



COURRIER

DES STATISTIQUES

N1 - Décembre 2018

Courrier des statistiques N1

SOMMAIRE

Éditorial <i>Jean-Luc Tavernier</i>	3
Avant-propos <i>Pascal Rivière</i>	4
Présentation du numéro <i>Lucile Olier</i>	5
Un système statistique intégré à l'administration centrale <i>Jean-Luc Tavernier</i>	7
• Dossier : le statisticien et les sources administratives	13
• Utiliser les déclarations administratives à des fins statistiques <i>Pascal Rivière</i>	14
• La DSN : nouvelle référence pour les échanges de données sociales des entreprises vers les administrations <i>Elisabeth Humbert-Bottin</i>	25
• Bien comprendre la déclaration sociale nominative pour mieux mesurer <i>Catherine Renne</i>	35
• De l'activité de la justice au suivi du justiciable : faire parler les données de gestion <i>Christine Chambaz</i>	45
Le <i>Legal Entity Identifier</i> : contexte international et rôle de l'Insee <i>Pierrette Schuhl</i>	58
Qu'entend-on par statistique(s) publique(s) ? <i>Michel Isnard</i>	69

Rédaction en chef

Lucile Olier

Contribution

Insee : Michel Isnard, Catherine Renne,
Pascal Rivière, Pierrette Schuhl,
Jean-Luc Tavernier

SSM Justice : Christine Chambaz

Gip-MDS : Elisabeth Humbert-Bottin

Directeur de la publication

Jean-Luc Tavernier

Directeur de la collection

Pascal Rivière

Rédaction

Pierre Glénat, Lucile Olier,
Pascal M. Rivière

Composition

Agence **LATITUDE** Nantes

0571/18

02 51 25 06 06

www.agence-latitude.fr

Photo de couverture :

Fotolia®

Éditeur

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

www.insee.fr

© Insee 2018 « Reproduction partielle
autorisée sous réserve de la mention
de la source et de l'auteur ».



ÉDITORIAL

En janvier 1977 paraissait le premier numéro du *Courrier des statistiques*, préfacé par Edmond Malinvaud. Il écrivait alors : « Grâce à cette revue, chaque statisticien devrait, même s'il a une fonction très spécialisée, être averti des travaux de ses collègues et des problèmes qu'ils rencontrent. Information générale donc sur les différentes activités du système statistique, mais aussi information plus pointue sur l'évolution des méthodes et des moyens de l'activité statistique. »

Cette aventure s'est interrompue en septembre 2011, après plus de 130 numéros publiés. À l'ère d'internet de l'Open data et de l'Open government, le projet d'origine semblait obsolète : chacun ne peut-il aujourd'hui trouver facilement l'information dont il a besoin en ligne ? L'Insee, les services statistiques ministériels, ont des sites internet très fournis, ils diffusent largement leurs travaux et leur méthodologie et donnent à voir leur organisation à qui prend la peine de chercher. Pourquoi cela ne suffirait-il pas ? L'étude de l'opportunité de (re)créer une revue professionnelle faite par l'Inspection générale de l'Insee fin 2017 a montré qu'il fallait revenir sur ce postulat. L'abondance de l'information peut être en elle-même un obstacle à son appropriation. Et surtout, nous nous illusionnons sur notre capacité collective à produire spontanément une information complète et partageable par tous – et d'abord au sein de notre communauté professionnelle – sur nos méthodes, nos enjeux, notre organisation et son contexte. Certes, nous produisons une documentation abondante, mais elle reste souvent très technique, confinée dans un petit cercle de spécialistes. L'étude d'opportunité de l'Inspection générale a montré que nombreux étaient les cadres du service statistique public qui regrettaient la disparition de la revue et qui estimaient qu'il manquait désormais un chaînon essentiel dans la chaîne d'information que nous produisons. Le besoin de faire circuler l'information au-delà des spécialités, de présenter de la méthodologie d'une manière plus synthétique et pédagogique que nous ne le faisons dans nos documents de travail, de donner à voir aussi notre organisation, le rôle de coordination du système statistique national consacré par le règlement européen n° 223/2009 et les évolutions dans la construction de la statistique européenne, le besoin enfin de garder la mémoire de ce que nous sommes et de ce que nous faisons, ces besoins sont toujours là. Et rien ne les satisfait spontanément.

Enfin, en cette ère de post-vérité, nous sommes plus que jamais tenus, devant nos concitoyens, à la rigueur et l'impartialité de nos méthodes. Le *Courrier des statistiques* permettra à tous ceux qui le souhaitent, enseignants, chercheurs, étudiants, citoyens curieux, de mieux s'informer sur la fabrique de nos statistiques et le cadre déontologique dans lequel elles se construisent. Ce qui garantit précisément que l'on ne fait pas dire aux chiffres ce que l'on veut.

C'est pourquoi le comité de direction de l'Insee a décidé de relancer le *Courrier des statistiques*. J'en ai confié la responsabilité à l'Inspection générale, qui en assume désormais la rédaction en chef. J'ai été heureux de constater que les bonnes volontés ont été nombreuses à se manifester pour alimenter ce premier numéro et les suivants. Le *Courrier des statistiques* est en effet notre bien commun. Sa pérennité repose sur la bonne volonté de tous.

Je vous souhaite une bonne lecture de ce premier numéro.

Jean-Luc Tavernier
Directeur général de l'Insee

“ Il faut que tout change
pour que rien ne change. ”

Lampedusa T., Le Guépard

Le Courrier des statistiques a accompagné des générations de statisticiens pour lesquels il était une référence. À sa lecture, on mesure à la fois la permanence des grandes problématiques du système statistique public et notre capacité collective à évoluer et innover pour coller aux besoins de notre temps. Sa mise en sommeil il y a quelques années avait laissé un certain nombre d'entre nous orphelins. Ce numéro inaugure donc le retour de cette revue emblématique.

Qui dit renouveau du Courrier des statistiques dit nouvelle numérotation (le lecteur en appréciera la subtilité), et surtout nouvelle maquette et nouveau format, avec une édition complètement numérique, accessible sur notre site internet. La Revue sortira deux fois par an, en décembre et en juin. Pour ne pas manquer un nouveau numéro, vous pouvez vous abonner aux avis de parution.

Mais pour l'essentiel, c'est-à-dire l'esprit, le Courrier des statistiques reste dans la lignée de son prédécesseur... tout en se renouvelant. Vous trouverez ainsi dans la revue des articles sur les méthodes de la statistique publique, ses grandes opérations, ses projets, son organisation, le cadre juridique dans lequel elle s'exerce, mais aussi désormais des éclairages plus fréquents sur l'activité d'autres Instituts nationaux de statistique, des réflexions prospectives, et une ouverture sur les nouvelles sources de données. Si le Courrier des statistiques ne sera pas systématiquement disponible en anglais, les articles que vous souhaitez traduire ou faire traduire, parce que susceptibles d'intéresser nos collègues étrangers, pourront aussi être mis en ligne.

La revue est ouverte à tous les statisticiens publics et aux experts français et étrangers qui s'y intéressent et qui ont des idées d'articles. Pour atteindre le niveau d'exigence souhaité en matière de rigueur et clarté du propos, la rédaction en chef vous accompagnera dans la phase d'écriture, de validation et de relecture. Un petit guide vous donnera les principales consignes de présentation et bonnes pratiques de rédaction à respecter afin que votre article informe et intéresse vos lecteurs, sur les sujets qui vous tiennent à cœur, sans perdre ni lasser.

Comme nous le rappelle le Directeur général de l'Insee dans son éditorial, nous avons envers nos collègues, mais aussi nos concitoyens, un devoir de pédagogie sur nos méthodes, et pas seulement sur nos statistiques et études. En écrivant dans le Courrier des statistiques, vous participerez à la transparence que nous leur devons. Qui mieux que vous peut contribuer à cette entreprise éminemment démocratique ?

Et maintenant, chers auteurs, à vos plumes et claviers...

Pascal Rivière
Directeur de la collection, Insee

PRÉSENTATION DU NUMÉRO

La statistique n'est-elle qu'un éternel recommencement ? Il est fascinant, en parcourant le sommaire du premier numéro du Courrier des statistiques, paru en janvier 1977, de constater la permanence d'un certain nombre de grands thèmes. Presque tous les sujets traités dans ce numéro y étaient déjà abordés, certes sous un angle assez différent de celui d'aujourd'hui. Il faut bien être de son temps. Mais au fond, ce sont les mêmes questions qui déjà se posaient au statisticien.

Michel Volle, dans un court article intitulé « Pourquoi faire l'histoire de la statistique ? », rappelait que « la statistique est impliquée, très naturellement, dans des problèmes de pouvoir ». Et c'est précisément cette question que traite dans le premier article de ce numéro Jean-Luc Tavernier, en s'interrogeant sur la juste distance de la statistique publique à l'exécutif. Il montre que les compromis nationaux peuvent finalement s'incarner dans des formes organisationnelles très variées et que le cas français est plutôt satisfaisant si on le juge à l'aune de la confiance que nos concitoyens accordent à la statistique publique.

D'une certaine façon, c'est aussi cette question que traite Michel Isnard dans l'article qui clôture ce numéro. Il donne à voir la façon dont ces compromis s'organisent différemment à l'échelon européen et national, contribuant à une définition somme toute relativement élastique de la statistique publique. Le lecteur pourra aborder ces questions finalement assez épineuses en mettant ses connaissances du sujet à l'épreuve d'un petit quiz.

Regarder plus de 40 ans en arrière permet aussi de mesurer le chemin parcouru, notamment en matière de statistique administrative. En 1977, un article était consacré au Cerfa, qui, on l'a oublié aujourd'hui, a été rattaché, à sa création, à l'Insee, compte tenu « de son expérience (...) dans l'établissement de questionnaires d'enquêtes et de l'intérêt statistique de certains documents administratifs », comme le rappelait l'auteur, René Marbot. L'Insee se félicitait d'ailleurs d'avoir réduit d'un quart en 10 ans le nombre de formulaires administratifs. Il cédait cette année-là ses prérogatives sur le Cerfa au Secrétariat général du gouvernement, afin que le gouvernement dispose de tous les outils nécessaires à l'entreprise de simplification administrative qu'il entendait intensifier.

Dans le dossier consacré dans ce numéro aux sources administratives, Élisabeth Humbert-Bottin, la directrice du Groupement d'intérêt public pour la Modernisation des déclarations sociales (Gip-MDS) nous explique en quoi consiste la Déclaration sociale nominative (DSN) et la simplification qu'elle a permise, au prix d'un changement complet de paradigme par rapport à celui des Cerfa : en se rapprochant du fait générateur, qu'est la paie, l'allègement de la charge administrative ne se traduit pas par une perte de qualité. Elle met aussi en évidence le rôle prééminent de la normalisation des flux d'échanges opéré par le Gip dans son rôle de concentrateur, rôle qui est largement monté en puissance depuis la généralisation de la DSN en janvier 2017.

Pascal Rivière souligne à son tour dans son article, qui aborde le sujet de l'exploitation statistique des déclarations administratives de manière générale, le rôle clé de l'organisme concentrateur. Il constate en particulier que le processus de traitement des déclarations administratives présente de nombreux points communs avec le processus de production statistique. Les statisticiens seraient bien avisés d'en tirer tous les avantages, ce qui passe par une forte collaboration avec l'organisme qui concentre les flux déclaratifs y compris dans la gestion fine des anomalies.

La DSN est d'ores et déjà incontournable pour le système statistique public, et Catherine Renne nous explique comment elle est utilisée. Pour mieux maîtriser la qualité des données, l'importance d'une participation active aux instances de gouvernance et d'une connaissance

fine de la norme y est soulignée. Mais l'auteur nous montre aussi la nécessité de vérifications et transformations effectuées en aval pour les besoins statistiques, et d'une véritable industrialisation des processus.

Ce dossier spécial comporte également un article sur les statistiques de la Justice de Christine Chambaz, sujet qui était également traité dans le tout premier numéro de 1977, par le chef du SSM de l'époque. La Chancellerie venait de se doter d'un Plan de rénovation de son dispositif statistique et s'attachait en priorité à la statistique judiciaire. L'article de Christine Chambaz fournit un précieux contrepoint aux articles du dossier. Ici, ni déclaration administrative ni formalisation rigoureuse des flux. Le statisticien est confronté à un ensemble complexe de données réparties entre silos, selon divers logiciels de gestion. Pour un usage statistique, elles requièrent mise en cohérence globale, normalisation *a posteriori* (définitions, unités statistiques, nomenclatures), et une façon de figer dans le temps des données vivantes, car liées à des besoins de gestion.

Enfin, l'article de Pierrette Schuhl illustre la façon dont les prérogatives de la statistique publique peuvent se renouveler et s'étendre en prenant appui sur des compétences anciennes et reconnues. En France, l'Insee a été proposé dès 2013 par le ministère de l'Économie et des Finances pour être émetteur local de *Legal Entity Identifier* (LEI), en raison de son expérience dans la gestion de répertoires et d'identifiants. L'article retrace les grandes étapes de la mise en place de ce dispositif mondial d'identifiant unique des intervenants sur les marchés financiers, au lendemain de la crise financière de 2007 et explique le rôle de l'Insee. Les données d'identification des entreprises ainsi collectées sont porteuses d'améliorations significatives de la connaissance de la composition des groupes, de leur assise financière et potentiellement de leur stratégie.

Lucile Olier
Rédactrice en chef, Insee

UN SYSTÈME STATISTIQUE INTÉGRÉ À L'ADMINISTRATION CENTRALE

Jean-Luc Tavernier*

Une préoccupation majeure en matière de statistiques publiques est l'indépendance, l'impartialité et l'objectivité de leur élaboration. Dans beaucoup de pays, cette indépendance a été consacrée en droit interne. Mais ce n'est pas le cas partout, notamment en France où l'Insee comme les services statistiques ministériels demeurent des services ministériels. L'indépendance institutionnelle est-elle une condition nécessaire à l'indépendance professionnelle qui finira par s'imposer ? La réponse apportée ici se veut nuancée. L'Insee a toujours relevé du pouvoir exécutif ; cela n'empêche pas le système statistique français d'être fiable et indépendant de toute intervention politique, et d'être perçu comme tel par le public. Il s'agit d'une indépendance de facto et non de jure, même si le cadre juridique protégeant l'indépendance professionnelle de la statistique publique doit continuer de se renforcer.

Depuis sa naissance il y a plus de 70 ans, l'Insee relève du pouvoir exécutif. En effet, l'Insee et les 16 autres principaux pourvoyeurs de statistiques publiques en France sont rattachés à divers ministères français. L'Insee lui-même est une direction du ministère de l'Économie et des Finances¹. Ce positionnement au sein des services mêmes de l'exécutif tend à devenir minoritaire, beaucoup de pays ayant conféré à leur institut statistique une indépendance institutionnelle pour mieux garantir son indépendance professionnelle. La majorité deviendra-t-elle la règle ? Ou peut-on imaginer que perdurent des instituts statistiques ayant le statut d'administration d'État. C'est à cette question qu'était consacrée une session du *World Statistics Congress* de l'*International Statistical Institute* en 2017². Et c'est mon intervention à ce congrès qui est développée dans cet article.

🕒 UNE INDÉPENDANCE DE L'INSEE FAIBLE DE JURE, FORTE DE FACTO

La notion d'indépendance est absente de la version initiale des textes fondateurs de l'Institut, que ce soient la loi de finances du 27 avril 1946, le décret³ d'application du 14 juin 1946 ou la loi⁴ du 7 juin 1951. C'est seulement en 2008 que la loi dite de modernisation de l'économie⁵ (ou loi LME) introduit pour la première fois la notion d'indépen-

* Directeur général, Insee.

1. Cf. l'article 32 de la loi de finances du 27 avril 1946 qui stipule qu'« il est créé au ministère de l'économie nationale un institut national de la statistique et des études économiques pour la métropole et la France d'outre-mer ».

2. La session était intitulée : « Does the production of official statistics need to be a separate branch of government ? ».

3. Décret n° 46-1432 du 14 juin 1946 portant règlement d'administration publique pour l'application des articles 32 et 33 de la loi de finances du 27 avril 1946 relatifs à l'institut national de la statistique et des études économiques pour la métropole et la France d'outre-mer.

4. Loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques.

5. Loi n° 2008-776 du 4 août 2008 de modernisation de l'économie.

dance professionnelle dans la loi de 1951 et crée l'autorité de la statistique publique (ASP). Plus précisément, l'article 1^{er} de la loi de 1951 modifiée stipule, à partir de cette date, que « *La conception, la production et la diffusion des statistiques publiques sont effectuées en toute indépendance professionnelle* ». Cependant, *de jure*, dans l'ordre juridique national, l'Insee et les services statistiques ministériels (SSM) restent des administrations comme les autres.

“ *En aucun cas les ministres ou leurs conseillers politiques ne demandent à l'Insee de modifier le contenu de ses communiqués de presse, qu'il s'agisse de chiffres ou de commentaires, ce qui a son importance, car l'Insee a une longue expérience de la « communication narrative* ». ”

Pour autant, l'Insee diffuse de longue date ses statistiques en toute indépendance, en annonçant ses publications à l'avance, en s'appuyant sur des embargos stricts et en présentant ses résultats aux médias de manière autonome. L'Insee adhère à la norme de diffusion des données statistiques (dite SDDS+) du FMI relative à la diffusion des indicateurs conjoncturels, et ce depuis 1993. Cette norme définit notamment les règles d'accès préalables à la diffusion : elle n'interdit pas l'accès sous embargo à certaines informations économiques mais

cet accès est limité, contrôlé et rendu public sur le site de l'Institut. Il permet au Gouvernement de prendre connaissance des indicateurs quelques heures avant leur publication ; cette pratique autrefois très majoritaire, est encore partagée par d'autres pays.

L'indépendance professionnelle de la statistique publique a été confortée par l'élargissement du calendrier prévisionnel des indicateurs macroéconomiques à toutes les statistiques publiées par l'Insee, puis, ces dernières années, à son extension à l'immense majorité des statistiques publiées par les SSM.

En aucun cas les ministres ou leurs conseillers politiques ne demandent à l'Insee de modifier le contenu de ses communiqués de presse, qu'il s'agisse de chiffres ou de commentaires, ce qui a son importance, car l'Insee a une longue expérience de la « communication narrative ».

S'il y a parfois eu des tentatives d'ingérence politique dans les travaux de l'Insee, elles n'ont jamais abouti. La dernière tentative remonte à 1990 et concerne l'indice des prix à la consommation. Le gouvernement fait alors adopter par l'Assemblée nationale française une loi contre la consommation de tabac et d'alcool, qui entraîne une forte hausse du prix du tabac (hausse qui se poursuit sans discontinuer depuis). Plusieurs types de revenus étant indexés sur l'indice des prix à la consommation, notamment le salaire minimum, le gouvernement ne souhaite pas que cette hausse du prix du tabac ait un impact inflationniste du fait de l'indexation des salaires. C'est ainsi que la loi finit par interdire à l'Insee de faire figurer le tabac dans la liste des produits inclus dans l'indice des prix à la consommation.

L'Insee proteste. Le ministre des Finances, auquel l'Insee rend compte, se joint à la protestation et la loi est rapidement modifiée début 1991. Plutôt que de retirer le tabac de l'indice des prix à la consommation, le Parlement décide de manière plus rationnelle d'indexer le salaire minimum – et certains autres types de revenus et prestations – sur la base d'un indice des prix à la consommation hors tabac. Ce dernier fait l'objet d'un calcul par l'Insee depuis 1992, parallèlement à l'indice courant des prix à la consommation. Depuis, c'est-à-dire au cours des 26 dernières années, il n'y a eu aucune intervention gouvernementale notable dans le travail de l'Insee.

DES STATISTIQUES DE L'INSEE LARGEMENT RECONNUES

Selon la dernière enquête représentative réalisée en 2016 par un institut privé indépendant (IPSOS) pour l'Insee auprès de deux échantillons de 1 000 personnes en deux vagues (mai et septembre) :

- 80 % des Français ont une opinion favorable de l'Insee ;
- 65 % des personnes interrogées estiment que l'information est indépendante du pouvoir politique, contre 60 % en 2015 ; bien que les Français ne connaissent pas précisément la configuration institutionnelle de l'Insee, ils ont bien conscience de la grande indépendance de fait dont il jouit ;
- environ deux tiers des Français ont confiance dans les chiffres publiés par l'Insee sur la situation économique et sociale de la France (un dernier pourcentage d'autant plus frappant que, si l'on pose la même question sur ces indicateurs économiques et sociaux sans préciser qu'ils proviennent de l'Insee, ce pourcentage chute à 42 % seulement).

Certes, ce niveau de confiance varie en fonction des indicateurs ; par exemple, seul un tiers des personnes interrogées estiment que les statistiques sur le taux de chômage produites par l'Insee reflètent fidèlement la situation du chômage en France. Mais globalement, cela confirme que la « marque » Insee est crédible aux yeux de la société française, et que les données et les études de l'Insee sont perçues comme étant réalisées de façon indépendante et sans intervention du pouvoir politique.

L'AUTORITÉ DE LA STATISTIQUE PUBLIQUE, GARDIENNE DE L'INDÉPENDANCE

L'Insee travaille donc *de facto* en toute indépendance, bien qu'il ne soit pas indépendant *de jure*. Cependant, le cadre juridique s'est renforcé, sous l'impulsion notamment des textes européens (*encadré*). La création d'une autorité indépendante en 2009, en application de la loi LME, a constitué une première évolution. L'article 1 de la loi de 1951 modifiée indique en effet que « (...) *L'Autorité de la statistique publique [qui] veille au respect du principe d'indépendance professionnelle dans la conception, la production et la diffusion de statistiques publiques ainsi que des principes d'objectivité, d'impartialité, de pertinence et de qualité des données produites* ». Elle est composée de 9 membres nommés pour 6 ans. Elle veille à la parfaite indépendance professionnelle de la préparation des statistiques publiques, conformément aux principes fondateurs du « Code de bonnes pratiques de la statistique européenne ». Elle peut, par exemple, intervenir publiquement en cas de violation des embargos par le gouvernement. Elle rédige un rapport annuel sur l'activité statistique publique, qui est soumis à l'Assemblée nationale française et rendu public. L'ASP a été créée avant la révision en 2015 du règlement n° 223/2009 relatif aux statistiques européennes. Il convient donc d'adapter et de renforcer sa mission afin de tenir compte de cette révision.

UNE COMMUNAUTÉ PROFESSIONNELLE SOUDÉE PAR UNE CULTURE ET DES VALEURS COMMUNES

Les garanties apportées par l'ASP sont-elles suffisantes ? La situation de l'Insee ne demeure-t-elle pas ambiguë au regard du droit national ? Certes, le statut d'administration centrale peut encore parfois être perçu – plutôt que vécu – comme limitant l'indépendance. Mais à tout prendre, entre une indépendance *de facto* dans un environnement juridique à parfaire, et une indépendance *de jure* de façade qui serait vulnérable aux changements politiques comme cela s'observe encore dans certains pays, la première option est de loin préférable.

Encadré. La notion d'indépendance des statistiques dans les textes européens

L'idée d'indépendance des statistiques européennes apparaît pour la première fois dans le traité d'Amsterdam* signé en 1997. Celui-ci stipule dans le second alinéa de son article 285 que « *L'établissement des statistiques se fait dans le respect de l'impartialité, de la fiabilité, de l'objectivité, de l'indépendance scientifique, de l'efficacité au regard du coût et de la confidentialité des informations statistiques; il ne doit pas entraîner de charges excessives pour les opérateurs économiques.* » Ces dispositions sont reprises dans les différents traités qui suivent : Nice, Lisbonne et les traités actuels. Elles figurent actuellement l'article n° 338 du traité de fonctionnement de l'Union européenne. Bien que cela ne soit pas explicite dans ces textes, cette disposition ne s'applique qu'aux statistiques entrant dans le champ de l'Union européenne.

En parallèle à cette disposition du traité d'Amsterdam, mais avant sa mise en œuvre effective qui se situe en 1999, le règlement n° 322/97 qui parle de statistiques européennes mentionne (article 10) des principes que la production de statistiques européennes doit suivre (impartialité, fiabilité, pertinence, coût-efficacité, secret statistique et transparence). Le terme « indépendance » n'est employé que dans le cadre du principe d'impartialité (impartialité est une manière objective et indépendante de produire des statistiques communautaires...). Là encore, on ne parle que de statistiques communautaires, c'est-à-dire de production de données dans le cadre du programme statistique communautaire (article 2). Ces deux textes aboutissent au Code de Bonnes Pratiques en 2005 qui est une recommandation de la Commission, mais que les règlements ultérieurs rendent plus ou moins obligatoires.

Dernière étape européenne : le règlement n° 223/2009 donne une existence légale au Code de Bonnes Pratiques et l'indépendance professionnelle est introduite comme principe autonome par un règlement de 2015 (n° 2015/759). Il y est explicitement indiqué que la production de statistiques européennes doit suivre les principes du code.

Rien n'est dit dans le règlement de 1997 concernant la nomination des dirigeants des INS, qui est une question sensible pour ce qui regarde l'indépendance des INS. Dans le code de bonnes pratiques de 2005, il est simplement dit que « *Le chef de l'autorité statistique a un rang hiérarchique suffisamment élevé pour lui permettre d'avoir des contacts à haut niveau au sein des administrations et organismes publics. Son profil professionnel doit être du plus haut niveau.* ». Le règlement de 2009 ne dit rien sur la nomination, mais lorsque le règlement de 2015 le modifie, il est créé un article 5bis qui précise les conditions de nomination et d'exercice.

*Le traité de Maastricht ne parle ainsi de statistiques qu'en lien avec l'Institut Monétaire Européen (ancêtre de la BCE), pour affirmer son indépendance.

L'indépendance repose-t-elle trop sur la personnalité du directeur général ? Bien entendu, celui-ci a un rôle emblématique à jouer en la matière. Toutefois, je souhaite insister sur le fait que l'indépendance de l'Insee est d'abord solidement ancrée dans les pratiques et dépend assez peu des individualités.

Les pratiques réglant la nomination du directeur général en sont un bon exemple. Certes, celle-ci est à la discrétion du gouvernement. Le directeur général de l'Insee est nommé en Conseil des ministres, comme tous les directeurs généraux des services ministériels français. Il peut être remplacé du jour au lendemain (ou d'un mercredi à l'autre, à l'occasion du Conseil des ministres hebdomadaire). Il ne rend compte qu'au ministre de l'Économie et des Finances. Mais il est, depuis toujours, un statisticien, ou un économiste avec

une expérience de statisticien. Des dispositions réglementaires nouvelles confortent depuis peu cette pratique ancienne. Le décret n° 2018-800 du 20 septembre 2018 modifie le décret n° 2009-250 du 3 mars 2009 relatif à l'ASP pour tenir compte en particulier de la révision en 2015 du règlement n° 223/2009 relatif aux statistiques européennes. Il introduit dans l'article 1^{er} un alinéa 3bis qui dispose que l'ASP « émet un avis, à l'occasion de la nomination du directeur général de l'Institut national de la statistique et des études économiques et de celle des responsables de services statistiques ministériels qui sont directeurs d'administration centrale, à l'attention du comité d'audition⁶ (...). Cet avis porte sur les compétences des personnes dont la nomination est envisagée au regard du principe d'indépendance professionnelle énoncé par le code de bonnes pratiques de la statistique européenne⁷ (...). Le sens de l'avis est publié au Journal officiel en même temps que l'acte de nomination ».

Mais la meilleure protection de l'institut est sans aucun doute la force de la cohésion et de l'homogénéité de son personnel : l'ensemble des responsables et des agents sont des fonctionnaires qui partagent la même formation, la même culture et la même éthique. En un mot, c'est la communauté des statisticiens-économistes qui est la mieux placée pour garantir la pérennité de ces valeurs. Cette culture s'enracine dans une histoire toujours très présente et se forge notamment dans le creuset des écoles de la statistique publique. Elle se transmet par l'exemple et la pratique dès les premiers postes et tout au long de la carrière. Elle fait l'objet d'un contrôle interne vigilant par la communauté de travail et les représentants du personnel.

