

## La statistique dans la cité n° 25 - juin 2021

*Lettre bimestrielle du groupe « Statistique et enjeux publics »*

### Sommaire du n°25 – juin 2021

Éditorial

Méthodes :

- *Le paradoxe des droits des personnes au regard des bases de données de santé et des bases pseudonymisées en général*

Outils :

- *Une nouvelle enquête pour mieux connaître les personnes sans domicile ?*
- *Croisement de fichiers grâce au NIR... sans le NIR !*

Vie des institutions :

- *Trois règlements européens sur les data en cours de préparation*
- *Menaces sur l'intégrité du système statistique brésilien*

Anniversaire :

- *Il y a trente ans, le 18 juin 1991, les Principes fondamentaux de la statistique officielle étaient adoptés ...*

Courrier des lecteurs :

- *À propos de l'article « Marges d'erreur et sondages pour l'élection présidentielle de 2022 » paru dans la dernière infolettre*
- *On nous écrit aussi du Québec*

Notes de lecture :

- *Lu dans Variances : « Faire parler les comptes nationaux : l'importance du vocabulaire »*
- *Lu dans l'édition numérique du magazine Le Particulier : « Devez-vous obligatoirement répondre aux enquêtes de l'Insee ? »*
- *Vu sur France 2 l'émission Cash Investigation : « Vos données personnelles valent de l'or »*

Annonces :

- *11e Colloque francophone sur les sondages - 6 au 8 octobre 2021*
- *Prochaines Journées de Méthodologie Statistique de l'Insee en 2022*
- *Les comptes de la Nation 2020*

Vie du groupe

**Tous les numéros de l'infolettre sont téléchargeables sur le site du groupe SEP ([cliquez ICI](#))**

### Éditorial

*La statistique dans la cité* s'apprête à prendre ses congés d'été et nous vous retrouverons à la mi-octobre avec son n° 26. Les deux mois qui viennent de s'écouler depuis la parution du n° 24 ont été encore marqués, directement ou indirectement, par la pandémie de la Covid19. Mais, sans doute en partie à cause de la publication fréquente d'indicateurs sur la situation sanitaire, les pouvoirs publics et les media ont aussi remis à l'ordre du jour les questions de sécurité des données. Ce numéro illustrera certains aspects de ces débats et de ces questions.

On trouvera notamment dans cette livraison de *La statistique dans la cité* un article de méthodes qui s'interroge sur le paradoxe des droits des personnes au regard des bases de données de santé et des

bases pseudonymisées en général. Un autre article présente une avancée significative permettant de combiner résultats d'enquêtes et données administratives avec une sécurité accrue grâce à l'utilisation d'identifiants non signifiants. Il est également fait part des propositions récentes de règlements de la Commission européenne destinés notamment à encadrer les activités des géants du Web, les Gafam. Enfin, on a « vu pour vous » sur France 2 le jeudi 20 mai l'émission Cash Investigation sur le thème : *Nos données personnelles valent de l'or ...* et un article à propos de cette émission dira les enseignements que les statisticiens et leurs utilisateurs peuvent en tirer.

Ce numéro contient également un article présentant différentes investigations destinées à mieux connaître les personnes sans domicile. N'oublions pas la célébration d'anniversaires (il y en a souvent dans notre domaine !) avec, il y a trente ans, le 18 juin 1991, l'adoption des *Principes fondamentaux de la statistique officielle* par la Commission économique pour l'Europe des Nations-Unies au cours de sa réunion annuelle au niveau ministériel. Un courrier des lecteurs, des notes de lecture et des annonces diverses complètent ce numéro. Faites-nous part de vos commentaires avant de partir en vacances !

Enfin, *La Statistique dans la Cité*, soucieuse comme toujours de défendre l'intégrité de la statistique publique et l'indépendance des statisticiens, a reçu quelques nouvelles, à la fois alarmantes et rassurantes, du Brésil dont il est fait part au chapitre '*Vie des institutions*'.

**Pour nous écrire :** [sep@sfds.asso.fr](mailto:sep@sfds.asso.fr)

## Méthodes

### **Le paradoxe des droits des personnes au regard des bases de données de santé et des bases pseudonymisées en général**

Nous avons signalé dans l'infolettre n°23 que les organismes gérant des bases de données de santé ayant vocation à être intégrées au Système national des données de santé (SNDS) ne peuvent pas en principe s'y opposer même si en pratique la plateforme <sup>(1)</sup> des données de santé a intérêt à les convaincre plutôt qu'à les contraindre.

Mais quid du droit d'opposition des personnes physiques concernées par les données du SNDS et de leurs autres droits (accès, rectification...) reconnus par le règlement général pour la protection des données <sup>(2)</sup> ?

