». Ce règlement précise les données à collecter auprès des entreprises par les États membres pour répondre aux besoins d'informations statistiques. Un système de seuils d'exemption et de simplification a été introduit, allégeant la charge de nombreuses entreprises.

S'agissant des statistiques du commerce extérieur avec les pays tiers (pays en dehors de l'Union européenne), le règlement « Extrastat »⁹ de 2009 en a défini les caractéristiques. S'appuyant sur les formulaires douaniers prévus pour les échanges de biens avec les pays tiers, ce règlement définissait les informations nécessaires au pilotage de la politique commerciale de l'Union européenne que chaque État membre devait collecter et transmettre à la Commission européenne.

Depuis 2022, les statistiques du commerce extérieur sont encadrées par le règlement « EBS » (European Business Statistics).

Depuis 2022, les statistiques du commerce extérieur sont encadrées par le règlement « EBS » (*European Business Statistics*)¹⁰ qui abroge les règlements « Intrastat » et « Extrastat ». Tout d'abord, le règlement « EBS » instaure l'échange de micro-données entre États membres sur leurs exportations intra-UE et sur certains flux extra-UE, afin d'améliorer la qualité des statistiques et alléger la charge statistique des entreprises. Par ailleurs, il ne couvre plus la collecte d'informations sur les importations intra-UE. Les États membres

peuvent utiliser les données miroir des autres États membres ou bien ils définissent un cadre réglementaire national pour la collecte de ces données : le DSECE a opté pour la seconde possibilité et mis en place en 2022 une enquête statistique. Enfin, le règlement « EBS » impose que les informations statistiques collectées au moyen d'enquêtes auprès des entreprises ou à partir de données administratives couvrent au moins 95 % de la valeur totale des exportations intra-UE de biens de chaque État membre vers l'ensemble des autres États membres (contre 97 % dans le règlement « Intrastat » révisé en 2004).

Eurostat publie les résultats des différents pays et agrège également les résultats au niveau européen pour diffuser les statistiques sur le commerce extérieur de l'ensemble de l'UE avec les pays tiers (hors UE).

⁸ Voir les références juridiques en fin d'article.

⁹ Voir les références juridiques en fin d'article.

¹⁰ Voir les références juridiques en fin d'article.

► Deux principales sources de données sont mobilisées —

La production de ces résultats nécessite d'exploiter des sources de données contenant des informations détaillées sur les flux de marchandises. Le DSECE s'appuie sur deux sources selon que les flux se font au sein de l'Union européenne ou avec des pays tiers :

- sur le champ des flux avec les pays tiers (extra-UE), une source administrative douanière est utilisée : le document administratif unique (DAU). Il s'agit de la déclaration de dédouanement que remplissent les entreprises qui effectuent des importations ou des exportations. Pour les besoins de l'administration, ces déclarations permettent la perception des droits de douane ;
- sur le champ des flux entre États membres : jusqu'à fin 2021, la direction générale des douanes et droits indirects (DGDDI) collectait une déclaration, la déclaration d'échanges de biens, qui contenait un volet statistique et un volet fiscal (état récapitulatif de TVA destiné à la direction générale des Finances publiques à des fins de contrôle de la TVA intracommunautaire). À partir de janvier 2022, une enquête statistique a été mise en place : l'enquête mensuelle sur les échanges de biens intra-UE (EMEBI)¹¹. Cette enquête a obtenu un avis d'opportunité, et le label d'intérêt général avec caractère obligatoire¹² par le Comité du label de la statistique publique (Christine et alii, 2020) ; elle a remplacé le volet statistique de la déclaration d'échanges de biens et porte à la fois sur les importations et les exportations intra-UE. Cela a été l'occasion de mieux séparer la collecte statistique de la déclaration fiscale. L'EMEBI contient une douzaine de questions : type de flux (importations ou exportations), valeur, masse, code produit, pays d'origine (à l'importation et à l'exportation), pays de provenance, pays de destination, identifiant TVA du destinataire (pour les exportations), département de départ et département d'arrivée de la marchandise, etc.

L'échantillon de l'enquête comporte près de 40 000 entreprises (unités légales). Les entreprises effectuant des flux de valeur inférieure à un seuil de 460 000 euros sont exclues de l'échantillon. Les flux inférieurs à ce seuil sont estimés à l'aide de sources fiscales (état récapitulatif de TVA et déclarations de chiffres d'affaires (CA3)).

Au total, avec l'enquête et les données administratives, environ 270 000 importateurs ou exportateurs sont suivis chaque année, et 35 millions de lignes (enregistrements) sont traitées par la chaîne statistique chaque mois (soit 420 millions par an).

► Des spécificités et des exigences particulières —

Une particularité des statistiques du commerce extérieur tient à leur nature intrinsèquement internationale : si la dimension internationale est en général un critère de comparaison dans les analyses, elle est le sujet même pour les statistiques du commerce extérieur. D'autres spécificités peuvent être soulignées sur le plan des concepts, de la diffusion ou de la méthodologie.

¹¹ <https://www.cnis.fr/enquetes/echanges-de-biens-intra-ue-emebi-enquete-mensuelle-sur-les-en-cours-de-labelisation/>.

¹² Sur le cheminement d'une enquête statistique publique, on peut consulter le site du Conseil national de l'information statistique : <https://www.cnis.fr/cheminement-dune-enquete-description/>.

Une forte proximité avec les concepts administratifs

La statistique publique s'appuie souvent sur les sources administratives¹³, mais elle construit ses propres concepts, généralement différents de ceux utilisés par les administrations, allant parfois jusqu'à réaliser des enquêtes spécifiques pour pouvoir



Ces statistiques comptabilisent généralement des flux de marchandises qui traversent la frontière française.



les mesurer (comme l'enquête Emploi pour le chômage au sens du Bureau international du travail). Au contraire, les statistiques du commerce extérieur produites et diffusées par le service statistique ministériel des douanes, en conformité avec le règlement EBS, sont fortement adhérentes aux concepts douaniers. Ces statistiques comptabilisent généralement des flux de marchandises qui traversent la frontière française. L'observation

de ces flux est plus proche de celle de la douane que de la comptabilité nationale, qui cherche par exemple à mesurer la valeur ajoutée lorsqu'un produit est exporté pour faire l'objet de travail à façon¹⁴ puis réimporté. De même, certains flux sont comptabilisés même s'ils ne s'accompagnent pas d'un transfert de propriété, contrairement à la comptabilité nationale.

Par ailleurs, il est nécessaire de retenir des conventions concernant la date à laquelle sont comptabilisés les flux. Dans le cas du commerce extérieur, les dates retenues sont essentiellement issues de critères administratifs : date de validation de la déclaration en douane pour le commerce extra-UE, date d'exigibilité de la TVA intra-communautaire pour le commerce intra-UE.

Un autre exemple de proximité avec les concepts douaniers est la notion de pays d'origine. Au-delà de leur comptabilisation, le DSECE décline les flux de commerce extérieur par produit, par pays de destination pour les exportations, et par pays d'origine pour les importations. Le pays d'origine peut différer du pays de provenance, qui est le dernier pays d'où est expédiée la marchandise importée en France. En effet, la marchandise peut avoir été fabriquée dans un autre pays que celui d'où il est expédié ou avoir fait l'objet de transformations dans plusieurs pays. Comment alors déterminer l'origine ? Les règles de détermination sont complexes, et le concept retenu pour les statistiques du commerce extérieur est le concept douanier. C'est selon celui-ci que les entreprises renseignent le pays d'origine dans leur déclaration ou dans l'enquête intra-UE. En cas de doute sur la valeur de cette variable renseignée par l'entreprise, le DSECE sollicite les spécialistes de l'origine au sein de la sous-direction du commerce international de la DGDDI.

Toutefois, il existe des cas particuliers dans le mode de comptabilisation des flux. Ainsi, sont exclus du champ les produits en transit : si une entreprise française effectue en France une déclaration en douane pour une exportation vers les États-Unis mais que les marchandises sont préalablement acheminées de la France vers un port aux Pays-Bas d'où aura lieu le départ vers les États-Unis, le flux de la France vers les Pays-Bas (où

¹³ Voir le dossier du Courrier des statistiques N° N1 sur « Le statisticien et les sources administratives » (2018).

¹⁴ La délivrance d'un travail à façon est définie comme la remise à son client par l'entrepreneur de l'ouvrage, d'un bien meuble qu'il a fabriqué ou assemblé au moyen de matières ou d'objets que le client lui a confiés à cette fin, que l'entrepreneur ait fourni ou non une partie des matériaux utilisés. Depuis le 1^{er} janvier 1996, le travail à façon est considéré comme une prestation de services (et non plus comme une livraison de biens).

les marchandises ne font que transiter) est hors champ, le flux comptabilisé est celui de la France vers les États-Unis. Autre exemple de flux non comptabilisé : lorsqu'un produit est expédié vers un autre pays pour réparation ou maintenance, le flux n'est pas comptabilisé dans les statistiques du commerce extérieur. De même, les œuvres d'art qui sont envoyées temporairement dans un autre pays pour y être exposées dans un musée ne sont pas comptabilisées dans les exportations. Enfin, certains produits spécifiques font l'objet de règles particulières de comptabilisation. C'est le cas notamment des satellites (**encadré 1**) mais aussi des métaux précieux, des timbres ou des pièces de monnaie ayant cours légal qui ne sont pas comptabilisés (l'or monétaire par exemple est considéré comme un actif financier et non comme un bien) alors qu'ils le sont s'ils n'ont plus cours légal (timbres et pièces de collection, par exemple).

► Encadré 1. Comptabiliser le commerce international de satellites

Comment sont comptabilisés les satellites dans le commerce extérieur ? Du fait de la nature de certains biens (satellites, avions, bateaux, etc.), il peut être problématique d'appliquer le critère du « mouvement physique ». Par exemple, pour les satellites, cela conduirait à gonfler les statistiques du pays de lancement. Ainsi, des dispositions particulières s'appliquent.

Un cas typique pourrait être le suivant : un satellite de télécommunication est fabriqué par une entreprise française. Il est vendu à une entreprise allemande en mars 2024 mais reste en France. Il est acheminé depuis la France vers Cap Canaveral, en Floride, en avril 2024 pour

être lancé par un lanceur américain le 19 juin 2024. Que comptabilise-t-on et quand ? Dans le cas particulier des satellites, le concept de commerce extérieur est celui du transfert de propriété : c'est donc une exportation de la France vers l'Allemagne. Par ailleurs, contrairement à la règle générale, la date de comptabilisation dans ce cas est celle du lancement du satellite et non celle de la validation de la déclaration douanière. L'exportation est donc comptabilisée dans les statistiques de juin 2024. Une veille est donc réalisée par le DSECE sur les dates de lancement, et la prise en compte des déclarations en douane correspondant à ces transactions est suspendue jusqu'à la date du lancement.

Des besoins de diffusion à un niveau très (très) détaillé

Le règlement EBS impose aux États membres de transmettre les résultats détaillés selon la nomenclature NC8 (nomenclature combinée à 8 positions) à Eurostat qui les diffuse à ce niveau. Le DSECE met également ces résultats détaillés à disposition des utilisateurs sur son site internet.

La NC8 est une déclinaison de la nomenclature harmonisée au niveau international, la SH6 (nomenclature du système harmonisé, sur 6 positions, développée par l'Organisation mondiale des douanes). Les deux chiffres supplémentaires par



Sur le plan opérationnel, les services douaniers ont besoin de connaître très précisément les produits échangés, pour suivre et contrôler les flux de marchandises.



rapport à la SH6 correspondent à une ventilation plus détaillée définie de façon harmonisée par l'Union européenne. Les codes produits en NC8 sont renseignés par les entreprises dans leurs déclarations de dédouanement pour les échanges extra-UE, et dans l'enquête EMEBI pour les échanges intra-UE.

La NC8 comporte environ 10 000 codes produits, ce qui représente une richesse d'information peu courante, et permet une analyse très fine des flux commerciaux, avec des règles plus

souples de secret statistique (**encadré 2**). Ce niveau de détail répond à des exigences de nature opérationnelle et stratégique.

Sur le plan opérationnel, les services douaniers ont besoin de connaître très précisément les produits échangés, pour suivre et contrôler les flux de marchandises, assurer la conformité aux normes de sécurité, sanitaires, environnementales, et autres régulations spécifiques s'appliquant à certains produits (quotas, embargos). Les codes détaillés permettent l'application des mesures tarifaires (par exemple des droits de douane à l'importation) et non tarifaires (par exemple des contrôles sanitaires) qui peuvent varier même entre des produits similaires.

Sur le plan stratégique, cette nomenclature très détaillée et harmonisée au niveau mondial permet un suivi des tendances émergentes et constitue une aide à l'évaluation de la compétitivité des secteurs économiques et à l'élaboration de politiques industrielles et commerciales (définition des tarifs douaniers et quotas, mise en place d'accords de libre-échange, etc.). Elle permet en outre aux fédérations professionnelles de réaliser des études de marché et à des économistes de mener des travaux de recherche.

À titre d'exemple, le degré de détail de la NC8 permet une mesure fine du commerce des animaux vivants : ainsi en 2023, la France a importé 122 autruches et émeus et elle est davantage importatrice qu'exportatrice d'abeilles vivantes !

Pendant, certains produits ne sont pas directement identifiables avec cette nomenclature. C'était le cas jusqu'en 2020 pour les masques de protection contre la Covid-19 qui étaient classés dans la même NC8 que d'autres produits de matières textiles à usage non sanitaire (Département des statistiques et des études du commerce extérieur, 2020). En 2021, de nouveaux codes ont été créés pour repérer plus directement les masques FFP2 et les autres masques de protection.

► **Encadré 2. Une vision atypique du secret : la règle du « secret passif »**

L'application du secret statistique dans les diffusions des statistiques du commerce extérieur font l'objet d'une règle spécifique, la règle dite du « secret passif » (Ader, 1978). Alors que la règle générale est de ne pas diffuser de résultats permettant d'identifier, directement ou indirectement, des entreprises (Redor, 2023), le règlement EBS (considérant 25 et article 19), et auparavant le règlement Intrastat, mentionnent une règle spécifique pour les statistiques du commerce extérieur, liée au niveau de détail de leur diffusion qui conduirait sans doute à blanchir trop de cases dans les tableaux avec les règles habituelles.

Ainsi, sont diffusés a priori l'ensemble des résultats statistiques détaillés y compris les cases comportant un faible nombre d'entreprises ou comprenant une entreprise dominante. Toutefois, si une entreprise demande la confidentialisation

de ses données, le DSECE vérifie si les critères usuels de confidentialité sont respectés pour cette entreprise dans les statistiques diffusées, à savoir si les cases concernées contiennent au moins trois entreprises et si une entreprise ne représente pas plus de 85 % de l'ensemble de la case en matière de flux considéré (importations ou exportations). Si les critères sont confirmés, la case est diffusée, sinon le DSECE ne diffuse plus la case.

Cette règle du secret passif ne concerne que les statistiques couvertes par le règlement européen. Les statistiques complémentaires publiées par la France, comme les statistiques régionales et départementales déterminées selon le lieu de départ et d'arrivée de la marchandise, ainsi que les statistiques sur les échanges entre les DROM et la métropole ou entre DROM, suivent les règles habituelles du secret.

Une absence de répertoire « officiel » des entreprises du commerce extérieur

Pour pouvoir construire l'échantillon des entreprises à interroger à l'enquête statistique sur les flux intra-UE, il est nécessaire de disposer d'une base de sondage contenant l'ensemble des entreprises qui effectuent du commerce extérieur intra-UE.

Contrairement à d'autres statistiques auprès des entreprises¹⁵, les statistiques du commerce extérieur ne peuvent s'appuyer directement sur un répertoire existant comme base de sondage ou comme outil de contrôle de l'exhaustivité. Pour pouvoir construire l'échantillon des entreprises à interroger à l'enquête statistique sur les flux intra-UE, il est nécessaire de disposer d'une base de sondage contenant l'ensemble des entreprises qui effectuent du commerce extérieur intra-UE, y compris celles réalisant très ponctuellement des transactions.

Ce référentiel d'entreprises doit donc être constitué par le DSECE. Le contrôle de l'exhaustivité du champ de l'enquête s'appuie à la fois sur l'observation des informations collectées précédemment (une entreprise ayant effectué du commerce extérieur en réalisera potentiellement de nouveau), mais aussi sur des sources externes (fiscales) : les états récapitulatifs de TVA (qui contiennent des informations agrégées sur les exportations intra-UE), les déclarations CA3 (informations agrégées sur les livraisons et les acquisitions intracommunautaires). Des contacts directs sont pris avec certaines entreprises pour valider leur inclusion ou leur exclusion du champ de l'enquête. Néanmoins, actuellement, ne sont pas incluses dans le champ les entreprises n'ayant jamais répondu à l'enquête (flux intra-UE) ni rempli de déclaration de dédouanement (extra-UE) et n'ayant pas non plus effectué de déclaration fiscale. Il peut s'agir d'entreprises voulant échapper aux taxes et droits de douane ou qui échangent des produits illicites (stupéfiants, contrefaçons).

La frontière, un miroir déformant

Enfin, les statistiques du commerce extérieur sont tout à fait originales du fait de l'existence de données « miroir ». En effet, lorsqu'une transaction a lieu entre deux entreprises de pays différents, l'entreprise exportatrice déclare cette exportation à l'enquête statistique (flux intra-UE) ou à la douane (flux extra-UE) de son pays, et l'entreprise importatrice déclare cette importation au service statistique ou à la douane de son pays. Chaque flux doit donc avoir son reflet parfait dans les données du pays partenaire.

Mais dans la pratique, ce miroir est « déformant ». Par exemple, les données des exportations françaises vers l'Allemagne publiées par la France font état de 80,3 milliards d'euros en 2022. De son côté, l'Allemagne publie un montant d'importations depuis la France de 69,8 milliards d'euros. Plusieurs facteurs expliquent cette déformation. En premier lieu, la plupart des asymétries entre la France et l'Allemagne s'expliquent par le concept de pays partenaire retenu dans les statistiques publiées nationalement. En effet, le concept de pays partenaire retenu pour les statistiques du commerce extérieur des pays européens, y compris les

¹⁵ Voir dans ce même numéro, l'article du Courrier des statistiques « La statistique annuelle d'entreprises: sa nature, son histoire, ses enjeux ».



Les exportations françaises vers l'Allemagne publiées par la France font état de 80,3 milliards d'euros en 2022. De son côté, l'Allemagne publie un montant d'importations depuis la France de 69,8 milliards d'euros.



statistiques allemandes et françaises, est le pays « d'origine » pour le pays partenaire à l'importation et le pays « de destination » pour le pays partenaire à l'exportation.

Cette différence de nature de pays partenaire, selon qu'il s'agisse d'un flux entrant ou d'un flux sortant, engendre mécaniquement des asymétries France-Allemagne : si une entreprise allemande exporte vers la France un produit originaire

de Chine, ce flux sera comptabilisé comme une exportation allemande vers la France du point de vue de l'Allemagne, et comme une importation de la France depuis la Chine du point de vue de la France.

Une bonne partie de cette asymétrie avec l'Allemagne disparaît si le pays d'importation retenu du point de vue de la France est l'Allemagne (pays « de provenance ») au lieu de la Chine (pays « d'origine »). C'est d'ailleurs le concept retenu par Eurostat pour les échanges intra-UE.

D'autres facteurs moindres peuvent aussi expliquer les écarts résiduels de mesure entre pays. Il peut s'agir de différences méthodologiques, de différences de valorisation (les importations étant évaluées au passage de la frontière du pays importateur (au prix CAF¹⁶ coût, assurance, fret) tandis que les exportations sont valorisées au passage de la frontière du pays exportateur (au prix FAB¹⁷ franco à bord), de décalage temporel dans l'enregistrement des flux commerciaux ou encore d'erreurs de déclaration (Eurostat, 2017 ; Javorsek et UN.Escap, 2016 ; Markhonko, 2014).

Pouvoir confronter les données miroir au niveau individuel permettrait d'affiner l'expertise des asymétries et de repérer d'éventuelles erreurs déclaratives. Le règlement EBS instaure et encadre l'échange entre États membres de micro-données relatives aux exportations intra-UE. Depuis janvier 2022, pour chaque flux d'exportation, figure l'identifiant TVA de l'entreprise destinataire de la marchandise. Il est alors possible d'apparier ces données avec celles des importateurs partenaires. Des mesures de sécurité informatique spécifiques (« bulle » sécurisée) ont été mises en place pour accueillir et expertiser les données des autres États membres.

► Vers une coproduction européenne des statistiques du commerce extérieur ?

Au-delà de l'expertise des données pour comprendre les asymétries, la demande d'Eurostat est d'aller encore plus loin dans l'utilisation de ces micro-données. Le but est l'intégration des données miroir dans le processus de production, avec une utilisation modulable de cette source miroir, afin d'améliorer la qualité des données et de réduire la charge statistique des entreprises. L'objectif recherché est de pouvoir :

- repérer des anomalies par confrontation de sources ;

¹⁶ <https://www.insee.fr/fr/metadonnees/definition/c1860>.

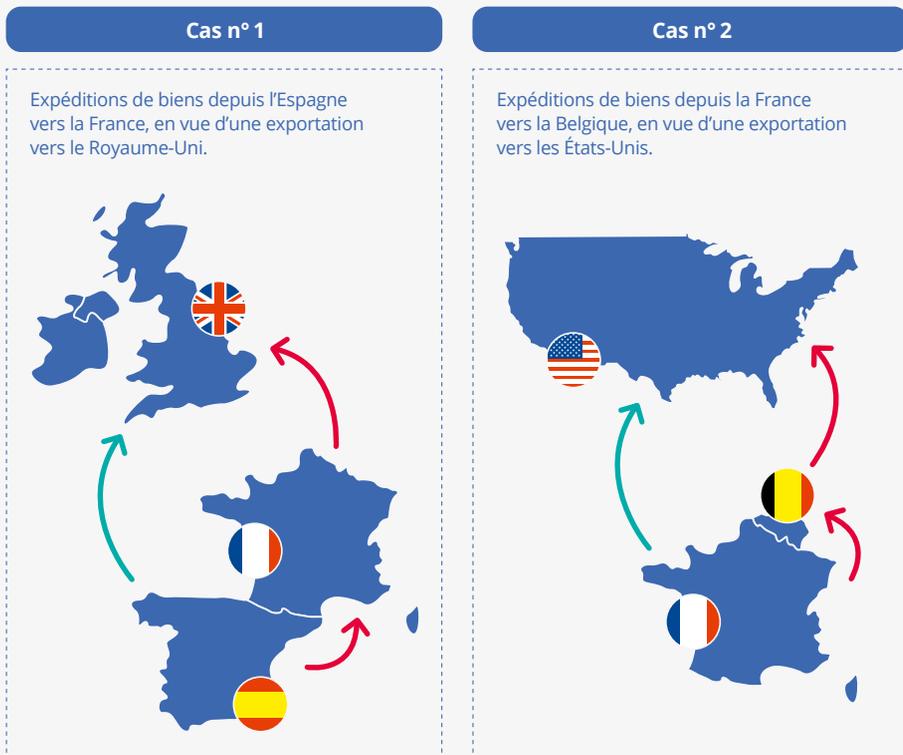
¹⁷ <https://www.insee.fr/fr/metadonnees/definition/c1347>.

- imputer des non-réponses ou redresser des données erronées ;
- imputer des données pour les flux inférieurs au seuil de collecte des importations (plusieurs États membres ont un seuil de collecte plus élevé pour les importations que pour les exportations) ;
- remplacer partiellement la collecte des importations par les données miroir, par exemple pour des entreprises pour lesquelles on n’observe pas d’asymétries ;
- remplacer intégralement la collecte des importations par les données miroir.

Ce changement radical ouvre la voie à une coproduction européenne des statistiques du commerce extérieur, pilotée activement par Eurostat.

► Figure 2 - Le quasi-export

Par exemple, il peut s’agir de l’expédition de biens depuis l’Espagne vers la France en vue d’une exportation vers le Royaume-Uni (cas n° 1) ou d’expéditions de biens depuis la France vers la Belgique en vue d’une exportation vers les États-Unis (cas n° 2).



Légende

- Flux enregistrés : il s’agit des flux collectés, non corrigés du quasi-export.
- Flux corrigés du quasi-export : il s’agit de donner les flux intermédiaires intra-Union européenne.

Depuis janvier 2024, le principe de coproduction est une réalité pour le traitement de certains flux particuliers de commerce extérieur. En complément des données miroir sur les flux intra-UE, sont échangées les données de certaines déclarations douanières, à des fins d'amélioration de la pertinence des statistiques produites. Cet échange, autorisé par le règlement EBS, vise notamment les données liées au quasi-export. Le quasi-export est l'expédition de biens depuis un État membre vers un autre État membre en vue d'une exportation vers un pays tiers (hors UE). Les échanges de micro-données douanières mis en place par Eurostat concernent les déclarations d'exportations vers les pays tiers, qui sont transmises par le pays intermédiaire au pays d'origine de l'exportation. L'objectif est de « gommer » les flux intermédiaires intra-UE pour ne faire apparaître dans les statistiques que les flux depuis les pays réels d'exportation vers les pays tiers (*figure 2*).

Adopter une ou plusieurs de ces utilisations nécessite des expertises préalables approfondies et un système d'information permettant de réaliser ces traitements conformément aux normes de sécurité requises. La diffusion des résultats est également affectée : chaque État membre fournit aux pays partenaires les cases à secrétiser compte tenu des demandes adressées par les entreprises (*encadré 2*) ; par ailleurs, des modalités spécifiques et harmonisées doivent être définies pour la diffusion de données aux chercheurs, qui peuvent sous certaines conditions accéder aux données non secrétisées.

Mais il requiert aussi une organisation de coproduction européenne : que se passe-t-il si un pays envoie ses données tardivement, ou si elles sont incomplètes ? Comment éviter les doubles comptes et les « trous » de collecte ?

► La digitalisation de l'économie et ses conséquences sur le commerce international



L'explosion de l'e-commerce a fortement modifié la structure du commerce international. De nombreux consommateurs achètent leurs biens via des plateformes qui importent les produits depuis d'autres pays.



Cette coproduction statistique au niveau européen est d'autant plus importante que le commerce international évolue du fait de la digitalisation croissante de l'économie. Pour accompagner ce mouvement, des dispositifs administratifs ont été élaborés au niveau européen, ce qui affecte la façon dont sont collectées les données du commerce extérieur.

L'explosion du e-commerce a fortement modifié la structure du commerce international. De nombreux consommateurs achètent leurs biens (en général de faible montant) via des plateformes qui importent les produits depuis d'autres pays.

Alors que les petits montants étaient exemptés de déclarations fiscales et douanières, la forte croissance du volume de biens de faible valeur échangés par les plateformes de e-commerce a conduit l'administration fiscale à supprimer l'exemption de TVA pour les petits montants : les importations sont devenues taxables dès le premier euro depuis le 1^{er} juillet 2021. Des déclarations douanières simplifiées (déclarations H7) ont



À plus longue échéance, la réforme en cours du code des douanes de l'Union européenne prévoit la mise en place d'une plateforme européenne de collecte de données douanières centralisées à la place des collectes nationales par les États membres.



été mises en place pour les petits montants (inférieurs à 150 euros) pour les importations de produits issus de pays tiers (extra-UE). Dès lors, comme le champ des statistiques du commerce extérieur est calé sur celui des déclarations douanières, ces flux doivent désormais être comptabilisés.

Par ailleurs, pour les importations et les exportations intra-UE effectuées dans le cadre du e-commerce, un guichet unique (*One stop shop*) a été mis en place afin de déclarer la TVA intracommunautaire pour les ventes à distance.

Les entreprises effectuant des échanges de

biens avec des particuliers en France qui optent pour ce guichet unique peuvent ne pas être assujetties en France : elles sont donc exclues du champ de l'enquête EMEBI ; il est alors nécessaire de récupérer les données de ce guichet unique pour ne pas perdre la connaissance de ces flux.

Ces nouvelles déclarations sont à intégrer dans le système d'information statistique de la Douane. Ces déclarations alternatives ne contiennent pas les mêmes informations que les sources de base (enquête EMEBI et DAU), et des traitements sont nécessaires pour estimer les données souhaitées. Les méthodes d'intégration de ces données aux statistiques de commerce extérieur font l'objet de discussions au niveau européen.

Enfin, la mise en place par la DGFiP¹⁸ de la facturation électronique à partir de 2026 pour les échanges intra-UE devrait conduire à la suppression des états récapitulatifs de TVA. Ne plus disposer de cette source est particulièrement gênant : un groupe de travail va se réunir sous l'égide d'Eurostat pour étudier les possibilités offertes par la facturation électronique en tant que source de données alternative.

À plus longue échéance, la réforme en cours du code des douanes de l'Union européenne prévoit la mise en place d'une plateforme européenne de collecte de données douanières centralisées (c'est-à-dire un hub de données), à la place des collectes nationales par les États membres. Ces transformations sont un défi pour le DSECE, qui doit s'adapter aux rapides évolutions du commerce international et à la simplification des démarches déclaratives pour les entreprises.

¹⁸ DGFiP : La direction générale des Finances publiques est une direction de l'administration publique centrale française qui dépend du ministère de l'Économie, des Finances et de l'Industrie.

Les statistiques du commerce extérieur, bien que datant de plus de trois siècles, sont toujours au cœur des politiques économiques. Ainsi au XVIII^e siècle, l'instauration des statistiques du commerce extérieur répondait au besoin de la politique mercantiliste de l'époque. Aujourd'hui, elles continuent de guider les politiques économiques, telles que la réindustrialisation de la France, dans le contexte des crises récentes (Covid, guerre en Ukraine) qui ont mis en évidence certaines vulnérabilités de l'économie et la nécessité d'une plus grande souveraineté sur des domaines stratégiques. Enfin, la digitalisation de l'économie et ses conséquences sur les procédures administratives des entreprises, notamment une plus grande intégration européenne, font évoluer les sources de données sur le commerce international. Ces changements constituent un défi : prise en compte de données nouvelles et multiples, coordination des productions statistiques au niveau européen. C'est également une opportunité : données plus riches, allègement potentiel de la charge statistique pour les entreprises, plus grande comparabilité des données au niveau international.

► Fondements juridiques

- **Règlement « Intrastat »** : Règlement (CE) n° 638/2004 du Parlement européen et du Conseil du 31 mars 2004 relatif aux statistiques communautaires des échanges de biens entre États membres et abrogeant le règlement (CEE) n° 3330/91 du Conseil. In : *site de l'Union européenne*. [en ligne]. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32004R0638>.
- **Règlement « Extrastat »** : Règlement (CE) n° 471/2009 du Parlement européen et du Conseil du 6 mai 2009 concernant les statistiques communautaires relatives au commerce extérieur avec les pays tiers et abrogeant le règlement (CE) n° 1172/95 du Conseil. In : *site de l'Union européenne*. [en ligne]. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A32009R0471>.
- **Règlement EBS (European Business Statistics)** : Règlement (UE) 2019/2152 du Parlement européen et du Conseil du 27 novembre 2019 relatif aux statistiques européennes d'entreprises, abrogeant dix actes juridiques dans le domaine des statistiques d'entreprises. In : *site de l'Union européenne*. [en ligne]. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32019R2152>.
- **Règlement d'exécution** (UE) 2020/1197 de la Commission du 30 juillet 2020 établissant des spécifications techniques et des modalités d'exécution en application du règlement (UE) 2019/2152 du Parlement européen et du Conseil relatif aux statistiques européennes d'entreprises, abrogeant dix actes juridiques dans le domaine des statistiques d'entreprises. In : *site de l'Union européenne*. [en ligne]. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32020R1197>.
- Loi no 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. In : *site de Légifrance*. [en ligne]. Mise à jour le 25 mars 2019. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573>.
- **Missions de la douane (dont la mission de production des statistiques du commerce extérieur)** : Décret n° 2021-328 du 26 mars 2021 modifiant le décret n° 2007-1664 du 26 novembre 2007 relatif à la direction générale des douanes et droits indirects. In : *site de Légifrance*. [en ligne]. Mise à jour le 25 mars 2019. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000043296997>.
- **Liste des principaux indicateurs économiques européens (PIEE)** : Communication de la Commission au Parlement européen et au Conseil concernant les statistiques de la zone euro « vers des méthodologies améliorées pour les statistiques et les indicateurs de la zone euro », 2002. In : *site de l'Union européenne*. [en ligne]. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52002DC0661>.

► Bibliographie

- ADER, Gérard, 1978. Secret et statistiques d'entreprises dans la communauté européenne. In : *Économie et statistiques*. [en ligne]. Mai 1978. Insee. N° 100, pp. 85-87. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06zm3s9/f1.pdf>.
- ASSOCIATION POUR L'HISTOIRE DE L'ADMINISTRATION DES DOUANES, 2024. In : *site de l'Association pour l'Histoire de l'Administration des Douanes*. [en ligne]. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://histoire-de-la-douane.org/>.
- BANQUE DE FRANCE, 2024. *Rapport de la balance des paiements et de la position extérieure de la France 2023*. [Consulté le 13 novembre 2024]. Disponible à l'adresse : <https://www.banque-france.fr/fr/publications-et-statistiques/publications/la-balance-des-paiements-et-la-position-exterieure-de-la-france-2023>.
- BEAUD, Michel, 1964. Le bureau de la balance du commerce (1781-1791). In : *Revue d'histoire économique et sociale*. Vol. 42, N° 3. pp. 357-377.
- BECUWE, Stéphane, BLANCHETON, Bertrand et ONFROY, Karine, 2019. « Base Montesquieu » les données du commerce extérieur français de 1836 à 1938. In : *Revue de l'OFCE*. N° 164, pp.87-109. [en ligne]. Avril 2019. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://doi.org/10.3917/reof.164.0087>.
- CHARLES, Loïc et DAUDIN, Guillaume, 2011. La collecte du chiffre au XVIII^e siècle : le Bureau de la balance du commerce et la production des données sur le commerce extérieur de la France. In : *Revue d'histoire moderne & contemporaine*. n° 58-1, pp. 128-155. [en ligne]. Janvier 2011. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://doi.org/10.3917/rhmc.581.0128>.
- 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 5 décembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/5008710/courstat-5.pdf>.
- CLINQUART, Jean, 1978. *L'Administration des douanes en France sous la Révolution*. Association pour l'Histoire de l'Administration des Douanes. ISBN 978-2-902691-01-2.
- DÉPARTEMENT DES STATISTIQUES ET DES ÉTUDES DU COMMERCE EXTÉRIEUR, 2020. Évolutions récentes des échanges extérieurs des produits liés à la lutte contre les crises sanitaires. In : *Le chiffre du commerce extérieur, un site de la direction générale des douanes et des droits indirects*. [en ligne]. Août 2020. Département des statistiques et des études du commerce extérieur. Études et éclairages N° 87. [Consulté le 23 septembre 2024]. Disponible à l'adresse : https://lekiosque.finances.gouv.fr/fichiers/etudes/tableaux/ee_87.pdf.
- DOUANES ROYALES DE FRANCE, 1825. *Tableau général du commerce de la France avec ses colonies et les puissances étrangères, pendant l'année 1825*. [en ligne]. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k32162212/f7.item>.

- ECONOMIC AND FINANCIAL COMMITTEE, 2008. *Status Report on Information Requirements in EMU*. [en ligne]. [Consulté le 23 septembre 2024]. 4 novembre 2008. Disponible à l'adresse : https://ec.europa.eu/eurostat/documents/4187653/6803935/EFC_REPORT_2008-EN.PDF/pdf/1672ac04-a032-49c9-abb7-99296a8c3358?t=1430227500000.
- EUROSTAT, 2017. *Transatlantic trade in services: Investigating bilateral asymmetries in EU-US trade statistics*. [en ligne]. 18 décembre 2017. Statistical reports. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/web/products-statistical-reports/-/ks-gq-17-016>.
- JAVORSEK, Marko et UN.ESCAP, 2016. Asymmetries in international merchandise trade statistics: A case study of selected countries in Asia and the Pacific. In : *ARTNeT Working Paper*. [en ligne]. Avril 2016. N°. 156. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://repository.unescap.org/handle/20.500.12870/1454>.
- JOURDAIN Roxane et VIGNE, Renaud, 2024. Le chiffre du commerce extérieur – Analyse annuelle 2023. In : *Le chiffre du commerce extérieur, un site de la direction générale des douanes et des droits indirects*. [en ligne]. Février 2024. Département des statistiques et des études du commerce extérieur. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://lekiosque.finances.gouv.fr/fichiers/Etudes/thematiques/A2023.pdf>.
- LE NINIVIN, David, 2024. Les opérateurs du commerce extérieur – Année 2023. In : *Le chiffre du commerce extérieur, un site de la direction générale des douanes et des droits indirects*. [en ligne]. Avril 2024. Département des statistiques et des études du commerce extérieur. [Consulté le 23 septembre 2024]. Disponible à l'adresse : https://lekiosque.finances.gouv.fr/fichiers/etudes/thematiques/A2023_operateurs.pdf.
- MARKHONKO, Vladimir, 2014. Asymmetries in official international trade statistics and analysis of globalization. In : *International Conference on Measurement of Trade and Economic Globalization*. [en ligne]. 1^{er} octobre 2014. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://unstats.un.org/unsd/trade/events/2014/mexico/default.asp>.
- REDOR, Patrick, 2023. Confidentialité des données statistiques : un enjeu majeur pour le service statistique public. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 46-63. [Consulté le 23 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635823?sommaire=7635842>.
- VIGNE, Renaud, 2023. Vulnérabilité énergétique de la France. In : *Le chiffre du commerce extérieur, un site de la direction générale des douanes et des droits indirects*. [en ligne]. Avril 2024. Département des statistiques et des études du commerce extérieur. Études et éclairages N°95. [Consulté le 23 septembre 2024]. Disponible à l'adresse : https://lekiosque.finances.gouv.fr/fichiers/etudes/tableaux/ee_95.pdf.

Les statistiques sur les causes de décès

Classer et coder... dans la classification internationale des maladies



Elise Coudin* et Aude Robert**

La statistique sur les causes de décès repose sur l'analyse des textes écrits par les médecins sur les certificats de décès, lesquels sont ensuite classés et codés d'après la classification internationale des maladies de l'Organisation mondiale de la santé. Cette statistique et la classification dans laquelle elle est codée se sont construites de concert depuis la fin du XIX^e siècle. La statistique sur les causes de décès est aujourd'hui produite par le Centre d'épidémiologie sur les causes médicales de décès (CépiDc) de l'Institut national de la santé et de la recherche médicale (Inserm). Depuis l'année de décès 2021, sa production articule trois modes de codage entre eux : codage automatique par système expert, codage interactif par équipe d'experts-nosologues et codage par prédictions de réseaux de neurones profonds entraînés sur l'historique des certificats de décès déjà codés. Sa mise en œuvre, qui s'appuie sur les critères de qualité en matière de statistique européenne, permet notamment de réduire les calendriers de production et d'avoir une démarche d'évaluation et une supervision statistiques systématiques.

 Causes-of-death (CoD) statistics are based on an analysis of the texts written by medical practitioners on death certificates, which are then classified and coded according to the World Health Organization's International Classification of Diseases. The statistics and the International statistical classification of diseases (ICD) classification have been developed together since the end of the 19th century. Today, causes-of-death (CoD) statistics are produced by the Centre for Epidemiology on Medical Causes of Deaths (CépiDc) of the National Institute of Health and Medical Research (INSERM). Since the year of death 2021, the CoD statistics production uses three different coding methods: automatic coding by an expert system, interactive coding by a team of expert nosologists, and predictive coding by deep neural networks trained on the history of previously coded death certificates. The implementation, which is based on European statistical quality criteria, makes it possible to shorten production delays and initiate systematic statistical evaluation and monitoring procedures.

* Directrice du Centre d'épidémiologie sur les causes médicales de décès, INSERM.
elise.coudin@inserm.fr

** Responsable de l'équipe automatisation, Centre d'épidémiologie sur les causes médicales de décès, INSERM.
aude.robert@inserm.fr

Pour hiérarchiser, suivre les problèmes de santé publique et analyser leurs disparités géographiques, temporelles ou sociales, les épidémiologistes, statisticiens et démographes utilisent les statistiques sur les causes de décès. Ces statistiques mobilisent peu de variables : tout au plus quelques informations démographiques et géographiques en complément du texte décrivant la ou les causes de décès. À cette apparente simplicité s'oppose une construction statistique plutôt complexe, résultat de l'histoire et de la spécificité de la source. La **codification** du texte rédigé par le médecin, en codes de nomenclature, c'est-à-dire l'attribution d'un « code » de la classification internationale des maladies (CIM) au texte médical décrivant la cause de décès pour obtenir des indicateurs statistiques pertinents, joue un rôle essentiel¹. Cette codification s'est complexifiée au cours du temps, avec la précision grandissante de la classification internationale des maladies, le besoin d'un traitement homogène et comparable entre pays. Elle repose aujourd'hui en France sur une articulation entre système expert à base de règles, équipe d'experts-nosologistes (la nosologie étant l'étude des classifications médicales) et depuis peu, inférence d'algorithmes d'intelligence artificielle. L'articulation entre ces trois modes a pour but d'optimiser la qualité du codage dans son ensemble, pour un usage statistique, en respectant les délais de production.

Cet article décrit l'histoire de cette statistique, sa construction et ses spécificités, en particulier la place de la codification. Puis, il présente les outils de codage mobilisés dans la production aujourd'hui, du traditionnel au plus innovant.

► De l'origine de la statistique sur les causes de décès au besoin de santé publique

Les prémices de la statistique sur les causes de décès sont très anciennes (Vallin et Meslé, 1988). Les grandes épidémies de peste auraient motivé des relevés et des comptages de décès par cause. Initialement « opérations de circonstance » (Bouvier-Colle et alii, 1990), ceux-ci sont mobilisés dans des études à l'origine des disciplines statistiques, épidémiologiques et démographiques. En France, les comptages par cause de décès s'établissent en parallèle du développement des registres paroissiaux, puis de l'état civil à partir de la fin du XVIII^e. Le XIX^e siècle marque le début d'une véritable construction statistique. Pour les besoins politiques de santé publique, de « salubrité publique » selon la terminologie de l'époque, le relevé des causes de décès se systématisait. En France, la statistique des causes de décès naît officiellement en 1886, à la suite d'une circulaire gouvernementale demandant aux maires des villes de plus de 10 000 habitants de relever une fois par quinzaine les décès dus à sept maladies infectieuses (fièvre typhoïde, variole, rougeole, scarlatine, coqueluche, diphtérie, diarrhée infantile) élargies à 27 rubriques dès l'année suivante.

¹ Pour une approche statistique historique voir le chapitre Classer et coder de *La politique des grands nombres* d'Alain Desrosières (Desrosières, 1993).

► Statistique et classification internationale se construisent de concert depuis la fin du XIX^e siècle —



Une bonne statistique est indispensable pour apprécier avec exactitude l'état sanitaire du pays, pour diriger avec efficacité la lutte contre les maladies, pour mesurer avec précision les résultats obtenus.



Les relevés par l'État sont généralisés à l'ensemble des causes de décès puis à l'ensemble du territoire à la fin du XIX^e et début du XX^e siècle. Cette collecte systématique requiert une codification dans une classification statistique commune et univoque. Alors qu'au cours du XIX^e siècle, une trop grande variété de listes et terminologies est utilisée (empêchant les comparaisons ou le suivi), les travaux en nosologie pour répondre aux besoins statistiques aboutissent, non sans détours², aux grands principes d'une classification commune. Le principe d'une classification internationale des maladies est adopté en 1893 par l'Institut

international statistique à partir de celle de synthèse proposée par Jacques Bertillon³, chef des travaux statistiques de la ville de Paris (en 14 chapitres, 161 rubriques détaillées). Sont construites ensuite puis adoptées par les pays les premières versions de la CIM. Cette dernière est gérée par l'Organisation mondiale de la santé (OMS) à partir de 1945 (**encadré 1**).

En France, on collecte l'ensemble des causes de décès sur l'ensemble du territoire, depuis 1906 (Aubenque et alii, 1978). Celles-ci alimentent les tableaux de Statistique sanitaire publiés par la Direction de l'assistance et de l'hygiène publique du ministère de l'Intérieur. Le simple comptage effectué par les mairies devient un recueil d'information organisé et systématisé quand, à partir de 1925, l'information sur la cause de décès est ajoutée au bulletin de décès de l'état civil. Ceci permet d'avoir des informations démographiques sur le défunt (sexe, groupe d'âge, département de décès puis de résidence). La statistique générale de France (SGF) est alors chargée du recueil et de l'élaboration de cette statistique. « Une bonne statistique est indispensable pour apprécier avec exactitude l'état sanitaire du pays, pour diriger avec efficacité la lutte contre les maladies, pour mesurer avec précision les résultats obtenus. » écrit Huber, directeur de la SGF. Ce verbatim résume ainsi les trois finalités de cette collecte, toujours d'actualité : veille sanitaire, politiques de santé publique et statistique.

► La naissance du certificat à caractère médical —

En 1937, est créé le certificat médical de décès, confidentiel, rempli par un médecin qui y déclare la cause en texte libre. Cette déclaration devient alors obligatoire. Le bulletin de décès anonymisé et le certificat médical cacheté sont transmis de la mairie à la Direction départementale de la santé à des fins de veille sanitaire. Le bulletin de décès sur lequel est reporté le texte de la cause est ensuite transmis à la SGF, laquelle codifie dans la classification des maladies et diffuse la statistique.

² Voir aussi l'historique très complet sur le site du CépiDc <https://www.cephidc.inserm.fr/causes-medicales-de-deces/cim-9/historique>.

³ Qui propose au même congrès une première classification sur les professions.

► Encadré 1. Classification internationale des maladies

La codification des causes de décès dans une nomenclature assurant comparabilité dans l'espace et le temps est un enjeu majeur et international dès le début du XX^e siècle. « Le but de la Classification statistique internationale des maladies et des problèmes de santé connexes est de permettre l'analyse systématique, l'interprétation et la comparaison des données de mortalité et de morbidité recueillies dans différents pays ou régions

et à des époques différentes (...). » (Organisation mondiale de la santé, 2008 ; Rey, 2016). La nomenclature doit assurer par son organisation une certaine stabilité temporelle et régionale de la codification des pathologies dans un contexte où les progrès de la médecine peuvent être rapides et les variations régionales fortes. Élaborée à partir de la classification Bertillon à la fin du XIX^e, adoptée internationalement, la CIM est réactualisée tous les

Évolution de la classification internationale des maladies : de la CIM 1 à la CIM 11*

CIM 1 - 1900

Première révision de la nomenclature internationale des causes de décès

179 rubriques, 14 chapitres

- I Maladies générales
- II Maladies du système nerveux et des organes des sens
- III Maladies de l'appareil circulatoire
- IV Maladies de l'appareil respiratoire
- V Maladies de l'appareil digestif
- VI Maladies de l'appareil génito-urinaire et de ses annexes
- VII Maladies puerpérales
- VIII Maladies de la peau et de ses annexes
- IX Maladies des organes de locomotion
- X Vices de conformation
- XI Maladies du 1^{er} âge
- XII Maladies de la vieillesse
- XIII Affections produites par des causes extérieures
- XIV Maladies mal définies

Morbidité



Organisation mondiale de la santé
Règles cause initiale

CIM 6 - 1948

Modèle standard de certificat, Règles de détermination de la cause initiale

17 chapitres, 765 rubriques, près de 2 000 codes

CIM 10 - 1992

21 chapitres, 12 000 rubriques, ~ 16 000 codes

- I A00-B99 Certaines maladies infectieuses et parasitaires
- II C00-D48 Tumeurs
- III D50-D89 Maladies du sang et des organes hématopoiétiques et certains troubles du système immunitaire
- IV E00-E90 Maladies endocrinienne, nutritionnelles et métaboliques
- V F00-F99 Troubles mentaux et du comportement
- VI G00-G99 Maladies du système nerveux
- VII H00-H59 Maladies de l'œil et de ses annexes
- VIII H60-H95 Maladies de l'oreille et de l'apophyse mastoïde
- IX I00-I99 Maladies de l'appareil circulatoire
- X J00-J99 Maladies de l'appareil respiratoire
- XI K00-K93 Maladies de l'appareil digestif
- XII L00-L99 Maladies de la peau et du tissu cellulaire sous-cutané
- XIII M00-M99 Maladies du système ostéo-articulaire, des muscles et du tissu conjonctif
- XIV N00-N99 Maladies de l'appareil génito-urinaire
- XV O00-O99 Grossesse, accouchement et puerpéralité
- XVI P00-P96 Certaines affections dont l'origine se situe dans la période périnatale
- XVII Q00-Q99 Malformations congénitales et anomalies chromosomiques
- XVIII R00-R99 Symptômes, signes et résultats anormaux d'examen clinique et de laboratoire, non classés ailleurs
- XIX S00-T98 Lésions traumatiques, empoisonnements et certaines autres conséquences de causes externes
- XX V01-Y98 Causes externes de morbidité et de mortalité
- XXI Z00-Z99 Facteurs influant sur l'état de santé et motifs de recours aux services de santé
- XXII U00-U99 Codes d'utilisation particulière

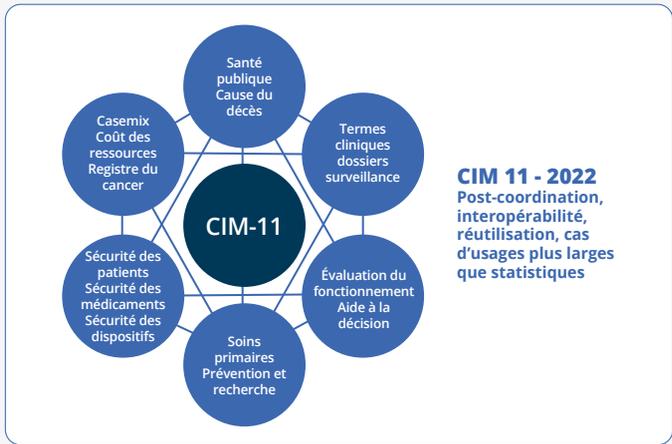
* <https://www.who.int/standards/classifications/classification-of-diseases>.

10 ans environ. Ces évolutions sont gérées par l'OMS depuis 1945. Les versions de la CIM sont de plus en plus riches et complexes. À l'origine, ce principe de classement reposait sur une distinction entre les maladies générales et celles localisées à un organe donné, plus facile à collecter que l'étiologie de la pathologie. La dimension étiologique s'est étendue au fil du temps et des progrès de la médecine, puis avec son usage pour élaborer les statistiques de

morbidité hospitalière. La première révision de la CIM (en 1900) comportait 179 rubriques, la CIM 9 (utilisée pour coder les causes de décès de 1979 à 1999) 6 000 rubriques, et la CIM 10 (utilisée depuis l'an 2000) compte 12 000 rubriques et 16 000 codes. La transition à la CIM 11, adoptée en 2022 par l'OMS et dont l'objectif d'interopérabilité et de réutilisation dépasse celui de la statistique de santé, constituera un défi pour les prochaines années.