La force de notre culture professionnelle ne doit pas être sous-estimée. Depuis toujours, la coutume est l'une des premières sources du droit, qu'elle précède souvent. Seltzer (1994) souligne d'ailleurs qu'une longue tradition d'intégrité statistique est l'une des clés d'une statistique indépendante, tout autant sinon plus que la loi. D'une certaine façon, les évolutions législatives et réglementaires de la dernière décennie entérinent et confortent des pratiques anciennes.

Voici un bon exemple de la force de notre culture professionnelle et de la réputation qui est la sienne bien au-delà des frontières du SSP : le Premier ministre et le ministre de l'Intérieur se sont publiquement et à plusieurs reprises félicités que le nouveau service statistique du ministère de l'Intérieur soit placé sous la responsabilité d'un directeur issu de l'Insee. Cela signifie, implicitement, que s'attacher les services d'un tel directeur garantit en soi l'indépendance du service statistique concerné. Les médias ont, par ailleurs, également relayé cette idée auprès du grand public.

LA CONFIANCE DE NOS CONCITOYENS, NOTRE MEILLEURE GARANTIE

En guise de conclusion, faut-il modifier l'organisation du système statistique français et le statut de l'Insee pour accroître notre indépendance ? Cette question mérite que l'on y réfléchisse à deux fois.

Il est bien évident qu'en dépit de cette indépendance *de facto* dont nous bénéficions, il serait préférable de consolider le cadre juridique existant. Cette garantie reste toutefois très théorique et nombre de commentateurs ne se privent pas de souligner que l'on ne peut en toute rigueur démontrer que l'indépendance institutionnelle est une garantie nécessaire et suffisante de l'indépendance professionnelle.

6. Ce comité est établi en application du décret n° 2016-663 du 24 mai 2016 portant création d'un comité d'audition pour la nomination des directeurs d'administration centrale.

7. Ce principe est prévu à l'article 2 du règlement européen (CE) n° 223/2009 du 11 mars 2009.

Cela aurait en contrepartie de graves inconvénients. Le premier, en éloignant les statisticiens publics du premier utilisateur (mais non le seul) de leurs travaux qu'est l'exécutif, est le risque d'affaiblir leur pertinence, notamment pour ce qui concerne les études économiques. Le second serait de compliquer le rôle de coordinateur du système statistique public qui est dévolu à l'Insee et qui a été inscrit dans le décret fondateur du 14 juin 1946 en application du règlement européen n° 223/2009.

En définitive, ce qui peut apparaître préjudiciable dans le statut d'administration centrale, ce n'est pas une limite à l'indépendance fonctionnelle, c'est surtout les contraintes et les lourdeurs liées à l'allocation et au suivi des moyens budgétaires. Dans un cadre d'administration centrale, l'annualité budgétaire est la règle ; le processus de discussion budgétaire se répète chaque année, jusqu'à l'épuisement. Les aléas de l'exécution budgétaire se manifestent, de plus en plus dans le contexte de plus en plus contraint pour les finances publiques ces dernières années. Or, pour un organisme tel qu'un institut statistique dont l'activité est stable, indépendante des conditions politiques, et doit nécessairement être programmée dans la durée, il est hautement souhaitable de disposer de plus de visibilité et de prévisibilité quant aux moyens disponibles. A cet égard, l'Insee est un candidat évident à l'expérimentation de la programmation pluriannuelle que propose le Gouvernement pour certaines administrations.

En tout état de cause, il ne faut pas perdre de vue que, quelle que soit la configuration institutionnelle retenue, nous sommes financés sur fonds publics. Notre indépendance professionnelle et notre capacité à produire une statistique publique de qualité dépendent donc *in fine* du consentement de nos concitoyens à la financer. Elle dépend donc de la confiance qu'ils nous font et de la qualité perçue du service que nous rendons. C'est à cela que nous devons nous attacher avec obstination.

BIBLIOGRAPHIE

Seltzer, W., « *Politics and Statistics : Independence, Dependence or Interaction ?* », *United Nations, Departement of Economics and Social Information and Policy Analysis, Working Paper* série n° 6, New York, 1994.



DOSSIER

LE STATISTICIEN ET LES SOURCES ADMINISTRATIVES

- Utiliser les déclarations administratives à des fins statistiques *Pascal Rivière* 14
- La DSN : nouvelle référence pour les échanges de données sociales
des entreprises vers les administrations *Élisabeth Humbert-Bottin* 25
- Bien comprendre la déclaration sociale nominative pour mieux mesurer *Catherine Renne* 35
- De l'activité de la justice au suivi du justiciable :
faire parler les données de gestion *Christine Chambaz* 45



UTILISER LES DÉCLARATIONS ADMINISTRATIVES À DES FINS STATISTIQUES

Pascal Rivière*

Pour compléter les données d'enquête ou pour s'y substituer, l'utilisation des données administratives par les statisticiens se généralise. Pourtant, étrangement, le sujet fait l'objet de très peu d'investigations méthodologiques générales. Le cas particulier des données issues de déclarations administratives est intéressant parce que le processus de production associé présente des similitudes avec l'appareillage statistique. Avec un acteur particulier : l'organisme chargé de la gestion des flux de données, souvent différent des administrations utilisatrices. Celui-ci épargne du travail aux statisticiens et le facilite sur certains plans... même si ces derniers perdent le pouvoir sur une partie des opérations. In fine le fichier de données élaboré résulte d'une co-construction statistico-administrative. Pour optimiser celle-ci, il faut souligner l'importance d'une relation suivie et formalisée avec l'organisme concentrateur, ce qui donne au statisticien plus de possibilités d'intervention qu'il n'imagine.

RICHESSE DES SOURCES ADMINISTRATIVES, FAIBLESSE DE LA MÉTHODOLOGIE

Dans l'arsenal du statisticien, les sources administratives occupent une place croissante, et ce pour des raisons variées. Elles permettent de compléter les données d'enquête en les enrichissant, dès lors que l'appariement est possible : la mise en place future du « NIR haché » offrira de ce point de vue des perspectives tout à fait nouvelles. Il arrive aussi que ces données se substituent aux enquêtes, dans un contexte conjoint d'explosion des possibilités en matière de données (Elbaum, 2018), mais aussi plus classiquement de baisse généralisée des taux de réponse¹ et de restrictions budgétaires. Enfin elles peuvent constituer une source nouvelle, différente des domaines explorés habituellement par l'appareil statistique.

Pourtant, les statisticiens qui utilisent ces données savent qu'il y a loin de la coupe aux lèvres, entre autres parce que les processus de production statistique et administratif diffèrent. Or s'il existe un vaste corpus méthodologique dédié aux statistiques d'enquêtes² (voir par exemple Lyberg *et alii*, 2012), on ne trouve rien d'équivalent pour celles qui découlent de sources administratives³. Certes, il existe une littérature abondante sur la manière de traiter tel ou tel gisement de données particulier, les problèmes rencontrés

* Chef de l'Inspection générale, Insee.

1. « *Plunging response rates to household surveys worry policymakers* », *The Economist*, 24 mai 2018.

2. Traitant de tirage d'échantillon, de calage sur marge, de calcul de variance, de traitement de la non-réponse, etc.

3. Alain Desrosières écrivait à leur sujet en 2004 : « *aucune théorie des erreurs de portée générale n'existe encore* ».

et les solutions apportées⁴. Mais il est difficile de trouver un article transversal donnant un cadre méthodologique général et opérationnel – à l'exception peut-être de Rouppert (2005) – comme le fait la théorie des sondages par exemple. Dans la littérature anglo-saxonne, on peut citer essentiellement un article de Hand (2018) qui effectue un tour d'horizon des problématiques soulevées par les données administratives.

Cette absence d'armature conceptuelle conduit parfois à une forme de bricolage dans l'organisation de la production, à un déficit de méthode. Et en parallèle, on ne s'interroge pas vraiment sur ce que les données administratives ont de spécifique, en les abordant, sans autre forme de procès, comme un tout indifférencié immédiatement disponible, un fichier de données à traiter.

Le but du présent article est d'interroger les spécificités d'un ensemble particulier de données administratives : celles qui sont issues de déclarations administratives. On observe en effet des proximités bienvenues avec la production statistique (**encadré**) et dans le même temps des limitations qu'il est essentiel de connaître.

UN STATUT ET DES USAGES SPÉCIFIQUES

Lorsqu'on associe l'adjectif « administrative » à « donnée », que veut-on dire, au fond ? Le sujet serait vaste à explorer. On va se limiter ici à trois spécificités : statut, usage, mouvement.

Le terme « administrative » revient d'abord à qualifier les données par leur origine, et plus que cela par leur statut. Il signifie qu'elles viennent de la sphère administrative en général : les administrations, au sens de la comptabilité nationale, plus généralement les organismes de droit public, et ce qui relève des devoirs de l'État par ailleurs. Il en découle souvent un degré important de centralisation, tout du moins en France, ce qui facilite la tâche des statisticiens.

Le caractère administratif des données nous informe également sur leur usage⁵ : elles s'inscrivent dans les processus de gestion d'un ou plusieurs organismes (que nous appellerons « administrations » par la suite), qui ont leur propre agenda et leurs propres objectifs. Contrairement aux données statistiques dont le seul but est d'informer, leurs consœurs administratives doivent leur existence aux actions qu'elles permettent de déclencher. Elles ne sont pas neutres. Par exemple, la donnée individuelle « montant d'une retraite », pour un individu donné, permet de calculer le montant effectif à payer, puis de procéder au paiement, donc d'agir concrètement dans le monde réel.

L'usage peut aller au-delà de l'administration concernée. Ainsi, les informations du Répertoire National Commun de la Protection Sociale (RNCPS) peuvent permettre de repérer des incohérences entre prestations sociales de diverses natures ; la déclaration sociale nominative est utilisée par la DGFIP, les institutions de prévoyance, les organismes de retraite, l'Acos, pour leurs propres usages de gestion ; les données de SIRENE servent de référence, de preuve pour les entreprises, elles sont utilisées par les chambres de commerce et d'industrie ou par les greffes des tribunaux de commerce.

4. Cf. tous les articles de la session 20 (« Appariements – fichiers administratifs ») des Journées de méthodologie statistique 2018.

5. Desrosières (2004) : « Une source administrative est issue d'une institution dont la finalité n'est pas de produire une telle information, mais dont les activités de gestion impliquent la tenue, selon des règles générales, de fichiers ou de registres individuels, dont l'agrégation n'est qu'un sous-produit ».

Encadré. Le processus de production statistique standard d'une enquête

Utilisé comme référence par l'ensemble des Instituts nationaux de statistiques (INS), le *Generic Statistical Business Process Model* (GSBPM)* explicite un découpage standard en activités à partir de 8 étapes principales : définition des besoins, conception, élaboration, collecte, traitement, analyse, diffusion, évaluation. Le GSBPM évoque aussi le traitement des sources administratives, mais on présente ici plus spécifiquement le processus fondé sur une enquête, de façon extrêmement simplifiée, dans les grands principes.

Commençons par les bases. Afin de mener à bien une enquête, le statisticien public a *grosso modo* besoin de trois éléments pour commencer : un cadre légal, la liste des données qu'il veut recueillir, et la liste des entités auprès desquelles les obtenir.

- En France, la liste des enquêtes est fixée chaque année par arrêté du Ministre de l'économie et des finances, après avis d'opportunité du CNIS et avis de conformité du comité du Label, qui permettent d'officialiser l'enquête et de formaliser notamment l'obligation de réponse.
- La détermination des données d'intérêt nécessite d'en avoir explicité le sens, de s'être mis d'accord sur les besoins d'information, sur une manière pertinente de questionner, et sur la possibilité même de les obtenir. De proche en proche, et via des comités d'utilisateurs notamment, on arrive ainsi à un questionnaire, puis à un support de collecte (questionnaire papier, questionnaire électronique pour l'enquêteur ou sur le net...), et on associe à ce dernier des contrôles.
- Définir les entités à interroger, c'est d'abord expliciter le champ, ce qui revient à définir un ensemble en intention : par exemple, les entreprises du secteur du commerce ayant exercé une activité en 2018, en France métropolitaine ; ou bien : les ménages habitant en région frontalière. On construit ensuite une liste en extension, qui va être la base de sondage (par exemple, l'échantillon-maître, ou toute base de sondage issue de Sirene), puis on tire l'échantillon, qui correspondra à l'ensemble des entités à interroger concrètement.

Disposant des trois outils que sont le cadre juridique, le support de collecte et l'échantillon, le processus de collecte peut commencer. Il s'effectue pendant une période bien déterminée, le service statistique prenant contact avec les unités de collecte. Selon les enquêtes, les modes de collecte peuvent être différents (face-à-face, téléphone, internet) et éventuellement être combinés.

Les données collectées sont soumises à des contrôles automatiques : on recherche ainsi les incohérences, ce qui permet de repérer des données ou combinaisons de données douteuses, par exemple une personne qui déclarerait simultanément être retraitée et avoir 20 ans. Les vérifications individuelles réalisées par la suite sont en lien avec ces incohérences potentielles repérées automatiquement. Elles peuvent nécessiter une interaction avec les enquêtés : en direct lors de la collecte pour les enquêteurs, *a posteriori* par les gestionnaires, via les rappels téléphoniques ou les relances dans le cas des enquêtes auprès des entreprises. Notons que le processus de collecte statistique assume l'existence de non-réponses, soit des non-réponses totales, soit des non-réponses partielles.

Il s'ensuit une phase de traitement et d'analyse : gestion de la non-réponse (imputation, repondération), contrôles supplémentaires (macro contrôles, notamment), tabulation, calculs d'évolution, etc. Puis les publications et études, donc la phase de diffusion et enfin l'évaluation.

*Voir sur le site d'Eurostat : <https://ec.europa.eu/eurostat/>

LA CRISTALLISATION DES DONNÉES, CLÉ DE LA STATISTIQUE ADMINISTRATIVE

Le fait que les données administratives découlent d'un processus de gestion a aussi une conséquence opérationnelle : elles se trouvent dans des bases de données de gestion, toujours susceptibles d'être mises à jour. Ce sont en quelque sorte des données vivantes, qui peuvent être modifiées en fonction d'événements, internes ou externes. Ils peuvent se produire n'importe quand et certains sont totalement exogènes à l'administration, imprévisibles.

Ceci n'est pas sans conséquence pour le statisticien, qui a l'habitude, avec les enquêtes, de récupérer un jeu de données fixées dans le temps et qui se rapportent à des unités statistiques précisément définies. Certes, les valeurs de ces données peuvent changer en fonction des vérifications, redressements... mais fondamentalement on veut parler de la même donnée. Le caractère sans cesse mouvant des bases de données administratives ne peut donc lui convenir : il ne peut travailler que sur des données figées, qui vont se présenter sous forme de fichier. Il s'agira ainsi nécessairement d'une photo à un instant t, mais aussi d'une sélection des variables et des unités qui l'intéressent. Il faut passer d'un monde en mouvement à une cristallisation de ces données sur une période.

Attention, la question de la temporalité n'est pas un sujet simple. Ainsi il faut distinguer au minimum deux niveaux de dates : la date d'obtention de l'information d'une part, et la date de référence de celle-ci d'autre part. Par exemple, on récupère en mars 2018 (date d'obtention) l'effectif salarié d'une entreprise au 31 décembre 2017 (date de référence). Les choses se compliquent lorsqu'on constate que la date de référence peut devenir une période de référence (ex : l'année 2017), et que celle-ci peut nécessiter tout un calcul, toute une reconstitution des données : par exemple, l'effectif moyen en équivalent temps plein de l'entreprise sur l'année 2017.

LA NOTION DE DÉCLARATION ADMINISTRATIVE

Les données administratives peuvent provenir de plusieurs origines. Elles sont souvent issues de processus de gestion interne à l'administration concernée⁶. Mais elles peuvent aussi découler de déclarations administratives.

De quoi s'agit-il ? Une déclaration administrative, c'est l'obligation qui est faite à un certain nombre d'entités (individus, entreprises, organismes publics) de fournir des informations respectant une certaine forme, selon certaines modalités (internet, papier) et temporalités. Par exemple, les différentes déclarations d'impôts (sur le revenu, foncier...) relèvent de dispositifs très précisément documentés, et ils sont obligatoires, sur une périodicité et une plage de temps déterminées (l'impôt sur le revenu est annuel), pour les personnes ou entreprises assujetties à ces impôts. Dans le monde des entreprises, la déclaration sociale nominative est mensuelle (exceptées les déclarations événementielles), et les règles qu'elle doit respecter sont consignées dans un cahier technique versionné et accessible en ligne⁷, car la normalisation des déclarations est essentielle.

L'existence d'une forme d'obligation pour les déclarations administratives, souvent supportée par des textes législatifs et réglementaires et adossée à une documentation très normée, est la bienvenue pour le statisticien. Cette obligation se traduit en effet par un pouvoir coercitif, pouvant prendre diverses formes, comme celle d'engager des poursuites... ce qui réduit, sans non plus l'annihiler, le risque de non-déclaration.

6. Voir article de C. Chambaz sur les statistiques de la justice dans ce même numéro.

7. Voir article de C. Renne sur l'utilisation de la DSN dans ce même numéro.

Attention cependant, la déclaration administrative elle-même n'est pas, *stricto sensu*, une source administrative : ce n'est pas un fichier, c'est un flux. Et la façon de construire un fichier figé de données à partir de ce flux est d'ailleurs un sujet en soi. En particulier, élaborer ce fichier ne se limite pas à empiler les déclarations. Ainsi, pour la DSN, l'entreprise peut faire des déclarations rectificatives, des déclarations sur des périodes antérieures, mais aussi des déclarations événementielles, qui conduisent à modifier des données. De ce point de vue, le processus déclaratif diffère du processus d'enquête car rien n'empêche d'avoir pendant la période « de collecte » des modifications à l'initiative du déclarant, liées ici à la vie de l'entreprise. Un travail de reconstitution et même de consolidation, parfois complexe, est donc nécessaire pour bâtir, à partir de l'ensemble des déclarations, une source administrative. Dans certains cas, la situation est encore compliquée par l'utilisation de plusieurs types de déclarations administratives.

L'ORGANISME CONCENTRATEUR, «HUB» DES DÉCLARATIONS

Cette logique de gestion de flux conduit à effectuer une distinction fondamentale entre deux entités : d'une part l'administration, utilisant les données pour sa propre gestion, d'autre part l'organisme concentrateur. Le concentrateur est celui qui pilote le processus déclaratif dans son ensemble : il réalise le système d'information nécessaire, le documente, anime les instances de gouvernance et groupes de travail, organise le contact avec les déclarants, gère les nomenclatures, assure la communication et effectue le suivi du processus. Il concentre l'ensemble des flux, entrants en provenance des déclarants, sortants en direction des administrations utilisatrices. Mais il ne fait pas nécessairement partie de ces administrations, son rôle de gestionnaire de flux étant distinct.

C'est typiquement le rôle du Gip-MDS⁸ avec la plate-forme net-entreprises qui permet d'effectuer des déclarations sociales comme la Déclaration sociale nominative (DSN)⁹. Mais on peut aussi citer l'ATIH¹⁰ avec la plate-forme e-PMSI¹¹, conçue pour le recueil et l'analyse des informations sur l'hospitalisation, ou le SANDRE¹², qui garantit l'interopérabilité des systèmes d'information relatifs à l'eau (même si dans les deux cas le formalisme déclaratif est moins affirmé que pour la DSN). Gérer de telles infrastructures de recueil d'information de façon industrielle est un métier en soi qui requiert une grande technicité (informatique, maîtrise d'ouvrage, droit), une organisation rodée et réactive, et une gouvernance incluant organismes utilisateurs et contributeurs. Les administrations utilisatrices sont par exemple les organismes fiscaux ou sociaux dans le cas des déclarations sociales, la Caisse Nationale d'Assurance Maladie (Cnam) et les hôpitaux dans le cas des données PMSI, l'Agence Française pour la Biodiversité ou les Agences de l'Eau dans le cas du SANDRE.

Avec la déclaration, nous sommes ainsi dans une logique de flux normés, clairement positionnés dans le temps, liés à une obligation déclarative pour des entités déterminées. Si le processus déclaratif est rigoureusement géré par le concentrateur, on sait ce qui arrive aux données depuis la source primaire unique d'information. En particulier, les contrôles sont réalisés à un endroit et un seul et ils sont documentés.

8. Groupement d'Intérêt Public pour la Modernisation des Déclarations Sociales.

9. Voir article d'E. Humbert-Bottin sur la DSN dans ce même numéro.

10. Agence Technique de l'Information sur l'Hospitalisation (<https://www.atih.sante.fr/>)

11. <https://www.epmsi.atih.sante.fr/> Le PMSI, Programme de Médicalisation des Systèmes d'Information, rendu obligatoire en 1996, vise à définir l'activité des unités du service public hospitalier pour calculer leurs allocations budgétaires.

12. Service d'Administration Nationale des Données et Référentiels sur l'Eau.

DES PROXIMITÉS AVEC LA STATISTIQUE D'ENQUÊTE : LE CHAMP, LES VARIABLES...

L'organisme concentrateur gère un processus qui présente des points communs avec le processus statistique (**encadré**), en commençant par le triptyque champ / variables / cadre légal.

Le champ est en principe formellement défini et connu : c'est l'ensemble des personnes, ou entités, assujetties à la déclaration. Il fait l'objet d'une définition en intention, avec un cadre général, mais aussi des exemptions, et des cas particuliers ; par exemple, les entreprises et cabinets « ayant des difficultés à assurer les paramétrages sur la partie des organismes complémentaires » ont été exemptés de DSN en phase 3.

La liste de données à recueillir peut faire l'objet d'âpres discussions entre les parties prenantes, car il s'agit de prendre en compte les besoins de chacun sans multiplier les données à collecter (c'était tout le problème avec les DADS). Ceci, en se fondant sur le droit et en recherchant la mutualisation via un modèle conceptuel de données commun, différence décisive entre DADS et DSN. Les données à recueillir sont définies et documentées, de même que la manière dont les échanges sont normés : on pense au cahier technique de la DSN¹³, mais aussi à la documentation très riche du Sandre (dictionnaire de données, modèle de données, scénarios d'échange, règles organisationnelles et techniques des données de référence). Par ailleurs, l'organisme concentrateur élabore, maintient et diffuse les nomenclatures de référence, ce qui est un travail considérable : ainsi, pour l'ATIH, la classification internationale des maladies (CIM), ou la classification commune des actes médicaux (CCAM).

“ La liste de données à recueillir peut faire l'objet d'âpres discussions entre les parties prenantes, car il s'agit de prendre en compte les besoins de chacun sans multiplier les données à collecter [...]. ”

... ET LE FONCTIONNEMENT DE LA COLLECTE

Reste à collecter les données. Notons d'abord que le support de la déclaration, à savoir un formulaire - type Cerfa - ou quelque chose d'équivalent, ressemble parfois furieusement à un questionnaire, proposé en ligne. La collecte est pour l'essentiel exhaustive sur le champ considéré. Sous cette réserve, l'exhaustivité est obtenue via un suivi rigoureux des non-déclarants, beaucoup plus puissant que le suivi des non-répondants : le pouvoir coercitif résultant de l'obligation de déclaration est sans commune mesure avec celui, limité, des instituts statistiques sur les non-répondants ; chacun connaît les conséquences d'une déclaration d'impôts non transmise ou en retard. On constate pourtant souvent une étrange sous-utilisation (et même non-utilisation) par le statisticien de l'information très riche issue de ce suivi des non-déclarants. C'est le cas par exemple pour la DSN.

À l'instar des questionnaires statistiques, les formulaires administratifs font l'objet de contrôles automatiques. Leur logique est légèrement différente des contrôles statistiques. Ainsi, lorsqu'on déclare une DSN et qu'un contrôle n'est pas satisfait par la déclaration, celle-ci ne « passe » pas, elle est bloquée, c'est comme si elle n'avait pas été envoyée ; le déclarant doit donc en envoyer une nouvelle, jusqu'à ce que sa déclaration « passe ». On trouve des contrôles d'appartenance à une liste prédéfinie (ex : la liste officielle des codes pays, celle des codes postaux), de cohérence formelle entre données (ex : égalité entre l'année de naissance et celle qui figure dans le NIR), de la nature d'une donnée (numérique, alphanumérique, date...), voire de sa structure fine (cf. les règles de forme complexes

13. Disponible sur le site dsn-info : www.dsn-info.fr

des adresses mail). Mais ces vérifications automatiques se limitent à des contrôles « durs », qui doivent absolument être vérifiés : on veille donc à ne pas multiplier les contrôles, car cela bloquerait le processus déclaratif.

❶ EN CONTREPARTIE, UNE PERTE DE MAÎTRISE POUR LE STATISTICIEN

La manière dont la collecte d'informations se prépare et s'organise, si elle présente des similitudes importantes avec le processus statistique (en gros les quatre premières étapes du GSBPM), a cependant un inconvénient majeur : toutes les étapes ainsi décrites sont sous la responsabilité du concentrateur, et échappent donc au service statistique. Champ, données, formulaire, tout cela peut évoluer dans le temps, sans que le statisticien ait le moindre contrôle sur ces aspects. Par exemple si une catégorie de population n'est plus assujettie à tel ou tel impôt et disparaît ainsi du champ, le statisticien n'a aucun pouvoir sur cette décision. Il n'a rien à dire non plus sur la liste des variables collectées ni sur leur définition même : celle-ci fait l'objet de multiples discussions avec les organismes utilisateurs de ces données¹⁴, dans lesquels les services statistiques ne sont pas ceux qui pèsent le plus. Il perd donc la maîtrise des concepts, du cadre de la collecte d'informations.

Il perd aussi la responsabilité de la collecte elle-même, au profit d'un tiers, l'organisme gestionnaire de flux. Les services de production statistique voient au passage disparaître un aspect essentiel de leur travail : le contact direct avec les unités de collecte, et la possibilité de vérifier l'information immédiatement auprès de l'enquêté. Cette interaction avec

“ Plus généralement, le statisticien ne décide pas des modalités d'obtention de ces données : l'organisation du processus déclaratif, les relations avec les déclarants [...], le système d'information dans son ensemble. Enfin et surtout, toute cette organisation peut aussi évoluer dans le temps, en lien avec de nouvelles dispositions législatives, sans qu'il soit lui-même consulté. ”

le terrain, cet ancrage au réel, s'évanouit. Au caractère vivant et dynamique de la collecte se substitue un fichier, projection statique et sans âme d'un processus externalisé.

Plus généralement, le statisticien ne décide pas des modalités d'obtention de ces données : l'organisation du processus déclaratif, les relations avec les déclarants (FAQ, communication), le système d'information dans son ensemble. Enfin et surtout, toute cette organisation peut aussi évoluer dans le temps, en lien avec de nouvelles dispositions législatives, sans qu'il soit lui-même consulté.

❷ DANS LES FAITS, UNE CO-CONSTRUCTION STATISTICO-ADMINISTRATIVE

Les déclarations administratives ne fournissent pas une information « clés en main » aux services statistiques. Il reste au statisticien tout un travail à faire en aval, car les données sont souvent inutilisables telles quelles. Pour aboutir à un fichier de données individuelles propre et adapté à leurs usages, les statisticiens doivent ainsi enclencher un second processus après le processus purement administratif. Le résultat est donc issu d'une co-construction entre les deux univers.

14. Cf. les nombreuses structures de gouvernance associées à la DSN, et auparavant aux DADS, faisant intervenir tous types d'organismes de protection sociale, services fiscaux... et statistiques.

Les statisticiens doivent ainsi réaliser plusieurs opérations :

- Des contrôles supplémentaires, qui n'ont pas été effectués côté administratif : on pense aux contrôles de vraisemblance, qui vérifient la plausibilité d'une combinaison de données sur des sujets sensibles sur le plan statistique. Par exemple, on va vérifier la cohérence entre le libellé de profession déclarée et le code PCS, ou vérifier la cohérence de montants de revenus entre eux.
- Des transformations des données : les données de la déclaration administrative sont en quelque sorte des données pivot dont chaque administration fait son miel. Cela vaut aussi pour les services statistiques : pour obtenir les vraies données utiles aux études, il faut parfois calculer de nouvelles données à partir de la déclaration. Ces « transformations » peuvent prendre des formes très simples (ex : calcul de l'âge à partir de la date de naissance, ou de la tranche d'âge à partir de l'âge). Elles peuvent résulter d'un calcul plus complexe (ex : la détermination des périodes d'activité à partir des dates de début et de fin se trouvant dans des déclarations séparées), ou nécessiter des transformations de nomenclatures, dont le niveau de finesse est parfois inutile pour ses usages (ex : on n'a pas toujours besoin de tout le détail de la NAF). Simples ou non, ces transformations doivent être définies rigoureusement pour être appliquées de façon systématique.
- Le passage à de nouvelles unités statistiques, ce qui relève d'un degré de complexité supérieur : par exemple la DSN raisonne sur des unités telles que le salarié, le contrat, alors que le besoin des statisticiens porte plutôt sur le poste de travail. Il faut donc reconstituer les données correspondantes, ce qui demande beaucoup de rigueur.