La loi sur les données de santé n'y a créé qu'une exception : on ne peut pas refuser que ses données soient versées au SNDS et utilisées pour certaines missions de service public d'organismes mentionnés dans un décret ; mais on peut s'opposer aux autres usages (c'est un droit de limitation des usages au sens du RGPD). Les droits d'accès et de rectification n'étant même pas mentionnés dans la loi, il semble aller de soi qu'ils s'appliquent. Et pourtant dans la pratique il est beaucoup plus simple et sûr de mettre en œuvre les droits d'opposition, de limitation ou d'effacement que les droits d'accès ou de rectification dans une base pseudonymisée.

Comme c'est contre-intuitif, il faut rappeler :

- que le SNDS réutilise des données administratives ou médicales à des fins de connaissance (statistiques, recherche...)
- qu'il n'enregistre pas l'identité (nom-prénoms, adresse, NIR...) mais qu'il a besoin de pseudonymes (numéros obtenus ici par un procédé irréversible à partir du NIR de chacun) pour pouvoir rattacher des données de différentes sources à la même personne.

Si un patient déclare à sa caisse ou à son hôpital qu'il ne veut pas que ses données servent à des recherches, un code sera inscrit dans son dossier et se retrouvera dans le dossier pseudonymisé ; le SNDS saura ainsi reconnaître et exécuter les instructions de non utilisation.

En revanche, si la personne veut accéder aux données qui la concernent, voire les faire rectifier, on devra réidentifier ces données alors qu'elles sont conservées sous un pseudonyme irréversible ! C'est faisable mais il faut pour cela révéler les données à des agents affectés à ce travail. Non seulement cela fragilise le secret mais la personne n'y a aucun intérêt puisque ces données, exactes ou non, ne pourraient être utilisées à son encontre que si elles étaient réidentifiées. Donc si on craint un mauvais usage des données, la solution est le droit d'opposition, de limitation ou d'effacement, non les droits d'accès et de rectification !

Ce paradoxe concerne le SNDS et toutes les bases de données pseudonymisées : il est plus facile et surtout plus utile et moins risqué pour la vie privée des personnes concernées d'utiliser les droits

d'opposition, limitation ou effacement que les droits d'accès et de rectification. Or l'article 11 du RGPD permet justement aux gestionnaires des bases de ne pas y conserver ou mettre en œuvre des moyens de réidentification sans autre utilité que le respect d'une disposition du RGPD lui-même. On informera les intéressés que les responsables de la base ne peuvent pas retrouver leurs données : aucune nouvelle loi n'est nécessaire.

(1) Cette plateforme est souvent désignée par le sigle anglais HDH (Health Data Hub)

(2) RGPD, règlement européen du 27 avril 2016 (chapitre III)

## **Outils**

### **Une nouvelle enquête pour mieux connaître les personnes sans domicile ?**

Lors du 119e Café <sup>(3)</sup> de la Statistique le 9 avril 2019, notre invité Stéphane Legleye (Insee) avait évoqué de manière très prudente l'hypothèse d'une reconduction de l'enquête « Sans domicile » dans les années 2024/2025 L'enquête « Sans domicile » Insee / Ined de 2012, s'appuyant sur une méthodologie rigoureuse, dénombrait 143 000 personnes sans domicile en France métropolitaine, en incluant les personnes accueillies dans les centres de demandeurs d'asile, soit 50 % de plus que lors de la première édition de 2001.

L'article « *Connaître les personnes sans domicile est encore plus important que les dénombrer* » publié le 13 avril 2021 sur le [blog de l'insee](#) nous apprend que la statistique publique mène des réflexions en vue d'une possible nouvelle enquête nationale, qui s'articulerait avec les différents dispositifs d'observation existants.

Le terme sans domicile recouvre à la fois les personnes sans-abri, c'est-à-dire celles dormant dans la rue ou dans un autre lieu non prévu pour l'habitation (hall d'immeuble, parking, jardin public, gare...), et celles mises à l'abri dans le cadre d'un dispositif d'hébergement (en centre collectif, à l'hôtel ou dans un logement ordinaire).

S'agissant des seules personnes sans abri, un dénombrement est effectué tous les cinq ans dans le cadre du recensement de la population (en 2016 27 000 personnes étaient dans cette situation en France). Mais, comme l'indique l'Insee, ce résultat n'est pas très fiable car le comptage dépend fortement des moyens et de l'implication des communes qui ne disposent que de deux jours pour réaliser cette opération.

Depuis quelques années, de nombreuses collectivités ont donc organisé leur propre recensement des personnes sans abri (Cf. n° 18 de février 2020 de *La statistique dans la cité*). Afin de faciliter la comparabilité entre les différentes villes et sur la base des échanges de pratiques déjà réalisées, l'Insee a produit avec la Dihal <sup>(4)</sup> un guide méthodologique afin de fournir un cadre commun aux collectivités souhaitant réaliser ce type d'opérations. Par ailleurs, dans son dernier rapport sur le mal logement, la Fondation Abbé Pierre, estime que cette population est d'environ 300 000 personnes en 2020 <sup>(5)</sup>.