CIM 10 - 2019

I	Certaines maladies infectieuses et parasitaires
II	Tumeurs
C00-C97	Tumeurs malignes
C00-C75	Tumeurs malignes, primitives ou présumées primitives, de siège précisé, à l'exception des tissus lymphoïde, hématopoïétique et apparentés
C00-C14	Tumeurs malignes de la lèvre, de la cavité buccale et du pharynx
...	...
C50-C50	Tumeur maligne du sein
C50	Tumeur maligne du sein
C50.0	Mamelon et aréole
C50.1	Partie centrale du sein
C50.2	Quadrant supéro-interne du sein
C50.3	Quadrant inféro-interne du sein
C50.4	Quadrant supéro-externe du sein
C50.5	Quadrant inféro-externe du sein
C50.6	Prolongement axillaire du sein
C50.8	Lésion à localisations contiguës du sein
C50.9	Sein, sans précision
...	...
C73-C75	Tumeurs malignes de la thyroïde et d'autres glandes endocrines
III	Maladies du sang et des organes hématopoïétiques et certains troubles du système immunitaire
...	...
XXII	Codes d'utilisation particulière



Ce circuit de recueil d'information statistique assure la confidentialité de l'information médicale transmise, tout en circulant par les mains des acteurs de la veille sanitaire au plus tôt. Le revers de la médaille est qu'il s'avère lourd, difficile à coordonner et peut conduire à des pertes d'informations. Ces contraintes existent toujours aujourd'hui du fait d'un circuit de transmission des certificats papiers quasi inchangé.

L'Insee remplace la SGF en 1945 et décentralise la codification des causes en direction régionale, introduisant une hétérogénéité dans les pratiques. Par ailleurs, l'information collectée devient plus précise et se standardise suivant les besoins de la classification. Dès 1955, le certificat médical permet au médecin de déclarer plusieurs causes de décès. Et dès 1958, le modèle de certificat en France suit le standard international élaboré dans la CIM et recommandé par l'OMS.

► Une production confiée à l'Inserm

En 1968, à l'occasion d'une nouvelle révision de la CIM, la mission de codification des causes de décès est confiée à l'Inserm afin qu'elle soit réalisée par un personnel médical qualifié et de manière centralisée. Jusqu'en 1987, les professionnels médicaux de la Section information en santé publique de l'Inserm, puis du Service commun d'information sur les causes médicales de décès (SC8) codifient des causes et reportent les codes sur les bulletins de décès compilés ensuite par l'Insee. À partir de 1988, avec le développement de l'informatisation, l'Inserm obtient la maîtrise complète du processus de production jusqu'à la diffusion : l'Insee met à disposition un extrait informatique de son fichier démographique de mortalité. Les experts-nosologues y adossent les codes des causes et l'Inserm peut ainsi publier des statistiques provisoires, puis définitives. Depuis 1997, le Centre d'épidémiologie sur les causes médicales de décès de l'Inserm (CépiDc, remplaçant le SC8 chargé du codage et de l'exploitation et le SC25 chargé de la saisie) assure la production et la diffusion de la statistique.

► Une statistique officielle européenne suivant les standards de l'OMS



La construction de la statistique sur les causes de décès est internationale dès sa création.



La construction de la statistique sur les causes de décès est internationale dès sa création. À partir de 2007-2008, avec la mise en place du système statistique européen, elle devient une statistique officielle européenne soumise au règlement 223⁴. Elle entre dans le champ d'application du règlement relatif aux statistiques communautaires de la santé

publique ainsi que de la santé et de la sécurité au travail, lequel découle du développement de programmes d'action communautaire dans le domaine de la santé publique. Ce règlement et son règlement d'application⁵ assurent la comparabilité de la statistique au

⁴ Voir les références juridiques en fin d'article.

⁵ CE 1338/2008 et UE 328/2011.

sein de l'Europe et définissent les conditions de collecte de l'information. Celle-ci s'appuie sur des certificats de décès nationaux conformes aux recommandations de l'OMS et sur un codage des causes dans la classification internationale des maladies de l'OMS. Les normes d'évaluation de la qualité sont celles du code des bonnes pratiques en matière de statistique européenne⁶ (pertinence, exactitude, actualité, ponctualité, accessibilité, comparabilité, cohérence) ; les concepts, les champs, les variables, leur ventilation, les périodes de référence et les délais de transmission sont fixés. En particulier, les États membres doivent transmettre à Eurostat les données d'une année dans un délai de 24 mois à compter de la fin de cette même année. Le CépiDc de l'Inserm devient en 2017 une « autorité statistique nationale » (ONA : *Other National Authority*) productrice de statistique officielle, hors service statistique public.

► Les grandes causes de décès en France et en Europe en 2021

Concrètement, l'indicateur le plus courant est la répartition des décès selon leurs causes initiales. La cause initiale du décès est « la maladie ou le traumatisme qui a déclenché l'évolution morbide conduisant directement au décès ou les circonstances dans le cas d'un traumatisme » (Organisation mondiale de la santé, 2008). Ainsi, en 2021, les cancers sont la première cause de décès en France avec plus d'un quart des décès ; en Europe ils représentent 22 % des décès. En moyenne en Europe, ce sont les maladies de l'appareil circulatoire qui se situent à la première place avec près d'un tiers des décès, contre 21 % en France (*figure 1*). Ceci n'est qu'un aperçu des tables de données sur les causes de décès mises à jour, et diffusées⁷ chaque année après collecte auprès des États membres par Eurostat⁸ (Eurostat, 2024), et en France, sur le site du CépiDc (Fouillet et alii, 2023, 2024 ; Cadillac et alii 2023, 2024). La base de données individuelles avec l'ensemble des causes, des textes et autres variables alimente chaque année le Système national des données de santé.

Cependant, à l'apparente simplicité de ces statistiques, ventilées par sexe, âge, localisation du décès ou lieu de résidence du défunt, s'oppose une complexité en pratique de la source et de son traitement. Et ceci, dès la collecte de l'information.

► La certification des décès en pratique aujourd'hui

Aujourd'hui, chaque décès donne lieu à la rédaction d'un certificat de décès par un médecin. La validation de ce certificat est nécessaire pour fermer le cercueil et procéder à l'inhumation. La forme de ce certificat et les finalités⁹ d'usage des informations qu'il

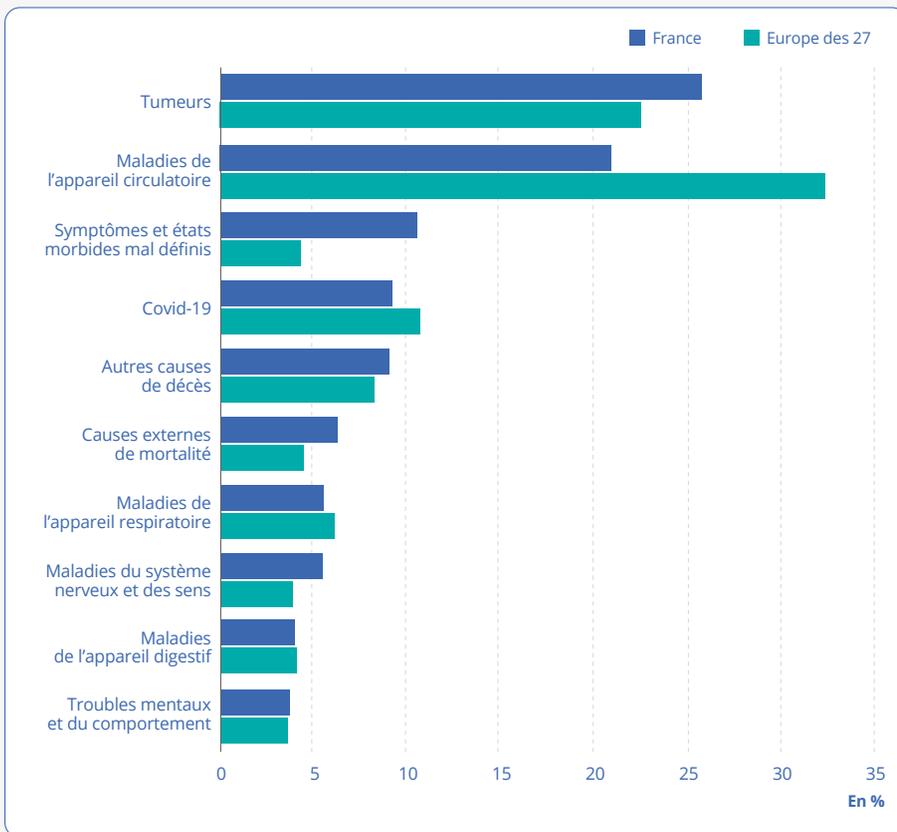
⁶ <https://www.insee.fr/fr/information/4137879>.

⁷ Voir les tables « Causes de décès » (hlth_cdeath) sur la base de données d'Eurostat : <https://ec.europa.eu/eurostat/web/health/database>.

⁸ Eurostat, l'Office statistique de l'Union européenne, est chargé de publier des statistiques et des indicateurs européens, permettant d'effectuer des comparaisons entre les pays et les régions.

⁹ Ces informations ne peuvent être utilisées que pour des motifs de santé publique, à des fins de veille et d'alerte, par l'État, les agences régionales de santé et l'Agence nationale de santé publique ; pour l'établissement de la statistique nationale des causes de décès et pour la recherche en santé publique par l'Inserm ; pour l'alimentation du Système national des données de santé et les traitements de données afférents concernant la santé ; et pour l'établissement de statistiques par l'Insee et le service statistique public dans le cadre de la loi de 1951. Voir les références juridiques en fin d'article.

► **Figure 1 - Les 10 causes de décès les plus fréquentes en 2021 en France et en Europe**



Lecture : 32 % des décès en Europe sont causés par les maladies de l'appareil circulatoire et 21 % en France.
 Champ : Tous les décès survenus en France et en Europe.
 Source : Eurodatabase, causes of death, hlth_cdeath, <https://ec.europa.eu/eurostat/web/health/database>.

contient sont définies dans le Code général des collectivités territoriales¹⁰. Un volet administratif avec les noms et prénoms est destiné aux opérateurs funéraires et à la mairie qui dressera l'acte de décès, et un volet médical confidentiel sans les noms et prénoms du défunt est transmis à l'Inserm. Depuis 2022, le certificat de décès doit obligatoirement être rempli sur support électronique dans les établissements de santé, sauf exceptions¹¹. Début 2024, un peu plus de 40 % des certificats étaient électroniques.

Le volet médical arrive au CépiDc au format électronique ou numérisé via un prestataire de saisie. Il suit le modèle international de l'OMS (*figure 2*). Dans une première partie (Partie 1), le médecin écrit sur quatre lignes l'enchaînement causal des maladies qui ont

¹⁰ Voir les références juridiques en fin d'article.

¹¹ Voir les références juridiques en fin d'article.

conduit au décès (aussi appelé « processus morbide »), de la cause immédiate à inscrire en première ligne, à la cause initiale. Le médecin est invité à remplir une cause en texte libre par ligne, chaque cause indiquée sur une ligne étant « due à » celle indiquée sur la ligne suivante (relation causale attendue). Dans la Partie 2, le médecin indique les autres états morbides, facteurs ou états physiologiques qui ont pu contribuer au décès sans être directement impliqués dans l'enchaînement conduisant au décès. Il consigne aussi sur le volet médical des informations notamment sur le lieu de décès, l'état de grossesse, les circonstances apparentes du décès (mort naturelle, mort subite, accident, suicide, etc.).

► **Figure 2 - Volet médical du certificat de décès***

VOLET MÉDICAL. À remplir et à clore par le médecin ayant constaté le décès – Renseignements confidentiels et anonymes			
INFORMATIONS RELATIVES AU DÉFUNT			
Commune de décès :	Code postal :	Date de décès : <input type="checkbox"/> date réelle OU <input type="checkbox"/> constatée	Sexe : <input type="checkbox"/> masculin <input type="checkbox"/> féminin
Commune de domicile :	Code postal :	Date de naissance :	
CAUSES DU DÉCÈS			
PARTIE I	Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès. <i>Il s'agit de la maladie, du traumatisme, de l'intoxication, de la complication ayant entraîné la mort (et non du mécanisme de décès comme une syncope, un arrêt cardiaque...).</i>		Intervalle entre le début du processus morbide et le décès <i>En heures, jours, mois ou ans</i>
	a) _____		_____
	due à ou consécutive à : b) _____		_____
	due à ou consécutive à : c) _____		_____
	due à ou consécutive à : d) _____		_____
	<small>La dernière ligne remplie doit correspondre à la cause initiale</small>		
PARTIE II	Autres états morbides, facteurs ou états physiologiques (grossesse...) ayant contribué au décès, mais non mentionnés en Partie I		

INFORMATIONS COMPLÉMENTAIRES (cocher la case appropriée pour chaque point)			
LIEU DU DÉCÈS	<input type="checkbox"/> Établissement de santé public	GROSSESSE La femme décédée était-elle enceinte ?	
<input type="checkbox"/> Domicile (du défunt ou autre)	<input type="checkbox"/> Établissement de santé privé	<input type="checkbox"/> non, pas au cours de l'année précédant le décès	<input type="checkbox"/> pas au moment du décès, mais grossesse terminée depuis plus de 42 jours et moins d'1 an
<input type="checkbox"/> EHPAD, maison de retraite	<input type="checkbox"/> Établissement pénitentiaire	<input type="checkbox"/> oui, au moment du décès	<input type="checkbox"/> ne sait pas
<input type="checkbox"/> Voie publique	<input type="checkbox"/> Autre lieu ou indéterminé	La grossesse a-t-elle contribué au décès ? <input type="checkbox"/> oui <input type="checkbox"/> non <input type="checkbox"/> ne sait pas	
MORT SUBITE S'agit-il d'un décès brutal et inattendu, évocateur de mort subite* ?	<input type="checkbox"/> oui <input type="checkbox"/> non <input type="checkbox"/> ne sait pas	ACTIVITÉ PROFESSIONNELLE Le décès est-il survenu lors d'une activité professionnelle* ?	
<small>* décès non traumatique (adulte, enfant, nourrisson) avec mode de survenue brutal (en moins d'une heure ou probablement) et inattendu (en l'absence de maladies chroniques ou stade terminal)</small>		<input type="checkbox"/> oui <input type="checkbox"/> non <input type="checkbox"/> ne sait pas	
CIRCONSTANCES APPARENTES DU DÉCÈS		<small>* toute activité source de revenu (y compris au domicile), les trajets domicile-travail, les déplacements professionnels, etc.</small>	
<input type="checkbox"/> Mort naturelle	<input type="checkbox"/> Faits de guerre	RECHERCHE DE LA CAUSE DU DÉCÈS	
<input type="checkbox"/> Accident	<input type="checkbox"/> Complications de soins médicaux, chirurgicaux	Une recherche de la cause du décès a-t-elle été demandée ?	
<input type="checkbox"/> Suicide	<input type="checkbox"/> Investigations en cours	<input type="checkbox"/> oui, recherche médicale <input type="checkbox"/> oui, recherche médico-légale <input type="checkbox"/> non	
<input type="checkbox"/> Atteinte à la vie d'autrui	<input type="checkbox"/> Indéterminées	<small>Si oui, un volet médical complémentaire sera établi ultérieurement par le médecin ayant réalisé le diagnostic des causes de décès</small>	
EN CAS DE MORT VIOLENTE (accidentelle, délictuelle, suicidaire, criminelle)		SIGNATURE Nom lisible et cachet obligatoire du médecin	
Précisez le lieu de survenue de l'événement déclencheur :		_____	
<input type="checkbox"/> Domicile	<input type="checkbox"/> Lieu de sport	_____	
<input type="checkbox"/> Commerce	<input type="checkbox"/> Local industriel, chantier	_____	
<input type="checkbox"/> Établissement accueillant du public	<input type="checkbox"/> Exploitation agricole	_____	
	<input type="checkbox"/> Autre lieu ou indéterminé	_____	

* Voir les références juridiques en fin d'article.

► Un texte médical riche, polysémique, parfois sans enchaînement causal

Le constat d'un décès relevant d'un diagnostic médical, le médecin qui en prend la responsabilité est libre dans sa rédaction, sans pré-remplissage. C'est d'ailleurs une recommandation de l'OMS. En conséquence, le texte¹² du processus morbide est tout aussi libre et contient de nombreux libellés possibles :

- toutes les maladies et pathologies, avec leurs étiologie, intensité, localisation et autres caractéristiques associées (telles que « cancer », « tabac », « génétique », « sévère », « stade 4 », « poumon », « gauche », « maladies osseuses ») ;
- des acronymes (comme « AVC », « ACR », « IRM », « BPCO »¹³), parfois polysémiques (« IRC » ou « IRA » peuvent signifier selon le contexte « insuffisance rénale » « chronique » ou « aiguë » ou « insuffisance respiratoire » « chronique » ou « aiguë », ou encore « TA » « tentative d'autolyse » ou « tension artérielle ») ;
- des abréviations (« K » pour cancer), parfois en grec (ψ , γ pour grossesse, etc.), des signes (ensemble vide pour « sans ») ;
- différentes appellations ayant le même sens (par exemple « cancer », « tumeur », « néo », « néoplasme », « carcinome », etc.).

Il peut aussi y avoir des incohérences dans l'enchaînement causal, voire entre la Partie I et la Partie II du certificat. On peut retrouver plusieurs entités sur la même ligne du formulaire avec des liens de causalité exprimés (par exemple par le terme « sur » : « métastases sur cancer pulmonaire ») ou pas (« infection sur cathéter »).



La spécificité des causes de décès provient bien de cette information textuelle riche en termes et faiblement structurée qu'il faut transformer en un langage statistique commun.



La spécificité des causes de décès provient bien de cette information textuelle riche en termes et faiblement structurée qu'il faut transformer en un langage statistique commun. Ainsi l'opération de codification des causes de décès se distingue pour plusieurs raisons de celles sur d'autres sources de la statistique publique, pour lesquelles il s'agit de classifier de simples libellés. Tout d'abord, la classification comprend plusieurs milliers de rubriques, sous forme d'une arborescence touffue. Par exemple, si l'étiologie d'une pathologie permet de la reclasser dans un chapitre définissant plus précisément son origine, on peut être amené à

changer de chapitre : ainsi une démence sera classée généralement en F « troubles mentaux » mais elle peut l'être dans d'autres chapitres si elle dérive d'autres pathologies (infectieuses, du système nerveux, etc.). En complément de la codification des causes, il faut vérifier la cohérence du processus causal comme préalable à la détermination de la cause initiale, ce qui est une opération complexe.

¹² Pour les certificats papier, le prestataire numérise le processus causal (parties 1 et 2) par saisie vocale.

¹³ « AVC » : accident vasculaire cérébral, « ACR » : arrêt cardio-respiratoire, « IRM » : imagerie par résonance magnétique, « BPCO » : bronchopneumopathie chronique obstructive.

En 2021, un certificat non vide comprend en moyenne 3,5 codes de causes, 15 % des certificats en ont 6 ou plus (et 1 %, 11 codes ou plus). De plus, 4 145 codes différents de la CIM ont été utilisés au moins une fois, dont 2 675 codes en cause initiale.

► Coder les causes de décès, un exercice complexe

Coder les causes de décès dans la CIM a deux finalités distinctes : déterminer dans le texte écrit par le médecin les entités nosologiques¹⁴ pour leur affecter un code de la nomenclature CIM et déterminer la cause initiale du décès.

Pour satisfaire ce double objectif, la classification internationale des maladies décrit en complément de la nomenclature (c'est-à-dire les codes) un ensemble de règles de codage, rassemblées dans un volume 2, le « guide de référence » (OMS, 2008). Ce manuel d'instruction au codage outille le codeur : une douzaine de règles appliquées sur la séquence causale selon un algorithme précis permet de déterminer la cause initiale du

décès, d'une façon systématique, tout en corrigeant des erreurs ou des incohérences possibles dans la chaîne causale déclarée ou encore en privilégiant certaines pathologies à suivre car d'intérêt de santé publique¹⁵ (figure 3).

Ce besoin d'homogénéité pour assurer la comparabilité de la statistique dans le temps et l'espace a motivé l'automatisation du codage.

Ce besoin d'homogénéité pour assurer la comparabilité de la statistique dans le temps et l'espace a motivé l'automatisation du codage. Le caractère systématique du raisonnement et la présence de règles précises décrites dans la CIM font de l'exercice un candidat idéal. Les outils de

cette automatisation ont évolué dans le temps, de l'interface interactive permettant un codage assisté par système de règles au batch automatique de ce même système, jusqu'aux réseaux de neurones profonds entraînés sur les multiples données déjà codées.

► Les systèmes experts nés du partage international des règles de décision

Les systèmes experts de codage des causes de décès sont utilisés en France depuis 2000, à l'occasion de la mise en œuvre de la CIM 10. Styx, développé en France, puis Iris à partir de 2012, sont des logiciels d'aide au codage avec une interface homme-machine, utilisant cette démarche (Pavillon et Laurent, 2003 ; Iris Institute, 2024). Ils s'appuient sur le modèle international du certificat de décès et intègrent les règles du volume 2 de la CIM 10 sous

¹⁴ Une entité nosologique est un terme classifiable dans la CIM.

¹⁵ Les experts du CépiDc sont membres du *Mortality reference group* (groupe de référence sur la mortalité), le groupe de l'Organisation mondiale de la santé (OMS) chargé du maintien et des évolutions des règles de codage de la CIM en matière de mortalité. Le centre collaborateur français de l'OMS sur la famille des classifications internationales présidé par l'Agence du numérique en santé regroupe l'Inserm-CépiDc (pour la mortalité), la Caisse nationale de l'assurance maladie (qui intervient sur la nomenclature de classement des remboursements, appelée l'*International Classification of Health Interventions* (ICHI)), l'Agence technique de l'information sur l'hospitalisation (pour la Classification internationale des maladies et ICHI en morbidité), l'École des hautes études en santé publique (pour la Classification internationale du fonctionnement, du handicap et de la santé).

► Figure 3 - Causes de décès indiquées par le médecin

Une des règles de codage indique que la séquence entre la cause immédiate et la cause initiale doit être entièrement causale. Ainsi l'état grippal peut entraîner une détresse respiratoire, mais une insuffisance cardiaque ne peut pas conduire à un état grippal : ce sera donc l'état grippal qui sera retenu comme cause initiale, même si le médecin a indiqué l'insuffisance cardiaque sur la dernière ligne de la Partie 1 (figure 3a).

Causes du décès		Intervalle entre le début du processus morbide et le décès (heures, jours, mois ou ans)
PARTIE I	Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès * <i>La dernière ligne remplie doit correspondre à la cause initiale.</i>	
a)	détresse respiratoire	
due à ou consécutive à : b)	état grippal	
due à ou consécutive à : c)	insuffisance cardiaque	
due à ou consécutive à : d)		
<i>* Il s'agit de la maladie, du traumatisme, de la complication ayant entraîné la mort (et non du mode de décès, ex. : syncope, arrêt cardiaque...)</i>		
PARTIE II	Autres états morbides, facteurs ou états physiologiques (grossesse...) ayant contribué au décès, mais non mentionnés en Partie I	
	diabète	

Le même raisonnement permet de retenir comme cause initiale le cancer pulmonaire, même si le médecin a indiqué tabac dans la ligne suivante (figure 3b).

Causes du décès		Intervalle entre le début du processus morbide et le décès (heures, jours, mois ou ans)
PARTIE I	Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès * <i>La dernière ligne remplie doit correspondre à la cause initiale.</i>	
a)	néphrose	
due à ou consécutive à : b)	cancer	
due à ou consécutive à : c)	tabac	
due à ou consécutive à : d)		
<i>* Il s'agit de la maladie, du traumatisme, de la complication ayant entraîné la mort (et non du mode de décès, ex. : syncope, arrêt cardiaque...)</i>		
PARTIE II	Autres états morbides, facteurs ou états physiologiques (grossesse...) ayant contribué au décès, mais non mentionnés en Partie I	

D'autres règles permettent d'aller rechercher la cause initiale en partie 2 si les causes en Partie 1 ne sont pas assez informatives. Ainsi, on ira chercher Alzheimer en cause initiale même s'il n'est indiqué qu'en partie 2 (figure 3c).

Causes du décès		Intervalle entre le début du processus morbide et le décès (heures, jours, mois ou ans)
PARTIE I	Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès * <i>La dernière ligne remplie doit correspondre à la cause initiale.</i>	
a)	pneumonie	
due à ou consécutive à : b)	hémorragie	
due à ou consécutive à : c)		
due à ou consécutive à : d)		
<i>* Il s'agit de la maladie, du traumatisme, de la complication ayant entraîné la mort (et non du mode de décès, ex. : syncope, arrêt cardiaque...)</i>		
PARTIE II	Autres états morbides, facteurs ou états physiologiques (grossesse...) ayant contribué au décès, mais non mentionnés en Partie I	
	Alzheimer	

Et on ira chercher « cancer » en cause initiale, dans l'exemple ci-dessous même si elle n'est pas précisée explicitement, car celle-ci est « évidente » (figure 3d).

Causes du décès		Intervalle entre le début du processus morbide et le décès (heures, jours, mois ou ans)
PARTIE I	Maladie(s) ou affection(s) morbide(s) ayant directement provoqué le décès * <i>La dernière ligne remplie doit correspondre à la cause initiale.</i>	
a)	maladies	
due à ou consécutive à : b)	tabac	
due à ou consécutive à : c)		
due à ou consécutive à : d)		
<i>* Il s'agit de la maladie, du traumatisme, de la complication ayant entraîné la mort (et non du mode de décès, ex. : syncope, arrêt cardiaque...)</i>		
PARTIE II	Autres états morbides, facteurs ou états physiologiques (grossesse...) ayant contribué au décès, mais non mentionnés en Partie I	

forme de relations entre codes CIM, rassemblées dans des « tables de décisions ». Ces « tables » ont été développées dès le début des années 1980 par le *National Center for Health Statistics* (NCHS, le centre national de statistiques sur la santé) américain dans le module « *Automated Classification of Medical Entities* » (ACME, Classification automatisée des entités médicales) du système expert américain *Mortality Medical Data System* (Système de données médicales sur la mortalité) (National Center for Health Statistics, 2024 ; Lu et Wu, 2005). Elles ont été partagées internationalement puis traduites de la CIM 8 à la CIM 9, et de la CIM 9 à la CIM 10 (Navarra et alii, 2016) et suivent les mises à jour annuelles officielles de l'OMS. Les systèmes experts comme Iris comportent deux étapes.

Tout d'abord, il faut traduire le texte libre rédigé par les médecins en codes CIM. Cette première étape dépend de la langue. Après des standardisations du texte du certificat (synonymes, abréviations, termes non pertinents¹⁶), il est mis en correspondance avec un dictionnaire contenant 157 000 expressions, pour lesquelles chaque terme est associé à un ou plusieurs codes CIM 10.

Il s'agit ensuite de sélectionner la cause initiale selon les règles et directives de la CIM. Cette seconde étape s'applique à l'aide des tables de décision¹⁷, maintenues depuis 2011 par un consortium international de pays au sein de l'Institut Iris. Depuis le développement du *Multicausal and Unicausal Selection Engine* (MUSE, moteur de sélection de cause initiale et de cause multiple), cette étape prend aussi en compte les circonstances de décès, l'âge et le sexe du défunt.

Une trentaine de pays utilisent le logiciel Iris (Allemagne, Canada, Italie, Pays-Bas, Grande-Bretagne, etc.).

Les systèmes experts étaient tout d'abord utilisés uniquement en interactif : tous les certificats étaient vus au moins une fois par un agent de l'équipe de codage¹⁸. L'organisation du travail de l'équipe de codage reste aujourd'hui encore très marquée par l'usage du système expert : les codeurs corrigent, mettent en forme, simplifient le texte sans en modifier le fond, pour permettre au système expert de coder en interactif et les nosologues, et en dernier lieu les experts tranchent sur les cas délicats.

► Du batch automatique au codage interactif

Le logiciel Iris/Muse est lancé en batch de façon hebdomadaire dès réception dans la base de codage. Même s'il échoue à coder certains cas, il apporte des informations sur le type de blocage rencontré (cause manquante, plusieurs causes initiales possibles, etc.), et guide le travail de l'expert-codeur. Le taux de codage automatique complet du certificat est de 63 % pour la France. À titre de comparaison, il est de 65 % pour les Pays-Bas et de 80 % pour l'Italie et l'Angleterre¹⁹. En France, la première étape de codage pêche souvent : pour 83 % des rejets du batch, Iris/Muse n'a pas réussi à coder l'ensemble des codes CIM 10 attendus. Ceci provient du fait de fautes d'orthographe, de textes non pertinents pour le codage

¹⁶ Au total près de 1 000 expressions normalisées.

¹⁷ Les tables de décision comprennent près de 30 millions de relations entre paires de codes CIM.

¹⁸ Depuis 2015, l'équipe de codage vérifie seulement les cas improbables ou sensibles.

¹⁹ Hors décès pour lesquels une enquête est ouverte (*inquest*, 5 % des décès). Dans ces cas en Grande-Bretagne, l'information est transmise sous la forme d'un compte-rendu médical et non d'un certificat de décès.



Maintenir les tables de décision et le dictionnaire est lourd et la capitalisation de l'expertise humaine doit se faire en parallèle de la campagne de codage.



sur les certificats (« médecin traitant », « smur », etc.), ou encore des expressions polysémiques mentionnées plus haut, qui peuvent s'ajouter aux difficultés de lecture du texte sur les certificats papiers. Si toutes les causes sont codées, Iris-Muse identifie le code de la cause initiale dans 90 % des cas.

Les systèmes de règles connaissent d'autres limites. Maintenir les tables de décision et le dictionnaire se révèle lourd et la capitalisation de l'expertise humaine doit se faire en parallèle de la campagne de codage. En France, les performances du système expert étaient insuffisantes pour respecter les délais de diffusion des données. Aussi, un troisième mode de codage a été introduit dans la production régulière de la statistique depuis la production des causes de décès 2021. Il repose sur des algorithmes d'apprentissage profond, lesquels ont l'avantage de pouvoir proposer une codification complète de tous les certificats qui leur sont soumis en un temps record.

► Une troisième technique, basée sur l'intelligence artificielle (IA)



Avoir recours à l'intelligence artificielle dans le codage des causes de décès est innovant.



Avoir recours à l'intelligence artificielle dans le codage des causes de décès est innovant²⁰. Cette innovation est possible grâce au développement récent des modèles d'apprentissage profond de type réseaux de neurones profonds (RNP) pour le traitement du texte, à celui des puissances de calcul, à leurs applications à l'informatique médicale et à la présence d'un historique conséquent de certificats déjà codés. Un RNP

va transformer l'enchaînement des termes médicaux décrivant le processus morbide (séquence d'entrée) en un enchaînement de codes de la CIM et proposer une cause initiale (séquence de sortie). On parle alors d'algorithme « *sequence to sequence* » (*seq-to-seq*). En traitement automatique des langues et en apprentissage statistique, la transformation de séquence s'apparente à une traduction et celle de chaque terme médical en code à un problème de « classification ». Les RNP mobilisés ici sont des « *Transformers* » (Vaswani et alii, 2017). Les « *Transformers* » ont révolutionné les tâches de traduction grâce à une prise en compte du contexte (liens entre les mots d'une phrase) tout en ne nécessitant pas de puissances de calcul trop importantes²¹. Les calculs sous-jacents sont hautement parallélisés et nécessitent moins de données d'entraînement. Enfin, ces RNP sont disponibles en librairies open source²².

²⁰ Cette avancée est notable au niveau international. Seuls les États-Unis et le Portugal en complément de la France utilisent ou ont utilisé des RNP pour leur production de statistiques officielles sur les causes de décès (National Center for Health Statistics, 2023 ; Pita Ferreira et alii, 2022).

²¹ À l'inverse des *recurrent neural networks* par exemple.

²² Ici, Keras et TensorFlow sont mobilisées (<https://keras.io/> ; <https://www.tensorflow.org/?hl=fr>).

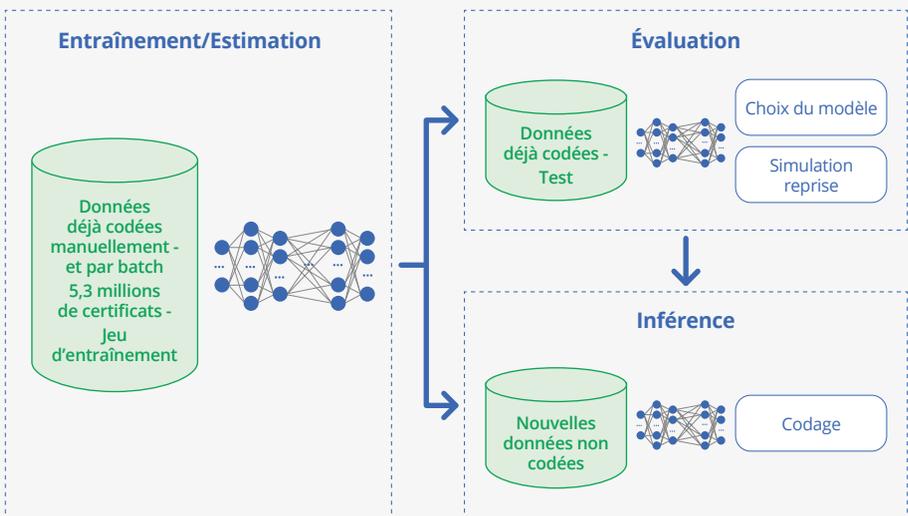
L'architecture des « *Transformers* » est de type encodeur/décodeur (Zambetta et alii, 2024 ; Hebbache et alii, 2024). Cela veut dire que la séquence d'entrée du modèle comme celle de sortie sont représentées dans un espace vectoriel de dimension beaucoup plus réduite que celles de leur corpus (tous les termes apparaissant au moins une fois dans la séquence d'entrée, respectivement celle de sortie). Cette étape de réduction de la dimension s'appelle « plongement lexical » ou *embedding* et s'appuie sur la proximité linguistique (Babet et alii, 2023). Par ailleurs, le « *Transformer* » modélise les éléments de contexte dans la phrase grâce au mécanisme d'« attention à plusieurs têtes » qui s'appuie sur des calculs parallélisés. En sortie, est calculée une distribution de probabilités qui permet de retenir la suite de codes la plus probable et composer ainsi la séquence de sortie.

Des travaux menés au CépiDc ont montré que les « *Transformers* » étaient performants dans le cas de la codification des causes de décès, dès lors que ces modèles étaient entraînés sur suffisamment de données déjà annotées (Falissard et alii, 2022). Après la preuve de concept et le potentiel démontré en recherche, l'enjeu a été de les intégrer dans la production courante, dans un cadre de qualité et en suivant les principes de la statistique européenne.

► Un mécanisme sophistiqué relevant de l'apprentissage statistique

En pratique, l'étape des pré-traitements permet de transformer le texte des certificats en entrée numérique interprétable par un RNP (*encadré 2*). Puis, comme en apprentissage statistique supervisé, l'étape d'entraînement consiste à estimer les millions de paramètres (« poids ») du réseau en utilisant l'historique des certificats déjà codés. Lors de l'étape de test, la performance du ou des réseaux à prédire avec

► **Figure 4 - Réseaux de neurones profonds - De l'entraînement à l'inférence**



exactitude la séquence de sortie est évaluée. Cette évaluation guide le choix du réseau estimé finalement retenu dans la production finale. Ce modèle, appliqué sur les textes d'un nouveau certificat, est désormais capable de prédire une séquence de codes et une cause initiale. Dans l'étape de prédiction, appelée aussi « inférence », on utilise donc ce réseau, en production, sur les certificats à coder (*figure 4*).

Ces RNP sont capables de « coder » environ 100 000 certificats par jour et par machine. Contrairement au système de règles qui rejette un certificat s'il n'a pas les règles pour le coder, un RNP prédira toujours une séquence de sortie quel que soit le certificat en entrée, pour peu que son texte suive correctement la structure spécifiée. Il associera aux termes de cette séquence prédite, des probabilités qui pourront par la suite être mobilisées pour estimer la qualité de la prédiction.

Les RNP sont utilisés pour prédire la séquence des codes et pour déterminer la cause initiale de certains certificats. Or, pour déterminer la cause initiale, il y a plusieurs stratégies : soit utiliser le code directement prédit par le modèle, soit appliquer le système expert Iris-muse sur la séquence des causes prédites par le modèle pour en déduire la cause initiale. L'approche retenue vise à retenir le meilleur des deux : elle consiste à entraîner un autre réseau de neurones capable de choisir entre les propositions possibles.

► Encadré 2. Le codage grâce aux réseaux de neurones profonds : des pré-traitements à l'inférence

Pour coder les causes de décès en CIM 10 en utilisant un réseau de neurones, on procède par étapes. Dans l'étape des pré-traitements, les données des certificats doivent être structurées en séquence interprétable par le modèle. On concatène les textes écrits sur chaque ligne du certificat de décès en ajoutant des variables de contexte. Pour une femme de 55 ans, décédée en 2017 et dont le certificat mentionnait « Ligne 1 : arrêt cardio respiratoire ; Ligne 2 : « dû à » épanchement pleural ; Ligne 3 : « dû à » métastases pulmonaires ; Ligne 4 : « dû à » cancer sein ; circonstances apparentes : mort naturelle, on obtient :

certificatpapier versioncertificat1997 femme age55ans annee2017 lignecause1 arrêt cardio respiratoire lignecause2 épanchement pleural lignecause3 métastases pulmonaires lignecause4 cancer sein lignecause7 mort naturelle causeinitiale

Les séquences en sortie ont quasiment la même structure sauf que les codes CIM 10 remplacent les termes. Elles se terminent par le code de la cause initiale, par exemple :

[start] certificatpapier versioncertificat1997 femme age55ans annee2017 lignecause1 r092 lignecause2 j90 lignecause3 c780 lignecause4 c509 lignecause7 causeinitiale c509 [end].

On découpe ensuite les séquences en « bouts », appelés « tokens » (*tokenizer*). Les tokens sont constitués pour la séquence d'entrée des mots qui se retrouvent au moins une fois dans le texte

d'un certificat (150 000 différents environ), et pour la séquence de sortie de codes de la CIM (6 300 différents).

« Inférence » veut dire « prédiction » : il s'agit d'appliquer le réseau de neurones à la séquence de textes d'un certificat pour en prédire la séquence de codes. En un jour, on prédit environ 100 000 certificats sur une machine avec une unité de traitement graphique (GPU).

La probabilité que la cause soit correctement prédite, ainsi que l'écart de probabilité entre le code le plus probable et le second code le plus probable, sont mobilisés dans l'étape de ciblage.

Au préalable il aura fallu « entraîner » le modèle, c'est-à-dire estimer les multiples paramètres qu'il contient pour l'adapter au mieux à la tâche de codage qu'on lui demande. L'« entraînement », c'est-à-dire l'estimation du modèle, minimise une fonction de coût (*loss fonction*) : on calcule les poids ou les paramètres (de l'ordre de 100 millions) en minimisant les écarts entre les séquences prédites par le modèle et les séquences données par la base d'entraînement. La base d'entraînement utilisée en début de campagne 2021 contient 5,3 millions de certificats. Elle est complétée en fin de campagne de la moitié* des certificats de 2021 codés manuellement pendant la campagne 2021 pour obtenir le codage final. Sur une machine avec une unité de traitement graphique de 48 Gigaoctets de RAM (GPU Nvidia RTX A6000), l'entraînement dure environ quatre jours pour trois millions de certificats.

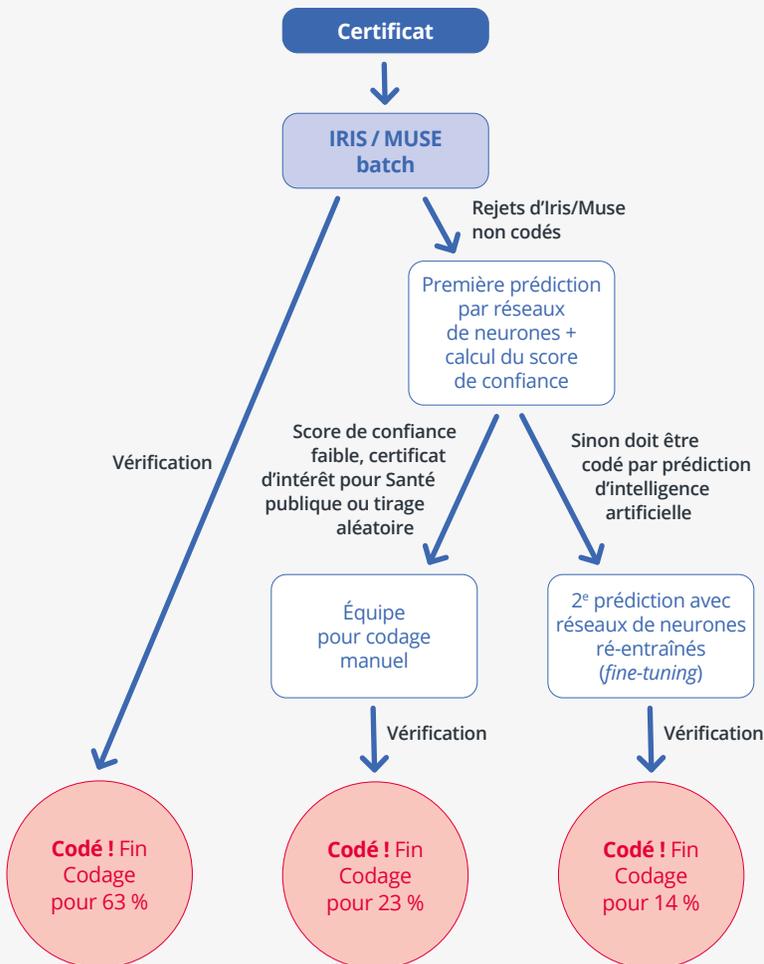
* L'autre moitié va servir à tester les modèles d'IA.

► Articuler les modes de codage et cibler l'expertise humaine

La campagne de production articule entre eux les trois modes de codage de façon à maximiser la qualité de la production dans ses différentes dimensions (*figure 5*). Les certificats qui ne sont pas codés entièrement automatiquement par batch vont être affectés soit à un codage manuel assisté, soit à un codage par prédiction de réseaux de neurones profonds en respectant les principes suivants :

- par campagne, le volume de certificats pouvant être codés manuellement est fixé à l'avance, en fonction des ressources humaines disponibles dans l'équipe de codage, de façon à respecter les délais de diffusion de la base statistique (93 000 certificats en 2021) ;

► **Figure 5 - Circuit de codage d'un certificat pendant la campagne de production**





La campagne de production articule entre eux les trois modes de codage de façon à maximiser la qualité de la production dans ses différentes dimensions.



- les situations pour lesquelles les prédictions des RNP sont moins vraisemblables, repérées sur la base d'un score de confiance modélisé et estimé pour chaque certificat, seront codées manuellement de façon ordonnée en commençant par les plus mauvaises. Il en sera de même pour certains cas d'intérêt pour la santé publique pour lesquels l'OMS requiert un suivi attentif : décès d'enfants en dessous d'un certain âge, morts maternelles, virus de l'immunodéficience humaine (VIH) ou syndrome d'immunodéficience acquise (SIDA) ;
- en complément, pour assurer l'entraînement et le contrôle réguliers des RNP, des échantillons tirés aléatoirement sont envoyés en reprise manuelle ;

- en fin de campagne, une partie de ces certificats codés manuellement, ainsi qu'une partie de ceux codés par batch automatique, sont utilisées pour ré-entraîner les RNP et obtenir la prédiction finale des certificats codés par RNP.

Ainsi, l'expertise humaine n'est plus mobilisée en bout de chaîne pour traiter les rejets du système de règles, mais dès le début pour alimenter les modèles et se concentrer sur les cas les plus complexes.

Au total, pour les décès en 2021, le batch aura codé 63 % des certificats, 23 % auront été codés en utilisant des prédictions de RNP (combinées ou non avec le système de règles), et 14 % auront demandé une intervention manuelle.

► Une évaluation statistique de la campagne annuelle de production

Introduire des RNP dans une production statistique nécessite de les « superviser », c'est-à-dire d'évaluer régulièrement leur performance en comparant leurs résultats à ceux des autres méthodes et de les réestimer régulièrement pour prendre en compte les nouveautés de vocabulaire, de description des pathologies ou les modifications dans les règles de codage.

La performance des RNP s'évalue en comparant la séquence de sortie prédite par le modèle avec la séquence véritablement observée sur un jeu de données déjà codées qui n'a pas été utilisé dans l'entraînement. Ce principe, essentiel en apprentissage statistique, permet de pallier le risque de sur-apprentissage de ces modèles qui par nature comprennent plus de paramètres à estimer qu'il n'y a d'observations dans la base d'entraînement.

La stratégie d'évaluation de la qualité d'une campagne de production repose aussi sur ce principe. Sur un jeu de test de référence n'entrant pas dans les entraînements des RNP et représentatif de la distribution des causes de décès en population générale, on peut estimer la cohérence de la campagne de codage d'une année donnée en comparaison

à une campagne traditionnelle en simulant l'articulation des modes de codage telle qu'elle a été réalisée lors de la production. Ainsi, en 2021, en prenant en compte le fait que 63 % des certificats ont été codés par batch (codage par rapport à une campagne classique identique), dans 95,7 % des cas, le code de cause initiale issu de la campagne à trois modes de codage est le même que celui que l'on aurait obtenu par une campagne classique de codage. Dans 97,3 % des cas, ce code apparaît dans la même rubrique de la shortlist (sélection) européenne²³ que celui obtenu par une campagne classique de codage. En outre, la campagne à trois modes de codage respecte les délais impartis en ne reposant que sur 14 % de codage manuel contre 37% dans une campagne traditionnelle.

► Un exemple d'usage de l'IA pour la production de statistique publique

L'usage des RNP entraînés sur l'historique des données déjà annotées dans la production de la statistique sur les causes de décès a permis de rattraper le retard de production, accumulé au cours du temps, en réduisant drastiquement le nombre de certificats devant être codés manuellement (Clanché et alii, 2023 ; Zambetta et alii, 2023). Ainsi, en septembre 2023, la France a été en mesure de fournir les données définitives des années 2018 et 2019 avec seulement 3 % de codage manuel²⁴, ciblé évidemment, tout en garantissant une cohérence de cause initiale par rapport à une campagne traditionnelle à 93,4 % au niveau le plus fin de la CIM, et de 95,6 % au niveau de la shortlist européenne. Depuis, le calendrier

régulier de production est raccourci, visant une fourniture des données 2023 dix-huit mois après la fin de l'année. Cet exemple témoigne de l'effet positif de l'intégration des procédures d'IA dans une production statistique régulière. Cette intégration est accessible et ouvre de nouvelles possibilités. Par exemple, un premier traitement purement automatique (RNP et batch) peut fournir des données provisoires²⁵ codées dès réception (*fast estimates* (premières estimations) utiles pour la veille sanitaire). Les précautions mises en œuvre pour guider le choix de la procédure s'appuient sur le code

Les précautions mises en œuvre pour guider le choix de la procédure s'appuient sur le code des bonnes pratiques de la statistique européenne : actualité, ponctualité, rapport coût-efficacité, etc.

des bonnes pratiques de la statistique européenne : actualité, ponctualité, rapport coût-efficacité, etc. En outre, les architectures des réseaux ont été choisies pour leur simplicité. Leur entraînement et leur inférence peuvent être réalisés sur des infrastructures conventionnelles. Il a été décidé de ne pas s'appuyer sur des modèles pré-entraînés, ni très complexes, pour garder entièrement le contrôle de la procédure statistique, depuis les données d'entraînement jusqu'aux modèles. On assure ainsi répliquabilité et transparence et on limite les risques de biais dans un souci d'impartialité et d'objectivité. Enfin, l'évaluation statistique qui permet de contrôler les modèles est inhérente à ces procédures. Leurs erreurs peuvent être mesurées, analysées, documentées.

²³ Principales marges de diffusion de la statistique européenne (Eurostat, 2012).

²⁴ Ce taux manuel pour 2018-2019 est faible comparé à celui pour 2021 (14 %) du fait d'un rattrapage pour les années 2018 et 2019, c'est-à-dire qu'elles ont été traitées en parallèle de la campagne courante, courant 2022 et 2023.

²⁵ Ainsi la diffusion des causes de décès en 2021 et en 2022 a été accompagnée de premières estimations sur les grandes causes de décès l'année suivante (Cadillac et alii, 2023, 2024).

Plusieurs axes de travaux se dessinent pour la suite : tout d'abord, sur l'explicabilité de ces modèles de façon à conforter la confiance des utilisateurs dans les statistiques produites à l'aide de l'intelligence artificielle, puis sur le maintien des modèles et des bases d'apprentissage (entraînement et test). Pour la statistique sur les causes de décès, le prochain défi sera aussi le passage à la version 11 de la CIM. Les modèles actuels sont totalement adhérents à la version de la nomenclature sur laquelle ils ont été entraînés, en l'occurrence la CIM 10. Passer à la CIM 11 implique d'adapter ou de changer les modèles, tout en maintenant la stratégie générale de la campagne de codage présentée ici.

► Fondements juridiques

- Règlement (CE) n° 1338/2008 du Parlement européen et du Conseil du 16 décembre 2008 relatif aux statistiques communautaires de la santé publique et de la santé et de la sécurité au travail. In : *site de l'Union européenne*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32008R1338>.
- Règlement (UE) n° 328/2011 de la Commission du 5 avril 2011 portant application du règlement (CE) n° 1338/2008 du Parlement européen et du Conseil relatif aux statistiques communautaires de la santé publique et de la santé et de la sécurité au travail, en ce qui concerne les statistiques sur les causes de décès. In : *site de l'Union européenne*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A32011R0328>.
- Article L2223-42 du Code général des collectivités locales. In : *site de Légifrance*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000023711931/2024-04-25/.
- Article R2213-1 du Code général des collectivités locales. In : *site de Légifrance*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006395864.
- Loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. In : *site de Légifrance*. [en ligne]. Mis à jour le 25 mars 2019. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573>.
- Arrêté du 17 juillet 2017 relatif aux deux modèles du certificat de décès. In : *site de Légifrance*. [en ligne]. Mis à jour le 11 mai 2020. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000035388290>.

► Bibliographie

- AUBENQUE, Maurice, DAMIANI, Paul et DERUFFE, Louise, 1978. La mortalité par cause en France de 1925 à 1974. In : *Journal de la Société statistique de Paris*. Tome 119, n° 3, pp. 276-295. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : http://www.numdam.org/item/JSFS_1978__119_3_276_0.pdf.
- BABET, Damien, DELTOUR, Quentin, FARIA, Thomas et HIMPENS, Stéphanie, 2023. Les réseaux de neurones appliqués à la statistique publique : méthodes et cas d'usages. In : *Document de travail*. [en ligne]. Février 2023. Insee. N° M2023/01. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/6801092>.
- BOUVIER-COLLE, Marie-Hélène, VALLIN, Jacques et HATTON, Françoise, 1990. Mortalité et causes de décès en France. In : *Les éditions de l'INSERM. Collection Grandes enquêtes en santé publique et épidémiologie*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k3325350b>.
- CADILLAC, Manon, FOUILLET, Anne, RIVERA, Cecilia, CLANCHÉ, François et COUDIN, Élise, 2023. Grandes causes de décès en France en 2021 : une année encore fortement marquée par le Covid-19. In : *Études et Résultats*. [en ligne]. Décembre 2023. N° 1288. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse/etudes-et-resultats/grandes-causes-de-deces-en-france-en-2021-une>.
- CADILLAC, Manon, FOUILLET, Anne, RIVERA, Cecilia et COUDIN, Élise, 2022. Les causes de décès en France en 2022 : recul du Covid-19 et hausse des maladies respiratoires. In : *Études et Résultats*. [en ligne] Octobre 2024. N° 1312. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse/etudes-et-resultats/241008_ER_les-causes-de-deces-2022.
- CLANCHÉ, François, RAZAKAMANANA, Nirintsoa, COUDIN, Élise et ROBERT, Aude, 2023. Les statistiques provisoires sur les causes de décès en 2018 et 2019 - Une nouvelle méthode de codage faisant appel à l'intelligence artificielle. In : *Drees Méthodes*. [en ligne]. Mars 2023. N° 8. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://drees.solidarites-sante.gouv.fr/publications-jeux-de-donnees-communique-de-presse/drees-methodes/les-statistiques-provisoires-sur>.
- DESROSIÈRES, Alain, 1993. *La politique des grands nombres. Histoire de la raison statistique*. La Découverte. ISBN 978-2707165046.
- EUROSTAT, 2012. Causes of death, Eurostat Shortlist. In : *site de Statistics Finland*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://stat.fi/en/luokitukset/kuolinsyyt/kuolinsyyt_11_20140101.
- EUROSTAT, 2024. Causes of death statistics. In : *site d'Eurostat*. [en ligne]. Mars 2024. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Causes_of_death_statistics.