Ce faisant on effectue une véritable mise en forme et mise aux normes de qualité statistique des données, qui vient se substituer à la mise en forme et aux normes administratives d'origine.

Au fond, l'organisme concentrateur réalise toute une partie du processus de production statistique à la place du service statistique (collecte, interaction avec les déclarants, suivi des non-déclarants, une part importante des contrôles) et ce dernier complète par d'autres contrôles, des redressements, des transformations et une mise aux normes statistiques. Avec un bémol important, la perte de maîtrise : on n'impose rien à l'organisme en question.

LA QUALITÉ DES DONNÉES, UNE QUESTION-CLÉ

Utiliser les données administratives à des fins statistiques requiert que celles-ci soient « de qualité »... L'article de David J. Hand sur le sujet mentionne à plusieurs reprises des « problèmes de qualité » sur les données administratives. Mais de quoi parle-t-on ? Les questions de non-qualité des données en général font l'objet de toute une littérature (par exemple McCallum 2012), mais pour un usage statistique elles se jouent à plusieurs niveaux : conformité de la population au champ souhaité, validité de la sémantique des données (qui peut diverger de ce que souhaite le statisticien), existence de contrôles suffisants (certaines données jugées peu importantes ne sont même pas vérifiées), problèmes de dates, d'unités statistiques...

Dans le cas des données issues de déclarations administratives, le fait que des contrôles aient déjà été effectués dans le processus déclaratif, et qu'ils soient précisément documentés et versionnés, constitue déjà un premier pas, une garantie de premier niveau. Chaque contrôle correspond au fond à une propriété que les données déclarées doivent respecter : appartenir à telle liste, être de nature numérique, être présent si telle donnée est présente... Mais il s'agit là de garanties de forme, de structure, qui n'assurent pas une qualité suffisante.

Car au-delà de la syntaxe, il faut se mettre d'accord sur une sémantique commune, clairement définie. Dans les exemples cités plus haut, comme la DSN ou le système d'information

de l'eau, il existe ainsi, bien avant toute considération sur la normalisation des échanges, un modèle conceptuel de données commun appliquant les standards du langage de modélisation unifié (UML). Il conduit à définir rigoureusement la signification des données, les concepts et liens entre concepts. Sur ces questions, il faut souligner que la Belgique a joué un rôle précurseur, avec la mise en place dès 1990 de la « Banque Carrefour de la Sécurité Sociale », institution publique de sécurité sociale chargée de l'échange de données entre les institutions de sécurité sociale... et exemple typique d'organisme concentrateur. Avec cette Banque Carrefour, véritable source d'inspiration pour la DSN, un modèle conceptuel de données a peu à peu été élaboré, mis en commun (Robben *et alii*, 2006) puis imposé jusque dans la loi qui lui a conféré un caractère opposable. Grâce à cela, tous les flux de données provenant des entreprises se fondent sur une même sémantique, sur les mêmes nomenclatures versionnées régulièrement, le tout étant intégralement documenté et mis en ligne. Ce système, qui était donc très en avance sur son temps, est aujourd'hui toujours opérant et efficace.

HYPOTHÈSE DU MONDE CLOS ET BACKTRACKING

La sémantique que nous évoquons ici caractérise un monde qui n'est pas immuable. Ainsi des combinaisons de données qui peuvent être jugées en anomalie à un instant *t* ne le seront peut-être plus à un instant ultérieur, parce que le monde a changé (exemple : les couples mariés de même sexe, situation impossible dans une certaine période, possible après). Supposer que l'on peut définir la qualité des données par une liste de propriétés formelles, que l'on peut vérifier via des contrôles (ex : vérifier que dans un couple les deux personnes sont de sexe opposé), revient à implicitement à faire « l'hypothèse du monde clos ». Or, comme on vient de le voir, celui-ci évolue inévitablement.

Comment donner un cadre opérationnel à ces évolutions et en tirer une approche opérationnelle et efficace ? La réponse nous vient à nouveau d'Outre-Quévrain : dans l'ouvrage de référence tiré de sa thèse, *Boydens (2000)* fournit un éclairage tout à fait original sur le sujet avec la notion de « temporalités étagées » issues de l'œuvre de l'historien Fernand Braudel. Les données sont soumises à trois temporalités : la temporalité juridique, la temporalité des bases de données, et la temporalité du réel. Les trois ne sont pas nécessairement ajustées, synchrones, loin de là : le juridique peut s'adapter avec retard au réel, les structures de bases de données n'ont pas la souplesse nécessaire pour s'adapter automatiquement aux évolutions.

Pour rapprocher ces temporalités, à défaut de les synchroniser, la méthode du *backtracking* (Redman 1996, Boydens 2000, 2018) offre une démarche opérationnelle efficace, que l'auteur a appliquée... à des données administratives, en l'occurrence des données de sécurité sociale.

Cette méthode consiste à identifier, à partir des violations de « business rules » les plus fréquemment observées à la source au sein de la base de données, les causes structurelles et y remédier en remontant les flux d'information vers l'amont, ce qui permet une action beaucoup plus rapide et beaucoup plus fiable puisque la vérification se fait à la source. Quel est le principe ? Les flux de données, qu'ils soient administratifs ou statistiques, font naturellement l'objet de contrôles. Par exemple, si telle prestation ne peut être accordée au-delà d'un certain plafond de revenus, on va contrôler la cohérence entre les données « existence de la prestation » et « revenu ». Et si l'on observe une incohérence entre les deux sur des données, ceci génère une anomalie. Jusque-là, rien de bien original.

L'élément novateur (entre autres) est de considérer les anomalies comme des objets d'intérêt en soi, et de suivre leur évolution dans le temps par type d'anomalie. Et dès que l'un de ces types d'anomalie se met à devenir fréquent, ceci nous informe, potentiellement,

sur l'évolution du réel sous-jacent (par exemple, dans le cas cité, sur l'existence de dérogations significatives au plafond de revenus). L'étape suivante est donc de revenir à la source pour mieux comprendre cela et agir auprès des fournisseurs de données, ou bien faire évoluer les contrôles.

Cette technique, utilisée en Belgique avec les administrations, permet réellement d'améliorer la qualité des données, via ces allers-retours pertinents avec les fournisseurs, avec « un ROI atteignant plus de 50 % en termes de diminution de la part de présomptions d'anomalies à traiter » (Boydens, 2018). Le processus est même allé beaucoup plus loin puisque la démarche de *backtracking* a été officialisée via un arrêté royal¹⁵.

UNE AUTRE FAÇON DE PENSER LE TRAVAIL DU STATISTICIEN

En conclusion, les déclarations administratives offrent de nombreux avantages pour les statisticiens, notamment en raison d'un processus de production ayant des similarités, mais elles présentent, comme les données administratives en général, un gros inconvénient : la perte de maîtrise par l'appareil statistique, qui n'est plus décisionnaire sur des aspects majeurs (champ, données collectées).

Pour remédier à ce problème ou tout du moins le contenir, il est crucial pour le statisticien de travailler de façon régulière et structurée avec l'organisme concentrateur des flux, bien au-delà de la fourniture de données : dans la conception des données et des contrôles, la mise à jour d'une documentation de référence partagée, ou la gestion des allers-retours avec les fournisseurs de données, dans l'esprit du *backtracking*. C'est là un changement culturel, mais il permet de limiter les incompréhensions entre monde statistique et univers de gestion. Une telle coopération peut enclencher un cycle vertueux favorable à une amélioration suivie et maîtrisée de la qualité des données... et des statistiques qui en sont issues.

L'utilisation de données administratives pour la statistique, vue ici à travers le cas particulier des déclarations donc de la gestion industrialisée de flux, ouvre ainsi des horizons nouveaux pour la statistique, d'autres modalités de travail, de nouveaux métiers, voire d'autres façons de penser notre activité, qui obligent à rompre avec les schémas habituels.

15. Arrêté royal du 2 février 2017 modifiant le chapitre IV de l'arrêté royal du 28 novembre 1969 pris en exécution de la loi du 27 juin 1969 révisant l'arrêté-loi du 28 décembre 1944 concernant la sécurité sociale des travailleurs.

Boydens I., « *Informatique, normes et temps* », Editions Bruylant, 2000.

Boydens I., « *Data Quality & « back tracking » : depuis les premières expérimentations à la parution d'un Arrêté Royal* », *Smals Research*, mai 2018.

Desrosières A., « *Enquêtes versus registres administratifs : réflexions sur la dualité des sources statistiques* », *Courrier des statistiques*, n° 111, septembre 2004.

Elbaum M., « *Les enjeux des nouvelles sources de données* », *Chroniques*, n° 16, CNIS, septembre 2018.

Hand D.J., « *Statistical challenges of administrative and transaction data* », *J. R. Statist. Soc. A* 181, Part 3, p. 555–605, 2018.

Journées de méthodologie statistique 2018, session 20 (« *Appariements – fichiers administratifs* ») <http://jms-insee.fr/programmejms2018/>.

Lyberg L., Biemer P., Collins M., De Leeuw E., Dippo C., Schwarz N., et Trewin D., « *Survey Measurement and Process Quality* », Wiley, 2012.

McCallum E. Q., « *Bad data handbook* », O'Reilly Media, 2012.

Redman T., « *Data quality for the information age* », Artech house, 1996.

Robben F., Desterbecq T. et Maes P., « *L'expérience de la Banque-carrefour de la Sécurité sociale en Belgique* », revue des politiques sociales et familiales, n° 86 sur le thème de « *La nouvelle administration. L'information numérique au service du citoyen* », pp. 19-31, 2006.

Rouppert B., « *Modélisation du processus de traitement d'une source administrative à des fins statistiques* », document de travail SGI, Insee, 2005

LA DÉCLARATION SOCIALE NOMINATIVE

NOUVELLE RÉFÉRENCE POUR LES ÉCHANGES DE DONNÉES SOCIALES DES ENTREPRISES VERS LES ADMINISTRATIONS

Élisabeth Humbert-Bottin*

La généralisation de la déclaration sociale nominative en 2017 est l'aboutissement d'un projet de simplification administrative qui s'est étalé sur près de dix ans. Elle repose sur un changement de paradigme qui a nécessité une simplification drastique des procédures déclaratives auxquelles sont soumises les entreprises. Mais en adossant la collecte des données au processus générateur de la collecte des cotisations sociales, c'est-à-dire le processus de paie, elle garantit une bien meilleure qualité de l'information recueillie. Au terme d'un déploiement progressif, la DSN concerne aujourd'hui environ 96 % des entreprises du secteur privé. Bien accompagnées grâce à un dispositif de concertation inédit, les trois quarts des entreprises sont globalement très satisfaites du nouveau dispositif. Le taux de couverture d'un certain nombre de procédures et la qualité des données progressent également rapidement, même si des points de vigilance demeurent, notamment en raison des délais d'adaptation des logiciels de paie. La couverture de la DSN est appelée à s'étendre dans les années à venir, ainsi que ses usages. Les porteurs du projet devront veiller à ce que ces extensions ne se traduisent pas par un alourdissement qui serait contraire à l'esprit initial du projet.

UNE SIMPLIFICATION DE LONGUE HALEINE

Indispensables pour assurer le financement de la protection sociale et permettre aux salariés d'exercer leurs droits (maladie, chômage, retraite...), les déclarations sociales font partie des tâches administratives imposées aux entreprises françaises. La déclaration sociale nominative (DSN), née à la fin des années 2000, a été instituée par la loi de simplification du 22 mars 2012, dite loi Warsman. Elle est obligatoire pour toutes les entreprises depuis début 2017.

Portée dès l'origine par les organismes de protection sociale (OPS), elle a été conçue pour satisfaire le plus grand nombre possible d'obligations déclaratives existant dans le champ social au sens large, tout en réduisant le nombre de demandes d'informations distinctes auxquelles sont confrontées les entreprises. Par son ampleur et son ambition, ce projet est unique. Sa mise en œuvre a impliqué volonté politique, conviction des dirigeants de l'entité chargée de le développer, le Groupement d'Intérêt Public de Modernisation

*Directrice, Groupement d'intérêt public pour la modernisation des déclarations sociales.

des Déclarations Sociales (Gip-MDS) et talent des équipes qui œuvrent chaque jour à sa réalisation, à tous les niveaux, depuis celles en charge du système jusqu'à celles en charge des systèmes d'information des organismes de protection sociale et des administrations concernées. Des années de préparation et d'instruction technique¹ ont été nécessaires (Vatinet, 2013) puis une concertation étroite avec tous les acteurs au cours de la phase de déploiement (**encadré 1**).

Encadré 1. Quelques repères concernant la gouvernance de la DSN

Une mission interministérielle rattachée au ministère du budget et au ministère des affaires sociales assure la maîtrise d'ouvrage stratégique du projet DSN depuis 2011. La maîtrise d'ouvrage opérationnelle de la DSN a été confiée au Gip-MDS. Créé en 2000 pour répondre au besoin de simplification des démarches administratives des entreprises, le Gip rassemble l'ensemble des acteurs de la protection sociale.

Le succès du projet de la DSN à ce jour tient à quatre éléments majeurs :

- La fixation et la tenue d'un périmètre clair, appuyé sur un modèle de données commun largement partagé et instancié dans un outillage automatisé qui en sécurise le fonctionnement,
- Une maîtrise d'ouvrage opérationnelle installée dans un collectif au sein de la protection sociale travaillant en lien étroit avec une maîtrise d'ouvrage stratégique, marquant les décisions de haut niveau nécessaires de la part de l'État,
- La structuration d'une base de connaissance partagée, creuset commun de l'apprentissage des nouvelles modalités et garant du patrimoine collectif sur les règles de fonctionnement,
- La constitution de très nombreuses structures de travail et de partage avec les acteurs du dispositif au sens large, les entreprises, les experts comptables, les éditeurs, les organismes et le système DSN avec l'entretien permanent du travail en collectif et de la mobilisation de chacun au profit de la simplification.

En particulier, la réussite du projet DSN et de ses évolutions repose sur une large concertation des acteurs (HCFIPS, 2017). Un comité des utilisateurs réunissant les représentants des entreprises, des experts comptables et des éditeurs vise à relayer notamment les décisions majeures structurant le projet en lien avec les fédérations professionnelles qui le souhaitent. Un groupe permanent d'expression des besoins des entreprises comprend des experts des ressources humaines et des experts comptables pour relayer de manière très concrète les attentes des entreprises. Un groupe d'échange avec les organisations syndicales a été mis en place pour faire le point sur toutes les questions en lien avec la DSN impactant potentiellement les droits ou les procédures des salariés. On peut citer également le club des pilotes (chargés de recenser notamment les difficultés rencontrées et la manière dont les entreprises devaient se préparer), les réunions mensuelles avec les éditeurs de paie (complétées par des ateliers qualité) et les réunions avec les entreprises de travail temporaire. Un important dispositif d'assistance téléphonique et une base de connaissance en ligne ont également été déployés pour prendre en charge les questions des entreprises.

1. On peut citer notamment l'étude de faisabilité conduite par le Gip-MDS en 2006, puis celle réalisée par l'IGAS et l'IGF en 2011.

① UN CHANGEMENT DE PARADIGME, UNE SIMPLIFICATION DRASTIQUE

La DSN instaure un fonctionnement en rupture avec celui qui prévalait auparavant. Les déclarations sociales reposaient sur des formulaires Cerfa dont le contenu était fixé par les textes fondant la collecte des données utiles aux organismes de protection sociale et à l'administration pour l'exercice de leurs missions. Non seulement les déclarants étaient amenés à fournir plusieurs fois la même information, mais ils devaient surtout fournir une information qui n'était pas naturellement produite par leur système de gestion, ce qui était source d'incohérences et d'erreurs dans les déclarations. La DSN met en œuvre une logique fondamentalement différente : elle s'approche au plus près du fait générateur des rémunérations et cotisations sociales dans le domaine de la protection sociale, la paie. Elle repose sur un modèle unique de cette dernière et un échange de données primaires de gestion entre l'émetteur, qui fait la paie, et tous les organismes et administrations qui ont besoin de ces données sociales pour recouvrer des cotisations et servir des droits. Elle opère donc un déplacement de la charge de traitement des données de l'amont (l'entreprise déclarante) vers l'aval (les OPS et administrations utilisatrices et leur opérateur, le Gip-MDS).

Un important travail de simplification a été réalisé en amont du déploiement de la DSN. Il a visé notamment : l'harmonisation des dates de déclaration et de paiement des cotisations ; la simplification des règles de décompte des effectifs pour les seuils sociaux avec l'adoption d'une définition unique ; la simplification des règles de calcul du plafond ; la clarification des règles d'assiette et du fait générateur ; des simplifications importantes sur les règles parfois très anciennes relatives aux indemnités journalières.

In fine, la collecte a été expurgée de toutes les variables calculées, les variables de contrôle et les variables à finalité purement statistique, ce qui a nécessité de longues discussions pendant la phase projet avec les destinataires finaux de l'information², sur la base d'un arbre de décision logique obligeant à situer la réelle valeur opérante de cette information. Pour donner un exemple, alors que la Déclaration Annuelle de Données Sociales (DADS) véhicule environ 800 rubriques déclaratives distinctes, la Cour des comptes (2016) mentionne le fait que la DSN, en phase de généralisation, en comprend 400, correspondant à environ 230 données³.

② UNE BONNE DSN EST AVANT TOUT UNE PAIE DE QUALITÉ

La déclaration se fait au niveau de chaque SIRET et repose sur un principe clé : chaque salarié de l'établissement déclaré doit apparaître dans la DSN.

La DSN se présente comme :

- une déclaration mensuelle unique⁴, issue des données de paie de l'entreprise, permettant d'assurer le recouvrement des cotisations, d'attribuer des droits et de livrer des informations à des fins statistiques ; elle est le reflet de la paie du mois M-1, à transmettre le 5 ou le 15 du mois M ;

2. Ainsi, la nécessité de collecter chaque donnée a fait l'objet d'un examen systématique, tant au regard des textes réglementaires que des processus métier.

3. Une rubrique est le niveau le plus élémentaire d'informations dans une déclaration. Par contre, une donnée peut apparaître dans plusieurs rubriques. Une entreprise n'est pas amenée à remplir l'ensemble de ces 400 rubriques.

4. Elle comprend : lieu d'activité et caractéristiques du contrat de travail, montants des rémunérations, des cotisations et contributions sociales, durée de travail retenue ou établie pour la paie de chaque mois, dates de début et de fin de contrat, de suspension et de reprise du contrat intervenant au cours de ce mois.

- une déclaration événementielle fournissant des informations ponctuelles sur la situation d'un ou plusieurs salariés pour des événements⁵ les concernant ; en cas d'événement particulier, les entreprises n'ont plus à transmettre les rémunérations des derniers mois afin que certaines prestations puissent être servies aux salariés concernés : elles n'ont plus qu'à signaler le salarié concerné et la nature de l'événement.

Le caractère innovant de la DSN par rapport aux précédentes démarches de simplification tient à ce qu'elle est fondée sur un flux de données émanant directement du logiciel de paie des entreprises. Ce flux entièrement dématérialisé est transmis par l'employeur via un site

Figure 1. Avant et après la DSN

Avant la DSN

Deux procédures distinctes manipulant des données communes mais fondées sur des codifications et des échéances différentes :

1 logique paie et 1 logique déclarative



Après la DSN

Un même traitement de la DSN : la paie et les déclarations sociales qu'elle sert.

En conséquence la paie devient la source directe de la DSN mensuelle



5. Arrêt de travail, fin du contrat de travail, radiation.

unique (net-entreprises). C'est le système de la DSN qui prend en charge la répartition des données entre les divers organismes, qui pesait auparavant sur les entreprises (**figure 1**). Les déclarations se font à travers deux portails de déclaration : le portail **net-entreprise.fr** pour l'ensemble des entreprises et le portail **msa.fr** uniquement destiné aux employeurs du régime agricole⁶.

Auparavant, les entreprises transmettaient des données multiples à diverses échéances et à différents organismes, globalisées par établissement. Avec la DSN, les employeurs entrés dans la démarche effectuent, à l'issue de la paie, une transmission mensuelle⁷ de données individuelles des salariés. Il faut bien comprendre ce principe majeur qui l'instaure pour l'utiliser de manière la plus optimale possible. En particulier, sauf erreur de masse dont le gestionnaire de l'entreprise déclarante se rend compte avant que l'échéance ne soit dépassée (déclaration en « annule et remplace » acceptée seulement jusqu'à cette échéance), une DSN ne se rectifie pas ; elle se corrige d'un mois sur l'autre, la paie étant elle-même à corriger le mois suivant.

L'IMPORTANCE DE LA NORMALISATION DES FLUX D'ÉCHANGE DE DONNÉES

La convention de projet DSN réunit de très nombreux organismes parmi lesquels l'Acos, les retraites complémentaires AGIRC-ARRCO, la Caisse nationale d'assurance maladie (Cnam), Pôle emploi, le Centre technique des institutions de prévoyance (CTIP), la Fédération française de l'assurance (FFA), la Fédération nationale de la mutualité française (FNMF), la Caisse nationale d'allocations familiales (Cnaf), la Caisse nationale d'assurance

vieillesse (Cnav), la Caisse centrale de la mutualité sociale agricole (CCMSA), la Caisse de retraite et de prévoyance des clercs et employés de notaires (CRPCEN), la Dares, l'Insee, la DGFiP, la Caisse des dépôts et consignations, la Caisse nationale des industries électriques et gazières, l'Union des caisses de France du réseau congés intempéries BTP (UCF CI-BTP), etc.

Tous ces systèmes d'informations reçoivent une liste spécifique de données, fixée par arrêté selon leurs missions et doivent se synchroniser au fur et à mesure de l'élargissement du périmètre, ce qui nécessite un réglage « fin » permanent ! Les destinataires des données DSN sont nombreux, puisqu'à ce jour elles remplacent 26 formalités. D'autres sont encore à venir.

La rationalisation des flux d'échanges d'information opérée par la DSN n'aurait pas été possible sans la mise en place d'un cadre rigoureux. La DSN est fondée sur une norme qui représente ce qu'est la paie. Cette norme d'échanges⁸, nommée NEODES, est portée dans un « cahier technique », document de référence qui décrit toutes les données, tous

« Tous ces systèmes d'informations reçoivent une liste spécifique de données, [...] et doivent se synchroniser au fur et à mesure de l'élargissement du périmètre, ce qui nécessite un réglage « fin » permanent ! Les destinataires des données DSN sont nombreux, puisqu'à ce jour elle remplace 26 formalités. D'autres sont encore à venir. »

6. Avant la DSN, la MSA constituait déjà un « guichet unique » pour le monde agricole. Changer de guichet unique pour cette profession au motif qu'on en instaurait un pour tous se serait éloigné de l'esprit même de la réforme qui était avant tout la simplification des démarches.

7. Décret n° 2013-266 du 28 mars 2013 relatif à la déclaration sociale nominative, JORF n° 0076 du 30 mars 2013 page 5421 texte n° 35.

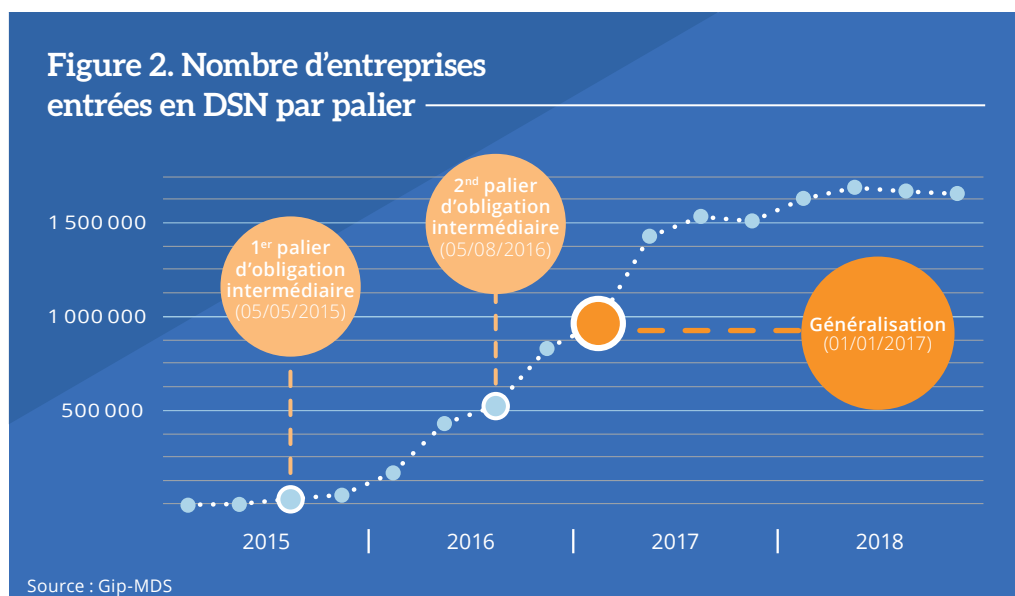
8. Dans le cas de la DSN, il s'agit de la norme NEODES : Norme d'Échange Optimisée des Données Sociales.

les contrôles et les usages mais aussi le modèle d'information, le modèle de message et la cinématique des échanges.

Ce cahier technique permet de générer automatiquement, grâce à un générateur nommé SATURNE, la documentation de la DSN, mais aussi une brique de contrôle (voir l'article de C. Renne dans ce numéro pour le détail des contrôles effectués) et un outil de validation (DSN Val). Cela garantit donc leur cohérence, ce qui n'était pas le cas dans l'ancien système. Cela est essentiel dans le cadre d'un système mensuel, qui évolue de façon régulière.

UNE MONTÉE EN CHARGE PRESQUE ACHEVÉE

Le déploiement progressif de la DSN depuis 2013 (**encadré 2**) a abouti à sa généralisation en janvier 2017. Fin 2017, 99,8 %⁹ de la cible déterminée sur les entreprises du secteur privé¹⁰ étaient entrées en DSN (**figure 2**).



Aujourd'hui, la DSN remplace et simplifie la majorité des déclarations sociales en automatisant mensuellement leur transmission à partir d'un flux de données issues de la paie. Elle remplace en particulier :

- les déclarations administratives de données sociales (DADS) ;
- les formalités de recouvrement pour les Urssaf, la retraite complémentaire, les organismes complémentaires santé / prévoyance, le régime agricole et certains régimes particuliers ou spéciaux ;
- les attestations de salaire pour le paiement des indemnités journalières (IJ) en cas de maladie, maternité, paternité, accident du travail et de maladie professionnelle ;
- les déclarations de mouvements de main-d'œuvre (DMMO / EMMO)
- les attestations employeur destinées à Pôle emploi ;

9. Compte tenu des créations / disparitions entreprises, l'Acoess estime l'atteinte de la cible « courante » à 96 %.

10. Sur le champ, un peu différent de celui retenu par l'Acoess, des entreprises du secteur privé, à l'exclusion des entreprises ayant déclaré des effectifs non nuls en fin de mois.

- les formulaires de radiation pour les organismes complémentaires de prévoyance ;
- le relevé mensuel de missions (pour les entreprises de travail intérimaire).