Au-delà du dénombrement, qui sont les individus sans-abri et pourquoi certains d'entre eux refusent-ils les propositions d'hébergement ? Des équipes de maraude <sup>(6)</sup> en janvier 2021 ont mené une enquête de terrain lors d'une nuit de janvier 2021 pour répondre à ces questions (Cf. Le Monde daté du vendredi 15 mai).

L'article du blog de l'Insee note que « le dénombrement des personnes sans domicile suppose de traiter de nombreux problèmes méthodologiques, mais l'objectif est bien de connaître les profils des personnes sans domicile et leurs parcours de vie, afin de servir les politiques publiques en direction de cette population spécifique ; il convient notamment de dépasser l'idée fautive d'une population homogène et de proposer aux personnes sans domicile un accompagnement et une aide adaptés ».

(3) Le thème de ce café était *Objectif Zéro SDF*

(4) Dihal : *Délégation Interministérielle à l'Hébergement et à l'Accès au Logement*

(5) Ce rapport s'appuie notamment sur le « rapport sur l'hébergement et le logement des personnes sans domicile pendant la crise sanitaire du printemps 2020 » de la *Cour des Comptes* ([voir ICI](#))

(6) *Fédération des acteurs de la solidarité (FAS) et Fédération nationale des SAMU sociaux (FMSS)*

## **Croisement de fichiers grâce au NIR... sans le NIR !**

Les statisticiens publics et les chercheurs souhaitent souvent combiner des fichiers de données individuelles, pour les enrichir mutuellement au niveau individuel et permettre des études croisant deux thématiques. On peut ainsi étudier l'insertion des apprentis en rapprochant des données de formation et des données sur l'emploi, étudier les revenus en rapprochant des données d'emploi obtenues par l'enquête emploi et des données de revenus issues des déclarations fiscales et les prestations sociales issues des données des organismes sociaux.

Pour cela, ils ont besoin d'un même identifiant dans chacun des fichiers, ou d'informations qui permettent d'identifier un individu de façon unique (comme les nom, prénom, date et lieu de naissance) pour appairer les données individuelles correspondantes.

Certains fichiers administratifs possèdent un identifiant signifiant, le [NIR](#) <sup>(7)</sup>, connu aussi sous le nom de numéro de sécurité sociale. Son utilisation est très fortement encadrée. Jusqu'en 2016, pour tout traitement portant sur des données parmi lesquelles figure le NIR mis en œuvre par une personne morale de droit public ou une personne morale de droit privé gérant un service public, il était nécessaire de passer par un décret en Conseil d'État. Dans la pratique, cette exigence s'est révélée insurmontable pour les organismes universitaires, puisque très rares sont ceux qui ont pu obtenir qu'un ministre prenne l'initiative de porter un décret en Conseil d'État pour permettre un appariement dans le cadre d'un projet de recherche. De même, pour la statistique publique, la lourdeur des démarches a fortement entravé le développement de travaux nécessitant la fusion de fichiers.

En 2016, la [loi pour une République numérique](#) a levé en partie ces difficultés en autorisant les appariements de fichiers pour un usage statistique fondés sur un code non signifiant, obtenu par transformation non réversible du NIR <sup>(8)</sup> à l'aide d'une fonction cryptographique. L'utilisation de ce code non signifiant permet d'éviter de faire cohabiter dans un même fichier le NIR et les données à analyser, renforçant ainsi la protection de la confidentialité.

- Pour les statisticiens publics, le même Code Statistique Non Signifiant (CSNS) peut être utilisé pour l'ensemble des travaux réalisés par le service statistique public (en dehors des « données sensibles », au sens de la loi « Informatique et libertés ») et devra être renouvelé au moins tous les dix ans. Pour ces traitements, une description de traitement dans le registre des traitements, accompagnée d'une analyse d'impact et assortie d'une information publique sur Internet conformément au RGPD suffit.
- Pour les chercheurs universitaires, le Code de Recherche Non Signifiant (CRNS) est spécifique à chaque projet de recherche.

Ces nouvelles dispositions apportent aux chercheurs de nouvelles perspectives de recherches. Pour la statistique publique, cette nouvelle façon de collecter les données en combinant des enquêtes et des [données administratives](#) permet d'alléger la tâche du répondant aux enquêtes ou d'éviter de recourir à une enquête.

L'Insee met en œuvre dès cette année les dispositions prévues par la loi et dont les textes d'application datent de [2016 pour le décret](#) et [2020 pour l'arrêté](#). Il s'agit de développer une offre de services facilitant les appariements de données individuelles au sein du service statistique public. Si le responsable de traitement ne transmet pas le NIR à l'Insee, parce qu'il n'en dispose pas, mais qu'il transmet des éléments d'état civil, une étape préalable pour retrouver le NIR sera réalisée par l'Insee avant de le transformer en CSNS. La mise en place du Code Statistique Non Signifiant a fait l'objet de communications régulières au bureau du Conseil national de l'information statistique (Cnis) les 9 décembre 2020 et 18 mars 2021.

