

Statistique et société

juin 2018

Volume 6, Numéro 1

**En mémoire
d'Alain Desrosières**

Sommaire

Statistique et société

Volume 6, Numéro 1

- 7 **Éditorial**
Emmanuel Didier
Rédacteur en chef de Statistique et société
- 9 **Dossier : Actualité de la sociologie de la quantification d'Alain Desrosières**
Michel Armatte
Président du groupe Histoire de la Statistique, des Probabilités et de leurs usages de la Société Française de Statistique
- 11 **Article du dossier : Approcher les données de gestion au moyen de la sociologie historique de la quantification**
Jean-Marie Pillon
Institut de recherches interdisciplinaires en sciences sociales, Université Paris Dauphine, Paris Sciences et Lettres
- 17 **Article du dossier : Une statistique de l'intimité**
Axel Pohn-Weidinger
Docteur en sociologie, Université de Göttingen, et associé au Centre Georg Simmel, Ecole des Hautes études en sciences sociales
- 21 **Article du dossier : Légitimer par les nombres, à propos d'une autre fonction des outils quantitatifs**
Emilien Ruiz
Historien, maître de conférences à l'université de Lille
-

Sommaire

Statistique et Société

Volume 6, Numéro 1

- 27 **Méthodes** : Addiction aux jeux de hasard et d'argent en ligne : comment mettre en place une prévention basée sur les preuves grâce aux Big Data ?
Amandine Luquiens
Psychiatre addictologue, Hôpital Paul Brousse, Villejuif.
- 31 **Société** : Protection des données personnelles et recherche : constats et perspectives d'évolution
Sophie Vulliet Tavernier
Directrice des relations avec les publics et la recherche - Commission Nationale de l'Informatique et des Libertés
- 37 **Histoire** : Éléments pour une histoire du « Numéro de sécurité sociale »
Gérard Lang
Statisticien retraité, SFdS
-

Statistique et société

Magazine quadrimestriel publié par la Société française de statistique. Le but de Statistique et société est de montrer d'une manière attrayante et qui invite à la réflexion l'utilisation pratique de la statistique dans tous les domaines de la vie, et de montrer comment l'usage de la statistique intervient dans la société pour y jouer un rôle souvent inaperçu de transformation, et est en retour influencé par elle. Un autre dessein de Statistique et société est d'informer ses lecteurs avec un souci pédagogique à propos d'applications innovantes, de développements théoriques importants, de problèmes actuels affectant les statisticiens, et d'évolutions dans les rôles joués par les statisticiens et l'usage de statistiques dans la vie de la société.

Rédaction

Rédacteur en chef : **Emmanuel Didier**, CNRS, France

Rédacteurs en chef adjoints :

Jean-Jacques Droesbeke, Université Libre de Bruxelles, Belgique

Chloé Friguet, Université de Bretagne-Sud, France

Jean-François Royer, SFdS - groupe Statistique et enjeux publics, France

Jean-Christophe Thalabard, Université Paris-Descartes, pôle de recherche et d'enseignement supérieur Sorbonne Paris Cité, France

Comité éditorial

Représentants des groupes spécialisés de la SFdS :

Ahmadou Alioum, groupe Biopharmacie et santé

Christophe Biernacki, groupe Data mining et apprentissage

Alain Godinot, groupe Statistique et enjeux publics

Delphine Grancher, groupe Environnement

Marthe-Aline Jutand, groupe Enseignement

Elisabeth Morand, groupe Enquêtes

Alberto Pasanisi, groupe Industrie

Autres membres :

Jean Pierre Beaud, Département de Science politique, UQAM, Canada

Corine Eyraud, Département de sociologie, Université d'Aix en Provence, France

Michael Greenacre, Department of Economics and Business, Pompeu Fabra
Université de Barcelone, Espagne

François Heinderyckx, Département des sciences de l'information, Université
Libre de Bruxelles, Belgique

Dirk Jacobs, Département de sociologie, Université Libre de Bruxelles, Belgique

Gaël de Peretti, INSEE, France

Theodore Porter, Département d'histoire, UCLA, États-Unis

Carla Saggiotti, INSEE, France

Patrick Simon, INED, France

Design graphique

fastboil.net

ISSN 2269-0271



Emmanuel DIDIER

Rédacteur en chef de Statistique et Société

Chers Lecteurs,

Je voudrais revenir sur le précédent numéro, intitulé « Ouvrir les données » car il a subi un sort très injuste. Très peu de temps après sa mise en ligne, un hacker est parvenu à s'introduire sur le site du CNRS qui nous héberge et a rendu l'accès à toutes les revues de la SFdS, y compris la nôtre, inaccessible. Nous ne sommes parvenus à résoudre le problème que très récemment. De telle sorte que ces fameuses données ouvertes sont en fait restées entièrement fermées pendant presque toute la durée d'exposition du numéro. Nous vous engageons donc vivement à en reprendre la lecture maintenant, c'était un excellent crû qui ne méritait en rien cette censure bien involontaire de notre part ! Que les auteurs acceptent nos excuses les plus sincères pour ce contretemps.

Le numéro que nous vous présentons aujourd'hui est d'abord composé d'un dossier comprenant des textes écrits par les lauréats de la première édition du prix Alain Desrosières décerné en 2016. L'introduction écrite par Michel Armatte, président du groupe Histoire des probabilités de la statistique et de leurs usages de la SFdS, rappelle que le prix récompense de jeunes auteurs utilisant les travaux de Desrosières dans leurs propres enquêtes. Pour cette première occurrence, ont été distingués Jean Marie Pillon pour un travail portant sur la mise en place de Pôle emploi ; Axel Pohn-Weidinger pour son article sur la catégorie du logement surpeuplé et enfin Emilien Ruiz pour son histoire des décomptes de fonctionnaires en France.

Hors dossier, nous présentons un article méthodologique d'Amandine Luquiens qui explique les techniques utilisées pour mettre en place une prévention contre l'addiction aux jeux en ligne basée sur les preuves grâce aux Big Data. Puis, un article de Sophie Vulliet-Tavernier, membre de la Cnil, qui expose les réflexions actuelles de son institution sur la protection des données personnelles et recherche. Enfin Gérard Lang, ancien administrateur de l'Insee, présente une histoire très riche du Numéro d'Inscription au Répertoire (NIR) autrement connu sous le nom de « Numéro de Sécu ». Cet article entame une sorte de feuilleton tenu par l'auteur qui continuera dans le prochain numéro en revenant sur la question des fichiers établis pendant la seconde guerre mondiale.

Bonne lecture !

Actualité de la sociologie de la quantification d'Alain Desrosières



Michel ARMATTE

Président du groupe *Histoire de la Statistique, des Probabilités et de leurs usages* de la Société Française de Statistique

Le Prix Alain Desrosières 2016 a récompensé trois jeunes chercheurs en socio-histoire de la quantification.

Le Prix Alain Desrosières, décerné tous les deux ans par la Société française de statistique sur la base d'un fonds constitué par ses amis et collègues, vise à distinguer des travaux de recherche actuels qui profitent aujourd'hui de la fécondité de la sociologie de la quantification que celui-ci a développé dans ses interventions et ses écrits jusqu'à la veille de sa disparition en 2013, aussi bien l'ouvrage historique fondateur qu'est la *Politique des grands nombres* (2000), que les nombreux articles réunis dans les deux tomes de *L'argument statistique aux Presses des Mines* (2008) et le recueil posthume *Prouver et gouverner à la Découverte* (2014).

Décerné une première fois en 2016 par le groupe « Histoire » de la SFdS, le Prix a récompensé trois jeunes chercheurs qui se sont distingués tous les trois par la qualité générale de leurs travaux et la référence faite à Alain Desrosières, mais que rien n'autorise à assimiler cependant, tant est différente la forme de ces travaux (articles ou thèse), la discipline dont ils relèvent (statistique, histoire, économie, sociologie), la problématique dont ils traitent, et la nature de ce qu'ils empruntent à Desrosières. Plutôt que de faire nous-mêmes une allusion imparfaite à leurs travaux, nous avons jugé utile de leur demander une courte contribution explicitant pour le lecteur de *Statistique et Société* la façon dont ils ont chacun mobilisé, utilisé, transformé ou dépassé les apports de Desrosières à la sociologie de la quantification.

Jean-Marie Pillon, primé pour sa thèse de 2014 en sociologie à Paris X sur « Le rendement du chômage. Travail de mesure et mesure du travail à Pôle emploi » témoigne de l'actualité de Desrosières en terme « d'approche, de programme et de surprise ». L'approche est celle de la typologie des rôles de l'Etat utile à la compréhension du changement de régime quand on est passé de l'ANPE à Pôle emploi. Le programme est celui d'une description des mutations des outils statistiques privilégiés quand on passe de l'Etat-ingénieur à l'Etat-néolibéral. Les surprises sont les nouveaux dispositifs que méthode et programme permettent d'analyser, comme par exemple le rôle des algorithmes si importants aujourd'hui.

Axel Pohn-Weidinger, primé pour son article « Une statistique de l'intimité. La catégorie du logement surpeuplé entre statistique et droit (1936-1962) », *Genèses*, 2013, a été séduit par l'idée que la statistique est à la fois méthode de preuve et de gouvernement, laquelle rend bien compte du fait que les conventions introduites pour la quantification du surpeuplement des logements sont largement dépendantes du mode de gouvernement de l'habitat, en terme de normes, de primes, d'allocations. L'auteur, qui est de nationalité allemande, ce qui montre la réception internationale du travail d'Alain Desrosières, prolonge l'idée desrosiérienne que quantifier c'est convenir puis mesurer par une proposition intéressante : il faut inverser l'idée

que la statistique fournit des appuis pour l'action publique ; c'est tout autant l'action publique qui détermine les conventions de la quantification.

Emilien Ruiz, primé pour sa thèse EHESS d'histoire en 2013 « Trop de fonctionnaires? Contribution à une histoire de l'Etat par ses effectifs (France, 1850-1950) », cherche à analyser les conditions de production du recensement des agents de la fonction publique en 1947, et il montre que cette quantification est davantage un outil de légitimation de nouvelles institutions (Insee, DFP), face à la direction du Budget, qu'un outil de preuve voire de gouvernement, et il propose de dépasser l'analyse de Desrosières par un déplacement de la focale à l'intérieur même de l'Etat en s'intéressant aux rapports de force entre administrations et grands Corps.

Illustration, prolongement, déplacement, inversion, dépassement, voilà bien en œuvre diverses appropriations et adaptations de la théorie de la quantification esquissée par Alain Desrosières, que nous souhaitons offrir au lecteur dans les textes qui suivent, et que nous allons réitérer comme critères pour la seconde édition du Prix en 2018.

Approcher les données de gestion au moyen de la sociologie historique de la quantification



Jean-Marie PILLON

Institut de recherches interdisciplinaires en sciences sociales, Université Paris Dauphine, Paris Sciences et Lettres

L'activité de Pôle emploi produit de nombreuses données, parmi lesquelles les plus connues sont les décomptes de demandeurs inscrits à chaque date. Mais d'autres données de gestion sont utilisées pour cadrer et contrôler le travail des agents. Et ces derniers agissent en retour sur les conventions qui permettent d'établir ces données.

Les travaux d'Alain Desrosières sont régulièrement classés dans la rubrique « sociologie des sciences », à la suite de Bruno Latour et Michel Callon dont il s'inspirait largement¹. Le cadre d'analyse qu'il a participé à développer avec Robert Salais, Laurent Thévenot, François Eymard-Duvernay ou encore Emmanuel Didier est néanmoins bien plus plastique et permet de s'attaquer à nombre d'outils de gouvernement et de preuve². J'ai cherché à éprouver ce cadre d'analyse au cours de mon travail de thèse sur les données de gestion utilisées dans le cadre de la lutte contre le chômage.

Une typologie des formes d'État et de connaissance.

A. Desrosières nous a d'abord légué une typologie des formes d'État : L'État libéral, l'État ingénieur, l'État néolibéral, autant de formes d'État qui mobilisent chacune des formes différentes de quantification³. Loin d'être figée, une telle typologie permet d'analyser les transformations de la raison d'État, c'est-à-dire la façon dont on passe d'une forme d'État à une autre. A la naissance de l'ANPE, les données produites par l'ANPE relèvent en grande partie d'un État ingénieur. L'État central cherche à planifier sur quatre ans la demande d'emploi, les offres d'emploi et le nombre de demandeurs d'emploi qui retrouvent un travail⁴. En fin de période, en revanche, les données quantifiées évoquent clairement un État néolibéral. Les agences Pôle emploi sont comparées entre elles, les directeurs peuvent consulter leurs résultats en temps réel, à la façon du benchmarking⁵. Toutefois, si l'on suit le processus qui a conduit d'une situation à l'autre, il paraît très difficile d'identifier un moment où l'on passerait soudain d'une forme d'État à une autre. On assiste plutôt à un amendement progressif des dispositifs de décompte, avec des périodes de forte lutte et des périodes plus apaisées mais qui conduisent à donner cops à l'instrumentation gestionnaire ou néo libérale.

1. Desrosières A. (1989), « Comment faire des choses qui tiennent : histoire sociale et statistique », *Histoire & Mesure*, vol. 4, n°3, pp. 225-242.
2. Desrosières A. (2011), « Les origines statisticiennes de l'économie des conventions : réflexivité et expertise », *OEconomia*, vol. 2011, n°2, pp. 299.
3. Desrosières A. (2003), « Historiciser l'action publique : l'État, le marché et les statistiques », in *Historicités de l'action publique*, PUF, Paris, pp. 207-221.
4. Pillon J.-M. (2014), « Un nouveau rôle pour Pôle emploi ? Évolution de ses stratégies au prisme des indicateurs de performance », *Connaissance de l'emploi - Centre d'études de l'emploi*, n°117, pp. 1-4.
5. Pillon J.-M. (2015), « Le pointage et le placement ? Quand Pôle emploi échoue à régler ses contradictions au moyen de logiciels », *Gouvernement et action publique*, vol. 4, n°2, pp. 81-103.

Cette typologie des formes d'État nous invite également à rechercher les soubassements théoriques et épistémologiques des instruments employés par les acteurs. Dans le cas du chômage on peut ainsi noter que dès 1967 le cadrage théorique qui justifie la naissance de l'ANPE n'est pas Keynésien. Sans se référer directement aux approches dites de « job search »⁶, qui n'avaient pas encore fait l'objet d'une importation en France, les approches du marché du travail mobilisées par les ingénieurs du Plan, et qui donneront naissance à l'ANPE, étaient comparables. Il s'agissait d'une approche informationnelle des rigidités du marché du travail. L'ANPE est ainsi pensée par la commission de main d'œuvre du Plan comme une bourse aux emplois chargée de fluidifier le marché du travail, tout en améliorant l'allocation de ce facteur sur l'ensemble du territoire et entre les secteurs d'activité⁷. Ce cadrage se trouve incorporé aux indicateurs de l'ANPE dès 1971 : l'État quantifie le nombre de chômeurs qui retrouvent un emploi. Et cette théorie se perpétue jusqu'à nos jours puisque l'enjeu reste le même en matière de performance : combien de demandeurs d'emploi sortent des listes ?