Encadré 2. Les étapes de la mise en place de la DSN

La DSN s'est mise en place progressivement à compter de 2013 selon trois phases de déploiement.

a) La phase 1 (2013-2014) était fondée essentiellement sur le volontariat. La DSN ne couvrait alors que six obligations déclaratives : la déclaration mensuelle des mouvements de main-d'œuvre (DMMO) ; l'enquête statistique sur les mouvements de main-d'œuvre ; l'attestation de salaire pour les indemnités journalières (IJ) de maladie et de maternité et les allocations de paternité et d'adoption (à l'exclusion de l'attestation de salaire AT-MP) ; l'attestation employeur destinée à Pôle emploi ; la formalité de radiation auprès des organismes de prévoyance collective obligatoire et, le cas échéant, celle portant sur l'option relative à la portabilité des droits.

b) La phase 2 (2015-2016) a étendu le champ de la DSN :

- aux déclarations liées au recouvrement de cotisations et contributions sociales par les Urssaf et caisses générales de sécurité sociale : déclaration unifiée de cotisations sociales (Ducs), bordereau récapitulatif de cotisations (BRC) et tableau récapitulatif annuel (TR),
- à la déclaration de salaire pour les indemnités journalières en cas d'accident du travail ou maladie professionnelle,
- au relevé mensuel de mission (RMM) pour les entreprises de travail temporaire.

Durant cette phase, la DSN est également devenue obligatoire depuis mai 2015 pour les entreprises les plus importantes.

c) La phase 3 du projet DSN a commencé en 2017. Il s'est agi :

- de rendre obligatoire l'utilisation de la DSN pour toutes les entreprises du régime général et du régime agricole ainsi que pour certains régimes spéciaux ;
- d'accroître la liste des usages et des destinataires des données issues de la DSN et d'en tirer les conséquences sur certaines déclarations existantes comme la déclaration annuelle de données sociales (DADS), qui a disparu fin 2017 pour les entreprises ayant satisfait les obligations de la phase 3 depuis janvier 2017.

La DSN est désormais généralisée à l'ensemble des employeurs du régime général et du régime agricole. Les employeurs publics entreront quant à eux au plus tard dans le dispositif début 2020.

LES TROIS QUARTS DES ENTREPRISES SATISFAITES

Le niveau de satisfaction des utilisateurs de la DSN est très encourageant après la généralisation : une étude menée en septembre 2017 par le Gip-MDS montre que 73 % des entreprises étaient d'ores et déjà satisfaites, alors que le démarrage a été compliqué pour une majorité d'entre elles. En outre, elles sont 80 % à bien comprendre le lien entre la paie et la DSN et à considérer que cette approche numérique constitue une réelle simplification à terme, une fois l'ensemble des processus rodés.

Pour les entreprises, les impacts du passage à la DSN les plus positifs concernent la transformation numérique et la simplification de la transmission des déclarations sociales.

Les cabinets d'experts comptables identifient les gains de la mise en place d'une norme unique et structurée pour tous les organismes de protection sociale induits par la DSN.

Le dispositif fonctionne bien en termes de qualité globale :

- Pour le régime général, 95 % des déclarations unifiées de cotisations sociales (DUCS) sont acceptées – les écarts restants ont majoritairement trait aux taux Accidents du travail et Versement transport puis aux régularisations de cotisations, encore mal maîtrisées dans de nombreux logiciels de paie.
- Pour le régime agricole, le taux de couverture DSN est désormais de l'ordre de 90 % et 95 % des DSN sont acceptées sans rejet.
- 93 % des Déclarations des salaires pour les indemnités journalières (DSIJ) sont automatiquement reconstituées à partir des signalements ; la moitié des DS IJ sont désormais issues du système d'information de la DSN et des gains de délai sur le paiement des assurés ont été notés (environ 6 jours).
- 79 % des Attestations employeurs sont également reconstituées, sachant que la moitié des fins de contrats « non courts » est désormais issue de la DSN.
- Le taux de conformité pour la Déclaration des mouvements de main-d'œuvre (DMMO), et désormais également de l'Enquête sur les mouvements de main-d'œuvre (EMMO) qui auparavant n'était que de 30 %, est désormais de 95%. Un bémol demeure néanmoins sur bémol qui demeure sur la disparition / apparition non justifiée de contrats du fait de la non-maîtrise du bloc changement par un certain nombre d'éditeurs et d'entreprises.
- Pour le secteur intérimaire, la DSN a constitué une avancée significative : le taux de qualité du Relevé mensuel de mission est de 94 % en moyenne et de 98 % pour les plus grandes entreprises du secteur ; en outre, le taux de qualité d'identification des salariés intérimaires est autour de 99 % (gain notable par rapport à la DADS-U).
- Le recouvrement de la retraite complémentaire porte désormais sur 87 % des entreprises et le taux d'intégration automatique est de 69 %, ce qui constitue un taux très significatif compte tenu de la jeunesse du dispositif de re-calcul systématique des cotisations au niveau individuel.
- Concernant les droits à retraite de base, une enquête réalisée par la Cnav a mis en évidence que seules 6 % des entreprises présentent un écart de qualité sur les données, un taux bien moindre que ce qu'il était en DADS U.

DES POINTS D'ATTENTION À SURVEILLER

Toutefois, ce bon fonctionnement d'ensemble ne doit pas masquer les difficultés dont les principales portent sur : la maîtrise des changements et des régularisations, le paramétrage et le bon fonctionnement du recouvrement des organismes complémentaires, lequel n'est pas encore à niveau industriel et, enfin, la gestion des fins de contrats avec Pôle emploi qui doit encore progresser.

Concernant les changements, il faut comprendre que puisque la DSN est une photo fin de mois, il est nécessaire que les logiciels de paie intègrent les informations nécessaires pour que l'on puisse dater un changement intervenu en cours de mois (par exemple une modification de la quotité de travail) sans ambiguïté. Certains logiciels ne le font pas systématiquement et de manière correcte.

Les régularisations de paie d'un mois sur l'autre ne sont pas non plus toujours bien prises en compte dans la DSN en raison de mauvais paramétrages au niveau de l'entreprise. En cas de rectification de la paie, corriger la DSN le mois suivant nécessite une gestion intermédiaire que tous les éditeurs de logiciels n'ont pas encore mise en place à ce jour.

Le recouvrement des cotisations pour les contrats complémentaires santé choisis par l'entreprise doit encore progresser. Son paramétrage doit être automatisé, ce qui est possible au niveau de la fiche de paramétrage des organismes complémentaires (OC). Mais tous les OC ne les ont pas encore mises en place et surtout tous les éditeurs ne les ont pas encore implémentées. Quand ce n'est pas fait, le gestionnaire de paie doit saisir ces informations à partir d'une fiche PDF. Quand on sait que le numéro de contrat est sur 14 caractères, on mesure le fort risque d'erreur à la saisie !

Enfin, la gestion des fins de contrat a été conçue sous l'hypothèse que les contrats longs (plus d'un mois) sont majoritaires. Dans la pratique, ce sont les contrats courts qui sont majoritaires dans les flux de fin de contrat depuis le début des années 2000 ; or le remplacement des attestations se fonde sur le fait d'avoir reçu les données de la DSN et donc du mois antérieur. Quand un contrat commence et finit le même mois, le système actuellement pensé ne marche pas. Une solution est en cours de conception.

Ainsi des chantiers sont en cours pour aboutir sur ces thèmes sachant qu'en parallèle, dans tous les organismes qui utilisent la DSN, le changement de « sens » et des modes de correction doit être compris, ce qui nécessite souvent de faire évoluer les systèmes d'information utilisant les données de la DSN.

DES BÉNÉFICES COMPLÉMENTAIRES ENCORE À VENIR

Dès 2019, la DSN sera le support du prélèvement à la source pour les salariés. L'administration fiscale calculera, selon les revenus 2017 déclarés au printemps 2018, le taux de prélèvement qui sera appliqué au salaire. Elle communiquera à l'employeur le taux de prélèvement retenu pour le contribuable, sauf s'il a opté pour le taux neutre. Dans ce dernier cas, la grille de taux par défaut sera appliquée. Dès octobre 2018, les entreprises récupéreront le fichier retour à intégrer dans la paie à partir de janvier 2019. Une DSN rodée en amont c'est la certitude d'un démarrage sans heurt de cette nouvelle procédure !

Par ailleurs, la DSN permet également de remplacer (depuis 2013 pour la radiation, et 2017 pour le recouvrement des cotisations et la prise en compte des droits) les formalités des organismes complémentaires santé et prévoyance ; ceux-ci sont nombreux et divers. L'automatisation pleine et entière évoquée ci-dessus apportera des gains importants en gestion car à l'heure actuelle une bonne partie des mutuelles fonctionnent encore en mode papier !

Le périmètre de la DSN a donc vocation à s'élargir encore, pour couvrir davantage de procédures et fonctionnalités et remplacer ainsi plus de déclarations, pour plus de simplification. Ainsi, les employeurs publics entrèrent au plus tard dans le dispositif début 2020. À venir également, en fonction des résultats des travaux en cours : déclaration d'emploi des travailleurs handicapés, attestation de salaires relative à un temps partiel thérapeutique, attestation d'emploi mensuel et fins de contrats des intermittents, formalités spécifiques à d'autres régimes spéciaux, formalités relatives à des conventions européennes, etc. Le calendrier de ces ajouts sera précisé une fois les grandes séquences d'extension statuées au niveau de l'État.

“ Un effort supplémentaire de réduction du nombre d'informations demandées aux entreprises restera donc à faire pour éviter que l'extension de la DSN ne conduise à un alourdissement qui serait préjudiciable au projet dans son ensemble. À cette fin, il est indispensable que les travaux de normalisation précèdent les extensions de périmètre pour éviter que la DSN ne devienne encore plus complexe que ce qui lui préexistait. ”

Un effort supplémentaire de réduction du nombre d'informations demandées aux entreprises restera donc à faire pour éviter que l'extension de la DSN ne conduise à un alourdissement qui serait préjudiciable au projet dans son ensemble. À cette fin, il est indispensable que les travaux de normalisation précèdent les extensions de périmètre pour éviter que la DSN ne devienne encore plus complexe que ce qui lui préexistait. Par ailleurs, les enjeux liés à la normalisation des données devraient être systématiquement pris en compte par tous les acteurs produisant les textes réglementaires comme conventionnels.

BIBLIOGRAPHIE

Cour des Comptes, « *Simplifier la collecte des prélèvements versés par les entreprises* », juillet 2016.

HCFIPS, « *Rapport sur les relations des entreprises avec les organismes de protection sociale* », juillet 2017.

Inspection générale des affaires sociales, Inspection générale des finances, « *Rapport sur la mise en œuvre de la nouvelle norme des déclarations annuelles de données sociales et les conditions de mise en œuvre d'une déclaration sociale nominative* », mars 2011.

Vatinet R., « *La Déclaration Sociale Nominative (DSN), simple évolution ou véritable révolution des relations entre les employeurs et les organismes de Protection sociale ?* », Regards, n° 43, Ecole nationale supérieure de sécurité sociale, pp. 91-97, janvier 2013.

BIEN COMPRENDRE LA DÉCLARATION SOCIALE NOMINATIVE POUR MIEUX MESURER

Catherine Renne*

L'Insee établit des statistiques sur l'emploi et les salaires depuis plus d'un demi-siècle à partir de différentes déclarations administratives. La mise en place de la déclaration sociale nominative n'a pas simplement consisté à fusionner les déclarations administratives existantes. Elle a également conduit à rationaliser et simplifier les données d'une part, et à définir rigoureusement le cadre dans lequel s'effectuent les échanges d'informations entre les entreprises et les organismes destinataires de la DSN, d'autre part. Elle a obligé l'institut à mener une refonte de grande ampleur de son système d'information en même temps qu'elle est venue questionner l'organisation des travaux. Les contrôles réalisés en amont, à l'occasion de la collecte des données, assurent en contrepartie de recevoir les déclarations avec un certain degré de qualité et donc avec une certaine valeur ajoutée. Pour tirer le meilleur parti de la DSN, le statisticien doit mieux connaître l'étape de collecte et de contrôle et s'impliquer davantage dans les nouvelles instances de gouvernance de la DSN. Même si la priorité à court terme a été d'intégrer la DSN dans son système d'information, la prochaine étape va consister pour l'Insee à construire un système suffisamment souple et modulaire qui permettra d'exploiter toutes les potentialités de cette nouvelle source d'information.

❶ LA DSN, UNE SOURCE INCONTOURNABLE POUR LE STATISTICIEN

Déclaration unifiée de cotisations sociales, déclaration mensuelle de mouvements de main-d'œuvre, relevé mensuel de missions, déclaration annuelle de données sociales... Toutes ces déclarations administratives auxquelles étaient soumis encore récemment la plupart des employeurs sont désormais remplacées par la déclaration sociale nominative (DSN) (voir l'article d'E. Humbert-Bottin dans le même numéro).

Mise en place progressivement depuis 2013, la DSN est bien plus qu'une nouvelle déclaration administrative, elle institue une nouvelle logique déclarative entre les entreprises et les organismes destinataires (Buhl, 2014). De fait, la DSN n'a pas vocation à servir un besoin spécifique, mais au contraire à couvrir différents usages.

Avant la DSN, les employeurs transmettaient des données parfois redondantes à une diversité d'acteurs avec des échéances et des formats différents. Désormais, ils transmettent une seule fois chaque mois les informations issues de leur logiciel de paie à un point

* Cheffe-adjointe du département de l'emploi et des revenus d'activité, Insee

de dépôt unique. Ces informations sont contrôlées puis redistribuées aux différents organismes destinataires, charge à eux de les transformer pour les mettre au format qui leur convient. D'une certaine façon, la DSN transfère une partie de la charge des entreprises vers les systèmes d'information en aval.

Pour l'Insee, la généralisation de la DSN se traduit progressivement par le remplacement des principales sources qui alimentent son système d'information sur l'emploi et les revenus d'activité (Siera). Dit autrement, la DSN devient « LA » source d'information majeure et incontournable et il est essentiel que les statisticiens en comprennent et maîtrisent les différents aspects de son traitement.

Lorsqu'on parle du traitement d'une déclaration administrative, on pense en général aux traitements statistiques qui sont mis en œuvre, notamment à l'Insee. Mais la phase amont de collecte et de contrôle de l'information qui, elle, relève largement d'acteurs extérieurs est tout aussi importante. Elle est souvent méconnue des utilisateurs finaux alors que c'est précisément cette étape qui leur assure de recevoir les déclarations avec un certain degré de qualité et donc avec une certaine valeur ajoutée.

Cette problématique n'est pas spécifique à la DSN et est largement généralisable à toute autre déclaration administrative. L'objet de cet article est de décrire, à travers l'exemple illustratif de la DSN, les étapes de gestion de la qualité d'un processus de traitement d'une déclaration administrative.

RATIONALISER ET SIMPLIFIER : DEUX OBJECTIFS AFFICHÉS DE LA DSN

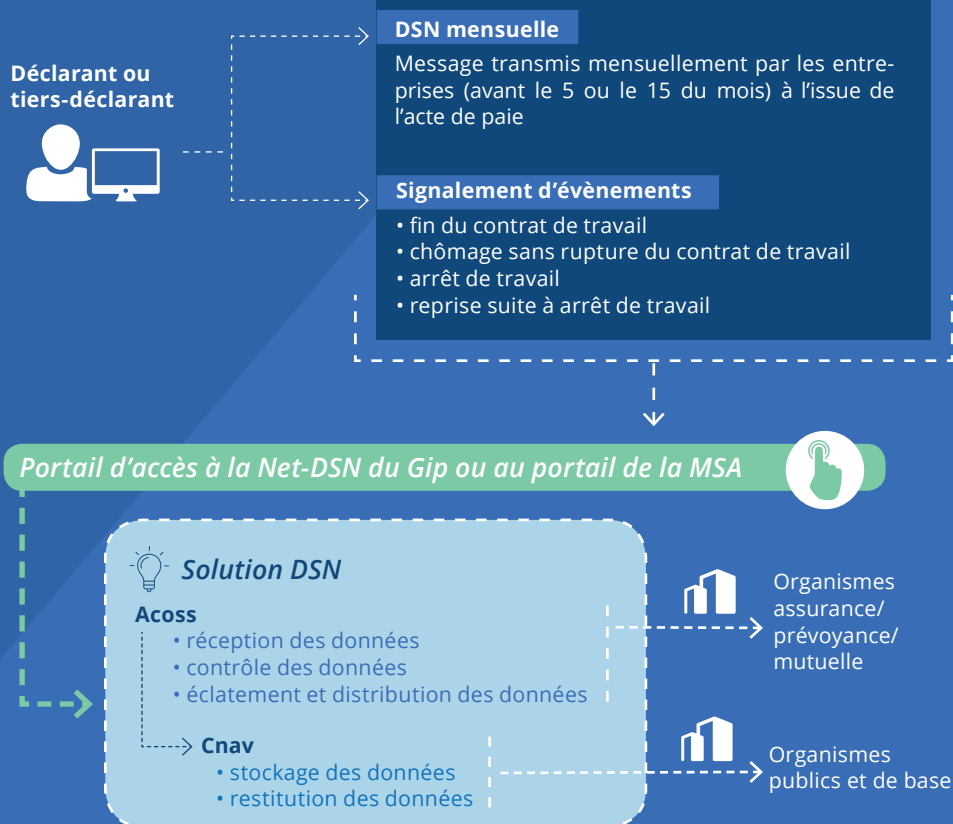
Avec la DSN, les entreprises transmettent trois fois moins de données que lorsqu'elles remplissaient leur déclaration annuelle de données sociales. En adoptant le principe du « Dites-le nous une fois », la DSN limite en effet les données demandées aux seules informations élémentaires qui ne peuvent être obtenues par un autre moyen. Par exemple, depuis janvier 2018, les entreprises n'ont plus obligation de fournir leur effectif salarié de fin de période, celui-ci pouvant être recalculé directement par les organismes destinataires à partir des informations individuelles transmises sur les salariés.

La rationalisation des données a permis également la mutualisation de l'information collectée pour un grand nombre d'organismes. D'une part, la DSN ne contient plus de rubriques spécifiques à tel ou tel organisme comme c'était le cas dans la déclaration annuelle de données sociales. D'autre part, le fait de rassembler toutes les informations dans une déclaration unique garantit la cohérence de l'information destinée à des usages différents. Par exemple, avec la DSN, on est sûr que les employeurs et les salariés sont identifiés de la même façon quel que soit l'organisme destinataire de l'information.

Cet objectif de rationalisation n'aurait pas pu être atteint sans une définition rigoureuse du cadre dans lequel s'effectuent les échanges d'informations entre les entreprises et les organismes destinataires de la DSN. Ce cadre appelé également norme d'échanges¹ définit le modèle d'information, le modèle de message et la cinématique des échanges. Leur description figure dans un document appelé « cahier technique » qui est mis à jour chaque année et versionné. Le modèle d'information contient la définition des différentes rubriques à renseigner, précise leur sémantique, c'est-à-dire comment elles sont organisées et à quoi elles sont rattachées (l'entreprise, le contrat, le salarié ?...) et vérifie que les données ne sont pas demandées plusieurs fois. Par ailleurs, lorsque les employeurs transmettent

1. Dans le cas de la DSN, il s'agit de la norme NEODeS : Norme d'Échange Optimisée des Données Sociales.

Encadré 1. La cinématique des échanges de la DSN



leur DSN, ils n'envoient pas les informations « en vrac » mais selon un modèle de message qui place les informations dans un certain ordre et selon une certaine structure.

La cinématique des échanges décrit, quant à elle, le processus d'échange des données entre les différents organismes partenaires. De façon très schématique, le processus de collecte de la DSN s'appuie sur deux portails de déclaration : le [portail net-entreprise.fr](https://net-entreprise.fr) pour l'ensemble des entreprises et le portail msa.fr uniquement destiné aux employeurs du régime agricole. Les informations transmises dans les DSN mensuelles sont d'abord réceptionnées et contrôlées par l'Acoss² qui les transmet à la Caisse nationale d'assurance vieillesse (Cnav) d'une part, et les filtre pour distribuer un extrait du message aux organismes d'assurance, de prévoyance et de mutuelle, d'autre part. La Cnav certifie l'identification des salariés (Nir) grâce à son système national de gestion des identifiants (SNGI), stocke les DSN pour ensuite les redistribuer de façon dématérialisée aux organismes de protection sociale et aux organismes publics, dont l'Insee (**encadré 1**).

2. Agence centrale des organismes de sécurité sociale.

S'IMPLIQUER DANS LA GOUVERNANCE DE LA QUALITÉ DE LA COLLECTE

L'utilisation de déclarations ou de sources administratives à des fins statistiques est préconisée par le code de bonnes pratiques de la statistique européenne dans le but d'alléger la charge statistique des déclarants. En France, ceci est rendu possible par la loi n° 51-711 de 1951 relative à l'obligation, à la coordination et au secret en matière de statistique et a été réaffirmé récemment par la loi n° 2016-1321 de 2016 pour une République numérique.

Si la loi facilite l'accès de l'Insee aux déclarations administratives, l'obtention des données se fait en général après un long temps d'échanges, de préparation et de formalisation avec l'organisme concentrateur. Cette phase préliminaire du processus ne doit pas être sous-estimée, car elle conditionne les étapes ultérieures du processus et l'utilisation que l'on pourra faire des données. Elle doit l'être d'autant moins que les usages statistiques de la DSN ne sont pas de même nature que les usages administratifs en fonction desquels le nouveau système a été conçu en priorité.

S'agissant de la DSN, l'Insee a conclu une convention tripartite en 2016 avec la Cnav et le Gip-MDS³, qui précise les modalités de livraison (date de livraison, contenu et format des fichiers, etc.). Un contrat de service a également été mis en place avec la Cnav dont le contenu précise de façon plus technique les modalités d'échanges des fichiers. Enfin, les traitements envisagés par l'Insee et les modalités de conservation des données ont fait l'objet d'une déclaration auprès de la Cnil.

Mais la bonne utilisation de la DSN nécessite également un suivi attentif des évolutions du cahier technique. Celui-ci est susceptible d'évoluer chaque année pour intégrer notamment de nouvelles rubriques. C'est le cas de la version 2019 du cahier technique qui intégrera de nouvelles rubriques spécifiques aux salariés de la fonction publique ou de la version 2020 qui intégrera quant à elle la Déclaration Obligatoire d'Emploi des Travailleurs Handicapés (DOETH). L'Insee doit alors signaler au Gip-MDS s'il souhaite ou pas recevoir ces nouvelles rubriques.

L'Insee est en effet régulièrement informé des modifications des rubriques de la norme via les instances de gouvernance, notamment le comité de normalisation des données sociales dont il est membre. Il a également des échanges réguliers avec le Gip-MDS qui assure la maîtrise d'ouvrage opérationnelle du projet DSN. L'institut peut, au même titre que les autres organismes, faire remonter les anomalies ou erreurs qu'il pourrait observer dans les données afin d'améliorer leur qualité. D'une certaine façon, le système DSN fonctionne comme une boucle vertueuse, dans laquelle l'institut doit prendre progressivement toute sa place.

UNE COLLECTE TRÈS CONTRÔLÉE

L'utilisation de déclarations administratives n'est pas toujours aisée pour le statisticien, car ce dernier n'est pas directement à l'origine des concepts sous-jacents ou de la collecte. Il se doit pourtant de comprendre les différentes étapes de la fabrication des données qu'il reçoit afin de juger de leur qualité, d'une part, et d'éviter de reproduire inutilement des traitements qui auraient déjà été réalisés en amont, d'autre part.

Lorsque l'Insee réceptionne chaque mois les déclarations sociales nominatives, celles-ci ont déjà fait l'objet d'une longue série de vérifications automatiques destinées à sécuriser le processus déclaratif. Cette notion de contrôle est à ce point essentielle dans le dispo-

3. Groupement d'intérêt public « Modernisation des déclarations sociales ».

sitif de collecte que les entreprises peuvent tester, à l'aide d'un outil d'auto-contrôle mis à leur disposition, le fichier contenant leur déclaration avant de le transmettre. Cet outil est cohérent avec celui de la plateforme de déclaration qui met en œuvre par la suite divers types de contrôles.

Des contrôles de structure et de syntaxe permettent de vérifier la conformité d'ensemble des fichiers envoyés. Il s'agit par exemple de vérifier que les variables déclarées respectent bien une certaine longueur et un certain format. Des contrôles de cohérence vérifient, quant à eux, la cohérence entre différentes rubriques. Par exemple, dans les éléments d'adresse, le renseignement du libellé d'une localité oblige à renseigner le code postal associé. Des contrôles sur des référentiels externes vérifient que la valeur prise par certaines données déclarées appartient bien à une nomenclature (catégorie socio-professionnelle, activité de l'établissement, code postal, etc.). Enfin, la validité de tous les numéros SIRET déclarés dans la DSN font l'objet d'une vérification, notamment le SIRET de l'établissement utilisateur et celui du lieu de travail lorsque celui-ci a été déclaré comme un établissement. Le SIRET des déclarants est, quant à lui, contrôlé dès le portail d'identification.

Tous ces contrôles sont bloquants, c'est-à-dire qu'ils entraînent le rejet complet de la déclaration tant que les erreurs ne sont pas corrigées. Ils représentent l'essentiel des vérifications automatiques faites par le processus de collecte de la DSN. Il existe par ailleurs des contrôles non bloquants qui ne sont présents qu'à titre d'information. Il n'était en effet pas envisageable de rendre bloquants tous les contrôles effectués sur les DSN au risque d'empêcher leur transmission. Les contrôles effectués sur la DSN sont uniquement des contrôles de flux ; ils ne permettent pas de vérifier la cohérence des informations avec une autre source ni celle des données transmises d'un mois ou d'une année sur l'autre. Tous ces contrôles sont documentés et décrits précisément dans le cahier technique de la norme d'échanges de la DSN et ont fait l'objet d'une réflexion concertée avec les différents organismes destinataires.

TRANSFORMER LES DONNÉES POUR RÉPONDRE AUX BESOINS DE LA STATISTIQUE

Pour le statisticien, disposer d'une source d'information documentée et contrôlée en amont, comme la déclaration sociale nominative, est plus que précieux. Cependant, son travail ne consiste pas à reproduire simplement l'information qu'il a reçue. Avant de la diffuser, il doit encore procéder à des contrôles et autres traitements afin de transformer cette matière brute en une information statistique pertinente pour l'analyse économique.

Le contrôle d'exhaustivité, souvent appelé de façon abusive « repérage des trous de collecte » fait partie des étapes essentielles du traitement statistique. S'agissant de la DSN, il ne s'agit pas de se substituer au Gip-MDS qui contrôle déjà l'exhaustivité de la collecte des déclarations en relançant les entreprises retardataires. À ce stade, il s'agit de vérifier que l'Insee a bien reçu toutes les déclarations qu'il était supposé recevoir, d'une part et qu'il a bien intégré dans son système d'information toute l'information qu'il a reçue, d'autre part. Toute la difficulté est de définir la notion de « déclarations attendues ».

Une première façon de procéder est d'effectuer des contrôles de vraisemblance « macro » destinés à repérer une éventuelle évolution anormale du nombre de déclarations reçues. En général ce type de contrôles se fait au moment de la réception des données, ce qui permet de réagir rapidement en cas de détection de déclarations manquantes conformément au contrat de service établi avec l'organisme émetteur de la déclaration. Le statisticien peut également effectuer des contrôles plus « micro » en confrontant les déclarations reçues au cours d'une période donnée avec celles de la période précédente ou à un référentiel externe. Avant la mise en place de la DSN, ce référentiel externe était construit

à partir des différentes déclarations administratives reçues à l'Insee. Le remplacement des différentes déclarations par une déclaration unique, la DSN rend progressivement difficile la construction d'un tel référentiel. Il serait en effet paradoxal de construire un référentiel externe à partir des données reçues en DSN alors que c'est précisément l'exhaustivité de cette information que l'on cherche à contrôler. Les contrôles de cohérence intertemporelle seront donc de plus en plus privilégiés au fur et à mesure de la montée en charge de la DSN.