- FALISSARD, Louis, MORGAND, Claire, ROUSSEL, Sylvie, IMBAUD, Claire, GHOSN, Walid, BOUNEBACHE, Karim et REY, Grégoire, 2020. A Deep Artificial Neural Network-Based Model for Prediction of Underlying Cause of Death From Death Certificates: Algorithm Development and Validation. In : *JMIR Medical Informatics*. [en ligne]. Avril 2020. Volume 8, n° 4. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://doi.org/10.2196/1712>.
- PITA FERREIRA, Patricia, GODINHO SIMÕES, Diogo, PINTO DE CARVALHO, Constança, DUARTE, Francisco, FERNANDES, Eugénia, CASACA CARVALHO, Pedro, LOFF, José Francisco, SOARES, Ana Paula, ALBUQUERQUE, Maria João, PINTO-LEITE, Pedro, PERALTA-SANTOS, André, 2022. Real-Time Classification of Causes of Death Using Artificial Intelligence - Sensitivity Analysis. In : *European Journal of Public Health*. [en ligne]. Octobre 2022. Volume 32, Supplément 3. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://doi.org/10.1093/eurpub/ckac129.574>.
- FOUILLET, Anne, GHOSN, Walid, RIVERA, Cecilia, CLANCHÉ, François et COUDIN, Élise, 2023. Grandes causes de mortalité en France en 2021 et tendances récentes. In : *Bulletin épidémiologique hebdomadaire*. [en ligne]. 19 décembre 2023. N° 26. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://beh.santepubliquefrance.fr/beh/2023/26/2023_26_1.html.
- FOUILLET, Anne, CADILLAC, Manon, RIVERA, Cecilia, et COUDIN, Élise, 2024. Grandes causes de mortalité en France en 2022 et tendances récentes. In : *Bulletin épidémiologique hebdomadaire*. [en ligne]. 8 octobre 2024. N° 18. [Consulté le 6 novembre 2024]. Disponible à l'adresse : http://beh.santepubliquefrance.fr/beh/2024/18/2024_18_1.html.
- HEBBACHE, Zina, BOULET, Pierre, ROBERT, Aude, ZAMBETTA, Elisa, RAZAKAMANA, Daniel, COUDIN, Élise et MARTIN, Diane, 2024. Rapport de production : année de décès 2021. In : *Document de travail du CépiDc*. [en ligne]. Mars 2024. N° 8. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.cepidc.inserm.fr/documentation/rapport-de-production-annee-de-deces-2021>.
- IRIS INSTITUTE, 2024. Iris software. In : *site de Federal Institute for Drugs and Medical Devices*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://www.bfarm.de/EN/Code-systems/Collaboration-and-projects/Iris-Institute/Iris-software/_node.html.
- LU, Tsung-Hsueh, TSAU, Shih-Ming et WU, Tzu-Chin, 2005. The Automated Classification of Medical Entities (ACME) system objectively assessed the appropriateness of underlying cause-of-death certification and assignment. In: *Journal of Clinical Epidemiology*. [en ligne]. Décembre 2005. Volume 58, n° 12, pp. 1277-1281. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <http://doi.org/10.1016/j.jclinepi.2005.03.017>.
- NATIONAL CENTER FOR HEALTH STATISTICS, 2023. MedCoder. In: *site du National Center for Health Statistics*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.cdc.gov/nchs/nvss/medcoder.htm>.
- NATIONAL CENTER FOR HEALTH STATISTICS, 2015. Mortality Medical Data System. In : *site du National Center for Health Statistics*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.cdc.gov/nchs/nvss/mmds.htm>.

- NAVARRA, Simone, CAPPELLA, Marisa, JOHANSSON, Lars Age, PELIKAN, László, FROVA, Luisa et GRIPPO, Francesco, 2016. Decision Table Editor: a web application for the management of the international tables for mortality coding. In: *Istat working papers*. [en ligne]. N° 6. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://www.istat.it/it/files/2016/04/IWP_06_2016.pdf.
- ORGANISATION MONDIALE DE LA SANTÉ, 2008. Classification statistique internationale des maladies et des problèmes de santé connexes, dixième révision, Volume 2. In : *site de l'Organisation mondiale de la santé*. [en ligne]. Édition 2008. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_fr_2008.pdf.
- PAVILLON, Gérard et LAURENT, Françoise, 2003. Certification et codification des causes médicales de décès. In : *Bulletin Épidémiologique Hebdomadaire*. [en ligne]. 8 juillet 2003, mis à jour le 30 août 2019, n° 30-31, p. 134-8. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.santepubliquefrance.fr/docs/certification-et-codification-des-causes-medicales-de-deces>.
- REY, Grégoire, 2016. Les données des certificats de décès en France : processus de production et principaux types d'analyse. In : *La Revue de Médecine Interne*. [en ligne]. Octobre 2016. Volume 37, n° 10, pp. 685-693. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <http://DOI.org/10.1016/j.revmed.2016.01.011>.
- VALLIN, Jacques et MESLÉ, France, 1988. Les causes de décès en France de 1925 à 1978. In : *Collection : Cahiers*. N° 115. ISBN 978-2-7332-0115-2.
- VASWANI, Ashish, SHAZEER, Noam, PARMAR, Niki, USZKOREIT, Jakob, JONES, Llion, GOMEZ, Aidan N., KAISER, Łukasz et POLOSUKHIN, Illia, 2017. Attention Is All You Need. In : *Advances in Neural Information Processing Systems*. [en ligne]. [Consulté le 6 novembre 2024]. Disponible à l'adresse : https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- ZAMBETTA, Élisabeth, RAZAKAMANANA, Nirintsoa, ROBERT, Aude, CLANCHÉ, François, RIVERA, Cecilia, MARTIN, Diane, HEBBACHE, Zina, FLICOTEUX, Rémi et COUDIN Élise, 2023. Codage des causes de décès de 2018 et 2019 en CIM10 - Approche combinant deep learning, système expert et codage manuel ciblé. In : *Document de travail du CépiDc N° 2*. [en ligne]. Septembre 2023. [Consulté le 6 novembre 2024]. Disponible à l'adresse : <https://www.cepidc.inserm.fr/documentation/codage-des-causes-de-deces-de-2018-et-2019-en-cim10-approche-combinant-deep-learning-systeme-expert-et-codage-manuel-cible-document-de-travail-cepidc-n22023>.
- ZAMBETTA, Élisabeth, RAZAKAMANANA, Nirintsoa, ROBERT, Aude, CLANCHÉ, François, RIVERA, Cecilia, MARTIN, Diane, HEBBACHE, Zina, FLICOTEUX, Rémi et COUDIN Élise, 2024. Combining deep neural networks, a rule-based expert system and targeted manual coding for ICD-10 coding causes of death of French death certificates from 2018 to 2019. In : *International Journal of Medical Informatics*. Août 2024. Volume 188.

Mesurer et qualifier l'insertion professionnelle des étudiants quittant l'enseignement supérieur

Le dispositif InserSup



Karl Even*

Le système d'information InserSup a été déployé par la sous-direction des systèmes d'information et des études statistiques (Sies) du ministère de l'Enseignement supérieur et de la Recherche fin 2023. L'objectif est de mesurer et qualifier l'insertion professionnelle des diplômés de l'enseignement supérieur. Ce dispositif s'appuie sur des données administratives exhaustives pour produire, tous les six mois, une batterie d'indicateurs à une granularité plus fine (7 000 formations fin 2024) que celle des enquêtes et assure une meilleure comparabilité des données.

Exposer des indicateurs largement scrutés exige une automatisation pour sécuriser le processus, assurer une traçabilité pour suivre le cheminement des fichiers et reproduire cette intégration de données. Cela nécessite également une coordination et une coopération sans faille avec les ministères en charge de l'éducation nationale, de l'emploi et l'Insee, producteurs de données.

Le système répond au besoin des étudiants de disposer de cette information pour guider leur choix de formation lorsqu'ils s'inscrivent sur les plateformes Parcoursup ou Mon Master où sont présentés les indicateurs. Il apporte une information riche et fine aux acteurs que sont le ministère de l'Enseignement supérieur et les universités (pour adapter l'offre de formation), les professionnels de l'orientation (pour l'adéquation emploi-formation) et le monde de la recherche (sur l'efficacité des formations).

 At the end of 2023, the ministry of Higher Education and Research implemented InserSup, the information system aiming to measure and characterise the professional integration of higher education graduates. This system based on exhaustive administrative data, produces a set of indicators every six months at a finer level of detail (7,000 training programs by the end of 2024) than that provided by surveys, ensuring better data comparability.

Disseminating widely scrutinized indicators requires automation to secure the process, ensure traceability to monitor the movement of files, and enable reproducible data integration. It also requires seamless coordination and cooperation with the ministries responsible for National Education, Employment, and Insee, the data producers.

The system meets students' need for this information to guide their training choice when registering on platforms like Parcoursup or MonMaster, where the indicators are presented. It provides rich, detailed information to key stakeholders – such as the ministry of Higher Education and universities (to adapt training offerings), career guidance professionals (for skills matching), and the research community (for evaluating training effectiveness).

* Directeur du projet InserSup et chef de la mission InserSup, sous-direction des systèmes d'information et des études statistiques (ministère de l'Enseignement supérieur et de la Recherche).
karl.even@enseignementsup.gouv.fr

L'insertion professionnelle est le débouché naturel des étudiants à l'issue de leurs études dans l'enseignement supérieur. Réussir cette insertion dans de bonnes conditions est au centre de leurs préoccupations. C'est même l'un des critères retenus lors de leur inscription dans la formation. Pour les étudiants, disposer de l'information sur l'insertion de leurs prédécesseurs ayant suivi tel ou tel cursus est une aide à la décision. Comme l'offre de formations dans le supérieur est très large, cette aide est utile aux différents jalons du parcours étudiant et plus particulièrement dès la terminale lors des inscriptions dans Parcoursup¹ ou encore à la fin de la seconde année de brevet de technicien supérieur (BTS²) pour ceux qui souhaitent s'orienter en licence professionnelle, mais aussi en troisième année de licence générale dans Mon Master³ pour ceux qui visent un master.

L'insertion professionnelle est tout aussi importante pour les établissements de formation. Ce critère objectif d'appréciation de leur activité figure parmi les indicateurs de performance annuels du ministère de l'Enseignement supérieur et de la Recherche (MESR). Il est aussi utilisé par les organismes d'évaluation des formations : le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur, la Commission des titres d'ingénieur et la Commission d'évaluation des formations et diplômes de gestion. L'insertion professionnelle fait également partie des contrats d'objectifs, de moyens et de performance entre le ministère et les établissements universitaires, qui visent notamment à renforcer le pilotage de l'offre de formation pour améliorer la réussite des étudiants et leur insertion professionnelle.



Depuis 2007, les établissements universitaires doivent rendre publiques les statistiques sur l'insertion professionnelle.



Depuis 2007⁴, les établissements universitaires doivent rendre publiques les statistiques sur l'insertion professionnelle. Depuis 2009, la sous-direction des systèmes d'information et des études statistiques (Sies) du MESR coordonne et exploite annuellement des enquêtes d'insertion professionnelle (IP) auprès des diplômés de licence professionnelle et de master réalisées par les établissements universitaires. Le Sies mène également une enquête auprès des diplômés de

doctorat avec le concours des écoles doctorales. Les résultats sont diffusés annuellement sur la plateforme d'open data du ministère d'une part et par les établissements sur leur périmètre, d'autre part.

La loi du 5 septembre 2018 pour la liberté de choisir son avenir professionnel⁵ prévoit notamment la publication de statistiques par centre de formation des apprentis ; ainsi, la Direction de l'évaluation, de la prospective et de la performance (Depp) a déployé le dispositif Inserjeunes en 2021 (Caron et Midy, 2022). Par la suite et dans cette continuité, le Sies a été sollicité à travers une demande interministérielle prévoyant la mise en production

¹ Parcoursup est une plateforme web destinée à recueillir et gérer les vœux d'affectation des futurs étudiants de l'enseignement supérieur français.

² Le brevet de technicien supérieur se prépare dans un lycée, après le baccalauréat. Cette formation permet d'obtenir un diplôme professionnalisé en deux ans.

³ Mon Master est une plateforme web destinée à recueillir les vœux d'affectation des futurs étudiants en master de l'enseignement supérieur français.

⁴ Loi LRU (loi relative aux libertés et responsabilités des universités) du 10 août 2007 (voir les références juridiques en fin d'article).

⁵ Voir les références juridiques en fin d'article.



Ce dispositif est basé sur des données exhaustives, et ouvre des possibilités supplémentaires à celles permises jusqu'ici par les enquêtes.



fin 2023 d'un système d'information pour permettre de calculer et d'exposer⁶ des indicateurs d'insertion professionnelle des étudiants quittant l'enseignement supérieur à partir de fichiers administratifs. Pour y répondre, au cours de l'année 2022, le Sies a construit un prototype développé et industrialisé par la suite. Le dispositif, nommé InserSup, a été mis en place et déployé à l'automne 2023 (**encadré 1**). Il a permis d'exposer les premiers indicateurs au grand public en décembre 2023.

Ce dispositif est basé sur des données exhaustives, et ouvre des possibilités supplémentaires à celles permises jusqu'ici par les enquêtes, en fournissant des indicateurs très précis et d'une granularité fine. À cet effet, il constitue une aide précieuse pour l'ajustement de l'offre de formation par le ministère ou les établissements de formation, ou encore pour documenter le sujet de l'adéquation des formations aux emplois occupés par les étudiants quittant l'enseignement supérieur (relation emploi-formation).

► **Encadré 1. Champ des établissements et des formations couvert par InserSup : calendrier d'extension**

Le périmètre couvert par InserSup doit s'étendre progressivement. Au moment de son déploiement, en décembre 2023, il couvrait les diplômes de licence professionnelle et de master délivrés par les universités, soit celui des enquêtes d'insertion professionnelle menées par le ministère de l'Enseignement supérieur. En juillet 2024, le champ a été étendu aux licences générales et aux établissements tels que l'Université Paris Dauphine-PSL, l'Institut national des langues et civilisations orientales (Inalco), les Instituts

d'études politiques, l'Observatoire de Paris, les Instituts catholiques, les universités de Nouvelle-Calédonie et de Polynésie. En décembre 2024, les écoles d'ingénieurs et les écoles de commerce et de gestion s'ajoutent au dispositif.

En 2025, les bachelors universitaires de technologie (première promotion en 2024) et les écoles sous tutelle du ministère de la Culture intégreront InserSup. En 2026, les doctorats compléteront le dispositif.

* L'Université Paris Dauphine-PSL fixe librement ses droits d'inscriptions, qui sont très élevés par rapport aux universités : elle se rapproche plus d'une école (sélection par les droits d'inscriptions).

► **Une coopération au sein du système statistique public** —

Un tel dispositif nécessite de mobiliser des informations produites dans différents environnements statistiques ou ministériels qui ont chacun leur logique et leur culture propre : l'information sur les étudiants relève du Sies (MESR) et celle sur les élèves de la Depp (ministère en charge de l'éducation nationale), tandis que l'information sur les personnes présentes sur le marché du travail est de la compétence de la Dares⁷ (ministère en charge de l'emploi et du travail) et les données sur les rémunérations proviennent de l'Insee (ministère en charge de l'économie) via la Dares. Sa mise en place présuppose une coopération entre

6 Ce terme recouvre la mise à disposition d'indicateurs aux personnes qui consultent les offres de formation sur les plateformes dédiées pour ensuite s'inscrire.

7 La Direction de l'animation de la recherche, des études et des statistiques (Dares) est le service statistique ministériel dans le domaine du travail.



Sa mise en place présuppose une coopération entre plusieurs ministères sur le plan logistique mais aussi pour se comprendre sur le sens et la finalité.

plusieurs ministères sur le plan logistique, certes, mais aussi pour se comprendre sur le sens et la finalité. Les périmètres du Sies et de la Depp sont proches et s'interpénètrent ; la coopération entre les deux univers est déjà à l'œuvre, car elle est historique. En revanche, celle avec la Dares était à construire.



Les données du monde du travail n'y sont pas parfaitement connues, contrairement à celles

du monde étudiant dont c'est le cœur de métier. Aussi, utiliser la source administrative sur l'emploi fournie par la Dares (la Déclaration Sociale Nominative, DSN⁸), nécessite de développer des compétences et de comprendre les concepts afférents au marché du travail. Elle oblige aussi à acquérir une connaissance du contenu de la source, portant à la fois sur la signification des données de gestion fournies et sur leur utilisation pour concevoir les indicateurs à calculer.

► Les sources mobilisées par InserSup

Les données sur les étudiants proviennent du Système d'information sur le suivi des étudiants (Sise)⁹. Elles sont constituées des remontées administratives des établissements publics et privés. Elles portent sur les inscriptions dans les différentes formations (Sise Inscriptions), d'une part, et sur la réussite aux diplômes préparés (Sise Résultats), d'autre part. Elles se présentent chacune sous la forme de six fichiers, un par type d'établissement formateur (universités, écoles d'ingénieurs, écoles de management, etc.) qui n'ont pas tous exactement le même dessin de fichier et qu'il convient d'unifier pour les besoins d'InserSup. Les variables de ces fichiers mobilisées pour InserSup portent sur la formation suivie (diplôme et mention du diplôme préparé, réussite/échec au diplôme), l'établissement ayant délivré le diplôme (identifiant, libellé et commune d'implantation), le régime d'inscription (apprentissage ou non). Figurent en outre l'année du baccalauréat, le type de baccalauréat obtenu et des données sociodémographiques des étudiants (date et commune de naissance, nationalité, sexe, profession et catégorie socioprofessionnelle (PCS) des parents).

Les données issues de Sise sont complétées par d'autres en provenance de la Depp sur les inscriptions en apprentissage dans les sections de techniciens supérieurs (STS), issues du système d'information sur la formation des apprentis (Sifa) et transmises par les centres de formation, et par les fichiers d'inscription en classes préparatoires aux grandes écoles dans les lycées (BPBAC).

Les données sur l'emploi sont extraites de la DSN – renseignée mensuellement par tous les employeurs pour chacun de leurs salariés – et fournies par la Dares à partir de son système d'information sur les mouvements de main-d'œuvre. Les informations utilisées portent sur l'occupation d'un emploi (dates de début et de fin de contrat) et les caractéristiques de l'emploi occupé (type de contrat, dispositif d'aide par l'État, quotité de travail, convention collective de branche couvrant le salarié) et de l'employeur

⁸ La déclaration sociale nominative (DSN) permet à l'employeur de déclarer et payer ses cotisations sociales via les données de l'entreprise sur ses salariés.

⁹ Voir les références juridiques en fin d'article.



Faire le lien entre des données administratives sur les étudiants et des données administratives sur le marché du travail est a priori impossible. Le rapprochement de ces fichiers est rendu possible, à partir des traits d'identité des étudiants.



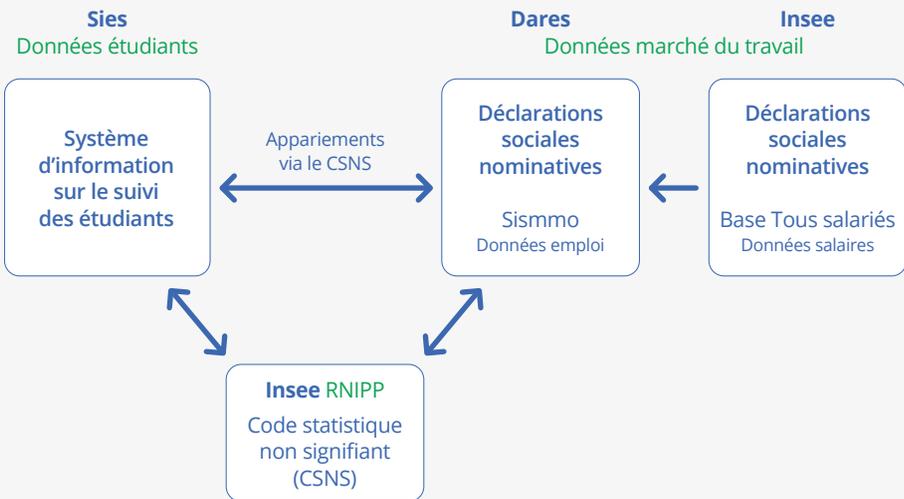
(secteur d'activité, numéro **Siret**¹⁰, forme juridique, taille de l'entreprise, localisation géographique). Elles sont complétées par les données de rémunérations, issues également de la DSN, reçues et redressées par l'Insee dans le cadre de la construction de la **Base Tous Salariés**. La jointure entre l'extrait de cette base et le fichier de la Dares est effectuée par la Dares.

Faire le lien entre des données administratives sur les étudiants et des données administratives sur le marché

du travail est a priori impossible. En effet, les données de ces deux univers n'ont pas d'identifiant commun permettant de les relier. L'identifiant national étudiant (INE) utilisé dans les fichiers du Sies ou de la Depp n'existe pas dans la DSN, et le numéro d'inscription au répertoire (NIR) ou l'identifiant de contrat de travail ne figurent pas

► Figure 1 - Le rapprochement des fichiers via le code statistique non signifiant

Le rapprochement des fichiers du Sies avec les fichiers du marché du travail s'effectue via le code statistique non signifiant (CSNS) obtenu au préalable auprès de l'Insee.



Sies : Sous-direction des systèmes d'information et des études statistiques : service statistique du ministère de l'Enseignement supérieur et de la Recherche.

Dares : Direction de l'animation, de la recherche, des études et des statistiques : service statistique du ministère en charge du travail.

RNIPP : Répertoire national d'identification des personnes physiques. Ce répertoire gère les numéros d'inscription au répertoire (NIR), appelés aussi « numéros de sécurité sociale ».

Sismmo : Système d'information sur les mouvements de main-d'œuvre.

¹⁰ <https://www.insee.fr/fr/metadonnees/definition/c1377>.

dans les fichiers Sise. Le rapprochement de ces fichiers est rendu possible, à partir des traits d'identité des étudiants (nom, prénom, date et lieu de naissance) par le service d'identification et de cryptage rendu par l'Insee depuis 2022, via le code statistique non significatif (CSNS) (Bénichou et alii, 2023) qui sert de clé d'appariement (Koumarianos et alii, 2024) (*figure 1*). Le Sies a été un des services statistiques ministériels (SSM) pilotes, fin 2021-début 2022, aux côtés de l'Insee, pour les expertises préalables à la mise en production du CSNS. Pour des raisons de sécurité, son utilisation génère des contraintes particulières sur les échanges des données et sur le stockage de l'information ; en effet, le CSNS ne doit pas être présent dans les fichiers des données d'origine.

► Le processus : des appariements séquentiels

Établir la population des étudiants quittant l'enseignement supérieur nécessite plusieurs étapes en amont. Repérer leur emploi en nécessite d'autres. Toutes ces étapes reposent sur des appariements réalisés séquentiellement. Le principe général est le suivant :

L'identification des étudiants quittant l'enseignement supérieur s'effectue dans un premier temps en repérant ceux en dernière année de leur formation¹¹ dans les six fichiers Sise Résultats concaténés¹² au préalable. On les apparie ensuite avec les fichiers Sise Inscriptions (également concaténés), afin de repérer ceux qui poursuivent des études en France à l'issue de leur dernière année de formation. Le repérage est effectué dans les fichiers des deux années universitaires qui suivent (*figure 2*) pour prendre en compte les éventuelles césures ou interruptions momentanées des études. L'opération est réalisée quels que soient la formation suivie et l'établissement délivrant le diplôme. Le même processus est ensuite appliqué aux fichiers Sifa pour les apprentis, puis à ceux de BPBAC pour les classes préparatoires aux grandes écoles dans les lycées. Ces repérages sont effectués en appariant sur l'identifiant national étudiant (INE). Mais, pour les promotions antérieures à 2020, un étudiant peut avoir différents INE au cours de son parcours étudiant (à l'occasion de changements d'académie ou d'établissement). Aussi un système a-t-il été mis en place au Sies pour retracer les différents INE attribués à un individu donné au fil du temps, système dans lequel les différents INE sont ancrés à un identifiant fixe.

À la fin de cette étape (22 fichiers mobilisés), une première liste d'étudiants quittant l'enseignement supérieur est arrêtée. Un dernier repérage de poursuites d'études en contrat d'apprentissage ou en convention industrielle de formation par la recherche (Cifre) est effectué par appariement avec la DSN.

Le repérage des situations d'emploi des étudiants quittant l'enseignement supérieur est ensuite réalisé à 6, 12, 18, 24 et 30 mois après leur sortie. Pour une promotion donnée, cette étape mobilise cinq fichiers supplémentaires extraits de la DSN, correspondant aux mois de décembre (pour l'insertion à 6, 18 et 30 mois) et juin (pour l'insertion à 12 et 24 mois) de trois années de DSN. Compte tenu des dates de délivrance des diplômes, qui s'étalent de juin à la fin de l'année, l'expression « insertion à 6 mois » est une commodité de langage.

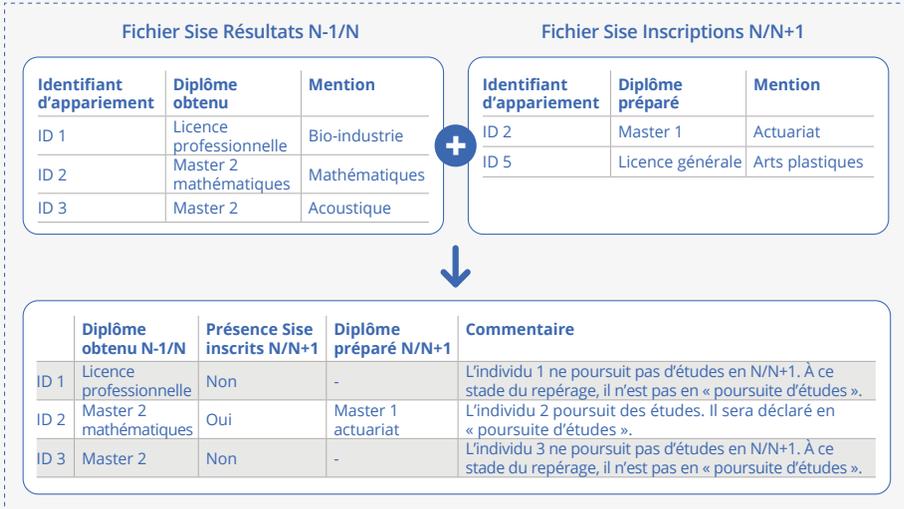
¹¹ Par exemple troisième année de licence professionnelle ou seconde année de master.

¹² Concaténation : enchaînement de deux listes ou de deux chaînes de caractères mises bout à bout.

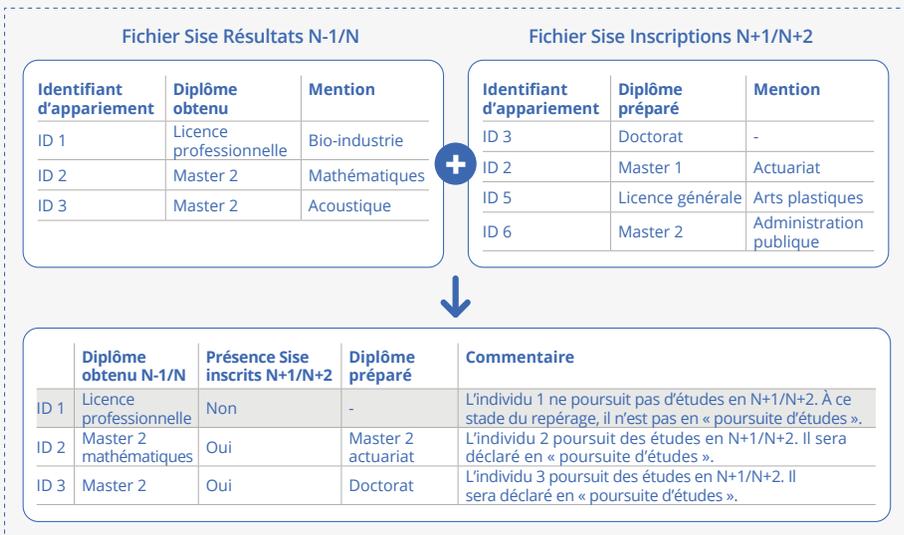
► Figure 2 - Zoom sur les deux premières étapes : appariements entre Sise Résultats et Sise Inscriptions

Les deux premières étapes de la phase d'identification de la population d'intérêt consistent à repérer au sein du fichier Sise Résultats de l'année N-1/N les diplômés poursuivant leurs études et ceux les arrêtant, repérables dans les fichiers Sise Inscriptions des années N/N+1 et N+1/N+2.

Appariement n° 1



Appariement n° 2



Les appariements se poursuivent entre fichiers Sise, puis avec les autres fichiers de formation (Sifa, BPBAC), et enfin ceux du marché du travail (DSN) (figure 3).

Tout au long du processus d'appariement, le repérage et également le non repérage des individus apportent une information signifiante : au cours des étapes de sélection de la population d'intérêt (les sortants du supérieur), ce sont précisément ceux qui ne sont pas repérés qui finissent par appartenir à cette population. Au cours de l'étape de repérage des situations d'emploi, être repéré est le signe que l'individu occupe un emploi.

Au total, 27 fichiers sont mobilisés pour une promotion donnée, auxquels s'ajoute un fichier de nomenclature « diplômés » et l'accès par interface de programmation d'application au répertoire des établissements (base centrale des établissements-BCE). Cela permet de constituer une base de données individuelles contenant toutes les informations des fichiers étudiants, auxquelles s'ajoutent des variables indicatrices informant sur l'appartenance ou non à la population d'intérêt aux différentes étapes du processus. Les informations des fichiers travail viennent compléter l'ensemble (*figure 3*).

► Processus de calculs des indicateurs, diffusion des fichiers et des résultats



Sur le seul champ universitaire, les indicateurs sont calculés sur près de 6 100 formations dispensées dans les établissements pour les diplômés de licence (générale ou professionnelle) et master.



Le calcul des indicateurs d'insertion professionnelle est effectué à partir de la base de données ainsi constituée. Elle est alimentée tous les six mois par la DSN et une fois par an par les fichiers étudiants. Cette fréquence ponctue le rythme des campagnes de calcul des indicateurs : une campagne au printemps et une campagne à l'automne. Plusieurs promotions d'étudiants sont traitées simultanément au cours de chaque campagne. A titre d'exemple, la campagne de l'automne 2024 a produit l'insertion à 6 et 12 mois des

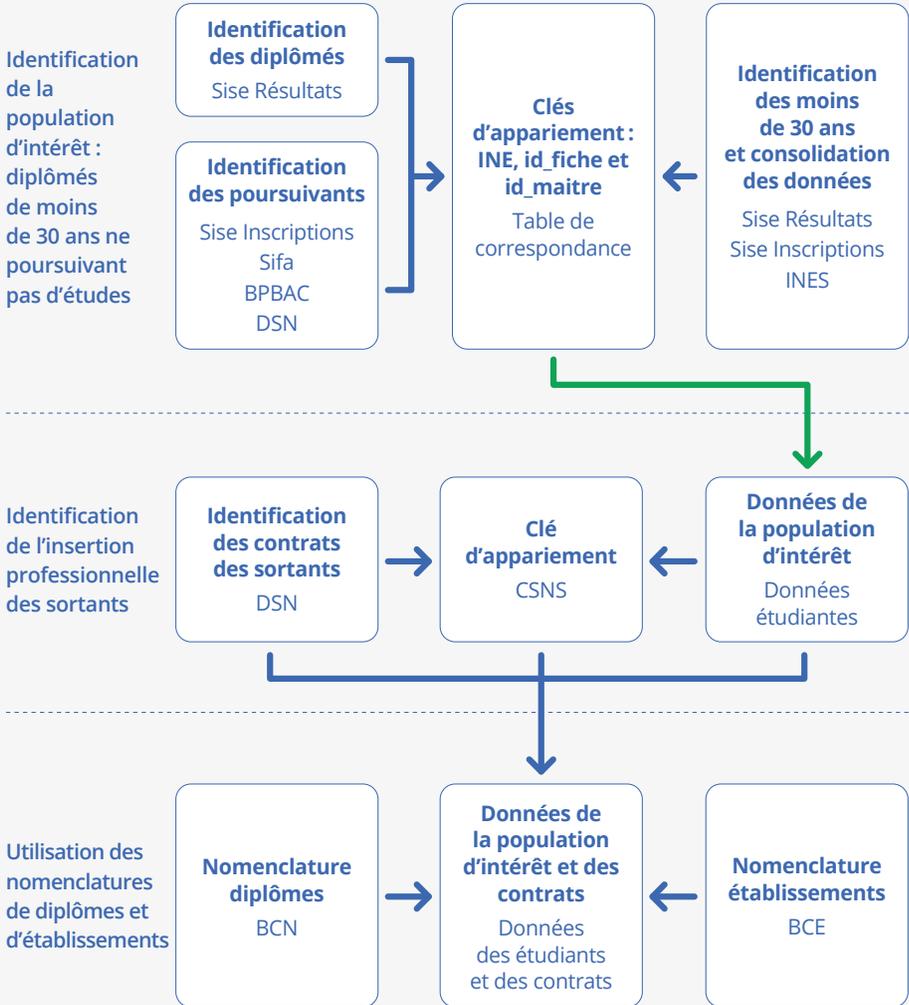
sortants du supérieur en 2023 et a complété l'insertion des sortants en 2022, qui avait été traitée au cours de la campagne précédente, en produisant les indicateurs à 24 mois à partir de l'information nouvelle parvenue entre temps. Chaque campagne nécessite un lourd travail de contrôle : vérification du contenu des nouveaux fichiers importés (examen des dessins de fichiers et des données elles mêmes par des procédures automatisées) et examen des indicateurs calculés jusqu'au niveau de la formation par établissement, en comparaison avec les chiffres calculés lors des campagnes précédentes pour une même promotion et entre promotions. Ces comparaisons suivent une procédure automatisée compte tenu du nombre important de sous populations pour lesquelles les indicateurs sont calculés. À titre d'exemple, sur le seul champ universitaire, les indicateurs sont calculés sur près de 6 100 formations dispensées dans les établissements pour les diplômés de licence (générale ou professionnelle) et master. Leur mise à disposition sur [la plateforme d'open data](#) du ministère de l'Enseignement supérieur et de la Recherche¹³ s'effectue à cette cadence. Ils sont par ailleurs exposés une fois par an sur différentes plateformes afin d'informer les élèves et étudiants (Parcoursup, Mon Master, Onisep¹⁴, etc.).

¹³ <https://data.enseignementsup-recherche.gouv.fr/pages/home/>.

¹⁴ Onisep : Office national d'information sur les enseignements et les professions. C'est un opérateur de l'État qui produit et diffuse toute l'information sur les formations et les métiers.

► Figure 3 - Les enchaînements d'appariement des différents fichiers mobilisés dans InserSup

La phase d'identification de la population d'intérêt repose sur une succession d'appariements de fichiers, réalisés dans un ordre établi. À la fin de cette phase, l'appariement avec la déclaration sociale nominative est réalisé. Une ultime phase consiste à accoler les libellés des formations, diplômes et établissements.



Sise : Système d'information sur le suivi des étudiants.
 Sise Inscriptions : Inscriptions dans les formations.
 Sise Résultats : Réussite aux diplômes préparés.
 Sifa : Système d'information sur la formation des apprentis.
 BPBAC : Fichiers d'inscription en classes préparatoires aux grandes écoles dans les lycées.
 DSN : Déclaration sociale nominative.
 INE : Identifiant national étudiant.
 INES : Identifiant national dans l'enseignement supérieur.
 CSNS : Code statistique non signifiant.
 BCN : Base centrale des nomenclatures.
 BCE : Base centrale des établissements.

La mesure de l'insertion professionnelle diffère de celle utilisée jusqu'à présent à partir des données d'enquête. Il s'agit du taux d'emploi et non plus du résultat du rapport entre le nombre de personnes en emploi et les actifs concernés (**encadré 2**). L'ensemble de la trajectoire d'insertion professionnelle au cours des 30 premiers mois après la sortie des études (Aubry et alii, 2024) est ainsi mis à disposition. Ce n'est donc pas seulement la quantité d'étudiants insérés mais aussi la vitesse d'insertion que le dispositif permet de documenter. Elle permet des comparaisons temporelles et entre formations ou établissements et est complétée par des indicateurs qualifiant l'insertion et la description de l'emploi occupé : le type de contrat (contrat à durée indéterminée (CDI), contrat à durée déterminée (CDD), etc.), la catégorie socio-professionnelle du salarié et sa rémunération et le secteur d'activité de l'entreprise. Les indicateurs d'insertion sont en outre déclinés selon que l'étudiant a suivi sa formation sous le régime de l'apprentissage ou non, selon sa nationalité et selon la réussite ou non au diplôme.

► Encadré 2. Une nouvelle mesure de l'insertion

La mesure de l'insertion professionnelle retenue par InserSup repose sur le taux d'emploi tel que défini et calculé par l'Insee, la Dares ou l'OCDE* dans ses comparaisons internationales. La Depp pour le dispositif InserJeunes y recourt également. Le taux d'emploi rapporte le nombre de personnes occupant un emploi à la population totale correspondante. Il peut être calculé pour une sous-population donnée, par exemple les titulaires d'un diplôme.

Pour la première version d'InserSup où, seul, l'emploi salarié en France est repérable dans les fichiers disponibles retraçant l'emploi, l'indicateur d'insertion professionnelle calculé est le taux d'emploi salarié en France, qui correspond à la proportion de diplômés en emploi salarié en France au sein des diplômés sortants de la promotion.

Ce taux diffère du taux d'insertion historiquement calculé avec les enquêtes d'insertion professionnelle (IP), qui rapporte le nombre de diplômés en emploi (quel qu'il soit et où qu'il soit) aux seuls diplômés actifs (en emploi ou au chômage). Il s'intéresse aux seuls étudiants sur le marché du travail, et mesure, en miroir, le taux de chômage de cette population en répondant à la

question : parmi ceux qui souhaitent un emploi, combien en ont un ? Il ne tient pas compte de l'inactivité qui s'élève à près de 5 % selon les enquêtes IP (et jusqu'à 10 % dans les disciplines de langues étrangères appliquées).

Enquête IP (taux d'insertion) :

$$\frac{\text{nb-sortants-occupants-un-emploi}}{\text{nb-sortants-actifs}}$$

Enquête IP modifiée 2023-2024 (taux d'emploi) :

$$\frac{\text{nb-sortants-occupants-un-emploi}}{\text{nb-sortants}}$$

InserSup en 2024 (taux d'emploi salarié en France) :

$$\frac{\text{nb-sortants-occupants-un-emploi-salarié-en-France}}{\text{nb-sortants}}$$

InserSup en 2025 (taux d'emploi en France) :

$$\frac{\text{nb-sortants-occupants-un-emploi-en-France}}{\text{nb-sortants}}$$

* OCDE : l'Organisation de coopération et de développement économiques est une organisation intergouvernementale d'études économiques (38 pays membres). <https://www.oecd.org/fr/>.

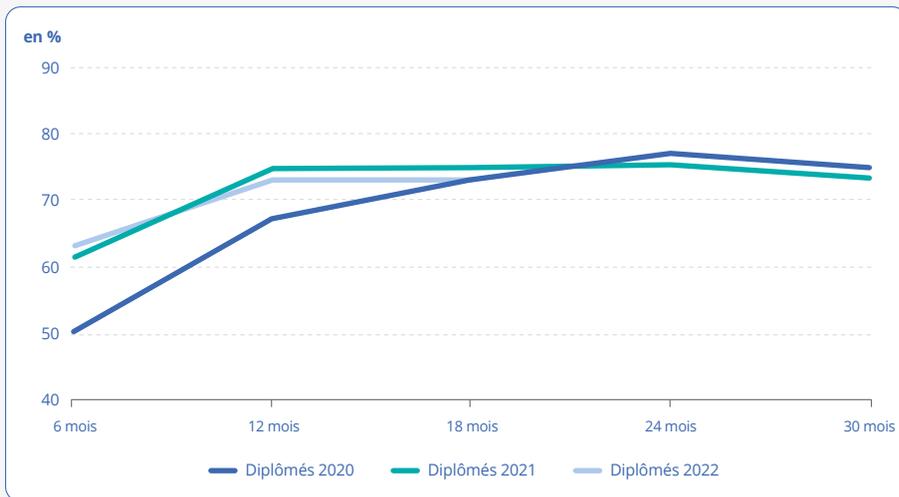
** Sortants actifs : sortants occupant un emploi ou au chômage.

Tous ces indicateurs sont mis à disposition du grand public et aux établissements. Ils sont déclinés par formation au sein de chaque établissement. Mais les exposer à un niveau fin (formation par établissement) ne va pas de soi : cela nécessite le respect de la confidentialité et aussi des mesures comparables aussi bien entre différents lieux (établissements) ou différentes formations que dans le temps (chiffrage entre deux promotions de diplômés, ou pour une même promotion entre deux dates).

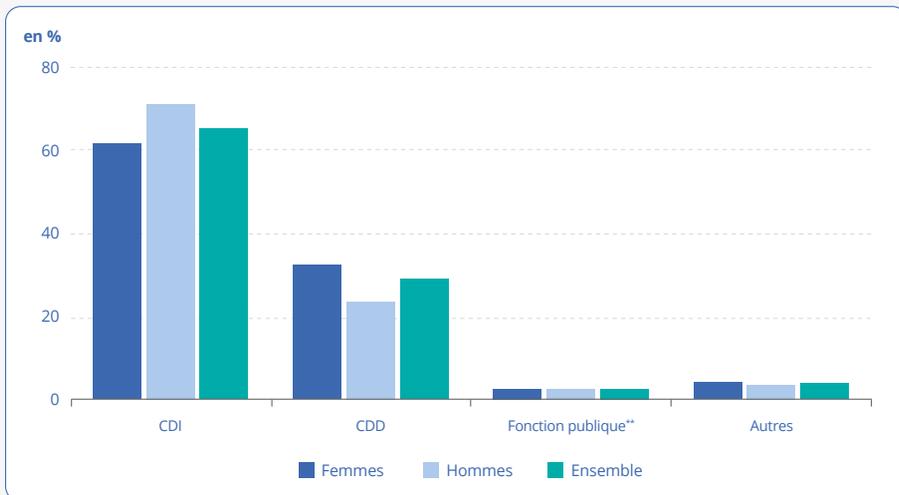
Dans cette perspective, les indicateurs sont affichés si la taille de la population concernée dépasse 19 individus, afin d'éviter des évolutions temporelles de l'indicateur de taux d'emploi salarié en France trop perturbées et ainsi assurer une fiabilité chronologique de la mesure de l'insertion. Cette règle dépasse donc le strict respect du secret statistique classique (Redor, 2023). Mais en fixant ce seuil, 80 % des mentions de licences professionnelles et 60 % des mentions de master ne pourraient pas être diffusées. Aussi, pour maximiser la diffusion

► Figure 4 - Insertion professionnelle des diplômés de master*

Taux d'emploi salarié en France des diplômés de master*, 6 à 30 mois après l'obtention du diplôme



Répartition des natures de contrat des diplômés de master* 18 mois après l'obtention du diplôme



* Hors master pour devenir enseignant.

** Titulaires de la fonction publique.

Source : Ministère de l'Enseignement supérieur et de la Recherche, sous-direction des systèmes d'information et des études statistiques (Sies). InserSup 2022.

d'indicateurs d'insertion des formations par établissement, la promotion précédente est cumulée avec celle de l'année en cours, dans le cas où l'effectif est inférieur à 20. Cela permet d'exposer 60 % des Licences Professionnelles et 70 % des Masters (*figure 4*). Cependant, un nombre important de cas pour lesquels l'effectif cumulé reste inférieur à 20 et pour lesquels aucun taux d'emploi salarié en France n'est affiché, demeure. Néanmoins, ces cas représentent une part marginale des étudiants inscrits dans l'ensemble des formations.

Au-delà de cet affichage, InserSup fournira au premier semestre 2025 des fichiers « détail » constitués de données individuelles, à destination des chargés d'études du système statistique public ou des chercheurs. Au Sies, ils permettent de mieux documenter les déterminants de l'insertion professionnelle en utilisant d'autres données disponibles : catégories socioprofessionnelles des parents, parcours des étudiants avant de quitter l'enseignement supérieur, mentions et choix de spécialités au baccalauréat, qualification en tant que boursier, etc.

► Un dispositif pérenne et sécurisé



L'automatisation est nécessaire pour sécuriser le processus, assurer une meilleure traçabilité des dépôts et retraits de fichiers et reproduire cette intégration de données.



La mise à disposition du grand public des indicateurs à un rythme semestriel engage le Sies sur une production permanente et régulière. Aussi, une automatisation du processus est nécessaire pour le pérenniser, le sécuriser (notamment face au risque de vol et d'interception des données), assurer une meilleure traçabilité des dépôts et retraits de fichiers et reproduire cette intégration de données. Cette exigence est d'autant plus forte qu'InserSup mobilise des données

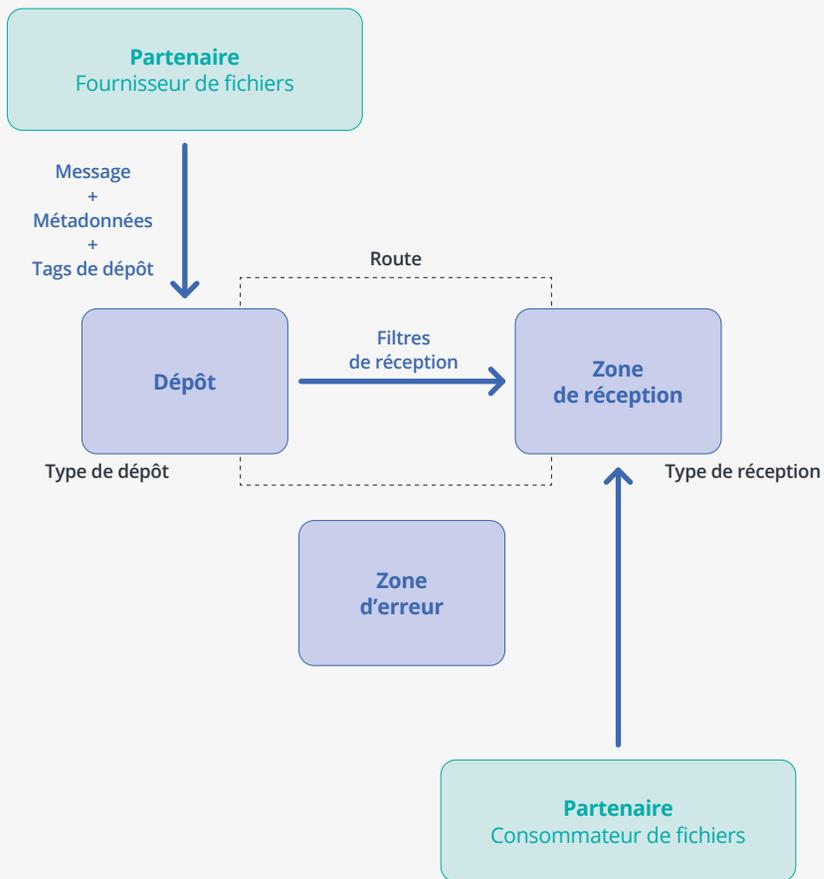
administratives individuelles dont certaines sont sensibles (rémunérations par exemple) et de surcroît en provenance de plusieurs services ministériels.

L'ensemble repose sur un système d'échange de fichiers mis en place dans les ministères de l'Éducation nationale et de l'Enseignement supérieur et de la Recherche qui gère le transport et la supervision des envois/réception de données, que celles-ci soient sous forme de fichiers ou soient l'objet d'une mise à disposition via une interface logicielle (API¹⁵). Le processus (*figure 5*) est le suivant :

1. Un chemin réseau indiquant que les fichiers partent du serveur A vers le serveur B est activé. Il associe un dépôt, une zone de réception et un filtre de réception localisés sur un serveur intermédiaire de la direction du numérique du ministère
2. Le producteur des données publie un message dans le dépôt en renseignant un tag de dépôt (bordereau d'envoi contenant l'identifiant de l'envoi) ;
3. Le message est délivré dans la zone de réception si l'identifiant du tag de dépôt est reconnu par le filtre de réception.

¹⁵ Une API (*application programming interface* ou « interface de programmation d'application ») 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.

► **Figure 5 - Le système d'échange de fichiers entre le Sies et la Dares**



Le consommateur dispose d'une ou de plusieurs zones de réception depuis lesquelles il récupère ses messages (un par un ou tous en une seule fois).

L'avancement du flux de données fait l'objet d'un traçage qui permet au producteur et au consommateur de suivre le cheminement. Les points de passage sont sauvegardés, notamment celui indiquant au producteur de fichier que son fichier est placé en zone d'erreur. Si le producteur est inconnu alors le consommateur ne reçoit pas les données. Le système a la possibilité de créer des rôles personnalisés pour les équipes intervenantes et d'y associer des habilitations selon les actions.

Dans le cadre de ce système d'échange, InserSup est principalement consommateur, mais il est aussi producteur dans l'échange avec la Dares ; il fournit à cette dernière la population d'intérêt avant qu'elle leur envoie les données associées de la DSN.

► Des difficultés particulières

Outre les spécificités inhérentes à toute élaboration d'un système d'information fondé sur des appariements de fichiers et de calculs d'indicateurs, concevoir InserSup nécessite de gérer des cas particuliers sur des situations « frontières » : adapter les changements de périmètre des établissements et ceux liés à la démographie des formations, traiter les situations d'emploi ou d'études à l'étranger.

Changements de périmètres des établissements et démographie des formations

Pour mettre à disposition les indicateurs, InserSup présente la particularité de reposer sur une nomenclature évolutive : une partie des formations proposées dans les établissements (mentions des diplômes) est modifiée chaque année dans certains établissements. En outre, ces derniers changent d'identifiants et/ou de contour à la faveur de fusions, d'apparitions ou de disparitions. Ces changements se sont accélérés au cours des années récentes depuis que le marché mondial de la formation se mesure à l'aune des classements internationaux. Pour intégrer ces classements et donner davantage de visibilité aux établissements français, une réorganisation du paysage universitaire est à l'œuvre. Elle vise à rassembler une masse critique suffisante en matière de recherche et se traduit par le regroupement d'établissements de l'enseignement supérieur (Guiselin, 2019 ; Musselin, 2017) : par exemple les universités avec les écoles ou les instituts. Les regroupements prennent la forme de communautés d'universités et d'établissements, d'établissements publics expérimentaux, ou de grands établissements qui disposent d'une autonomie administrative, pédagogique, scientifique et financière renforcée. Le statut d'établissement public expérimental offre davantage de latitude ; en effet, les établissements regroupés dans cet établissement peuvent conserver leur personnalité morale et sont nommés « établissements-composantes » de celui-ci.