Cette composante initiale du mandat de l'ANPE structure de manière importante le rôle attribué à Pôle emploi par les politiques de l'emploi. Du point de vue de leurs indicateurs, l'ANPE comme Pôle emploi n'ont jamais eu pour mission de faire baisser le chômage. Il demeure toujours un écart entre le nombre de chômeurs et le nombre de postes vacants et Pôle emploi n'a pas les moyens d'intervenir sur le taux de chômage. Pour autant, lorsque l'on conduit une campagne d'entretiens depuis les agences Pôle emploi jusqu'au ministère du Travail, cette interprétation tend progressivement à disparaître du discours des acteurs. Les agents assument cette vision, tout comme leur encadrement. Au sein même de la direction de Pôle emploi il s'agit d'une opinion assez peu discutée, bien qu'elle puisse être occultée. Dans la relation entre Pôle emploi et le ministère du Travail, en revanche, quelque chose change⁸. L'agenda politique vient ainsi régulièrement bousculer les rapports entre Pôle emploi et ses tutelles. L'opérateur public de placement se retrouve alors enrôlé dans la lutte contre le chômage, à travers différents plans d'action : « chômeurs longue durée », « plans rebonds », etc. Alors même que Pôle emploi ne peut que constater l'existence de ces chômeurs, il conserve la possibilité de modifier l'ordre de la file d'attente ou d'élever le niveau de qualification des demandeurs d'emploi. Mais l'influence de ces interventions sur le niveau de l'emploi demeure marginale.

Ainsi en resituant les indicateurs de performance par rapport au mandat confié à Pôle emploi, l'on peut rendre intelligibles les tensions qui secouent quotidiennement cet établissement entre une injonction politique impossible à mettre en œuvre et une situation de l'emploi et de l'embauche particulièrement sinistrée⁹.

La sociologie historique de la quantification comme méthode

A travers ses écrits, Alain Desrosières propose ensuite une méthode. Celle-ci suppose de préciser à quel type de statistiques nous faisons face sur nos terrains et ce de manière très concrète : qu'est-ce que l'on compte, dans quel but, en lien avec quelle forme de gouvernement ? C'est dans cette perspective que je me suis saisi des outils de la sociologie historique de la quantification pour étudier des objets illégitimes, du moins beaucoup moins légitimes que les statistiques publiques : les indicateurs de gestion utilisés pour quantifier l'efficacité d'une administration, en l'occurrence Pôle emploi. Comprendre le sens de ces indicateurs suppose d'abord d'étudier la construction du chômage en tant que problème public. C'est-à-dire à la fois les mobilisations

-
6. Mortensen D.T. (1987), « Job search and labor market analysis », in *Handbook of labor economics*, Amsterdam, Elsevier, pp. 849-919. ; Stigler G.J. (1962), « Information in the labor market », in *Investment in Human Beings*, The Journal of Political Economy Vol. LXX, No. 5, Part 2 (University of Chicago Press), pp. 94-105
 7. Pillon J.-M. (2015a), « La transparence du marché du travail – un travail en soi », *Revue de l'Ines*, vol. 1, n°84, pp. 31-58.
 8. Pillon J.-M. et Vivés C. (2016), « La fermeture du service public de l'emploi est-elle structurelle ? Objectiver les difficultés à produire des connaissances sur les politiques de l'emploi. », in *Les frontières de la recherche sur le social*, La Découverte, Paris.
 9. Pillon J.-M. (2017), *Pôle emploi : gérer le chômage de masse*, Presses universitaires de Rennes, France (Res-Publica), 272 p.

sociales qui ont conduit à son « invention »¹⁰, mais également les définitions théoriques qui ont été retenues pour penser, décrire et intervenir sur le phénomène. Le cadrage du traitement du chômage par des indicateurs est d'abord historique et théorique et relève de l'articulation entre mode de gouvernement et mode de connaissance.

Les indicateurs utilisés pour mesurer l'efficacité de Pôle emploi varient dans le temps mais les plus stables concernent le nombre de chômeurs reçus, les prestations dont ils ont bénéficié ou le nombre de chômeurs qui retrouvent un emploi. Le travail que j'ai mené porte donc sur les chiffres du chômage. Il s'agit toutefois de chiffres qui font l'objet de moins de publicité que la demande d'emploi en fin de mois, largement commentée dans les médias. Les données dont j'ai étudié la construction et l'usage sont surtout utilisées au sein de Pôle emploi. Elles servent aux contrôleurs de gestion pour éclairer les décisions des directeurs : directeur d'agence, directeur départemental, directeur régional, directeur général. L'élément le plus frappant lorsque l'on s'intéresse à ce type de quantification est le fait que malgré leurs défauts méthodologiques, ces indicateurs tiennent. Ils tiennent au sens qu'Alain Desrosières donnait à ce terme¹¹ : ils circulent de scène en scène ; ils font agir les acteurs qui s'y rapportent comme à une boussole. Les outils de quantification de l'efficacité de Pôle emploi cadrent même très directement le travail des conseillers à l'emploi et, partant, la façon dont les chômeurs sont traités au guichet de l'État social¹².

Le cadrage du traitement du chômage par les indicateurs s'inscrit alors au creux de dispositifs techniques bien concrets. Les catégories de découpage de la réalité incorporées aux logiciels des conseillers reprennent les visions du monde privilégiées par les indicateurs. Une telle circulation des formes et des théories n'est pas fortuite. En étudiant le développement de ces logiciels, l'on perçoit bien comment les rapports de force entre l'équipement logiciel de l'encadrement et l'équipement logiciel des agents conduit à privilégier in fine les catégories de jugement usitées parmi la hiérarchie¹³. Cette situation n'est néanmoins jamais stabilisée. Il existe des arbitrages budgétaires entre différentes fonctionnalités des outils qui peuvent, à la faveur de coalitions de circonstance, tourner en faveur des agents et de la façon dont ils définissent leur travail.

Des effets de rétroaction

Le cadrage du traitement du chômage par les indicateurs passe enfin par ce qu'Alain Desrosières désignait par le terme d'effets de « rétroaction », à savoir une intervention directe ou indirecte des acteurs sur l'instrument¹⁴. En matière d'indicateurs de gestion, les phénomènes de rétroaction sont explicitement recherchés, ils ne constituent pas un « biais » imprévu.

Les effets de rétroaction des indicateurs se jouent d'une part à travers des processus organisationnels classiques : Lors des réunions d'équipe, les conseillers se voient signifiés les résultats de la semaine précédente. Par ce biais les catégories de pensée propres aux indicateurs se diffusent jusqu'aux agents. La discussion collective revient souvent à chercher une traduction entre des actes professionnels et la façon dont les indicateurs les figurent. Contrairement à ce que l'on pourrait attendre, ce type de socialisation aux catégories des indicateurs n'est pas toujours violent, ni descendant. On peut assister à des échanges assez apaisés à propos des

10. Salais R., Baverez N. et Reynaud-Cressent B. (1986), *L'Invention du chômage*, Paris, Presses Universitaires de France, 288 p. ; Topalov C. (1994), *Naissance du chômeur, 1880-1910*, Paris, Albin Michel.

11. Cf. note 1

12. Pillon J.-M. (2016), « Hiérarchiser les tâches, classer les chômeurs. La gestion du chômage assistée par ordinateur », *Réseaux*, vol. 1, n°195, pp. 197-228.

13. Pillon J.-M. (2016), « Gérer ou manager ? L'hésitation des outils numériques de traitement du chômage », *Pyramides. Revue du Centre d'études et de recherches en administration publique*, n°25, pp. 107-126.

14. Desrosières A. (2012), « Est-il bon, est-il méchant ? Le rôle du nombre dans le gouvernement de la cité néolibérale », *Nouvelles perspectives en sciences sociales*, vol. 7, n°2, pp. 261-295.

façons de coder tel ou tel acte dans l'ordinateur pour qu'il représente les « meilleurs » résultats. Il s'agit là d'un des éléments les plus marquants de l'enquête : la prégnance des données de gestion conduit à une optimisation de la mise en scène du travail. Il ne s'agit pas là de bidonnage. Bien entendu, l'on ne se situe pas non plus dans la pureté méthodologique. Il s'agit d'une forme spécifique de mise en forme des actes en fonction des critères d'évaluation les plus courus¹⁵.

Les effets de rétroaction colorent d'autre part la division des tâches : les directeurs d'agence découpent le travail des conseillers en différentes plages d'activités qui correspondent à un indicateur en particulier qu'il s'agit d'améliorer. Certes, la plupart du temps, les conseillers demeurent relativement autonomes dans leur planning. Néanmoins, dès que les chiffres virent au rouge, l'activité se trouve réorganisée en fonction des indicateurs à rétablir¹⁶. Ce faisant, dans les agences qui se situent dans des zones économiquement sinistrées, ce tropisme des indicateurs structure très largement le travail, et ce tout au long de l'année.

Ces différentes formes de cadrages de l'activité ne sont pas toujours synchrones. La pression politique ou les usages professionnels peuvent faire entrer en résonance ces différentes influences ou au contraire les désynchroniser, ouvrant ou refermant différents espaces de liberté pour les acteurs. Quand l'ensemble des cadres du travail entrent en résonance (hiérarchie, indicateurs, logiciel, politique), les marges de manœuvre se réduisent à peau de chagrin, notamment pour les conseillers et partant pour les chômeurs.

De l'actualité d'Alain Desrosières

Pour conclure, il me semble important de rappeler les raisons pour lesquelles A. Desrosières demeure un sociologue des plus actuels. A travers son œuvre il nous lègue une approche, un programme et des surprises.

Une approche d'abord. A. Desrosières utilise régulièrement la métaphore du bébé et de l'eau du bain pour rendre raison à sa vision des statistiques, équidistante entre relativisme et réalisme. Il disait en substance : En 1968, nous, militants à l'Insee, étions tous révolutionnaires mais on ne voulait pas « jeter le bébé » statistique avec « l'eau du bain » de la critique de l'État. Dit autrement : « *Les faits sociaux peuvent être traités comme des choses seulement s'ils sont durcis d'une façon ou d'une autre [...], comme le sont les institutions, les lois, la religion, les coutumes* »¹⁷. C'est là une clé de lecture de son œuvre : il a essayé de comprendre pourquoi les statistiques tiennent alors même qu'elles sont le reflet d'un pouvoir politique lui-même fragile et discutable. A travers une telle approche, Alain Desrosières nous rappelle que l'on fait fausse route lorsque l'on assène l'idée selon laquelle l'on peut faire dire n'importe quoi aux chiffres. Chaque mode de quantification autorise différent discours, en facilite certains, en interdit d'autres.

Un programme ensuite. En proposant d'aller voir partout où cela était possible les points de contact entre les grandeurs chiffrées et le monde social, A. Desrosières a lancé un programme impossible à achever, offert ainsi aux générations qui lui succèdent. Mieux que quiconque il a analysé l'histoire entremêlée des formes de gouvernement et les évolutions de l'épistémologie des statistiques. Mais il reste un pan entier de son programme qu'il n'a pas eu le temps de réaliser. Son travail s'est concentré sur les liens amont entre statistiques et État. Les points de contact aval entre les statistiques et les citoyens ont en revanche eu moins de place dans son travail. Or, avec le développement des indicateurs de gestion dans les grandes organisations, il reste là un champ complet de recherche à arpenter. Que l'on pense à des chercheurs comme

15. Pillon J.-M. (2016), « Le crédit des gestionnaires. Les contrôleurs de gestion de Pôle emploi entre conviction et prescription », *Sociologie du Travail*, vol. 58, n°1, pp. 32-56.

16. Pillon J.-M. et Quéré O. (2014), « La planification de l'action publique à l'échelle du planning », *Temporalités.*, n°19, pp. 2-15.

17. Cf. note 1

Hugo Bertillot¹⁸ ou Pierre-André Juven¹⁹ sur l'hôpital, Vincent Arnaud Chappe sur l'égalité²⁰, François Régis Puyou sur les contrôleurs de gestion²¹, Yann Lelann sur les calculs de la protection sociale²², Marion Gilles sur les risques psychosociaux²³ ou encore Bilel Benbouzid sur les statistiques de la délinquance²⁴, nombre de jeunes chercheurs creusent le sillon ouvert par la sociologie historique de la quantification.

De nombreuses surprises enfin. Si A. Desrosières a davantage réalisé la partie amont de son programme, le travail qu'il a réalisé avec Laurent Thévenot sur la rencontre des catégorisations profanes et statistiques du réel demeure d'une très grande richesse²⁵. De plus, son œuvre regorge de fulgurances qui permettent de penser des objets quantifiés émergents qu'il n'avait pas lui-même eu le loisir d'étudier. Je pense notamment à un grand nombre de réflexions sur la constitution des statistiques administratives qui sont d'une grande aide pour penser les indicateurs conçus à partir de registre informatique²⁶. Aujourd'hui, la sociologie historique de la quantification fait toutefois face à un gigantesque défi, à travers le développement des algorithmes. Mais il est certain que se nichent dans son travail des éléments pour penser ce nouvel âge de la quantification. Les sociologues, auront en tout cas besoin de gens d'une grande finesse pour faire face à l'explosion de ces algorithmes sans pouvoir nous tourner directement vers le regard perçant d'Alain Desrosières pour les orienter.

-
18. Bertillot H. (2014), La rationalisation en douceur : sociologie des indicateurs qualité à l'hôpital, Thèse de doctorat, Paris, Institut d'études politiques.
 19. Juven P.-A. (2013), « Codage de la performance ou performance du codage: Mise en chiffre et optimisation de l'information médicale », *Journal de gestion et d'économie médicales*, vol. 31, n°2, pp. 75-91.
 20. Chappe V.-A. (2016), « Conflits et tensions autour de l'usage des indicateurs d'égalité professionnelle en entreprise : à quoi sert le « rapport de situation et comparée entre femmes et hommes »,
 21. Puyou F.-R. (2009), Le contrôle de gestion comme scène du gouvernement des groupes de sociétés : la place du calcul économique dans l'action organisée, Thèse de doctorat, Paris, Institut d'études politiques, 265 p.
 22. Le Lann Y. (2013), Les frontières comptables du social : la protection sociale et le salaire dans la comptabilité nationale (1944-2010), Thèse de doctorat, Nanterre, France, Paris Ouest Nanterre, 374 p.
 23. Gilles M. (2016), « Figures for what purposes? The issues at stake in the struggles to define and control the uses of statistics », *in The Social Sciences of Quantification*, Springer, pp. 149-160.
 24. Benbouzid B. (2010), « L'enquête de victimation américaine : de l'étude compréhensive à l'analyse des facteurs de risque (1965-1985) », *Revue Française de Socio-Économie*, vol. 5, n°1, pp. 63-82.
 25. Desrosières A. et Thévenot L. (2010), *Les catégories socioprofessionnelles*, Paris, La Découverte.
 26. Desrosières A. (2005), « Décrire l'État ou explorer la société : les deux sources de la statistique publique », *Genèses*, vol. 58, n°1, pp. 4-27 ; Revil H. (2015), « L'informatisation du non-recours aux droits maladie en France, The informatization of the non take-up (NTU) of benefits », *Gouvernement et action publique*, n°2, pp. 57-80.