Ensuite, le statisticien va s'assurer de la vraisemblance des variables qu'il juge structurantes pour son système d'information. Dans cette phase du processus de traitement de la déclaration administrative, la détection d'éventuelles anomalies constitue un enjeu important. Plusieurs questions se posent alors : comment distinguer, parmi ces anomalies, celles qui s'avèrent être des erreurs et doivent donc être corrigées ? Dit autrement, si on trouve dans les déclarations sociales des salaires horaires inférieurs au Smic ou à l'opposé des salaires horaires extrêmement élevés, s'agit-il d'une erreur de déclaration ou le reflet de la réalité ? Ces erreurs peuvent-elles être traitées de façon automatique ou nécessitent-elles un traitement manuel ? Toutes les erreurs

Encadré 2. Le contrat de travail : une notion centrale en DSN

Les caractéristiques du contrat :

- libellé d'emploi
- catégorie socio-professionnelle
- nature du contrat (CDD, CDI, intérim...)
- motif de recours (pour les CDD et les missions)
- dispositif de politique publique (contrats aidés)
- établissement utilisateur (pour les contrats d'intérim)
- lieu de travail
- multi-activité du salarié.

Les revenus rattachés au contrat :

- durée de la période d'activité
- éléments de rémunération
- primes, gratifications et indemnités.

La vie du contrat :

- date de début
- date prévisionnelle de fin de contrat (CDD)
- date de fin
- motif de rupture
- arrêts de travail (motif, durée).

Les cotisations sociales et les taxes :

- bases assujetties
- cotisations individuelles
- exonérations
- assujettissement fiscal (taxe d'apprentissage, taxe sur les salaires...).

“ *Comment distinguer, parmi ces anomalies, celles qui s'avèrent être des erreurs et doivent donc être corrigées ?* ”

doivent-elles être corrigées et si non, sur quels critères définit-on les priorités de traitement et à quel moment doit-on s'arrêter ? Du fait de la multiplicité des usages parfois antagonistes, il est souvent difficile de fixer un critère d'arrêt unique. On sera en effet plus ou moins exigeant sur le critère d'arrêt selon qu'on souhaitera publier

des données individuelles ou des estimations agrégées pour lesquelles on peut tout à fait accepter qu'elles reposent sur des données individuelles imparfaites. Le contrôle de la vraisemblance des données se poursuit donc en aval de la chaîne de traitement de la déclaration administrative. Dans le domaine de l'emploi et des salaires, le processus de validation des données produites à partir des déclarations administratives est un processus long et itératif qui implique, outre l'Insee, de nombreux services statistiques ministériels (Dares, DGAFP, Drees, DGCL, OED...).

Mais la principale difficulté pour le statisticien lorsqu'il travaille à partir d'une déclaration administrative, c'est que celle-ci n'a pas été conçue pour son usage exclusif. La déclaration peut donc contenir des informations dont il n'a pas besoin et surtout ne pas reposer sur ses unités statistiques de référence. Dans le domaine de l'emploi et des revenus d'activité par exemple, on va s'intéresser à des notions telles que le poste de travail ou les périodes d'activité. La DSN véhicule quant à elle les informations relatives à des contrats de travail et des périodes de rémunération, d'une part et à des signalements d'événements comme les arrêts de travail, d'autre part (**encadré 2**).

🕒 UN CHANGEMENT D'ÉCHELLE ET DE TEMPORALITÉ DE LA PRODUCTION STATISTIQUE

Avec la mise en place de la DSN, c'est bien plus que le remplacement progressif des déclarations administratives qui alimentaient le système d'information de l'Insee qui est à l'œuvre. L'institut doit désormais se donner les moyens de traiter chaque mois un volume d'information équivalent à celui d'une déclaration annuelle de données sociales. La volumétrie des données devient un enjeu important ; il s'agit en effet de stocker et traiter chaque mois un peu moins d'une centaine de gigaoctets de données, soit plus d'un téraoctet par an. Vu sous cet angle, la DSN aurait pu être considérée uniquement par l'Insee comme une contrainte technique et traitée comme telle, ce qui aurait été une erreur.

Sans nul doute, la DSN vient bouleverser les systèmes d'information construits depuis longtemps autour des différentes déclarations administratives. Pour l'Insee, il y a là une formidable opportunité de repenser et documenter ses processus de production afin de gagner en efficacité et en cohérence. Même si la priorité à court terme a été d'intégrer la DSN, il faut dès à présent anticiper la demande qui sera adressée à l'Insee à moyen terme, dans le domaine de l'emploi et des revenus d'activité, afin de construire un système suffisamment souple et modulaire qui permettra d'y répondre.

Le fait de disposer de données mensuelles permet de traiter désormais les données au fil de l'eau, ce qui devrait à terme réduire les délais de production. Devoir traiter chaque mois environ 2 millions de déclarations contenant les déclarations sociales nominatives de 20 millions de salariés oblige à se montrer parcimonieux dans les traitements manuels confiés aux gestionnaires. Les nouveaux processus mis en place se fondent donc largement sur un principe de capitalisation afin d'éviter de reproduire chaque mois les mêmes traitements. Par exemple, si un gestionnaire traite une anomalie présente dans la catégorie socio-professionnelle d'un salarié un mois donné, et que l'anomalie subsiste dans la déclaration du mois suivant, le système reproduira de façon automatique le traitement effectué

par le gestionnaire le mois précédent. Les travaux menés sur la qualité de la localisation des emplois, dits de « dégroupement », seront désormais conduits au sein d'un seul processus au sein du Siera. Ils étaient pris en charge jusqu'à présent par différents processus au sein et à l'extérieur du Siera. Enfin, la ré-ingénierie des processus se fera avec la volonté d'améliorer la cohérence entre données macro et données micro, d'une part et entre statistiques en évolution et en niveau, d'autre part.

PIRÉNÉS⁴ : UN PROJET DE RÉINGÉNIERIE AMBITIEUX

L'Insee a décidé de lancer en 2012 un ensemble de projets de ré-ingénierie des processus du Siera, dont le premier, le projet Pirénés, a pour objectif d'accueillir et de traiter la déclaration sociale nominative. Le processus de production a été organisé autour d'applications remplissant chacune un rôle bien précis.

Il s'agit tout d'abord d'être en mesure d'accueillir chaque mois environ 2 millions de fichiers. La première étape pour l'Insee consiste à les intégrer selon son propre modèle de données en ayant au préalable transcrit les données qu'ils contiennent en « langage statistique ». Il peut s'agir d'un simple renommage des variables ou d'une transformation plus complexe. Cette opération dite de « mapping », qui concerne une centaine de variables, permet aux phases ultérieures du processus de ne jamais faire appel aux données sources mais uniquement à ces nouvelles variables. Cette étape du processus s'appuie sur l'application Arc⁵. À l'origine, l'application Arc avait été conçue pour réaliser également un certain nombre de contrôles de conformité mais l'expérience a montré que la quasi-totalité d'entre eux étaient inutiles puisque déjà réalisés par le processus de collecte de la DSN.

Une fois les données intégrées dans le système d'information, des traitements sont effectués sur chaque DSN mensuelle dans l'application Artemis⁶. Il s'agit à ce stade de s'assurer de la qualité de variables telles que l'identification des employeurs ou encore l'identification, le lieu de résidence et la catégorie socio-professionnelle des salariés (PCS). Le principe est de déterminer pour chaque information unitaire s'il y a ou non anomalie. S'agissant de l'identification des employeurs et de la commune de résidence des salariés, les informations déclarées sont comparées à celles figurant dans des référentiels (référentiel d'unités statistiques « Sirus » et répertoire « Sirene », code officiel géographique). Le nombre d'anomalies détectées sur un mois est très faible au regard du nombre de déclarations reçues à savoir de l'ordre d'un pour dix mille pour le SIRET de l'employeur et de l'ordre d'un pour mille pour le code de la commune de résidence du salarié. Pour la catégorie socio-professionnelle (PCS), on compare le code déclaré par l'employeur à celui calculé par l'outil Sicore⁷ à partir notamment du libellé de profession déclaré. Dans 90 % des cas, le code profession déclaré par l'employeur est confirmé par l'outil de codification automatique, ce qui signifie que la majorité des employeurs déclarent des codes et des libellés de profession cohérents. Malgré tout, ce sont environ 1,5 million d'anomalies qui doivent potentiellement être traitées chaque mois, ce qui n'est matériellement pas possible. Des arbitrages coût-qualité ont donc été faits dans un premier temps, qui conduisent à ne traiter que les divergences de codification portant sur les deux premières positions du code. Cela représente tout de même 40 % des anomalies à traiter. Compte tenu de leur nombre encore important, les cas à traiter sont regroupés en grappes homogènes. Le gestionnaire établit alors une règle de traitement au niveau de la grappe qui sera ensuite appliquée de façon auto-

4. Projet informatique de refonte sur l'emploi et les salaires.

5. Accueil Réception Contrôle.

6. Application de reprise et de traitements élémentaires de l'emploi et des salaires.

7. Système Informatique de COdage des Réponses aux Enquêtes.

matique à l'ensemble des cas à traiter dans la grappe. Ces règles sont capitalisées par l'application et seront appliquées de façon automatique lors du traitement des déclarations des mois suivants, ce qui doit conduire à diminuer à terme le nombre de cas à traiter par les gestionnaires. Les gains engendrés par ce processus de capitalisation seront réemployés pour traiter les cas de divergence portant sur les deux dernières positions du code PCS.

“ Des arbitrages coût-qualité ont dû être faits et seules les anomalies les plus importantes sont soumises à l'expertise des gestionnaires. ”

L'étape suivante consiste à analyser les incohérences internes aux DSN d'une même année entre les variables sur le volume de travail et celles relatives aux éléments de la rémunération, une fois les 12 mois de déclarations agrégés. Là encore, des arbitrages coût-qualité ont dû être faits et seules les

anomalies les plus importantes sont soumises à l'expertise des gestionnaires. Les traitements de détection et de correction des anomalies sont effectués dans l'application Diane⁸. Les méthodes de détection d'anomalies sont parfois relativement complexes et reposent pour certaines sur des modèles économétriques.

À ce stade des traitements, il reste encore à reconstituer l'information selon l'unité statistique d'intérêt (le poste de travail) et de calculer un certain nombre de variables statistiques. C'est l'objet de l'application Thésée⁹. Cette application permet également de produire un fichier sur l'ensemble des salariés c'est-à-dire en incluant les informations issues des systèmes d'information sur l'emploi et les salaires des agents de la fonction publique d'une part, et les salariés des particuliers employeurs d'autre part.

❶ APRÈS DEUX ANS D'UTILISATION, QUEL BILAN ET QUELLES PERSPECTIVES ?

La mise en place de la DSN ouvre des perspectives nouvelles pour la production des statistiques sur l'emploi et les revenus d'activité. Les attentes des utilisateurs en la matière sont importantes, en atteste le rapport du Conseil national de l'information statistique (Cnis) de mars 2016 sur la diversité des formes d'emploi. Il est cependant prématuré de fixer l'échéance à laquelle l'Insee sera en mesure d'exploiter toutes les potentialités de cette nouvelle source d'information. Par ailleurs, il faudra attendre 2022 et l'entrée de tous les employeurs publics en DSN pour avoir une couverture quasi exhaustive du champ des employeurs. À ce stade, il est tout de même possible de dresser un premier bilan.

Comme c'est souvent le cas dans les projets de cette ampleur, les coûts de transition ont été sous-estimés. L'Insee a dû composer avec un projet externe dont il ne maîtrisait pas le calendrier. Le rythme de montée en charge de la DSN a été plus lent qu'initialement prévu et a obligé à accueillir des DADS en nombre important et à faire cohabiter les deux déclarations administratives.

L'Insee a dû apprendre à évoluer dans un environnement qui n'est pas le sien, avec des acteurs habitués à travailler selon un certain formalisme faisant appel à la notion de norme d'échanges et avec des exigences en termes de réactivité. En quelque sorte, la DSN est venue bousculer la temporalité de nos travaux. Pour autant, l'Insee doit continuer à développer les échanges en amont avec ces acteurs afin de profiter pleinement de la boucle vertueuse que le projet externe DSN a su mettre en place. Un mode de travail collaboratif similaire mériterait d'être mis en place, cette fois en aval du processus,

8. Dispositif Informatique pour l'Agrégation et la Normalisation de l'Emploi.

9. Traitement Harmonisé des Salaires et de l'Emploi.

avec les acteurs du Système Statistique Public (SSP) ou du péri-SSP également destinataires de la DSN, avec qui l'Insee partage des préoccupations communes. Par exemple, la question de l'utilisation de la DSN pour alléger le questionnaire de certaines enquêtes statistiques se pose à l'Insee pour l'enquête coût de la main-d'œuvre et structure des salaires (Ecmoss) mais aussi à la Dares¹⁰ pour l'enquête activité et conditions d'emploi de la main-d'œuvre (Acemo), notamment.

En interne à l'Insee, il reste maintenant à mettre en place des processus industrialisés, documentés et sécurisés. Le traitement de la DSN à l'Insee doit s'appuyer sur une démarche qualité, ce qui suppose une description fine des processus, la mise en place d'un cadre d'assurance qualité et une revue régulière des processus. Ceci est d'autant plus important que le projet DSN n'est pas encore terminé qu'un nouveau chantier de simplification voit le jour. Il doit permettre aux organismes délivrant des revenus de remplacement (retraites, allocations chômage, etc.) de le faire à leur tour selon les mêmes principes que la DSN, c'est-à-dire selon une norme d'échanges précise. Dès lors, la démarche mise en place à l'Insee pour le traitement de la DSN à des fins statistiques devrait être facilement généralisable au traitement de cette nouvelle déclaration administrative.

BIBLIOGRAPHIE

Bonnet O., Cordier-Villoing M., Deroyon T., Djiriguian J. et Sakarovitch B., « 5 324 euros de l'heure : outlier ou footballeur ? Méthodes d'apprentissage non supervisé pour la détection d'anomalies : application au cas de la Déclaration sociale nominative », 13^e Journées de méthodologie statistique de l'Insee (JMS 2018), juin 2018.

Buhl J.L., « La Déclaration sociale nominative (DSN) et l'accès aux droits », Regards n° 46, EN3S, 2014.

Gazier B., Picart et C., Minni C., « La diversité des formes d'emploi », Rapport du Cnis, n° 142, juillet 2016.

Millin K., « Reconstitution des mouvements de main-d'œuvre depuis 1993 », Document d'études Dares, n° 221, juin 2018.

Rivière P., « Approche coût-qualité pour l'amélioration des processus de production statistique », Courrier des statistiques, n° 105-106, pp. 65-75, juin 2003.

Rouppert B., « Modélisation du processus de traitement d'une source administrative à des fins statistiques », Document de travail Insee, n° C2005/02, 2005.

Pour en savoir plus sur la DSN :

Site internet du Gip-MDS : www.gip-mds.fr

Site internet de la DSN : <http://www.dsn-info.fr/>

Cahier technique de la norme Neodes : <http://www.dsn-info.fr/dsn-norme-phase3.htm>

10. Direction de l'Animation de la Recherche, des Études et des Statistiques du ministère du travail.

DE L'ACTIVITÉ DE LA JUSTICE AU SUIVI DU JUSTICIABLE

FAIRE PARLER LES DONNÉES DE GESTION

Christine Chambaz*

La statistique sur la justice est une statistique bâtie essentiellement sur l'exploitation des données administratives issues des applicatifs de gestion des juridictions et des services. Elle ne s'appuie que très rarement sur des enquêtes auprès des personnes, ne serait-ce que parce qu'elle ne dispose pas de façon naturelle d'une base de sondage qui permettrait de les organiser. Il n'existe en effet pas de fichier général des justiciables, et la plupart des sources mobilisables sont expurgées des éléments nominatifs avant d'être transmises au service statistique. Le positionnement de cette statistique par rapport à la production de données décisionnelles est ambigu, quand bien même les deux portent des objectifs différents. La diversité des processus de collecte reflète la diversité des sources à exploiter. Au-delà des questions techniques, les difficultés sont liées au passage de données de gestion vivantes, à travers des concepts métiers également vivants, à des données statistiques datées, adossées à des concepts et des référentiels plus stables. Mais le véritable défi pour les années à venir reste de passer d'un système rendant compte de l'activité de la justice à un système d'observation du parcours des justiciables. L'évolution des systèmes d'information et l'extension des possibilités d'appariement ouvrent de nouvelles perspectives, mais des évolutions législatives seront également indispensables.

LE JUSTICIALE : L'INVENTION D'UNE UNITÉ STATISTIQUE

La statistique sur la justice a longtemps été centrée sur le suivi du volume d'activité des juridictions, civiles ou pénales, dans une approche uniquement quantitative. Elle s'attache désormais de plus en plus à décrire les contentieux qu'elle traite, voire les justiciables qui les portent, dans une approche plus explicative des comportements humains. Ce glissement a été permis par l'évolution des sources mobilisables, elle-même sous-tendue par un intérêt accru pour la connaissance économique et sociologique de la justice, et un souci d'évaluation des politiques publiques. Il s'accompagne d'un changement d'unités statistiques qui peut être porteur d'ambiguïté pour qui s'intéresse à des séries historiques. Il s'inscrit aussi dans le développement simultané, aux côtés d'indicateurs de statistique publique, d'indicateurs décisionnels, tournés vers le suivi de la performance et dont les propriétés peuvent être différentes, en termes de fraîcheur mais aussi de complétude.

* Cheffe de la sous-direction de la statistique et des études, ministère de la Justice

S'intéresser aux justiciables suppose de pouvoir faire le lien entre les affaires et les parties prenantes à ces affaires. Et cela est plus ou moins aisé. Contentieux par contentieux, et affaire par affaire, cela semble évident : un divorce se prononce généralement à l'encontre de deux ex-conjoints, un contentieux locatif oppose un bailleur et un locataire, une agression compte un (des) auteurs(s) et une (des) victime(s), etc. La réalité peut cependant être plus compliquée, notamment en matière pénale : supposons ainsi une affaire qui arrive au parquet avec plusieurs personnes (ou « auteurs ») impliquées. Le traitement du dossier par le parquet peut conduire la justice à poursuivre certaines de ces personnes, tandis que d'autres seront mises hors de cause ou verront l'affaire classée, pour ce qui les concerne ; a contrario, l'instruction du dossier peut conduire à identifier d'autres « auteurs » qui viendront grossir les parties prenantes de l'affaire, en cours de traitement. On voit ici que le nombre d'auteurs de cette affaire, selon les étapes considérées de la procédure judiciaire, pourra fluctuer. Compter les justiciables, ou les « auteurs » impliqués dans les affaires suppose donc non pas tant de pondérer les affaires par le nombre de leurs auteurs que de suivre ces auteurs eux-mêmes tout au long de leur aventure judiciaire. Ce qui implique la construction d'un système d'information statistique structuré selon ces étapes, au plus près de la pratique des juridictions, qui permette de rendre compte des « parcours judiciaires ». Produire une statistique sur les parcours judiciaires nécessite de comprendre la procédure judiciaire et ses pratiques. C'est une statistique fortement imprégnée du « métier », des processus de production de la justice.

Par ailleurs, l'intérêt porté aux justiciables fait naître rapidement une problématique de suivi de ces justiciables au cours du temps, au-delà d'une affaire bien définie qui les a fait rencontrer la justice, à un moment donné. Sont-ce les mêmes qui reviennent pour régler des différends ou des actes successifs ? Les auteurs d'infractions récidivent-ils au sens commun - qui suggère simplement un retour vers la justice rendant une réponse pénale - plutôt qu'au sens légal¹ plus restrictif ?

En l'état du système d'information, le suivi des personnes est encore impossible, faute d'identifiant commun dans les différents applicatifs de gestion, et faute de disposer de coordonnées nominatives qui permettraient la construction d'un tel identifiant dans les fichiers statistiques. En matière civile, la statistique reste jusqu'à présent une statistique relative aux affaires. En matière pénale, la volonté de produire des statistiques sur les justiciables force à imaginer une unité statistique originale, intermédiaire entre l'affaire et l'auteur : « l'auteur-affaire ». Ne pouvant repérer le retour d'un auteur dans des affaires successives, le statisticien estime des volumes de fréquentation de la justice selon différents critères, de l'orientation par les parquets à la filière de jugement en passant par les infractions commises et les peines prononcées. De la même façon, il approche le nombre des victimes ou des parties civiles par des « victimes-affaires ».

UNE PRODUCTION EN TENSION ENTRE DÉCISIONNEL ET STATISTIQUE

L'introduction du justiciable comme unité de compte était censée lever l'ambiguïté entre indicateurs décisionnels pour le pilotage opérationnel et indicateurs statistiques de plus long terme pour l'évaluation des politiques de justice, dans une approche plus sociologique de la justice. Pourtant, même si le point peut sembler anodin, l'acrobatie consistant à produire des statistiques sur les justiciables via des unités *ad hoc* contribue à entretenir

1. La récidive légale est la réitération d'une infraction similaire ou proche d'une infraction précédemment et définitivement condamnée. Les règles de la récidive varient selon que l'infraction est contraventionnelle, délictuelle ou criminelle.

cette ambiguïté. Le suivi des auteurs peut en effet être compris, même si ce n'est pas son but premier, comme la possibilité de pondérer les affaires par le nombre de personnes qu'elles impliquent, et indirectement renseigner sur la charge de travail associée ; une affaire avec trois auteurs devrait selon toute logique donner plus de travail de gestion qu'une affaire avec un seul auteur. À partir de là, l'indicateur construit pour un usage de suivi de parcours est récupéré pour un usage d'évaluation de l'activité.

Le ministère a souhaité assurer une cohérence entre décisionnel et statistique, à travers une qualification partagée de la donnée utilisée. Ce souhait sous-tend le lancement du projet de Système d'Information Décisionnel et statistique (SID) inscrit au schéma directeur du ministère pour les années 2009 à 2013, et dont le développement est encore en cours. Économiquement raisonnable, la construction de cette donnée « unique » en termes de sources et de concept renforce la perméabilité entre décisionnel et statistique. Elle fait ainsi naître une difficulté nouvelle : celle d'expliquer les différences entre les deux, voire de justifier leur coexistence. L'argument le plus fort devient alors celui de la qualité de la donnée, dans les processus de production et de diffusion statistique, et dans la permanence du chiffre qui n'est rendu public qu'une fois corrigé d'un certain nombre d'erreurs de mesure et au moment où l'alimentation des applicatifs de gestion est supposée suffisante.

UNE COLLECTE EN SILOS, MIROIR DE L'ORGANISATION DE LA JUSTICE

L'organisation de la justice repose sur un partage de compétences complexe entre ordres, juridictions et instance (**encadré**). L'exercice de la justice, organisée en silos métiers, est adossé à un système d'information également en silos. La gestion de chaque type de contentieux est en effet enregistrée dans un applicatif de gestion *ad hoc*, non conçu initialement pour communiquer avec les autres. La « statistique justice » produite par le Service Statistique du Ministère (SSM) de la justice exploite les données des juridictions judiciaires, en termes de décisions et d'accès au droit (à travers l'accès à l'aide juridictionnelle). Depuis peu, elle utilise également les données des services chargés de l'application des peines, de l'insertion et de la probation et prochainement les données des établissements pénitentiaires. Enregistrées pour des fins de gestion en silos, ces données sont collectées également en silos pour la statistique, concourant à produire une statistique morcelée.

Au-delà même du morcellement lié à la multiplicité des contentieux et à l'organisation judiciaire et ses outils, les données de gestion de la justice présentent une complexité qui reflète la variété des conditions de réception des différentes affaires et l'évolution dans le temps des affaires. Au civil, la justice est souvent directement saisie par le justiciable ou ses représentants. Au pénal, les affaires arrivent au parquet soit à partir d'un procès-verbal établi par la police, la gendarmerie ou une autre administration, soit d'une plainte ou d'une dénonciation directe. De plus en plus souvent, les affaires transmises par la police ou la gendarmerie le sont par échanges interapplicatifs, le parquet ayant possibilité d'importer directement certains attributs des affaires, renseignés par les services de sécurité intérieure. Peut ainsi être récupérée une indication de la nature d'infraction telle que définie par ces services avant même la qualification de l'affaire par le parquet (qui seule a valeur légale, sur le champ restreint des affaires poursuivies). L'information est ainsi plus riche, mais avec une qualité hétérogène sur une variable clé : à un stade précoce, la nature d'infraction attachée à une affaire pourra en effet être certifiée par la justice, notamment dans les cas de traitement en temps réel où l'officier de police judiciaire décide de cette variable en lien direct avec le parquet, mais aussi n'être qu'une indication provisoire. Autrement dit, ce qui à un stade avancé d'une affaire poursuivie constitue

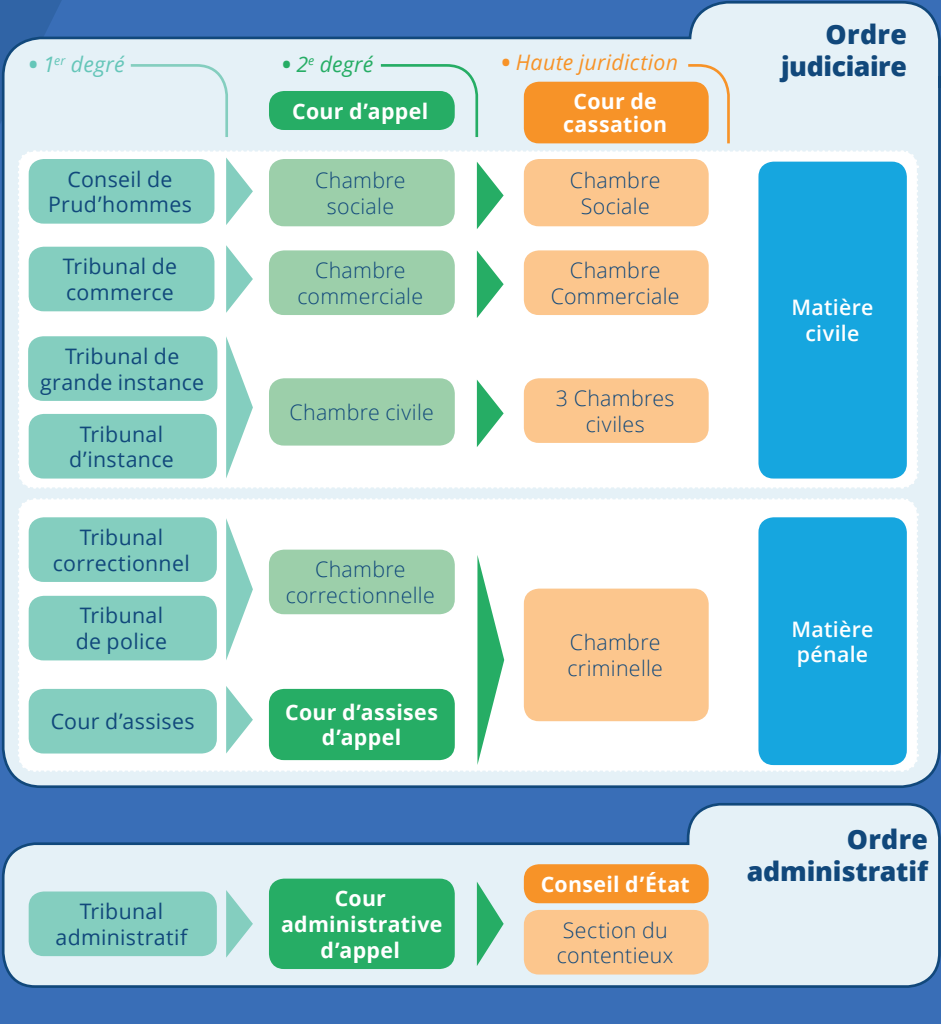
Encadré. L'organisation du système judiciaire

La Justice est organisée en deux ordres :

- un ordre judiciaire pour résoudre les conflits entre les personnes privées (individus, associations, entreprises, etc.) et sanctionner les infractions à la loi pénale ; en son sein, on distingue la matière civile, spécialisée (sociale ou commerciale) ou non, et la matière pénale (contraventions, délits et crimes) ;
- un ordre administratif pour les litiges entre un particulier et une personne publique (administration, collectivité territoriale, personne privée chargée d'une mission de service public) ou entre administrations.

Chaque ordre comporte trois degrés de juridictions.