(7) NIR : numéro d'inscription au répertoire national d'identification des personnes physiques, géré par l'Insee

(8) Sur ce sujet, voir l'ouvrage « [Le secret statistique](#) », publié dans la collection « Le monde des données » de la SFdS ou l'article du journal de la SFdS : « [L'appariement sécurisé de fichiers d'étudiants grâce au 'hachage' des identifiants](#) »

## **Vie des institutions**

### **Trois règlements européens sur les data en cours de préparation**

La Commission européenne a décidé de faire de la décennie qui s'ouvre la « décennie numérique » de l'Europe. Un des tout premiers actes de cette stratégie européenne est la proposition de trois

projets de règlements <sup>(9)</sup> dont certaines dispositions auront un impact sur la statistique publique. La Commission a ainsi soumis au Parlement et au Conseil des Ministres de l'Union européenne les projets "Data Governance Act (DGA)", "Data Market Act (DMA)" et "Data Services Act (DSA)". Ces deux dernières propositions concernent plus les plateformes numériques, que ce soit en termes d'obligations à remplir selon la taille de la plateforme (DSA) ou la libre concurrence (DMA). En revanche, le DGA peut concerner la statistique publique en tant que productrice ou utilisatrice de données.

Le DGA traite en premier lieu des données couvertes par un secret et qui ne sont pas disponibles en Open Data (le cas du secret statistique est explicitement mentionné). Si le projet ne revient pas sur les dispositions déjà existantes, qu'elles soient européennes ou nationales, il impose un certain nombre de conditions que devra respecter l'accès à ces données : transparence de la décision, absence d'exclusivité (sauf en cas de nécessité), réponse dans un délai de deux mois, existence d'un point d'accès unique, ... Il conviendra d'attendre la version finale pour connaître l'impact sur l'organisation des données confidentielles en France. Le projet définit ensuite la notion d'intermédiaires de données qui sont des entreprises privées dont le rôle est de permettre l'accès à des données détenues par des entreprises privées et entoure ce rôle d'un certain nombre de précautions comme l'impossibilité d'utiliser les données concernées à d'autres fins que celles pour lesquelles l'accès a été autorisé. La question du modèle économique de ces entités et de la possibilité des administrations à obtenir, dans un cadre légal, ces données "gratuitement", est en cours de discussion au Conseil européen. La troisième partie évoque "l'altruisme des données" : il s'agit de la possibilité, pour une personne physique ou une personne morale, de confier volontairement des données à une entité spécifique dont le rôle serait de les communiquer à des fins d'intérêt public. Là encore, le modèle économique de ces entités intermédiaires est en cours de discussion. Enfin, le projet crée un Comité européen de l'innovation dans le domaine des données dont le rôle est de conseiller la Commission dans la mise en œuvre des dispositions du règlement.

Les trois projets devraient, d'ici la fin de l'année 2021, être complétés par un projet appelé "Data Act" qui entrerait dans les détails des données concernées.

Les trois projets sont en cours de discussion au Parlement qui devrait les adopter, vraisemblablement amendés, cet été. La discussion sera sans doute plus difficile du côté du Conseil des ministres qui a déjà organisé des « discussions politiques » sur le projet de la Commission. Il est courant que, pour ce qui concerne l'examen des textes régulant la gestion des données et du numérique, les débats entre Etats fassent ressortir des divergences entraînant souvent de longs débats. On peut notamment relever que Cédric O, le secrétaire d'État au numérique, a indiqué le 27 mai que la France souhaitait durcir certaines mesures.

*(9) Les projets de la Commission sont consultables en suivant les liens suivants :*

[DGA](#) ("Data Governance Act")

[DMA](#) ("Data Market Act")

[DSA](#) ("Data Services Act")

### **Menaces sur l'intégrité du système statistique brésilien**

Dans son n° 16, paru en octobre 2019, *La Statistique dans la Cité* posait la question : *l'intégrité du système statistique brésilien est-elle menacée ?* La conclusion de cet article indiquait notamment que « *les orientations politiques de la nouvelle administration brésilienne ... nous imposent de veiller au grain et d'être prêt à apporter notre aide à nos collègues de l'agence brésilienne pour la géographie et la statistique (IBGE) si cela devient nécessaire* ».