Une statistique de l'intimité



Axel POHN-WEIDINGER

Docteur en sociologie

Université de Göttingen

et associé au Centre Georg Simmel, Ecole des Hautes études en sciences sociales

Qu'est-ce qu'un logement « habitable » ? Quand un logement est-il convenablement habité, ou au contraire surpeuplé ? Pendant longtemps, la statistique publique a beaucoup peiné pour répondre à ces questions. C'est seulement lorsque des normes juridiques ont été définies et diffusées dans la société que la mesure statistique est devenue possible.

Le nom d'Alain Desrosières, je l'ai entendu pour la première fois dans une réunion du département de sociologie de Paris 8, où, entre enseignants, nous discutons de lectures utiles pour préparer un cours intitulé « Décrire la société française ». On m'a conseillé *La politique des grands nombres*. Ce fut une révélation pour moi, jeune sociologue formé à Vienne, où l'on voyait la statistique avant tout comme une méthode, un outil pour rendre intelligible le social. La proposition d'Alain Desrosières d'envisager la statistique non seulement comme une méthode de *preuve* mais aussi, et simultanément, comme une méthode de *gouvernement* était très séduisante pour moi, sans doute aussi parce que découvrais au même moment les cours de Michel Foucault sur la biopolitique. J'ai donc découvert Alain Desrosières, comme sans doute beaucoup de jeunes collègues, à travers l'enseignement, pour préparer des cours de statistique qui avaient pour objectif d'apprendre aux étudiants non seulement le sens des chiffres mais aussi ce que les chiffres font au monde social.

Cette fascination initiale, j'ai essayé de la transmettre, en traitant avec les étudiants un certain nombre de thèmes pour lesquels les sociologues mobilisent des statistiques publiques, de l'emploi au chômage, en passant par la mobilité sociale, la formation, etc., en interrogeant à chaque fois les catégories de mesure et les effets qu'elles produisent sur la compréhension des phénomènes en question. Dans une séance sur la statistique du logement, je me suis aperçu que l'Insee calculait, pour chaque logement en France, un taux d'occupation, en regardant de très près si la taille du logement était adaptée à la composition du ménage ou si, au contraire, le logement était « surpeuplé ». Pour ce faire, l'Insee compte avec beaucoup de méticulosité les pièces habitables et décrit les membres du ménage, pour déterminer si chacun dispose d'un minimum d'espace à soi, notamment les enfants qui, à partir de 6 ans, devraient avoir une chambre à eux s'ils sont de sexe différent. L'indice de surpeuplement, c'est une véritable *statistique de l'intimité*, dont l'existence m'a frappé parce que je travaillais en même temps, dans le cadre de ma thèse, sur les demandes de logement social ; or, dans les lettres que les gens écrivaient à la mairie, je lisais parfois « je demande une chambre pour ma fille de 8 ans ». J'ai été saisi par le fait de retrouver, dans deux espaces si différents, en quelque sorte aussi bien au sommet qu'à la base de l'État, cette catégorie d'une chambre à soi, à la fois comme une sorte d'équipement minimal de l'individu, nécessaire à son épanouissement et responsable d'une souffrance en cas de manque, et comme un droit, que la puissance publique se devrait de garantir à tous les membres de la société et que l'on peut faire valoir individuellement.

Pour comprendre le surgissement socio-historique de cette statistique de l'intimité, j'ai entamé une généalogie de la catégorie du logement surpeuplé, en analysant de près les recensements et les enquêtes sur le logement effectués depuis les années 1930 en allant jusqu'aux années 1960, période où se stabilise l'indice de surpeuplement que l'Insee utilise jusqu'à nos jours. Or, je me suis rendu compte que les deux catégories cruciales qui composent la mesure du surpeuplement, « la pièce » et « l'habitant », ont soulevé pendant très longtemps des problèmes de mesure importants, à tel point que les chiffres sur le (sur)peuplement du parc immobilier étaient inopérants, et ne permettaient donc pas de déterminer le besoin en logement que le gouvernement souhaitait tant connaître pour reconstruire le pays. En effet, quatre types de problèmes traversent la mise en place de cette statistique du logement, tous liés à une façon populaire d'investir l'habitat, et en opposition avec une définition « moderne » de l'habitat en tant qu'espace rigoureusement différencié selon les fonctions du travail professionnel, du travail domestique, de la vie commune et de l'intimité.

C'est que, premièrement, la statistique publique a pendant longtemps eu des énormes difficultés à déterminer si un espace servait à l'habitation ou non. Rappelons que dans la France de l'après-guerre, les populations sinistrées couchent dans des garages, caves ou greniers, les travailleurs immigrés dans des taudis : s'agit-il de logements, avec des pièces habitables ? Puis, dans les immeubles, il est difficile de différencier, à l'intérieur du logement, les pièces qui servent à un usage personnel et celles dont l'usage est professionnel, car les fonctions du travail et de la vie privée ne sont pas dissociées dans les pratiques populaires. C'est le cas des commerçants qui couchent dans leur arrière-boutique, des ouvriers agricoles qui couchent à l'écurie, des travailleurs à domicile, et, en ville, des domestiques et des concierges pour qui le logement est indissociable du lieu de travail. Deuxièmement, les recenseurs ont du mal à distinguer la cuisine de la salle commune, car, surtout à la campagne, un certain nombre de petites maisons ne comptent qu'une pièce où la famille travaille et couche, parfois en compagnie de bêtes. Un troisième problème concerne les habitants eux-mêmes : faut-il compter les personnes qui ne sont pas présentes au moment du recensement, comme par exemple les malades soignés à l'hôpital, les travailleurs saisonniers, les appelés, les marins, etc. comme faisant partie du ménage, ou, comme le feront finalement les statisticiens, comme « temporairement absents » ? Enfin, les statisticiens doivent aussi statuer sur le besoin d'espace nécessaire à chaque membre du ménage : faut-il compter une pièce pour chaque membre du ménage ? Et si non, quels membres du ménage peuvent coucher dans la même pièce ?

Schématiquement, il y avait deux options pour expliquer cette difficulté à produire des chiffres qui « tiennent ». On pouvait penser qu'il s'agissait de simples problèmes techniques, qui tenaient au fait que l'appareil catégoriel du recensement n'était pas encore suffisamment élaboré, ou mal appliqué sur le terrain. Une deuxième option consistait à penser que ces incertitudes étaient plus fondamentalement liées au fait qu'il n'y avait aucune « convention d'équivalence » permettant de penser le logement et ses occupants comme des entités stables, commensurables. C'est la distinction, importante, qu'Alain Desrosières fait entre « mesure » et « quantification ». Adopter cette perspective voulait dire explorer ce qui s'est passé pour que ce que Desrosières appelle « les conditions sociales de la quantification » soient réunies.

Dans « une statistique de l'intimité », je défends l'hypothèse que la quantification de l'espace intime devient possible au moment où les normes juridiques du peuplement sont créées et diffusées à travers la société française. La définition statistique de l'espace habitable et du besoin d'espace des habitants est en effet précédée de la réglementation juridique qui se met en place dans l'après-guerre, et qui a été en partie préparée par la jurisprudence des années 1930. Par l'intermédiaire des allocations de logement, des allocations familiales, de la prime de déménagement et des normes de peuplement de l'habitat à loyer modéré, ou encore des subventions pour l'accession à la propriété, qui atteignent les classes populaires aussi bien que les classes moyennes, une vision légitime et unifiée de ce qu'est un logement, une chambre et

un occupant se généralise lentement. C'est sur cette nouvelle vision que peuvent s'appuyer non seulement les concepteurs du recensement, mais aussi l'agent recenseur et le ménage recensé, lorsqu'ils classent la réalité qui est sous leurs yeux. En un mot, la stabilisation de la statistique du logement apparaît comme une conséquence de l'intégration des normes juridiques dans le sens commun, que mobilisent le recensé et l'agent recenseur lors de la passation du questionnaire, réduisant par-là les multiples « erreurs de mesure » et confusions qui hantaient les premières études, et dont cet article rend compte. C'est donc finalement la relation entre le droit et la statistique qui est explorée. En inversant l'idée selon laquelle la statistique publique fournit des appuis pour l'action publique, mon travail montre justement que c'est l'action publique qui doit fournir des appuis pour que la statistique puisse décrire la réalité.

En dernier lieu, à travers le prisme du surpeuplement, c'est aussi un pan de l'histoire de la vie privée que ce travail éclaire, en saisissant la période où l'intimité dans le logement est mise en chiffres, précisément parce qu'elle accède alors au statut de problème public à traiter, et donc à quantifier. La généalogie de cette catégorie statistique très ténue, de cet incroyable effort de produire des chiffres sur l'inimité des ménages, renvoie évidemment à l'histoire plus large de l'État providence français qui, pendant les « trente glorieuses », portait le projet et l'espoir d'un logement digne pour toute la population, d'un minimum d'espace à soi, tout comme celui d'un salaire minimum. Une histoire dont nous sommes sans doute sortis, mais dont nous avons hérité un certain nombre de catégories de pensées, des sensibilités et des souffrances ; c'est pourquoi revenir sur le moment où nos évidences se sont formées et ont été stabilisées, en mobilisant les concepts et les outils qu'Alain Desrosières nous a laissés, m'a semblé nécessaire.

Je remercie vivement la Société française de statistique et le jury du prix Alain Desrosières d'avoir récompensé ce travail.

Légitimer par les nombres, à propos d'une autre fonction des outils quantitatifs



Émilien RUIZ

Historien, maître de conférences à l'université de Lille

Combien y a-t-il de fonctionnaires dans chaque ministère ? L'émergence après la deuxième guerre mondiale de nouveaux outils de mesure des effectifs de la fonction publique met en évidence un rôle souvent méconnu des chiffres. Un travail statistique peut aussi aider le service producteur à occuper une place dans un paysage administratif pas encore figé.

Introduction

Dans l'un des derniers textes qu'il publia, Alain Desrosières résumait ainsi l'un des principaux *leitmotivs* de son œuvre : « les outils quantitatifs ne sont pas seulement des outils de preuve, utilisés par les scientifiques pour étayer leurs arguments, mais aussi des outils de *coordination*, ou de *gouvernement*¹. »

De fait, l'un des apports majeurs du paradigme du « gouvernement par les nombres » a été de permettre une forme de démystification de la neutralité et de l'objectivité intrinsèque des statistiques publiques. Il ne faudrait toutefois pas tomber dans le travers opposé consistant à ne chercher que la volonté de contrôle derrière toute production quantitative ministérielle, au risque de manquer d'autres motivations ou usages.

À partir d'une analyse des conditions de production du premier recensement des agents des services publics en France, l'objet de cet article est de montrer qu'une troisième fonction peut parfois être assignée aux quantifications officielles : celle d'être un outil de légitimation institutionnelle de leurs commanditaires et producteurs.

Quantifier les agents des services publics en 1947

Entre 1946 et 1949, c'est le jeune Institut national des statistiques et des études économiques (Insee) qui mena le premier véritable recensement des agents publics. Amorcée sous la forme d'un « dénombrement » en 1946² cette initiative fut entérinée par la réalisation d'un « recensement général » en 1947, dont les résultats furent publiés en 1949³.

Mené à la demande de la direction de la Fonction publique (ci-après DFP), le recensement fut réalisé principalement à partir des informations relatives aux traitements. En janvier 1947, des imprimés furent mis en place par l'Imprimerie nationale et envoyés aux trésoriers payeurs

1. Alain Desrosières (2012) « Est-il bon, est-il méchant ? Le rôle du nombre dans le gouvernement de la cité néolibérale », *Nouvelles perspectives en sciences sociales*, 2012, vol. 7, no 2, p. 264 - souligné dans le texte.
2. Direction de la Fonction publique et INSEE, *Résultats du dénombrement des agents des services publics effectué en Janvier 1946 (fonctionnaires civils)*, Paris, Imprimerie nationale, 1946
3. INSEE, *Recensement général des agents des services publics*, Paris, Imprimerie nationale & PUF, 1949

généraux (TPG), aux Préfets pour les services départementaux et communaux, et directement à certains ministères (Armée, PTT, Colonies, Affaires étrangères pour les agents à l'étranger, Affaires allemandes et autrichiennes)⁴. La quantification du nombre des agents des services publics devait ainsi être réalisée à partir des « éléments retenus pour le calcul et la mise en paiement des rémunérations allouées aux dits fonctionnaires et agents au titre du mois de mars »⁵. Il s'agissait là, pour le directeur de la Fonction publique, du meilleur moyen pour « obtenir des renseignements précis et détaillés et dont l'exactitude ne soit pas contestable »⁶.

Après le test de 1946, c'est la première fois qu'un recensement des agents publics fut réalisé par un organisme public de statistique à la demande d'une direction ministérielle. Dès l'introduction de la brochure, l'initiative était d'ailleurs présentée comme inédite car « *avant 1941, aucun recensement général des fonctionnaires n'avait été entrepris en vue de réunir* » tant des informations relatives aux effectifs et à leurs rémunérations que des éléments dont la connaissance était jugée « *indispensable pour mener à bien toute étude concernant les fonctionnaires* » (sexe, âge, nombre d'enfants...). Pourtant, dans la brochure publiée en 1949, comme il l'avait fait en 1946, l'Insee proposait une « *évolution des effectifs des fonctionnaires et agents de l'État depuis 1914* ». Les informations chiffrées relatives aux années 1914, 1922, 1936 et, parfois, 1941 étaient en effet citées comme les résultats de « *recensements précédents* ».

Or, aucune des informations ainsi intégrées à la rétrospective de l'évolution des effectifs de l'État n'avait été produite par un recensement. Les données de 1914 et 1922 correspondaient en effet aux chiffres cités dans un rapport déposé à la Chambre en septembre 1921 au nom de la commission des finances. Pour 1914 il s'agissait du résultat d'une enquête budgétaire réalisée par le rapporteur au Budget et pour 1922 il s'agissait de prévisions ministérielles⁷. Les données 1936 et 1941 étaient quant à elles tirées d'une brochure de 1944⁸ à partir d'une enquête budgétaire réalisée à la demande d'Yves Bouthillier en 1941. Il avait alors été demandé aux ministères de fournir des chiffres pour 1941 et 1936, sans que l'on sache très bien comment ils furent établis⁹.