En conséquence, InserSup, qui calcule et affiche des indicateurs d'insertion d'étudiants quittant l'enseignement supérieur au cours d'une année N et les expose en N+2 au niveau établissement, se heurte à une correspondance imparfaite avec les formations présentes sur les plateformes, en vue des inscriptions dans les formations en cours (année N+2).

Les changements d'identifiant des établissements, à contour inchangé, sont gérés par un système qui retrace l'historique des identifiants d'un établissement donné et qui les relie à un établissement (identifiant « maître »). Les créations/disparitions de formations se traduisent par une absence de données dans InserSup, au moment de l'appariement des données avec celles des plateformes. En outre, les composantes d'établissements se rassemblant en vue de former des ensembles plus grands, tout en souhaitant conserver leur autonomie et un affichage des indicateurs de leur périmètre, ne facilitent pas la tâche du statisticien.

Le cas des étudiants en formation et en emploi à la fois

InserSup vise à mesurer l'insertion professionnelle de personnes quittant l'enseignement supérieur qui délaissent donc leur statut étudiant pour un autre statut à établir au regard de leur situation sur le marché du travail. Les différentes situations possibles pour ce

nouveau statut sont : occuper un emploi, être demandeur d'emploi ou être en situation d'inactivité. Dans ce cadre, toute personne est soit en poursuite d'études après le diplôme pour lequel elle a été repérée (même si éventuellement elle occupe un job étudiant), soit sur le marché du travail, ou inactive, mais pas à la fois étudiante et sur le marché du travail. Or des situations, à la frontière de l'emploi et de la formation, posent question pour InserSup pour le repérage de la poursuite d'études : l'alternance et les doctorants.

Ces situations sont adossées à un contrat de travail rémunéré. Ces cas ont fait l'objet d'une réflexion qui a débouché sur la décision que la période d'apprentissage et la préparation d'une thèse sont considérées comme une poursuite d'études et non comme un emploi. En revanche le contrat de professionnalisation, qui s'apparente davantage à de la formation continue, n'a pas été considéré comme tel. Ces choix ont notamment été dictés pour assurer une cohérence avec les enquêtes existantes sur l'insertion professionnelle et le dispositif InserJeunes.

Les situations d'emploi ou d'études à l'étranger

L'insertion professionnelle des étudiants quittant l'enseignement supérieur ne s'effectue pas uniquement sur le territoire français. Certains d'entre eux occupent un emploi hors de France. C'est le cas, par exemple, de 7 % des diplômés issus de master en 2022 dont un peu plus de la moitié dans l'Union européenne. D'autres (1 % de la promotion 2022 de diplômés) poursuivent leurs études à l'étranger. L'utilisation de fichiers administratifs couvrant exclusivement le territoire français pour mesurer l'ensemble du champ de l'insertion professionnelle constitue donc une limite.



Pour compléter le panorama couvert par InserSup, les enquêtes d'insertion professionnelle (IP) captent l'information sur les diplômés en emploi à l'étranger.



Pour compléter le panorama couvert par InserSup, les enquêtes d'insertion professionnelle (IP) captent l'information sur les diplômés en emploi à l'étranger. Ces enquêtes deviennent complémentaires au dispositif InserSup. Cet ensemble constitué de données administratives et d'une enquête permet des mesures et des analyses couvrant l'ensemble de l'insertion professionnelle (Bachbauer et alii, 2022). Pour consolider le tout, le recours aux fichiers issus du Répertoire Statistique des Individus et des Logements (Résil, (Lefebvre, 2024)) pour repérer les étudiants quittant l'enseignement supérieur ne

résidant pas en France est envisagé. Il permettrait un cadrage quantifié de la population de sortants de l'enseignement supérieur en emploi ou qui poursuivent leurs études à l'étranger et par différence, du nombre de demandeurs d'emploi et d'inactifs.

Tous les pays recourant aux fichiers administratifs pour la mesure de l'insertion professionnelle sont confrontés à la difficulté de ne pouvoir repérer, dans les fichiers administratifs nationaux, les emplois occupés à l'étranger. La Commission européenne a entrepris depuis quelques années d'homogénéiser, au sein des États membres, les pratiques de recueil de l'information en pilotant, dans un premier temps, la conduite d'enquêtes sur le suivi des diplômés (Commission européenne, 2020). Elle réfléchit désormais à une initiative permettant l'échange de données entre États membres sur l'emploi de diplômés ressortissant d'un autre État membre.

► Une comparabilité améliorée



InserSup assure une comparabilité totale, dans le temps et entre établissements, qui n'existait pas.



En appliquant exactement le même traitement des données à tous les diplômés pour tous les établissements de formation, InserSup assure une comparabilité totale, dans le temps et entre établissements, qui n'existait pas parfaitement à partir des enquêtes d'insertion professionnelle. C'est un progrès par rapport à la situation précédente où chaque établissement conduisait sa propre enquête auprès de « ses » étudiants diplômés. Ces enquêtes

sont coordonnées par le ministère et adossées à un questionnaire commun, mais laissent la place à des conduites différentes. De plus, le Sies agit en producteur neutre d'indicateurs au regard des résultats produits pour chaque établissement, car il n'a aucun intérêt lié à la valeur des indicateurs. Cette neutralité est importante pour les utilisateurs, qu'il s'agisse du grand public ou encore des acteurs du monde de l'enseignement supérieur.

► Une couverture complète à terme

À partir de décembre 2024, InserSup élargit son champ pour couvrir les écoles d'ingénieurs et les écoles de commerce et de gestion, apanage des conférences des grandes écoles via leurs enquêtes. Le périmètre d'InserSup est appelé à s'agrandir au cours des deux prochaines années, par extension à l'ensemble des diplômés et établissements de formation du supérieur, qu'ils relèvent du champ de l'université ou de celui des écoles. Exposer des indicateurs, largement utilisés et scrutés, tous les semestres pour près de 7 000 sous-populations d'étudiants quittant l'enseignement supérieur, exige une automatisation rodée, une sécurité informatique solide, une organisation efficace de la production et une coordination avec les services producteurs de données sans faille et pérenne. Le dispositif apporte une information riche sur l'avenir professionnel des étudiants et relève de l'intérêt public (Fack et Huillery, 2021). Le grand public peut d'ores et déjà consulter les indicateurs produits sur les plateformes Parcoursup, Mon Master mais aussi sur SupTracker¹⁶, l'Étudiant.fr¹⁷, Onisep.

De nombreux autres acteurs s'en emparent également : les professionnels de l'orientation, l'Association pour l'emploi des cadres (Apec), les branches professionnelles, les politiques, le ministère dans le cadre des contrats d'objectifs, de moyens et de performance¹⁸ et pour le pilotage de l'offre de formation. Le monde de la recherche, notamment pour des travaux sur l'efficacité des formations de l'enseignement supérieur, attend également les fichiers détails que le dispositif produira annuellement à compter de 2025.

InserSup remplit déjà les objectifs assignés en 2022, sur un périmètre plus large que celui que la statistique publique couvrait jusqu'à maintenant à travers ses enquêtes d'insertion professionnelle. Celles-ci doivent maintenant être revues en se limitant à une stricte complémentarité avec les informations que les fichiers administratifs fournissent.

¹⁶ SupTracker est un outil de visualisation des données en open data issues des plateformes Parcoursup et Mon Master. <https://beta.suptracker.org/>.

¹⁷ <https://www.letudiant.fr/>.

¹⁸ Contrat d'établissement sur lequel s'applique le financement à la performance.

► Fondements juridiques

- LOI n° 2007-1199 du 10 août 2007 relative aux libertés et responsabilités des universités. In : *site de Légifrance*. [en ligne]. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT00000824315>.
- LOI n° 2018-771 du 5 septembre 2018 pour la liberté de choisir son avenir professionnel. In : *site de Légifrance*. [en ligne]. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037367660>.
- Arrêté du 30 juillet 2018 portant création par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation d'un traitement automatisé de données à caractère personnel dénommé « Système d'information sur le suivi des étudiants » (SISE). In : *site de Légifrance*. [en ligne]. Mise à jour le 22 juin 2023. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000037359616/>.

► Bibliographie

- AUBRY, Antonin, BAH, Souleymane et HERZBERG, Kendrick, 2024. Le taux d'emploi salarié en France des diplômés en 2022 de licence professionnelle et de master à 6, 12 et 18 mois. In : *Note Flash*. [en ligne]. Juillet 2024. SIES. N° 20. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2024-07/nf-sies-2024-20-33923.pdf>.
- AUBRY, Antonin, BAH, Souleymane et HERZBERG, Kendrick, 2024. Le taux d'emploi salarié en France des diplômés en 2022 de licence générale à 6, 12 et 18 mois. In : *Note Flash*. [en ligne]. Juillet 2024. SIES. N° 21. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2024-07/nf-sies-2024-21-33926.pdf>.
- BACHBAUER, Nadine, WOLF, Clara et FUSS, Daniel, 2022. Education and Employment Trajectories in NEPS-ADIAB: The Survey Data of the National Educational Panel Study Linked to Administrative Data of the Institute for Employment Research. In : *European Sociological Review*. [en ligne]. Août 2022. Vol. 38, n° 4, pp. 663-676. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://academic.oup.com/esr/article/38/4/663/6486883>.
- BÉNICHOU, Yves-Laurent, ESPINASSE, Lionel et GILLES, Séverine, 2023. Le code statistique non signifiant (CSNS) : un service pour faciliter les appariements de fichiers. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 64-85. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635825?sommaire=7635842>.
- CARON, Nathalie, MIDY, Loïc, 2022. Enjeux statistiques du système d'information Inserjeunes sur l'insertion professionnelle. In : *Journées de méthodologie statistique de l'Insee 2022*. [en ligne]. DEPP. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://journées-methodologie-statistique.insee.net/enjeux-statistiques-du-systeme-dinformation-inserjeunes-sur-linsertion-professionnelle/>.
- DIRECTION GÉNÉRALE DE L'ÉDUCATION, DE LA JEUNESSE, DU SPORT ET DE LA CULTURE (COMMISSION EUROPÉENNE), ICF CONSULTING, BEADLE, Shane, VALE, Patricia, HANNAH, Arthur et ZAIDI, Ali, 2020. Graduate tracking: a 'how to do it well' guide. In : *Office des publications de l'Union européenne*. [en ligne]. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://data.europa.eu/doi/10.2766/263936>.
- FACK, Gabrielle et HUILLERY, Élise, 2021. Enseignement supérieur : pour un investissement plus juste et plus efficace. In : *Les notes du conseil d'analyse économique*. [en ligne]. Décembre 2021. N° 68. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/staticfiles/pdf/cae-note068.pdf>.
- GUISELIN, Emmanuel-Pie, 2019. Les regroupements d'établissements dans l'enseignement supérieur et la recherche : enjeux politiques et cadrage juridique. In : *Revue française d'administration publique*. [en ligne]. 2019/1, n° 169, pp. 37-50. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://doi.org/10.3917/rfap.169.0037>.

- 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 7 octobre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8203044?sommaire=8203072>.
- 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 7 octobre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8203040?sommaire=8203072>.
- MUSSELIN, Christine, 2017. *La Grande Course des universités*. Presses de Sciences Po.
- REDOR, Patrick, 2023. Confidentialité des données statistiques : un enjeu majeur pour le service statistique public. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 46-63. [Consulté le 7 octobre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635823?sommaire=7635842>.

Des chiffres pour la culture

60 ans d'objectivation scientifique au service des politiques culturelles



Amandine Schreiber* et Laure Turner**

Dès le début des années 60, dans un contexte de développement sans précédent du temps de loisirs et pour répondre aux enjeux d'observation liés à la planification économique, le ministère de la Culture se dote d'un service des études et recherches. Produire des statistiques sur la culture s'avère tout autant nécessaire que complexe, du fait de l'étendue du champ culturel, de sa diversité et du renouvellement constant des offres comme des pratiques.

L'enquête phare du service sur les pratiques culturelles des Français, menée pour la première fois en 1973 et étendue aux territoires ultra-marins lors de sa dernière édition en 2018, est conçue pour témoigner sur longue période de l'évolution de pratiques sans cesse diversifiées par les innovations technologiques successives et les évolutions sociétales.

La diversité des acteurs, publics comme privés, et des professionnels aux statuts spécifiques, constitue autant d'autres défis à relever pour le statisticien. Afin de mesurer l'activité économique de ces acteurs et de suivre les professionnels de la culture, un cadre conceptuel harmonisé au niveau européen et des nomenclatures concertées sur le plan international, qu'il convient de faire régulièrement évoluer, ont été élaborées.

 As soon as the early 1960s, in a context of unprecedented development of leisure time and to meet the observation challenges related to economic planning, the Ministry of Culture set up a study and research department.

Producing statistics on culture is both necessary and complex, given the breadth of the cultural field, its diversity and the constant renewal of cultural offer and practices.

The department's flagship survey on the French cultural practices, first conducted in 1973 and extended to overseas territories during its latest edition in 2018, is designed to testify over a long period of time to the evolution of practices constantly diversified by successive technological innovations and societal changes.

The diversity of public and private actors and that of professionals with specific statutes is another challenge for the statistician. In order to measure the economic activity of these actors and to report on the characteristics of cultural professions, a harmonised conceptual framework at European level and internationally agreed nomenclatures, which need to be regularly updated, have been developed.

* Cheffe du Département des études, de la prospective, des statistiques et de la documentation (DEPS), ministère de la Culture.
amandine.schreiber@culture.gouv.fr

** Adjointe à la cheffe du Département des études, de la prospective, des statistiques et de la documentation (DEPS), ministère de la Culture.
laure.turner@culture.gouv.fr

Télévision, radio, presse, livres, écoute de musique enregistrée, films, séries ou encore musées, théâtres, concerts, festivals mais aussi pratiques en amateur : qu'elle se pratique au quotidien ou qu'elle sublime nos temps de loisirs, la culture occupe une place essentielle dans notre société.

L'étendue du champ culturel, la diversité des offres culturelles comme des pratiques et leur évolution rapide au gré des changements sociétaux et des innovations technologiques soulèvent de nombreux défis pour le statisticien. Cet article décrit les enjeux relevés depuis plusieurs décennies.

► Planification économique et sociale et premières statistiques culturelles



Dans le contexte de croissance sans précédent des Trente Glorieuses... une attention inédite est portée aux pratiques et aux consommations culturelles.



La nécessité d'établir des statistiques sur la culture et de promouvoir des travaux de recherche en sciences sociales s'impose rapidement après la création du ministère des Affaires culturelles, confié à André Malraux en 1959. L'impulsion est donnée par la Commission de l'équipement culturel et du patrimoine artistique du IV^e Plan, créée en 1961, qui déplore l'absence de données statistiques à des fins de planification¹. Dans le contexte de croissance sans précédent des Trente Glorieuses, d'augmentation du niveau de vie et du pouvoir d'achat, de généralisation de l'instruction

et d'augmentation du temps de loisirs, une attention inédite est portée aux pratiques et aux consommations culturelles. Les pouvoirs publics doivent répondre aux nouveaux besoins de la population en matière de loisirs et de vie culturelle ; la culture devient ainsi une catégorie d'intervention publique à part entière (Détrez, 2020 ; Dubois, 1999).

► Une culture du chiffre pour mettre fin à « l'ère des goûts et des couleurs »

Produire des chiffres pour la culture n'avait pourtant rien d'évident et nécessitait une petite révolution, tant sur le plan administratif que sur le plan scientifique et conceptuel. En 1963, une cellule préfiguratrice du Service des études et recherches (SER) est créée au ministère des Affaires culturelles et confiée à Augustin Girard² (DEPS, 2010), qui l'administrera pendant trente ans, sous ses diverses appellations administratives³ (Martin, 2012), et en pilotera les travaux d'études et de recherche (**encadré 1**). Il prône de « quantifier le domaine culturel » à une époque où certains perçoivent cette entreprise de quantification comme anticulturelle

¹ La planification de la France, débutée en 1946 à l'initiative de Jean Monnet, prévoyait plusieurs plans de modernisation successifs.

² Il dirige le SER jusqu'à son départ à la retraite en 1993, date à laquelle il crée le Comité d'histoire de la culture dont il est le président jusqu'en 2007.

³ Département des études et de la prospective (DEP) en 1986, puis des statistiques (DEPS) en 2004, et de la documentation depuis 2021. Il devient service statistique ministériel (SSM) de la Culture en 1995.

► Encadré 1. Des activités d'études et de recherche au service du développement culturel



En octobre 1969, est publié le premier numéro de *Développement culturel*, bulletin mensuel du Service des études et recherches (SER) diffusé à tous les services du ministère et aux membres de la Commission des affaires culturelles du Plan.

Le titre de ce bulletin, qui laissera la place en 2007 à quatre nouveaux titres (Culture études, Culture chiffres, Culture prospective et Culture méthodes) témoigne de l'ambition du chef du service de l'époque, Augustin Girard, d'influencer directement – en les éclairant le mieux possible – les décisions prises en matière de politique culturelle.

Parmi les orientations suggérées dans les diverses études et rapports du SER, l'une d'elles reste emblématique : les résultats de travaux inédits sur les pratiques amateurs sont à l'origine de la création, par Jack Lang, de la première « fête de la musique » en juin 1982.

Premier numéro du bulletin périodique « Développement culturel » du Service des études et recherches.

Source : Archives DEPS.

par essence (Girard, 1965) et jugent sacrilège de chercher à mathématiser les « mystères sacrés de l'art ». À ses détracteurs, il répond que ce n'est pas la culture qu'il s'agit de mesurer, mais ses moyens, afin d'établir un fondement objectif et durable à l'affectation des crédits culturels publics et mettre ainsi fin « à l'ère des goûts et des couleurs ». Cette expression fait référence aux périodes antérieures au cours desquelles l'octroi des crédits dépendait avant tout des sensibilités personnelles et du bon vouloir des décideurs politiques successifs.

► Des premières enquêtes auprès des publics aux enquêtes sur les pratiques culturelles des Français

Dans cette perspective, il s'agit de recenser les offres (équipements, financements, etc.) mais aussi d'analyser les pratiques. Le Service des études et recherches (SER) du ministère des Affaires culturelles commande une série d'enquêtes dans les années 1960, dont celles menées par Alain Darbel – administrateur de l'Insee – et Pierre Bourdieu – sociologue – auprès des publics des musées⁴. Leur analyse des déterminants sociaux de « l'amour de l'art » (Bourdieu et Darbel, 1966), puis celle sur le rôle de la culture dans les mécanismes de domination et de reproduction sociale (Bourdieu, 1979), posent les bases des problématiques traitées par la suite par les sociologues de la culture pour la France (Lahire, 2004, Coulangeon, 2005 et 2011, Glevarec, 2013), dans un contexte, toujours d'actualité, d'idéal politique de démocratisation culturelle.

⁴ Le SER contribue à la naissance puis à la structuration de tout un pan de la discipline sociologique dédié à la culture.

Après plusieurs enquêtes menées sur les publics (des musées mais aussi des théâtres, des salles de cinéma ou des maisons de la culture) et dans la foulée de l'enquête menée en 1967 par l'Insee sur les loisirs, la première grande enquête Pratiques culturelles est réalisée par le SER en 1973. Elle est reconduite environ tous les dix ans, en 1981, 1988, 1997, 2008 et 2018.

Les six éditions de cette enquête décennale constituent un dispositif d'observation unique en France sur les pratiques culturelles. L'enquête aborde le sujet au niveau national et de façon transversale sur l'ensemble du champ (spectacle vivant, industries culturelles, patrimoines, etc.). Elle permet de décrire les différentes formes de participation à la vie culturelle : consommations culturelles (télévision, radio, écoute de musique, lecture de livres, jeux vidéo, films, vidéos, etc.), fréquentation des équipements culturels (musées, cinémas, théâtres, bibliothèques, etc.) ou des manifestations culturelles (festivals, spectacles de rue, etc.) et pratiques en amateur (musique, théâtre, danse, arts plastiques, etc.).

La sixième et dernière édition, menée en 2018, bénéficie d'innovations méthodologiques et de moyens étendus. L'échantillon est tiré par l'Insee. Le nombre de personnes interrogées est multiplié par deux et atteint 9 000 personnes de 15 ans et plus en France métropolitaine. L'enquête est conduite pour la première fois dans les cinq départements et régions d'Outre-mer : Guadeloupe, Guyane, Martinique, La Réunion et Mayotte. Le questionnaire a évolué, en particulier depuis l'édition de 2008 : une place importante est accordée aux nouvelles technologies liées au développement du numérique. Mais le dispositif est resté suffisamment stable, tant sur le plan méthodologique que dans la formulation des questions, pour mener des analyses longitudinales et suivre l'évolution des comportements culturels des Français au fil des générations.

► Un demi-siècle d'observation des pratiques culturelles des Français



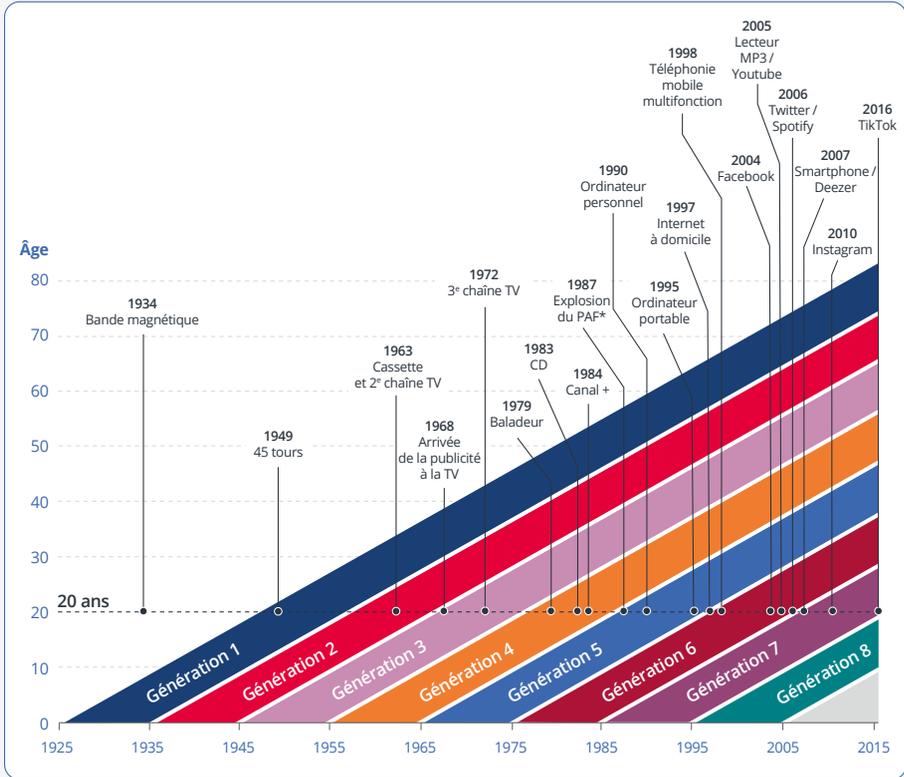
Parmi l'ensemble des pratiques analysées et pour ne donner qu'un seul exemple, l'écoute de musique hors radio est celle qui a connu le développement historique le plus dynamique.



Dans la lignée des travaux menés par Olivier Donnat (Donnat et Lévy, 2007) ont été mises en lumière les évolutions des pratiques culturelles observées de 1973 à 2018 (Lombardo et Wolff, 2020). Parmi l'ensemble des pratiques analysées et pour ne donner qu'un seul exemple, l'écoute de musique hors radio est celle qui a connu le développement historique le plus dynamique, porté notamment par l'évolution des technologies, de l'arrivée des chaînes hi-fi dans les ménages aux baladeurs et jusqu'aux smartphones, désormais premier terminal

culturel nomade (**figure 1**). Parmi les personnes âgées de 15 ans et plus résidant en France métropolitaine, 81 % ont écouté de la musique au cours des douze derniers mois en 2018 contre 66 % en 1973. L'écoute quotidienne de musique s'est tout particulièrement développée : en 2018, cela concerne 57 % des personnes contre 34 % en 2008 et seulement 9 % en 1973.

► **Figure 1 - Principaux effets d'offre dans le domaine audiovisuel**



* PAF : Paysage Audiovisuel Français.

Note : Les bandeaux de couleur renvoient aux différentes générations retenues pour mener l'approche générationnelle.

Lecture : La génération 1 née entre 1925 et 1934 avait environ une vingtaine d'années quand est diffusé le 45 tours à la fin des années 1940.

Source : Bipe/DEPS, ministère de la Culture et de la Communication, 2011. Figure actualisée extraite de Donnat, 2011.

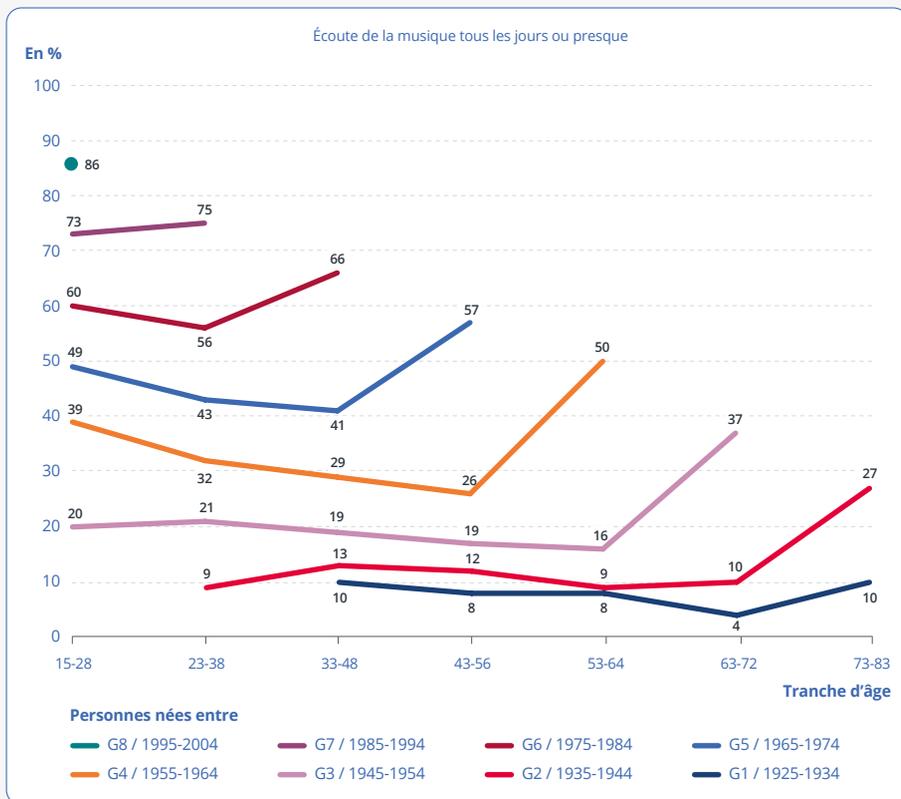
► Analyser les effets d'offre, d'époque et de génération

Analyser l'évolution des pratiques culturelles sur longue période nécessite de prendre en compte les transformations structurelles profondes de la société française depuis les années 1970 (élévation du niveau de formation, féminisation de l'emploi, tertiarisation de l'économie française, vieillissement de la population, etc.) et celles qui ont concerné les conditions d'accès à la culture. Ces pratiques ont été profondément modifiées sous l'effet des politiques publiques d'aménagement culturel du territoire et d'enrichissement de l'offre, et grâce aux innovations technologiques successives qui ont révolutionné l'accès aux divers contenus culturels (images, musiques, textes, vidéos).

Parmi les effets d'offre qui influent sur les pratiques culturelles, certaines peuvent être qualifiées d'« effets d'époque », car concernant indifféremment toutes les catégories de la population, tandis que d'autres relèvent d'« effets de génération ». Dans les faits, il est rare que toutes les catégories de la population s'emparent de la même façon des différentes innovations : les plus jeunes sont souvent les plus prompts à s'approprier les nouveautés, tant technologiques qu'artistiques⁵, et la majorité d'entre eux conservent en général tout au long de leur vie les habitudes prises pendant leur jeunesse (Donnat, 2011). Ainsi, l'approche générationnelle adoptée pour l'analyse de la série des enquêtes Pratiques culturelles se révèle être particulièrement pertinente.

Pour reprendre l'exemple de l'écoute quotidienne de musique, la diffusion massive de cette pratique s'appuie sur une dynamique générationnelle à l'œuvre depuis les années 1970 : au fil des éditions successives de l'enquête, chaque génération se distingue de la précédente par un taux d'écoute, à tout âge, systématiquement supérieur à celui de la

► **Figure 2 - Cinquante ans de pratiques culturelles en France**



Lecture : Entre 15 et 28 ans, 86 % de la génération née entre 1995 et 2004 écoutaient de la musique quotidiennement alors que 73 % de la génération née entre 1985 et 1994 le faisaient au même âge. Source : Enquête sur les pratiques culturelles, 1973-2018, DEPS, ministère de la Culture, 2020. Figure extraite de Wolff et Lombardo, 2020.

5 Nouveaux genres musicaux ou nouvelles technologies de l'information et de la communication.

génération précédente au même âge (*figure 2*). Elle s'appuie également sur les effets du développement des usages numériques et tout particulièrement de la consommation de musique dématérialisée qui contribue à la nette progression d'écoute de musique entre 2008 et 2018, notable pour toutes les générations, y compris les plus âgées (près de 20 points d'augmentation des taux d'écoute quotidienne pour les générations nées entre 1935 et 1965). En 2018, 36 % des personnes âgées de 15 ans et plus écoutent de la musique en ligne, en flux (aussi couramment appelé « streaming »). Les plus jeunes y ont le plus recours (73 % des 15-24 ans), sans pour autant que les moins jeunes en soient exclus : plus d'un tiers des 40-59 ans (34 %) utilisent ces technologies, et 12 % des plus de 60 ans.

► Des points intermédiaires d'observation pour un suivi conjoncturel

La série des enquêtes Pratiques culturelles constitue le principal instrument d'observation de l'évolution structurelle des pratiques, et les dynamiques générationnelles qu'elle permet d'étudier lui confère des vertus prospectives particulièrement utiles. Elle ne permet toutefois pas de répondre aux questions plus conjoncturelles auxquelles les décideurs publics sont souvent confrontés. Des épisodes inédits comme celui de la crise de Covid-19, ou de façon générale, le rythme sans cesse plus rapide de diffusion des innovations technologiques imposent de mettre en place des dispositifs d'observation complémentaires (*encadré 2*).

► Encadré 2. Le suivi des pratiques culturelles depuis la crise de Covid-19

Lorsqu'en mars 2020 la crise sanitaire liée à la pandémie de Covid-19 vient bousculer le quotidien des Français, le service statistique du ministère de la Culture met en place plusieurs dispositifs d'enquêtes en ligne, selon la méthode des quotas.

Du 20 avril au 4 mai 2020, une enquête en ligne sur les pratiques culturelles en temps de confinement est confiée au CRÉDOC* et réalisée auprès d'un échantillon de 3 000 personnes, représentatif des Français âgés de 15 ans et plus. Elle donne lieu à une description détaillée de la façon dont les Français occupent leur temps de loisirs pendant cette période inédite (Jonchery et Lombardo, 2020).

Après deux périodes de fermeture des établissements culturels (mi-mars à juin 2020 puis novembre 2020 à mai 2021) et dans un contexte de réouverture progressive des lieux, limitée par des protocoles spécifiques (jauges, respect des

mesures de couvre-feux, réservations obligatoires, distanciation entre visiteurs ou spectateurs, port du masque obligatoire puis mise en place du pass sanitaire à partir du 21 juillet 2021), une enquête flash est réalisée fin août 2021 par l'institut Harris Interactive afin de connaître l'impact de la crise sanitaire sur les sorties culturelles. En janvier 2022, puis de nouveau en octobre 2023, un ensemble de questions sur les sorties culturelles est inséré dans l'enquête Conditions de vie et aspirations du CRÉDOC (Müller et Schreiber, 2022).

Appelée à être renouvelée au cours des prochaines années, cette collaboration avec le CRÉDOC préfigure la mise en place d'un baromètre annuel sur les pratiques culturelles – à l'image du baromètre de l'Institut national de la jeunesse et de l'éducation populaire (Injep) sur les pratiques sportives (Vicard, 2023).

* Le CRÉDOC, Centre de Recherche pour l'Étude et l'Observation des Conditions de Vie, est un organisme d'études et de recherche au service des acteurs de la vie économique et sociale.

La multiplication des dispositifs d'observation soulève des questions. Conçus pour répondre à des besoins différents, les résultats obtenus sont difficilement comparables entre eux dès lors que les méthodes et protocoles d'enquêtes diffèrent (modes de collecte, période de référence et calendrier de l'enquête, taille des échantillons,

méthode probabiliste ou méthode des quotas⁶, formulation des questions, etc.). Il est d'autant plus nécessaire de faire preuve de pédagogie et de transparence sur les sources et les méthodes, que le champ culturel se caractérise par une grande multiplicité d'acteurs, commanditaires de sondages ou producteurs de données administratives (Autorité de la statistique publique⁷, 2024).

► Une multitude de données mobilisables, une multitude de producteurs de statistiques

En complément des données d'enquêtes, un grand nombre de données administratives ou de données privées (**encadré 3**) produites par différents acteurs peut être mobilisé. Le ministère de la Culture compte en particulier près de 80 établissements sous tutelle dont la grande diversité de champs d'action reflète l'étendue du champ culturel lui-même : musées nationaux, opéras et théâtres nationaux, bibliothèque nationale, écoles nationales d'art et d'architecture, conservatoires nationaux, etc. Si la majorité de ces organismes produisent des données essentiellement pour piloter leur activité, d'autres exercent une mission d'animation de leur filière et d'observation (le Centre national du cinéma et de l'image animée (CNC), le Centre national de la musique (CNM) ou le Centre national du livre (CNL) notamment).

► Encadré 3. L'exploitation des données privées

La digitalisation à la fois de certains produits culturels et de leurs canaux de vente, ou encore la généralisation des smartphones qui permettent la localisation des publics, ouvrent des perspectives pour le suivi statistique de la consommation culturelle comme de la fréquentation des institutions culturelles.

Par exemple, les données privées du panel de la société GfK^{*}, issues des sorties de caisse hebdomadaires de plus de 4 000 distributeurs (points de vente physiques et numériques), permettent d'analyser de façon précise le volume

des ventes de produits culturels sur l'ensemble du territoire français pour quatre marchés : jeux vidéo physiques pour console et ordinateur personnel, livres physiques, musique enregistrée physique et numérique, vidéos physiques et vidéos à la demande. Le DEPS détient ces bases sur une quinzaine d'années. Plusieurs projets de recherche exploitant ces données ont été lancés et des études menées, en particulier sur la diversité de l'offre du marché du livre ou de la musique (Donnat, 2018) ou du marché de la musique en streaming (Maillard, 2023).

* GfK (« Growth from Knowledge ») est une société de conseil spécialisée dans la collecte et l'analyse de données, le suivi de panel de distributeurs et les études de marché.

Organismes de gestion collective, syndicats professionnels ou associations sont également nombreux à produire des données et à publier des informations sur leurs champs respectifs d'intervention, les statistiques publiques nationales ne les reflétant qu'imparfaitement. Cartographier l'ensemble de ces acteurs et proposer des modalités de coordination de leurs productions statistiques constituent un vaste chantier. Celui-ci doit être mené en parallèle de l'exploitation des sources de la statistique publique, à laquelle le service statistique ministériel (SSM) s'attelle dans un cadre conceptuel harmonisé au niveau européen.

6 La méthode des quotas, moins performante d'un point de vue statistique car plus sujette aux biais de sélection, est moins coûteuse et plus rapide à implémenter que la méthode probabiliste, systématiquement privilégiée par la Statistique publique (Ardilly, 2023).

7 L'Autorité de la Statistique Publique veille au principe d'indépendance professionnelle dans la conception, la production et la diffusion de statistiques publiques et assure une vigilance quant à la qualité des statistiques publiques, que ce soit en termes de rigueur méthodologique ou de respect de différents principes déontologiques.

► Une définition européenne du champ statistique de la culture

Si la production de statistiques culturelles n'est soumise à aucun règlement européen, les coopérations internationales et européennes sont anciennes⁸ et les travaux réguliers pour définir un champ culturel harmonisé au niveau européen et le faire évoluer. Eurostat⁹ initie dès 1997 plusieurs travaux auxquels le SSM de la Culture contribue activement¹⁰. En septembre 2009, le réseau ESSnet¹¹ « Statistiques culturelles », financé par la Commission européenne, est créé à l'initiative d'Eurostat pour une période de deux ans, afin d'améliorer la connaissance des secteurs culturels et de répondre aux demandes croissantes de données comparables au niveau européen. Le SSM de la Culture y partage son expertise nationale en pilotant le travail consacré au cadre conceptuel des statistiques de la culture (Deroin, 2011).

Le cadre défini... repose sur l'identification de dix domaines culturels... puis sur le croisement des domaines culturels avec six fonctions économiques du système productif.

Les activités culturelles relèvent de plusieurs secteurs économiques (industrie, services, information, communication, arts et spectacle, commerce, enseignement, etc.). Elles sont par ailleurs liées à diverses pratiques reconnues comme culturelles selon des conventions sociales qui elles-mêmes évoluent. Il est dès lors nécessaire de définir un cadre conceptuel sur lequel s'appuyer pour déterminer l'appartenance d'une activité au champ culturel. Le cadre défini par le groupe ESSnet (ESSnet-Culture,

2012), dans la continuité des travaux européens et de l'approche retenue par le service statistique français, repose sur l'identification de dix domaines culturels : le patrimoine culturel, les archives, les bibliothèques, le livre et la presse, les arts visuels (arts plastiques, photographie et design), l'artisanat d'art, le spectacle vivant (théâtre, danse, musique, autre), l'audiovisuel (radio, télévision, film, vidéo, musique enregistrée) et le multimédia, l'architecture, la publicité. Puis, le croisement des domaines culturels avec six fonctions économiques du système productif (création, production, diffusion et commercialisation¹², préservation, éducation, administration et management culturels), permet de définir les activités culturelles (*figure 3*).

8 Membre de la Délégation française à la Conférence générale de l'UNESCO (Organisation des Nations unies pour l'éducation, la science et la culture), Augustin Girard est par la suite directeur du programme « Développement culturel » du Conseil de l'Europe et membre de la Délégation française aux conférences internationales des ministres de la Culture.

9 Eurostat, l'Office statistique de l'Union européenne, est chargé de publier des statistiques et des indicateurs européens permettant d'effectuer des comparaisons entre les pays et les régions.

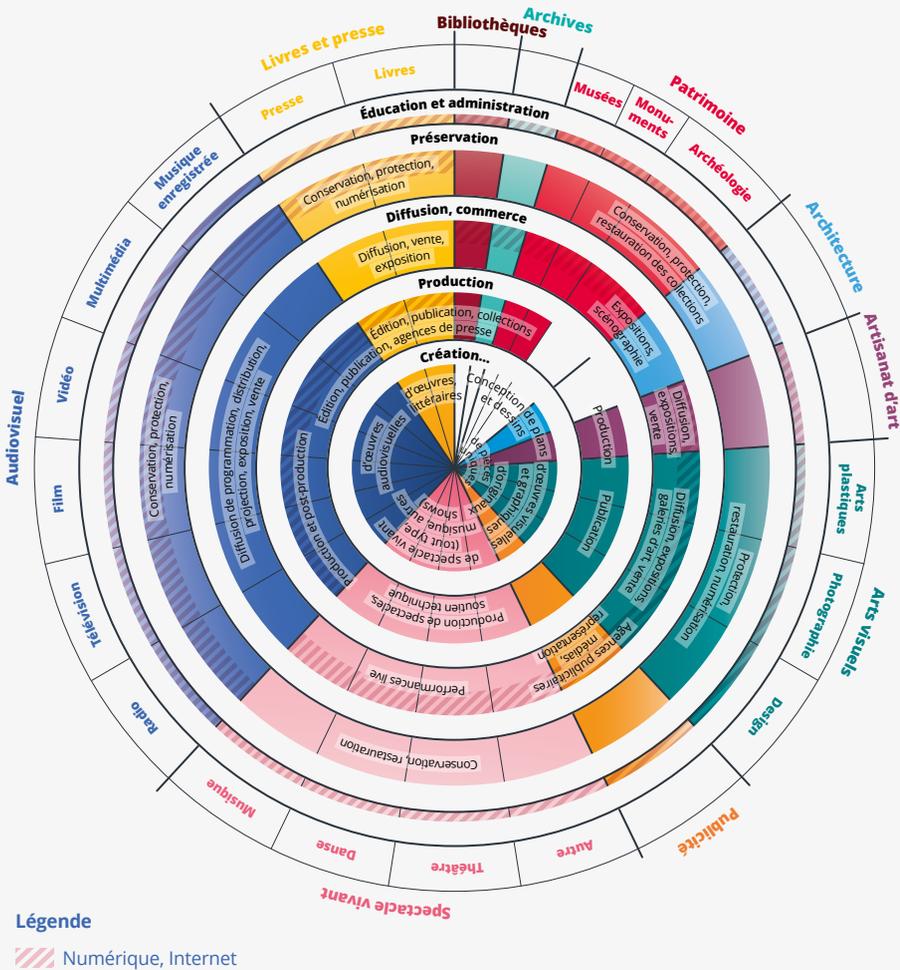
10 Membre du *Leadership Group* de 1997 à 2004, il participe à la mise en place en 2002 du Groupe européen sur les statistiques des musées (*European Group on MUseum Statistics – EGMUS*). Eurostat lui confie la réalisation de *Cultural Statistics*, premier ouvrage sur les statistiques culturelles harmonisées des 27 États membres diffusé en 2007.

11 ESSnet : *European Statistical System network*.

12 Augustin Girard milite en France dès la fin des années 1970 pour que ces fonctions de production/diffusion/commerce soient retenues dans le périmètre de la culture, insistant sur le rôle déterminant des industries culturelles (Girard, 1978).

Une fois ce cadre conceptuel posé, le champ statistique de la culture est défini dans une approche pragmatique, à partir des nomenclatures statistiques d'activités partagées au niveau international : la nomenclature des activités européenne (la NACE¹³), s'emboîte avec la nomenclature d'activités française (NAF¹⁴) plus détaillée, mais aussi la nomenclature internationale produite par l'ONU (la CITI : classification internationale type par industrie), plus agrégée. Chaque activité identifiée comme culturelle est ainsi reliée à un code de la NACE et de la NAF (*Tableau*).

► **Figure 3 - Mandala des activités économiques culturelles**



Source : Deroin, 2011.

¹³ La NACE est la Nomenclature statistique des Activités économiques dans la Communauté Européenne : <https://ec.europa.eu/eurostat/fr/web/products-manuals-and-guidelines/-/KS-RA-07-015>.

¹⁴ <https://www.insee.fr/fr/information/2406147>.

► **Tableau : Liste des activités culturelles selon la nomenclature statistique française d'activités (NAF Rév. 2, 2008)**

Domaine culturel	Code Niveau 4 (sous-classe)	Libellé
Patrimoine	91.01Z 91.02Z 91.03Z	<ul style="list-style-type: none"> • Gestion des bibliothèques et des archives • Gestion des musées • Gestion des sites et monuments historiques et des attractions touristiques similaires
Livres et presse	58.11Z 58.13Z 58.14Z 63.91Z	<ul style="list-style-type: none"> • Édition de livres • Édition de journaux • Édition de revues et périodiques • Activités des agences de presse
Arts visuels	90.03A 90.03B 74.20Z (p) 74.10Z	<ul style="list-style-type: none"> • Création artistique relevant des arts plastiques • Autre création artistique • Activités photographiques • Activités spécialisées de design
Architecture	71.11Z	<ul style="list-style-type: none"> • Activités d'architecture
Spectacle vivant	90.01Z 90.02Z 90.04Z	<ul style="list-style-type: none"> • Arts du spectacle vivant • Activités de soutien au spectacle vivant • Gestion de salle de spectacles
Audiovisuel/ Multimédia	59.11A 59.11B 59.11C 59.12Z 59.13A 59.13B 59.14Z 58.21Z 59.20Z 60.10Z 60.20A 60.20B	<ul style="list-style-type: none"> • Production de films et de programmes pour la télévision • Production de films institutionnels et publicitaires • Production de films pour le cinéma • Post-production de films cinématographiques, de vidéo et de programmes de télévision • Distribution de films cinématographiques • Édition et distribution vidéo • Projection de films cinématographiques • Édition de jeux électroniques • Enregistrement sonore et édition musicale • Édition et diffusion de programmes radio • Édition de chaînes généralistes • Édition de chaînes thématiques
Publicité	73.11Z (p)	<ul style="list-style-type: none"> • Activités des agences de publicité
Éducation/ Formation	85.52Z	<ul style="list-style-type: none"> • Enseignement culturel

(p) : activité partiellement culturelle.

Note : Les activités culturelles sont présentées ici selon la nomenclature statistique française d'activités en vigueur de 2008 à 2025 (la NAF rév. 2).

Source : DEPS, ministère de la Culture et de la Communication, 2011. Deroin, 2011.

► Des limites inhérentes à toute classification des activités

Les nomenclatures statistiques ne sont pas des nomenclatures dédiées à circonscrire des filières et peuvent donc agréger activités culturelles et activités non culturelles, de sorte que le champ statistique de la culture comporte des codes « partiellement culturels ». Par exemple, les activités photographiques et celles des agences de publicité sont prises en compte dans la définition européenne de la culture en vigueur alors qu'elles sont partiellement culturelles, englobant aussi des activités techniques non culturelles : les activités des laboratoires et du traitement de films pour la photographie et les activités de marketing pour la publicité.

Inversement, certaines activités culturelles sont classées dans des catégories plus larges d'activités majoritairement non culturelles et ne peuvent pas être prises en compte dans le champ de la culture. Par exemple, l'activité de fouilles archéologiques est classée avec la recherche en sciences humaines et sociales ; dans la NAF 2008, celle des plateformes d'écoute de musique en streaming est classée dans la programmation informatique avant sa refonte en 2025 qui la place désormais dans le champ culturel (code 59.20).

► Une limite liée à la spécificité de la création artistique et culturelle



Certains champs de la culture sont caractérisés non pas par le produit final réalisé, mais par la créativité mise en œuvre pour sa fabrication. Or les nomenclatures mettent l'accent sur le produit final et non pas sur le processus de fabrication.



Une autre limite de la définition du champ culturel à partir des nomenclatures d'activité, tient au rôle central de la création dans la définition du cadre conceptuel du champ culturel. Certains champs de la culture sont caractérisés non pas par le produit final réalisé, mais par la créativité mise en œuvre pour sa fabrication. Or les nomenclatures mettent l'accent sur le produit final et non pas sur le processus de fabrication. Ainsi, un bien issu de la fabrication du verre, qu'il soit réalisé dans une démarche artistique par un maître-verrier ou de manière industrielle par une grande entreprise, est classé dans le même code d'activité de fabrication du verre. L'opérabilité

d'une nomenclature rend nécessaire cette approche : au moment du classement des unités économiques au sein de la nomenclature d'activité, l'évaluation de la créativité de leur production ne peut être établie ni contrôlée par les statisticiens. Par conséquent, à moins de disposer d'un code de la nomenclature dédié à leur activité, certains champs culturels ne peuvent pas être identifiés comme tels. Pour la même raison, le fait que les travaux soient exécutés par des machines ou à la main, dans une usine ou dans un ménage, n'est pas pris en compte pour la définition des activités manufacturières. Cela revient en pratique à classer les activités artisanales au sein des activités réalisées de manière industrielle et en série¹⁵.

¹⁵ En France, avant le 1^{er} janvier 2023, une entreprise ou un auto-entrepreneur de l'artisanat pouvait obtenir l'inscription de son activité principale au Répertoire des Métiers selon la Nomenclature d'Activités Française du secteur des métiers de l'Artisanat (NAFA). Ce répertoire est depuis remplacé par le registre national des entreprises (RNE) et seul le code APE (activité principale exercée) identifie l'activité de l'entreprise.

Ce critère d'opérabilité affecte le contenu du champ statistique de la culture, auquel certains pays européens auraient souhaité en particulier adjoindre les activités d'artisanat visant la création d'objets artistiques originaux. Une centaine de codes de la nomenclature englobent de telles activités, dans différentes sections et notamment celle de l'industrie. Les organisations professionnelles pour la promotion des métiers d'art¹⁶ auraient souhaité bénéficier d'une ou de plusieurs sous-classes identifiées de la nomenclature qui isoleraient ces activités, afin de leur donner une visibilité dans les statistiques et leur comptabilisation lors de la délimitation du champ culturel. Mais cette demande, exprimée à l'occasion de la refonte de la NACE 2025 (**encadré 4**) ne pouvait aboutir en raison du critère d'opérabilité décrit.

► **Encadré 4. La culture dans la refonte des nomenclatures d'activités 2025**

La refonte récente des nomenclatures d'activités, qui a donné lieu à une nouvelle NACE et à une nouvelle NAF qui entreront en vigueur au 1^{er} janvier 2026, avait pour but de refléter plus fidèlement la réalité des activités économiques dans la classification. Il était en particulier nécessaire de prendre en compte l'impact de la mondialisation et de la digitalisation sur de nombreuses activités économiques fournissant des biens et des services, de même que le poids croissant dans l'économie mondiale de nouvelles activités (par exemple des activités spécialisées visant à protéger l'environnement ou des activités d'intermédiation) et le poids moindre d'autres activités (par exemple, la vente de support physique d'écoute musicale).

La nomenclature internationale produite par l'Organisation des Nations unies (la CITI : classification internationale type par industrie) en amont de la NACE et de la NAF 2025 avec lesquelles elle forme un système emboîté, a posé, parmi ses objectifs pour la refonte, l'amélioration du recueil des statistiques culturelles. Il en a résulté, dans les trois nomenclatures, une restructuration des divisions 90 « Activités créatives, artistiques et de spectacle » et 91 « Bibliothèques, archives, musées et autres activités culturelles », avec la constitution de trois nouveaux groupes d'activités pour les arts et le spectacle, d'un nouveau groupe pour la conservation-restauration du patrimoine matériel et la mention explicite pour la première fois dans la nomenclature, du patrimoine immatériel, qui

considère les langues et toutes les expressions et traditions culturelles vivantes comme des biens culturels*.

En outre, la section « Information et Communication » a été scindée en deux pour aboutir à la création de la nouvelle section J « Activités d'édition, de radiodiffusion, de production et de distribution de contenus », dont la structure vise à prendre en compte trois éléments nouveaux :

- les activités d'édition, de radiodiffusion, de production et de distribution de contenu (livre et presse, cinéma, télévision, vidéo, musique, etc.), désormais souvent exercées ensemble ;
- le développement rapide des activités de distribution de contenu utilisant un mélange de médias et de plateformes ;
- les produits numériques, y compris le téléchargement et le streaming.

Pour mesurer l'impact de ces évolutions sur l'appréhension statistique du champ culture, un groupe de travail européen a été constitué pour réévaluer le champ statistique de la culture et tenir compte des transformations qui ont modifié les codes de la classification, mais en conservant l'approche par domaines et fonctions retenue jusqu'à présent. Au niveau international, les recommandations pour partager un cadre conceptuel commun pour les statistiques s'intensifient et l'OCDE a choisi d'adopter le cadre conceptuel d'Eurostat-ESSnet (OECD, 2022).

* La France a ratifié en 2006 la Convention de l'UNESCO du 17 octobre 2003 pour la sauvegarde du Patrimoine culturel immatériel.