De récentes attaques gouvernementales contre la tenue du recensement de la population ont eu lieu récemment mais la communauté scientifique et les médias ont su réagir, ainsi que le montre l'extrait ci-après de l'éditorial <sup>(10)</sup> du dernier numéro <sup>(11)</sup> de *Estadística*, l'infolettre de l'Institut Interaméricain de la Statistique (IASI) :

*Le recensement au Brésil devait avoir lieu en 2020 et a été reporté, en principe à 2021, en raison de l'intensité de la Covid-19 dans ce pays. En 2021, l'incidence de la maladie a fortement augmenté dans le pays, la vaccination se déroulant à un rythme lent, ce qui a conduit à débattre sur l'opportunité de réaliser le recensement en 2021 en toute sécurité. Tout le monde a défendu sa réalisation mais n'était pas d'accord sur la date la plus appropriée pour sa réalisation.*

*Au cours de ce débat, un budget fédéral a été voté qui a rendu le recensement irréalisable, que ce soit en 2021 ou en 2022, indiquant clairement que la classe politique et les autorités*



gouvernementales responsables du budget approuvé ne comprenaient pas ou n'acceptaient pas le rôle fondamental du recensement au sein des dirigeants brésiliens. De nombreux acteurs et les médias ont tout fait pour remettre en cause cette décision, qui conduirait à un black-out statistique inacceptable. La Cour suprême du Brésil a finalement tranché en faveur du recensement brésilien.

Le débat intense, ainsi que l'existence de parlementaires favorables au recensement, amènent à la conclusion que ceux qui l'ont majoritairement nié étaient motivés par des considérations politiques. La réaction des différents acteurs, conscients de la nécessité du Recensement, à son tour, a sans aucun doute renforcé l'initiative et la décision de la Cour suprême. L'IBGE, l'agence brésilienne pour la géographie et la statistique, historiquement soucieuse d'informer et d'éduquer ses partenaires sur les statistiques publiques a récolté les fruits de ses nombreuses actions en faveur de la maîtrise de la statistique. De nombreux acteurs se sont coalisés pour démontrer que la statistique est un bien public qui, comme la communication scientifique, au-delà d'être limitée aux programmes de formation professionnelle nécessaires, résulte d'une action quotidienne pour éclairer les différents acteurs sociaux.

(10) Le texte original de cet article a été publié en espagnol et en anglais

(11) N° 129 - juin 2021

## Anniversaire

**Il y a trente ans, le 18 juin 1991, les Principes fondamentaux de la statistique officielle <sup>(12)</sup> étaient adoptés ...**

... par la Conférence <sup>(13)</sup> des statisticiens européens pendant sa 39e session au Palais des Nations à Genève. Ces principes avaient été préparés dès le début de l'année 1990 à la demande de plusieurs pays d'Europe Centrale et Orientale (les PECO) qui amorçaient leur transition vers des régimes démocratiques et une économie de marché. Ces pays étaient conscients que les défis qu'ils allaient avoir à relever nécessitaient un nouveau paradigme pour leurs systèmes statistiques et qu'ils avaient besoin de nouveaux repères, non pas tellement d'un point de vue technique, mais surtout en raison de la nécessité de se référer à une nouvelle vision du rôle de la statistique dans la société. Ils reconnaissaient que la production et la diffusion de statistiques se devaient d'être légitimes et crédibles et que de nouveaux concepts comme l'impartialité, la fiabilité, la pertinence et la transparence devaient désormais guider leur activité. Le groupe de rédaction de ces principes était présidé par le GUS, l'office statistique polonais, et la coordination des travaux avait été confiée à l'Insee, qui assurait la vice-présidence du groupe ; deux autres PECO <sup>(14)</sup> (Bulgarie et Roumanie) et deux autres pays d'Europe de l'Ouest (Espagne et Suisse), ainsi que la Turquie, Eurostat et l'Institut international de statistique, y participaient.

On peut noter qu'à l'époque ces principes portaient le nom de *Principes fondamentaux de la statistique officielle dans la région* <sup>(15)</sup> *de la CEE/ONU*. Ce n'est que trois ans plus tard que, après que les autres régions des Nations-Unies eurent reconnu que ce texte avait une portée universelle, que la Commission de statistique des Nations-Unies a adopté ces dix principes au niveau mondial.

On peut trouver une histoire détaillée de la naissance de ces principes au chapitre 5 de l'ouvrage publié en 2003 à l'occasion du 50e anniversaire de la Conférence des statisticiens européens (pages 63 à 76) ; cet ouvrage peut être librement consulté [ICI](#).

(12) La Statistique dans la cité aurait préféré utiliser l'expression 'statistique publique', mais tous les documents des Nations-Unies ont fait le choix de traduire 'official statistics' par 'statistiques officielles'. Nous ne pouvons que respecter ce choix {NdR}

(13) La Conférence des statisticiens européens est un organisme subsidiaire à la fois de la Commission de statistique des Nations-Unies et de la Conférence économique pour l'Europe des Nations-Unies (CEE/ONU)

(14) PECO est l'acronyme permettant de désigner les Pays d'Europe Centrale et Orientale

(15) Le système des Nations-Unies comprend cinq régions : outre l'Europe (qui comprend aussi le Canada, les Etats-Unis et Israël), les autres régions sont l'Amérique Latine et les Caraïbes (siège à Santiago, Chili), l'Afrique (siège à Addis-Abeba), l'Asie occidentale (siège à Beyrouth) et l'Asie et le Pacifique (siège à Bangkok)

## Courrier des lecteurs

### **À propos de l'article « Marges d'erreur et sondages pour l'élection présidentielle de 2022 » paru dans la dernière infolettre**

On nous écrit à propos de l'article « *Marges d'erreur et sondages pour l'élection présidentielle de 2022* » paru dans la dernière infolettre, dans sa section Méthodes

Cet article mentionne un amendement à une loi organique relative à l'élection du Président de la République. Cet amendement vise à obliger les instituts de sondage à publier la marge d'erreur d'une enquête lors de toute diffusion, et non seulement à la première diffusion des résultats. Louable obligation, à première vue : les sondages sont une source importante d'information lors d'une campagne électorale, et tout ce qui peut contribuer à leur transparence ne peut qu'être jugé positif.