Désigner ces informations comme étant issues de « *recensements précédents* » ne fut pas sans conséquences. De la sorte, l'Insee en faisait des « *chiffres officiels* » et, surtout, des données comparables. Ainsi, un article de *La tribune économique* s'appuya sur la brochure issue du dénombrement de 1946 pour formaliser l'évolution des effectifs depuis le Front populaire dans un graphique dont la légende vantait l'exactitude de la comparaison puisque « *les recensements des trois années citées [avaient] porté [sur] les mêmes catégories de fonctionnaires*¹⁰. » La méprise était d'autant plus prévisible que, ni dans la brochure issue du dénombrement de 1946, ni dans celle du recensement de 1947, les sources des statistiques concernant les années 1914-1941 ne furent indiquées.

4. Lettre n°248 / 2-E de Francis-Louis Closon, directeur général de l'INSEE, aux directions régionales de l'INSEE et à la direction de la Fonction publique, instruction n°78.448 sur le recensement des agents des services publics en 1947, Paris, 30 janvier 1947, CAEF, Fonds Statistiques et études financières, INSEE, B55439.

5. Lettre n°1119 F du directeur de la Comptabilité publique aux trésoriers payeurs généraux, Paris, 14 février 1947, CAEF, Fonds Statistiques et études financières, INSEE, B55439.

6. Lettre du secrétaire général du gouvernement [en-tête de la direction de la Fonction publique] à Monsieur le ministre de..., Paris, 30 octobre 1946, AN, F60 980.

7. Rapport n°3160 fait au nom de la commission des finances chargée d'examiner le projet de loi portant fixation du budget général de l'exercice 1922 par M. Maurice Bokanowski, Annexe au procès-verbal de la 2e séance du 12 juillet 1921 [déposé à la Chambre le 30 septembre 1921].

8. SNS, *Contribution à la statistique des fonctionnaires de l'État (effectifs, traitements, pensions de retraites)*, Paris, Imprimerie nationale, 1944., p. 49 et 51.

9. Lettre n°V/2904 du ministre secrétaire d'État à l'Économie nationale et aux Finances à messieurs les ministres, secrétaires d'État et secrétaires généraux, 25 octobre 1941, CAEF, Fonds Budget, B11856/60.

10. J.M. Gernigon, « L'administration française est malade. Y a-t-il un remède ? », *La Tribune économique, financière, sociale, politique*, 1^{er} février 1946, Dossiers de presse de la FNSP, La fonction publique sous la IV^e République, 1944-1946, France 114 mf1.

Rivalités administratives et quantifications concurrentes

Dès lors, une question s'impose au chercheur : pourquoi les statisticiens de l'Insee nommèrent-ils ces enquêtes précédentes des « *recensements* » en éludant leur source réelle ?

Peut-être y eut-il une part de facilité de langage à faire de « *recensement* » un synonyme de toute « *comptabilisation statistique* », mais les statisticiens de l'Insee étaient généralement plus attentifs à bien distinguer un dénombrement budgétaire par exemple, d'un recensement des effectifs réels... D'autant plus qu'ils réclamaient, depuis de nombreuses années, l'application d'une méthode de comptage des agents plus fiable que celles qui avaient été mises en œuvre jusqu'ici. D'ailleurs, les concepteurs de l'enquête réclamée en 1941 par Y. Bouthillier ne s'y trompaient pas. Dans l'introduction – dont on ne sait si elle fut rédigée par Gaston Duon, chargé de l'étude en question, ou par Raymond Rivet, qui la dirigeait –, il était stipulé que cette enquête ne présentait « pas tous les avantages des fiches individuelles préconisées dès 1905 par la commission compétente du Conseil supérieur de la statistique ». Ces précautions prises, Henri Bunle ajoutait toutefois qu'il s'agissait « d'une approximation suffisante pour répondre aux besoins intérieurs des services publics, à qui ce tirage [était] strictement et confidentiellement destiné »¹¹.

Éluder le caractère nettement plus approximatif des données antérieures à celle du recensement proprement dit relevait plus certainement d'une volonté de légitimer la comparaison diachronique ainsi proposée. Ce qui permet d'énoncer cette hypothèse, ce sont les transformations administratives qui, au sortir de la Seconde Guerre mondiale, occasionnèrent certaines modifications dans le champ des compétences relatives à la Fonction publique. En effet, si l'Insee fut amené à gommer les biais de sa rétrospective, c'est parce qu'au sein de l'État, il n'était pas le seul producteur de données chiffrées sur le sujet. En effet, en parallèle de l'élaboration du dénombrement de 1946, puis du recensement de 1947, le ministère des Finances produisit un « *inventaire de la situation financière* » qui traitait aussi de l'évolution des effectifs depuis 1913. Soulignant la constance de l'augmentation du nombre des agents de l'État au cours de cette période, les auteurs précisait qu'ils avaient entrepris d'en retracer l'évolution « *malgré l'insuffisance de la documentation statistique* » en ce domaine, « *fait connu depuis longtemps et toujours critiqué* ». Or, si c'est le même constat qui avait conduit l'Insee et la direction de la Fonction publique à tenter un dénombrement sur la base des traitements, les auteurs de l'inventaire considéraient que « *par les conditions mêmes* » de sa réalisation, l'opération contenait « *nécessairement des erreurs ou des insuffisances* ». La direction du Budget préféra comptabiliser elle-même les « *emplois budgétaires* »¹². Cette opération – que le ministère des Finances avait été quasiment incapable de réaliser depuis le début du siècle – fut sans aucun doute facilitée par les rivalités entre le Budget et la Fonction publique.

En effet, née de l'ordonnance du 9 octobre 1945, la DFP fut d'emblée chargée de « *préparer les éléments d'une politique d'ensemble de la fonction que publique* », et « *d'établir ou de faire établir une documentation et des statistiques d'ensemble [la] concernant* ». Cela explique donc, en premier lieu, le rôle que joua la DFP dans la réalisation et la publication du dénombrement de 1946, ainsi que sa volonté de voir réaliser un véritable recensement en 1947. Mais les conditions de cette création permettent aussi d'expliquer la publication, au même moment, de l'inventaire des Finances. Si, toujours en vertu de l'ordonnance de 1945, elles gardaient un droit de regard sur « *tous les projets élaborés ou examinés par la présidence du gouvernement* » en ce qui concerne les règles statutaires, les rémunérations, le régime de prévoyance des personnels et, plus généralement, tout texte ayant vocation à avoir des « *répercussions financières* »¹³, elles n'étaient

11. SNS, *Contribution à la statistique des fonctionnaires de l'État*, op. cit., p. 4 et 49.

12. Ministère des Finances, *Inventaire de la situation financière (1913-1946)*, établi sur les instructions de M. Robert Schuman, ministre des Finances, Paris, Imprimerie nationale, 1946, Annexe I-Q, p. 349-353.

13. Ordonnance n°452283 du 9 octobre 1945 relative à la formation, au recrutement et au statut de certaines catégories de fonctionnaires, et instituant une direction de la Fonction publique et un conseil permanent de l'administration civile, Journal Officiel, 10 octobre 1945, p. 6378-6380 (ART. 15 - citations p. 6380).

plus chargées de la production des chiffres officiels sur la question. D'autant moins que l'Insee ne relevait pas de la tutelle du ministère des Finances mais de celle de l'Économie nationale¹⁴. En l'occurrence, alors même que les années 1946-1949 furent le théâtre de nombreuses politiques de suppressions d'emplois, ce n'est pas par les chiffres de l'Insee que le Parlement fut éclairé sur l'évolution des effectifs, mais par les mises à jour de l'inventaire¹⁵.

Une opération de légitimation institutionnelle

Si les quantifications produites par l'Insee et la DFP ne furent pas directement des outils de gouvernement, ils ne furent toutefois pas sans utilité pour leurs producteurs. Le dénombrement puis le recensement jouèrent un rôle important dans le processus de légitimation des deux nouveaux organismes.

Fondé en 1946, l'Insee avait d'emblée dû faire face à de nombreuses hostilités. Béatrice Touchelay a bien montré les difficultés budgétaires auxquelles l'organisme fut confronté dès sa naissance¹⁶. En outre, l'Institut fit régulièrement l'objet de procès en inutilité ou en illégitimité.

Deux exemples permettent d'en prendre la mesure. Alors que les politiques d'économies étaient à l'ordre du jour, le 1^{er} juillet 1946, François-Didier Gregh, directeur du Budget, informa Francis-Louis Closon, directeur de l'INSEE, des jugements sans équivoque portés par un inspecteur des Finances « *à l'esprit vif et délié* » sur les travaux du « *Service des Statistiques* ». Il affirmait ainsi que, « *muni d'un puissant matériel et de crédits trop importants* » l'Insee accepterait et susciterait même « *des besognes dont l'utilité ne paraît pas immédiate* » uniquement pour « *occuper* » les agents de l'Institut. Convaincu que les cadres du service étaient « *beaucoup trop nombreux* » et le rendement individuel des agents « *médiocre* », l'inspecteur des Finances considérait qu'il s'agissait là du « *type même du service où des coupes sombres seraient indispensables* », d'autant plus que les « *Directeurs régionaux [étaient] sans valeur ; il n'y a que les machines qui en [avaient] une...* »¹⁷. Les critiques ne venaient pas uniquement des Finances. Directeur de l'Ined – institut né lui aussi après la guerre dans un subtil mélange de rupture et de continuité avec les structures de l'État français – Alfred Sauvy ne cessa de fustiger l'inefficacité de l'Insee. En novembre 1947 par exemple, dans un article sur le rôle de la statistique dans les processus de décision, il rappelait que l'Insee était né en partie des développements donnés à la statistique publique « *sous le gouvernement de Vichy* » et qu'en dépit de son caractère « *quelque peu monstrueux* », il s'en fallait « *malheureusement de beaucoup que le rendement de l'organisme ait augmenté dans les mêmes proportions* » que son matériel et ses effectifs¹⁸.

Francis-Louis Closon, soucieux de l'image de l'organisme et de la défense de sa légitimité dans l'État, prenait systématiquement soin de répondre aux attaques, mais il prenait aussi souvent les devants. On trouve ainsi, parmi les archives conservées au CAEF, de nombreux documents datant des premières années d'existence de l'Insee qui visaient à présenter ses origines, son fonctionnement et le bilan de ses activités. En pleine période de refondation d'une politique de la Fonction publique (non seulement par la création de la DFP, mais aussi par l'adoption du statut et la mise en place de la grille indiciaire¹⁹) il présentait le recensement général des agents des services publics comme un témoignage de l'utilité des travaux de l'Institut. Peu importait,

14. Articles 32 et 33 de la loi n°46-854 du 27 avril 1946, Journal officiel, 1er mai 1946, p. 3630.

15. Ministère des Finances et des Affaires économiques, Statistiques et documents pour la mise à jour de l'inventaire de la situation financière (publié en 1946). Annexe au projet de loi de finances pour l'exercice 1950 (déposé sous le n°8336 à la séance du 10 novembre 1949 de l'Assemblée nationale), Paris, Imprimerie nationale, 1949.

16. Béatrice Touchelay, L'INSEE des origines à 1961 : évolution et relation avec la réalité économique, politique et sociale, Thèse de doctorat en histoire économique, Université Paris 12, 1993, p. 222, 262-263 et 275.

17. Lettre n°8-750 du directeur du Budget au directeur de l'INSEE, Paris, le 1er juillet 1946, CAEF, Fonds Statistiques et études économiques, B55423.

18. Alfred Sauvy, « Gouverner, c'est prévoir : le rôle de la statistique », Revue de défense nationale, vol. 3, novembre 1947, p. 624

19. Jeanne Siwek-Pouydesseau, « La genèse du statut de 1946 et de la grille indiciaire », La revue administrative, vol. 48, 1995, p. 35-40.

finalement, qu'il soit vraiment utilisé : l'organisme statistique avait fait sa part en donnant « *au gouvernement les éléments dont il avait besoin* »²⁰.

De son côté la DFP fit aussi de ces initiatives un témoignage de l'utilité et de l'efficacité de ses services. Roger Grégoire, son directeur jusqu'en 1954, citait ainsi le dénombrement de 1946 et le recensement de 1947 dans tous ses rapports d'activité. En 1951 par exemple, il signalait que pour « satisfaire à ses obligations tendant à l'établissement d'une documentation relative aux personnels de l'État », la DFP avait « pris l'initiative » de faire réaliser « selon ses directives » trois recensements détaillés du personnel en février 1946, mars 1947 et avril 1950²¹.

Conclusion

À la fin des années 1940, la quantification des effectifs publics fut ainsi principalement un outil de légitimation institutionnelle, tant pour l'Insee que pour son alliée de circonstance, la DFP.

Si l'on ne saurait généraliser à toute production quantitative cette conclusion relative à un cas particulier, il faut toutefois noter que l'on retrouve cette fonction de légitimation à chaque tournant de l'histoire du comptage des agents publics en France. Victor Turquan tenta ainsi, par un premier « essai de recensement » en 1896, tant de renforcer sa position dans les cercles libéraux anti-étatistes que de légitimer son expertise statisticienne. Dix ans plus tard, le petit réseau de membres actifs de la Société de statistique de Paris qui inventa la « statistique des fonctionnaires » renforça tant sa position que celle de la statistique publique dans l'État²². Or, aucune des données qu'ils produisirent ne fit l'objet d'un usage parlementaire ou ministériel. Sur ce terrain la quantification fut toujours plus un outil de légitimation qu'un instrument de gouvernement. Il ne faudrait pas pour autant en conclure que l'on n'aurait jamais gouverné avec ces chiffres. En effet, si le recensement de 1947 ne fut pas mis au service de la politique de réduction des effectifs, et bien que cet usage n'ait pas été anticipé par ses commanditaires, ses résultats provisoires contribuèrent, entre 1948 et 1950, à la mise à l'agenda gouvernemental du « *problème de l'auxiliarat* » par la DFP qui, contre la direction du Budget, préconisait une facilitation des titularisations²³.

Dès lors, tandis que, dans le texte cité en introduction, Alain Desrosières proposait une typologie des quantifications associées à cinq formes d'État²⁴, le présent article voudrait modestement appeler à prolonger l'analyse. Les opérations statistiques pouvant être non seulement des outils de preuve et de gouvernement mais aussi des instruments de légitimation, il apparaît désormais nécessaire de déplacer la focale à l'intérieur de l'État. Traversé de rapports de forces (administratifs et/ou politiques, entre ministères et/ou entre corps de fonctionnaires, etc.) l'État est aussi un espace de légitimation où la capacité (ou non) à compter et où la concurrence des quantifications jouent parfois un rôle fondamental.