Lorsqu'une ou plusieurs personnes ne sont pas satisfaites du premier jugement, elles peuvent faire appel. La chambre compétente de la Cour d'appel réexamine alors l'affaire. Enfin, la Cour de cassation vérifie que les lois ont été correctement appliquées par les tribunaux et les cours d'appel. Cette juridiction ne juge donc pas l'affaire une troisième fois. Il y a une Cour de cassation pour toute la République, car son rôle est de faire en sorte que la loi soit appliquée de la même manière sur tout le territoire.



une variable d'intérêt statistique majeur pour l'analyse du traitement par la justice de la délinquance n'est à un stade plus précoce qu'une variable de transition entre le traitement d'une affaire par les forces de sécurité et le traitement de la même affaire par la justice. La remontée plus fréquente de cette variable devrait toutefois permettre, à terme, de faciliter les comparaisons entre les statistiques de la délinquance enregistrée et les statistiques de la justice pénale... du moins sur le champ des affaires arrivées via les services de police et de gendarmerie.

La construction des données administratives de la justice, dont se nourrit la statistique justice, repose donc sur des dispositifs multiples, où la généalogie des affaires et des données est variable. Elle s'appuie néanmoins de plus en plus sur des flux applicatifs organisant la remontée d'extractions depuis les bases de gestion des juridictions vers le système d'information statistique.

POUR LE CIVIL, UNE STATISTIQUE CORSETÉE PAR LES LOGICIELS DE GESTION

Le système de flux applicatifs est particulièrement développé en matière de statistique sur la justice civile. Il couvre les affaires enregistrées dans le Répertoire Général Civil (RGC) des tribunaux d'instance, des tribunaux de grande instance, des cours d'appel, des CPH, mais aussi l'information sur les affaires relatives à la protection des majeurs et des mineurs, aux injonctions de payer, à l'acquisition de la nationalité française, à l'octroi de l'aide juridictionnelle... La justice des mineurs est également concernée. Jusqu'en octobre 2017, l'enregistrement des PACS (pactes civils de solidarité) par les greffes remontait par cette voie vers le service statistique. Au total, ce système de flux applicatifs collecte les données de 11 logiciels, en miroir de l'organisation judiciaire.

“ Cette transformation suppose de nombreux contrôles et redressements, notamment par confrontation avec les nomenclatures de gestion en vigueur. Contrairement à ce qu'on peut observer dans des enquêtes, où le respect des référentiels est un *a priori* de la collecte, la normalisation des données intervient ici *a posteriori* et suppose donc une intervention spécifique du statisticien. ”

La cartographie des flux se trouve par ailleurs compliquée par la composante locale des bases interrogées. Au-delà du millier de points de collecte impliqués dans les extractions statistiques mensuelles du RGC (avec plus de 5 000 fichiers réceptionnés et traités par mois, dont 85 % au cours de la première quinzaine du mois), la complexité du dispositif tient aussi à l'existence de référentiels parfois ouverts (locaux) et à des pratiques de greffes hétérogènes pour l'enregistrement et la codification des affaires et des procédures.

Seules des informations codées sont extraites, et stockées dans des tables primaires ; les documents numérisés relatifs aux différentes procédures ne sont pas exploités par la statistique. Du fait des adaptations locales qui ont pu être faites, les données en provenance des différentes juridictions n'ont pas forcément la même qualité. La première étape de la production statistique consiste alors en une normalisation des données, pour transformer les données brutes de gestion en des données qualifiées pour la statistique. Cette transformation suppose de nombreux contrôles et redressements, notamment par confrontation avec les nomenclatures de gestion en vigueur. Contrairement à ce qu'on peut observer dans des enquêtes, où le respect des référentiels est un *a priori* de la collecte, la normalisation des données intervient ici *a posteriori* et suppose donc une intervention spécifique du statisticien.

Au final, dans le champ du civil, si le système de flux applicatifs permet de disposer de données de détail sur les affaires enregistrées par les juridictions, leur champ reste cependant limité, prédéfini, si bien que la statistique est difficilement évolutive. L'évolution des flux statistiques suppose en effet une évolution des logiciels de gestion eux-mêmes et doit être arbitrée et programmée en concurrence avec les maintenances exigées par le métier. L'automatisation des remontées de flux envisagée dans le cadre de la refonte à l'horizon 2022 des applicatifs civils du ministère (projet Portalis) devrait toutefois s'accompagner d'une plus grande souplesse dans la définition du périmètre des extractions.

POUR LE PÉNAL, LES RESSOURCES D'UN ENTREPÔT DE DONNÉES CENTRALISÉ

Les progrès ainsi enclenchés dans le domaine de la statistique civile sont déjà partiellement opérants dans celui de la statistique pénale. Pendant de nombreuses années, la statistique pénale s'appuyait sur deux types de données : les compteurs d'activité renseignés par les juridictions à partir de leurs infocentres métiers dans des questionnaires *ad hoc*, et l'exploitation des données de détail du casier judiciaire national. La possibilité d'exploiter directement les données du logiciel Cassiopée, en substitution de la collecte par questionnaire, a changé en profondeur le travail du statisticien pénal. En lui donnant accès à un champ élargi de données de détail, collectées de façon centralisée selon des normes communes, elle a permis l'arrivée d'une nouvelle unité de compte (l'auteur-affaire, voir plus haut) et favorisé les études sur la filière pénale : Cassiopée couvre en effet la procédure depuis l'enregistrement de la plainte jusqu'au service de l'exécution des peines. D'abord expérimentale, l'exploitation de Cassiopée s'est rapidement inscrite dans le cadre plus général du projet de système d'information décisionnel et statistique (SID) du ministère. L'architecture visée pour le SID est celle d'un entrepôt de données central alimenté par l'ensemble des applications de gestion et proposant des magasins de données qui répondent aux différents besoins : pilotage stratégique, statistique, pilotage opérationnel, analyses par métier. Le SID n'est aujourd'hui que très partiellement en place : aux côtés de Cassiopée, présente dès 2015 dans l'entrepôt, ont été chargées en 2018 les données du logiciel de gestion Application des Peines, Probation et Insertion (APPI), et la réflexion sur le chargement prochain des données du logiciel de suivi des détenus (Genesis) ou d'autres sources relatives au pénal n'est pas terminée.

L'alimentation du SID repose sur une extraction de données brutes des applicatifs de gestion. Cette extraction se fait soit sur le périmètre global de l'applicatif concerné (c'est le cas pour APPI), soit sur une sélection de tables, dérivant d'une sélection opérée en amont (c'est le cas pour Cassiopée où seules les tables alimentant aussi l'infocentre décisionnel du même nom ont été reprises). Le volume des données chargées et manipulées se chiffre en millions d'enregistrements, les données étant historisées. Le processus de traitement prévoit l'existence de plusieurs couches de données de qualité variable pour la statistique : les données brutes sont d'abord complétées avec les informations issues du système de référence de la justice, des nomenclatures propres au SID et des tables de paramétrage du SID, puis corrigées, transformées et historisées dans des tables dites de « données qualifiées ». Ce sont ces tables qualifiées qui servent à la fois pour le décisionnel et pour la statistique ; la qualification des données est complexe, devant répondre aux besoins, en partie contradictoires, du décisionnel (indicateurs prédéfinis calculés sur des données vivantes) et de la statistique (données de référence mais dont l'exploitation doit rester souple). Des couches de publication, sans historisation et contenant des clés fonctionnelles, sont ensuite créées pour les besoins de la statistique. Ce nouveau mode de collecte des données administratives devrait donner le cadre de référence pour l'évolution

à venir de la statistique de la justice, permettant de limiter voire de supprimer certaines collectes par questionnaire, dans un horizon de cinq ans.

❶ DES ENQUÊTES POUR COMPLÉTER LES DONNÉES DE GESTION

Les données codées enregistrées dans les logiciels de gestion restent toutefois insuffisantes pour la connaissance et la compréhension d'un certain nombre de sujets. Seules les variables nécessaires à la gestion des affaires sont en effet enregistrées ; les variables « périphériques », quand elles sont prévues, sont généralement assez mal saisies dans la mesure où la qualité de l'information n'engage pas le devenir du dossier. Pour pallier les limites des données applicatives, la statistique s'appuie sur un autre type de dispositif : les « enquêtes décisions », dont la collecte permet de disposer d'informations détaillées présentes dans les jugements ou décisions mais pas dans les logiciels de gestion. Les thèmes abordés sont variés, depuis l'organisation de la résidence des enfants en cas de divorce jusqu'aux condamnations pour des infractions à caractère raciste ou la motivation des décisions d'emprisonnement. Ces enquêtes consistent à collecter auprès des juridictions un échantillon de décisions, sous format papier ; une grille d'analyse est définie à partir d'un sous-échantillon de ces décisions, en lien avec les services métiers à l'origine de la demande ; elle est utilisée pour la saisie des principales informations en une base structurée permettant l'exploitation statistique. Les échantillons sont généralement définis comme l'ensemble des décisions prises pendant une période donnée relativement au contentieux étudié ; la période est fixée de sorte à pouvoir disposer d'un échantillon suffisant, au minimum de 3 000 dossiers.

❷ D'UNE DONNÉE DE GESTION VIVANTE À UNE DONNÉE STATISTIQUE MILLÉSIMÉE

Les données de nature administrative mobilisées pour la construction de la statistique justice ont été organisées primitivement pour servir des besoins de gestion des affaires, pas ceux de la production statistique. Leur transformation en données statistiques impose la définition de concepts « juridico-statistiques », puis l'interprétation d'événements administratifs dans ces concepts afin de pouvoir décrire les procédures et les parcours de justiciables.

La première difficulté du statisticien est de poser des dates de référence pour l'observation et le dénombrement des données conjoncturelles. L'enregistrement des affaires est en effet réalisé en continu par les services des greffes (ou du casier judiciaire pour ce qui concerne ses données), à mesure de l'avancée du traitement des dossiers, mais aussi sous la double contrainte des urgences et des moyens disponibles. Les estimations de flux comme de stocks pour un trimestre ou une année donnée peuvent donc être affectées par une variation du rythme de saisie des dossiers en juridiction. S'agissant par exemple des données relatives aux affaires pénales traitées par les tribunaux correctionnels, on constate ainsi un écart de 3 % à 5 % entre les nombres d'auteurs poursuivables un trimestre donné selon qu'on regarde les indicateurs publiés en M+4 (provisoires) et ceux publiés en M+8 (« définitifs », même si les bases peuvent encore évoluer marginalement) ; les données publiées correspondent cependant, dans les deux cas, aux volumes dénombrés². Dans le cas des statistiques relatives aux condamnations, le choix est en revanche fait de corriger

2. On sait toutefois que la donnée provisoire sera nécessairement réévaluée à la hausse.

« Les estimations de nombre d'affaires nouvelles, terminées, et a fortiori des stocks, mais aussi les estimations de délais de procédures peuvent ainsi être entachées d'erreurs de mesure qu'il convient de réduire au maximum. »

de la sous-estimation anticipée : ainsi, on estime qu'environ 14 % des condamnations de 2016 n'ont pas encore été prises en compte en août 2017 au Casier judiciaire national, et les estimations provisoires publiées à l'automne 2017 sur les condamnations de 2016 inscrites au casier judiciaire national sont recalées pour tenir compte de ce sous-enregistrement. En 2016, la correction était moindre, de l'ordre de 12 %. Ces écarts devraient toutefois se réduire avec le développement des procédures

d'échanges inter-applicatifs, qui favorisent une entrée en continu des nouveaux dossiers dans les applicatifs de gestion.

Les écarts de dénombrement correspondant à des délais de saisie des dossiers peuvent par ailleurs porter soit sur l'enregistrement des affaires nouvelles, soit sur celui d'événements particuliers, dont la fin du traitement des dossiers. Les estimations de nombre d'affaires nouvelles, terminées, et a fortiori des stocks, mais aussi les estimations de délais de procédures peuvent ainsi être entachées d'erreurs de mesure qu'il convient de réduire au maximum. Le service statistique est de ce fait attentif au rythme de remontée des affaires. Les variations de volume inattendues donnent lieu à des échanges avec la direction des services judiciaires, voire les juridictions elles-mêmes, pour déterminer si la variation correspond à une variation d'activité ou à une variation du rythme de saisie (éventuellement du fait d'un simple changement dans le processus d'enregistrement des affaires...) qu'il sera difficile voire impossible de corriger. Dans le domaine pénal, on a ainsi constaté une croissance des transmissions inter-applicatives des affaires des services de police et de gendarmerie vers la justice ; celle-ci pourrait, au moins partiellement, expliquer la baisse constatée par ailleurs de la proportion des affaires non enregistrées, c'est-à-dire arrivées au parquet sous format papier et qui n'ont pas fait l'objet d'un enregistrement, car les infractions sont de faible gravité et l'auteur inconnu ou non identifiable.

DE L'ENREGISTREMENT DES FAITS À L'OBSERVATION DES COMPORTEMENTS

L'exemple précédent illustre aussi que, hors échanges inter-applicatifs, l'enregistrement même des faits n'est pas forcément automatique, et n'est réalisé que s'il est porteur de conséquences pour l'administration de la justice. L'analyse des séries historiques des condamnations, portées au casier judiciaire national, montre ainsi l'impact des amnisties présidentielles de 1988, 1995 et de 2002.

D'une part, ces dernières ont entraîné une baisse importante des condamnations prononcées au cours de ces années. Compte tenu des délais de transmission, une partie des condamnations prononcées l'année qui précède est d'autre part arrivée au Casier judiciaire après la promulgation de cette loi, et n'a pas été inscrite. Les données des années ultérieures sont enfin également touchées pour des faits commis avant la loi et sanctionnés l'année qui la suit. La méconnaissance de cette pratique peut conduire à une interprétation erronée des évolutions constatées statistiquement.

Une autre difficulté vient du fait que les applicatifs enregistrent des éléments d'un processus de gestion, qu'il faut réorganiser pour les besoins de la statistique. Une affaire est constituée d'une chaîne d'événements. En fonction du type de l'affaire, les événements ne seront pas les mêmes. En matière pénale, des concepts de filière, de phase et d'étape ont été créés dans le SID pour mieux suivre les différents déroulements possibles d'une affaire au Parquet. Une filière est constituée d'une succession de phases (saisine, orien-

tation, décision...), elles-mêmes rassemblant une ou plusieurs étapes, propres ou non à la filière considérée. La notion d'étape est proche de la notion d'événement, mais sans en être synonyme, seuls les événements « marquants » signalent une étape. Il s'agit alors de bien associer les différentes étapes possibles aux filières sur lesquelles on produit des statistiques, ce qui suppose un dialogue régulier avec les professionnels de la justice. Jusqu'en avril 2017, les premières orientations en Comparution sur Reconnaissance Préalable de Culpabilité (CRPC) n'étaient ainsi comptabilisées qu'à partir d'une seule étape, celle de requête en homologation de peine sur CRPC. Un travail méthodologique a conduit à élargir le champ des étapes associées à la CRPC et à réviser de façon très significative (+27 000, soit +46 %) le nombre de premières orientations en CRPC ; la révision a été documentée lors de la publication des résultats. En matière civile, de la même façon, la statistique s'appuie sur une nomenclature des activités civiles dont il s'agit de bien accompagner les évolutions pour conserver une statistique pertinente.

DES NOMENCLATURES DE GESTION AUX NOMENCLATURES STATISTIQUES

Comme dans tout système d'information, les nomenclatures et référentiels fournissent l'ancrage du système d'information de la justice, auquel est adossée la statistique de la justice. Ces référentiels présentent toutefois des catégories potentiellement en mouvement du fait de leur double dépendance aux codes juridiques et à l'organisation judiciaire. Les évolutions législatives successives peuvent en effet créer de nouvelles catégories d'affaires en matière civile, de nouvelles infractions pénales, ou en altérer la qualification. Car les nomenclatures de gestion de la justice peuvent être complexes, fortement adhérentes aux codes et donc au moment de l'observation (le contenu des codes pouvant changer), et surtout incomplètement structurées. La nomenclature des affaires civiles (NAC) épouse une structure en trois niveaux pour identifier les demandes dont sont saisies les juridictions civiles à titre principal, dans les différents secteurs des relations juridiques. En revanche, en matière pénale, les codes associés aux différentes natures d'infractions (NATINF) ne sont pas significatifs. La base de ces codes, qui recense la plupart des infractions pénales, évolue au gré des modifications législatives et réglementaires. Son usage pour la statistique nécessite de ce fait une analyse préalable, par un professionnel de la justice pénale, pour opérer les regroupements les plus opérants en fonction de la problématique traitée.

Les statistiques fondées sur les catégories juridiques des nomenclatures de gestion, dont les contours évoluent avec les textes, en sont nécessairement affectées. Tant qu'il s'agit de mesurer l'activité de la justice, cela conserve un sens ; à partir du moment où l'objet d'intérêt devient le comportement des justiciables, il importe de pouvoir tracer l'évolution des catégories, afin de ne pas imputer une variation statistique à un changement de comportement alors qu'elle ne traduirait qu'un glissement d'une catégorie vers une autre... voire l'apparition ou la disparition d'une nouvelle catégorie. En matière pénale, on relève ainsi une augmentation de 17 % entre 2004 et 2016 du nombre de condamnations prononcées en matière délictuelle à l'encontre des personnes majeures. Cette progression résulte principalement de la transformation, en 2004, de plusieurs contraventions en délits³. Le fait de dépendre de nomenclatures de gestion, vivantes, ne dédouane pas de disposer de nomenclatures statistiques plus stables. L'enjeu est de construire et maintenir le lien entre

3. Défaut de permis de conduire, défaut d'assurance automobile ou encore blessures involontaires par le conducteur avec incapacité totale de travail (ITT) inférieure à huit jours.

les deux types de nomenclatures. Afin d'assurer une continuité dans les regroupements – qui doivent néanmoins être mis à jour en continu, du fait de la production en continu des textes de loi –, des nomenclatures de diffusion statistique ont été construites au fil du temps. Dans les années 1980, le service statistique ministériel a ainsi élaboré la « NOMINF » pour l'exploitation des seules condamnations inscrites au casier judiciaire ; il s'agit d'une nomenclature en quatre positions, organisée selon la notion juridique de l'intérêt protégé qui est celle du code pénal, facilitant l'affectation des NATINF. En 1988, une nomenclature spécifique de diffusion a été créée, la DIFINF, avec un nombre de postes restreint (environ 200). Construite sur 5 caractères, elle fait apparaître en premier lieu la grande catégorie d'infraction (crime/délit/contravention) ; puis elle reprend la présentation par intérêts protégés, le domaine d'infraction et en dernier lieu les infractions les plus fréquentes auxquelles sont agrégées, dans un même poste, des infractions plus rares. Cette nomenclature est utilisée dans toutes les publications du service statistique ministériel.

L'usage de ces nomenclatures de diffusion reste cependant limité aux publications du service statistique ministériel. Or, une nouvelle nomenclature internationale, la Classification internationale des infractions à des fins statistiques (ICCS), est amenée à être de plus en plus mobilisée pour produire et comparer les données statistiques des différentes institutions de justice pénale des différents pays. L'ICCS peut être appliquée à toutes les formes de données pénales recueillies aux différents stades de la procédure (police, parquet, tribunaux et administration pénitentiaire) ou au cours des enquêtes de victimation. L'objectif aujourd'hui est d'adapter l'ICCS à la situation française dans une Nomenclature Française des Infractions (NFI), puis de mettre celle-ci en œuvre dans le système statistique public français. Un groupe de travail se penche sur le sujet depuis fin 2016 ; il est piloté par l'Insee et associe les services statistiques de la justice et de la sécurité intérieure, l'observatoire national de la délinquance et des réponses pénales (ONDRP) et le pôle d'évaluation des politiques pénales de la direction des affaires criminelles et des grâces.

UNE DÉMARCHE QUALITÉ INDISPENSABLE

La statistique de la justice présente un certain nombre de particularités qui, au final, sont sûrement partagées avec nombre de systèmes adossés à une remontée de données administratives : diversité des dispositifs de collecte et, avec elle, variété des unités statistiques, de la granularité des données, de leur périodicité ; adhérence des données collectées au cadre légal et réglementaire en vigueur et, avec elle, difficulté à produire des séries homogènes – à moins de rester sur une description de l'activité des services – voire de suivre des phénomènes qui, à un moment donné, peuvent sortir du champ du ministère de rattachement du service statistique. En définitive, leur exploitation nécessite une bonne connaissance des pratiques sociales et professionnelles dont elle entend rendre compte, pour reconstruire des concepts propres à être compris par l'ensemble des acteurs de la justice, justiciable compris. Elle nécessite aussi de disposer d'une documentation précise des processus de collecte et de production statistique, qui permet de tracer les évolutions de contexte et les changements méthodologiques opérés et permet *in fine*

“ Le service statistique du ministère de la Justice investit fortement [...] dans une démarche qualité d'amélioration des processus qui contribue aussi à distinguer la statistique du décisionnel. ”

les analyses intertemporelles de séries fortement heurtées dans leur construction. L'élaboration de cette documentation et sa mise à disposition font partie intégrante du travail du statisticien. Le service statistique du ministère de la Justice investit fortement dans cet aspect et s'inscrit ainsi dans une démarche qualité d'amélioration des processus qui contribue aussi à distinguer la statistique du décisionnel.

BÂTIR UN SYSTÈME D'OBSERVATION DU PARCOURS DES JUSTICIABLES...

Autant sinon plus que sur la connaissance de l'activité de la justice, la demande exprimée – tant par les décideurs que par les citoyens – porte aujourd'hui sur la compréhension des processus à l'œuvre. Les besoins de connaissance et d'évaluation se multiplient, par exemple en matière pénale sur la durée de mise à exécution des peines, l'impact des alternatives aux poursuites sur les retours devant la justice, en amont même de condamnations qualifiant la récidive ou la réitération en matière pénale ; ou en matière civile sur la médiation familiale ou les mesures de protection juridique des majeurs. Progresser sur ces sujets suppose de pouvoir suivre les justiciables à travers le temps ; les suivre à travers le temps suppose souvent d'approcher des informations collectées dans des applicatifs disjoints.

Or, ce rapprochement est aujourd'hui impossible sur le plan statistique, faute de disposer, à une étape ou une autre de la construction des fichiers statistiques, de données identifiantes communes à chacune des sources. Le service statistique n'a accès qu'à des sources expurgées des données nominatives : dans le domaine pénal, du fait de l'article 48-1 du code de procédure pénale qui réduit l'accès aux données par les statisticiens aux seules données non nominatives ; dans le domaine civil, du fait d'une habitude et des travaux à engager pour faire évoluer les extractions pour la statistique (voir supra). Une expérimentation conduite en 2016-2017 d'appariement de sources pénales (Cassiopée, APPI) à partir d'un identifiant anonyme, construit par les services informatiques du ministère de la Justice, en amont de la livraison au service statistique a été réalisée. Cet identifiant anonyme dérive des noms des personnes phonétisés et cryptés, la même clé ayant été utilisée dans les deux sources. L'expérimentation a mis en évidence la possibilité de réaliser l'appariement entre Cassiopée et APPI certes sans recours aux noms et prénoms phonétisés et cryptés, mais aussi que la présence de ces variables a un effet positif sur le taux et la qualité de cet appariement. Cette possibilité ne s'applique cependant qu'aux sources Cassiopée et APPI, qui contiennent l'une et l'autre des données relatives aux jugements. Les noms et prénoms demeurent ainsi indispensables pour apparier les données d'individus présentes dans plusieurs affaires, comme c'est le cas au sein de la source Cassiopée. Il serait par ailleurs indispensable qu'ils figurent en clair dans les fichiers transmis à la statistique, afin de pouvoir vérifier et contrôler la qualité des appariements réalisés, en confrontant les données appariées aux données primaires. Cela suppose toutefois un travail législatif pour modifier le code de procédure pénale.

... MIEUX CONNAÎTRE LES USAGERS DE LA JUSTICE...

Un second enjeu pour une meilleure compréhension des processus judiciaires serait de développer la connaissance des usagers de la justice, utile à la définition des politiques de prévention et/ou d'accès au droit (une meilleure connaissance des niveaux de revenu du public permettant par exemple d'évaluer l'impact tant social que budgétaire d'une réforme de l'aide juridictionnelle). Or, les fichiers statistiques tirés des applicatifs de gestion sont relativement pauvres en descripteurs socio-économiques (niveau de diplôme, niveau de revenu...). Le progrès passerait ici par le rapprochement avec des sources externes au ministère de la Justice. Mais ce rapprochement suppose également de disposer d'informations communes avec ces dernières donc, en l'absence notamment du NIR, de données nominatives permettant la création d'un « pseudo-NIR ».

Un troisième enjeu sera d'être en capacité de produire une statistique sur le règlement des contentieux, voire sur la réponse au besoin de justice exprimé par les citoyens, qui dépasse le cadre des activités prises en charge par les juridictions.

La loi de modernisation de la Justice du XXI^e siècle « J21 » publiée au Journal officiel du 19 novembre 2016 visait à « *permettre aux juges de se recentrer sur leur mission essentielle : trancher des litiges* », en favorisant notamment les modes alternatifs de règlement des conflits, en réservant le passage devant le juge aux affaires familiales, lors des divorces, aux seuls divorces contentieux et, pour les divorces par consentement mutuel, à ceux où un enfant a demandé à être entendu. Le projet de loi de programmation 2018-2022 et de réforme pour la justice⁴, débattu au Parlement à l'automne 2018, porte également l'objectif « de développer les modes alternatifs de résolution des différends afin que ne soient portées devant le juge que les affaires les plus contentieuses, pour lesquelles les parties n'ont pu trouver ensemble de solution amiable (...) ». Le mouvement ainsi amorcé de déjudiciarisation du règlement des contentieux va ainsi faire sortir du champ de l'administration des activités dont la connaissance ne repose aujourd'hui que sur des sources de données administratives.

Cette évolution pose la question de la capacité du système statistique public, dépendant aujourd'hui des données produites par l'administration, à accompagner ce mouvement de déjudiciarisation. On relève en 2017 une forte baisse du nombre de divorces par consentement mutuel dans l'activité des tribunaux de grande instance ; cette baisse correspond à la déjudiciarisation du divorce par consentement mutuel, mentionnée plus haut. Pour autant, les comportements des Français en termes de divorce ont-ils évolué ? Un peu plus d'un an après la mise en œuvre de J21, et en l'absence d'un système de remontée d'information sur les divorces enregistrés par les notaires, il est devenu impossible d'estimer simplement le nombre total de divorces prononcés en France, et *a fortiori* d'en décrire les populations concernées et les modalités d'organisation.

De plus en plus, l'intérêt du public se porte sur le service rendu quels qu'en soient ses acteurs, davantage que sur les acteurs qui rendent le service. Ce besoin s'exprime aussi bien au niveau national qu'international. Citons, par exemple, le suivi des objectifs de développement du millénaire de l'ONU, qui nécessite de produire des indicateurs sur l'accès de tous à la justice, et les travaux de l'OCDE qui considère l'accès à la justice comme un facteur d'inclusion sociale. L'émergence de ces besoins nouveaux, couplée à l'évolution du champ de compétence de l'administration, pose une limite à la statistique élaborée à partir de données administratives, alors que le champ de production du service par l'administration ne recouvre plus que partiellement le champ du service étudié.

4. L'étude d'impact est disponible sur le site www.justice.gouv.fr/art_pix/20180420_pjljustice_pjl_etude_impact.pdf

Belmokhtar Z., « *Tuteurs et curateurs des majeurs : des mandataires aux profils différents* », Infostat justice, n° 162, juin 2018.

Camus B., « *La statistique pénale, une statistique gestionnaire et une statistique criminelle : esquisse d'histoire* », communication pour le colloque Insee du 29 juin 2016.

Godin G., « *Première application du Plan de rénovation des statistiques de la Justice : la statistique judiciaire* », Courrier des statistiques, n° 1, pp.13-14, 1977.

Houllé R. et Vaney G., « *La mise à exécution des peines d'emprisonnement ferme aménageables avant toute incarcération* », Infostat justice, n° 166, septembre 2018.

Munoz-Perez B., « *Les statistiques judiciaires civiles, sous-produit du répertoire général des affaires civiles* », Droit et société, n° 25, pp.351-360, 1993.

Saglio A., « *La statistique Justice : vue d'ensemble* », Courrier des statistiques, n° 97, 2001.

SDSE, « *Références Statistiques Justice - Année 2016* », janvier 2018.

SDSE et SSM-SI, « *Pour une méthodologie d'analyse comparée des statistiques Sécurité et Justice : l'exemple des infractions liées aux stupéfiants* », Document d'études, décembre 2016.