16 Métiers caractérisés par la maîtrise de gestes, de techniques et de savoir-faire complexes en vue de la transformation de la matière et produisant des œuvres, ouvrages ou objets avec un apport artistique.

► Une limite liée à la définition sociale de la création —

Une autre limite de la classification – qui nécessite son actualisation régulière – résulte du fait que les pratiques culturelles sont reconnues comme telles selon des conventions sociales qui évoluent. Fondés sur la notion d'œuvre et en partie aussi sur la notion de contenu intellectuel novateur, les contours de la culture sont naturellement amenés à évoluer en fonction de l'élargissement des techniques de création, de leur hybridation, de l'évolution des goûts et des usages, de la reconnaissance sociale de l'originalité des créations, etc. De fait, l'UNESCO (UNESCO, 2009) ou d'autres organisations internationales et certains organismes de statistique nationaux se différencient quelque peu d'Eurostat dans leur approche¹⁷. La délimitation statistique du champ culturel est donc régulièrement remise en cause. La NACE 2025 fait une place dans les codes « historiquement culturels » de la création artistique (groupe 90) aux activités des influenceurs apparaissant dans les vlogs¹⁸, étant donné leur nouvelle importance portée par le développement des réseaux sociaux. Au sein de l'audiovisuel, l'ensemble des activités de distribution de contenus vidéo est pris en compte, que ce soit via les plateformes, les sites en ligne, les réseaux sociaux. De nouveaux champs apparaissent aussi, comme celui du cinéma d'animation.

Quoi qu'il en soit, parallèlement aux nomenclatures d'activités et devant les limites décrites, d'autres nomenclatures sont donc indispensables à l'analyse socio-économique de la culture, par exemple celle des professions qui permet d'isoler les effectifs des métiers considérés comme spécifiquement culturels, qu'ils s'exercent ou non dans des secteurs culturels. Les approches « métiers » et « secteurs » se recoupent en effet partiellement (Higgs et Cunningham, 2008) : certaines professions artistiques s'exercent dans des secteurs non culturels (par exemple, un designer dans l'industrie automobile) et inversement¹⁹.

► Le suivi des professions culturelles, autre défi statistique —

Les démarches et difficultés relevées pour délimiter le champ statistique des activités culturelles se retrouvent lorsqu'il s'agit de définir de façon univoque le champ des professions culturelles et s'adapter aux évolutions de la société (Menger, 2010). En outre, le suivi des professions culturelles est particulièrement complexe de par l'existence de statuts professionnels spécifiques à la culture, du poids exceptionnel du non-salariat et du recours fréquent à la pluriactivité. La combinaison de ces caractéristiques, associée à de petits effectifs sur certaines professions, rend cette comptabilisation au travers des dispositifs de la statistique publique plus difficile.

Les statistiques sur les effectifs de l'emploi culturel s'appuient actuellement majoritairement sur l'enquête Emploi de l'Insee, car elle fournit une mesure des concepts d'activité (emploi, chômage et inactivité) harmonisée entre pays et stable dans le temps, ainsi que des

¹⁷ Par exemple, l'Organisation mondiale de la propriété intellectuelle (OMPI) compte les logiciels et bases de données, l'UNESCO, le patrimoine naturel, etc. (Measuring CCS Consortium, 2022).

¹⁸ Un « vlog », mot-valise issu des termes « vidéo » et « blog », consiste à faire des vidéos de sa vie quotidienne pour les diffuser sur des plateformes dédiées comme YouTube.

¹⁹ En 2020, 701 600 personnes exercent en France une profession culturelle au titre de leur emploi principal, soit 2,6 % de l'emploi total, et 739 800 personnes travaillent dans le secteur culturel, soit 2,7 % de l'emploi total (DEPS, Chiffres clés 2023, 2024).



Le suivi des professions culturelles est particulièrement complexe de par l'existence de statuts professionnels spécifiques à la culture, du poids exceptionnel du non-salariat et du recours fréquent à la pluriactivité.



informations détaillées sur les caractéristiques des personnes (âge, sexe, niveau de diplôme, etc.) et les caractéristiques de l'emploi principal (type de contrat de travail, temps de travail, catégorie socioprofessionnelle, etc.) (Merchaoui et Picard, 2024). Cependant, le manque d'information sur les activités secondaires conduit à une probable sous-estimation de l'emploi culturel, et la spécificité de certaines professions impose de recourir à des dispositifs d'observation complémentaires.

► Des dispositifs d'observation spécifiques pour les artistes-auteurs...

Les artistes-auteurs (plasticiens, écrivains, compositeurs, etc.) perçoivent des droits d'auteur au titre de leur création artistique. Les conditions d'exercice de leur profession font l'objet d'une attention particulière²⁰. Jusqu'en 2022, les données produites par deux organismes de gestion auxquels les artistes-auteurs pouvaient s'affilier²¹ ont permis au DEPS de produire des analyses régulières sur ces populations (Patureau et Sinigaglia, 2020 ; Tholozan et Thoumelin, 2022). Le récent transfert de la gestion de l'ensemble du recouvrement des cotisations et contributions des artistes-auteurs et de leurs diffuseurs à l'Urssaf du Limousin ouvre de nouvelles perspectives de coopération entre le DEPS et la direction statistique de l'Urssaf caisse nationale, et pourrait à terme conduire à leur intégration par l'Insee dans la « **base non-salariés**²² ».

Ce transfert a eu pour conséquence un élargissement du périmètre de l'observation de cette population, les individus étant désormais enregistrés « au premier euro » de droit d'auteur perçu, y compris lorsque le métier artistique n'est pas leur métier principal, alors que l'affiliation traduisait une sorte de permanence d'exercice dans la profession. Par ailleurs, ce transfert a induit une perte d'information sur la profession exercée, à la suite de la mise en place d'un formulaire de déclaration très succinct. La mise en ligne d'un nouveau formulaire depuis le 1^{er} janvier 2023 doit permettre de poursuivre l'analyse des contours de la population des artistes-auteurs et celle de l'évolution de leurs revenus.

► ... et pour les intermittents du spectacle

Les intermittents du spectacle (artistes du spectacle, techniciens et ouvriers) sont les salariés des entreprises du secteur du spectacle et de l'audiovisuel recrutés avec un contrat à durée déterminée particulier appelé CDD d'usage, un contrat de travail qui peut ne durer que quelques jours et être renouvelé sans limite, à la différence d'un

²⁰ Un rapport de mission (Racine, 2020) a conduit le ministère de la Culture à mettre en place un « Plan artistes-auteurs » pour améliorer leurs conditions d'exercice.

²¹ L'Association pour la gestion de la sécurité sociale des auteurs (Agressa) et l'association nationale des artistes des arts visuels de France (la Maison des Artistes, MDA).

²² <https://www.insee.fr/fr/metadonnees/source/serie/s1037>.

CDD ordinaire. Pour compenser la succession parfois discontinue des contrats courts, les intermittents du spectacle peuvent bénéficier d'une indemnisation dans le cadre du régime de l'Assurance chômage. France Travail (anciennement Pôle emploi) récolte les données détaillées liées à ce régime et permet d'établir les statistiques d'emploi (Pôle emploi, 2022) et d'indemnisation de cette population (Pôle emploi, 2023). La base « Tous salariés » de l'Insee permet également de repérer les intermittents, mais elle ne fournit pas d'information sur leur indemnisation.

► La pluriactivité, plus fréquente parmi les professionnels de la culture, nécessite d'apparier les sources entre elles



Un autre défi pour observer l'emploi est celui de la pluriactivité. Dans les secteurs culturels, un actif sur trois exerce à la fois un travail salarié et non salarié.



Un autre défi pour observer l'emploi est celui de la pluriactivité. Dans les secteurs culturels, un actif sur trois exerce à la fois un travail salarié et non salarié. Cette pluriactivité peut se doubler d'un cumul de contrats dans le salariat ou le non-salariat. Cette donnée complique le décompte de l'emploi culturel, car il est, en l'état actuel des sources, impossible de suivre les trajectoires individuelles des professionnels de la culture dans une base consolidée des salariés, non-salariés, et artistes-auteurs. Par exemple, les artistes-auteurs qui exercent aussi une profession salariée pour

compléter leur revenu seront repérés dans le dispositif d'observation des salariés au titre de leur emploi salarié complémentaire (exercé éventuellement dans un secteur non culturel), mais également dans le dispositif d'observation construit pour les artistes-auteurs, sans qu'il ne soit établi qu'il s'agisse du même individu. Il est ainsi impossible de fournir une vision complète des conditions d'emploi et de rémunération des professionnels de la culture (secteur d'activité, type de profession et catégorie socioprofessionnelle (PCS), nature du contrat, rémunérations, etc.).

► La refonte de la nomenclature des PCS en 2020 impose de recréer des catégories ad hoc

Le dernier enjeu des statistiques de l'emploi culturel concerne la nomenclature des professions et catégories socio-professionnelles (PCS). La refonte en 2020 de la version de 2003, qui conduit à un nombre réduit de rubriques (Amossé, 2020), induit une perte de précision sur le champ des professions culturelles souvent de petits effectifs. Certains nouveaux codes et libellés de la PCS 2020 les regroupent avec des professions non culturelles, ce qui empêche de suivre les professions culturelles incluses dans le nouveau code. D'autres codes correspondent à des regroupements de deux codes et libellés de la PCS 2003, et donc à une perte de détail sectoriel. Par exemple, les PCS en 2003 dédiées aux métiers du patrimoine sont désormais dans la PCS 2020 groupées avec les « Cadres spécialistes de la formation », les formateurs non cadres, les directeurs de centres de formation, qui n'exercent pas une profession culturelle.

Les « Traducteurs et interprètes » (PCS 2003) se retrouvent dans une classe large « Assistants de la communication, de la publicité et des relations publiques » (PCS 2020). Les artistes dramatiques sont réunis avec les artistes de la danse, du cirque et des spectacles divers alors que chacun de ces domaines disposait d'un code spécifique dans la PCS 2003. Le DEPS doit désormais travailler avec l'Insee pour constituer un champ « culture » des professions à partir des libellés en clair des professions présents dans les déclarations à la source.

► Des enjeux et problématiques multiples qui invitent à nouer de nombreuses collaborations au sein du service statistique public

Les quelques exemples de défis méthodologiques propres aux statistiques culturelles ne couvrent pas l'ensemble des thématiques qui relèvent du champ culturel. Son étendue et la diversité des problématiques soulevées invitent le service statistique ministériel à tisser de nombreuses collaborations au sein du service statistique public. En particulier, les liens sont étroits entre le DEPS et la DEPP²³, avec laquelle plusieurs dispositifs d'enquête auprès des enfants ont été conçus, tant pour le suivi des politiques d'éducation artistique et culturelle (EAC) que pour l'analyse des pratiques culturelles des plus jeunes. La cohorte Elfe suivie par l'Ined²⁴ et l'Inserm²⁵, avec la contribution scientifique et financière de nombreux partenaires, dont le DEPS, offre un autre outil précieux de suivi longitudinal des pratiques depuis le plus jeune âge. Avec l'Injep²⁶, les centres d'intérêt commun relèvent tout autant de la jeunesse que du volet associatif, les associations culturelles pesant parmi l'ensemble des acteurs de la culture. Sur le volet économique, le DEPS a également mis en place une méthode de mesure de la valeur ajoutée produite par les branches culturelles, à partir des données et concepts de la comptabilité nationale élaborée à l'Insee. Les données des finances publiques et des collectivités locales permettent d'étudier le financement de la culture, tant du côté des particuliers ou des entreprises (crédits d'impôt, mécénat) que du côté des collectivités territoriales. Les données sur le commerce extérieur apportent également des informations précieuses sur le commerce des œuvres culturelles et d'objets d'art. Enfin, si les échanges avec l'Insee et la Dares²⁷ permettent de suivre les professions culturelles, les échanges avec le Sies²⁸, qui délègue au DEPS la collecte des données sur les formations culturelles de l'enseignement supérieur, permettent de suivre les étudiants qui y sont inscrits²⁹.

23 DEPP : La Direction de l'évaluation, de la prospective et de la performance est le Service statistique ministériel (SSM) du ministère de l'Éducation nationale.

24 Ined : Institut national d'études démographiques.

25 Inserm : Institut national de la santé et de la recherche médicale.

26 Injep : Institut national de la jeunesse et de l'éducation populaire est un service à compétence nationale du ministère de l'Éducation nationale et de la Jeunesse, qui comprend le service statistique ministériel, la Mission enquêtes, données et études statistiques (Medes).

27 Dares : La Direction de l'animation de la recherche, des études et des statistiques est le SSM du ministère chargé du travail.

28 Sies : La sous-direction des systèmes d'information et études statistiques est le SSM du ministère de l'Enseignement supérieur et de la Recherche.

29 Voir le [site internet du DEPS](#) répertoriant les nombreuses études.

“

La nécessité de mieux comprendre les liens entre consommations culturelles et offres culturelles et les dynamiques territoriales nécessitent de mettre à disposition des données de qualité via des dispositifs innovants.

”

Ainsi, les statistiques culturelles, dans leur diversité, tentent de répondre aux enjeux de quantification dans des domaines variés, suivant les évolutions sociétales, mais les défis restent nombreux. La nécessité de mieux comprendre les liens entre consommations culturelles et offres culturelles et les dynamiques territoriales nécessitent de mettre à disposition des données de qualité via des dispositifs innovants (site interactif³⁰). Les évolutions technologiques rapides, la diversification des pratiques culturelles et l'émergence de nouvelles formes de consommation culturelle exigent une adaptation continue des méthodes de collecte et

d'analyse des données, des nomenclatures et des sources mobilisées. Cependant, il est indispensable de garantir en parallèle la stabilité des dispositifs d'observation permettant d'analyser des évolutions de long terme, ainsi que l'harmonisation des concepts aux niveaux européen et international.

30 <https://atlasculture.fr/>.

► Bibliographie

- 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 27 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4497076/courstat-4-5.pdf>.
- ARDILLY, Pascal, 2023. Peut-on se fier aux sondages empiriques ? In : *Courrier des statistiques*. [en ligne]. 11 décembre 2023. Insee. N° N10, pp. 113-132. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7722109?sommaire=7722116>.
- ASP, 2024. Rapport annuel 2023 de l'Autorité de la Statistique Publique. In : *site de l'Autorité de la Statistique Publique*. [en ligne]. 19 mars 2024. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.autorite-statistique-publique.fr/rapports-dactivites-annuels/>.
- BOURDIEU, Pierre et DARBEL, Alain, 1966. *L'Amour de l'art*. Les Éditions de Minuit. ISBN 9782707300287.
- BOURDIEU, Pierre, 1979. *La Distinction. Critique sociale du jugement*. Les Éditions de Minuit. ISBN 9782707302755.
- COULANGEON, Philippe, 2005. *Sociologie des pratiques culturelles*. Éditions La découverte, collection Repères. ISBN 9782707190604.
- COULANGEON, Philippe, 2011. *Les métamorphoses de la distinction. Inégalités culturelles dans la France d'aujourd'hui*. Paris, Grasset, coll. « Mondes vécus ». ISBN 9782246769712.
- DEPS, Comité d'histoire du ministère de la culture, 2010. L'invention de la prospective culturelle – Textes choisis d'Augustin Girard. In : *Culture prospective*. [en ligne]. Janvier 2010. DEPS. 2010-1. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/nous-connaître/Decouvrir-le-ministere/histoire-du-ministere/ressources-documentaires/publications/Les-collaborations/l-invention-de-la-prospective-culturelle.-textes-choisis-d-augustin-girard>.
- DEPS, 2024. Chiffres Clés 2023 de la culture et de la communication. In : *site du ministère de la Culture*. [en ligne]. 14 mars 2024. DEPS. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/Collections-d-ouvrages/chiffres-cles-statistiques-de-la-culture-et-de-la-communication-2012-2023/Chiffres-Cles-2023-de-la-culture-et-de-la-communication>.
- DEROIN, Valérie, 2011. Approche statistique européenne de la culture – Synthèse des travaux européens ESSnet-Culture, 2009-2011. In : *Culture Études*. [en ligne]. 29 décembre 2011. DEPS. 2011-8. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-etudes-2007-2024/Approche-statistique-europeenne-de-la-culture.-Synthese-des-travaux-europeens-ESSnet-Culture-2009-2011-CE-2011-8>.

- DÉTREZ, Christine, 2020. *Sociologie de la culture*. Armand Colin, « Coursus ». ISBN 9782200626969.
- DONNAT, Olivier et LÉVY, Florence, 2007. Approche générationnelle des pratiques culturelles et médiatiques. In : *Culture prospective*. [en ligne]. 3 juin 2007. DEPS. 2007-3. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/Culture-prospective-2007-2014/Approche-generationnelle-des-pratiques-culturelles-et-mediatiques-CP-2007-3>.
- DONNAT, Olivier, 2011. Pratiques culturelles, 1973-2008. Questions de mesure et d'interprétation des résultats. In : *Culture méthodes*. [en ligne]. 30 novembre 2011. DEPS. 2011-2. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/Culture-methodes-2007-2014/Pratiques-culturelles-1973-2008.-Questions-de-mesure-et-d-interpretation-CM-2011-2>.
- DONNAT, Olivier, 2018. Évolution de la diversité consommée sur le marché du livre, 2007-2016. In : *Culture Études*. [en ligne]. 26 septembre 2018. DEPS. 2018-3. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-etudes-2007-2024/Evolution-de-la-diversite-consommee-sur-le-marche-du-livre-2007-2016-CE-2018-3>.
- DUBOIS, Vincent, 1999. *La politique culturelle. Genèse d'une catégorie d'intervention publique*. Belin Éditeur. ISBN 9782701164137.
- ESSnet-CULTURE, 2012. ESSnet-CULTURE Final Report. In : *site de la Commission européenne*. [en ligne]. Mai 2012. [Consulté le 5 août 2024]. Disponible à l'adresse : https://ec.europa.eu/assets/eac/culture/library/reports/ess-net-report_en.pdf.
- GARCIA, Léa, JONCHERY, Anne et THOUMELIN, Claire, 2024. Les sorties culturelles des Français et leurs pratiques en ligne en 2023 – Cinéma, concert et théâtre. In : *Culture Études*. [en ligne]. 8 avril 2024. DEPS. 2024-2. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-etudes-2007-2024/les-sorties-culturelles-des-francais-et-leurs-pratiques-en-ligne-en-2023-cinema-concert-et-theatre-ce-2024-2>.
- GIRARD, Augustin, 1965. « Pourquoi des recherches ? La fin de « l'ère des goûts et des couleurs », dossier « Des chiffres pour la culture. Colloque de Bourges ». In : *Revue trimestrielle de l'Association d'étude pour l'expansion de la recherche scientifique*. Décembre 1965. n° 21. [en ligne]. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/content/download/20516/file/2-Pourquoi%20des%20recherches.pdf>.
- GIRARD, Augustin, 1978. Industries culturelles. In : *Futuribles*. Septembre-octobre 1978. N° 17. [en ligne]. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/fr/content/download/20508/file/7-Industries%20culturelles.pdf>.
- GLEVAREC, Hervé, 2013. *La culture à l'ère de la diversité*. Éditions de l'Aube. ISBN 9782815906760.

- HIGGS, Peter et CUNNINGHAM, Stuart, 2008. Creative industries mapping: Where have we come from and where are we going? In : *Creative Industries Journal*. Volume 1, n° 1, pp. 7-30.
- JONCHERY, Anne et LOMBARDO, Philippe, 2020. Pratiques culturelles en temps de confinement. In : *Culture Études*. [en ligne]. 13 décembre 2020. DEPS. 2020-6. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-etudes-2007-2024/Pratiques-culturelles-en-temps-de-confinement-CE-2020-6>.
- LAHIRE, Bernard, 2004. *La culture des individus*. Dissonances culturelles et distinction de soi. La Découverte. ISBN 9782707149282.
- LOMBARDO, Philippe et WOLFF, Loup, 2020. Cinquante ans de pratiques culturelles en France. In : *Culture Études*. [en ligne]. 10 juillet 2020. DEPS. 2020-2. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-etudes-2007-2024/Cinquante-ans-de-pratiques-culturelles-en-France-CE-2020-2>.
- MAILLARD, Sophie, 2023. Un aperçu du marché de la musique en streaming de 2020 à 2022 : une analyse de données hebdomadaires détaillées de marché. In : *Rapport du sénateur Julien Bargeton, relatif à la stratégie de financement de la filière musicale en France*. [en ligne]. Avril 2023. Annexe IV. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/Rapports/Rapport-du-senateur-Julien-Bargeton-relatif-a-la-strategie-de-financement-de-la-filiere-musicale-en-France2>.
- MARTIN, Laurent, 2012. Du SER au DEP, ou la constitution d'une socio-économie de la culture et d'une prospective culturelle au service de l'action (1959-1993). In : *Revue historique*. [en ligne]. 2012/3, n° 663, pp. 683-704. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.cairn.info/revue-historique-2012-3-page-683.htm>.
- MENGER, Pierre-Michel, 2010. Les artistes en quantités. Ce que sociologues et économistes s'apprennent sur le travail et les professions artistiques. In : *Revue d'économie politique*. [en ligne]. Volume 120, n° 1, pp. 205-236. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.cairn.info/revue-d-economie-politique-2010-1-page-205.htm>.
- MEASURING CCS CONSORTIUM, 2022. Measuring the Cultural and Creative Sectors in the EU Final report. In : *site de Measuring the Cultural and Creative Sectors EU*. [en ligne]. 22 novembre 2022. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.measuring-ccs.eu/>.
- MERCHAOUI, Wided et PICARD, Sébastien, 2024. Vingt-cinq ans d'évolution de l'emploi dans les professions culturelles : 1995-2019. In : *Culture Chiffres*. [en ligne]. 14 juin 2024. DEPS. 2024-1. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-chiffres-2007-2024/vingt-cinq-ans-d-evolution-de-l-emploi-dans-les-professions-culturelles-1995-2019-cc-2024-1>.

- MÜLLER, Jörg et SCHREIBER, Amandine, 2022. Les sorties culturelles des Français après deux années de Covid-19. In : *Culture Études*. [en ligne]. 20 juillet 2022. DEPS. 2022-6. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-etudes-2007-2024/Les-sorties-culturelles-des-Francais-apres-deux-annees-de-Covid-19-CE-2022-6>.
- OECD, 2022. The Culture Fix: Creative People, Places and Industries. In : *Local Economic and Employment Development (LEED)*. [en ligne]. OECD Publishing, Paris. [Consulté le 5 août 2024]. Disponible à l'adresse : https://www.oecd-ilibrary.org/urban-rural-and-regional-development/the-culture-fix_991bb520-en.
- PATUREAU, Frédérique et SINIGAGLIA Jérémy, 2020. *Artistes plasticiens : de l'école au marché*. Éditions Presses de Sciences Po, co-édition Ministère de la Culture - DEPS. ISBN 9782724625950.
- PÔLE EMPLOI, 2022. L'emploi intermittent dans le spectacle au cours de l'année 2021. In : *Statistiques et Indicateurs*. [en ligne]. 25 octobre 2022. 22.033. [Consulté le 5 août 2024]. Disponible à l'adresse : https://www.francetravail.org/files/live/sites/peorg/files/documents/Statistiques-et-analyses/S%26l/si_22.033_intermittents_2021.pdf.
- PÔLE EMPLOI, 2023. Les allocataires indemnisés au titre des annexes 8 et 10 en 2021. In : *Statistiques et Indicateurs*. [en ligne]. 27 janvier 2023. 23.002. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://statistiques.pole-emploi.org/indem/indempub/216735>.
- RACINE, Bruno, 2020. L'auteur et l'acte de création. In : *Rapport au ministre de la Culture*. [en ligne]. 22 janvier 2020. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.vie-publique.fr/rapport/272913-lauteur-et-lacte-de-creation>.
- THOLOZAN, Léa et THOUMELIN, Claire, 2022. Les artistes-auteurs en 2018. In : *Culture Études*. [en ligne]. 11 février 2022. DEPS. 2022-2. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.culture.gouv.fr/espace-documentation/statistiques-ministerielles-de-la-culture2/publications/collections-de-synthese/culture-etudes-2007-2024/les-artistes-auteurs-en-2018-ce-2022-2>.
- UNESCO, 2009. The 2009 UNESCO Framework for Cultural Statistics (FCS). In : *site de UNESCO Institute for Statistics*. [en ligne]. [Consulté le 5 août 2024]. Disponible à l'adresse : https://uis.unesco.org/sites/default/files/documents/unesco-framework-for-cultural-statistics-2009-en_0.pdf.
- VICARD, Augustin, 2023. Quantifier la pratique sportive : une approche sociologique et sanitaire. In : *Courrier des statistiques*. [en ligne]. 11 décembre 2023. Insee. N° N10, pp. 53-69. [Consulté le 5 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7722093?sommaire=7722116>.

La statistique annuelle d'entreprises : sa nature, son histoire, ses enjeux



Sylvain Moreau*

Les données annuelles sur les entreprises, produites par l'Insee, ainsi que les études correspondantes, sont des rubriques du site de l'Insee peu consultées, avec un faible impact médiatique : elles s'adressent plus particulièrement à des experts ou des professionnels qui sont à même d'en comprendre les enjeux. Ces travaux constituent une des activités les plus stratégiques de l'Institut. Ils apportent des éclairages objectifs réguliers et très détaillés sur la situation économique du pays, tout en constituant le principal intrant pour l'élaboration des comptes nationaux. Leur production mobilise un grand nombre d'agents, dont le métier, statisticien d'entreprises, reste méconnu. Les problématiques liées à leurs activités sont souvent très spécifiques et peuvent nécessiter des compétences comptables, fiscales ou juridiques.

Pour répondre aux différents enjeux de cette production de données, la statistique publique a conçu et élaboré dès le début des années 70 un système d'information structuré, profitant de la disponibilité de sources administratives, enrichies par des enquêtes. Une telle infrastructure a permis de construire des bases de données microéconomiques sur les entreprises d'une très grande richesse. Ce système a considérablement évolué depuis, mais il fait preuve d'une extraordinaire résilience. Les principes définis à l'époque sont inchangés et permettent d'observer tout au long de cette période les comportements des entreprises pour s'adapter au contexte économique. Cependant, alors que celles-ci font face à des questions de plus en plus complexes, et que les organisations mises en place évoluent constamment et aboutissent à des solutions originales voire novatrices, se pose la question de l'efficacité de ce système d'observation et de son adaptabilité.

 The annual business data, produced by INSEE, and their corresponding studies, are disseminated on the Institute's website, in sections that are rarely consulted and with little media impact. These data and studies are specifically intended for experts or professionals who are in a position to understand the issues at stake. This work is one of the Institute's most strategic activities. It provides regular, objective and very detailed insights into the country's economic situation while representing the main input for drawing up the national accounts. Their production involves a large number of agents, whose profession – that of business statistician – remains little known. The issues associated with their activities are often very specific and may require special accounting, tax or legal know-how.

To respond to the various challenges stemming from this data production, Official Statistics designed and developed a structured information system in the early 1970s, taking advantage of the availability of administrative sources enriched by surveys. This made it possible to build up very rich databases of microeconomic data on businesses. This system has changed considerably since then, yet has continued to be extraordinarily resilient. The principles defined at the time remain unchanged, and have enabled us to observe how companies have adapted to the economic context throughout this period. However, while companies are beginning to face increasingly complex challenges, and as the organisations they put in place are constantly changing and leading to original and often innovative solutions, the question arises as to the effectiveness of this observation system and its continuing adaptability.

* Directeur des statistiques d'entreprises, Insee.
sylvain.moreau@insee.fr

En 2024, plusieurs articles dans la presse ont évoqué la croissance de l'enseigne « Grand Frais » qui comprend maintenant plus de 300 magasins en France. Cette enseigne a une organisation très particulière : chaque magasin est une société à part entière, organisée sous forme de GIE (groupement d'intérêt économique) et chaque GIE est partagé entre ses membres, ceux-ci étant donc des partenaires qui interviennent dans ses rayons. Une autre singularité est l'APE (activité principale exercée) des GIE : pour une majorité d'entre eux, celle-ci ne relève pas du commerce de détail, et le constat est identique pour celles des partenaires. Concernant le commerce alimentaire, analyser l'activité de « Grand Frais » et comparer son évolution avec celles des grands acteurs du secteur impliquent d'identifier au préalable chaque entreprise concernée, puis de récupérer pour chacune d'elles son chiffre d'affaires ainsi que les flux financiers entre toutes les entreprises qui dépendent de l'enseigne. La statistique publique a accès aux données fiscales et peut, une fois les entreprises identifiées, calculer un chiffre d'affaires consolidé. Mais elle doit définir le contour de la population d'entreprises utilisée pour ce calcul. Or, ce travail n'est pas automatisable ; il demande d'examiner les documents mis à disposition par l'entreprise. Sans ce travail d'expertise, le risque est de ne pas attribuer les résultats de « Grand Frais » à la bonne activité, voire de ne pas identifier les doubles comptes dans l'estimation du chiffre d'affaires.

Analyser l'industrie automobile française, l'activité qu'elle génère sur le territoire français, le volume de sa production, le nombre de salariés qu'elle emploie, requiert une connaissance fine de la structure des grands constructeurs français de véhicules.

Analyser l'industrie automobile française, l'activité qu'elle génère sur le territoire français, le volume de sa production, le nombre de salariés qu'elle emploie, requiert une connaissance fine de la structure des grands constructeurs français de véhicules, d'autant plus que ces organisations sont particulièrement complexes. Renault SAS en est un exemple. Celui-ci est un donneur d'ordre industriel. Il achète les intrants et les livre gratuitement à ses usines, et il achète également un service industriel d'assemblage à celles-ci. Il est donc propriétaire des véhicules automobiles dès le premier boulon. Les usines facturent un service d'assemblage à Renault SAS. Il n'y a pas d'intermédiaire commercial, Renault SAS vend

directement les véhicules à ses réseaux de distribution. PSA Automobiles est quant à lui un donneur d'ordre commercial. Il achète les intrants et les vend à ses usines, il leur facture la mise à disposition de personnel et il leur achète les véhicules (et pièces automobiles). Les usines achètent les intrants et vendent leur production à PSA Automobiles, qui fait ensuite appel à deux intermédiaires commerciaux : Peugeot Automobiles et Citroën Automobiles pour distribuer lesdites marques.

Pour comparer l'activité et les performances de ces deux entreprises, il est nécessaire de comprendre leurs organisations respectives, d'identifier les différentes unités qui les composent et de consolider leurs résultats pour ces deux ensembles, de façon à obtenir des résultats comparables qui évitent, là aussi, les doubles comptes et qui reflètent leur activité et leurs performances économiques. Et pour connaître l'activité du secteur automobile, il faut faire ce même travail pour l'ensemble des entreprises de la construction automobile. On pourra alors agréger leurs données et estimer avoir une vision correcte du secteur automobile en France.

Ces travaux ont deux écueils : un nécessaire accès à de nombreuses sources d'information et un coût élevé. Ils sont néanmoins indispensables pour disposer d'éléments pertinents permettant d'analyser le secteur. On pourra alors s'appuyer sur ces données exhaustives et sans double compte pour tout diagnostic économique, voire pour toute mise en œuvre de politique publique, mais aussi pour la comptabilité nationale. Ces travaux doivent être actualisés régulièrement, les entreprises, surtout les plus grandes, évoluant constamment et exigeant une veille constante.

Cette observation des entreprises, qu'on appelle le système productif, est une des missions importantes de l'Insee, et plus largement de la statistique publique (Rivière, 1998). Dans cet article, est présenté le dispositif mis en place pour décrire et comprendre le monde des entreprises. Les premiers besoins des comptes nationaux au sortir de la guerre et les outils développés à ce moment-là pour collecter les données ont conduit au cours des années à un dispositif maintenant opérationnel. Enfin, sont décrits les enjeux auxquels il devra faire face dans les années qui viennent.

► Observer les entreprises, pour quoi faire ?

L'objectif est de comprendre leur fonctionnement et leur organisation, de connaître leur production ainsi que le processus et les moyens mis en œuvre pour assurer cette production, toutes les informations nécessaires au calcul d'indicateurs économiques pertinents et à même d'enrichir le débat public et d'aider à l'élaboration des politiques publiques. Décrire le système productif est l'objet même des travaux du statisticien d'entreprises.

Le système productif peut être défini comme « l'ensemble des facteurs et des acteurs concourant à la production, à la circulation et à la consommation de richesses » (Carroué, 2013). C'est le mode d'organisation de la production en tant que mise en cohérence d'un mode d'organisation du travail et d'un système technique. Les dimensions auxquelles l'Institut s'intéresse sont multiples : production, activité, emploi, consommation d'intrants, voire émissions de gaz à effet de serre, gestion des déchets, etc. Cela donne lieu à la publication d'indicateurs spécifiques suivant des périodicités différentes, à des études ad hoc, etc.

La production de données d'entreprises comprend des données de périodicité diverse. La demande pour disposer de données infra-annuelles, généralement mensuelles, rapidement disponibles est forte. Néanmoins, la production de données annuelles (**encadré 1**), qui permet de disposer de données de cadrage très détaillées, présente

► Encadré 1. Données annuelles/Données structurelles

Dans la terminologie habituelle, on distingue les données conjoncturelles des données structurelles, c'est-à-dire les données infra-annuelles des données annuelles. L'idée sous-jacente est la suivante : les évolutions annuelles sont suffisamment lentes, les phénomènes mis en œuvre prenant des années, voire des décennies pour se réaliser, ainsi une photo de l'appareil productif une année donnée reflète des caractéristiques de structure a priori stables pour une période pluriannuelle. Ceci était vrai dans les

années 70 et 80 mais l'est sans doute moins en 2024. On constate depuis plusieurs années des changements plus rapides dans l'organisation du système productif, avec des forts impacts sur l'activité et l'économie des entreprises. Il semble que depuis 2020, les évolutions se sont encore accélérées avec les différentes crises auxquelles ont dû faire face les acteurs économiques. Aussi, on parle de données annuelles d'entreprises plutôt que de données structurelles.



Un des principaux objectifs est de fournir des chiffres servant d'intrants à la comptabilité nationale.



des enjeux plus forts en matière de production et a fait l'objet d'investissements importants et d'une réflexion particulièrement aboutie. C'est à leur présentation que le reste de l'article sera consacré.

La production annuelle de données d'entreprises n'a pas pour seul objectif de construire des indicateurs permettant de comprendre l'organisation du système productif. Un des principaux objectifs est de fournir des chiffres servant d'intrants à la comptabilité

nationale. De nombreux choix conceptuels et organisationnels faits depuis le début l'ont été pour répondre aux demandes de celle-ci, ce qui a complètement modelé la statistique annuelle d'entreprises.

► Encadré 2. Quelques définitions

Établissement identifié par un numéro Siret* :

L'établissement est une unité de production géographiquement individualisée, mais juridiquement dépendante de l'unité légale. Il produit des biens ou des services : ce peut être une usine, une boulangerie, un magasin de vêtements, un des hôtels d'une chaîne hôtelière, la « boutique » d'un réparateur de matériel informatique, etc. L'établissement, unité de production, constitue le niveau le mieux adapté à une approche géographique de l'économie.

Remarque : La population des établissements est relativement stable dans le temps et est moins affectée par les mouvements de restructuration juridique et financière que celle des entreprises.

Unité légale ou entreprise-unité légale identifiée par un numéro Siren** :

L'unité légale est une entité juridique de droit public ou privé. Cette entité juridique peut être :

- une personne morale, dont l'existence est reconnue par la loi indépendamment des personnes ou des institutions qui la possèdent ou qui en sont membres ;
- une personne physique qui, en tant qu'indépendant, peut exercer une activité économique.

Elle est obligatoirement déclarée aux administrations compétentes (greffes des tribunaux de commerce, Sécurité sociale, DGFIP, etc.) pour exister. L'existence d'une telle unité dépend du choix des propriétaires ou de ses

créateurs (pour des raisons organisationnelles, juridiques ou fiscales). L'unité légale est l'unité principale enregistrée dans Sirene***.

Remarque : Cette définition de l'unité légale ne doit pas être confondue avec celle de l'entreprise, considérée comme unité statistique (Hachid et Leclair, 2022).

Entreprise :

L'entreprise est la plus petite combinaison d'unités légales qui constitue une unité organisationnelle de production de biens et de services jouissant d'une certaine autonomie de décision, notamment pour l'affectation de ses ressources courantes.

Groupe de sociétés :

Un groupe de sociétés est une entité économique formée par une société contrôlante et l'ensemble des sociétés qu'elle contrôle.

Contrôler une société, c'est avoir le pouvoir de nommer la majorité des dirigeants.

Le contrôle d'une société A par une société B peut être direct (la société B est directement détentrice de la majorité des droits de vote au conseil d'administration de A) ou indirect (B a le contrôle de sociétés intermédiaires C, voire D, E, etc., à qui elle peut demander de voter d'une même façon au conseil d'administration de A, obtenant ainsi la majorité des droits).

La définition statistique française actuellement en vigueur retient comme critère de contrôle pour définir les contours des groupes la majorité absolue des droits de vote.

* <https://www.insee.fr/fr/metadonnees/definition/c1841>.

** <https://www.insee.fr/fr/metadonnees/definition/c2047>.

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

► Les entreprises, un monde complexe avec des défis d'observation

Les enjeux de la production statistique sont nombreux et rendent sa mise en œuvre particulièrement complexe :

- **Quelle est l'unité d'observation ?** Définir ce qu'est une entreprise est une opération délicate, et d'ailleurs l'unité d'observation pourra être différente suivant la question à laquelle on veut répondre. Pour l'emploi et les effets potentiels de la réindustrialisation au niveau des territoires, l'unité d'observation sera l'établissement, c'est-à-dire l'implantation locale. En revanche, pour des questions plus stratégiques, comme l'innovation ou la politique de recherche et développement (R&D¹), l'unité pertinente est soit l'unité légale, soit plus probablement le groupe (**encadrés 2 et 3**).
- **Quelle est l'unité de collecte ?** Dans le dispositif existant, selon qu'on s'intéresse à la production, à l'emploi, voire à la consommation d'énergie, l'unité de collecte la plus pertinente est souvent l'établissement, car c'est là que les informations sont le plus facilement mobilisables. En revanche, la collecte des données financières implique de contacter le service comptable qui opère souvent au niveau de l'entreprise pour plusieurs établissements.
- **Comment mesurer son activité ?** Chaque unité juridique recensée dans le répertoire Sirene a une activité principale exercée (APE). L'APE d'une unité est, parmi ses activités, celle qui génère le plus de valeur ajoutée. Elle est déterminée en fonction de la ventilation des différentes activités de l'unité statistique concernée (entreprise, unité légale, établissement). Ce code permet d'identifier le secteur d'activité de l'unité concernée mais il masque toute autre activité que pourrait avoir l'entreprise et qui peut intéresser le statisticien. Ainsi, pour reprendre l'exemple de l'industrie automobile, collecter auprès des constructeurs leurs données de production ne suffit pas, il est nécessaire de disposer d'informations sur leur activité commerciale afin de ne pas passer à côté d'une grande partie du commerce de véhicules. Cela a un impact sur la façon dont sont conçus les questionnaires et dont est organisée la collecte des données.
- **Quel est le champ géographique ?** Le champ est celui des unités productives ayant une activité sur le territoire national : l'Insee ne s'intéresse qu'à la « trace française » de ces entreprises. On ne dispose d'ailleurs généralement que de données nationales. Dans plusieurs secteurs d'activité, certains acteurs économiques parmi les plus influents sont des firmes multinationales ; la compréhension de leur organisation et de leur mode de production n'a de sens qu'à un niveau supranational. De plus en plus de travaux internationaux, menés souvent en collaboration avec d'autres instituts nationaux de statistiques (INS), sont lancés, généralement sous le parrainage d'Eurostat², mais les différences importantes entre les systèmes d'observation nationaux restent un facteur de complexité (Colin, 2019).

1 <https://www.insee.fr/fr/metadonnees/definition/c1174>.

2 Eurostat, l'Office statistique de l'Union européenne, est chargé de publier des statistiques et des indicateurs européens, permettant d'effectuer des comparaisons entre les pays et les régions.

► Encadré 3. Unité d'observation

Il n'y a pas de statistiques sans « unités statistiques » (Volle, 1984). Or, dans le cas des entreprises, l'unité d'observation est une question difficile. L'observation du système productif est fondée à l'origine sur l'exploitation de sources administratives ; historiquement l'unité d'observation était soit l'unité légale (UL), identifiée par son numéro Siren, soit l'établissement, l'unité territoriale, identifiée par son numéro Siret, voire le groupe qu'il était possible de reconstituer via les enquêtes sur les liaisons financières (LIFI¹). Les enquêtes ont ainsi été conçues, jusqu'aux années 2020, pour collecter des données au niveau de l'UL ou au niveau de l'établissement.

Néanmoins, dès la fin des années 1990, les limites de cette approche sont devenues patentes et il est apparu nécessaire de définir un objet d'étude, différent de l'UL, qui permette d'avoir une image plus appropriée de l'organisation et du fonctionnement des entreprises. Dans les années 1990, les grandes entreprises automobiles se sont restructurées, avec des UL productives et des UL gérant les moyens de production qu'elles mettaient à disposition des premières. Seule une approche d'ensemble permettait d'avoir une vision

cohérente de l'activité de construction automobile en France. Au début des années 2000, cette demande qui permet de déterminer les contours des groupes et les déclinaisons en entreprises s'est peu à peu étendue à d'autres groupes. En se généralisant et s'industrialisant, cette activité s'est appelée **profilage**² (Haag, 2019). L'objectif est de construire une unité, soit un ensemble d'UL, qui regroupe l'ensemble des facteurs de production et dispose d'une autonomie de décision notamment pour l'affectation de ses ressources courantes. Cette définition a été reprise dans la loi de modernisation de l'économie (LME) de 2008³. Cette unité s'appelle désormais « entreprise » (**figure encadré**). Cette notion d'entreprise est adaptée au travail de production d'indicateurs économiques qui se focalisent sur l'activité productrice et s'intéressent majoritairement aux données financières. Si on s'intéresse à d'autres aspects du système productif, par exemple les questions de gestion des ressources humaines, de l'innovation et de la R&D, ce partage n'est pas forcément le plus pertinent.

Une APE est calculée au niveau de l'entreprise, suivant les mêmes règles que celles appliquées au niveau de l'UL⁴.

Unités légales et entreprises de l'industrie, appartenant à un groupe, en 2022

Nombre d'unités légales (UL)	46 500 UL de l'industrie	UL hors industrie	
70 900 UL des entreprises de l'industrie	37 600	16 800 Hors financier	16 500 du financier
UL des entreprises hors industrie	8 900		

Lecture : 46 500 UL de l'industrie appartiennent à un groupe, et parmi elles 37 600 sont filiales d'une entreprise de l'industrie.

Champ : UL de l'industrie appartenant à un groupe et unités légales appartenant à un groupe de l'industrie. Les 282 500 entreprises « mono-unité légale » sont hors champ.

Source : Insee, É sane 2022 (données individuelles).

- 1 Cette enquête a été abandonnée en 2012 et remplacée par la source fiscale équivalente qui permet de disposer d'informations sur la totalité du champ. <https://www.insee.fr/fr/metadata/source/serie/s1038>.
- 2 <https://www.insee.fr/fr/metadata/definition/c1665>.
- 3 Voir les références juridiques en fin d'article.
- 4 L'APE d'une unité est l'activité qui génère le plus de valeur ajoutée. En pratique, la ventilation du chiffre d'affaires selon les branches est souvent utilisée comme critère de détermination.

• **Quel est le champ en matière de secteur d'activité ?** Le champ traditionnel de la statistique annuelle d'entreprises est celui des sociétés et entreprises individuelles, marchandes et productives ou participant au système productif, non agricoles et non financières. Ce champ n'a quasiment pas changé depuis 50 ans. Il s'explique à l'origine par une organisation de la collecte des données très différente pour les entreprises agricoles (ministère en charge de l'agriculture) et pour les entreprises relevant du secteur financier (sous la responsabilité de la Banque de France). Cependant, dans le cadre du règlement européen sur les statistiques structurelles (*Structural Business Statistics, SBS*), le champ demandé comprend le secteur financier³.

► Les premiers dispositifs pour répondre aux préoccupations de l'après-guerre

Des données souvent très détaillées ont depuis longtemps permis de construire les politiques économiques, par exemple afin de connaître et d'anticiper la production sylvicole pour répondre aux besoins de la marine royale du temps de Colbert, ou encore de prévoir et d'estimer la production houillère pour les besoins croissants de l'industrie pendant le Second Empire. Mais c'est à la veille de la Seconde Guerre mondiale que l'absence de données sur les activités industrielles et les produits est fortement ressentie par les milieux économiques et politique et donne lieu à la création des premières enquêtes de branches (**encadré 4**), qui se sont généralisées sous le régime de Vichy dans un but de répartition des matières premières.

► Encadré 4. Secteur d'activité versus branche d'activité

La statistique annuelle d'entreprises permet de disposer de résultats économiques par secteur d'activité. Un secteur regroupe des entreprises de fabrication, de commerce ou de service qui ont la même activité principale (au regard de la nomenclature d'activité économique considérée). L'activité d'un secteur n'est donc pas homogène et comprend des productions ou services secondaires qui relèvent d'autres nomenclatures que celui du secteur considéré. Au contraire, une branche d'activité regroupe des unités de production homogènes par définition.

Ces deux approches ne répondent pas aux mêmes besoins et se complètent. Si vous vous intéressez au transport de marchandises, vous regarderez l'évolution de la branche d'activité concernée. En revanche, si vous voulez faire un focus sur les entreprises de transport de marchandise et notamment l'impact de l'évolution du transport pour compte propre* sur leur activité, vous utiliserez les données sectorielles.

* c'est-à-dire les entreprises de commerce qui assurent elles-mêmes la livraison des marchandises.

Après la guerre, les questions autour du marché du travail deviennent plus prégnantes et débouchent sur la mise en œuvre des enquêtes sur l'activité et les conditions d'emploi de la main-d'œuvre (**ACEMO**⁴) en 1946, avec un questionnaire autour de la durée hebdomadaire du travail, des effectifs employés et du taux de salaire horaire. De leur côté, dès 1948, les comptables nationaux commencent à exploiter les déclarations fiscales sur les bénéfices industriels et commerciaux. Dans les années suivantes, d'autres outils

³ Voir les références juridiques en fin d'article.

⁴ <https://www.insee.fr/fr/metadonnees/source/serie/s1189>.

d'observation ou de collecte sont développés. En ces temps de pionniers, la formule « un besoin, une source » résume la façon dont le système statistique public répond aux besoins d'information sur le système productif. La nécessité d'une coordination des systèmes de production n'apparaît pas spontanément.



Une caractéristique structurante de la statistique d'entreprises est l'utilisation massive des sources administratives.



Mais une caractéristique structurante de la statistique d'entreprises est déjà présente : l'utilisation massive des sources administratives. Outre les sources fiscales, on peut aussi citer les déclarations annuelles de données sociales (DADS⁵), les permis de construire, les déclarations de taxe sur la valeur ajoutée (TVA) (CA3⁶), les déclarations en douane ; toutes ont été rapidement sollicitées par les statisticiens ou les comptables nationaux. L'exhaustivité de ces sources a toujours

été considérée comme un atout majeur pour leur exploitation à des fins statistiques et, par ailleurs, leur utilisation permet de limiter la charge de réponse des entreprises.

Très rapidement, ces systèmes construits de façon autonome montrent leurs limites. Les incohérences constatées par les comptables nationaux entre les informations disponibles par les enquêtes de branches et les sources fiscales ainsi que le constat de lacunes dans les dispositifs existants conduisent les statisticiens publics à mettre en place des collectes de données. Dans un premier temps, il est décidé de réaliser des recensements (industrie en 1962, transports et distribution en 1966) (Volle, 1982). Un des objectifs de ces recensements est d'expliquer les incohérences entre les sources existantes et de permettre de les réconcilier autant que faire se peut.

À cette époque, la confrontation des résultats des recensements avec les sources fiscales avait mis en évidence plusieurs causes de divergence : les différences de champs d'observation, la définition des variables et surtout le classement sectoriel des unités. Mais ces travaux avaient aussi permis de confirmer la fiabilité et la qualité des déclarations fiscales ainsi que la pertinence de leur usage pour élaborer les comptes nationaux. La mise en place d'un système d'information sur les entreprises construit autour des sources fiscales, complétées par des enquêtes ad hoc, apparaît alors comme une solution adaptée pour répondre aux demandes de la comptabilité nationale mais aussi à tous les besoins d'information sur le système productif.

► L'évidence d'une réponse systémique

Les recensements, tels qu'ils avaient été menés dans les années 60, se révélaient néanmoins être des solutions coûteuses ne permettant pas de répondre à la forte demande de données annuelles. Cette nécessité de disposer d'observations plus fréquentes a conduit à mettre en place les enquêtes annuelles sectorielles, en commençant par celle sur l'industrie en 1970 (après des expérimentations sur les secteurs de la chimie, de la fonderie et de l'imprimerie en 1966), permettant ainsi

⁵ <https://www.insee.fr/fr/metadonnees/source/serie/s1163>.

⁶ <https://www.insee.fr/fr/metadonnees/source/serie/s1284>.

d'abandonner l'idée même de recensement. Ces enquêtes avaient pour objectif d'enrichir et de compléter les sources fiscales, mais aussi de comparer les variables communes (compte de résultat et investissements notamment), afin de s'assurer que les données fiscales correspondaient aux concepts que la statistique publique voulait mesurer. Des redressements étaient ensuite réalisés pour les rendre cohérents avec ceux de la comptabilité nationale. La mise en place de ces enquêtes a duré près de dix ans pour obtenir ce qui devaient devenir les enquêtes annuelles d'entreprises (EAE) (Raulin, 1992) ; la totalité du champ n'a été couvert qu'en 1983.

Les données collectées via les EAE concernaient principalement l'activité de l'entreprise et les facteurs de production. Elles étaient complétées par des questions sur les ventes par famille de produits ou par activité à un niveau fin et par une interrogation associée des établissements (effectifs, rémunérations). Selon les secteurs, les questionnaires permettaient de collecter des données spécifiques d'activité. À l'origine, ces enquêtes étaient de la responsabilité des services enquêteurs dépendant des services statistiques ministériels (SSM) concernés : industrie, construction, transports et agriculture. Néanmoins, l'Insee était en charge des EAE « Services » et « Commerce ».



Pour assurer la cohérence des statistiques issues de plusieurs sources, avec l'expérience des recensements, l'utilité de la comparaison avec les sources fiscales au niveau individuel était évidente.