20. Marcel Brichler, « Le recensement général des agents des services publics effectué en 1947 », *Journal de la société de statistique de Paris*, vol. 90, 1949, p. 263.

21. Rapport sur le fonctionnement de la direction de la Fonction publique. Bilan d'activités 1954-1951 et perspectives d'avenir, par Roger Grégoire, Paris, 30 juin 1951, AN, 20040366-ART1.

22. Sur ces deux moments, je me permets de renvoyer à Émilien Ruiz, « Statistique des fonctionnaires et histoire de l'État par ses effectifs (France, 1850-1950) », *Statistique et société*, 4(1), mai 2016, p. 45-49 et « Compter. L'invention de la statistique des fonctionnaires en France (années 1890 - années 1930) », *Sociologie du travail*, 52(2), avril-juin 2010, p. 212-233.

23. PV de la séance du jeudi 28 octobre 1948, conseil supérieur de la Fonction publique, Paris, le 16 novembre 1948, AN, 20030468-ART2 ; Loi n°50-400 du 3 avril 1950 portant autorisation de transformation d'emplois et réforme de l'auxiliarat, *Journal officiel*, 6 avril 1950, p. 3704-3705 ; Loi n°50-400 du 3 avril 1950 portant autorisation de transformation d'emplois et réforme de l'auxiliarat, *Journal officiel*, 6 avril 1950, p. 3704-3705.

24. Alain Desrosières, « Est-il bon, est-il méchant ? », op. cit. – texte repris dans *Prouver et gouverner. Une analyse politique des statistiques publiques*, Paris, La Découverte, 2014, p. 33-59.

Addiction aux jeux de hasard et d'argent en ligne : comment mettre en place une prévention basée sur les preuves grâce aux Big Data ?



Amandine LUQUIENS

Psychiatre addictologue, Hôpital Paul Brousse, Villejuif

L'addiction aux jeux d'argent en ligne est devenue un grave problème de santé publique. Pour y faire face, il est possible d'utiliser les données laissées par les joueurs sur les sites de jeu. Ces données permettent de repérer l'apparition de comportements à risque, et de proposer aux intéressés des outils favorisant un jeu responsable. Peut-être faut-il, pour bénéficier d'une prévention efficace, accepter les inconvénients d'une surveillance systématisée.

Big Data et prévention

L'engouement sociétal actuel pour le Big Data n'épargne pas la médecine, où l'on oscille entre fascination, fantasmes, vraies avancées et perspectives. Il faut tout d'abord discerner la e-santé et la digitalisation. La e-santé correspond à l'intégration de technologies d'information et de communication dans les soins. La digitalisation correspond aux processus sous-jacents d'innovation et de changement, en particulier le développement de logiciels et de matériel visant in fine l'aide à la décision médicale¹. La masse des données exploitables en médecine est en augmentation exponentielle comme dans les autres secteurs. Différents types de données déjà utilisées en médecine ont été recensés, qu'elles viennent du dossier patient et du soin ou d'autres sources : (a) données biologiques, (b) données d'environnement, (c) dossiers médicaux dématérialisés, (d) enregistrements via des objets connectés (exemples : géo localisation, mais aussi enregistrements d'équipements médicaux comme ceux d'un service de réanimation) (e) publications sur les réseaux sociaux, (f) données issues des téléphones portables². La diversité des données disponibles permet donc aussi bien de suivre à distance le rythme cardiaque d'un patient portant un pace maker, que de dépister un épisode dépressif à travers la publication de posts sur Twitter. Pour ce dernier exemple, il s'agit donc de dépister de manière indirecte un trouble mental chez un individu ne l'ayant pas forcément sollicité, à partir d'un versant de son comportement, les échanges sur un réseau social. D'autres données issues d'un comportement sur internet permettent de faire de la prévention. La pratique des jeux de hasard et d'argent en ligne en fait partie. Mais la nature des données disponibles constitue une configuration exceptionnelle : les données permettent de décrire précisément et dans son intégralité un comportement potentiellement problématique. On a donc non pas une reconstitution indirecte du comportement, mais sa lecture directe.

1. Gehring H., Rackebrandt K. & Imhoff M. 2018. E-Health and reality - what are we facing in patient care? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.

Keasberry J., Scott, I. A., Sullivan, C., Staib, A. & Ashby, R. 2017. Going digital: a narrative overview of the clinical and organisational impacts of eHealth technologies in hospital practice. Aust Health Rev, 41, 646-664.

2. Mooney, S. J. & Pejaver, V. 2017. Big Data in Public Health: Terminology, Machine Learning, and Privacy. Annu Rev Public Health.

Données de compte joueur et prévention du jeu problématique

La pratique des jeux de hasard et d'argent en ligne en France est régulée par l'Autorité de régulation des jeux en ligne (ARJEL)³, qui demande aux opérateurs de jeu l'enregistrement de nombreuses données de jeu issues des comptes joueurs. L'ARJEL stocke elle-même ces données sur un serveur frontal⁴, et est autorisée à les utiliser dans une optique de protection du joueur et de prévention de l'addiction depuis 2016. Par ailleurs, ces données peuvent être utilisées par chaque opérateur de jeu dans une optique également de protection du joueur et de prévention de l'addiction. Elles constituent l'opportunité de dépister les joueurs problématiques et de réaliser de la prévention ciblée. En effet, si la prévalence du jeu problématique est en hausse et a été estimée en 2014 à 1,9 % de la population générale française de 15 à 75 ans⁵, l'accès aux soins pour ces personnes est très médiocre. Des initiatives de prévention primaire (éviter l'entrée dans la maladie) et secondaire (favoriser l'accès aux soins et limiter les dommages liés à la maladie) sont donc nécessaires. Deux outils utilisant des algorithmes ont été développés dans le passé pour proposer aux joueurs de repérer un jeu problématique notamment à travers un changement dans leur comportement de jeu. Le plus utilisé, Playscann, a été développé à partir de données de comptes joueurs, mais en toute opacité, sans validation par rapport à une échelle classique de dépistage clinique. Un premier outil de dépistage des joueurs problématiques de poker en ligne a été développé et validé par notre équipe à partir d'un modèle statistique de régression expliquant le score clinique dans l'index canadien du jeu excessif (ICJE) recueilli en ligne, par les données de comportement de jeu enregistrées en routine par l'opérateur de jeu Winamax⁶. Nous avons alors travaillé sur des données agrégées sur le mois. Cette initiative a permis de montrer la faisabilité d'un repérage fiable à partir des données de compte joueur seules, chez des joueurs de poker d'un opérateur. Elle a abouti à une étude fille au sein de l'ARJEL portant sur l'ensemble des activités de jeu en ligne sur les sites français. Cette étude réalisée en 2016 n'a pourtant pas encore eu de traduction en mesure pratique de prévention, probablement du fait d'un paradigme complexe entre liberté de jouer et défiance face au spectre de la surveillance d'une part, argument parapluie porté par les enjeux économiques, et nécessité de mettre en place la prévention ciblée face à une maladie lourde en terme de poids socio-sanitaire, d'autre part.

Au-delà du repérage

Si la mise en place du repérage est complexe car soulevant des questionnements de société, les données de compte joueur permettent aussi de mieux appréhender le processus du jeu problématique, et de mesurer l'efficacité de mesures existantes de jeu responsable, c'est à dire visant à limiter les risques et les dommages des jeux de hasard et d'argent. Plusieurs outils dits de jeu responsables peuvent être utilisés pour aider un joueur à ajuster son comportement de jeu en cas de pratique problématique. Un des outils proposés aux joueurs est l'auto-exclusion, temporaire ou « définitive » (trois ans), ne concernant qu'un opérateur de jeu, ou l'interdiction volontaire de jeu (trois ans également, mais concernant tous les opérateurs de jeu en ligne régulés par l'ARJEL, et les casinos terrestres). Les données disponibles dans la littérature sur les joueurs en ligne s'étant auto-exclus ont montré des résultats peu robustes du fait de problèmes méthodologiques : la majorité des auteurs recrutaient des joueurs volontaires pour participer à l'étude et non l'intégralité des joueurs s'étant auto-exclus, et les interrogeaient *a posteriori* sur leurs motivations. Plusieurs auteurs se sont accordés à dire que les joueurs

3. Assemblée Nationale 2010. Loi n° 2010-476 du 12 mai 2010 relative à l'ouverture à la concurrence et à la régulation du secteur des jeux d'argent et de hasard en ligne- Article 26. In: Sénat, A. N. E. (ed.) NOR: BCFX0904879L.

4. Machine qui reçoit les connexions des utilisateurs

5. Costes J-M., Eroukmanoff V., Richard J-B., Tovar M-L. (2015) Les jeux d'argent et de hasard en France en 2014 *Les notes de l'observatoire des jeux* n°6. Sont classés « problématiques » les joueurs « à risque modéré » (1,5 %) et les joueurs « excessifs » (0,4 %).

6. Luquiens A., Tanguy M. L., Benyamina A., Lagadec M., Aubin H. J. & Reynaud M. (2016). Tracking online poker problem gamblers with player account-based gambling data only. *Int J Methods Psychiatr Res*, 25, 333-342.

s'étant auto- exclus avaient majoritairement une pratique de jeu plus intense que les autres joueurs en ligne sur le même site⁷, mais d'autres ont décrit que seulement une partie d'entre eux présentait un jeu problématique, exploré *a posteriori*⁸. Aucune donnée dans la littérature n'a documenté les motifs d'auto- exclusion au moment de sa mise en œuvre, permettant de lever le biais de rappel et les modifications de comportement à posteriori. A la différence d'études antérieures, et dans la lignée de partenariats précédents entre académiques et opérateurs de jeu (Bwin et « the transparency project »), notre équipe, constituée de professionnels de santé chercheurs (APHP-Inserm) et de mathématiciens du Centre de Mathématiques Appliquées de l'Ecole Polytechnique, travaille sur les données de compte joueur grâce à un partenariat formalisé par une convention laissant clairement la liberté de publication sans droit de regard de l'opérateur de jeu. Les données à notre disposition documentent à la fois les motifs d'auto-exclusion, permettant de repérer les joueurs s'auto-excluant pour un motif évoquant un jeu problématique, et les données de jeu de poker avec une forte granularité, c'est-à-dire un niveau de détail fin de l'information présentée. Le travail sur des données à forte granularité soulève pourtant des questions méthodologiques. En effet, nous sommes face à des mesures répétées dans le temps un grand nombre de fois. D'autre part, le matériau même du jeu est complexe car il est en partie dû au hasard, rendant les variables financières très dispersées : même un joueur pathologique peut gagner sur une main « chanceuse ». La question est donc grossièrement : à quel zoom travailler ? Y a-t-il un vrai intérêt à travailler avec une granularité fine sur des données mettant en jeu en partie le hasard ? Comment gérer les variables à très fort écart type ? Comment prendre en compte l'historique du joueur dans son comportement, à savoir que chaque action de jeu est faite par le joueur en connaissance des précédentes, et que ses prises de décisions sollicitent le fonctionnement cérébral et émotionnel ? Le machine learning se distinguerait de l'approche statistique par le fait qu'il générerait des hypothèses, plutôt que de chercher à les tester. Pourtant, si la masse des données à disposition semble ouvrir ici une boîte de Pandore, le garde-fou revient, comme dans les autres domaines de la médecine et des approches statistiques classiques, à connaître la nature des variables, savoir comment elles ont été obtenues, ce qu'on veut en faire, c'est-à-dire quelle est la finalité pratique de l'analyse.

Enjeux éthiques

En effet, la question « que veut-on faire des données ? » se heurte nécessairement aux enjeux éthiques soulevés par l'exploitation de données issues de comportements en ligne. La balance entre le risque de violation de la vie privée et les intérêts publics doit être faite, et la démarche de l'étude doit nécessairement viser à des bénéfices sociétaux directs ou rapidement applicables⁹. Dans notre cas, plusieurs questions sont soulevées. L'aboutissement des recherches doit-il être un dépistage systématique ou la possibilité de proposer systématiquement un dépistage aux joueurs ? Mais dans ce deuxième cas, à qui et quand faut-il proposer un dépistage ? La faisabilité méthodologique et la création d'un algorithme engage-t-elle éthiquement à son utilisation si elle peut permettre de sauver des vies, sachant le sur-risque suicidaire porté par les joueurs pathologiques ? Posée ainsi, la question reçoit une réponse positive évidente à première vue, qui rencontre le spectre de la « surveillance ». Dans le cas qui nous concerne, la surveillance à visée de prévention est déjà une obligation imposée aux opérateurs de jeu, qui doivent, tous les ans, documenter leurs initiatives de jeu responsable auprès de l'ARJEL dans un rapport dédié, nécessaire pour renouveler leur agrément ; à ceci près que la surveillance est aujourd'hui artisanale, et donc quasi inefficace (quelques dizaines de cas par ans repérés pour certains opérateurs). Toute forme de veille à visée de prévention utilise et collige les données

7. Dragicevic S., Percy C., Kudic A. & Parke J. 2015. A descriptive analysis of demographic and behavioral data from Internet gamblers and those who self-exclude from online gambling platforms. *J Gambl Stud*, 31, 105-32.

8. Griffiths M. D., Wood R. T. & Parke J. 2009. Social responsibility tools in online gambling: a survey of attitudes and behavior among Internet gamblers. *Cyberpsychol Behav*, 12, 413-21.

9. De Lusignan S., Liyanage H., Di Iorio C. T., Chan T. & Liaw S. T. 2016. Using routinely collected health data for surveillance, quality improvement and research: Framework and key questions to assess ethics, privacy and data access. *J Innov Health Inform*, 22, 426-32.

Mittelstadt B. D. & Floridi L. 2016. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Sci Eng Ethics*, 22, 303-41.

issues des prises en charge individuelles, pour avoir un tableau global, épidémiologique. Ainsi, le pic épidémique de grippe est-il surveillé chaque année en relevant le nombre de cas dépistés et pris en charge. La « surveillance » systématisée est ainsi prévention systématisée si elle est au service de la santé. Paradoxalement, dans le cadre légal actuel, le risque n'est pas à l'utilisation par les opérateurs de ce type d'algorithme à mauvais escient, mais plutôt à leur réticence à le mettre en place, pour ne pas détourner les joueurs de leur pratique. Le deuxième risque est plutôt un risque par omission, c'est-à-dire la non mise en place par les autorités de régulation ou de santé, possiblement pour le même motif économique, alors que l'État bénéficie largement des recettes des jeux. Enfin, l'exemple de notre partenariat, comme d'autres à l'étranger, illustre la faisabilité d'un partenariat académique-privé sur l'exploitation de données. La nécessité d'une convention qui protège les académiques sur la publication sans réserve des résultats s'impose. Il apparaît recommandable d'éviter tout financement direct par l'opérateur afin d'éviter une position de conflit d'intérêt. Dans un domaine où l'État lui-même est en conflit d'intérêt, les professionnels de santé devraient pouvoir se positionner en sentinelles innovantes afin de mobiliser l'État sur des initiatives de prévention dont ils auront démontré la faisabilité et l'efficacité.