LE LEGAL ENTITY IDENTIFIER

CONTEXTE INTERNATIONAL ET RÔLE DE L'INSEE

Pierrette Schuhl*

La volonté de créer un dispositif mondial d'identifiant unique des intervenants sur les marchés financiers s'est renforcée lors de la crise de 2007. Lancée par le G20 en 2011, l'initiative Legal Entity Identifier (LEI) s'est concrétisée par la mise en place d'un annuaire mondial pouvant servir de référentiel commun fin 2012-début 2013. Les codes LEI sont depuis lors utilisés au niveau international pour les obligations déclaratives des entités effectuant des transactions sur des instruments dérivés appliquées aux États-Unis ou en Europe. En France, l'Insee a été proposé dès 2013 par le ministère de l'Économie et des Finances pour être émetteur local d'identifiants LEI, en raison de son expérience dans la gestion de répertoires et d'identifiants. Il a été confirmé dans ce rôle en 2018 par son accréditation par la Global LEI Foundation (GLEIF). Cette nouvelle activité conforte ainsi sa position prépondérante dans les processus d'identification des entités économiques françaises. La seconde étape de l'initiative, lancée mi-2017, consiste à compléter les données d'identification des demandeurs de LEI par celles de leurs détenteurs direct et ultime. Ces données dites « de niveau 2 » sont porteuses d'améliorations significatives de la connaissance de la composition des groupes, de leur assise financière et potentiellement de leur stratégie.

UN MANDAT POLITIQUE FORT ISSU DE LA CRISE FINANCIÈRE

Le monde académique, les régulateurs et les superviseurs ainsi que l'industrie financière ont depuis longtemps débattu des attraits et des avantages d'une codification internationale permettant d'identifier les entités juridiques de manière unique. La crise de 2007 a été un accélérateur de l'initiative *Legal Entity Identifier* (LEI). La première pierre en a été posée au sommet du G20 à Cannes en novembre 2011. Les dirigeants du G20 ont ainsi confié au Conseil de stabilité financière (CSF, ou FSB en anglais) la mission de proposer des recommandations permettant de développer un LEI mondial.

« Nous soutenons la création d'un identifiant international pour les entités juridiques (*Legal Entity Identifier* – LEI), qui identifie de manière unique les contreparties aux transactions financières. Nous invitons le FSB à prendre l'initiative en aidant à coordonner les travaux de la communauté des régulateurs afin d'élaborer, d'ici au prochain sommet, des recommandations sur la mise en œuvre d'un cadre de gouvernance approprié, tenant compte de l'intérêt général, pour cet identifiant international. » (Déclaration finale du G20, 2011)

* Cheffe du département des répertoires, infrastructures et statistiques structurelles, Insee.

Dans un premier temps, un groupe d'experts composé sous l'égide du FSB et associant des représentants des banques centrales, des ministères des finances et autorités de contrôle a défini l'architecture cible, posé les grands principes de gouvernance et précisé le mode de financement. Ce groupe d'experts a établi un ensemble de 35 recommandations (Rapport FSB, 2012) qui serviront de principes fondateurs à l'initiative, approuvées par les dirigeants du G20 au sommet de Los Cabos de juin 2012.

« Nous approuvons les recommandations du FSB concernant le cadre d'élaboration d'un système d'identifiant international pour les entités juridiques (*Legal Entity Identifier* – LEI), qui identifie les contreparties aux transactions financières, avec un cadre de gouvernance mondial représentant l'intérêt public. Le système LEI sera lancé d'ici mars 2013. » (Déclaration du G20, juin 2012).

Dès cette étape, sont posées les bases qui vont permettre le développement de l'initiative LEI. Il est ainsi affirmé que l'association avec le secteur privé est essentielle à son succès ; le groupe d'experts a ainsi bénéficié de l'interaction avec les représentants de l'industrie financière sous forme d'ateliers de travail ou d'échanges interactifs continus. En établissant qu'un système d'identification mondial est un bien public, le groupe d'experts en a tiré des grands principes d'organisation de l'initiative : transparence et gratuité de l'accès à l'information, indépendance financière de l'agence centrale gérant le système la mettant ainsi à l'abri des risques de dépendance vis-à-vis de groupes d'intérêt, neutralité politique...

Le groupe d'experts a laissé la place à un groupe chargé de la mise en œuvre opérationnelle des recommandations qui, de juin 2012 à décembre 2012, a posé les fondations de l'initiative. En particulier, la construction juridique a reposé sur une Charte (novembre 2012) proposée à la signature des autorités de tous les pays intéressés, après approbation par le FSB (5^e rapport d'étape) et le G20 (Communiqué du G20, novembre 2012). Cette charte définissait l'architecture cible : une agence centrale (COU : *Central Operating Unit*), un réseau d'opérateurs locaux (LOU : *Local Operating Unit*), un comité de surveillance réglementaire (ROC : *Regulatory Oversight Committee*) et un comité d'évaluation et des standards (CES).

DES OBJECTIFS AMBITIEUX

Née des réflexions engagées après la crise des subprimes survenue en 2007, l'initiative LEI a affiché dès le lancement une grande ambition : une vocation mondiale, une mise en œuvre accélérée, un champ large (le LEI s'adresse potentiellement à toute entité juridique pouvant engager des transactions financières) et des usages potentiels multiples.

De fait, l'initiative a été marquée par ses origines et la prise en compte de besoins de surveillance macro et micro-prudentielles.

Améliorer l'identification des contreparties permet de mieux gérer la prise de risque par les acteurs et favorise ainsi une allocation plus efficace des risques dans le système financier mondial. Un cadre standardisé d'identification des intervenants sur les marchés financiers permet de disposer des bases pour la collecte d'informations sur les positions individuelles et ainsi de mieux appréhender l'exposition au risque tant au niveau individuel que de l'ensemble du système. La base de cette collecte est l'obligation qui est faite aux entités souhaitant intervenir sur les marchés financiers de demander un identifiant LEI. Par le passé, si les bénéfices potentiels d'un meilleur système d'identification étaient reconnus par tous, les progrès dans ce domaine avaient été traditionnellement limités par deux principaux facteurs :

- Premièrement, les systèmes déclaratifs existants, construits en réponse à des besoins essentiellement locaux, étaient largement intégrés dans les processus opérationnels. De fait, le changement devenait potentiellement très coûteux.

- Deuxièmement, les incitations à réformer ces systèmes locaux étaient trop limitées pour les acteurs individuels, illustrant ainsi un problème classique d'action collective et de premier pas à faire.

La crise financière survenue en 2007 dans les économies développées a permis une prise de conscience collective du besoin d'agir et de la nécessité pour les secteurs public et privé de travailler ensemble pour résoudre les problèmes d'action collective.

La mise en place d'un annuaire mondial pouvant servir de référentiel commun a été la première étape de l'initiative (fin 2012 - début 2013). La seconde étape, lancée mi-2017, consiste à utiliser ce socle pour mieux appréhender les relations entre les différentes entités et mettre en place un système d'identification des liens entre entités légales pour établir une cartographie des groupes d'entreprises et institutions financières. Ces données dites « de niveau 2 » sont porteuses d'amélioration significative de la connaissance de la composition des groupes, de leur assise financière et potentiellement de leur stratégie. Leur collecte a débuté mi-2017 auprès des entités détentrices ou demandeuses de LEI.

UN SYSTÈME FONDÉ SUR L'AUTO-ENREGISTREMENT DES DÉCLARANTS

La première étape de l'initiative consiste donc à attribuer un identifiant (LEI) qui fournit une identification unique et exclusive du déclarant. La stratégie suivie, très pragmatique, emprunte tout à la fois à l'approche juridique et à la démarche statistique. Annuaire et référentiel, la base de données des LEI offre un socle juridique et statistique qui repose sur un même besoin de définition de l'objet traité (notamment de la notion d'entité légale).

Le déclarant peut être une institution financière, une société non financière, un fonds et, plus généralement, toute entité avec personnalité juridique morale partie d'une transaction financière soumise à déclaration réglementaire. Les fonds sans personnalité juridique mais soumis à la réglementation européenne sur les produits dérivés (*European Market Infrastructure Regulation* ou EMIR, juillet 2012), aux obligations déclaratives définies par la loi américaine « Dodd Franck » ou à toute autre obligation déclarative ont également la possibilité d'obtenir un LEI et ainsi d'être autorisés à intervenir sur les marchés financiers.

Les personnes physiques ne sont pas autorisées à obtenir un LEI, notamment à des fins de protection des données personnelles (les informations renseignées par les demandeurs sont en effet en libre accès). Pour autant, la question reste posée pour les individus exerçant une activité professionnelle pour leur compte et plus généralement pour les déclinaisons juridiques multiples au niveau mondial de la notion d'entreprise individuelle.

Quel que soit le statut juridique de l'entité, le LEI attribué est un code alphanumérique de 20 caractères conforme à la norme d'identification ISO 17442 de l'organisation internationale de normalisation et ne comporte aucune information sur le déclarant ou l'entité **(voir Graphique 1)**.

Chaque code est associé à des données de référence identifiant l'entité légale (données dites « de niveau 1 ») telles que le nom de l'entité, l'adresse et le numéro d'enregistrement de l'entreprise dans le registre du commerce local (en France, chaque LEI de société est associé à un numéro d'enregistrement dans le répertoire des entreprises Sirene).

Le système est fondé sur l'auto-enregistrement, ce qui fait du déclarant le premier responsable de la qualité des données associées au LEI. Les opérateurs locaux attribuant les LEI ont cependant le devoir de valider les données déclarées (par exemple en s'appuyant sur le registre local) et les procédures de contrôle font l'objet de vérification par l'agence centrale.

Graphique 1. Le code LEI



Le système est financé, sur le principe de coûts recouverts, par les frais d'enregistrement et la cotisation annuelle versés aux LOU par les entités qui s'enregistrent. Les LOU reversent quant à eux une redevance à la GLEIF (voir ci-après) calculée sur le nombre des LEI qu'ils gèrent (17 USD par LEI en 2018). Le principe de coûts recouverts permet de contenir ou faire baisser progressivement les coûts pour les entités qui s'enregistrent.

UNE ORGANISATION DÉCENTRALISÉE, UNE COORDINATION EXIGEANTE

La mise en place d'un réseau décentralisé de LOU a nécessité une coordination et une harmonisation des pratiques, notamment en termes de qualité des données. Ceci a été assuré dans un premier temps par les autorités publiques en charge de la supervision du système via la définition de standards applicables dans le système. Depuis juin 2014, cette coordination est assurée par une fondation de droit suisse, la *Global LEI Foundation* (GLEIF) créée dans ce but. Cette dernière est également en charge de la vérification du respect des standards qu'elle a définis, ainsi que du processus d'accréditation des émetteurs de LEI.

Ainsi, chaque LOU émetteur de LEI doit prouver sa capacité à mettre quotidiennement à disposition du public (et des autres émetteurs) un fichier standardisé des codes LEI émis localement et des informations qui leur sont associées, afin notamment de permettre une centralisation des LEI émis et d'éviter l'émission de doublons. Il doit également assurer la diffusion de ces données en open data, afin d'en permettre une utilisation large.

Une telle procédure suppose en premier lieu de pouvoir gérer globalement la multiplicité des langues utilisées localement et chaque entité doit donc fournir une version romanisée du nom local. Cette translittération associée à un code unique est une première et forte assurance contre la duplication des codes.

L'attention portée à la qualité des données recensées est un des éléments qui s'inspire de la démarche statistique. Le principe de gratuité et de transparence de l'information offerte en open data permet d'intégrer les utilisateurs dans le système d'évaluation de la qualité des données, chacun pouvant « contester » auprès des opérateurs locaux les informations diffusées et demander une actualisation ou une rectification des données.

VERS LE SYSTÈME MONDIAL D'IDENTIFICATION D'ENTITÉS LÉGALES

Six ans après le sommet de Los Cabos qui marquait le début de la réflexion opérationnelle sur la mise en œuvre du LEI, l'état des lieux est plus qu'encourageant.

Début 2013, les autorités qui avaient signé la Charte se sont réunies pour la première fois à Toronto dans le cadre du Comité de surveillance réglementaire (ROC en anglais). Le Comité, qui est l'organe décisionnaire et qui est responsable de la gouvernance du système dans l'intérêt public, est ouvert à toutes les autorités signant la Charte. Mi-2013, le Comité accueillait 53 autorités membres et 19 observateurs. Il accueille aujourd'hui 72 autorités membres (dont 28 de l'Union européenne) et 18 observateurs.

L'utilisation du LEI s'est développée sur la base d'un engagement volontariste des autorités publiques. L'obtention d'un LEI par les acteurs opérant sur les marchés financiers a été rendue obligatoire d'abord pour les produits dérivés par la loi Dodd-Franck aux États-Unis dès février 2013 et la réglementation EMIR en Europe à partir de février 2014. Par ailleurs, l'Autorité bancaire européenne a imposé le LEI dans le contexte des obligations déclaratives bientôt suivie en cela par l'Autorité européenne des assurances et des pensions professionnelles. En France, l'instruction 2013-I-16 de l'Autorité de contrôle prudentiel et de résolution rend l'obtention d'un LEI obligatoire pour tous les organismes assujettis (établissements de crédit, entreprises d'investissement...). Et depuis janvier 2018, la révision de la directive européenne sur les marchés d'instruments financiers (MIF2) a élargi le champ des transactions sur instruments financiers nécessitant l'obtention d'un LEI. En dehors de l'Union européenne, de plus en plus de réglementations en imposent également l'obtention.

Les autorités membres du ROC ont posé les bases du succès de l'initiative en garantissant aux détenteurs d'un LEI la possibilité d'utiliser leur identifiant dans leur juridiction. Tout LEI, émis par un opérateur local agréé, peut être accepté dans le cadre des obligations déclaratives imposées par les autorités membres du ROC.

L'INSEE, UNE EXCEPTION DANS LE MONDE DES LOU

Début 2015, 31 LOU étaient déjà sponsorisés par les membres du ROC mais seuls 22 étaient opérationnels. En France, dès février 2013, l'Insee a été désigné par le ministère de l'Économie et des Finances pour être opérateur local pour les entités de droit français (c'est-à-dire pour la juridiction française) : fonds enregistrés par l'Autorité des marchés financiers (AMF) et personnes morales immatriculées au répertoire Sirene. Cette décision a été justifiée par :

- le savoir-faire de l'institut en matière de gestion de répertoires et d'identifiants (Sirene), et de ce fait par sa capacité à mettre rapidement en place l'infrastructure informatique nécessaire ;
- la proximité avec Sirene, pour vérifier l'existence et les informations d'identité des entités demandeuses ;
- le souci d'une participation active de l'administration française à la mise en place de cette initiative mondiale ;
- le souhait d'éviter un coût trop élevé pour les entités économiques françaises.

L'intérêt pour l'Insee était double :

- l'assurance de continuer à jouer un rôle prépondérant dans les processus d'identification des entités économiques françaises (et de garder la main sur la table de passage entre les identifiants LEI et SIREN) ;
- l'apport financier de cette activité pour le budget de l'Insee : il est en effet permis de dégager un excédent par cette activité, dans le contexte de la gratuité à venir des données Sirene.

En 2015, la GLEIF a mis en place une procédure d'accréditation des LOU en lieu et place de la sponsorship par les autorités financières. Gage d'une augmentation et d'une harmonisation de la qualité des données collectées, cette procédure a été imposée à tous les LOU déjà opérationnels ainsi qu'aux nouveaux organismes souhaitant devenir émetteurs de LEI. Mi-2018, la GLEIF a ainsi accrédité 32 LOU dont l'Insee¹ en France (en janvier 2018). L'Insee est un LOU, mais son statut d'administration ne joue aucunement dans l'activité qui en découle.

Il est à noter que, sauf exceptions, les instituts statistiques nationaux ne gèrent pas les répertoires d'entreprises. Les LOU sont majoritairement des opérateurs privés. Ils ne disposent pas d'un monopole sur une région donnée et chaque entité a la possibilité de choisir l'opérateur qui émettra ou maintiendra son LEI. Dans certains pays plusieurs opérateurs peuvent exister (Allemagne, États-Unis...). Néanmoins, tous les LOU ne sont pas accrédités pour opérer sur toutes les juridictions, soit par choix, soit par difficulté d'accès à certains registres de sociétés ou de fonds. À ce jour, seuls 10 LOU sont accrédités pour opérer sur la juridiction française. L'accréditation est renouvelée chaque année.

UN SYSTÈME OPÉRATIONNEL AU NIVEAU INTERNATIONAL

Fin septembre 2018, plus de 1 280 000 LEI ont été émis pour des entités situées dans 223 pays, dont près de 920 000 en Europe. Si l'on se concentre sur les LEI valides (c'est-à-dire émis, certifiés et actifs) au niveau mondial, on en dénombre 1 045 000 dont près de 766 000 en Europe. La liste de l'ensemble des LEI émis, actualisée quotidiennement, est disponible en accès libre sur le site de la GLEIF (www.gleif.org).

Les principaux pays développés et émergents sont associés à l'initiative, mais la répartition géographique des codes émis reste déséquilibrée. L'essentiel des codes émis correspond à des entités européennes et nord-américaines. Ceci est lié, en partie, à la répartition géographique des intervenants sur les marchés financiers mais également à des calendriers réglementaires différents. Le LEI s'inscrit naturellement dans un agenda européen d'intégration toujours plus poussée de la sphère financière, mais est plus dépendant aux États-Unis des stratégies des agences en charge de la régulation financière. L'accréditation des LOU chinois, coréen et japonais

“ Les principaux pays développés et émergents sont associés à l'initiative, mais la répartition géographique des codes émis reste déséquilibrée. L'essentiel des codes émis correspond à des entités européennes et nord-américaines. Ceci est lié, en partie, à la répartition géographique des intervenants sur les marchés financiers mais également à des calendriers réglementaires différents. ”

1. Cette accréditation a été confirmée à l'issue notamment d'un audit de l'Inspection générale de l'Insee mené en 2017, prévu dans l'accord cadre signé avec la GLEIF.

permet de rattraper petit à petit le retard pris dans le continent asiatique et souligne encore plus la faible pénétration du LEI en Amérique Latine. Il est notable que les centres financiers tels que le Luxembourg, les Îles Cayman, les Îles Vierges, Hong Kong ou Singapour hébergent un nombre significatif de LEI. Enfin, l'Afrique, traditionnellement à l'écart des initiatives dans le domaine financier, a pris sa part, certes modeste mais réelle, dans le développement du LEI. Quatre pays (Afrique du Sud, Angola, Nigeria et Île Maurice) ont ainsi signé la charte du ROC.

LA FRANCE EN 7^E POSITION DES ÉMISSIONS DE LEI

Fin septembre 2018, la France se situe en septième position en termes de LEI émis (environ 69 000 dont 54 000 valides), la très grande majorité émis par l'Insee (environ 62 800 dont 49 400 valides), ce qui situe l'Insee au huitième rang des émetteurs de LEI. Le système français, fondé sur un adossement du LOU au répertoire Sirene qui inclut le registre du commerce et au répertoire des fonds de l'AMF, a été longtemps un cas particulier. S'il permet d'atteindre le plus haut niveau de qualité (tout changement d'état civil est quasi immédiatement répercuté sur les LEI correspondants), il ne peut en revanche s'appliquer qu'aux entités légales de droit français enregistrées dans la base Sirene de l'Insee et aux fonds de droit français enregistrés ou déclarés à l'AMF. D'autres LOU ont choisi comme l'Insee de n'opérer que pour les entités de leurs propres juridictions. On en dénombre à ce jour neuf en plus de l'Insee. Il s'agit principalement de chambres de commerce ou d'émetteurs de l'identifiant unique national des entités de leur juridiction.

Du fait de la mise en œuvre de la directive MIF2 annoncée pour le 3 janvier 2018, le nombre de demandes de création et de renouvellement de LEI s'est fortement accru depuis l'été 2017 pour atteindre un pic en décembre 2017 et janvier 2018 puis décroître à partir de février 2018, tout en restant à un niveau plus élevé qu'auparavant (**voir Graphique 2**).

En effet, jusqu'en mai 2017, le nombre de nouveaux LEI attribués (créés) par l'Insee chaque mois était compris entre 200 et 350. En juin, juillet et août 2017, il se situait plutôt entre 400 et 500. Puis, à l'approche de la mise en œuvre effective de MIF2, il a augmenté très fortement, pour atteindre 5 700 en décembre 2017 et dépasser 6 400 en janvier 2018 avant de redescendre aux alentours de 3 000 en février et mars 2018, aux environs de 1 500 en juillet 2018 et se situer juste au-dessus de 800 en septembre 2018.

De ce fait, fin septembre 2018, le stock de LEI valides en gestion à l'Insee atteignait presque le triple de son niveau d'un an auparavant, avec 48 948 LEI valides dont 40 775 sociétés et 7 786 fonds.

Si assez naturellement, les fonds constituaient au démarrage de l'initiative la plus grosse population des LEI enregistrés par l'Insee (environ 40 % début 2015 et encore 38 % en juin 2017), ils ne représentent plus que 16 % des LEI émis suite à la mise en œuvre de la directive MIF2 (**Tableau 1**).

Le nombre de sociétés parmi les LEI valides émis par l'Insee a presque quadruplé entre fin juin 2017 et fin septembre 2018 et les caractéristiques des entités concernées ont évolué. Les sociétés n'appartenant pas à un groupe y ont un poids plus fort (54 % fin septembre 2018, contre 35 % fin juin 2017). Au regard de la catégorie d'entreprise définie par la loi de modernisation de l'économie (LEM) de 2008, les parts des Grandes entreprises (GE) et des Entreprises de taille intermédiaire (ETI) ont diminué au profit des Petites et moyennes entreprises (PME) et plus particulièrement au profit des Microentreprises (ME) (**Tableau 2**).

De manière similaire, au regard du chiffre d'affaires (CA) et des effectifs, c'est également la part des plus petites qui a crû. En effet, fin septembre 2018 plus de la moitié (63 %) des sociétés ayant un LEI valide ont moins de 1 million d'euros de CA alors qu'elles n'étaient

Graphique 2. Créations et renouvellements mensuels de LEI par l'Insee de janvier 2017 à septembre 2018

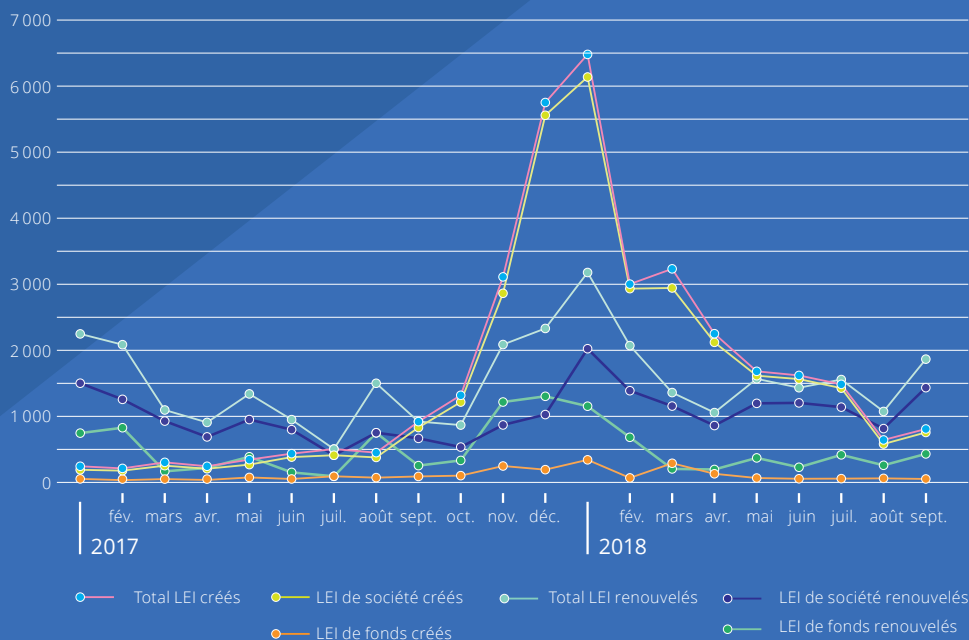


Tableau 1. Répartition par type d'entité

Nombre de LEI validés	Fonds	Sociétés	Total
Fin juin 2017	6 568	10 729	17 297
Répartition	38 %	62 %	100 %
Fin septembre 2018	7 786	41 162	48 948
Répartition	16 %	84 %	100 %

Tableau 2. Répartition des LEI des sociétés par catégorie d'entreprise au sens de la LME

Catégorie d'entreprise au sens de la LME	Fin sept 2018	Fin juin 2017
Grandes entreprises (GE)	6 %	17 %
Entreprises de taille intermédiaire (ETI)	13 %	22 %
Petites et moyennes entreprises (PME)	81 %	61 %
- Microentreprises (ME)	47 %	22 %
- Petites et moyennes entreprises (PME) hors ME	33 %	39 %
Total	100 %	100 %

que 30 % dans ce cas fin juin 2017. Et 62 % de celles enregistrées fin septembre 2018 avaient moins de 5 salariés au 31 décembre 2015 contre 48 % de celles enregistrées fin juin 2017. La mise en application de la réglementation MIF2 a également induit une modification de la répartition sectorielle des sociétés ayant un LEI valide. Si le secteur des activités financières et d'assurance reste en tête en maintenant sa part à 26 %, le secteur du commerce et de la réparation d'automobiles et de motocycles voit sa part diminuer de 23 % à 14 % entre juin 2017 et septembre 2018 et passe en 3^e position après celui des activités immobilières qui remonte en 2^e position avec une part de 16 % contre 14 % en juin 2017. Le secteur des activités spécialisées, scientifiques et techniques reste quant à lui en 4^e position en doublant presque sa part pour atteindre 13 %.

Fin septembre 2018, un peu moins de 1 % du nombre total de sociétés du secteur productif marchand avaient un LEI. En pondérant par les effectifs, ces sociétés regroupent 11 % des effectifs du secteur productif marchand, et en pondérant par le CA, 16 % du CA de l'ensemble du secteur productif marchand.

UNE ACTIVITÉ À PART ENTIÈRE POUR L'INSEE

L'Insee est engagé dans l'attribution de LEI depuis juillet 2013, avec l'ouverture du site <https://lei-france.insee.fr>. Depuis juillet 2014, ce site permet aux demandeurs de réaliser la certification et le renouvellement en ligne via identifiant et mot de passe, ainsi que de payer en ligne les frais associés. On y trouve également les fichiers de LEI émis par l'Insee en open data. Outre la gestion de répertoires (attribution d'identifiants sans doublon, validation des données déclarées...), l'activité d'émetteur de LEI comporte également des volets non négligeables de réponse à la demande et de facturation. L'ensemble nécessite aussi bien entendu des travaux de spécification et de développement informatique, ainsi que des relations très étroites avec la GLEIF et une forte réactivité dans la prise en compte des évolutions pilotées au niveau mondial.

Pour sa mise en œuvre, l'Insee s'appuie sur des ressources en interne et en externe. Conformément aux prescriptions de la GLEIF, l'Insee a mis en place en janvier 2013 une unité dédiée à son activité de LOU émetteur de LEI. Cette unité, située au sein de la division qui gère le répertoire Sirene, est rattachée à la Direction des statistiques d'entreprises. Elle gère l'application LEI, qui se compose du site internet et d'un ensemble de traitements permettant en particulier d'attribuer les LEI et de valider les données de référence par appariement au fichier Sirene géré par l'Insee et au fichier des fonds géré par l'AMF. Elle veille à la mise à jour de ces données en cas de modification des informations détenues dans les fichiers de référence. Elle est par ailleurs dotée des outils budgétaires et comptables nécessaires pour :

- établir le calcul de ses coûts, directs et indirects, afin de pouvoir en justifier la réalité auprès de la GLEIF ;
- élaborer des projections budgétaires glissantes à trois ans permettant de déterminer le barème de tarification de ses services (émission et renouvellement de LEI) pour l'année à venir².

En externe, l'Insee peut mobiliser les compétences et ressources de la Banque de France, de l'Autorité des marchés financiers (AMF) et de la Direction générale du Trésor, toutes trois autorités financières membres du ROC (et membre du comité exécutif du ROC pour ce qui est de la Banque de France).