Pour assurer la cohérence des statistiques issues de plusieurs sources, avec l'expérience des recensements, l'utilité de la comparaison avec les sources fiscales au niveau individuel était évidente. L'idée d'apparier (Koumarios et alii, 2024) et de confronter systématiquement les déclarations fiscales avec les résultats des EAE pour chaque unité élémentaire faisait partie du projet originel (Quélenec, 1986). Dès 1971, des données des EAE ont été rapprochées des données fiscales, posant ainsi les bases du système

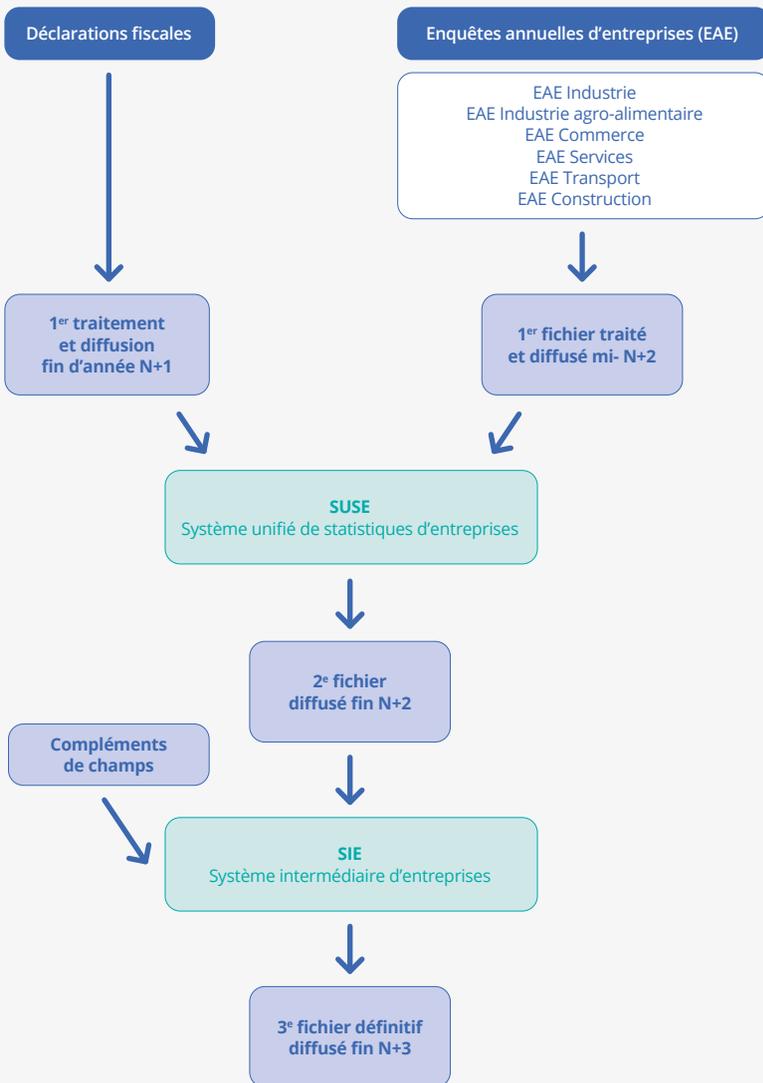
unifié de statistiques d'entreprises (SUSE⁷, **figure 1**) (Camus et alii, 1983). L'objectif était de vérifier et d'assurer la cohérence des données individuelles des EAE et des sources fiscales, pour élaborer les comptes nationaux. Ce système statistique cohérent, stable dans le temps, a été modelé à l'origine pour répondre aux besoins de la comptabilité nationale et a donné le cadre normatif dans lequel il s'inscrit. Entièrement construit pour apparier et rapprocher des données individuelles il pouvait être aussi considéré comme un outil de synthèse microéconomique.

Les objectifs initiaux étaient ambitieux et allaient au-delà de la simple confrontation des sources fiscales et des EAE. Dès le début des années 1970, il avait été envisagé d'étendre le projet SUSE à d'autres sources d'entreprises (déclarations TVA et DADS notamment). Si l'intention première de SUSE était particulièrement simple (le rapprochement de deux fichiers : EAE et source fiscale annuelle), sa mise en œuvre s'est révélée longue et laborieuse. Les incohérences identifiées lors des travaux de rapprochement entre les recensements et les sources fiscales ont exigé beaucoup d'études complémentaires. Ces dernières ont permis de clarifier la gestion des identifiants dans le répertoire Sirène (Coste, 1978). Elles ont permis de définir des règles pour préciser les modalités de

⁷ <https://www.insee.fr/fr/metadonnees/source/serie/s1272>.

création ainsi que les évènements jalonnant la vie des entreprises et des établissements et également de stabiliser le calcul de l'activité principale exercée. Sa mise en œuvre a aussi exigé la conception et la réalisation d'une batterie de contrôles et de redressements au niveau individuel, ainsi qu'au niveau agrégé pour assurer la cohérence : interne au sein de chaque source, en évolution entre deux exercices pour chaque unité, mais aussi au niveau sectoriel. Et ce n'est qu'à la fin des années 80 que SUSE est vraiment entré en rythme de croisière, en se contentant uniquement du rapprochement source

► **Figure 1 - SUSE : le système unifié de statistiques d'entreprises, en vigueur des années 1980 aux années 2010**



fiscale annuelle/EAE. Le rapprochement avec d'autres sources administratives ne s'est concrétisé que beaucoup plus tard, avec la mise en place du système d'élaboration des statistiques annuelles d'entreprises (**Ésane**⁸) dans le courant des années 2010. Néanmoins, cette première étape a permis de sécuriser le processus de production des comptes nationaux en permettant de qualifier et redresser la source fiscale et de collecter des informations au niveau branche.

Sur un plan technique, plusieurs facteurs fondamentaux ont favorisé, dans le courant des années 70 et au début des années 80, l'exploitation de ces données :

- L'introduction des identifiants du répertoire Sirene dans tous les fichiers administratifs. Cette utilisation d'un unique identifiant pour l'ensemble des administrations, géré par l'Insee, est une particularité française et a beaucoup facilité les appariements de fichiers et donc la mobilisation des données administratives pour la production de données statistiques ;
- La normalisation comptable qui aboutira au Plan comptable révisé de 1982⁹. L'introduction de nouveaux concepts, proches de ceux de la comptabilité nationale, a représenté un net progrès pour l'analyse économique et renforcé considérablement la statistique d'entreprises. Enfin, l'harmonisation européenne a permis la comparaison de statistiques venant de différents pays dès que celles-ci s'appuyaient sur les données comptables ;
- L'aboutissement des réflexions autour des nomenclatures d'activité (Grouthier et Lachize, 2008). Cela a permis de construire une nomenclature d'activité commune et partagée ;
- La standardisation des flux de données entre administrations.

► La mise en place d'Ésane, aboutissement du projet initial

Le système tel qu'il avait été conçu et réalisé durant toutes ces années a évolué. L'idée était d'accroître sa cohérence d'ensemble et notamment d'améliorer la collecte des EAE, en favorisant une plus grande coordination des services enquêteurs. L'interrogation différenciée par secteur d'entreprises pouvait contraindre certaines d'entre elles, à cheval sur plusieurs activités (industrie/commerce, commerce/service), à répondre à plusieurs questionnaires annuels. Cela complexifiait la collecte mais aussi le traitement des données et s'avérait être peu lisible pour les entreprises enquêtées. Une véritable coordination de l'ensemble des EAE devenait indispensable. Dans un monde où les entreprises se réorganisent de plus en plus fréquemment, le besoin de mieux appréhender les changements de contour des unités enquêtées était aussi devenu de plus en plus important.

⁸ <https://www.insee.fr/fr/metadonnees/source/serie/s1188>.

⁹ Le Plan comptable général (PCG) est une réglementation de normalisation comptable en France. Il définit les règles comptables applicables aux entreprises domiciliées en France, ainsi que la présentation de leurs comptes financiers, bilans, comptes de résultats, rapports et annexes.



Dans un souci de réduction de la charge auprès des entreprises, les questionnaires ont été simplifiés et se focalisent désormais sur la ventilation du chiffre d'affaires par branche.



Ces diverses réflexions ont abouti au début des années 2000 à la mise en place d'Ésane (*figure 2*), en remplacement du dispositif SUSE/EAE. C'est le système en vigueur en 2024. Si le principe de base est toujours le même, soit la confrontation d'une enquête annuelle et d'une source fiscale au niveau de l'unité élémentaire, la gestion des EAE a été profondément transformée. Dans un souci de réduction de la charge des entreprises, les questionnaires ont été simplifiés et se focalisent désormais sur la ventilation du chiffre d'affaires par branche.

Ne sont demandées que des données non disponibles par ailleurs. Ainsi ne sont plus collectées les données :

- comptables, celles-ci pouvant être mobilisées via les liasses fiscales ;
- d'emploi, provenant des sources administratives DADS puis DSN (déclarations sociales nominatives), et non plus de l'enquête annuelle.

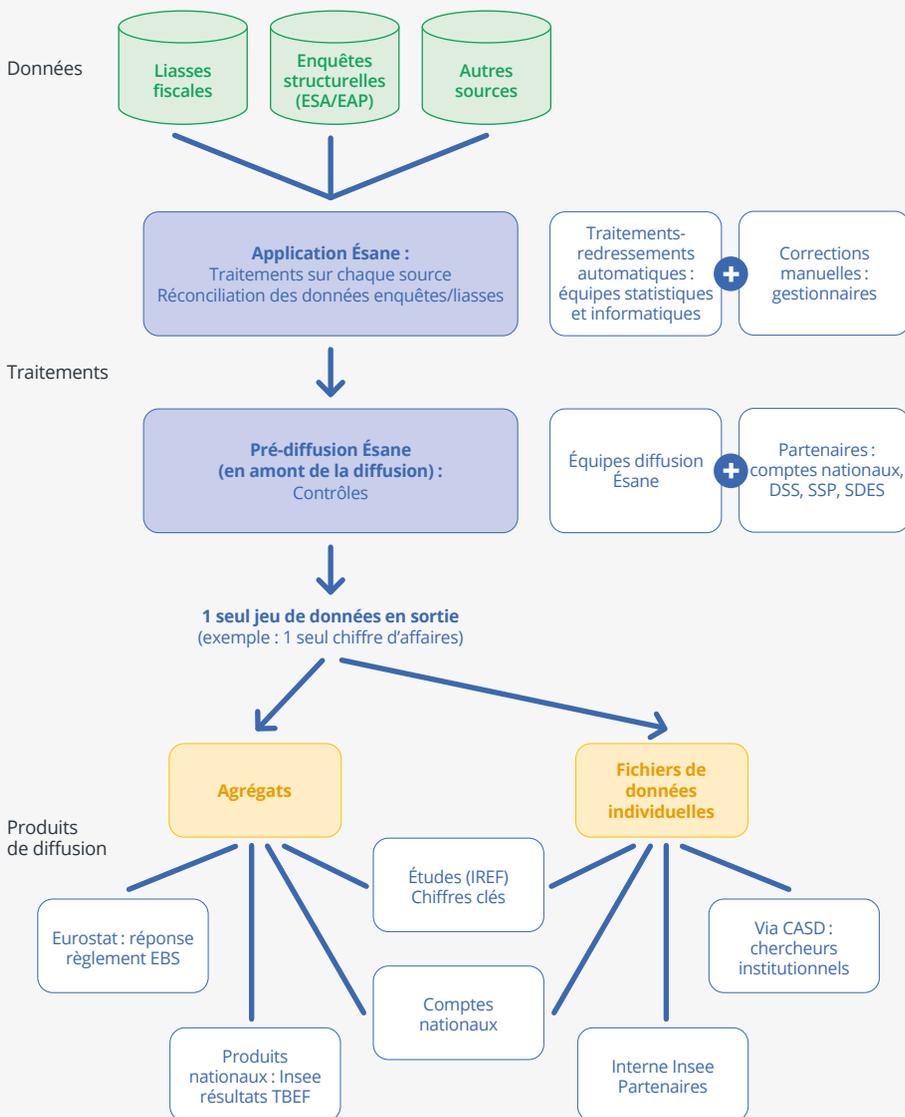
Et enfin la gestion de l'enquête a été entièrement centralisée à l'Insee.

Outre la diminution de la charge de collecte et la simplification des travaux de production statistique, l'objectif du projet était de raccourcir sensiblement les délais de mise à disposition des données et de rationaliser leur diffusion. En effet jusqu'ici, que ce soit au niveau des agrégats ou au niveau des données individuelles, étaient diffusés des fichiers issus des sources fiscales et de l'EAE, et des fichiers comprenant des données résultant du rapprochement des deux. Suivant les usages possibles, travailler sur l'un ou l'autre fichier pouvait avoir un sens pour des experts mais ce système rendait l'accès aux informations complexe et peu compréhensible pour des utilisateurs non experts. Désormais ne sont diffusés que des jeux de données issues du rapprochement entre la source fiscale et les enquêtes.

► Une nouvelle unité d'observation

Enfin, un changement majeur s'opère par rapport aux systèmes précédents : l'unité d'observation n'est plus systématiquement l'unité disposant d'un numéro Siren (l'unité légale, UL), mais l'entreprise au sens de la LME (encadré 3). Ceci nécessite lors des campagnes de production Ésane de définir le contour des entreprises dans un premier temps et d'élaborer des comptes consolidés sur ces contours dans un deuxième temps. En 2022, environ 726 000 UL sont concernées. Une partie de ce travail est faite automatiquement, via des algorithmes, mais pour les plus grandes entreprises susceptibles d'avoir un impact sur le calcul des agrégats et d'avoir une activité à l'international, un travail « individuel » doit être fait par un expert, semblable à celui décrit précédemment pour l'enseigne Grand Frais ou les grands constructeurs automobile. C'est ce qu'on appelle le profilage, qui consiste donc à identifier des contours d'un groupe et de son organisation en unités économiques autonomes, soit les entreprises au sens de la LME (Béguin et Haag, 2017 ; Moreau, 2023).

► **Figure 2 - Ésane : l'élaboration des statistiques annuelles d'entreprises**



EAP : Enquête annuelle de production.
 ESA : Enquête sectorielle annuelle.
 DSS : Département des synthèses sectorielles de l'Insee.
 SSP : Service de la statistique et de la prospective du ministère chargé de l'agriculture.
 SDES : Service des données et études statistiques du commissariat général au développement durable.
 Règlement EBS (*European business statistics*) : Statistiques européennes d'entreprises.
 TBEF : Tableau de bord de l'économie française.
 IREF : Insee Références : les entreprises en France.
 CASD : Centre d'accès sécurisé aux données.



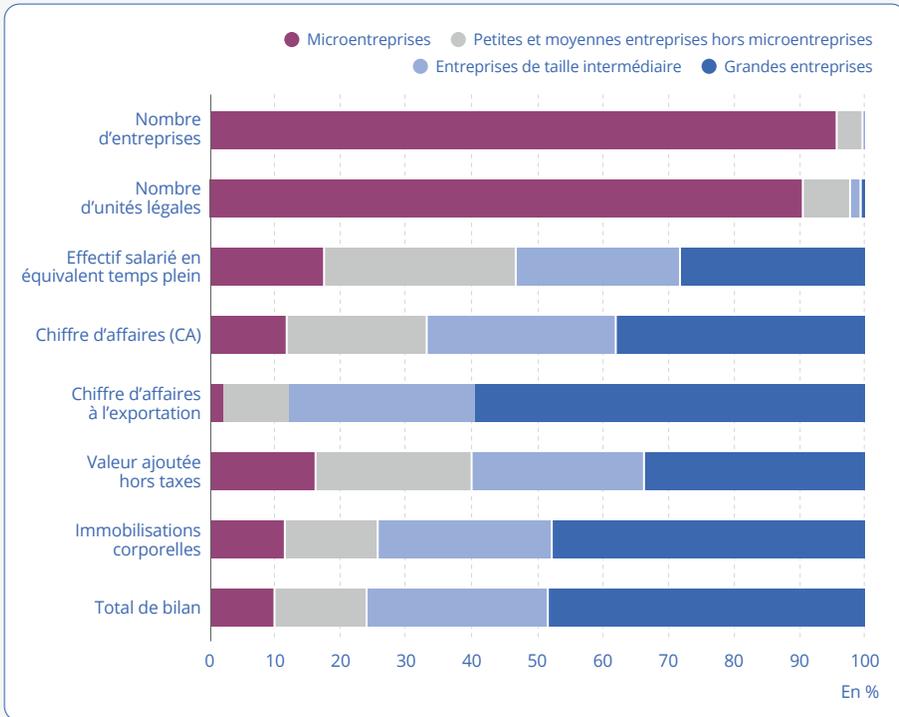
Avec la nouvelle approche de l'entreprise au sens économique, la photographie du tissu productif a été en partie renouvelée. Ce dernier apparaît plus concentré, avec une contribution accrue des grandes entreprises.



Avec la nouvelle approche de l'entreprise au sens économique, la photographie du tissu productif a été en partie renouvelée. Ce dernier apparaît plus concentré, avec une contribution accrue des grandes entreprises. En particulier, les performances de l'industrie sont modifiées : le taux de valeur ajoutée (rapport de la valeur ajoutée sur le chiffre d'affaires) est rehaussé de près de trois points¹⁰, une partie des activités de services et de commerce adhérentes

aux unités industrielles étant, avec la mise en place du concept d'entreprise, rattachée à l'industrie (*figure 3*).

► **Figure 3 - Répartition de différents agrégats selon la catégorie d'entreprises en 2022**



Lecture : En 2022, les grandes entreprises réalisent 34 % de la valeur ajoutée.

Champ : France, entreprises des secteurs principalement marchands non agricoles et non financiers.

Source : Insee, É sane 2022.

¹⁰ Cette hausse s'explique par la prise en compte dans les entreprises du secteur de l'industrie d'unités légales relevant des services ou du commerce qui travaillent majoritairement pour une UL industrielle.

Cinquante ans après le début de sa conception, le Système Statistique d'Entreprises (SSE) tel qu'il avait été rêvé par les statisticiens d'entreprises est maintenant totalement achevé. En 2024, on dispose d'un système au niveau des UL pour apparier différentes sources administratives (déclarations fiscales, DSN, liaisons financières (LIFI)) avec une enquête unifiée qui collecte un tronc commun de variables pour l'ensemble des secteurs d'activité ainsi que quelques variables sectorielles pour répondre aux besoins des SSM, permettant ainsi d'enrichir les données disponibles par ailleurs. Ces fichiers représentent toujours un des principaux intrants de la mécanique des comptes nationaux.

Afin de coller à l'actualité économique, de décrire de façon pertinente le système productif, mais aussi de saisir toute opportunité d'accès à des sources de données jusqu'ici inaccessibles et qui pourraient enrichir la vision du système productif, ce dispositif d'élaboration des données évoluera de nouveau probablement. Les questionnaires de l'enquête annuelle sont modifiés (**encadré 5**). Ces dernières années, ils ont été profondément simplifiés. En conséquence, les questions sectorielles, héritage des anciennes EAE, ont en grande partie disparu, et d'autres sont apparues pour répondre à des besoins nouveaux (le détail des consommations d'énergie pour l'ensemble des secteurs est à l'étude). Jusqu'ici, seuls les secteurs des transports et de l'industrie faisaient l'objet d'une collecte détaillée sur ses aspects.

► Encadré 5. Quelques chiffres

La statistique annuelle d'entreprises en 2024, c'est :

- 141 000 questionnaires envoyés ;
- 83 000 questionnaires renseignés ;
- 35 000 questionnaires traités manuellement par les gestionnaires ;
- 4 millions de déclarations fiscales, dont 1 million relevant de microentreprises ;
- 5 millions de lignes et 200 variables diffusées chaque année ;
- une centaine d'agents en charge de la production des données.

► Un gisement de données encore sous-utilisé

S'ils améliorent la qualité des données, ces enrichissements constants ont aussi pour conséquence une rupture statistique qui rend difficiles les comparaisons temporelles. Les sources se développant continuellement, les méthodes évoluant pour tenir compte des nouvelles technologies, la comparaison interannuelle des résultats s'avère quasiment impossible sans retraitement important des fichiers individuels.

Les comptes nationaux prennent en compte ces changements méthodologiques, considérés comme des gains de qualité, au moment des changements de base. Il est donc nécessaire, dans la production courante des comptes, de retraiter les données annuelles d'entreprises pour disposer de données comparables d'une année à l'autre. Dans le cadre de la diffusion des résultats d'Ésane, les fichiers de données individuelles ou agrégées prennent en compte les évolutions de méthodes au moment où elles sont mises en œuvre. Ainsi la connaissance des liens financiers entre les entreprises s'est beaucoup enrichie depuis 2018 et l'information disponible a été prise en compte immédiatement dans ESANE alors que dans le cadre des comptes, elle n'a été prise que lors de la bascule dans la base 2020.

Pour permettre les comparaisons entre deux années successives, chaque millésime est diffusé deux fois : une première fois, l'année de leur millésime avec les processus tels qu'ils existent, et l'année suivante, en prenant en compte les évolutions méthodologiques apparues depuis, ce qui permet de les comparer aux données de l'année suivante. En outre, les contours des unités d'observation des données recalculées l'année suivante le sont suivant les nouveaux contours des entreprises. Cette solution, mise en œuvre pour pallier les ruptures d'évolution d'une année à l'autre, est difficile à appréhender pour des utilisateurs non experts mais permet de répondre, bien qu'imparfaitement, aux demandes d'évolutions sectorielles à des niveaux fins de la nomenclature. Ces données sont utilisées à des fins d'étude sur des sujets aussi variés que les huiles essentielles (Boniou, 2022), la sécurité privée (Gallot, 2021) ou encore l'industrie du meuble (Biot, 2019).



**Néanmoins,
la valorisation
de ces données reste
un des challenges
des années à venir.**



Néanmoins, la valorisation de ces données reste un des challenges des années à venir. Intrinsèquement, la réalité des entreprises est complexe, les modes d'organisation sont de plus en plus sophistiqués, les enjeux auxquels elles doivent faire face sont multiples, et par ailleurs, les données qui permettent d'enrichir notre connaissance du monde productif sont nombreuses et compliquées à traiter et à comprendre. Par nature, elles s'adressent à un public d'experts.

Sans être aussi « grand public » que d'autres productions de l'Institut, ces résultats sont néanmoins à même d'intéresser de nombreux acteurs professionnels, d'alimenter le débat public et de permettre d'objectiver les questions qui font débat aujourd'hui, comme celles autour de la réindustrialisation, de l'évolution de la productivité ou de l'estimation des marges. L'Insee a un rôle de diffusion de ces données, mais aussi d'aide à leur utilisation.

Les fichiers de données individuelles sont mis à disposition des chercheurs via le CASD¹¹ et font partie de ceux les plus utilisés. Ils présentent une richesse d'informations quasi sans équivalent au monde¹². Ils permettent de réaliser des études longitudinales sur des problématiques variées. De plus, il est possible de remonter jusqu'aux années 70. Bénéficiant de l'exhaustivité des sources fiscales, qu'on peut enrichir par des données sur des thématiques spécifiques, les informations issues du système de production de données annuelles d'entreprises, qu'elles soient individuelles ou agrégées, offrent des possibilités d'exploitation à des niveaux géographiques fins et à des niveaux très détaillés de la nomenclature d'activité. Cependant, leur exploitation nécessite une connaissance fine des richesses et des limites de ces données ; à cet effet, l'Insee anime régulièrement des formations à destination de communautés de chercheurs pour aider à leur utilisation.

¹¹ CASD : Centre d'accès sécurisé aux données. Ce groupement d'intérêt public a pour objet principal d'organiser et de mettre en œuvre des services d'accès sécurisé pour les données confidentielles à des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation (Gadouche, 2019).

¹² Dans de nombreux pays, les données fiscales ne sont pas accessibles aux statisticiens et le système d'identifiants uniques pour les entreprises n'existe pas. La statistique d'entreprises sert à alimenter la comptabilité nationale et se base sur des enquêtes échantillonnées. L'exhaustivité des données disponibles et accessibles aux statisticiens et aux chercheurs en France est quasiment unique au monde.



Dans les années à venir, d'autres sources de données administratives pourraient bouleverser le paysage.



Par ailleurs, les opportunités de travail à partir des fichiers administratifs de la DGFIP¹³ ont crû de façon importante. Ainsi, pour la détermination des groupes, et par là même du contour en entreprise, le passage de l'enquête LIFI aux fichiers des impôts du même nom a considérablement enrichi les données. Dans les années à venir, d'autres sources de données administratives pourraient bouleverser le paysage. Ainsi, la facturation électronique, quand elle sera totalement opérationnelle, devrait permettre de

disposer via les sources fiscales d'un niveau de détail plus fin qui pourrait enrichir sensiblement l'offre de données annuelles d'entreprises mais aussi considérablement modifier le système de collecte des données.

Dans le futur, la statistique d'entreprises doit faire face à deux enjeux fondamentaux :

- La mesure de l'impact des questions environnementales sur le fonctionnement et l'activité des entreprises est une question à laquelle l'Insee, et plus généralement la statistique publique, s'est intéressée depuis plusieurs années. Il existe des enquêtes spécifiques sur ces sujets : l'enquête sur les consommations d'énergie dans l'industrie (EACEI¹⁴), l'enquête sur les investissements dans l'industrie pour protéger l'environnement (Antipol¹⁵), les enquêtes sur la production de déchets au niveau sectoriel. Elles collectent des données sur des échantillons représentatifs d'entreprises et permettent de construire des indicateurs nationaux. La transposition en France de la directive européenne CSRD¹⁶ doit permettre de disposer de données pour un ensemble plus large d'entreprises. Enrichir les données issues du processus É sane au niveau individuel avec ces données thématiques en profitant de la richesse du champ É sane est un prochain challenge. Si on veut analyser l'impact de normes environnementales ou leur mise en œuvre à des niveaux fins de l'activité, utiliser et apparier les données collectées via la directive européenne CSRD est sans aucun doute une solution à étudier.
- L'autre point à investir est celui du champ de la statistique d'entreprises. En effet, celui-ci regroupe les sociétés et entreprises individuelles, marchandes et productives ou participant au système productif, non agricoles et non financières (mais y compris les *holdings* et les auxiliaires financiers et d'assurances). Ce périmètre historique est la conséquence des modes de production des données. Or l'observation du champ productif en concept d'entreprises conduit à travailler sur des unités qui relèvent certes du champ historique mais peuvent avoir des activités financières ou agricoles traditionnellement hors champ. Par ailleurs, disposer d'une vision de l'ensemble de l'activité économique sur le territoire national, des liens entre les différents acteurs, des modes de financement des entreprises, est un enjeu fort auquel la statistique d'entreprises doit apporter des réponses.

¹³ DGFIP : La Direction générale des Finances publiques est une direction de l'administration publique centrale française qui dépend du ministère chargé de l'économie.

¹⁴ <https://www.insee.fr/fr/metadonnees/source/serie/s1214>.

¹⁵ <https://www.insee.fr/fr/metadonnees/source/serie/s1232>.

¹⁶ Applicable depuis le 1^{er} janvier 2024, la directive européenne *Corporate Sustainability Reporting Directive* (CSRD) fixe de nouvelles normes et obligations de reporting extra-financier. Elle concerne les grandes entreprises et les PME cotées en bourse. L'objectif de cette directive est d'encourager le développement durable des entreprises. Voir les références juridiques en fin d'article.

“

Depuis plusieurs années, l'explosion des plateformes et des services d'intermédiation et plus généralement la prise en compte de l'immatériel posent question aux statisticiens.

”

Ces champs d'investissement semblent pouvoir être investigués sans que la nature même du dispositif d'information tel qu'il a été conçu il y a plus de 50 ans soit remis en cause. La possibilité d'apparier des données fiscales et des enquêtes statistiques via un identifiant unique et largement partagé permet a priori de répondre à ces demandes sur l'impact environnemental ou sur

l'élargissement du champ. Le système est suffisamment résilient pour prendre en compte de nouvelles sources, voire de nouvelles unités, à condition qu'elles s'articulent avec les unités déjà présentes dans le système. Or, depuis plusieurs années, l'explosion des plateformes et des services d'intermédiation et plus généralement la prise en compte de l'immatériel posent question aux statisticiens. Ces entreprises n'ont pas d'ancrage territorial (Booking par exemple). Comment rendre compte de leur activité sur le territoire national ? Des liens qu'elles ont avec d'autres entreprises productrices ? De leur apport à la richesse nationale ? Ces enjeux sont importants pour la comptabilité nationale, et par voie de conséquence, pour ce système d'observation des entreprises, conçu à l'origine pour répondre aux besoins des comptes nationaux. C'est là un des défis majeurs de la statistique d'entreprises dans les prochaines années.

► Fondements juridiques

- Règlement (UE) 2019/2152 du Parlement européen et du Conseil du 27 novembre 2019 relatif aux statistiques européennes d'entreprises, abrogeant dix actes juridiques dans le domaine des statistiques d'entreprises. [en ligne]. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1730207160102&uri=CELEX%3A32019R2152>.
- Directive (UE) 2022/2464 du Parlement européen et du Conseil du 14 décembre 2022 modifiant le règlement (UE) n° 537/2014 et les directives 2004/109/CE, 2006/43/CE et 2013/34/UE en ce qui concerne la publication d'informations en matière de durabilité par les entreprises. [en ligne]. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32022L2464>.
- Loi n° 2008-776 du 4 août 2008 de modernisation de l'économie. In : *site de Légifrance*. [en ligne]. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000019283050>.

► Bibliographie

- BÉGUIN, Jean-Marc et HAAG, Olivier, 2017. Méthodologie de la statistique annuelle d'entreprises – Description du système « É sane ». In : *Insee Méthodes*. [en ligne]. Octobre 2017. Insee. N° N130 [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/3056089>.
- BIOT, Sylvie, 2019. L'industrie française du meuble face à la concurrence européenne et chinoise. In : *Insee Focus*. [en ligne]. 9 octobre 2019. Insee. N° 164. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4225112>.
- BONIOU, Anne, 2022. La France, second pays producteur d'huiles essentielles de l'Union européenne. In : *Insee Focus*. [en ligne]. 27 avril 2022. Insee. N° 266. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/6434102>.
- CAMUS, Benjamin, FERRÉ, Thérèse, ROUSSET, Mireille et TAMISIER, Marie-Hélène, 1983. SUSE, système unifié de statistiques d'entreprises (Sources, méthodes et apports). In : *Les collections de l'Insee*. [en ligne]. Septembre 1983. Insee. Série E, Entreprises N° 86 [Consulté le 13 août 2024]. Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k97544683/f1.item>.
- CARROUÉ, Laurent, 2013. *La France : les mutations des systèmes productifs*. Collection U, Armand Colin. ISBN 978-2200287962.
- COLIN, Christel, 2019. FRIBS : un nouveau cadre commun pour les statistiques d'entreprises européennes. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 110-124. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254231?sommaire=4254170>.
- COSTE, Hubert, 1978. L'opération SIRENE – Système Informatique pour le Répertoire des Entreprises et des Établissements. In : *Courrier des statistiques*. [en ligne]. Janvier 1978. Insee. N° 5, pp. 21-29. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06z97k0/f1.pdf>.
- 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. Insee. N° N3, pp. 76-92. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254227?sommaire=4254170>.
- GALLOT, Philippe, 2021. La sécurité privée modérément affectée par la crise, des taux de marge toujours faibles. In : *Insee Focus*. [en ligne]. 26 octobre 2021. Insee. N° 253. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5758520>.
- GROUTHIER, Jean-Philippe et LACHIZE, Jean-Paul, 2008. Le changement de nomenclatures d'activités et de produits dans l'industrie. In : *Courrier des statistiques*. Novembre - Décembre 2008. Insee. N° 125, pp. 67-70. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06xt47h/f1.pdf>.

- HAAG, Olivier, 2019. Le profilage à l'Insee : une identification plus pertinente des acteurs économiques. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. Insee. N° N2, pp. 86-102. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4168409?sommaire=4168411>.
- HACHID, Ali et LECLAIR, Marie, 2022. Sirius, le répertoire d'entreprises au service du statisticien. In : *Courrier des statistiques*. [en ligne]. 29 novembre 2022. Insee. N° N8, pp. 115-130. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/6665192?sommaire=6665196>.
- 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 13 août 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8203044?sommaire=8203072>.
- MOREAU, Gérard, 2023. Les grandes entreprises sous leur meilleur profil. In : *Le Blog de l'Insee*. [en ligne]. 17 novembre 2023. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://blog.insee.fr/les-grandes-entreprises-sous-leur-meilleur-profil/>.
- QUÉLENNEC, Michel, 1986. Les statistiques d'entreprises (Réalités observées et méthodes d'observation). In : *Les collections de l'Insee*. Juin 1986. Insee. Série E, Entreprises. N° 101. [Consulté le 27 septembre 2024]. Disponible à l'adresse : <https://gallica.bnf.fr/ark:/12148/bpt6k9757017b?rk=21459;2>.
- RAULIN, Emmanuel, 1992. Pour une nouvelle génération d'enquêtes annuelles d'entreprises. In : *Courrier des statistiques*. [en ligne]. Décembre 1992. Insee. N° 64, pp. 29-38. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06z98v2.pdf?download=1>.
- RIVIÈRE, Pascal, 1998. En quoi la statistique d'entreprise est-elle spécifique ? In : *Courrier des statistiques*. [en ligne]. Juin 1998. Insee. N° 85-86, pp. 3-18. [Consulté le 13 août 2024]. Disponible à l'adresse : <https://www.bnsp.insee.fr/ark:/12148/bc6p06z99f7/f1.pdf>.
- VOLLE, Michel, 1982. *Histoire de la statistique industrielle*. Economica. ISBN 2-7178-0520-6.
- VOLLE, Michel, 1984. *Le métier de statisticien*. Economica. ISBN 2-7178-0824-8.

L'économie racontée par les données bancaires

Ce que nos relevés de comptes disent de nous



Odran Bonnet* et Tristan Loisel**

Depuis la crise sanitaire en 2020, l'Insee bénéficie d'un accès à des échantillons anonymisés de comptes bancaires de La Banque Postale et du Crédit Mutuel Alliance Fédérale. Ces données, riches en informations, ont permis de documenter l'évolution de la situation financière des ménages en temps réel lors de la crise inédite de la Covid-19, puis lors de l'épisode inflationniste en 2022. Elles ont également permis de documenter la situation quotidienne des ménages, mettant en lumière des épisodes de découvert en fin de mois, et d'évaluer une politique publique en mesurant les impacts financiers, distributifs et environnementaux de la remise à la pompe instaurée en 2022 à la suite de la hausse des prix du carburant.

Les données bancaires sont une mine d'informations précieuses, mais leur utilisation soulève de nombreux défis. Après quatre années d'utilisation, un premier bilan de leur exploitation est proposé dans cet article. Comment construire des concepts qui ont un sens économique à partir de ces données ? Comment s'assurer de leur représentativité ? Quels sont les apports de ces données pour la statistique publique ? Si elles ne permettent pas de remplacer les données d'enquêtes et fiscales, elles complètent les analyses conjoncturelles. En outre, elles permettent de répondre à des questions anciennes sur la consommation et l'épargne des ménages auxquelles les sources traditionnelles ne pouvaient apporter de réponses. Dans cet article, sont détaillés les enjeux et les difficultés pour utiliser des données qui proviennent d'acteurs privés et n'ont pas été produites à des fins statistiques, mais à des fins de gestion par les banques.

 Since the health crisis in 2020, the National Institute of Statistics and Economic Studies (INSEE) has had access to anonymized samples of bank accounts from La Banque Postale and Crédit Mutuel Alliance Fédérale. These data, rich in information, have enabled the analysis of the real-time evolution of households' financial situation during both the unprecedented COVID-19 crisis and the inflationary episode in 2022. They have also paved the way for the analysis of households' daily situation, highlighting instances of overdrafts at month-end, and for the evaluation of a public policy by measuring the financial, distributive, and environmental impacts of the fuel subsidy implemented in 2022 following the rise in fuel prices.

Bank data are a goldmine of valuable information, but their use raises many challenges. After four years of use, a first assessment of their exploitation is presented in this paper. How can we build economically meaningful concepts from these data? How can we ensure their representativeness? What are the benefits of these data to public statistics? While they do not replace survey and tax data, they complement cyclical analyses. Moreover, they allow us to address long-standing questions about household consumption and savings that traditional sources could not answer. This paper details the challenges and difficulties in using private data that were not produced for statistical purposes but for the management of bank operations.

* Chargé d'études, Département des études économiques, Dese, Insee.
odran.bonnet@insee.fr

** Chargé d'études, Département des études économiques, Dese, Insee.
tristan.loisel@insee.fr

Les relevés de comptes bancaires en disent beaucoup sur nos vies. Lorsque les fins de mois sont difficiles, les soldes des comptes sont au plus bas et les découverts plus fréquents. Lorsque nous perdons notre emploi, un virement régulier de salaire peut être remplacé par un virement de France Travail. Même les départs en vacances ne passent pas inaperçus ; les dépenses de transport et d'hébergement augmentent, de même que les dépenses réalisées à l'étranger. Les données bancaires, très riches en information, permettent donc des exploitations variées.

Dans cet article, nous reviendrons d'abord sur le contexte qui a permis l'accès de l'Insee à ces données ; le mouvement d'ouverture vers des données d'origine privée a été accéléré par la crise sanitaire. Ensuite, nous expliciterons comment ces données sont produites, depuis les transactions des clients jusqu'aux graphiques publiés dans les collections de l'Insee. Enfin, nous détaillerons les défis à relever pour les exploiter et expliquerons les nouvelles informations qu'elles apportent à la statistique publique : comment complètent-elles le diagnostic conjoncturel ? À quelles nouvelles questions économiques permettent-elles de répondre ? Enfin, nous reviendrons sur leurs limites, notamment en matière de représentativité.

► Une aubaine pour décrire la situation financière des ménages pendant la crise sanitaire

En 2020, la crise sanitaire et le confinement à grande échelle de la population ont provoqué un choc sur l'économie aussi soudain qu'inédit. À l'Insee, le suivi de la conjoncture en a été bouleversé, conduisant notamment au recours accru à des données nouvelles, plus rapidement disponibles que les données traditionnelles (Pouget, 2020). Parmi les nouvelles sources mobilisées, les données de comptes bancaires se sont avérées précieuses car très détaillées et rapidement disponibles. Pour l'Insee, ces



Parmi les nouvelles sources mobilisées, les données de comptes bancaires se sont avérées précieuses car très détaillées et rapidement disponibles.



données offraient une réelle opportunité pour conforter et enrichir les informations provenant des sources traditionnelles d'enquête avant que les données de sources fiscales ne soient disponibles. En outre, elles permettaient d'éclairer la situation financière des ménages sous un jour nouveau, grâce à des indicateurs atypiques pour la statistique publique mais très parlants pour le grand public, comme les montants présents sur les comptes en fin de mois ou la proportion de ménages à découvert.

L'intérêt de l'Insee pour ces données s'est concrétisé lorsque certaines banques, dans le contexte de la crise sanitaire, ont souhaité contribuer à l'analyse de la situation économique par la statistique publique. Crédit Mutuel Alliance Fédérale (CMAF) a ainsi permis un accès sécurisé à des données anonymisées de comptes bancaires d'un échantillon de sa clientèle dès 2020¹. Des premières études sur l'évolution de la situation

¹ Le Conseil d'analyse économique (CAE) a demandé et obtenu un accès à des données anonymisées de Crédit Mutuel Alliance Fédérale. À la suite de ce partenariat, l'Insee a formulé la même demande.

financière des ménages pendant la crise sanitaire ont résulté de cette collaboration (Insee Note de conjoncture, 2021). Cependant, utiliser des données provenant d'un seul réseau bancaire pouvait être insuffisant pour décrire l'ensemble de la population. La clientèle de CMAF étant plus aisée que la moyenne, ces données pouvaient notamment s'avérer inadéquates pour documenter la situation financière des plus fragiles. Or, le sujet était brûlant à l'époque, en plein cœur de la crise sanitaire. L'Insee a pris contact avec un second réseau : La Banque Postale (LBP). Il s'agit de l'établissement le plus à même de fournir des informations sur les clients bancarisés les plus fragiles, car le législateur lui a confié une mission d'accessibilité bancaire. Les informations de cette banque ont permis de montrer la stabilité des revenus, malgré la crise, y compris pour les plus modestes. Cela a conforté les estimations obtenues par le modèle de microsimulation Ines² qui concluait à un taux de pauvreté stable. Ce constat, à rebours des discours publics³, rendait le recours à des données externes particulièrement précieux pour convaincre de la solidité des résultats obtenus avec les méthodes usuelles (Tavernier, 2022). Ainsi, la publication sur les données LBP a été programmée le même jour que la publication annuelle portant sur le taux de pauvreté, soit une dizaine de jours avant la journée mondiale de la pauvreté.

La crise inflationniste succédant à la crise sanitaire, les données bancaires ont de nouveau été mobilisées afin d'éclairer l'évolution de la situation financière des ménages dans ce nouveau contexte en 2022 et 2023.

► Une nouvelle utilisation de « big data » privées dans la statistique publique

Si l'accès aux données bancaires a été accéléré par la crise sanitaire, il s'inscrit dans un contexte plus large de nouvelles données massives d'origine privée, comme les données de caisse.

Si l'accès aux données bancaires a été accéléré par la crise sanitaire, il s'inscrit dans un contexte plus large d'utilisation par les instituts nationaux de statistiques de nouvelles données massives d'origine privée, comme les données de caisse (Leclair, 2019). Ces données, appelées communément « big data », peuvent inclure des informations aussi variées que celles d'annonces de logements, de téléphonie mobile ou de compteurs de gaz et d'électricité. Ces données procurent de nombreux avantages par rapport aux sources usuelles comme la haute fréquence ou la granularité (Blanchet et Givord, 2018). Mais il existe également de nombreux obstacles dans leur utilisation. Outre les enjeux de

représentativité, ces données sont peu structurées, et leurs exploitations nécessitent donc d'importants investissements.

² <https://www.insee.fr/fr/information/2021951>.

³ Pendant la crise sanitaire de la Covid, le chiffre d'un « million de pauvres supplémentaires » était repris dans la presse (Le Monde, 2020). Ce chiffre provenait d'associations caritatives. Les outils déployés par la statistique publique pour mesurer le taux de pauvreté (finalement stable en 2020) sont détaillés par la suite (Tavernier, 2022).

Avant d'être transformée en information utile au débat public, la donnée bancaire subit de nombreux traitements, du terminal de paiement aux tableaux ou graphiques des publications, en passant par les systèmes d'information des banques. Ce long chemin n'est pas propre aux données bancaires ou aux données privées et fait écho aux traitements réalisés lors d'utilisation de données administratives (Cotton et Haag, 2023), par exemple celles de propriétés immobilières (André et Meslin, 2022).

Nous allons décrire ce parcours de la donnée et les retraitements réalisés par la banque puis par le statisticien.

► Parcours de la donnée bancaire : du système de gestion privé à une meilleure compréhension des comportements économiques

L'exploitation des données bancaires repose sur des partenariats étroits avec les banques. Le statisticien public ne se contente pas de récupérer et d'exploiter une base préexistante, il doit la construire progressivement et conjointement avec le partenaire bancaire. L'objectif est de transformer des données issues des processus de gestion de comptes bancaires en des tables de données individuelles agrégées au niveau ménage. Cette entreprise nécessite de nombreux échanges avec les banques. D'un côté, le statisticien doit expliciter ses besoins, de l'autre la banque doit détailler le contour des données qu'elle peut mettre à disposition. Les données s'enrichissent au fur et à mesure du partenariat au gré des besoins du statisticien et des contraintes des banques (**encadré 1**). Le parcours de la donnée depuis les opérations réalisées par le client jusqu'à la publication des analyses statistiques se réalise en quatre grandes étapes.

► Encadré 1. Des partenariats étroits avec les banques qui facilitent la montée en puissance de l'utilisation des données bancaires à des fins statistiques

Des échanges réguliers avec les banques ont permis à l'Insee d'améliorer significativement l'exploitabilité des données depuis le début du partenariat. Ces collaborations étroites se sont d'ailleurs traduites par des co-publications avec les banques (de l'Insee avec La Banque Postale et du Conseil d'Analyse Économique (CAE) avec Crédit Mutuel Alliance Fédérale).

Les partenariats avec les banques ont également conduit à des projets d'expériences aléatoires sur certains clients. L'Insee mène actuellement une expérimentation en collaboration avec La Banque Postale, dans le but d'évaluer un dispositif d'accompagnement budgétaire, qui s'inscrit dans le cadre de la mission publique d'accessibilité bancaire de La Banque Postale, destiné aux clients en situation de fragilité financière. Le CAE

a également estimé des propensions marginales à consommer, c'est-à-dire les montants dépensés par les ménages qui bénéficient d'une hausse de revenu, à partir d'une expérience sur des clients du CMAF (Boehm et alii, 2023).

Apparier à d'autres sources (tout en préservant l'anonymat) peut ouvrir d'autres perspectives d'études : grâce à un appariement avec des données publiques de diagnostic de performance énergétique (DPE)*, le CAE a étudié la performance énergétique des logements (Astier et alii, 2024). Au Danemark, des chercheurs ont également étudié la réaction des ménages lors d'une perte d'emploi en appariant les données bancaires des clients d'une grande banque danoise avec les données sur leurs employeurs** (Andersen et alii, 2023).

* Appariement réalisé par la banque à partir de l'adresse des clients.

** L'appariement repose sur une identification de l'employeur par le compte bancaire utilisé pour verser les salaires.

Les deux premières étapes correspondent aux traitements effectués par la banque elle-même. La première consiste à récolter et construire les informations sur ses clients à des fins de gestion à partir des différents types d'opérations réalisées chaque jour sur les comptes de la banque. La seconde étape vise à réaliser les traitements nécessaires afin de mettre à disposition un sous-ensemble de données exploitable par les statisticiens. La table finale doit contenir les variables pertinentes pour le statisticien, sur un champ défini à l'avance, et doit également respecter les critères d'anonymat de la clientèle.

Les deux étapes suivantes sont effectuées par le statisticien. Il va d'abord effectuer des retraitements spécifiques à la source de données afin de constituer une base mobilisable pour une étude. La dernière étape n'est pas spécifique aux données bancaires : une fois les retraitements opérés, le statisticien peut réaliser calculs et graphiques comme il le fait à partir de n'importe quelle autre source.

► Du côté de la banque : des données de gestion privées aux bases de données statistiques

D'où viennent les données bancaires ?

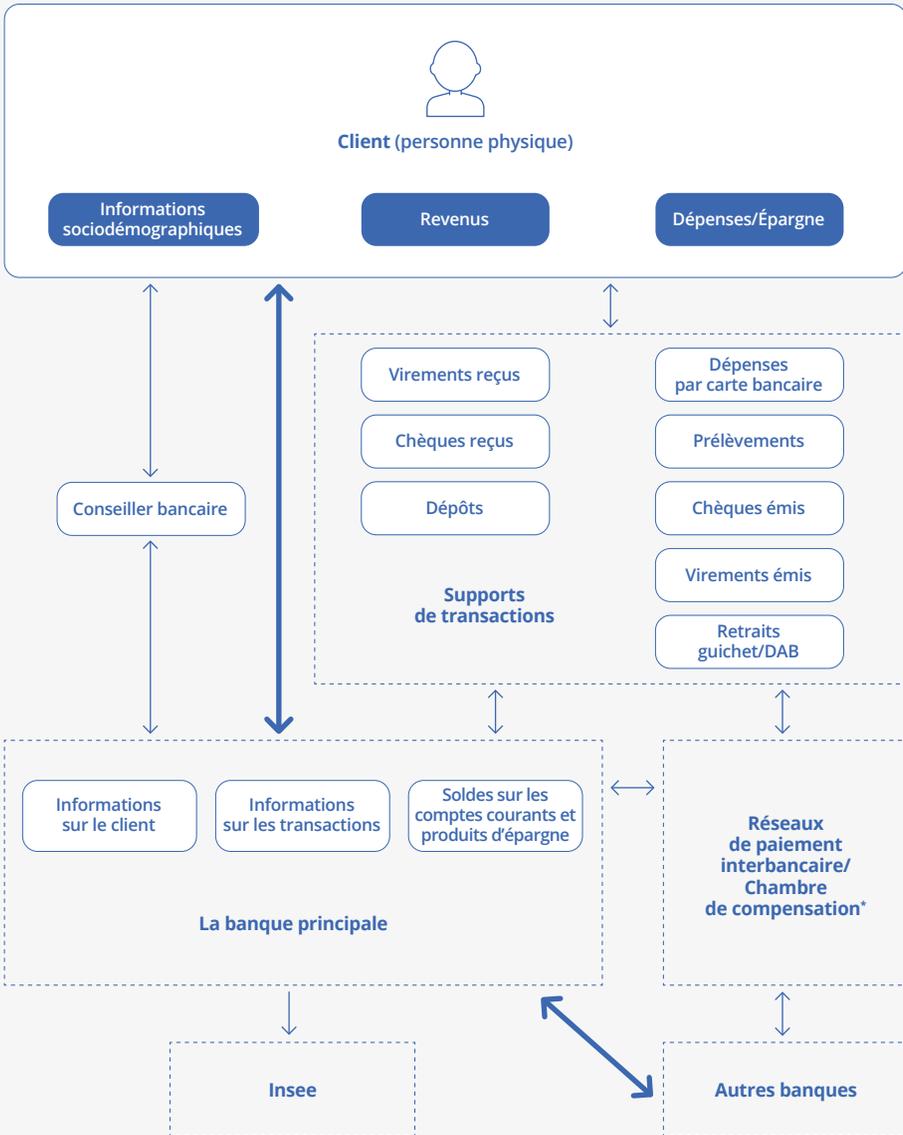
Chaque jour, des millions d'opérations bancaires ont lieu. Toutes ces opérations sont enregistrées par le système d'information bancaire afin que l'argent circule entre les comptes des ménages et des entreprises. Une transaction bancaire génère un flux de données entre la banque émettrice et la banque acquéreur, en passant par plusieurs intermédiaires (*figure 1*). Par exemple, lors d'un paiement par carte chez un commerçant, le traitement du paiement se déroule généralement en trois étapes : l'autorisation, la compensation et le règlement. La phase d'autorisation a lieu au moment où le client insère sa carte dans le terminal de paiement du magasin. La banque du commerçant (banque acquéreur) interroge alors la banque du client (banque émettrice) via le réseau de cartes pour savoir si la transaction est approuvée. La banque émettrice renvoie ensuite une réponse, positive ou négative, qui parvient jusqu'au terminal de paiement et qui permet de valider ou refuser l'opération. En fin de journée, le commerçant transfère à sa banque un fichier récapitulatif

toutes les transactions enregistrées (début de l'étape de compensation). Celle-ci regroupe les informations pour l'ensemble de ses clients et les transmet à la banque émettrice, par l'intermédiaire du réseau de cartes, afin de préparer l'étape de règlement. Lors de cette dernière étape, trois règlements ont lieu (de manière indépendante, et pas nécessairement dans cet ordre) : la banque acquéreur reçoit les fonds de la banque émettrice, verse le montant dû au commerçant, tandis que la banque émettrice débite le compte du client.

Une transaction bancaire génère un flux de données entre la banque émettrice et la banque acquéreur, en passant par plusieurs intermédiaires.

Au-delà de ces flux, les banques collectent également des variables sociodémographiques afin de mieux connaître leur clientèle (profession, nombre d'enfants par exemple). Cela leur permet de proposer des services adéquats. Ces

► **Figure 1 - Représentation des flux de données : du client de la banque jusqu'au statisticien**



Légende

↔ Échanges d'informations

↔ Flux monétaires

* Les chambres de compensation sont des infrastructures privées, qui déterminent pour chaque banque le solde à livrer ou à recevoir en fonction de l'ensemble des transactions produites. Ce mécanisme permet d'éviter la multiplication des flux entre les banques.

informations sont traditionnellement collectées lors des rendez-vous avec les conseillers en agence mais également de plus en plus via des questionnaires envoyés par courriel.

Sélection et transformation des données pour répondre aux besoins de la statistique publique

De manière itérative, à la suite de nombreux échanges visant à clarifier les données disponibles et l'expression des besoins de l'Insee, la banque sélectionne les données pertinentes, les anonymise et les transforme selon les étapes suivantes.