Mais cette obligation va-t-elle vraiment atteindre son but ? Il est permis d'en douter. Nos législateurs, en l'occurrence, font la confusion entre deux notions statistiques différentes : le biais et la précision.

La variance des résultats des enquêtes de différents instituts montre que la précision des sondages est supérieure à celle que l'on calcule à partir de la formule standard, qui sert de base actuellement à la publication des marges d'erreur (le lecteur intéressé pourra [lire cette analyse](#), pour plus de détail). On peut conjecturer que cette précision accrue est due aux redressements effectués par les instituts, redressement rendus nécessaires par le biais des données collectées.

Mais aucune formule statistique ne permet de mesurer le biais (sinon, il n'y aurait plus de biais....). Or, ce biais est bien la principale difficulté à laquelle sont confrontés les sondeurs : on l'a vu encore récemment avec l'élection de Joe Biden en novembre dernier. Lors de l'élection de François Hollande en 2012, les marges d'erreur publiées des sondages sur le deuxième tour auraient dû interdire à tout commentateur, avant le jour de l'élection, de donner le nom du vainqueur. Alors qu'il n'y avait en fait aucun doute. A contrario, lors de cette même élection, au premier tour, les scores de Marine Le Pen et Jean-Luc Mélenchon étaient clairement en-dehors des marges d'erreur publiées.

En obligeant à la publication d'un indicateur à la fois erroné et non pertinent, on ne renforce pas la transparence des sondages : on en affaiblit la crédibilité.

### **On nous écrit aussi du Québec**

Le Rédacteur en chef du journal *Convergence* de l'Association des statisticiennes et statisticiens du Québec (ASSQ) a particulièrement apprécié l'article de notre n° 24 sur la définition de la statistique due à Eugène Labiche et nous a demandé l'autorisation de le reproduire dans le prochain numéro de sa revue qui sortira en septembre 2021. Bien entendu, c'est avec grand plaisir que nous lui avons donné notre accord et nous nous réjouissons que *La statistique dans la cité* soit lu outre-Atlantique dans la 'Belle province'.

**La rédaction de *La Statistique dans la Cité* invite ses lecteurs à participer à la vie de leur infolettre en écrivant à l'adresse : [sep@sfds.asso.fr](mailto:sep@sfds.asso.fr)**

## Notes de lecture

### **Lu dans *Variations*, le Web magazine d'Ensaë Alumni** <sup>(16)</sup>, daté du 12 mai 2021 : **« Faire parler les comptes nationaux : l'importance du vocabulaire », ...**

... Didier Blanchet, ancien directeur des études et synthèses économiques de l'Insee revient sur un problème qu'il avait déjà abordé en 2019, celui du « bond troublant » du produit intérieur brut (PIB) de l'Irlande en 2015 <sup>(17)</sup>. Cette année-là, le PIB de l'Irlande affichait une croissance exceptionnelle : + 32 % aux prix courants, + 24 % à prix constants (« en volume »). Cela ne résultait pas d'une erreur statistique mais d'une décision d'une entreprise multinationale qui avait choisi d'attribuer à sa filiale irlandaise la propriété de tous ses brevets à compter de cette date. Cette filiale était donc désormais attributaire de redevances, qui rentrent dans le PIB du pays, en conformité avec les règles du système de comptabilité nationale des Nations-Unies adopté en 2008 (SCN08). Traditionnellement, le PIB d'un pays en volume est considéré comme le meilleur indicateur synthétique de son activité économique. On a du mal à croire que l'activité économique de l'Irlande ait progressé aussi fortement en un an alors même que l'emploi, les salaires et toutes les autres variables macroéconomiques n'ont pas varié autant et de loin. D'où l'embarras des commentateurs.

Dans sa très riche contribution, Didier Blanchet tire aussi les leçons de la crise sanitaire actuelle au

cours de laquelle la mesure du PIB s'est heurtée à des difficultés supplémentaires, en particulier dans les services publics <sup>(18)</sup>. Il suggère de privilégier la lecture du PIB en termes de revenus distribués aux différents facteurs de production : ainsi considéré, le « bond irlandais de 2015 » reste étonnant, mais cesse de paraître absurde. Corrélativement, l'analyse des variations de l'activité économique du pays ne gagne pas, selon lui, à être menée globalement avec un indicateur très conventionnel comme le PIB en volume : mieux vaut considérer « les productions » que « la production ».