2. Pour l'Insee, les frais d'inscription sont de 70 euros (50 euros pour la certification annuelle). Ces montants ont été fixés par arrêté du 28 juin 2013 (publié au JORF le 06 juillet 2013) modifié le 1^{er} juin 2018.

DES PERSPECTIVES DE DÉVELOPPEMENT INTÉRESSANT LES STATISTICIENS

Jusqu'à présent, l'initiative a été guidée par des préoccupations essentiellement micro et macroprudentielles. Dans les pays ayant déjà imposé l'usage du LEI, les entités identifiées sont autant de partisans d'une extension du LEI à d'autres usages. Par ailleurs, les autorités n'ayant pas encore intégré l'utilisation du LEI dans leur système de supervision, pourront s'appuyer sur les exemples réussis de mise en œuvre réglementaire (loi « Dodd-Frank », règlements EMIR, MIF2). Enfin, l'extension géographique progressive se poursuit et devrait accroître mécaniquement le nombre de LEI émis.

Bien qu'ambitieux et multiples, les objectifs évoqués par le groupe d'experts et repris dans la Charte ne font pour autant qu'une place limitée aux apports de l'initiative en termes de statistiques qui constituent un domaine naturel de développement.

En effet, les objectifs d'amélioration de la gestion du risque recouvrent largement les préoccupations statistiques. Il y a de fait, trois dimensions pour lesquelles l'adhérence entre les objectifs des statisticiens et de la gestion du risque est certaine :

- l'identification des entités ;
- la classification des instruments financiers ;
- l'identification des transactions.

Le LEI s'est attaché au démarrage essentiellement au premier aspect, et d'ores et déjà, il peut faciliter la production de statistiques en offrant un référentiel pour l'identification des contreparties dans des opérations financières. Complété par un code secteur de comptabilité nationale, il permet d'offrir une matrice de rattachement des entités qui pourrait être mise à disposition des déclarants statistiques (essentiellement les banques). Dans le même esprit, l'identification des agents non financiers dans les contreparties de la masse monétaire en serait facilitée et le recensement des opérations transfrontalières dans les balances des paiements plus aisé. De manière générale, le LEI peut servir de base à un référentiel statistique permettant un passage plus aisé des données individuelles aux statistiques agrégées. Dans un contexte où les statistiques puisent de manière croissante dans les données granulaires, le LEI apparaît comme un élément important de développement.

La deuxième étape de l'initiative LEI lancée mi-2017 (la déclaration des parents direct et ultime au sens de la consolidation comptable et des liaisons financières) devrait par ailleurs permettre la construction d'une cartographie des groupes d'entreprises par tous les acteurs publics et privés (et compléter celle réalisée par l'Insee à travers le répertoire LIFI des liaisons financières). Elle conforte ainsi le rôle du LEI en tant qu'outil statistique, objet d'étude et de recherche. D'ores et déjà, l'analyse de la dispersion géographique des LEI est porteuse d'enseignements sur l'intégration dans l'économie mondiale de certaines zones géographiques³. À l'avenir, l'analyse des structures des groupes et de leur évolution permettra d'élargir un domaine de recherche pour l'instant insuffisamment exploré.

3. Fin septembre 2018, sur les 1 280 000 LEI émis, 920 000 l'ont été dans l'Union européenne, 170 000 aux États-Unis, 86 000 dans les « centres financiers/paradis fiscaux » et 7 500 au Japon.

BIBLIOGRAPHIE

Sommet du G20, « *Déclaration finale* », 4 novembre 2011.

FSB, « *Report Global Legal Entity Identifier for Financial Markets* », 8 juin 2012.

Sommet du G20, « *Déclaration des chefs d'États et de gouvernement* », 18-19 juin 2012.

FSB, « *Charter of the Regulatory Oversight Committee for the Global Legal Entity Identifier (LEI) System* », 5 novembre 2012.

FSB, « *Fifth progress note on the Global LEI Initiative* », 11 janvier 2013.

Sommet du G20, « *Final Communiqué, Meeting of Finance Ministers and Central Bank Governors* », 4-5 novembre 2012.

Le Parlement européen et le Conseil de l'Union européenne, « *Règlement n° 648/2012 sur les produits dérivés de gré à gré, les contreparties centrales et les référentiels centraux* », 4 juillet 2012.

QU'ENTEND-ON PAR STATISTIQUE(S) PUBLIQUE(S)?

Michel Isnard*

Méfiez-vous des termes « statistique(s) » ou « statistique(s) publique(s) » ! Ils ont, selon le contexte, des significations très différentes et ne concernent pas les mêmes personnes ou organisations ou les mêmes indicateurs. En particulier, les textes établissent un lien plus ou moins étroit entre la production de statistiques et l'organisation mise en place pour cette production. Le quiz introductif vous permettra d'éprouver la géométrie variable de ces notions, parfois bien éloignée de la perception spontanée qu'en a le statisticien public. La définition que donnent de la statistique publique les textes français et européens est ainsi très différente. Et le domaine couvert par le secret statistique ne coïncide pas non plus exactement avec celui des statistiques publiques. Vous aurez, après la lecture de l'article, la solution du quiz.

DES EMPLOIS DU MOT «STATISTIQUE»

Le mot « statistique », généralement accolé à celui de « publique », recouvre des réalités bien différentes selon qu'il est employé au singulier ou au pluriel, comme un substantif ou un adjectif. Le terme de service statistique public fait référence à un aspect institutionnel alors que celui de « statistiques publiques » fait référence à des productions statistiques, c'est-à-dire à des chiffres, ayant un lien – parfois assez distendu comme on le verra – avec le premier terme. Quant au « secret statistique », si son rôle est fondamental dans la confiance qu'accordent les enquêtés au service statistique public, on en verra, là encore, des acceptions différentes.

Ces différentes notions peuvent se recouper et les distinguer n'est pas simple. Pour vous convaincre du caractère non trivial de cette distinction, commençons par un petit quiz, dont vous trouverez les réponses en fin d'article, après des explications plus détaillées sur les différents sens du mot « statistique ». Ne trichez pas !

LE QUIZ

Indiquer si chacune de ces propositions est vraie ou fausse.

Proposition n° 1

Le service statistique public ne produit que des statistiques publiques

☐ **Vrai ?** ou ☐ **Faux ?**

* Inspection générale, Insee.

Proposition n° 2

Les statistiques publiques ne sont produites que par le service statistique public

☐ **Vrai ?** ou ☐ **Faux ?**

Proposition n°3

Les statistiques européennes ne sont produites que par Eurostat et les autorités nationales statistiques

☐ **Vrai ?** ou ☐ **Faux ?**

Proposition n°4

Le secret statistique s'applique à toute donnée collectée directement ou indirectement à des fins de production de statistiques publiques

☐ **Vrai ?** ou ☐ **Faux ?**

Proposition n°5

Les statistiques publiques sont produites en toute indépendance professionnelle

☐ **Vrai ?** ou ☐ **Faux ?**

Proposition n°6

La libre communication de données couvertes par le secret statistique à des fins de statistique publique est possible à l'intérieur du service statistique public

☐ **Vrai ?** ou ☐ **Faux ?**

Proposition n°7

Il est possible de produire des statistiques publiques à partir de données non couvertes par le secret statistique

☐ **Vrai ?** ou ☐ **Faux ?**

Au travers de ces questions/réponses, nous avons souhaité mettre en évidence les recoupements mais aussi les différences entre ces différents concepts.

LE SERVICE STATISTIQUE PUBLIC

Parler en premier du « service statistique public » permet de fixer le cadre institutionnel de la statistique publique, mais aussi de commencer par un concept simple, puisque le service statistique public est clairement défini dans les textes réglementaires.

La France a souhaité organiser¹ son service statistique public autour de l'Insee dans le cadre de l'article 1^{er} de la loi du 7 juin 1951 qui définit le service statistique public. Il est constitué par l'Insee et les services statistiques ministériels dont la liste est fixée dans l'annexe du décret relatif à l'autorité de la statistique publique. Cette annexe est modifiable par un arrêté du ministre chargé de l'économie. Le principe est celui de la « décentralisation fonctionnelle » : au sein de quasiment chaque ministère, se trouve un (ou plusieurs)

1. D'autres organisations sont possibles : aux Pays-Bas, le service statistique public est constitué par le Centraal Bureau voor de Statistiek (CBS), tandis qu'en Allemagne, la structure fédérale de l'État a favorisé une décentralisation géographique.

service(s) statistique(s) ministériel(s) (**encadré 1**) dont le rôle est notamment de produire des statistiques dans le domaine de compétence du ministère. Cette organisation favorise les échanges entre les producteurs de statistiques et les décideurs publics, du fait de leur proximité. Elle facilite l'utilisation des déclarations ou des sources administratives en permettant aux statisticiens de mieux connaître la culture propre à chaque ministère et donc le contenu de chacune des sources.

Reste dans l'absolu un problème de coordination : il s'agit que le service statistique public (SSP) fournisse des données complètes (sans omission d'une partie de la société ou de l'économie), sans doublon (pour une gestion efficace des deniers publics) et cohérentes (pour que l'utilisateur puisse retrouver des sous-populations comparables dans les différentes

Encadré 1. Liste des services statistiques ministériels

Agriculture : Service de la statistique et de la prospective (SSP) et Services régionaux de l'information statistique et économique (Srise) - Ministère de l'agriculture et de l'alimentation

Collectivités locales : Département des études et des statistiques locales (DESL) - Ministère de l'intérieur et Ministère de la cohésion des territoires

Culture : Département des études, de la prospective et des statistiques (Deps) - Ministère de la culture

Défense : Observatoire économique de la défense (OED) - Ministère des armées

Développement durable : Service de la donnée et des études statistiques (SDES) - Ministère de la transition écologique et solidaire

Douanes : Département des statistiques et des études économiques (DSEE) - Ministère de l'action et des comptes publics

Éducation : Direction de l'évaluation, de la prospective et de la performance (Depp) - Ministère de l'éducation nationale et de la jeunesse et Ministère de l'enseignement supérieur, de la recherche et de l'innovation

Enseignement supérieur, recherche : Sous-direction des systèmes d'information et des études statistiques (Sies) - Ministère de l'enseignement supérieur, de la recherche et de l'innovation

Finances publiques : Bureau des études statistiques en matière fiscale (Bureau GF3C) - Direction générale des finances publiques - Ministère de l'action et des comptes publics

Fonction publique : Département des études, des statistiques et des systèmes d'Information (Dessi) - Ministère de l'action et des comptes publics

Immigration : Département des statistiques, des études et de la documentation (DSED) - Ministère de l'intérieur

Jeunesse et sports : Mission des études, de l'observation et des statistiques (Meos) - Ministère de l'éducation nationale et de la jeunesse

Justice : Sous-direction de la statistique et des études (S/D SE) - Ministère de la Justice

Santé et solidarités : Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) - Ministère des solidarités et de la santé

Sécurité intérieure : Service statistique ministériel de la sécurité intérieure (SSMSI) - Ministère de l'intérieur

Travail : Direction de l'animation de la recherche, des études et des statistiques (Dares) - Ministère du travail.

publications). Cette coordination repose sur des institutions spécifiques, le Conseil national de l'informatique statistique et le Comité du label de la statistique publique. Elle passe aussi par des outils comme les nomenclatures communes. Enfin, le règlement n° 223/2009 (voir infra) fait de l'autorité statistique nationale l'organe chargé de coordonner l'ensemble des activités menées au niveau national pour le développement, la production et la diffusion de statistiques européennes. Son article 5bis, introduit par une modification en 2015, confie notamment au directeur général de l'Insee le soin de mettre en place des lignes directrices de qualité s'appliquant à l'ensemble des autorités statistiques nationales.

Les membres du service statistique public bénéficient d'une disposition importante qui facilite énormément le travail du statisticien. Ce sont les seuls organismes à pouvoir mettre en œuvre l'article 7bis de la loi du 7 juin 1951. Cet article leur permet de se faire communiquer, à des fins d'élaboration de statistiques publiques, tout fichier de gestion d'une administration ou d'une personne privée gérant un service public, dès lors que le Conseil national de l'information statistique a été consulté et que la demande émane du ministre chargé de l'économie (en pratique du directeur général de l'Insee). Cette mesure, insérée dans la loi du 7 juin 1951 par la loi du 26 décembre 1986, a permis une exploitation large des données administratives et ainsi un allègement de la charge des répondants.

LES INSTITUTS NATIONAUX STATISTIQUES ET LES AUTRES AUTORITÉS NATIONALES STATISTIQUES

Alors qu'en France, l'embryon du service statistique public est en place dès la fin de la Seconde Guerre mondiale, il émerge plus tardivement au niveau européen. Il est d'abord défini à travers sa production. La notion de « statistiques européennes » apparaît dans l'article 285 du traité d'Amsterdam (1997) et est évoquée dans le cadre du règlement n° 322/97 du 17 février 1997 relatif à la statistique communautaire, sans être précisément définie.

C'est le règlement européen n° 223/2009 qui va en donner une définition précise dans son article 1^{er} : « *Les statistiques européennes sont des statistiques pertinentes nécessaires à l'accomplissement des activités de la Communauté. Les statistiques européennes sont régies par le programme statistique européen.* ». Le développement, la production et la diffusion de ces statistiques européennes sont mis en œuvre par les différentes entités constituant le système statistique européen, qui est également nommé et défini pour la première fois. Il comprend :

- Eurostat, une Direction générale de la Commission européenne en charge de la coordination du système à laquelle les traités européens ont confié le droit d'initiative sur l'ensemble des règlements statistiques ;
- Un institut national de statistiques dans chaque État membre, point de contact d'Eurostat et dont le dirigeant joue un rôle clé dans la production de statistiques européennes (voir l'article 5bis du règlement) ;
- les autres autorités nationales statistiques dont la liste est fixée par chaque État membre (**encadré 2**).

Chaque État membre est ainsi laissé libre de l'organisation de la composante nationale du service statistique européen, mais la Commission souhaite avoir un point de contact unique afin de favoriser les relations entre les différentes parties du système. Même si le chef de l'institut national de statistique ne peut être responsable de la statistique européenne dans l'État membre, il a un rôle de coordination de la partie nationale du service statistique européen.

Encadré 2. Liste des autorités nationales statistiques françaises

- Institut national de la statistique et des études économiques (**Insee**) ;
- Service de la donnée et des études statistiques (**SDES**) ;
- Direction de l'animation de la recherche, des études et des statistiques (**Dares**) ;
- Direction de la recherche, des études, de l'évaluation et des statistiques (**Drees**) ;
- Service de la statistique et de la prospective (**SSP**) ;
- Direction de l'évaluation, de la prospective et de la performance (**Depp**) ;
- Service des systèmes d'information et des études statistiques (**SIES**) ;
- Département des statistiques et des études économiques (**DSEE**) ;
- Département des statistiques, des études et de la documentation (**DSED**) ;
- Sous-Direction de la statistique et des études (**SDSE**) ;
- Le Service statistique ministériel de la sécurité intérieure (**SSMSI**) ;
- Le Centre d'accès sécurisé aux données du Genes* (**CASD**) ;
- Le Centre d'épidémiologie sur les causes médicales de décès de l'Inserm (**CepiDC**).

*Groupe des écoles nationales d'économie et statistique.

Comme on peut le constater, cette liste est différente de la liste des services statistiques ministériels. Un certain nombre de ces services ne sont pas des autorités nationales statistiques françaises, car ils ne développent, ne produisent, ni ne diffusent aucune statistique européenne. Deux petits nouveaux font en revanche leur apparition : le CASD, car il diffuse des statistiques européennes ; le CepiDC, car il est responsable du domaine « causes de décès » tant pour les statistiques françaises que pour les statistiques européennes concernant la France.

L'article 2 du règlement européen n° 223/2009 précise que le développement, la production et la diffusion de statistiques européennes obéissent à des principes généraux d'indépendance, d'impartialité, d'objectivité, de fiabilité, de secret statistique et de rapport « coût-efficacité ». Ces 6 principes généraux sont mis en œuvre dans le cadre d'un code de bonnes pratiques, adopté en 2005. Ils ont été repris de manière formelle dans le règlement n° 223/2009 sous la forme de 16 principes, qui sont déclinés en 84 indicateurs. Ce code joue un rôle structurant pour le fonctionnement de la statistique européenne en homogénéisant les pratiques entre les différents États membres. Une « évaluation par les pairs », réalisée tous les 6-7 ans, vise à évaluer la conformité des pratiques des autorités nationales statistiques vis-à-vis des principes du code de bonnes pratiques de la statistique européenne.

LES STATISTIQUES PUBLIQUES FRANÇAISES

Si la statistique publique européenne a d'abord été définie par sa production et plus tardivement par les institutions qui y concourent, c'est l'inverse pour la statistique publique française.

Ainsi, dans la version initiale de la loi du 7 juin 1951, les « statistiques publiques » ne sont pas définies et le terme n'est pas utilisé dans le texte. Leur définition est relativement récente : elle date de la loi du 4 août 2008 relative à la modernisation de l'économie, qui l'inscrit dans la loi de 1951.

Cette définition est très large. Elle comprend « *l'ensemble des productions issues d'une part des enquêtes statistiques et, d'autre part, de l'exploitation, à des fins d'information générale, de données collectées par des administrations, des organismes publics ou des organismes privés chargés d'une mission de service public* ».

Le législateur français a voulu cette définition large des statistiques publiques, qui va bien au-delà de la seule production du service statistique public. Le nombre de participants à une manifestation (sociale par exemple) à Paris est bien, dès lors qu'il est publié, une exploitation, à des fins d'information générale de données collectées par la préfecture de police. À ce titre, il s'agit bien d'une statistique publique.

Mais la version actuelle de la loi du 7 juin 1951 impose des contraintes aux statistiques publiques. Tout d'abord, elle oblige que le développement, la production et la diffusion de ces statistiques soient conformes aux dispositions du Code de Bonnes Pratiques de la statistique européenne qui ont été détaillées plus haut. Cette disposition demande aux statistiques publiques françaises le même niveau de qualité que les statistiques européennes.

Afin de veiller à cette conformité, la loi crée l'Autorité de la statistique publique (ASP). Celle-ci contrôle la conformité au code de bonnes pratiques des productions statistiques du service statistique public. Elle a par ailleurs mis en place une procédure spécifique, permettant de « labelliser » un certain nombre de productions de statistiques publiques (**encadré 3**) à la demande d'organismes n'appartenant pas au service statistique public. Elle est aidée dans cette tâche par le comité du label de la statistique publique.

Encadré 3. Liste des productions de statistiques publiques labellisées par l'Autorité de la statistique publique

- Statistique trimestrielle du prix du logement en province ;
- Données issues de la base nationale des causes de décès produites par le Centre d'épidémiologie sur les causes médicales de décès (CepiDC) ;
- Statistiques issues de la base des pensions du service des retraites de l'État (SRE) ;
- Statistiques de l'Agence centrale des organismes de sécurité sociale (Acos) ;
- Statistiques mensuelles de demandeurs d'emploi inscrits à Pôle emploi (DEFM) ;
- Données statistiques de la Caisse nationale de l'assurance maladie (Cnam) ;
- Statistiques de la Caisse nationale d'assurance vieillesse (Cnav) ;
- Statistiques des accidents de la route produites par l'Observatoire national interministériel de la sécurité routière (ONISR) ;
- Statistiques produites par la Mutualité sociale agricole (MSA) ;
- Statistiques de bénéficiaires de prestations légales individuelles sur le champ de la Caisse nationale d'allocation familiale (Cnaf) ;
- Statistique trimestrielle du prix du logement en Ile-de-France.

Le législateur n'a donc pas voulu lier les deux concepts de « service statistique public » et de « statistiques publiques », comme cela l'est dans le cadre de la statistique européenne. Il a préféré opter pour un cadre plus large et plus souple – incluant notamment toutes les administrations – tout en affichant des objectifs de qualité élevés.

LE SECRET STATISTIQUE

Dernier volet à être traité dans le cadre de cet article, le « secret statistique » est la pierre angulaire de la confiance qui est accordée par le public à la statistique publique. Comme vous allez le voir, le domaine de ce dernier concept ne recouvre pas exactement les domaines vus précédemment.

La loi du 7 juin 1951 interdit la communication de données permettant d'identifier l'unité ou toute autre personne en dehors de l'organisme producteur, que ces données soient des données individuelles ou des données agrégées. La loi restreint ce principe de secret statistique aux données issues d'enquêtes statistiques (article 6) et aux données administratives collectées dans le cadre de l'article 7bis. Ainsi, le secret statistique ne s'applique pas à l'ensemble des productions du service statistique public (qui peut mettre en œuvre des opérations administratives, comme la tenue des répertoires par exemple), ni des statistiques publiques (puisque qu'elles ne se limitent pas aux statistiques issues des enquêtes statistiques ou des fichiers collectés par l'article 7bis). Enfin, ce sujet déborde le service statistique public puisque des enquêtes statistiques peuvent être mises en œuvre en dehors de ce service.

Ces mêmes articles prévoient des dérogations à cette obligation de secret :

1. Au bout d'un délai de 75 ans pour les données « ménages » et de 25 ans pour les données « entreprises » (délai d'incommunicabilité), aucune restriction ne peut être opposée à la communication des données. C'est ainsi que les données du recensement de 1936 sont actuellement communicables.
2. Une procédure dérogatoire est mise en place : le demandeur dépose un dossier au comité du secret statistique qui donne un avis sur la communication. Cet avis est alors transmis au producteur des données qui donne ou non son accord. En cas d'accord du producteur, la décision de communication des données est transmise à la direction des Archives. La jurisprudence créée autour du comité du secret statistique fait qu'un avis favorable du comité du secret entraîne dans la quasi-totalité des cas une décision favorable de communication des Archives. La communication des données « ménages » ne peut être faite qu'à des fins de statistique publique ou de recherche scientifique ou historique. En 2017, le nombre total de dossiers traités par le comité du secret statistique est de 329 (dont environ 2/3 font l'objet d'une procédure électronique simplifiée). Ce nombre est en constante augmentation au cours de ces dernières années (287 en 2016).
3. Des dérogations existent aussi en cas de réquisitions à la demande d'un juge ainsi qu'en cas d'information du procureur de la République par un agent public qui acquiert, au cours de l'exercice de ses fonctions, la connaissance d'un crime ou d'un délit.

La communication aux SSM ou aux chercheurs de données individuelles ne peut qu'améliorer la qualité des statistiques publiques produites. Un équilibre doit donc être trouvé entre la contrainte du secret statistique et cette opportunité. En France, la jurisprudence du comité du secret statistique (consultation par voie électronique, dérogation permanente pour les services statistiques ministériels) permet de faciliter l'accès des statisticiens et des chercheurs aux données individuelles, tout en garantissant le respect du secret statistique mentionné dans la loi.

Cette prégnance du secret statistique n'a pas été entamée par les lois récentes sur l'ouverture des données publiques ou par le règlement général sur la protection des données. Même s'il ne s'impose que sur un domaine limité, le secret statistique reste une pierre angulaire incontournable du système statistique public.

Vous avez maintenant toutes les clés pour vérifier vos réponses au quiz **(encadré 4)** !

Encadré 4. Les réponses au quiz

Proposition n° 1

Le service statistique public ne produit que des statistiques publiques.

Faux

Aucune disposition légale, que ce soit au niveau européen ou au niveau national, n'oblige un membre du service à ne produire que des statistiques publiques. C'est notamment le cas à l'Insee (qui produit différents répertoires administratifs, comme le RNIPP ou SIRENE), à la Drees (qui produit notamment le répertoire national des établissements sanitaires et sociaux - Finess). Bien entendu, les données collectées à des fins statistiques ne peuvent en aucune manière servir à des fins de gestion administrative, mais l'autre sens est possible (voir l'utilisation du numéro SIREN dans les enquêtes statistiques auprès des entreprises).

Proposition n° 2

Les statistiques publiques ne sont produites que par le service statistique public.

Faux

Les deux sources des statistiques publiques (enquêtes statistiques et production à des fins d'informations générales) peuvent être mises en œuvre aussi bien par les membres du service public que par d'autres administrations. Le terme « administrations » doit être entendu au sens large, puisqu'il inclut des personnes morales de droit privé dans le cadre d'une mission de service public qui leur serait confiée.

Proposition n° 3

Les statistiques européennes ne sont produites que par Eurostat et les autorités nationales statistiques.

Vrai

Le règlement n° 223/2009 du Parlement européen et du Conseil du 11 mars 2009 relatif aux statistiques européennes prévoit explicitement dans son article 4 que l'institut national de la statistique et les autres autorités nationales sont responsables du développement, de la production et de la diffusion des statistiques européennes. Comme prescrit par ce règlement, la liste de ces organismes est visible sur le site Internet d'Eurostat à l'adresse suivante : http://ec.europa.eu/eurostat/documents/747709/753176/20180314_List_ONAs_CZ/feb336f7-f8e6-483b-895d-9da7951f55d1. On peut noter que pour la France on y retrouve l'Insee, 10 services statistiques ministériels auxquels viennent s'ajouter le centre d'accès sécurisé aux données du Genes et le centre d'épidémiologie sur les causes médicales de décès de l'Institut national de la santé et de la recherche médicale.

Proposition n° 4

Le secret statistique s'applique à toute donnée collectée directement ou indirectement à des fins de production de statistiques publiques.

Faux

Comme indiqué ci-dessus, l'application du secret statistique est limitée aux données issues d'une part des enquêtes statistiques et d'autre part des données collectées par le service statistique public en application de l'article 7bis de la loi du 7 juin 1951. Or, les statistiques publiques peuvent être produites à partir de n'importe quelle autre source. Ce qui implique que, dans ce cas, les données ne sont pas couvertes par le secret statistique. Elles restent néanmoins éventuellement couvertes par d'autres secrets : secret fiscal, secret défense, secret médical, secret professionnel...

Proposition n° 5

Les statistiques publiques sont produites en toute indépendance professionnelle.

Vrai

Il s'agit des conséquences de l'article 1^{er} de la loi du 7 juin 1951.

Proposition n° 6

La libre communication de données couvertes par le secret statistique à des fins de statistique publique est possible à l'intérieur du service statistique public.

Faux

Les deux articles concernant le secret statistique décrivent la procédure qu'il est nécessaire de mettre en œuvre pour communiquer des données couvertes par le secret statistique en dehors de l'organisme dépositaire : il s'agit d'un avis du comité du secret statistique, d'un accord du service producteur et d'une décision de l'administration des Archives. Cette procédure est aussi valable pour les membres du service statistique public. Toutefois, en raison de la fréquence des échanges entre ces membres et de leur nécessaire pérennité pour assurer une production continue dans le temps, le comité du secret statistique a mis en place, avec l'accord de l'administration des Archives, une autorisation permanente à une source par un service statistique ministériel.

Proposition n° 7

Il est possible de produire des statistiques publiques à partir de données non couvertes par le secret statistique.

Vrai

Seules les données collectées par une enquête statistique et celles communiquées au titre de l'article 7bis à l'Insee ou à un service statistique ministériel sont couvertes par le secret statistique. Or, le domaine des données utilisées pour produire des statistiques publiques est beaucoup plus large. Voir la proposition 4.



Présentation de la revue

L'Insee relance le *Courrier des statistiques*, une revue emblématique qui a accompagné des générations de statisticiens pour lesquels il constituait une référence. L'ambition de la revue est d'aborder, à travers la variété des approches et des sujets traités, l'ensemble des grandes problématiques du système statistique public. Elle vise également à illustrer notre capacité collective à évoluer et innover pour coller aux besoins de notre temps.

Dans le premier article de ce numéro, le Directeur général de l'Insee s'interroge sur la juste distance de la statistique publique à l'exécutif, compatible avec son indépendance. Un dossier de quatre articles éclaire la problématique de l'utilisation des sources administratives en statistique dans différents domaines et selon divers points de vue. Puis on change de registre avec la mise en place du dispositif mondial d'identifiant unique des intervenants sur les marchés financiers (*Legal Entity Identifier*, LEI), et le rôle qu'y joue l'Insee. Enfin, le dernier article présente de façon pédagogique la notion de statistique publique sous ses différentes facettes, tant au niveau français qu'européen.

ISSN 2107-0903
ISBN 978-2-11-139689-0



9 782111 396890