Tout d'abord, la banque opère une **sélection des tables et des variables**. Au sein de la masse de données stockées au sein de la banque à des fins de gestion (des centaines, voire des milliers de tables contenant chacune quelques dizaines de variables), la banque sélectionne les informations sur les transactions, les soldes, les crédits, les incidents bancaires, ainsi que les caractéristiques sociodémographiques de ses clients. Les données sont **figées** un mois donné, il arrive donc qu'il manque certaines transactions remontant avec un délai, notamment celles effectuées les derniers jours du mois.

Ensuite, la banque procède au **tirage d'un échantillon**. Une fois ces tables et variables rassemblées, la banque tire un échantillon de clients à mettre à disposition selon une méthodologie préalablement définie. Cette opération d'échantillonnage qui est monnaie courante dans la statistique publique est rare dans le monde bancaire. Le tirage se fait uniquement sur les clients dont la banque estime héberger les revenus et la majorité des opérations de consommation.

Puis la banque opère l'**anonymisation des données**. L'anonymisation (*figure 2*) implique la suppression de variables identifiantes comme le(s) nom(s) et adresse des clients mais également le(s) nom(s) d'émetteurs et les libellés des virements. Or, ces dernières variables sont utiles pour les analyses puisqu'elles permettent de repérer les salaires perçus par les clients, mais aussi les prestations sociales, les allocations chômage, les pensions de retraites, etc. Avant la suppression, la banque extrait donc les informations utiles et non identifiantes de ces variables (type d'émetteur, régularité de l'émetteur, libellé contenant des mots-clés comme « salaire », etc.). Il s'agit d'un compromis qui permet de garantir l'anonymat des clients, tout en limitant la perte d'information (Redor, 2023).

Enfin, la banque opère le **transfert des données**. Elle transmet les données et les métadonnées à l'Insee dans des environnements sécurisés, au format CSV via le Centre d'Accès Sécurisé aux Données (CASD) pour LBP, et sous la forme d'une base de données relationnelle via une machine virtuelle pour CMAF.

► Du côté du statisticien : de la base de données à la publication d'une étude économique

Les données transmises comportent de riches informations contenues dans de multiples tables (*encadré 2*) ; le statisticien doit les retraiter afin de réaliser des études économiques.

► **Figure 2 - L'anonymisation des données**

Identifiant anonyme du client bénéficiaire (clé d'appariement)	Compte émetteur	Compte bénéficiaire	Nom de l'émetteur	Nom du bénéficiaire	Libellé	Date	Montant	Catégorie
1	Compte émetteur 1	Compte bénéficiaire 1	France Travail	Baron Danglars	Allocation	01/02/2024	1256,34	Chômage
2	Compte émetteur 2	Compte bénéficiaire 2	CAF	La marquise de Saint-Méran	Prestation CAF Marseille	05/02/2024	607,75	Prestations sociales
3	Compte émetteur 3	Compte bénéficiaire 3	CNAV	Louis Dantès	Pension	09/02/2024	1482,73	Retraite
4	Compte émetteur 4	Compte bénéficiaire 4	Gaspard Caderousse	Fernand Mondego	Remboursement dettes de jeu	27/02/2024	443,87	Inconnu
5	Compte émetteur 5	Compte bénéficiaire 5	Armateur Morel	Edmond Dantès	Salaire février	27/02/2024	2978,36	Salaire
5	Compte émetteur 6	Compte bénéficiaire 6	Mercédès Herrera	Edmond Dantès	Vide	28/02/2024	200,00	Inconnu

↓ Anonymisation

Identifiant anonyme du bénéficiaire	Date	Montant	Catégorie
1	01/02/2024	1256,34	Chômage
2	05/02/2024	607,75	Prestations sociales
3	09/02/2024	1482,73	Retraite
4	27/02/2024	443,87	Inconnu
5	27/02/2024	2978,36	Salaire
5	28/02/2024	200,00	Inconnu

Légende

- Variables créées par la banque pour anonymiser les données tout en limitant la perte d'information
- Informations sensibles ou variables identifiantes à supprimer avant la mise à disposition des données

Transformer les données transmises en bases statistiques facilement utilisables

Une fois les données reçues, la première étape consiste en des **macro-contrôles internes**. Cela consiste à produire une série de vérifications sur la volumétrie des données afin de s'assurer que le dernier rafraîchissement ne comporte pas d'erreurs grossières. Concrètement, cela revient à vérifier dans les différentes tables que les évolutions du nombre de clients et du nombre de transactions ne sont pas erratiques. Le statisticien ne maîtrisant pas le processus amont, les macro-contrôles sont primordiaux. Il se peut par exemple que les dernières données transmises au statisticien soient incomplètes (comme des transactions manquantes en fin de mois).

► Encadré 2. Au cœur des finances des clients : que contiennent les données transmises par la banque ?

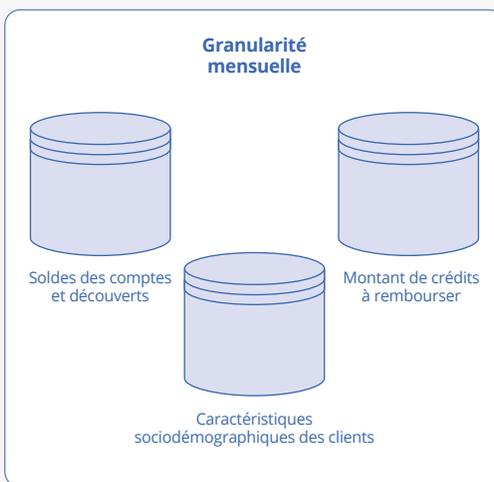
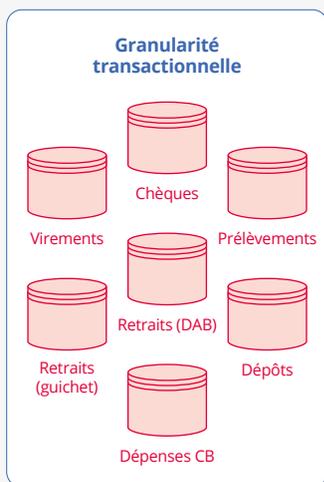
Les données transférées se trouvent dans plusieurs tables selon leurs origines (*figure encadré*) : une table contient le solde des comptes bancaires, d'autres contiennent les transactions par type de mouvement (carte, prélèvement, virement, retraits) et d'autres les caractéristiques des individus titulaires des comptes. Elles couvraient initialement des panels anonymisés de 300 000 clients/groupes familles*.

Pour les deux réseaux bancaires, les données contiennent peu ou prou l'équivalent anonymisé de ce que chacun peut retrouver sur son relevé de compte. Plus précisément, elles incluent : les soldes des comptes en fin de mois pour différents types de comptes (comptes courants individuels ou joints, livrets, assurances-vie et comptes-titres), toutes les transactions effectuées (montants et dates des opérations par carte bancaire, chèques, virements, prélèvements, retraits et dépôts). En outre, elles contiennent quelques informations sociodémographiques connues de la banque : âge, sexe, département, statut marital, type d'habitat et catégorie socioprofessionnelle**.

Cependant, pour des questions de respect de l'anonymat (cf. supra), les données mises à la disposition de l'Insee ne sont pas aussi riches que celles présentes sur nos relevés personnels. Afin de pallier cette perte d'information, les

banques mettent à disposition une catégorisation des opérations. Les dépenses par carte sont catégorisées selon une nomenclature du système bancaire*** qui identifie le type d'établissement récipiendaire du paiement (tel que supermarché, station essence, boulangerie, etc.). Cette nomenclature de 650 postes permet notamment de savoir comment les consommateurs ont modifié leur panier de consommation pendant la crise sanitaire. Les catégorisations de prélèvement permettent d'identifier également les consommations d'énergie et de mesurer l'impact de la crise énergétique sur le portefeuille des ménages français. De plus, depuis début 2020 dans les données de CMAF et mi-2020 dans les données de LBP, les virements entrants provenant de France Travail, des Caisses d'Allocations Familiales (CAF) et des caisses de retraites sont identifiés ; cela permet de repérer les versements de prestations de chômage, de solidarité et les pensions de retraite. Des virements entrants correspondant à des salaires sont également isolés.

Au total, ces informations riches impliquent une volumétrie élevée : de l'ordre d'une dizaine de millions d'opérations mensuelles. Cela requiert d'adapter les traitements à ces données volumineuses.



* La banque regroupe les clients qu'elle considère appartenir à un même ménage dans ce qu'elle nomme un groupe famille. Ces groupes sont constitués à partir de la détention de comptes joints et de l'information collectée par les conseillers.

** Les catégorisations de professions reposent sur des nomenclatures proches de celles utilisées à l'Insee, mais elles ne les recourent pas parfaitement au niveau le plus fin à quatre positions.

*** La nomenclature internationale se nomme Merchant Category Code (MCC).

Ensuite, le statisticien construit à partir des données bancaires **des variables économiques** : les variables correspondant aux concepts à étudier. Par exemple, le patrimoine financier brut d'un groupe famille/foyer est reconstitué à partir de la somme des soldes sur ses comptes.

Les données sont structurées selon une **représentation rectangulaire** familière au chargé d'étude, dans laquelle une ligne correspond à une observation, c'est-à-dire un groupe famille/foyer un mois donné (ou un jour donné selon la fréquence étudiée). Les colonnes contiennent les variables d'intérêt : revenus, dépenses, patrimoine financier et caractéristiques sociodémographiques.

L'échantillon est ensuite **restreint** à la population d'intérêt et un **calage**⁴ est mis en place. Différents filtres peuvent être appliqués afin de conserver uniquement l'échantillon de clients souhaité. De plus, l'échantillon peut être calé sur plusieurs marges (âge quinquennal, sexe, population par département) afin de le rendre davantage représentatif de la population française.

Enfin de nouveaux **macro-contrôles externes** sont réalisés sur les données finales. L'objectif est de confronter les informations et les messages issus des données bancaires à ceux provenant d'autres sources. Idéalement, les données bancaires doivent être confrontées à des données portant sur les mêmes concepts, pour permettre des comparaisons plus directes, et surtout de même fraîcheur. Une première confrontation naturelle et fondamentale est donc la comparaison des données des deux banques. Mais d'autres sources peuvent également être mobilisées, comme les indicateurs de la Banque de France (publiés mensuellement), ou encore les données du Groupement d'Intérêt Économique des Cartes Bancaires (GIE CB), qui permettent de tracer des évolutions agrégées de dépenses par carte bancaire.

Au bout du compte, des données comme les autres ?

Les données bancaires sont alors traitées par le statisticien de la même manière que n'importe quelle autre source de données. Il peut donc effectuer les opérations statistiques et économétriques nécessaires pour répondre à sa problématique. La dernière étape consiste à extraire les résultats agrégés du CASD (pour LBP) ou de la banque (pour CMAF) après vérification de leur conformité (notamment l'absence de données individuelles).

Ces données permettent de répondre à plusieurs problématiques, du suivi conjoncturel à l'évaluation de politiques publiques.

⁴ Le calage sur marges est une technique statistique visant à améliorer la précision des enquêtes par sondage. Elle consiste à modifier les poids de sondage des individus de l'échantillon afin que les totaux pondérés sur l'échantillon de certaines variables correspondent aux totaux connus pour ces variables sur l'ensemble du champ d'observation (la population, le parc de logements, les entreprises). Voir aussi « Miscellanées sur le calage » : <https://www.insee.fr/fr/information/2387498>.

► Voir vite et net : des données transformées pour améliorer le suivi conjoncturel

Pour éclairer la conjoncture, ces données ont le quadruple avantage de la fraîcheur, de la granularité fine, de la grande taille de l'échantillon et de la diversité des informations fournies.

Pour éclairer la conjoncture, ces données ont le quadruple avantage de la **fraîcheur**, de la **granularité fine**, de la grande **taille** de l'échantillon et de la **diversité** des informations fournies. En effet, elles sont mobilisables rapidement : les données du mois M sont disponibles à la fin, voire au milieu du mois M+1. Les opérations sont disponibles au niveau journalier, permettant une analyse en temps réel des réactions des clients face à un choc soudain, comme la baisse des taxes sur le carburant. En outre, les échantillons de clients sont ici suffisamment grands pour focaliser l'étude sur des populations particulières, comme les

bénéficiaires du RSA⁵ pendant la crise sanitaire ou les grands consommateurs de carburant au moment du déclenchement de la guerre en Ukraine. Avec les enquêtes traditionnelles, ces populations ne sont souvent pas suffisamment nombreuses dans les échantillons pour être étudiées spécifiquement. Cet avantage pour la statistique publique a également été souligné lors de l'utilisation de données administratives exhaustives qui permettent d'aller au-delà des décompositions habituelles en dixième de patrimoine et permettent de zoomer jusqu'au centième, voire jusqu'au millième (André et Meslin, 2022). Enfin, les variables sont suffisamment riches pour à la fois mesurer les concepts habituellement mobilisés à l'Insee comme les revenus, la consommation ou l'épargne mais également des grandeurs parlantes pour le grand public comme le solde à la fin du mois.

Ces atouts ont été pleinement exploités lors de la crise sanitaire et de l'épisode inflationniste qui a suivi. Tout d'abord, les données bancaires ont permis de documenter la hausse de l'épargne du fait de la chute de la consommation pendant les confinements (Insee, 2021). Ensuite, elles ont permis de montrer que les revenus de la plupart des clients ont été affectés seulement de manière limitée et temporaire par la crise, mais que certaines populations en marge du marché de l'emploi, comme les allocataires du RSA, ont été davantage pénalisées (Bonnet et alii, 2021). Par ailleurs, plusieurs indicateurs ont été construits permettant de conclure que la précarité financière n'a pas augmenté significativement au cours de la crise sanitaire (**figure 3**). Au début de la période inflationniste en 2022 et de nouveau en 2023, deux études ont montré que la précarité financière, mesurée par quelques indicateurs sur les découverts bancaires, augmente depuis début 2021, mais reste inférieure à son niveau d'avant-crise sanitaire (Adam et alii, 2022, Bonnet et alii, 2023).

En dehors de l'Insee, d'autres exploitations conjoncturelles des données bancaires ont été publiées en France. Par exemple, les données CMAF ont été mobilisées par le Conseil d'Analyse Économique pour étudier l'évolution de la situation financière des ménages (Astier et alii, 2024, Fize et alii, 2022a, Fize et alii, 2022b) mais aussi des entreprises (Epaillard et alii, 2021, Fize et alii, 2022b).

⁵ RSA : Le revenu de solidarité active est une prestation de protection sociale française, qui complète les revenus d'une personne démunie ou aux ressources faibles, afin de lui garantir un revenu minimal.

► **Figure 3 - Précarité financière sur le panel de clients de La Banque Postale entre janvier 2019 et juin 2021**



Note : Les bandes vertes correspondent aux périodes de confinement.

Lecture : Fin mars 2020, 44 % des clients ont moins de 3 000 euros sur leurs comptes (précarité en patrimoine). Au cours de ce mois, 6,3 % des clients ont cumulé au moins 5 jours débiteurs hors découvert autorisé (précarité en découvert), 27,6 % ont perçu moins de 1 000 euros (précarité en revenus) et les revenus sont de 1 740 euros en moyenne.

Champ : France, échantillon de clients bancarisés principaux de La Banque Postale présents sur toute la période janvier 2019-juin 2021 après filtrage des comptes inactifs.

Sources : La Banque Postale ; Insee, calculs des auteurs.

► **Mesurer la diversité des ménages : des données massives qui complètent les enquêtes sans pouvoir les remplacer**

En Espagne, des chercheurs soulignent l'importance de ces données, qui permettent d'augmenter la précision des mesures d'inégalités de consommation, lesquelles reposent sinon sur des enquêtes. Leur apport est affirmé dans un document de travail de l'Université de Cambridge : « Une fois organisées selon les principes de la comptabilité nationale, [les données bancaires peuvent] reproduire les statistiques officielles actuelles sur la consommation globale au niveau national avec un haut degré de précision. En raison de la richesse des données de transaction, elles peuvent en outre produire de nouveaux comptes distributionnels de la consommation, qui révèlent des inégalités de consommation plus importantes que ne le suggèrent les enquêtes. » (Buda et alli, 2022).

Elles permettent également d'aller au-delà des limites des données d'enquêtes de consommation (comme l'enquête [Budget de Famille](#)⁶), grâce à la possibilité de suivre les ménages en panel et d'étudier certaines populations spécifiques, qui, sinon, ne seraient pas en effectifs suffisants. De plus, ces enquêtes ne recensent la consommation de certains biens que sur une courte période, ce qui limite la connaissance de la distribution de la consommation. Par exemple, si un ménage répondant n'a pas consommé un bien pendant la période d'interrogation, il est impossible de savoir s'il ne le consomme jamais ou seulement rarement. Disposer d'informations très précises et détaillées est ainsi fondamental pour mesurer la diversité des pertes liées à une hausse des prix du carburant ou de l'énergie, qui peuvent varier très fortement au cours de l'année.

Enfin, les données bancaires peuvent partiellement remédier aux limites liées aux erreurs de déclarations, comme la sous-déclaration du patrimoine financier.

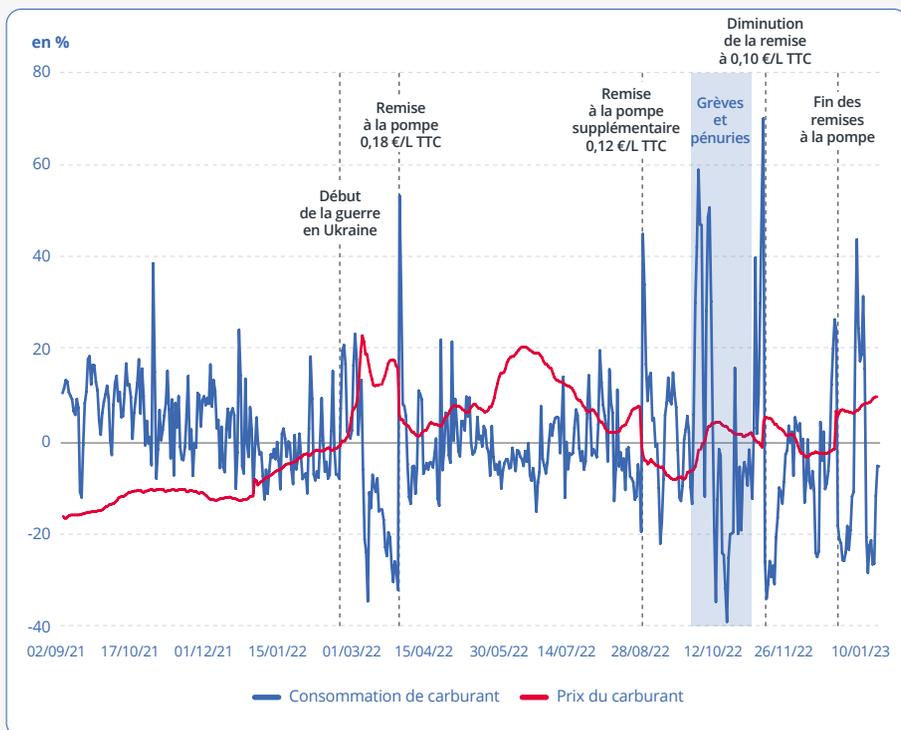
► **Analyser l'évolution de la situation financière des ménages à la suite de chocs de prix ou de revenus**

Les données bancaires ont l'avantage décisif de contenir simultanément des informations quotidiennes sur le revenu, la consommation et l'épargne d'un panel de ménages observé sur plusieurs années.

Pour les études économiques, les données bancaires ont l'avantage décisif de contenir simultanément des informations quotidiennes sur le revenu, la consommation et l'épargne d'un panel de ménages observé sur plusieurs années. L'apport de ces données est souligné pour mesurer la réaction des ménages à des chocs de revenus (Baker et alli, 2020), qu'ils soient permanents (hausse de salaire, passage à la retraite) ou temporaires (épisode de chômage, allocation exceptionnelle, prime inflation). Il est possible de savoir si les ménages peuvent puiser dans leur épargne ou doivent se résoudre à diminuer leur consommation à la suite

6 <https://www.insee.fr/fr/metadonnees/source/serie/s1194>.

► **Figure 4 - Prix et consommation de carburant entre septembre 2021 et janvier 2023, variations quotidiennes par rapport à la moyenne sur la période**



Lecture : Le 2 septembre 2021, les prix du carburant étaient inférieurs de 16,4 % à leur valeur moyenne sur la période (de septembre 2021 à janvier 2023) et les quantités achetées supérieures de 10,6 % aux quantités achetées moyennes sur la même période.

Champ : Échantillon de clients actifs et automobilistes Crédit Mutuel Alliance Fédérale.

Source : Crédit Mutuel Alliance Fédérale, calculs Insee.

d'un choc de revenu. Elles participent donc à l'évaluation de politiques publiques. À ce titre, ces données ont déjà permis d'évaluer les impacts financiers, distributifs et environnementaux des remises sur le carburant mises en place par le gouvernement à la suite des fortes variations de prix consécutives à la guerre en Ukraine (Adam et alii, 2023, *figure 4*).

Ces données ont également permis d'étudier l'évolution de la situation financière des ménages au jour le jour (Bonnet et alii, 2023), illustrant les contraintes de liquidité qui pèsent sur ceux-ci au fil du mois : parmi eux, combien sont dans le rouge la veille du jour de paie (*figure 5*) ? Enfin, elles permettent de mesurer les achats à l'étranger des résidents français, et donc d'étudier les questions de tourisme, d'achats transfrontaliers et d'évitement des taxes pour des biens tels que le tabac (Hillion, 2024) ou le carburant (Adam et alii, 2024).



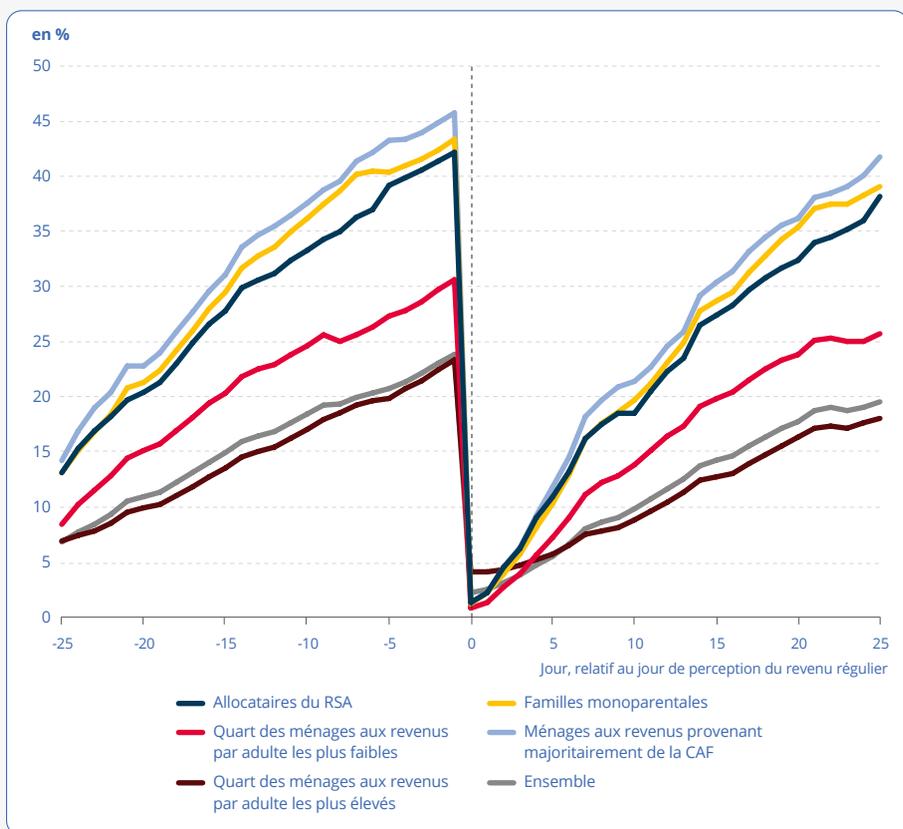
Les données bancaires sont donc une source d'information riche, et ont le grand avantage, par rapport aux enquêtes, d'épargner au statisticien public (et aux comptes publics) les coûts liés au processus de collecte.



Les données bancaires sont donc une source d'information riche, et ont le grand avantage, par rapport aux enquêtes, d'épargner au statisticien public (et aux comptes publics) les coûts liés au processus de collecte. L'inconvénient principal est la perte de la maîtrise du processus (Rivière, 2018). Tout comme les données administratives, les données bancaires existent à des fins de gestion, indépendamment des besoins statistiques. Par conséquent, le champ et les

variables de ces données ne répondent pas nécessairement aux besoins du statisticien. La construction conjointe des bases avec les banques ne permet que partiellement de résoudre ce problème du fait de la nature des données d'origine.

► **Figure 5 - Part des ménages à découvert au fil du mois**



Lecture : 20,7 % des ménages de La Banque Postale percevant un revenu régulier les trois derniers mois sont à découvert 5 jours avant la perception de leur revenu régulier.

Champ : France métropolitaine, échantillon de clients bancarisés principaux à La Banque Postale entre janvier 2019 et août 2023, après filtrage des comptes inactifs et repérage d'un revenu régulier. Période : Mai 2022 - août 2023.

Source : La Banque Postale, calculs Insee.

► La représentativité : étudier la santé financière des Français ou simplement celle d'une clientèle particulière ?

Le champ étudié dans les données bancaires correspond aux clients de la banque et n'est donc pas représentatif de la population française. Il exclut par définition les personnes non titulaires d'un compte bancaire, mais chaque banque peut aussi se spécialiser différemment sur certains segments de la population. Avoir les données de deux banques aux clientèles différentes est un atout pour répondre à cet enjeu. Des comparaisons avec des sources externes permettent également de quantifier l'ampleur du biais. En termes de revenus, de patrimoine et de consommation, la clientèle de CMAF apparaît en moyenne un peu plus aisée que la population générale et celle de LBP un peu moins. Cependant les clientèles des deux banques présentent une diversité importante qui semble couvrir une large partie du spectre économique des ménages français. L'analyse des catégories socioprofessionnelles disponibles dans les fichiers bancaires pointe une surreprésentation des étudiants et une sous-représentation des retraités ; cela peut résulter d'un biais de représentativité ou d'un délai de mise à jour de la situation professionnelle.

Afin de corriger ce biais potentiel, l'échantillon peut être calé sur l'âge quinquennal, le sexe et le département. Ces variables sont les seules pour lesquelles il est certain que la définition est la même entre l'échantillon bancaire et la source de référence. Le calage contribue à tendre vers un échantillon plus représentatif de la population ; mais faute d'avoir les données de l'ensemble des banques, l'échantillon ne peut pas respecter les standards de représentativité des enquêtes. Il est donc impossible de produire des chiffres « officiels » sur l'évolution des revenus, de l'épargne ou de la consommation à partir de ces données bancaires.

► Un texte à trou : les comptes bancaires, une vision partielle du patrimoine des ménages

Outre la question de la représentativité, les informations sur les ménages et leur situation économique provenant des données bancaires sont incomplètes (Baker et alii, 2020). Certains clients peuvent disposer de plusieurs comptes dans différentes banques, ce qui peut conduire à sous-estimer leurs revenus, leurs dépenses, mais surtout leur patrimoine financier (notamment en cas de détention d'assurances-vie). L'enquête *Histoire de Vie et Patrimoine*⁷ mesure l'ampleur du phénomène qui se concentre surtout chez les ménages aux plus hauts revenus : les 10 % de ménages aux plus hauts revenus détiennent 60 % de leur patrimoine financier⁸ au sein de leur banque principale, contre 90 % pour les 10 % aux revenus les plus faibles. De plus, le patrimoine immobilier est inconnu. Pour réduire ce problème de complétude, des chercheurs utilisent parfois des données d'application qui agrègent les comptes des clients des différentes institutions bancaires (Olafsson et Pagel, 2018). Cependant, les utilisateurs d'une telle application bancaire sont probablement moins représentatifs de la population générale que les clients d'une banque dans leur ensemble.

7 <https://www.insee.fr/fr/metadonnees/source/serie/s1005>.

8 La proportion s'élève à 75 % en se restreignant au patrimoine détenu en banque.

► Des statistiques éloquentes, mais éloignées des mesures usuelles de la statistique publique

Contrairement aux enquêtes, les données bancaires ne permettent pas une mesure directe des concepts usuels de la statistique publique. Alors que le statisticien souhaite mesurer les revenus, l'épargne et la consommation, la banque enregistre des flux entrants ou sortants par type de moyen de paiement. Et alors que le statisticien souhaite calculer des statistiques par ménage, les données bancaires permettent, dans le meilleur des cas, un regroupement des membres du ménage bancarisés au sein de l'établissement bancaire considéré.

Précisément, la notion de **ménage**⁹ est approchée par celle de groupe famille, qui repose sur des informations transmises au conseiller et sur la détention de comptes joints. La taille des ménages est donc sous-estimée, car le conjoint et les enfants ne sont pas toujours intégrés dans le ménage, notamment s'ils n'ont pas de compte ouvert dans cette banque¹⁰. Les **revenus** mensuels sont mesurés à partir des flux entrants, mais cette mesure inclut des transferts qui ne correspondent pas tous à des revenus, tels que des transferts entre comptes d'un même individu dans des banques différentes. La catégorisation des virements (pension de retraite, allocation de chômage, prestation sociale, salaire, etc.) distingue certains types de revenus au sein des virements entrants. Mais si le repérage des revenus de remplacement semble fiable, comme les revenus d'allocations chômage qui proviennent de France Travail, celui des revenus d'activité l'est moins, tout comme celui des revenus du capital ; par exemple, les loyers versés entre particuliers ne peuvent être distingués de transferts réguliers entre membres d'une même famille n'appartenant pas à la même banque. Les **dépenses** mensuelles peuvent être mesurées par la somme des dépenses par carte¹¹, des retraits (au distributeur ou au guichet), et de certains prélèvements¹². Le fait d'exclure les chèques et les virements émis conduit à sous-estimer les dépenses. Cependant, inclure tous les virements, chèques et prélèvements conduirait à surestimer les dépenses, tel qu'en cas de virement à soi-même ou de prélèvement

de remboursement de crédit. La nature des dépenses n'est pas toujours facile à identifier. La catégorisation des dépenses par carte ne permet qu'imparfaitement d'identifier la structure du panier de consommation. Seul le type de vendeur est connu mais pas le détail des biens achetés. Or, un supermarché, par exemple, vend des biens très divers.

Les exploitations des données bancaires sont encore récentes. Elles se sont révélées précieuses lors de la période exceptionnelle de la crise sanitaire, mais la statistique publique doit encore prouver sa capacité à les utiliser dans le cadre d'un usage conjoncturel pérenne.

Les exploitations des données bancaires sont encore récentes. Elles se sont révélées précieuses lors de la période exceptionnelle de la crise sanitaire, mais nous devons encore prouver notre capacité à les utiliser dans le cadre d'un usage conjoncturel pérenne.

9 <https://www.insee.fr/fr/metadonnees/definition/c1879>.

10 En outre, à CMAF, un nouveau groupe client est automatiquement créé quand un enfant atteint l'âge de 18 ans, même s'il réside toujours chez ses parents.

11 La nomenclature MCC identifie le type d'établissement récipiendaire de la transaction (station essence, boulangerie, mais aussi supermarché, etc.) : le détail des biens consommés au sein de ces établissements n'est pas observé.

12 Les prélèvements relatifs aux crédits ou aux impôts sont exclus, car il ne s'agit pas de dépenses de consommation.

► **En temps de crise les évolutions écrasent le bruit des données, mais que peut-on mesurer par temps calme ?** —

Les questions de volatilité des indicateurs, de représentativité partielle, ou de complétude peuvent apparaître secondaires dans l'étude d'un choc aussi violent que la crise sanitaire ou la crise inflationniste. Le bruit provenant de dynamiques saisonnières et d'imprécisions inhérentes à ces données est alors de deuxième ordre par rapport aux variations subies par l'économie. En période de moins grandes turbulences, ces limites sont plus problématiques et pourraient conduire à tirer des enseignements erronés sur la conjoncture. L'enjeu est d'autant plus important que la profondeur temporelle de ces données demeure faible. À terme, avec des séries de plus longue période, il sera possible de neutraliser une partie du bruit en désaisonnant¹³. La désaisonnalisation pourrait même être journalière, plutôt que mensuelle, afin de bénéficier de la grande fréquence de ces données.

► **Des études ponctuelles ou un partenariat pérenne ?** —

Un autre inconvénient d'utiliser des données privées par rapport à des données d'enquêtes est l'absence de garantie sur la pérennité des partenariats : les conventions précisent que les deux partenaires peuvent mettre fin à la collaboration à tout moment.

Au-delà de ces incertitudes, la méthode d'échantillonnage doit également permettre un rafraîchissement de la clientèle étudiée. Le premier échantillonnage avait été pensé pour des études ponctuelles sur la crise sanitaire. Les premières publications de l'Insee à partir des données bancaires reposaient sur un échantillon de clients issus de la clientèle de 2019-2020 de chaque banque. Pour une exploitation pérenne, il est cependant nécessaire de rafraîchir l'échantillon par l'ajout de nouveaux clients, pour éviter un biais d'attrition¹⁴ (certains clients disparaissent) et de vieillissement (avec un risque de décalage de la distribution des revenus observés). Ces deux biais s'aggravant avec le temps, une nouvelle méthodologie d'échantillonnage inspirée du plan de sondage de l'échantillon démographique permanent (EDP) a été mise en place. Cette nouvelle méthode permet un rafraîchissement régulier, mensuel, en ajoutant au fil du temps les clients qui entrent dans le champ du plan de sondage. Ce nouvel échantillon ouvre des perspectives pérennes à l'utilisation de ces données.

► **Un fort potentiel pour les études économiques et des limites potentiellement surmontables** —

Les études réalisées depuis quatre ans à partir des données bancaires ont permis de mieux cerner leur potentiel et leurs limites. Elles renseignent en temps réel sur les revenus, la consommation et le patrimoine de centaines de milliers de ménages et permettent ainsi de documenter l'évolution de la situation financière des ménages à

¹³ Désaisonnaliser consiste à appliquer un traitement statistique pour éliminer les effets dus aux phénomènes saisonniers.

¹⁴ Biais d'attrition : perte de représentativité de l'échantillon à cause de la disparition de certains clients de l'échantillon.

la suite de chocs conjoncturels ou individuels. En revanche, malgré leur richesse, elles ne peuvent pas remplacer les enquêtes. Les principales limites sont les suivantes : la clientèle d'une banque n'est jamais totalement représentative de la France entière (certains ménages ne détiennent d'ailleurs aucun compte en banque), les différents types de biens achetés dans une enseigne sont inconnus des banques, les données d'une seule banque ne donnent qu'une vision partielle des flux et du patrimoine des clients multi-bancarisés. Par ailleurs, pour l'instant, le statisticien ne dispose pas de suffisamment de recul pour désaisonnaliser ces données et donc pour exploiter tout le potentiel qu'offre leur très grande fréquence.

À terme, certaines limites peuvent cependant être partiellement surmontées. Ainsi, pour mieux connaître leurs clients, les banques enrichissent continûment leurs données. Chaque amélioration de la catégorisation des flux sur les comptes est un pas supplémentaire vers une identification des revenus et des dépenses tels qu'usuellement définis dans la statistique publique. Chaque information collectée par le conseiller améliore également le regroupement des clients d'un même ménage. Nourrir les partenariats existants, et en développer de nouveaux, permettra de gagner en représentativité et d'acquérir une vision plus complète des comptes des personnes multi-bancarisées. Enfin, avec des partenariats plus anciens, la profondeur temporelle des données augmente : de deux ans à la naissance des partenariats, l'historique disponible des clients est désormais de plus de cinq ans.

L'Insee remercie La Banque Postale et Crédit Mutuel Alliance Fédérale pour leur disponibilité et pour avoir permis l'accès à des données de comptes bancaires dans un cadre garantissant l'anonymat des clients.

Nos partenaires souhaitent rappeler les éléments suivants :

Pour La Banque Postale : ce partenariat contribue pleinement à la réalisation des objectifs environnementaux et sociaux que La Banque Postale a défini dans ses statuts en tant qu'entreprise à mission. Les données bancaires, communiquées anonymement, fournissent des points de vue inédits et complémentaires sur les situations financières des ménages, permettant d'enrichir les outils de statistique publique, et ainsi plus largement le débat public.

Pour Crédit Mutuel Alliance Fédérale : Crédit Mutuel Alliance Fédérale, première banque à adopter le statut d'entreprise à mission, participe à ces études dans le cadre des missions qu'elle s'est fixées :

- Contribuer au bien commun en œuvrant pour une société plus juste et plus durable : pour Crédit Mutuel Alliance Fédérale, participer à l'information économique c'est contribuer au débat démocratique ;
- Protéger l'intimité numérique et la vie privée de chacun : Crédit Mutuel Alliance Fédérale veille à la protection absolue des données de ses clients. Toutes les analyses réalisées dans le cadre de cette étude ont été effectuées sur des données strictement anonymisées et sur les seuls systèmes d'information sécurisés du Crédit Mutuel et hébergés en France.

► Bibliographie

- ADAM, Marine, BONNET, Odran et LOISEL, Tristan, 2022. Avec l'inflation, une précarité financière en légère hausse, mais inférieure en août 2022 à son niveau d'avant-crise sanitaire. In : *Insee Analyses*. [en ligne]. 13 octobre 2022. Insee. N° 76. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/6542263>.
- ADAM, Marine, BONNET, Odran, FIZE, Étienne, RAULT, Marion, LOISEL, Tristan et WILNER, Lionel, 2023. How does fuel demand respond to price changes? Quasi-experimental evidence based on high-frequency data. In : *Documents de travail*. [en ligne]. 19 juillet 2023. Insee. N° 2023-17. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/en/statistiques/7644836>.
- ADAM, Marine, BONNET, Odran, FIZE, Étienne, RAULT, Marion, LOISEL, Tristan et WILNER, Lionel, 2024. Cross-border shopping for fuel at the France-Germany border. In : *Documents de travail*. [en ligne]. 7 mai 2024. Insee. N° 2024-08. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/en/statistiques/8183838>.
- ANDERSEN, Asger L., JENSEN, Amalie S., JOHANNESSEN, Niels, KREINER, Claus T., LETH-PETERSEN, Soren et SHERIDAN, Adam, 2023. How Do Households Respond to Job Loss? Lessons from Multiple High-Frequency Datasets. In : *American Economic Journal: Applied Economics*. Octobre 2023. Vol. 15, N° 4, pp. 1-29.
- ANDRÉ, Mathias et MESLIN, Olivier, 2022. Patrimoine immobilier des ménages : enseignements d'une exploitation de sources administratives exhaustives. In : *Courrier des statistiques*. [en ligne]. 20 janvier 2022. Insee. N° N7, pp. 107-125. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/6035944?sommaire=6035950>.
- ASTIER, Jeanne, FACK, Gabrielle, FOURNEL, Julien, MAISONNEUVE, Flavie et SALEM, Ariane, 2024. Performance énergétique du logement et consommation d'énergie : les enseignements des données bancaires. In : *Focus*. [en ligne]. Janvier 2024. CAE. N° 103. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/performance-energetique-du-logement-et-consommation-d-energie-les-enseignements-des-donnees-bancaires>.
- BAKER, Scott R., FARROKHANIA, Robert A., MEYER, Steffen, PAGEL, Michaela et YANNELIS, Constantine, 2020. How Does Household Spending Respond to an Epidemic? Consumption during the 2020 COVID-19 Pandemic. In : *The Review of Asset Pricing Studies*. [en ligne]. Décembre 2020. Volume 10, N° 4, pp. 834-862. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://academic.oup.com/raps/article/10/4/834/5874450>.
- BLANCHET, Didier et GIVORD, Pauline, 2018. Les Big Data : quelles perspectives pour la statistique publique ? In : *Enjeux numériques*. [en ligne]. Juin 2018. N° 2. Big Data : économie et régulation. [Consulté le 29 juillet 2024]. Disponible à l'adresse : https://www.anales.org/enjeux-numeriques/2018/en_02_06_18.html.

- BOEHM, Johannes, FIZE, Etienne et JARAVEL, Xavier, 2023. Five Facts about MPCs: Evidence from a Randomized Experiment. In : *site github de Johannes Boehm*. [en ligne]. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://jmboehm.github.io/Helicopter.pdf>.
- BONNET, Odran, LOISEL, Tristan et OLIVIA, Tom, 2021. Impact de la crise sanitaire sur un panel anonymisé de clients de La Banque Postale – Les revenus de la plupart des clients ont été affectés de manière limitée et temporaire. In : *Insee Analyses*. [en ligne]. 3 novembre 2021. Insee. N° 69. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5760458>.
- BONNET, Odran, DIOP, Ouleye, LOISEL, Tristan, OLIVIA, Tom et WILNER, Lionel, 2023. La situation financière des ménages au jour le jour. In : *Insee Analyses*. [en ligne]. 5 décembre 2023. Insee. N° 90. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/7721885>.
- BUDA, Gergely, CARVALHO, Vasco M., HANSEN, Stephen, ORTIZ, Alvaro, RODRIGO, Tomasa et RODRIGUEZ MORA, José V., 2022. National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. In : *Cambridge Working Papers in Economics*. [en ligne]. Juillet 2022. 2244. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://ideas.repec.org/p/cam/camdae/2244.html>.
- COTTON, Franck et HAAG, Olivier, 2023. L'intégration des données administratives dans un processus statistique – Industrialiser une phase essentielle. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 104-125. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635829?sommaire=7635842>.
- EPAULARD, Anne, FIZE, Étienne, LE CALVÉ, Titouan, MARTIN, Philippe, PARIS, Hélène, PARRA RAMIREZ, Kevin et SRAER, David, 2021. La situation financière des PME/TPE en août 2021 au vu de leurs comptes bancaires. In : *Focus*. [en ligne]. Septembre 2021. CAE. N° 65. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/la-situation-financiere-des-pme-tpe-en-aout-2021-au-vu-de-leurs-comptes-bancaires>.
- FIZE, Étienne, PARIS, Hélène et RAULT, Marion, 2022. Quelle situation financière des entreprises et des ménages deux ans après le début de la crise Covid ? In : *Focus*. [en ligne]. Mars 2022. CAE. N° 83. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/quelle-situation-financiere-des-entreprises-et-des-menages-deux-ans-apres-le-debut-de-la-crise-covid>.
- FIZE, Étienne, MARTIN, Philippe, PARIS, Hélène et RAULT, Marion, 2022. La situation financière des ménages en début de crise énergétique. In : *Focus*. [en ligne]. Juillet 2022. CAE. N° 88. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.cae-eco.fr/la-situation-financiere-des-menages-en-debut-de-crise-energetique>.
- HILLION, Mélina, 2024. Une évaluation des achats transfrontaliers de tabac et des pertes fiscales associées en France. In : *Documents de travail*. [en ligne]. 16 avril 2024. Insee. N° 2024-06. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/8172202>.

- INSEE, 2021. En 2020, la chute de la consommation a alimenté l'épargne, faisant progresser notamment les hauts patrimoines financiers : quelques résultats de l'exploitation de données bancaires. In : *Note de conjoncture*. [en ligne]. 11 mars 2021. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/5232043>.
- 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 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254225?sommaire=4254170>.
- LE MONDE, 2020. Covid-19 : la crise sanitaire a fait basculer un million de Françaises et de Français dans la pauvreté. In : *Le Monde*. [en ligne]. 6 octobre 2020. [Consulté le 29 juillet 2024]. Disponible à l'adresse : https://www.lemonde.fr/societe/article/2020/10/06/un-million-de-nouveaux-pauvres-fin-2020-en-raison-de-la-crise-due-au-covid-19_6054872_3224.html.
- OLAFSSON, Arna et PAGEL, Michaela, 2018. The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software. In : *The Review of Financial Studies*. [en ligne]. 25 avril 2018. Volume 31, N° 11, pp. 4398-4446. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://ideas.repec.org/a/oup/rfinst/v31y2018i11p4398-4446..html>.
- POUGET, Julien, 2020. Nouvelles données pour suivre la conjoncture économique pendant la crise sanitaire : quelles avancées ? Quelles suites ? In : *Le blog de l'Insee*. [en ligne]. 28 juillet 2020. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://blog.insee.fr/nouvelles-donnees-pour-suivre-la-conjoncture-economique-pendant-la-crise-sanitaire-quelles-avancees-quelles-suites/>.
- REDOR, Patrick, 2023. Confidentialité des données statistiques : un enjeu majeur pour le service statistique public. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 46-63. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635823?sommaire=7635842>.
- RIVIÈRE, Pascal, 2018. Utiliser les déclarations administratives à des fins statistiques. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 14-24. [Consulté le 29 juillet 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/3647013?sommaire=3647035>.
- TAVERNIER, Jean-Luc, 2022. Crise Covid et mesure de la pauvreté. In : *Constructif*. [en ligne]. Juin 2022. N° 62 [Consulté le 29 août 2024]. Disponible à l'adresse : http://www.constructif.fr/bibliotheque/2022-6/crise-covid-et-mesure-de-la-pauvrete.html?item_id=6819.

Les revues par les pairs : un levier au service de la statistique publique



Mylène Chaleix* et Jean-Michel Durr**

La confiance dans les statistiques officielles est devenue un enjeu majeur au XXI^e siècle. Pour donner des gages sur leur qualité, le système statistique européen s'est doté en 2005 d'une référence : le Code de bonnes pratiques. Il s'est également donné les moyens d'évaluer la conformité à ce Code des systèmes statistiques nationaux et européen en organisant des revues par les pairs tous les sept à huit ans. Ces revues visent non seulement à en vérifier la conformité, mais également à améliorer les systèmes européens et nationaux.

Après avoir détaillé le fonctionnement des revues, des exemples illustrent l'efficacité du dispositif pour faire progresser l'ensemble des pays européens tout en respectant leurs spécificités. Il contribue à la fois à renforcer la gouvernance et l'organisation des systèmes statistiques et à harmoniser et diffuser des bonnes pratiques dans les différents pays.

En conclusion, sont évoquées les réflexions en cours sur l'évolution possible de la méthodologie des revues et du Code de bonnes pratiques pour renforcer la confiance dans les statistiques officielles.

 Confidence in official statistics has become a major issue in the 21st century. To guarantee their quality, the European Statistical System adopted the Code of Practice in 2005. It has also developed the tools to assess the compliance of national and European statistical systems with the Code by organising peer reviews every seven to eight years. The purpose of the peer reviews is not only to check compliance with the Code, but also to improve European and national statistical systems.

After presenting the peer review operating mechanism, examples are given to illustrate the effectiveness of the system in helping European countries improve their compliance with the Code, while respecting their specific characteristics. Peer reviews contribute both to strengthening the governance and organisation of statistical systems and to harmonising and disseminating best practices among the countries. The paper concludes with a discussion of possible changes to the peer review methodology and to the Code of Practice to strengthen confidence in official statistics.

* Cheffe de l'unité Qualité, DMCSI, Insee.
mylene.chaleix@insee.fr

** Inspecteur général honoraire de l'Insee, pair pour les 2^e et 3^e cycles de revues.
jmdurr@caos-consulting.com

Les statisticiens publics partagent depuis longtemps des valeurs éthiques sur l'indépendance professionnelle, la rigueur scientifique et la neutralité, le secret statistique et la confidentialité, la comparabilité. Ces valeurs sont souvent reprises dans les lois statistiques nationales. Les premières chartes ou déclarations ont été formalisées par l'Institut international de statistique¹ (ISI) en 1985 (International Statistical Institute, 2010) ou par la commission statistique de l'ONU adoptée en 1994 (United Nations, 2014). Les statisticiens français ont publié un premier code de déontologie en 1986.

La construction européenne a très rapidement mis l'accent sur la nécessité de disposer de statistiques de qualité². Dès 1959, Eurostat est créé et les principes d'impartialité, comparabilité, objectivité, indépendance scientifique et secret statistique, ainsi que la limitation de la charge des répondants sont présents dans la première loi statistique européenne de 1997 après la parution d'une première politique pour les informations statistiques en 1988. Mais en 2005, avec notamment l'invalidation des comptes nationaux grecs³ (Ernoul, 2012), le besoin de garantir la crédibilité des statistiques est de plus en plus présent et le Comité du programme statistique⁴ va adopter un cadre partagé alliant déontologie (valeurs) et qualité (des statistiques produites), première étape pour aller au-delà des simples intentions des déclarations précédentes. Le Code de bonnes pratiques de la statistique européenne⁵ (CoP⁶) est né le 24 février 2005, reste à lui donner un statut officiel.

En 2009, le CoP acquiert un statut réglementaire avec le Règlement 223/2009⁷. Ce texte définit les termes, il précise la gouvernance et la coopération au sein de la statistique européenne et avec les chercheurs, mais également les principes éthiques (indépendance professionnelle, impartialité, objectivité, fiabilité, secret statistique, rapport coût/efficacité) dans son article 2. Le Code de bonnes pratiques devient la référence pour les statisticiens (article 11) et la qualité des statistiques produites (principes de pertinence, exactitude, ponctualité, accessibilité et clarté, comparabilité, cohérence) dans son article 12. Le code actuel comporte 16 principes regroupés selon 3 domaines : un domaine reprenant les principes attendus pour la qualité des produits (orienté utilisateur), un domaine sur la qualité des processus (orienté producteur) et enfin le domaine institutionnel qui reprend les principes éthiques (**encadré**).

► La déclinaison opérationnelle du Code

La statistique européenne est organisée autour d'Eurostat et des systèmes statistiques nationaux des États membres. Il est indispensable de s'assurer aux niveaux national et européen de la qualité des statistiques produites en partageant le même référentiel. Le CoP a ainsi été décliné en indicateurs plus opérationnels pour permettre d'évaluer la conformité aux principes du code.

1 Créé en 1985, à la suite des premiers congrès de statistique initiés par Adolphe Quetelet, l'ISI s'inscrit dès sa création dans les valeurs d'indépendance et de rigueur scientifique, regroupant des individus et non des représentants des gouvernements.

2 Traité sur le fonctionnement de l'Union européenne (voir les références juridiques en fin d'article).

3 Voir les références juridiques en fin d'article.

4 Remplacé par le Comité du système statistique européen (CSSE) en 2009 avec l'adoption de la « loi statistique » européenne (règlement 2009/223). Le système statistique européen (SSE) réunit Eurostat, les autorités statistiques des États membres de l'Union européenne et ceux de l'Association européenne de libre-échange (Islande, Liechtenstein, Norvège et Suisse).

5 <https://ec.europa.eu/eurostat/web/products-catalogues/-/ks-02-18-142>.

6 CoP : Code of Practice.

7 Voir les références juridiques en fin d'article.

► Encadré - Le Code de bonnes pratiques de la statistique européenne, en vigueur en 2024

Utilisateurs

Qualité des produits statistiques

Pertinence
Exactitude et fiabilité
Actualité et ponctualité
Cohérence et comparabilité
Accessibilité et clarté

Producteurs

Qualité des processus statistiques

Méthodologie solide
Procédures statistiques adaptées
Charge non excessive pour les déclarants
Rapport coût-efficacité

Environnement institutionnel

Qualité du cadre institutionnel
du système statistique

Indépendance professionnelle
Coordination et coopération*
Mandat pour la collecte et l'accès aux données
Adéquation des ressources
Engagement sur la qualité
Secret statistique
Impartialité et objectivité

* En 2017, lors de la deuxième révision du CoP, ce principe a été ajouté.



Les autorités statistiques disposent d'un mandat légal clair les habilitant à collecter et à accéder à des informations issues de sources de données multiples pour les besoins des statistiques européennes.