Cette position, que Didier Blanchet avait déjà exposée en 2019, ne fait pas l'unanimité chez les spécialistes de la comptabilité nationale et chez les utilisateurs. Dans un article également publié par *Variations* en 2019 <sup>(19)</sup>, François Lequiller, ancien directeur des comptes nationaux à Eurostat, suggérerait pour sa part de corriger le SCN08 de façon que les redevances ne soient plus enregistrées comme des rémunérations d'un facteur de production mais comme des flux financiers ; à ce prix, la vision traditionnelle du PIB pourrait être conservée. Ainsi présenté, le débat semble technique et il l'est en grande partie, mais il est lourd d'enjeux idéologiques et politiques au moment où la « réindustrialisation » ou « la souveraineté économique » sont des thèmes d'actualité. Toutes ces notions ne peuvent prendre sens que si on sait donner un contour précis à ce dont on parle quand on parle de « l'économie d'un pays ». Didier Blanchet ne récuse pas ces débats lorsqu'il écrit : « Il faut des indicateurs mesurant les revenus générés par la véritable base productive du pays, ce qu'on avait l'habitude d'attendre du PIB et qu'il faut donc chercher à préserver ». Mais le lecteur reste sur sa faim : qu'est-ce que la « véritable base productive du pays » ? Comment aborder « les productions » (au pluriel) dans une analyse générale ?

(16) Cette revue électronique de l'Association des anciens élèves de l'Ensaie est en accès libre : <http://variances.eu>

(17) La statistique dans la cité avait aussi abordé cette question dans son n°7 (décembre 2017)

(18) Voir par exemple La statistique dans la cité n° 21 (octobre 2020) :

"Confinement : peut-on comparer aujourd'hui le PIB de la France et celui de l'Allemagne ? "

(19) Lire [Les multinationales rendent-elles le PIB obsolète ?](#)

**Lu dans l'édition numérique du magazine *Le Particulier (groupe du Figaro)* en date du 31 mai 2021 : « Devez-vous obligatoirement répondre aux enquêtes de l'Insee ? », ...**

... ce court article explique clairement de façon très pédagogique à un large public pourquoi les enquêtes de l'Insee auprès de ménages ou des entreprises sont obligatoires et comment elles sont collectées et dans quels buts ; il rappelle également que le secret statistique garantit que les données recueillies à l'occasion de ces enquêtes sont strictement confidentielles et ne peuvent en aucun cas être utilisées à d'autres fins que la fourniture de statistiques ou à la recherche scientifique. Elle met de façon simple en garde sur le risque de se retrouver face à de faux enquêteurs. L'article est accessible [ICI](#).

**Vu sur France 2 le jeudi 20 mai l'émission *Cash Investigation* d'Élise Lucet sur le thème : « Vos données personnelles valent de l'or » ...**

Cette émission (visible en rediffusion sur le site France TV) a présenté une série de courts reportages illustrant par des exemples frappants le thème « Vos données personnelles valent de l'or » :

- Des listes comportant des centaines de millions de « contacts » (noms, prénoms, adresses, numéros de téléphone, emploi...) y compris de ministres, responsables de la police etc. sont en vente, illégalement, par abonnement, sur Internet. Explication : des applications sur votre téléphone mobile demandent l'autorisation d'accéder à vos contacts. Si vous acceptez et si les responsables à l'origine de ces applications sont peu scrupuleux, ils collectent les coordonnées de vos correspondants pour les revendre ; et si des amis et correspondants donnent accès à leurs contacts, alors vos coordonnées sont à portée de clic...
- D'autres applications sur téléphone ou des sites web collectent aussi des informations sur vous quand vous les interrogez ; un reportage donne l'exemple de l'application *Ma Grossesse* proposée par le site *Doctissimo*, qui alimente nous dit-on une foule de marchands de données (data brokers) sans l'accord de l'utilisateur ; *Doctissimo* a démenti mais a refusé de répondre aux questions des journalistes, et l'application *Ma Grossesse* a curieusement suspendu son fonctionnement ;
- Un reportage est consacré au recueil par la branche française de la société américaine IQVIA d'une information très complète sur les achats des Français en pharmacie, avec collecte des données sur le prescripteur, la prescription, le taux de remboursement et les données d'identification issues de la carte *Vitale*. La moitié des pharmacies françaises seraient impliquées dans ce recueil, autorisé par la



Cnil en 2018. Les pharmaciens y gagnent quelques statistiques et un logiciel fourni. Les patients et les prescripteurs sont censés être informés par des affichettes (que les journalistes n'ont aperçues nulle part dans plusieurs centaines de pharmacies visitées) et disposer d'un droit d'opposition (qui n'était pas techniquement possible à la date de l'émission). Le président de cette filiale française d'IQVIA <sup>(20)</sup> a indiqué que les données étaient pseudonymisées et que sa société vendait des études d'intérêt général et non des données. La Cnil a fait savoir qu'elle procéderait à des contrôles.