Protection des données personnelles et recherche : constats et perspectives d'évolution



Sophie VULLIET-TAVERNIER

Directrice des relations avec les publics et la recherche -
Commission Nationale de l'Informatique et des Libertés (Cnil)

La recherche moderne a souvent besoin d'utiliser des données personnelles dont la confidentialité est protégée par la loi. Cette contradiction peut être surmontée. La Cnil encourage les dispositifs qui permettent aux chercheurs de travailler tout en fournissant les garanties indispensables.

Protection des données personnelles et recherche sont-elles inconciliables ? Non assurément. Mais l'application de la loi informatique et Libertés dans le monde de la recherche, c'est encore beaucoup de malentendus, d'idées reçues et d'incompréhensions à lever... Ce sont à l'évidence des notions à clarifier, une démarche à expliquer, une meilleure prise en compte des besoins spécifiques des chercheurs...

Ainsi, la frontière entre données personnelles, anonymat et pseudonymat n'est pas toujours perçue de façon claire, nonobstant les conseils et avis émis sur le sujet (cf. par exemple l'avis du groupe européen des autorités de protection des données mai 2014¹).

De même la démarche d'analyse « informatique et libertés » qui doit accompagner tout projet de recherche portant sur des données personnelles et qui s'apparente pourtant à une démarche éthique, n'est souvent pas ressentie comme telle, les chercheurs n'y voyant parfois qu'une procédure administrative de plus (il faut déclarer à la Cnil...).

Enfin, alors même que les spécificités de la recherche sont d'une certaine façon prises en compte par le corpus juridique de la protection des données, tant au niveau européen que national, leur caractère peu visible et complexe ne permet pas toujours leur réelle appréhension par les chercheurs et ce, quelle que soit la discipline.

Face à ces constats, quelles ont été et quelles sont les réponses apportées par la Commission Nationale de l'Informatique et des Libertés (Cnil) et le nouveau cadre de régulation qui se dessine avec l'entrée en application, en mai 2018 du règlement européen sur la protection des données personnelles, le RGDP ? Quels sont, aujourd'hui et demain, les nouveaux enjeux pour la protection des données dans le domaine de la recherche ?

La Cnil et la recherche

Depuis sa création, à la fin des années 70, la Cnil a beaucoup évolué...
Au-delà des chiffres d'activité (cf. encadré sur la Cnil en bref) et de l'exercice des missions

1. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_fr.pdf

traditionnelles de contrôle a priori (via les formalités déclaratives) et a posteriori (via les contrôles et les sanctions), la Cnil s'est donnée aujourd'hui pour objectif de :

- mieux répondre aux besoins des usagers via notamment son service besoin d'aide, ses 500 FAQ et sa réponse en ligne, via aussi ses tutoriels et conseils pratiques disponibles sur ses sites www.cnil.fr et www.educnum.fr et ses fils Twitter et Facebook...
- accompagner la conformité au moyen des labels (+ de 90), des packs de conformité, du service dédié aux correspondants informatique et libertés (CIL)
- avoir une politique de contrôles et de sanctions plus ciblée en particulier en fonction des enjeux liés à la mondialisation du numérique et des sujets de préoccupation et d'inquiétudes remontant du public (spam, droit à l'oubli, vidéosurveillance...)
- accompagner l'innovation et la recherche.

La Cnil, la recherche et les statistiques : une longue histoire :

La Cnil a, très tôt, entretenu des relations avec le monde de la recherche, relations certes parfois tendues mais toujours passionnantes et riches d'enseignements.

Cette prise en compte des spécificités de la recherche s'est tout d'abord marquée dans l'organisation même de la Cnil.

Un membre de la Cnil est ainsi spécifiquement en charge des questions de recherche et des statistiques. Au sein des équipes de la Cnil, une direction² en lien avec d'autres services (et notamment le pôle innovation et prospective) assure l'interface avec les chercheurs qui sollicitent la Cnil. Par ailleurs, le service santé suit particulièrement les projets dans le domaine de la recherche médicale.

Enfin, il doit être relevé que la composition actuelle de la Cnil fait une assez large part au monde académique puisque 4 commissaires sur 17 en sont issus.

Sur le fond, la commission a, quand elle l'estimait nécessaire au regard des enjeux soulevés par telle ou telle problématique de recherche, créé en son sein ou avec les acteurs concernés des groupes de travail permettant ainsi une réelle concertation : ce fut le cas, par exemple, pour les registres épidémiologiques sur le cancer dans les années 80, la recherche sur le SIDA dans les années 90, les « statistiques ethniques » dans les années 2000, l'open data au début des années 2010...les enjeux éthiques des algorithmes et de l'IA, les objets connectés, les smart cities, ...

Sur d'autres grands dossiers « historiques », le NIR, le recensement, les données génétiques,... elle n'a pas hésité à faire appel à des expertises extérieures (experts en statistiques, en génétique, comité national d'éthique...) pour mûrir sa réflexion.

La Cnil et la recherche aujourd'hui :

Aujourd'hui, la Cnil et la recherche, ce sont des partenariats : avec l'Inria³, l'Institut Mines-Télécoms, la Conférence des Présidents d'Universités (et le réseau « Supcil » des correspondants informatique et libertés dans les universités⁴), la Conférence des Grandes Ecoles...

C'est aussi l'accompagnement de la recherche, et en particulier, le conseil sur les projets de recherche (Big Data, web social...), la participation au comité du secret statistique du CNIS, à certains comités de l'ANR... et la conception, en cours, d'un guide pratique, vade-mecum pour

2. La direction des relations avec les publics et la recherche.
3. Institut national de recherche en informatique et automatique
4. <https://groupes.renater.fr/wiki/supcil/>

les chercheurs, en liaison notamment avec le CNRS.

C'est enfin la promotion de la recherche sur la protection des données. Ainsi, la Cnil est partenaire de la chaire de recherche de l'institut mines-télécoms sur « Valeurs et politiques des informations personnelles⁵ ». Elle participe également à des travaux de recherche (exemple du projet mobilitics sur les smartphones avec l'Inria).

Et depuis deux ans elle organise avec l'Inria un Prix européen « privacy » visant à récompenser un article scientifique paru sur cette thématique.

La contribution de la Cnil à la réflexion éthique et prospective

Même si la Cnil a toujours intégré la dimension éthique dans ses réflexions, la loi pour la République numérique⁶ a élargi son champ d'intervention en lui confiant expressément la mission de conduire une réflexion sur les problèmes éthiques et les questions de société soulevés par l'évolution des technologies numériques.

Cette mission a été engagée avec le lancement début 2017 d'un débat sur les enjeux éthiques des algorithmes qui s'est récemment conclu par la publication, en décembre 2017, d'un rapport⁷ intitulé « comment permettre à l'homme de garder la main ? ».

La publication de cet ouvrage s'inscrit dans une politique volontariste de la Cnil visant à assurer une plus grande visibilité à ses réflexions : c'est ainsi qu'a été créé en 2017, en collaboration avec la documentation française une nouvelle collection « le point Cnil » dont le premier numéro est consacré aux données génétiques.

Par ailleurs, la Commission participe aux travaux conduits respectivement par la Commission de réflexion sur l'éthique de la recherche en sciences du numérique (Cerna) de l'alliance Allistene et par le comité d'éthique de Teralab (GENES – institut Mines Telecoms)⁸.

Enfin, sur un plan plus prospectif, a été créé, en 2012, un Comité de la prospective, réunissant membres de la Cnil et personnalités extérieures pour développer la réflexion prospective sur les nouveaux enjeux sociétaux à venir à l'aune des développements technologiques et des usages numériques.

Ce nouveau dimensionnement s'est également concrétisé par l'ouverture d'un laboratoire d'innovation numérique au sein de la Cnil, le labo linc⁹, d'un site dédié <http://linc.cnil.fr/> et des publications : les cahiers IP dont le dernier numéro est consacré aux smart cities¹⁰.

Les actions ainsi entreprises sont autant de signaux en direction de la recherche et qui doivent être poursuivies et amplifiées à l'aune de la nouvelle régulation européenne.

La prise en compte des spécificités de la recherche dans les lois de protection des données

La loi informatique et Libertés, reprenant en cela les dispositions de la directive européenne de 1995¹¹ reconnaît expressément la possibilité de traiter ultérieurement les données initialement

5. <https://cvpip.wp.imt.fr/accueil/>

6. Loi du 7 octobre 2016.

7. <https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de>

8. <https://www.teralab-datascience.fr/fr/>

9. <https://linc.cnil.fr/propos-de-linc>

10. <https://linc.cnil.fr/la-plateforme-dune-ville-explore-les-enjeux-de-la-smart-city>

11. Directive 95/46CE du 24 octobre 1995 remplacée par le RGPD

recueillies pour une finalité donnée (par ex de gestion) à des fins statistique ou de recherche scientifique ou historique (art 6). De même les données peuvent être conservées en vue d'être traitées à de telles fins (art 36). En outre, des dérogations à l'obligation d'information sont également prévues (art. 32, III). Enfin, le droit d'accès n'est pas applicable aux données qui sont conservées sous une forme excluant manifestement tout risque d'atteinte à la vie privée des personnes concernées et ce sous contrôle de la Cnil.

Le recueil et le traitement, à des fins de recherche ou de statistiques, des données dites sensibles (origine raciale ou ethnique, opinions politiques, religieuses, philosophiques, appartenance syndicale, données de santé, vie sexuelle) est également possible sous certaines conditions.

De façon plus spécifique, le besoin d'un accès plus large des chercheurs aux bases de données administratives a progressivement été pris en compte à la suite de plusieurs interventions du législateur :

- 1988 : accès aux données de la statistique publique
- 1994 : accès aux données de santé pour la recherche
- 2013 : accès aux données fiscales
- 2016 : accès aux données de santé (SNDS)
- 2016-2017 : accès aux autres données produites par les administrations

S'agissant des évolutions récentes, la loi de modernisation de notre système de santé (art 193) a introduit de nouvelles modalités pour l'accès et la réutilisation des données de santé à des fins de recherche et d'évaluation en modifiant notamment dans la loi Informatique et Libertés, le circuit des procédures de formalités applicables en instaurant notamment un guichet unique via l'Institut National des Données de Santé (INDS) et sans entrer dans les détails des procédures plus ou moins allégées selon les types de recherches¹².

La loi pour une République numérique a quant à elle défini les conditions d'utilisation du NIR chiffré à des fins de recherche et de statistiques en prévoyant également une simplification des procédures.

S'agissant des droits des mineurs en matière de recherche médicale, cette loi prévoit sous certaines conditions, que l'information d'un des parents peut suffire ; elle introduit aussi la possibilité pour les mineurs de plus de 15 ans de s'opposer à l'information des parents.

Le règlement européen et la recherche :

De façon générale, le règlement général sur la protection des données (RGPD) qui entre en application en mai 2018¹³ poursuit un triple objectif :

- renforcer les droits des personnes pour développer la confiance et contribuer à l'essor de l'économie numérique
- assurer une plus grande harmonisation des règles de protection des données tout en renforçant la responsabilité des entreprises (et en réduisant fortement les formalités) ;
- développer le rôle des autorités de protection des données (APD) et du groupe européen des APD en renforçant leur pouvoir de sanction et en instaurant un mécanisme de coopération au plan européen.

S'agissant de la recherche, les dispositions générales contenues dans la directive et dans la loi sont largement reprises dans le règlement européen (article 89 notamment) mais le champ est élargi puisque le RGPD ne vise pas seulement les réutilisations (« traitements ultérieurs ») de données à des fins de recherche, statistiques mais aussi les traitements ayant pour finalité première la recherche.

12. Pour plus d'information, cf <https://www.cnil.fr/fr/recherche-medicale-queelles-formalites-pour-les-theses-et-les-memoires>

13. Pour plus d'information, cf <https://www.cnil.fr/fr/comprendre-le-reglement-europeen>

Le règlement précise le régime juridique applicable aux traitements de données réalisés à des fins archivistiques dans l'intérêt public, ce qui constitue une nouveauté par rapport à la directive, à des fins de recherche scientifique ou historique ou à des fins statistiques. Reconnaisant les spécificités de ces types de traitements, il prévoit ainsi la nécessité de garanties appropriées et introduit des possibilités de dérogations à certaines dispositions du règlement. Les garanties appropriées insistent sur le principe de minimisation des données et consistent notamment en l'adoption de mesures techniques et organisationnelles portant notamment sur des méthodes de pseudonymisation.

Plusieurs articles du règlement comportent également des dispositions spécifiques à la recherche ou aux statistiques. Il en est ainsi de la dérogation à l'obligation d'information, dès lors que cette dernière serait susceptible de rendre impossible ou de compromettre gravement la réalisation des objectifs dudit traitement (à cet égard, le considérant 62 apporte un éclairage : devraient être pris en considération le nombre de personnes concernées, l'ancienneté des données).

Un projet de loi modifiant la loi informatique et liberté¹⁴ sera discuté au Parlement au début de l'année 2018. Il permettra l'application effective du RGPD.

Un article est consacré aux traitements à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique et historique, ou à des fins statistiques (art 12). Il s'agit de préciser les marges de manœuvre relatives aux mesures concernant de tels traitements permises par l'article 89 du règlement, lequel autorise le droit national à prévoir des dérogations à certains droits des personnes concernées (droits d'accès, de rectification, d'opposition, à la limitation du traitement, portabilité), sous certaines conditions, dans la mesure où ces droits rendent impossible ou entravent sérieusement la réalisation des finalités spécifiques et où de telles dérogations sont nécessaires pour atteindre ces finalités.

Des dispositions particulières sont également prévues pour les traitements à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé.

En tout état de cause, à l'aune de ce nouveau cadre de régulation, il apparaît d'autant plus indispensable de concevoir et de diffuser, en concertation avec les milieux de la recherche, un guide de bonnes pratiques permettant d'expliquer au mieux les règles de protection des données applicables à ce secteur. Il importe aussi de développer des actions de formation et de sensibilisation adaptées tant auprès des délégués à la protection des données désignés au sein des organismes de recherche qu'auprès des chercheurs intéressés à se faire les relais de cette sensibilisation.

Protection des données personnelles et recherche : quels enjeux aujourd'hui ? Quelles priorités ?

Au-delà, les gisements de données personnelles de plus en plus massifs qui se constituent via nos pratiques numériques (web social, objets connectés...) suscitent des convoitises aussi de la part des équipes de recherche désireuses d'y accéder à des fins, légitimes, de connaissance.