Enfin, le dispositif a également été complété en 2011 par un Cadre d'assurance qualité⁸ qui propose des méthodes, outils et bonnes pratiques à mettre en œuvre pour illustrer les principes et les indicateurs.

Ainsi, le principe « Mandat pour la collecte et l'accès aux données » s'énonce de la manière suivante : « Les autorités statistiques disposent d'un mandat légal clair les habilitant à collecter et à accéder à des informations issues de sources de données multiples pour les besoins des statistiques européennes. À la demande des autorités statistiques, les administrations, les entreprises et les ménages ainsi

que le public en général peuvent être contraints par la loi à permettre l'accès à des données ou à fournir des données pour l'établissement de statistiques européennes. »

Ce principe se décline en quatre indicateurs :

1. Le mandat des autorités statistiques les habilitant à collecter des informations issues de multiples sources de données et à y accéder pour l'élaboration, la production et la diffusion des statistiques européennes est inscrit dans le droit.

⁸ <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V2.0-final.pdf>.

2. Les autorités statistiques sont autorisées par la législation à accéder rapidement et gratuitement aux données administratives et à les utiliser à des fins statistiques. Elles participent, dès le démarrage, à la conception, au développement et à l'abandon des fichiers administratifs, afin, le cas échéant, de rendre leurs données plus adaptées à des finalités statistiques.

3. Les autorités statistiques peuvent rendre obligatoire la réponse aux enquêtes statistiques en se fondant sur un acte juridique.

4. L'accès à d'autres données, telles que les données privées, à des fins statistiques est facilité, dans le respect du secret statistique et de la protection des données.

Parmi les bonnes pratiques proposées concernant les données administratives, on peut citer : la disponibilité de la législation, l'existence de procédures pour explorer, collecter et accéder aux données, le caractère obligatoire de la transmission et le principe de gratuité, l'existence de métadonnées⁹, les accords et la coopération.

Certains engagements peuvent relever de plusieurs principes. Par exemple, l'annonce des calendriers de diffusion des publications statistiques relève à la fois du principe d'impartialité et d'objectivité (« Les autorités statistiques décident en toute indépendance, de la date de parution et du contenu des publications statistiques [...] ») ou encore « Les dates et heures de parution des statistiques sont annoncées à l'avance. ») et du principe d'indépendance professionnelle (« Les responsables des instituts nationaux de statistique, d'Eurostat et, le cas échéant des autres autorités statistiques sont les seuls compétents pour décider [...] du contenu et de la date de diffusion des publications statistiques. »).

► La naissance des revues par les pairs

Avec le CoP, la statistique européenne s'est dotée d'une référence. Pour en assurer la crédibilité, il faut pouvoir prouver son application. Il est donc nécessaire de mettre en place un mécanisme de vérification de la conformité au CoP. Sous quelle forme ?

Les revues par les pairs du système statistique européen ont pour premier objectif de vérifier la conformité aux principes du CoP et de garantir la qualité des statistiques européennes, dans l'esprit d'un audit.

Inspirées des revues par les pairs du milieu scientifique (évaluation d'article, projet ou laboratoire de recherche par d'autres chercheurs, les « pairs », sur la base d'un référentiel partagé), les revues par les pairs du système statistique européen ont pour premier objectif de vérifier la conformité aux principes du CoP et de garantir la qualité des statistiques européennes, dans l'esprit d'un audit. Au-delà de la conformité, elles visent à mettre en œuvre des plans d'actions d'amélioration et de diffusion des bonnes pratiques et ainsi participer à leur harmonisation. Pour cela, les revues les mettent en avant et formulent des recommandations tournées vers l'avenir. L'objectif est d'aider les instituts nationaux de statistique (INS), les autres autorités nationales

⁹ Une métadonnée est une donnée qui fournit de l'information sur une autre donnée.

(ONA : *other national authority*¹⁰) qui conçoivent, produisent et diffusent des statistiques européennes et Eurostat à améliorer et à développer les systèmes statistiques nationaux et européen d'un cycle à un autre. Le caractère public des rapports permet également d'encourager les gouvernements des États membres à soutenir la mise en œuvre des plans d'actions. Dans le cadre du programme de coopération statistique panafricain financé par l'Union européenne, les pays africains se dotent eux aussi d'un mécanisme de revue par les pairs pour évaluer la maturité de leurs systèmes statistiques nationaux.

Les revues par les pairs évaluent un système statistique national et non l'élaboration d'une statistique donnée. D'autres dispositifs existent pour juger de la qualité des statistiques sous règlement européen, avec notamment la fourniture de rapports Qualité par les États membres. Le cas échéant, des évaluations, ponctuelles ou prévues dans les règlements sectoriels, peuvent être décidées sur une opération statistique donnée, pour l'ensemble ou certains pays. Par exemple, le dispositif de contrôle de la conformité de l'IPCH¹¹ (Indice des prix à la consommation harmonisé) repose sur la fourniture d'une documentation détaillée par les États membres, une analyse approfondie des données, des métadonnées et des méthodes par Eurostat, ainsi que sur un dialogue ouvert avec les experts nationaux avec parfois une visite, ce qui fut le cas pour la France en 2024. Les rapports sont publiés sur le site d'Eurostat et peuvent comporter des recommandations.

Le protocole de revue par les pairs d'Eurostat est adapté à son rôle particulier et la « revue » est réalisée depuis le 2^e cycle par le Conseil consultatif européen pour la gouvernance statistique créé en 2008 l'ESGAB – European Statistical Governance Advisory Board –, équivalent européen de l'ASP – Autorité de la statistique publique¹² – pour la France (Bureau, 2020).

► Mais comment procéder ?

Le périmètre

Un cycle de revue par les pairs s'appuie sur l'évaluation des principes du CoP et, depuis le 2^e cycle (2014-2017), il porte sur l'ensemble du système statistique national des États membres. L'exigence sur la conformité aux principes augmente d'un cycle à un autre.

Le principe de la revue nécessite donc de clarifier ce que l'on entend par système statistique national. Il est composé de l'INS et des ONA. Le nombre de ces derniers varie selon les pays : de zéro ONA pour les Pays-bas à une trentaine pour l'Allemagne. En France, 10 services statistiques ministériels (SSM) sur les 16¹³ participent directement à des statistiques européennes ainsi que le CépiDc¹⁴ pour les statistiques sur les causes

10 À noter que les banques centrales, qui disposent de leur propre processus d'évaluation, peuvent être exclues de la liste des ONA.

11 <https://ec.europa.eu/eurostat/documents/272892/272977/Compliance+Monitoring+description/>.

12 L'ASP veille à l'indépendance professionnelle dans la conception, la production et la diffusion de statistiques publiques et assure une vigilance quant à la qualité des statistiques publiques, que ce soit en termes de rigueur méthodologique ou de respect de différents principes déontologiques. <https://www.autorite-statistique-publique.fr/>.

13 Pour le service statistique public (SSP) français, l'ensemble des SSM sont concernés par la conformité au Code de bonnes pratiques et participent à la stratégie Qualité quinquennale définie à l'issue de chaque cycle de revue.

14 L'article L.2223-42 du Code Général des Collectivités Territoriales et les articles R2213-1-1 – R2213-1-4 confient au CépiDc, centre d'épidémiologie sur les causes médicales de décès, la mission de produire la base de données des causes médicales de décès et de la diffuser.

médicales de décès¹⁵, le CASD¹⁶ (Gadouche, 2019) étant le douzième ONA. Tous les ONA ne sont pas concernés par la revue. Les INS ont un rôle de coordination dans le système statistique national et décident quels ONA sont sélectionnés pour la phase d'autoévaluation et invités à la visite d'évaluation par les pairs, sur la base des critères de sélection prédéfinis tels que l'importance de la production de statistiques européennes, en volume et en signification.

Pour le 3^e cycle (2021-2023), trois ont ainsi directement participé à la revue française : la Drees¹⁷ (chargé de la santé et des solidarités), le Sdes¹⁸ (chargé de l'énergie, du logement, des transports et de l'environnement) et le SSP¹⁹ (chargé de l'agriculture et de l'alimentation).

Les principes retenus



La troisième série d'évaluations par les pairs visait à identifier les progrès réalisés par rapport à la revue précédente dans l'adhésion aux principes du Code de bonnes pratiques et également les domaines dans lesquels la collaboration au sein du SSE pouvait être améliorée.



Ils ont fait l'objet d'un accord commun au sein du système statistique européen (SSE), sur les objectifs, la portée et les modalités de mise en œuvre. La méthode pour le 3^e cycle²⁰ a été développée par un comité d'experts animé par Eurostat sur la base des enseignements du 2^e cycle. Elle s'organise autour d'une autoévaluation et d'une visite par une équipe d'experts (les pairs).

L'approche retenue est une combinaison d'une démarche de type audit et d'examen par les pairs. Elle consiste en l'implication d'experts externes pour garantir :

- la crédibilité et l'objectivité des examens par les pairs ;
- la fourniture de documents comme preuves des déclarations ;
- la pleine responsabilité des recommandations de l'équipe d'experts chargée de l'examen par les pairs ;
- et le droit pour les INS d'exprimer des opinions divergentes sur les recommandations d'amélioration formulées par l'équipe d'experts.

Les éventuelles opinions divergentes sont publiées avec le rapport d'examen par les pairs, dans une annexe. Enfin, dans son plan d'actions, l'INS formule les actions d'amélioration retenues en réponse aux recommandations.

¹⁵ Voir l'article de Coudin et Robert sur les statistiques de causes de décès dans ce même numéro.

¹⁶ Centre d'accès sécurisé aux données, groupement d'intérêt public, il a pour mission d'organiser et de mettre en œuvre des services d'accès sécurisé pour les données confidentielles, notamment statistiques, à des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation.

¹⁷ Drees : Direction de la recherche, des études, de l'évaluation et des statistiques.

¹⁸ Sdes : Service des données et études statistiques.

¹⁹ SSP : Service de la statistique et de la prospective.

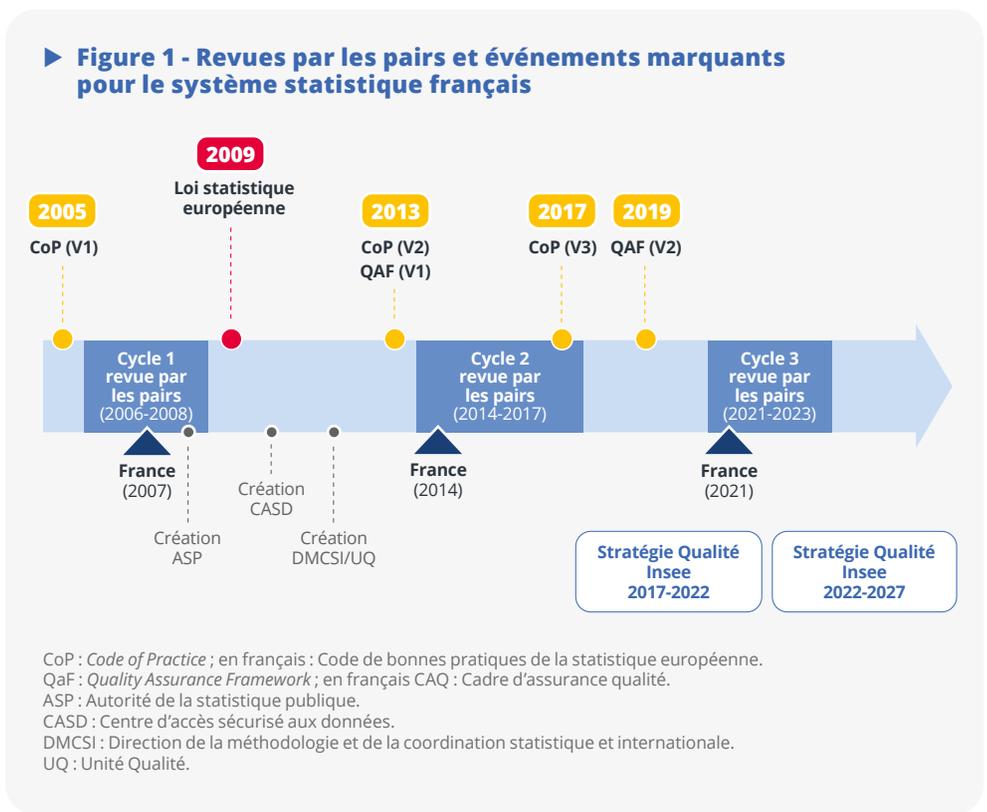
²⁰ https://ec.europa.eu/eurostat/documents/64157/4372828/1_Methodology+for+peer+reviews.pdf.

En outre, la troisième série d'évaluations par les pairs visait à identifier les progrès réalisés par rapport à la revue précédente dans l'adhésion aux principes du Code de bonnes pratiques et également les domaines dans lesquels la collaboration au sein du SSE pouvait être améliorée.

La fréquence

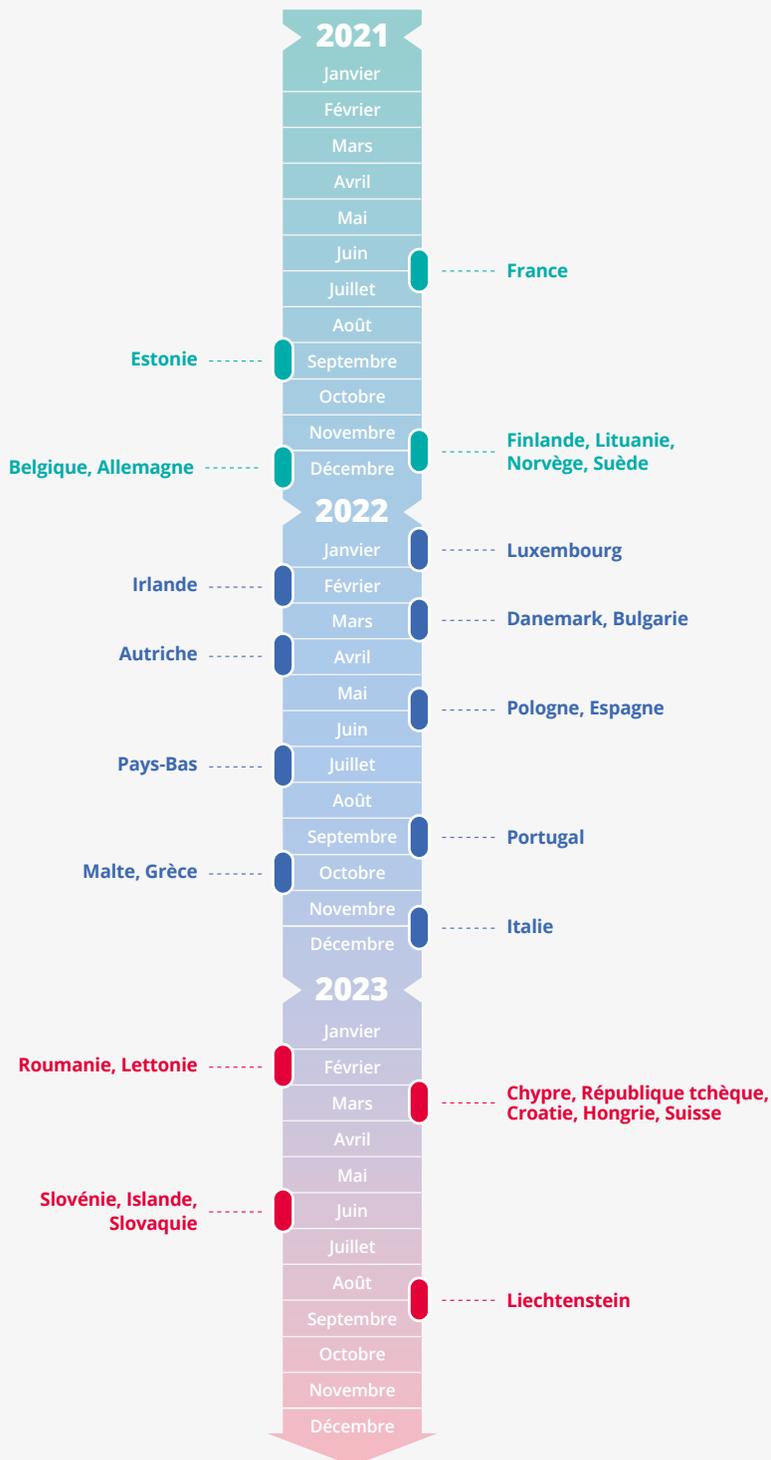
L'exercice est exigeant. Pour chaque pays, de l'autoévaluation à la validation du plan d'actions, il s'écoule environ deux ans ; quant au plan d'actions, il s'étale généralement sur une période de cinq ans. La revue ne peut donc être renouvelée que tous les sept à huit ans.

Au niveau européen, la partie opérationnelle de chaque cycle dure au moins trois ans (*figure 1*) ; les « visites dans les pays » s'étalent sur deux ans et demi (*figure 2*).



Entre deux cycles et pendant l'étape de suivi des plans d'action, la préparation du cycle suivant s'organise avec la révision du Code de bonnes pratiques et celle du cadre d'assurance Qualité mais également avec l'évolution de la méthodologie, la préparation de la communication et le recrutement des pairs du prochain cycle.

► **Figure 2 - Calendrier du troisième cycle de revues par les pairs**



► Le 3^e cycle (2021-2023)

Le processus général...

Chaque revue par les pairs dans un pays est confiée à une équipe composée de quatre experts, dont au moins un expert externe au système statistique européen et un expert d'Eurostat, afin de permettre une combinaison équilibrée de compétences, de connaissances et d'aptitudes : expérience en tant que cadre supérieur dans un INS/ONA, connaissance du fonctionnement d'un système statistique national, connaissance des développements stratégiques en matière de statistiques aux niveaux national, européen et international, connaissance des évolutions récentes dans le SSE, expertise en matière de statistiques et d'activités de modernisation, participation active à des activités liées au SSE. Huit équipes de quatre pairs ont ainsi été constituées. L'exercice est intense et stimulant pour les membres de l'équipe des pairs. Il s'agit de comprendre rapidement le fonctionnement, mais également les forces et les faiblesses d'un système statistique national dans la mise en application du CoP et de formuler des recommandations pertinentes (Eurostat, 2023).

Pour ce 3^e cycle de revues, les pairs devaient tout particulièrement être vigilants sur les principes relatifs à l'indépendance professionnelle et la coordination du système statistique national et sur les principes porteurs de modernisation (notamment sur l'accès aux données, les engagements Qualité, la méthodologie statistique robuste et adaptée, l'accessibilité et la clarté de l'accès aux statistiques).

Pour chaque pays, le processus comportait quatre étapes :

- 1.** Une autoévaluation par le système statistique national portant sur la conformité des 84 indicateurs, les forces et faiblesses²¹ sur les 16 principes et un questionnaire complémentaire plus prospectif sur les adaptations et innovations du système statistique national notamment en réponse à la crise sanitaire de 2020-2021. Pour les ONA inclus dans la revue, le questionnaire était simplifié.
- 2.** Une visite d'une semaine de l'équipe des pairs pendant laquelle ces derniers rencontraient des experts et des responsables de l'INS et des ONA inclus dans la revue, ainsi que des parties prenantes externes, notamment des journalistes, des associations professionnelles, des chercheurs, des pouvoirs publics et des organismes consultatifs ou encore quelques jeunes statisticiens. Lors d'un débriefing à la fin de la visite, l'équipe des pairs présentait ses recommandations aux équipes de direction de l'INS et des ONA ayant fait partie de la revue. Deux sortes de recommandations étaient formulées : des recommandations fondamentales pour assurer la conformité avec le CoP et des recommandations moins critiques, ou plus techniques, suggérant des améliorations dans la mise en œuvre du CoP.
- 3.** Dans les semaines suivant la visite, un rapport d'évaluation rédigé par l'équipe des pairs. Il contenait, outre ses recommandations, une présentation du service statistique national, une évaluation des progrès réalisés depuis cinq ans, les forces du service statistique dans la conformité au CoP, y compris les innovations réalisées, et les recommandations regroupées autour de questions ou de thèmes généraux

²¹ Selon la méthodologie SWOT : *Strengths, Weaknesses, Opportunities and Threats*, soit forces, faiblesses, opportunités, menaces.

identifiés, avec une référence au(x) principe(s) et à l'(aux) indicateur(s) concerné(s). Ce rapport était ensuite transmis à Eurostat pour validation et harmonisation ainsi qu'à l'INS et aux ONA concernés pour réaction et si nécessaire expression de vues divergentes dans un chapitre spécifique.

4. Un plan d'actions d'amélioration établi par le service statistique national, une fois le rapport finalisé.



Par souci de transparence, les rapports et les plans d'actions sont publiés sur les sites internet des INS et sur celui d'Eurostat.



Par la suite, un suivi annuel de la mise en œuvre de ces actions d'amélioration était piloté par Eurostat. Par souci de transparence, les rapports et les plans d'actions sont publiés sur les sites internet des INS et sur celui d'Eurostat.

Une revue par les pairs est un moment important pour le service statistique d'un pays. Cet exercice est exigeant et complexe par son approche transversale et non par processus métier.

... et en France

Dans un contexte contrarié par la crise sanitaire du Covid, la France a été le premier pays concerné par le 3^e cycle (Insee, 2022). Revue pilotée par l'Insee²², le choix a été fait d'associer pendant la phase de préparation l'ensemble du service statistique public (l'Insee et les 16 SSM) ainsi que le CépiDc et le CASD (ONA pour l'Europe) ou encore la Banque de France pour « embarquer » l'ensemble des « parties prenantes » dans l'exercice. L'organisation s'est appuyée sur des relais : un cercle rapproché de conseillers²³ et des ambassadeurs dans les services²⁴ ; elle a également associé les institutionnels : ASP, Cnis²⁵ (Anxionnaz et Maurel, 2021) et le Comité du Label (Christine et Roth, 2020).

L'établissement des documents préparatoires (présentation générale du système statistique, questionnaires d'autoévaluation et complémentaire) a été l'occasion d'un exercice collectif démarré fin 2019 à la fois pour confirmer les éléments de preuves et les bonnes pratiques identifiées, et pour proposer des argumentaires et des positions pour l'autoévaluation. L'exercice s'est poursuivi avec l'établissement de synthèses par principe et la cotation de la conformité au CoP. Les échanges au sein du comité stratégique de la qualité²⁶ ont permis de dégager les forces et faiblesses. Cette approche a permis de travailler dès le début sur les axes de progrès identifiés en interne.

Les trois SSM participant à la revue ont renseigné le questionnaire d'autoévaluation dédié avec l'appui de l'unité Qualité. Après échanges avec les SSM, les documents ont été validés en comité de direction de l'Insee et envoyés à Eurostat en mars 2021.

²² Par trois membres de l'unité Qualité : Dominique Bonnans, Laurent Di Carlo et Mélanie Vanderschelden.

²³ Les conseillers étaient des membres de la Direction de la méthodologie et de la coordination statistique et internationale ainsi que des membres de l'Inspection Générale de l'Insee.

²⁴ Directions ou départements de l'Insee et des SSM.

²⁵ Le Conseil national de l'information statistique (Cnis) assure la concertation entre les producteurs et les utilisateurs de la statistique publique.

²⁶ Composé des directeurs de l'Insee, de deux directeurs régionaux et des chefs de SSM audités.

L'étape suivante a été la préparation de la visite, programmée du 28 juin au 2 juillet 2021. L'équipe des pairs était composée de la directrice-adjointe de l'INS irlandais accompagnée par la responsable des standards et nomenclatures de l'INS slovène, d'un consultant en statistique publique et du conseiller de la directrice générale d'Eurostat²⁷. Des allers-retours avec les pairs ont été nécessaires pour préciser leurs attentes et pour convier les bonnes personnes aux différentes sessions organisées par principe, par acteurs (institutionnels, utilisateurs) et pour chaque ONA audité. Les échanges ont été à la fois libres et bienveillants.

À la fin de la visite, les pairs ont donné les grandes lignes de leur rapport qui a été envoyé la semaine suivante, et diffusé officiellement en septembre 2021.

La rédaction du plan d'actions a été grandement facilitée par tout le travail préparatoire à la revue. Il a été finalisé au second semestre 2021 et, après échanges avec Eurostat, publié en mars 2022. Le choix a été fait de proposer des actions très proches des recommandations du plan d'actions européen et d'élargir au niveau de la stratégie Qualité 2022-2027 en ajoutant des axes et thèmes jugés importants qui n'avaient pas été pointés par les pairs (par exemple sur l'attention à porter aux publics pour ne pas limiter les actions à celles en faveur des chercheurs).

Mais concrètement, quels sont les résultats du cycle 2021-2023 de la revue par les pairs ?

► Résultats du 3^e cycle

En Europe...

Les revues ont mis en évidence les progrès notables réalisés sur la gouvernance des systèmes statistiques nationaux. Dans de nombreux pays, attirer un personnel diplômé et développer les compétences restent des enjeux importants.

Sur le management de la qualité (c'est-à-dire gouvernance de la qualité, outillage, revues de processus, etc.), les progrès réalisés après le 2^e cycle se poursuivent ; certains pays comme le Portugal ont choisi de demander une certification ISO9001²⁸.



L'accès aux données privées a été repris par de nombreux pays comme axe d'amélioration pour les prochaines années, accès qui devrait être facilité par l'évolution de la loi statistique européenne.



Concernant l'accès et l'utilisation des données administratives pour élaborer les statistiques, des progrès ont été réalisés pour que les services statistiques soient mieux insérés dans les processus de décisions de recueils administratifs et informés des évolutions des recueils permettant un recours généralisé à ce type de données en Europe. La crise sanitaire de Covid a été un levier supplémentaire pour investir dans de nouveaux recueils (administratifs ou privés²⁹). L'accès aux données

²⁷ Jennifer Banim, Tina Steenvoorden, Michail Skaliotis et Pierre Bischoff.

²⁸ Norme internationale qui définit les exigences pour établir et améliorer un système de management de la qualité <https://www.iso.org/standard/62085.html>.

²⁹ Voir l'article de Bonnet et Loisel sur les données de comptes bancaires dans ce même numéro.

privées a été repris par de nombreux pays comme axe d'amélioration pour les prochaines années, accès qui devrait être facilité par l'évolution de la loi statistique européenne (Règlement 223³⁰).

De nombreux pays ont entrepris une modernisation substantielle de leur production statistique, grâce à des portails pour la collecte des données via des questionnaires en ligne, le traitement automatisé des données, la standardisation des processus ou encore le recours à des méthodes d'apprentissage automatique. On peut citer également la mise en place de systèmes d'information statistique modernes basés sur l'activation des métadonnées et couvrant tout ou partie de l'ensemble de la chaîne de production.



Dans le domaine de la diffusion, des approches plus centrées sur l'utilisateur se sont développées en proposant différents types de canaux en fonction des types d'utilisateurs.



Enfin dans le domaine de la diffusion, des approches plus centrées sur l'utilisateur se sont développées en proposant différents types de canaux en fonction des types d'utilisateurs. Les INS ont également investi les questions de littératie statistique en développant des partenariats avec le monde éducatif et universitaire ainsi qu'avec les médias à l'instar de la compétition européenne de statistique pour les 14-18 ans. La crise sanitaire a favorisé l'innovation avec la diffusion de statistiques dites expérimentales (c'est-à-dire produites à partir de nouvelles données ou faisant appel à de nouvelles méthodes) sans attendre de disposer du recul nécessaire pour en assurer la robustesse et les rendre officielles.

Les pairs ont également identifié le rôle d'intendant des données (*data stewardship*) dans quelques INS, positionnés comme acteurs clés de l'écosystème national des données, compte tenu de leur expertise dans les pratiques de gestion, d'accès et de partage des données dans le secteur public. Le rôle des INS dans l'écosystème des données peut être très différent, depuis un rôle minimal de service national de statistique à un fonctionnement en tant que responsable des données du secteur public doté de l'autorité de gouvernance requise.

Au total, 581 recommandations ont été émises lors du 3^e cycle, soit une moyenne de 19 par pays (**tableau**). Avec un degré d'exigence supérieur à celui des revues précédentes, seules 12 % des recommandations concerne un manquement au CoP, signifiant que le niveau de conformité des autorités statistiques nationales est élevé. Ces 68 recommandations concernent principalement les principes du domaine institutionnel : 21 sur l'indépendance professionnelle (en particulier sur les processus de nomination des responsables des INS ou des ONA), 16 sur la coordination et 12 sur l'impartialité et l'objectivité (gestion des accès anticipés et calendriers de diffusion notamment).

Toutes les autres recommandations, soit près de 88 %, portent sur des améliorations possibles de la mise en œuvre du CoP sur la quasi-totalité des principes du Code.

30 Voir les références juridiques en fin d'article.

► **Tableau : Les recommandations du 3^e cycle par principe du CoP pour les 31 pays**

Principes	Recommandations		
	Total	Pour amélioration au-delà du CoP	Pour conformité au CoP
Total	581	513	68
Coordination et coopération	80	64	16
Indépendance professionnelle	68	47	21
Accessibilité et clarté	66	63	3
Adéquation des ressources	59	59	0
Méthodologie solide	45	45	0
Mandat pour la collecte et l'accès aux données	43	41	2
Impartialité et objectivité	40	28	12
Procédures statistiques adaptées	38	37	1
Pertinence	35	35	0
Engagement sur la qualité	33	28	5
Rapport coût-efficacité	29	29	0
Secret statistique	17	12	5
Charge non excessive pour les déclarants	17	16	1
Exactitude et fiabilité	5	5	0
Actualité et ponctualité	5	3	2
Cohérence et comparabilité	1	1	0

Source : Bilan du 3^e cycle de revue par les pairs - Eurostat.

... et en France

Les pairs sont venus, ont vu et ont conclu... pour la France à un « niveau élevé de conformité aux principes du CoP dans tout le SSP », tout particulièrement grâce à la loi de 1951³¹ qui a su s'enrichir au fil des dernières décennies, à une structure et une gouvernance de la statistique publique française très efficaces et à des ressources adaptées. Les pairs ont formulé 16 recommandations pour poursuivre la démarche d'amélioration continue de la qualité, dont une seule de conformité.

Sur l'environnement institutionnel, la recommandation de conformité porte sur la procédure de nomination du directeur général de l'Insee et des chefs de SSM³². En effet,

³¹ Voir les références juridiques en fin d'article.

³² Les directeurs d'administration centrale sont, en France, nommés par le président de la République par décret en Conseil des ministres.

le CoP est particulièrement vigilant sur les procédures de nomination des responsables pour assurer l'indépendance professionnelle des INS et des ONA. Les pairs ont jugé que les progrès réalisés avec la mise en place d'un comité d'audit et la possibilité pour l'ASP de rendre un avis public sur les compétences des candidats étaient insuffisants, en particulier sur le début du processus de recrutement et la publicité sur la vacance des postes, et la procédure peu connue. Selon le cas, il s'agit surtout de gagner en lisibilité en portant à la connaissance du public la procédure ou d'élargir le rôle de l'Autorité de la statistique publique au cas des chefs de SSM qui ne sont pas directeurs d'administration centrale (actions réalisées en 2021). Par ailleurs, les pairs ont recommandé de mettre en place et de promouvoir un logo spécifique « Service statistique public », projet également en cours au sein du SSP.

Les pairs ont estimé que le service statistique public français (SSP) devrait poursuivre l'élargissement des conditions d'accès aux données privées au-delà de ce qui est déjà prévu par l'article 3bis de la loi de 1951 (remplacement d'enquêtes statistiques obligatoires), mais également mobiliser l'Insee et les SSM dans la conception, le développement et l'éventuelle évolution des sources administratives afin de les rendre plus adaptées aux usages statistiques (Hand, 2018).

Sur la qualité et la documentation des processus, les pairs encouragent le SSP à poursuivre ses efforts de formalisation et de généralisation pour rester sur une dynamique d'amélioration des processus (démarches Qualité, documentation, métadonnées, etc.) et de facilitation des transmissions de compétences lors des mobilités.



Sur l'accès aux données, les pairs encouragent le SSP français à poursuivre ses initiatives pour rendre les données de plus en plus accessibles à tous les publics.



Enfin sur l'accès aux données, les pairs encouragent le SSP français à poursuivre ses initiatives pour rendre les données de plus en plus accessibles à tous les publics.

À partir de ces recommandations, le SSP a construit un plan d'actions d'amélioration en 36 actions. Par exemple, pour la recommandation d'« améliorer la navigation sur son site web pour tous types d'utilisateurs et optimiser le moteur de recherche de son site

pour faciliter l'accès aux statistiques diffusées », le plan comporte deux actions. La première concerne l'amélioration du moteur de recherche et de l'accès aux données sur le site internet (réalisée en 2023), la seconde la publication d'un catalogue de jeux de données de l'Insee sur son site, réalisée en 2024 avec le déploiement du projet Mélodi (Mauguin et Sagnes, 2024).

Le dispositif des revues par les pairs est très formalisé et peut apparaître bureaucratique par certains aspects (lourdeur du questionnaire d'autoévaluation et éléments de preuve à fournir notamment). Il est cependant important à la fois de garantir l'objectivité des réponses et d'assurer l'équité entre pays. Il s'avère en pratique très efficace pour faire progresser l'ensemble des pays européens en respectant la culture et le rythme de chacun des pays.

► Les revues, un levier pour progresser

Des progrès significatifs en Europe

Sur les vingt dernières années et après trois cycles de revues par les pairs, des progrès significatifs ont été constatés. Prenons trois exemples emblématiques.

- Premier exemple : **le renforcement de la gouvernance et de la coordination des systèmes statistiques nationaux et du SSE.**

L'inscription des principes d'indépendance professionnelle et de coordination du système statistique national s'est progressivement imposée dans les lois statistiques nationales.

L'inscription des principes d'indépendance professionnelle et de coordination du système statistique national s'est progressivement imposée dans les lois statistiques nationales. Certains pays le font à un rythme régulier en fonction des opportunités, d'autres à intervalles plus longs.

Sur la gouvernance de la statistique publique, les pays qui n'en disposaient pas se sont progressivement dotés de comités de concertation avec les utilisateurs, et d'une

autorité ou d'un dispositif indépendant permettant de veiller au respect des principes du Code. Cependant, l'organisation des systèmes statistiques nationaux et le rattachement administratif des INS ou des ONA restent très hétérogènes et liés à l'organisation étatique des pays (double niveau fédéral/régional, administration rattachée à un ministère ou au Conseil des ministres, agence indépendante).

En Allemagne (Alexandre et Cling, 2019) : l'organisation du système statistique national est adaptée à un système fédéral avec un office fédéral central (Destatis), 14 offices statistiques³³ indépendants pour les 16 Länder. Le règlement 223 et la revue par les pairs en 2015 ont été des leviers pour renforcer le rôle de coordination de Destatis. Cette coordination se traduit par une méthodologie unique pour les statistiques fédérales (un tiers sont directement produites par Destatis) et une coopération sur des outils mutualisés, complétées par la mise en place d'un pilotage par la qualité pour les offices des Länder.

L'Italie est organisée différemment. Mis en place en 1989, le système statistique italien est très décentralisé³⁴. Seuls 13 de ces multiples organismes sont actuellement qualifiés d'ONA participant directement à l'élaboration de statistiques européenne. La coordination est pilotée par Istat (INS italien) et s'appuie sur une politique Qualité à deux niveaux d'exigence (Istat et ONA avec la conformité au Code, les autres offices statistiques avec un « CoP italien » – version simplifiée du Code).

- Deuxième exemple : **le développement des bonnes pratiques dans les INS avec des conséquences possibles sur les relations avec les chercheurs.**

³³ L'office de Berlin-Brandebourg (basé à Berlin) et celui de Hambourg-Schleswig-Holstein (basé à Hambourg) couvrent chacun deux Länder.

³⁴ Il comprend des bureaux statistiques dans les ministères, divers organismes locaux des régions aux municipalités et d'autres organismes publics ou privés qui produisent des statistiques, soit environ 3 400 organismes, dont 3 000 bureaux statistiques des municipalités.

Les relations avec le monde académique relèvent de plusieurs principes du Code de bonnes pratiques. Avec les universitaires, elles sont encouragées à la fois au titre de la coopération mais également pour améliorer la méthodologie, l'efficacité des méthodes employées et encourager le développement d'outils performants (méthodologie solide). L'accès aux micro-données pour les chercheurs est présent dans le principe d'accessibilité et de clarté avec en contrepartie des protocoles stricts pour assurer le respect du secret statistique et de la protection des données individuelles du principe éponyme.



Les relations avec le monde académique relèvent de plusieurs principes du Code de bonnes pratiques. Avec les universitaires, elles sont encouragées à la fois au titre de la coopération mais également pour améliorer la méthodologie.



La diffusion des bonnes pratiques ou d'innovations d'un pays à un autre a permis des progrès significatifs alors qu'existaient des situations très disparates (accès interdits, procédures non formalisées, accueils sur site, premiers sites sécurisés à distance notamment au Danemark et aux Pays-Bas). En France, cela s'est traduit par l'évolution de la loi statistique en 2009 et la création du CASD en 2010. Des avancées similaires ont eu lieu dans d'autres pays, par exemple en Allemagne, Espagne,

Portugal, etc. Le règlement européen (n° 557/2013³⁵) permet désormais aux chercheurs d'accéder aux données harmonisées. Enfin, certains centres sécurisés se sont organisés en réseau (Idan) en 2019 et en Autriche, l'interdiction de transmettre les données aux chercheurs a été levée avec l'ouverture d'un centre d'accès en 2023.

- Troisième exemple : **les pratiques de diffusion.**

Le partage entre systèmes statistiques nationaux et européen des bonnes pratiques, des argumentaires et une formalisation plus forte des calendriers de diffusion, de la gestion des accès anticipés ou encore de la gestion des révisions et des erreurs contribuent aux principes de transparence et d'impartialité. Si l'indépendance des INS sur le choix des dates de parution et des contenus des publications statistiques ne fait pas débat, les accès privilégiés dans certains pays sont beaucoup plus discutés. Faut-il les interdire ?

Faut-il les conserver pour éviter de mauvaises interprétations ou la remise en cause des chiffres en ne laissant pas aux décideurs politiques le temps de construire leur argumentaire ? Le Code ne l'interdit pas mais fixe des règles de plus en plus strictes : justification, limitation des accès anticipés pour quelques décideurs publics et quelques journalistes totalement identifiés et en cas de rupture d'embargo (statistique révélée avant l'heure de diffusion officielle), l'existence de procédures pour garantir l'égalité de traitement. Près de la moitié des pays européens avaient renoncé aux accès anticipés avant le 3^e cycle de revue.

Dans d'autres domaines, les revues peuvent ne pas être adaptées compte tenu de leur méthodologie et de leur rythme tous les sept à huit ans. Dans le cas du recours aux données privées, le levier a plutôt été les *task forces* ou les ESSnet³⁶.

³⁵ Voir les références juridiques en fin d'article.

³⁶ Groupe de travail ou projet de recherche financé par Eurostat.

Des conséquences concrètes en France

La conformité du service statistique public français a toujours été soulignée lors des revues, mais il peut pécher parfois par manque de formalisme. À cet égard, les revues ont été un soutien décisif à l'amélioration de la statistique publique française depuis l'adoption du CoP et une réelle opportunité pour obtenir un levier positif dans certains domaines. En effet, le caractère engageant des plans d'actions et le suivi réalisé annuellement permet d'objectiver les progrès même si après quelques années, on a parfois oublié la situation précédente : 15 ans après sa création, le rôle de l'ASP est connu et reconnu par les statisticiens publics.

Le soutien à l'amélioration est particulièrement visible sur la formalisation de l'indépendance professionnelle ou sur l'accès des chercheurs aux données de la statistique publique (1^{re} revue). Il a également permis de formaliser des approches de la qualité et les cadres d'assurance (2^e revue) et de progresser dans l'accès aux données administratives et privées (lors du 3^e cycle).

L'opportunité d'impliquer tous les acteurs concernés s'est traduite par la mise en œuvre d'une stratégie Qualité quinquennale au niveau du SSP à l'issue des deux dernières revues.

L'adhésion aux règles internationales, notamment à celles du Fonds Monétaire International (FMI) qui imposent aux pays de fournir une description des conditions de construction et de diffusion des données, et d'élaborer un calendrier de diffusion sur les quatre prochains mois, est ancienne. Mais l'harmonisation des règles au sein du SSP a bénéficié des revues par les pairs : politiques de gestion des révisions et des erreurs, publication des accès anticipés, partage au niveau du SSP d'un cadre de référence public sur le processus de diffusion des publications statistiques avec en particulier des accès anticipés plus limités et des transmissions plus sécurisées.

► Les enseignements



L'organisation du service statistique public en France avec l'Insee et ses directions régionales, les SSM, ses formations initiales et ses mobilités reste quasi unique en Europe.



La revue par les pairs s'apparente pour un pays à un arrêt sur image qui lui révèle ce qu'il est et ses forces et priorités d'amélioration.

En effet, le rapport élaboré à la suite d'une revue pointe les forces du système statistique national, forces révélées grâce au regard extérieur : l'organisation du service statistique public en France avec l'Insee et ses directions régionales, les SSM, ses formations initiales et ses mobilités reste quasi unique en Europe, alors que les gouvernances ternaires (service statistique

national, autorité indépendante, comité de concertation avec les utilisateurs) se sont généralisées. Le Danemark a, par exemple, renforcé sa loi en 2018, en clarifiant les rôles et les missions de l'INS, du statisticien national et du conseil de surveillance de Statistics Denmark. La Suisse a mis en place un Conseil de l'éthique pour les statistiques dès le début des années 2000 au rôle proche d'une autorité indépendante mais sans moyens de sanction.

Un autre enseignement est celui de la diversité des organisations en place, mais avec pour point commun le rôle central de l'Institut national de la statistique. Certains INS, à l'instar de l'Insee, ont également des missions régaliennes ; par exemple les INS de République tchèque et de Slovaquie sont responsables de l'établissement des résultats des élections, l'INS estonien est l'intendant général des données pour l'administration de son pays.

Réunis à Bucarest en novembre 2023, les pairs et les coordonnateurs nationaux ont partagé leurs bilans du 3^e cycle : maturité, honnêteté de l'exercice et bienveillance des pairs. Si la nécessité de pouvoir non seulement illustrer mais également prouver la qualité des systèmes nationaux et des statistiques officielles est partagée, les réflexions sur l'évolution de la méthodologie sont encore à mener.

Un des points de discussion réside dans les recommandations qui s'adressent aux gouvernements dont on constate la répétition d'une revue à l'autre, ce qui à terme risque d'affecter leur crédibilité. N'étant pas de la responsabilité directe du système statistique, le suivi par Eurostat de leur mise en œuvre est relativement inopérant. Ces recommandations sont de deux types : celles qui visent à renforcer l'indépendance du système statistique (que ce soit en la mentionnant explicitement dans la loi, ou dans les procédures de nomination/révocation du directeur de l'INS) et celles qui visent à accroître ou au moins assurer des ressources suffisantes au système statistique. Concernant les premières, on lui objecte souvent des raisons juridiques, telles que la constitution ou le statut de la fonction publique, pour ne pas les mettre en œuvre. Cependant, l'inscription dans la loi européenne de la nécessaire indépendance des institutions statistiques et la répétition des recommandations lors des revues successives a conduit plusieurs pays à modifier leurs lois nationales en ce sens.

L'impact des recommandations financières est plus complexe à analyser. La recherche de gains de productivité au travers de la modernisation des méthodes ou l'utilisation de nouvelles sources nécessite des investissements coûteux avant de produire leurs effets ; il est naturel de le mentionner dans le rapport de la revue si les pairs constatent que les institutions statistiques ne disposent pas des moyens, humains et financiers, pour réaliser ces investissements. Cependant, une telle recommandation, adressée aux autorités nationales, a peu de chances d'être audible dans le cadre de la préparation budgétaire nationale au cours de laquelle chaque administration exprime peu ou prou les mêmes arguments.

► Perspectives

Le thème retenu pour la conférence Qualité 2024³⁷ à Estoril était « le rôle des statistiques officielles en tant que pilier de la démocratie ». Pour atteindre cet objectif, il est indispensable que la confiance dans la qualité des statistiques produites mais également envers les institutions statistiques soit au rendez-vous.

37 <https://www.q2024.pt/>.



Le thème retenu pour la conférence Qualité 2024 à Estoril était « le rôle des statistiques officielles en tant que pilier de la démocratie ».



Dans le contexte actuel dominé par la multiplication des chiffres, des canaux et des informations plus ou moins vérifiées, les enjeux de confiance dans les statistiques officielles nécessitent plus de transparence sur nos valeurs mais également sur les données mobilisées, sur les processus et sur les méthodes pour chacun des pays et pour le collectif européen.

Deux questions se posent :

- Faut-il faire évoluer le Code de bonnes pratiques ?
- Faut-il reconduire les revues à l'identique ou les faire évoluer ?

Sur la révision du CoP, le constat est partagé sur l'aspect peu communicant en externe du CoP avec ses 16 principes et ses 84 indicateurs. Dès lors, deux approches s'opposent. La première préconise de conserver le Code comme boussole unique pour les statisticiens publics européens et la communication externe, et par conséquent de continuer à l'enrichir. La seconde propose de séparer le CoP en deux : une version simplifiée avec une approche sur les valeurs destinée à l'extérieur et une version en interne au système statistique européen, plus opérationnelle, comme référence pour la qualité des systèmes statistiques nationaux. Concernant les valeurs, les six préconisées par l'UNECE³⁸ en 2023 (indépendance professionnelle, transparence, pertinence, impartialité, respect de la confidentialité et collaboration) apparaissent comme plus adaptées pour communiquer.

Quel avenir des revues par les pairs ? Faut-il continuer sur le mode actuel ou conduire l'évaluation des systèmes statistiques nationaux autrement ? Le risque n'est-il pas que les systèmes s'adaptent à la méthodologie en réalisant des améliorations cosmétiques plutôt qu'en profondeur ? On pourrait par exemple imaginer des revues ciblées et approfondies sur les seuls points négatifs ou simplement à améliorer mis en évidence par la précédente revue, plutôt qu'une revue de l'ensemble des principes du Code. Le récent cycle a quelque peu ouvert la voie en donnant la possibilité aux pairs de renforcer la revue sur certains principes plus faibles lors de l'analyse du questionnaire d'autoévaluation et des documents support. De même, faut-il laisser le soin à l'INS de choisir les personnes invitées lors des réunions, avec le risque de n'avoir lors de ces réunions que des « supporters » de l'INS, même s'il semble que ce risque ne se soit pas réellement avéré lors des précédents cycles ?

Ces travaux ne font que commencer au niveau européen pour préparer le prochain cycle.

38 UNECE : Commission économique des Nations unies pour l'Europe (et l'Amérique du nord).

► Fondements juridiques

- Traité sur le fonctionnement de l'Union européenne (version consolidée), article 338 paragraphe 2. In : *site de l'Union européenne*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12016M/TX>.
- Règlement (CE) n° 223/2009 révisé du Parlement européen et du Conseil du 11 mars 2009 relatif aux statistiques européennes. In : *site de l'Union européenne*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1724682778358&uri=CELEX%3A32009R0223>.
- Règlement (UE) n° 557/2013 de la Commission du 17 juin 2013 mettant en œuvre le règlement (CE) n° 223/2009 du Parlement européen et du Conseil relatif aux statistiques européennes en ce qui concerne l'accès aux données confidentielles à des fins scientifiques. In : *site de l'Union européenne*. [en ligne]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32013R0557&qid=1727019849934>.
- Loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. In : *site de Légifrance*. Mise à jour le 25 mars 2019. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573>.
- Recommandation relative à une décision du Conseil sur l'existence d'un déficit excessif en Grèce, 2004. In : *site de Légifrance*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1724740719253&uri=CELEX%3A52004SC0815>.
- Code Général des Collectivités Territoriales - Articles L.2223-42 et R2213-1-1 – R2213-1-4. In : *site de Légifrance*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006070633/LEGISCTA000006192272/2022-08-01/?anchor=LEGIARTI000038886319#LEGIARTI000038886319.

► Bibliographie

- ALEXANDRE, Véronique et CLING, Jean-Pierre, 2019. Le système statistique allemand – Une évolution facilitée par la réglementation européenne. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 140-150. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254168?sommaire=4254170>.
- 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 11 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5398693?sommaire=5398695>.
- 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 11 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5008696?sommaire=5008710>.
- 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 11 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5008698?sommaire=5008710>.
- ERNOULT, Thomas, 2012. Maquillage des comptes publics grecs : statistiques à la dérive ? In : *Regards croisés sur l'économie*. N° 11, pp.113-115. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://shs.cairn.info/revue-regards-croises-sur-l-economie-2012-1-page-113?lang=fr&ref=doi>.
- EUROSTAT, 2023. Revue par les pairs: cycle actuel 2021-2023. In : *site de Eurostat*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://ec.europa.eu/eurostat/web/quality/peer-reviews/current-round-2021-2023>.
- 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. Insee. N° N3, pp. 76-92. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254227?sommaire=4254170>.
- HAND, David J., 2018. Statistical challenges of administrative and transaction data. In : *J. R. Statist. Soc. A* 181, Part 3, pp. 555–605. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : https://www.researchgate.net/publication/340548882_Statistical_challenges_of_administrative_and_transaction_data_we_participated_in_the_Discussion_on_the_Paper_by_Hand.
- INSEE, 2022. La troisième revue européenne par les pairs en 2021. In : *site de l'Insee*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/5357958>.
- INTERNATIONAL STATISTICAL INSTITUTE, 2010. Declaration on Professional Ethics. In : *site de l'ISI*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://isi-web.org/declaration-professional-ethics>.

- MAUGUIN, Jocelyne et SAGNES Nicolas, 2024. Faciliter l'accès aux données de l'Insee – Cubes, catalogue et métadonnées. In : *Courrier des statistiques*. [en ligne]. 8 juillet 2024. Insee. N° N11, pp. 31-50. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8203036?sommaire=8203072>.
- UNITED NATIONS, 2014. Fundamental Principles of National Official Statistics. In : *site des Nations Unies*. [en ligne]. [Consulté le 11 septembre 2024]. Disponible à l'adresse : <https://unstats.un.org/fpos/>.



PRÉSENTATION DU NUMÉRO N12

Dans ce numéro N12, plusieurs services statistiques ministériels (SSM) présentent leur domaine d'activité. Le premier article nous fait découvrir les statistiques du commerce extérieur, de son origine à nos jours, un récit ponctué d'anecdotes surprenantes. Puis le CépiDc (Centre d'épidémiologie sur les causes médicales de décès) de l'Inserm explique pourquoi la mesure précise des causes de décès est essentielle. Pour cela, il est nécessaire de coder les certificats de décès en utilisant des méthodes de plus en plus automatisées. Comment réaliser des indicateurs sur l'insertion professionnelle des étudiants quittant l'enseignement supérieur ? Un troisième article décrit le dispositif InerSup, fondé sur des appariements successifs de fichiers.

On change d'environnement avec un article sur les statistiques de la culture. Depuis plus de 60 ans, les enquêtes objectivent les pratiques culturelles afin d'aider aux politiques. Les innovations technologiques et les enjeux économiques sont décrits minutieusement. Retour à l'Insee avec un article socle sur les statistiques structurelles d'entreprises et leur évolution depuis les années 70 à nos jours.

L'article suivant illustre l'utilisation des données bancaires à des fins statistiques pour documenter la situation financière des Français et met en évidence les opportunités et limites des données privées pour la statistique publique.

Enfin, un article institutionnel explique le processus de la revue par les pairs, démarche menée au niveau européen pour s'assurer du respect du Code des bonnes pratiques.



ISSN 2107-0903

ISBN 978-2-11-162461-0