- Un expert explique aussi pourquoi il ne suffit pas de supprimer les noms, prénoms et NIR en les remplaçant par un pseudonyme pour rendre les données anonymes : avec très peu d'informations sur une personne, on la retrouve sans difficulté dans une base de données.

*(20) Le président de cette filiale française de la société IQVIA est Jean-Marc Aubert qui a été en 2018 et 2019 le directeur de la DREES (direction statistique du ministère de la santé) ; avant 2018 il était directeur à la société QuantileIMS, ancien nom d'IQVIA, et y est donc retourné après les deux années qu'il a passées à la DREES*

## **Annonces**

### **11e Colloque francophone sur les sondages - 6 au 8 octobre 2021**

L'Université libre de Bruxelles (ULB) accueille du mercredi 6 au vendredi 8 octobre 2021 le *11e Colloque international francophone sur les sondages*. Les lundi 4 et mardi 5 octobre 2021 sont consacrés à des ateliers de formation. Ce colloque se tiendra sous un format « hybride » en privilégiant les communications en présentiel mais en permettant un certain nombre de communications faites « à distance » par les conférenciers qui ne pourront pas faire le déplacement jusqu'à Bruxelles, notamment en raison de la situation sanitaire.

Cette rencontre scientifique internationale est organisée depuis 1997, tous les deux à trois ans, sous l'égide de la Société française de statistique (SFdS). Elle réunit des chercheurs et praticiens, issus des instituts publics ou du secteur privé, qui réalisent ou utilisent des enquêtes par sondage et permet de faire le point sur l'état des pratiques et de la recherche dans les divers domaines de la méthodologie des enquêtes et des sondages. Plus d'informations et inscription sur le [site du colloque](#).

### **Prochaines Journées de Méthodologie Statistique de l'Insee en 2022**

L'Insee organisera la quatorzième édition des Journées de Méthodologie Statistique de l'Insee (JMS) dans la deuxième quinzaine de mars 2022. L'organisation matérielle (date et lieu exacts) sera précisée ultérieurement. En effet, le contexte sanitaire actuel rend encore incertaine la possibilité de réaliser les JMS uniquement en présentiel. Une organisation hybride est pour l'instant privilégiée, avec une partie en présentiel et une partie en distanciel et un basculement éventuel sur une seule des deux configurations. La décision sur le mode de réalisation sera prise en octobre au plus tard.

Organisées depuis 1991, les Journées de Méthodologie Statistique de l'Insee sont un lieu d'échanges très apprécié. La dernière édition tenue en juin 2018 avait enregistré plus de 600 inscrits. Les JMS ouvertes sur des domaines d'intérêt larges, s'adressent à un public diversifié, tant à l'intérieur du service statistique public qu'à sa périphérie, y compris au monde de l'enseignement et de la recherche et au secteur privé, ainsi qu'à nos collègues de l'étranger. La participation est libre de frais d'inscription.

Pour en savoir plus, voir la page des JMS sur le site de l'Insee : [jms-insee.fr](https://jms-insee.fr).

### **Les comptes de la Nation 2020**

Le n° 1860 d'Insee Première consacré à la présentation des comptes de la nation pour 2020 est paru le 28 mai 2021 : <https://www.insee.fr/fr/statistiques/5387891>

## **Vie du groupe**

Depuis la parution de l'infolettre n° 24 en avril dernier, deux *Cafés de la Statistique* ont été organisés, en visioconférence :

- le mardi 11 mai sur le thème de la mesure de l'activité hospitalière avec Roland Cash, médecin, conseiller scientifique du Haut conseil pour l'avenir de l'assurance maladie, ancien responsable scientifique de la mission Tarification à l'activité ;
- le mardi 8 juin sur le thème de l'évaluation de la compétence des élèves avec Thierry Rocher, adjoint

au sous-directeur des évaluations et de la performance scolaire à la DEPP, ministère de l'éducation nationale, président de l'IEA (International Association for the Evaluation of Educational Achievement).

Les élections pour le renouvellement des sièges de membres du Conseil de la SFdS et des bureaux de ses groupes ont été organisées par voie électronique en mai. Les résultats de ce vote ont été annoncés pendant l'Assemblée Générale de la SFdS tenue le 8 juin. Pour ce qui concerne le groupe *Statistiques et Enjeux Publics*, un seul siège était à pourvoir ; Erik Zolotoukhine a été réélu.

Le bureau du groupe se réunira fin juin pour répartir les tâches de ses différents membres et pour préparer le programme de rentrée des Cafés.

*Responsable de l'infolettre : Chantal Cases, présidente du groupe SEP*

*Rédacteur en chef : Jean-Louis Bodin*

*Secrétaire de rédaction : Jean-Pierre Le Gléau*

*Webmestre : Érik Zolotoukhine*