Bien que partagées et souvent publiées de façon ouverte, ces données n'en conservent pas le moins le caractère de données personnelles dont l'exploitation à des fins de recherche appelle des garanties adaptées. La Cnil y réfléchit. Dans le contexte international, il s'agit là d'enjeux majeurs pour la recherche française.

14. <https://www.cnil.fr/fr/la-cnil-publie-son-avis-sur-le-projet-de-loi-relatif-la-protection-des-donnees-personnelles>

De même il est essentiel de développer plus fortement les travaux de recherche sur le chiffrement, les méthodes d'anonymat et de pseudonymat ; les dispositifs d'accès sécurisé aux données à l'égal du centre d'accès sécurisé aux données (CASD), dispositif exemplaire à bien des égards.

Enfin, mieux garantir les droits des personnes et l'aspiration de chacun à une plus grande maîtrise de ses données constitue un impératif que la recherche devra sans doute plus prendre en compte à l'avenir pour assurer ainsi une réelle participation citoyenne.

La CNIL en bref

- **Une autorité administrative indépendante**

- 18 membres + le défenseur des droits
- Services: 195 personnes
- Budget 2017: 17 millions d'euros

- **Une quadruple mission**

- Contrôle : déclarations, contrôles sur place et en ligne
- Sanction
- Information, conseil, éducation
- Réflexion éthique et veille prospective

- + de 90 000 déclarations/an;
- 800 autorisations recherche santé
- Correspondants Informatique et libertés: +18000 organismes
- guides pratiques, tutoriels
- 7000 plaintes/an
- + de 400 contrôles/an



CNIL.

Éléments pour une histoire du « Numéro de sécurité sociale »



Gérard LANG

Statisticien retraité, SFdS

Ce que nous appelons couramment « numéro de sécurité sociale » porte officiellement le nom de « numéro d'inscription au répertoire » (NIR). Le répertoire dont il s'agit est le « Répertoire national d'identification des personnes physiques » (RNIPP). L'histoire de ces deux institutions n'est pas de la « petite histoire » : elle contribue à éclairer l'histoire du pays, depuis l'entre-deux-guerres jusqu'à nos jours, en passant par les années d'occupation et par la création de la Commission Informatique et libertés (Cnil).

1 La conception du numéro national d'identification des personnes physiques

1.1 En décembre 1934, René Carmille, polytechnicien de la promotion 1906 et membre du contrôle général des armées, invente un identifiant numérique des personnes physiques basé sur l'état civil, dans le cadre de l'étude d'un fichier destiné à faciliter la mobilisation de l'armée française à l'aide de puissants moyens mécanographiques. Ce numéro matricule, attribué aux garçons dès leur naissance, doit en effet permettre d'organiser l'appel sous les drapeaux des classes d'âges successives.

1.2 Cet identifiant numérique, formé de 13 chiffres, comprend 6 composantes de longueurs respectives 1, 2, 2, 2, 3 et 3, dont seule la dernière est non-significative.

- **1°** : La première composante indique le sexe de la personne concernée (1 pour les hommes et 2 pour les femmes), information théoriquement redondante pour assurer correctement l'identification de la personne une fois les autres composantes connues, mais capitale dans l'objectif d'un fichier de mobilisation.
- **2°** : Les seconde et troisième composantes, chacune de longueur 2, sont relatives à la date de naissance de la personne concernée et formées respectivement des deux derniers chiffres de l'année de naissance (le millésime, qui indique l'âge de la personne concernée, information indispensable pour la mobilisation) et du mois de naissance, codé par un chiffre compris entre 01 et 12.
- **3°** : Les quatrième et cinquième composantes sont relatives au lieu de naissance de la personne concernée et de longueurs respectives 2 et 3. Leur construction part du constat qu'un identifiant numérique de longueur 5 suffit à identifier tous les lieux de naissance possibles du fait que :
 - **a/** la France métropolitaine est composée de moins de 100 départements (qui peuvent donc être codés chacun par un numéro « minéralogique » allant de 01

à 90), chacun composé de moins de 1000 communes (qui peuvent donc, au sein de chaque département et après avoir été rangées dans l'ordre alphabétique, être codées par un numéro compris entre 001 et 999)

- **b/** les autres composantes du territoire de l'Empire français, formant la France d'outre-mer, peuvent être rangés sous le numéro 98, puis répartis entre les groupes 1 (Afrique du Nord), 2 (Afrique occidentale française), 3 (Afrique équatoriale française), 4 (Madagascar et dépendances, Côte française des Somalies), 5 (Indochine, Etablissements français dans l'Inde), 6 (Colonies de l'Océan Pacifique), 7 (Colonies de l'Océan Atlantique) et 0 (Levant : Liban et Syrie, pour mémoire) comprenant chacun moins de 10 territoires ;
- **c/** les autres pays du monde, qui sont à la fois moins de 1000 en tout et moins de 100 dans chacun des 5 continents, peuvent être rangés sous le numéro 99, puis répartis entre les 5 continents, numérotés 1 (Europe), 2 (Asie) 3 (Afrique), 4 (Amérique) et 5 (Océanie), et codés au sein de chaque continent par un numéro compris entre 01 et 99.

- Ces deux composantes forment ce qui deviendra le « code officiel géographique ».
- **4°** : La sixième et dernière composante, de longueur 3, indique le rang de naissance de la personne concernée dans le lieu de naissance et le mois de naissance de celle-ci, donnés par les quatre composantes précédentes (ce qui confirme que la première composante n'est pas indispensable pour l'identification de la personne), remarque étant faite que (une fois les communes de Paris, Lyon et Marseille divisées respectivement en 20, 7 et 12 arrondissements et cantons) dans tous les cas cela représente moins de 1000 naissances.

1.3 Ce projet fait l'objet d'un rapport officiel de la Troisième République rendu le 5 février 1935 par René Carmille. À la suite celui-ci, qui est également professeur à l'École libre des Sciences politiques, publie en 1936 chez Sirey la première édition d'un ouvrage intitulé « De la mécanographie dans les Administrations ».

2 La constitution initiale du fichier d'identification des personnes physiques

2.1 Fin 1940, René Carmille est devenu directeur du « Service de la démographie » créé par la loi du 14 novembre 1940. Une Instruction de René Carmille du 18 mars 1941 sur la constitution initiale du fichier national d'identification des personnes physiques et une note de René Carmille du 11 avril 1941, relative à l'exécution par MM. Les Greffiers des Tribunaux de première instance du relevé des registres des actes de naissance (des personnes nées en France de 1881 à 1940) pour l'établissement d'un répertoire d'identification destiné au service de la Démographie, lancent la constitution du fichier national d'identification des personnes physiques. Le répertoire sera effectivement créé à partir de 1941.

2.2 L'identifiant numérique significatif à 13 chiffres des personnes physiques développé par ce fichier s'appuie sur un volume de 440 pages publié à Lyon par la direction de la Démographie, et interne au service, intitulé « Code officiel géographique établi à la date du 1er juin 1941 (Imprimerie Emmanuel Vitte) ».

2.3 Le statut « occulte » de ce fichier concernant la mobilisation a été approuvé par une correspondance militaire secrète relative à la création du service de la Démographie et à sa mission militaire camouflée¹.

1. Un article ultérieur sera consacré aux fichiers de personnes créés ou envisagés entre 1940 et 1944

3 L'utilisation à la Libération des fichiers d'identification des personnes physiques et des entreprises et établissements

3.1 L'ordonnance n° 45-2250 du 4 octobre 1945 porte organisation de la Sécurité sociale. L'ordonnance n° 45-2454 du 19 octobre 1945 fixant le régime des assurances sociales applicables aux assurés des professions non agricoles précise le régime de la Sécurité sociale créée par l'ordonnance du 4 octobre. Son article 6 écrit : « Un décret en forme de règlement d'administration publique fixe les modalités selon lesquelles est effectuée l'immatriculation aux assurances sociales des travailleurs remplissant les conditions requises pour être affiliés. ». Ce décret, pris le 29 décembre 1945 (décret 45-0179) précise dans son article 6 qu'un arrêté fixe les conditions dans lesquelles les caisses de sécurité sociale procèdent à l'immatriculation des assurés sociaux.

3.2 Les articles 32 et 33 de la loi n° 46-854 du 27 avril 1946, loi de finances pour 1946, portent création au ministère de l'économie et des finances d'un « Institut national de la statistique et des études économiques pour la métropole et la France d'outre-mer » dont le décret n° 46-1432 du 14 juin 1946 fixe les attributions et l'organisation. Cet Institut prend la place du Service national des statistiques (SNS) de René Carmille², dont il reprend pratiquement l'organisation. Le décret n° 47-834 du 13 mai 1947 relatif à l'organisation des services centraux et des directions régionales de l'Insee (dont, pour la première fois le titre, mais non pas le corps du texte, oublie « pour la métropole et la France d'outre-mer ») précise encore les choses et son article 2 décrit les services centraux de l'Institut : y figurent les inventaires démographiques et les inventaires économiques, c'est-à-dire les fichiers de personnes physiques et les fichiers d'entreprises et d'établissements hérités de Carmille. De ces fichiers, Alfred Sauvy, alors directeur de l'Ined, entreprend une critique si violente qu'elle oblige Francis-Louis Closon, premier directeur général de l'Insee, à faire une mise au point publique et à publier une lettre très sévère du 2 décembre 1947, dans laquelle il reproche à Alfred Sauvy son manque de solidarité avec ses collègues statisticiens publics tout en lui confirmant sa décision irrévocable de poursuivre les travaux de gestion du fichier national d'identification des personnes physiques (et du fichier des établissements).

3.3 En réalité, il n'est plus temps en 1947 d'attaquer la gestion par l'Insee des grands fichiers légués par Carmille, qui sont déjà devenus indispensables pour accompagner la reconstruction du pays et de son économie. En effet, dès le début de 1946, les arrêtés d'application relatifs à l'immatriculation des assurés sociaux par les caisses de sécurité précisent que cette immatriculation se fait à partir du numéro d'identification du SNS, qui devient dès lors connu comme le numéro de sécurité sociale. Toutefois, je n'ai pas réussi à retrouver ces arrêtés d'application, ni à déterminer à quelle date le numéro d'identification à 13 chiffres du SNS a été complété par une clé à deux chiffres (calculée comme étant le complément à 97 du reste de la division par 97 du nombre à 13 chiffres qu'est l'identifiant du SNS). Par ailleurs, peu après la création de l'Insee, les articles 1 et 2 de la loi n° 46-1889 du 28 août 1946 confient à l'Institut la charge de « tenir un fichier général des électeurs et électrices en vu du contrôle des inscriptions sur les listes électorales. ». Ce fichier, qui comprend désormais les femmes auxquelles le droit de vote a été reconnu par une ordonnance du 21 avril 1944, est constitué à partir du fichier des personnes physiques.

2. Le SNS avait été créé en octobre 1941 par fusion du Service de la démographie et de la Statistique générale de la France, et René Carmille en avait été nommé directeur général. René Carmille a été déporté en février 1944 et est mort en déportation.

4 Le passage de l'administration à l'informatique et la loi du 7 janvier 1978

4.1 Vingt ans plus tard, l'Insee décide d'entreprendre l'informatisation du registre d'identification des personnes physiques. Ce projet est annoncé dans le numéro 10 de la revue mensuelle de l'Institut, Économie et statistique, du mois de mars 1970, dans laquelle figure une chronique de 3 pages de Jacques Desabie, directeur de la statistique générale, intitulée « L'Insee entreprend d'automatiser le répertoire des personnes ». Dans cet article, l'auteur décrit en toute candeur les nombreux avantages de l'informatisation du fichier, permettant à son identifiant numérique traditionnel de longueur 13 auquel est ajouté une clé de contrôle numérique de longueur 2 (constituée du complément à 97 du reste de la division des 13 chiffres par le nombre premier 97) de réaliser le projet initial d'identifiant unique du citoyen dans ses relations avec l'administration.

4.2 Mais le choix fait en 1973 par le ministère de l'Intérieur du nom du projet d'informatisation du répertoire des personnes physiques « Système Automatisé pour les Fichiers Administratifs et le Répertoire d'Identification » est pour le moins malheureux, car l'acronyme beaucoup plus court sous lequel il est connu « Safari » (on ne se méfie jamais assez de l'inconscient de ses informaticiens) va contribuer à lancer immédiatement un vif débat, nourri de violentes réactions s'opposant à son côté « orwellien ».

4.3 C'est un article publié par Philippe Boucher dans le journal « Le Monde », dont Jacques Fauvet est le directeur, et intitulé « SAFARI ou la chasse aux Français » qui lance le débat et conduit le premier ministre Pierre Messmer à demander un rapport à ce sujet, rédigé par une commission dont le rapporteur général est le conseiller d'État Bernard Tricot. Ce rapport est publié en 1975 et conduit au vote par le Parlement de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dont le Garde des sceaux Alain Peyrefitte dira qu'elle est la fille naturelle du rapport.

4.4 D'ailleurs la loi crée la Commission nationale de l'informatique et des libertés (Cnil), premier exemple en France d'une autorité administrative indépendante, dont Bernard Tricot devient le premier secrétaire général. La loi du 6 janvier 1978 cible très clairement le secteur public, considérant que seul celui-ci dispose à ce moment du moyen de collecter de grands fichiers relatifs aux personnes physiques et d'ordinateurs assez puissants pour les traiter. En fait, la loi distingue les procédures suivantes :

- **1°** L'article 15 dispose que « Hormis les cas où ils doivent être autorisés par la loi, les traitements automatisés d'informations nominatives opérés pour le compte de l'Etat, d'un établissement public ou d'une collectivité territoriale ou d'une personne de droit privé gérant un service public, sont décidés par une loi ou par un acte réglementaire pris après avis motivé de la Cnil. Si l'avis de la Cnil est défavorable, il ne peut être passé outre que par un décret pris sur avis conforme du Conseil d'État ou, s'agissant d'une collectivité territoriale, en vertu d'une décision de son organe délibérant approuvée par décret sur avis conforme du Conseil d'État.
- **2°** L'article 16 dispose que les traitements effectués pour le compte de personnes autres que celles visées à l'article 15 doivent, préalablement à leur mise en œuvre, faire l'objet d'une déclaration auprès de la Cnil.
- **3°** L'article 17 dispose que pour les catégories les plus courantes de traitements à caractère public ou privé, qui ne portent manifestement pas atteinte à la vie privée, la Cnil établit et publie des normes simplifiées.
- **4°** L'article 18 écrit :
 - « L'utilisation du répertoire national d'identification des personnes physiques (RNIPP) en vue d'effectuer des traitements nominatifs est autorisée par décret en Conseil d'État

pris après avis de la Cnil. »

- Ainsi, la loi prévoit 4 procédures distinctes, dont les plus simples sont la norme simplifiée (dont trois, numéros 18, 19 et 26, seront spécifiques à la statistique publique, s'appliquant pour les enquêtes auprès des entrepreneurs individuels et enquêtes auprès des ménages, ainsi qu'aux exploitations de fichiers administratifs sur les personnes physiques) et la déclaration.
- Sachant que l'hypothèse de la loi prévue par l'article 15 est rarissime, et que (à ma connaissance) le gouvernement n'est jamais passé outre à un avis défavorable de la Cnil, la procédure de l'article 15 se réduit pratiquement à un arrêté. Aussi la procédure de l'article 18, réservée spécifiquement au RNIPP, est bien la plus lourde et en fait le fichier paradigmatique soumis au contrôle de la Cnil.

4.5 Cette mise sous surveillance suspicieuse de l'Insee par la Cnil concernant le RNIPP n'instaure pas initialement un climat favorable à la coopération entre les deux institutions, dont le « choc de cultures » va se manifester à propos du recensement de la population de 1982. Après quelques années, la statistique publique réagit, d'une part, en régularisant sur le plan juridique des pratiques tenues depuis longtemps pour évidentes, et, d'autre part, en pacifiant ses relations avec la Cnil et ses services par une meilleure compréhension mutuelle, grâce à l'entremise d'Anne-Marie James pour l'Insee et de personnes comme Fatima Hamdi dans les services de la Cnil, ainsi que des commissaires chargés de la statistique publique, comme Michel May et Guy Rosier. Cette évolution demande évidemment un certain temps. En ce qui concerne le RNIPP, on compte parmi ses premières étapes :

- **1°** Le décret n° 82-103 du 22 janvier 1982 relatif au répertoire national d'identification des personnes physiques (RNIPP) ;
- **2°** Le décret n° 83-101 du 15 février 1983 autorisant l'utilisation du RNIPP en vue de la tenue du fichier électoral ;
- **3°** Le décret n° 84-393 du 23 mai 1984 autorisant l'utilisation du RNIPP pour le traitement automatisé de l'échantillon démographique permanent ;
- **4°** Le décret n° 84-394 du 23 mai 1984 autorisant l'utilisation du RNIPP pour la gestion du répertoire national d'identification des entreprises et de leurs établissements (Sirene).
- **5°** le décret n° 85-1343 du 16 décembre 1985 instituant un système de transfert de données sociales (TDS) utilisant le NIR.

La Cnil reste néanmoins inflexible sur le principe du cantonnement de l'utilisation du RNIPP, notamment en refusant l'extension de l'utilisation de son identifiant, rebaptisé « numéro d'inscription au répertoire (NIR) » au domaine de l'éducation nationale et en résistant le plus possible à son utilisation dans les domaines de la santé (voir ainsi pour le RNIAM le décret n° 96-793 du 12 décembre 1996 relatif à l'utilisation du NIR et à l'institution du répertoire national inter-régimes des bénéficiaires de l'assurance-maladie) ou des impôts (voir l'article 107 de la loi n° 98-1236 de finances pour 1999 qui gouverne l'utilisation du NIR en matière fiscale).

Au total, on compte une cinquantaine de décrets de l'article 18 de la loi de 1978 autorisant des traitements à recourir au RNIPP entre 1982 et 2004.

4.6 Un autre point intéressant de cette époque regarde la première composante du numéro d'identification, qui code le sexe de la personne.

Signalons d'abord que l'un des premiers actes d'Yvette Roudy, nommée le 21 mai 1985 ministre des Droits de la femme, donc ministre de plein exercice (et c'est une première en France) a été d'écrire une lettre au directeur général de l'Insee pour protester contre la scandaleuse discrimination résultant du fait que le codage de la variable sexe fait passer les hommes (codés 1)

avant les femmes (codées 2). La réponse officielle est qu'il s'agit d'une pratique internationale, d'ailleurs sanctionnée dès 1976 par la norme internationale ISO/CEI 5218 « Codes for the representation of the human sexes » qui prescrit le codage suivant :

- 0 : not known
- 1 : male
- 2 : female
- 9 : not applicable

Une réponse officieuse faite au cabinet de la ministre est la suivante : « Le code 0 est disponible, et coder le sexe féminin par 0 fait passer le sexe féminin devant le sexe masculin. C'est à vous de décider si symboliquement le code 0 est préférable au code 2 pour le sexe féminin »

Une autre difficulté résulte du fait que le Conseil d'État a jugé que l'affichage par le numéro de sécurité sociale du sexe de naissance d'une personne transsexuelle qui considère qu'elle se reconnaît psychologiquement de l'autre sexe est susceptible de lui causer un préjudice moral. Toutefois à cette époque le mouvement de « changement de sexe » n'est pas reconnu par l'état civil français et le seul moyen de modifier la première composante du numéro d'identification d'une personne physique consiste à trouver un juge de l'état civil de bonne volonté qui accepte de faire un faux en « corrigeant » le sexe de naissance de la personne concernée.

5 La situation issue de la loi de transposition de 2004

5.1 Au-delà du droit proprement français, le droit communautaire s'intéresse également à la protection des données personnelles. En effet, dans le prolongement de la convention européenne des droits de l'homme et des libertés fondamentales adoptée à Rome le 4 novembre 1950, le Conseil de l'Europe a adopté à Strasbourg le 28 janvier 1981 la convention n° 108 pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, précisée le 30 septembre 1987 par la recommandation R(97) 18 concernant la protection des données à caractère personnel collectées et traitées à des fins statistiques. De son côté, après avoir adhéré à la Convention n° 108, la Communauté européenne a adopté la directive n° 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995 relative à la protection des personnes à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, puis le règlement n° 322/97 du Conseil du 17 février 1997 relatif à la statistique communautaire. A cause de deux alternances politiques, qui ont entraîné la remise sur l'ouvrage complète de trois projets de loi distincts mobilisant 15 réunions interministérielles (j'ai participé à 14 d'entre elles, ce qui doit être le record) étalées sur 8 ans, le délai de transposition de la directive de 1995 a été largement dépassé et la loi de transposition n'a été finalement publiée que par le Journal officiel de la République française du 7 août 2004.

5.2 Le texte « simplificateur » du chapitre « Formalités préalables à la mise en œuvre des traitements » de cette loi substitue aux 3 ou 4 procédures (déclaration, norme simplifiée, et loi ou acte réglementaire, dont décret en Conseil d'État pour le RNIPP) délimitées très clairement dans la loi initiale de 1978 un véritable labyrinthe comportant au moins 5 procédures distinctes (déclaration, norme simplifiée, autorisation, arrêté, décret en Conseil d'État ou arrêté) dont les limites sont beaucoup plus complexes. Pour ce qui concerne l'utilisation du RNIPP, on constate qu'une autorisation remplace le décret en Conseil d'État lorsque le NIR ne figure pas parmi les données nécessaires au traitement, ce qui constitue une certaine amélioration de la situation. Mais, dès lors que le numéro d'identification devient nécessaire au traitement, la procédure demeure celle du décret en conseil d'État, sauf l'exception ouverte au 4° du II de l'article 27 concernant les télé services de l'administration. A ce stade, il ne s'agit vraiment que d'un mince progrès, même s'il ouvre une piste éventuellement prometteuse.

5.3 Sur la question relative au codage de la première composante de l'identifiant, il convient

désormais de tenir compte du fait que le sexe des personnes physiques est devenu une matière juridique de rang constitutionnel, dans la mesure où la loi constitutionnelle n° 2008-724 du 23 juillet 2008 a inséré dans l'article 1er de la Constitution de la République française un second alinéa qui écrit « La loi favorise l'égal accès des femmes et des hommes aux mandats électoraux et fonctions électives, ainsi qu'aux responsabilités professionnelles et sociales. ».

6 Les avancées du droit à partir de 2012

6.1 Constatons à ce stade que les trois instruments fondés par le SNS de René Carmille, devenus au sein de l'Insee respectivement le code officiel géographique (COG), le répertoire national d'identification des personnes physiques (RNIPP) et le répertoire national d'identification des entreprises et de leurs établissements (Sirene), y constituent trois fichiers intimement liés qui contribuent non seulement à l'encadrement général des activités de la statistique publique mais également à l'exercice de la souveraineté (numérique ?) de l'État français pour lequel ils produisent :

- un code pour la représentation du nom des circonscriptions administratives et des collectivités territoriales de la République, ainsi que des pays et territoires étrangers ;
- un code pour la représentation de l'état civil des personnes physiques nées ou vivant sur le sol national ;
- un code pour la représentation de l'identité des entreprises ayant leur siège sur le sol national, ainsi que de leurs établissements.
- L'intimité des liens entre les trois fichiers résulte :
- du fait que le COG code la localisation du siège des entreprises et le lieu d'activité de leurs établissements dans Sirene, et le lieu de naissance des personnes physiques dans le RNIPP tout en fournissant 5 des 13 chiffres de l'identifiant NIR de celui-ci ;
- du fait que l'identité de tout entrepreneur individuel inscrit dans Sirene est vérifiée dans le RNIPP, et que réciproquement tout entrepreneur individuel inscrit dans Sirene né en dehors du territoire de la République française est alors inscrit dans le RNIPP.
- C'est cette armature fondamentale qui est en passe de se voir désormais consacrée par les textes récents pour prendre en compte l'explosion des services rendus par l'administration électronique.

6.2 Au milieu de la décennie 2010-2020, l'encadrement donné par le droit européen est modifié par le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données personnelles (RGPD), qui abroge et remplace la directive de 1995 qu'il modifie parfois sensiblement, et qui devient applicable à compter du 25 mai 2018 pour entrer directement dans le droit national des 28 Etats membres de l'Union européenne. L'article 87 de ce règlement dispose :

- « Traitement du numéro national d'identification Les Etats membres peuvent préciser les conditions spécifiques du traitement d'un numéro national d'identification national ou de tout autre identifiant d'application général. Dans ce cas, le numéro d'identification national ou tout autre identifiant d'application générale n'est utilisé que sous réserve des garanties appropriées pour les droits et libertés de la personne concernée adoptées en vertu du présent règlement. »

Il est bien clair que le numéro d'inscription au RNIPP (le NIR) ne peut pas être considéré autrement que comme le numéro d'identification national régi par l'article 87 (justifiant ainsi son appellation originelle de NNI : numéro national d'identité), cependant que l'identifiant SIREN, concernant plusieurs millions d'entrepreneurs individuels est un autre identifiant d'application générale.

6.3 Il convient d'ailleurs de préciser que, considérant les développements exponentiels des capacités de stockage et de traitement de l'information depuis l'an 2000, l'interconnexion de données relatives à une même personne distribuées dans plusieurs fichiers distincts réalisée

par « matching » sur l'état civil de cette personne n'a plus rien d'une prouesse technique. Le recours à l'identification de cette personne par son NIR ne fait que faciliter quelque peu cette opération, et donner une plus grande sécurité dans cette identification.

6.4 Pour sa part, la loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé a également modifié la loi du 17 janvier 1978 pour permettre la mise en place du système national de données de santé créé par le I de son article 193, insérant un titre VI nouveau dans le livre IV de la première partie du code de la santé publique. À la suite, le VII de cet article 193 réécrit comme suit le début de l'article L. 1111-8 du même code : « Le numéro d'inscription au répertoire national d'identification des personnes physiques est utilisé comme identifiant de santé des personnes pour leur prise en charge à des fins sanitaires et médico-sociales ». Il s'agit là d'une véritable révolution, si l'on considère la résistance historique acharnée de la Cnil à l'extension de l'utilisation du NIR dans les matières fiscales, de l'éducation et de la santé.

6.5 Le décret n° 2017-412 du 27 mars 2017 relatif à l'utilisation du numéro d'inscription au répertoire national d'identification des personnes physiques comme identifiant national de santé, pris en application de cet article 193 vient encore valider la thèse de la reconnaissance progressive du NIR comme identifiant national unique du citoyen dans ses relations (électroniques) avec l'administration française.

6.6 Concernant la première composante de l'identifiant, il convient de remarquer qu'un arrêt de la première chambre civile de la Cour de cassation du 4 mai 2017 a récemment refusé la requête d'une personne inscrite à l'état civil comme née de sexe masculin de rectifier son acte de naissance pour y substituer à l'indication « sexe masculin » celle de « sexe neutre » ou à défaut « inter sexe » au motif que la binarité des sexes masculin et féminin poursuit un but légitime car elle est nécessaire à l'organisation sociale et juridique dont elle constitue un élément fondateur. Ajoutons, a contrario, qu'un jugement de la Cour constitutionnelle allemande du 8 novembre 2017 a donné à la Chambre des députés un délai allant jusqu'à la fin de l'année 2018 pour voter la légalisation de la prise en compte, sur les registres de naissance allemands, d'un « troisième sexe », pratique encouragée par un rapport de 2015 du Conseil de l'Europe et reconnue depuis 2014 en Australie. Pour leur part, le Talmud et la Halaka (loi juive classique ancienne) reconnaissent six sexes.

7 L'aboutissement de 2018

7.1 A l'heure où cet article est rédigé, le Parlement français délibère sur un projet de loi sur les données personnelles destinée à modifier encore la loi « Informatique et libertés », pour y intégrer les conséquences de l'entrée en vigueur, au 25 mai 2018, du RGPD, et pour en exploiter les possibilités d'adaptation nationale.

7.2 L'article 22 du projet de loi est relatif aux traitements portant sur des données comportant le numéro d'inscription des personnes au répertoire national d'identification des personnes physiques qui nécessitent un décret d'autorisation en Conseil d'État, à l'exception

- **a/** du cas de finalités de statistique publique (après une opération cryptographique substituant au NIR un identifiant statistique non signifiant dont l'utilisation est restreinte au service statistique public) ;
- **b/** du cas de finalité de recherche scientifique ou historique (après une opération cryptographique) ;
- **c/** du cas de mise à disposition des usagers des l'administration un ou plusieurs télé services de l'administration électronique.

L'article 27 est relatif aux traitements de données à caractère personnel mis en œuvre pour le compte de l'Etat et qui portent sur des données génétiques ou sur des données biométriques nécessaires à l'identification ou au contrôle de l'identité des personnes. Ces traitements nécessitent un décret d'autorisation en Conseil d'État.

7.3 Un autre article important, voté dans des termes peu différents par les deux assemblées parlementaires est l'article 13, qui réécrit le chapitre IX « Traitements de données à caractère personnel dans le domaine de la santé » de la loi Informatique et libertés. La section 2 de cet article 13 « Dispositions particulières relatives aux traitements à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé » renforce encore le rôle de l'Institut national des données de santé.

7.4 Compte tenu de tout ce qui précède, on peut donc de bonne foi considérer que l'objectif affiché en 1970 par l'article de Jacques Desabie est pleinement en voie de réalisation en cette année 2018, où nous assistons à l'éclatement, qu'il va falloir savoir maîtriser, de l'utilisation de « l'intelligence artificielle » et dans laquelle nous célébrons le 40e anniversaire de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dont résulte la création de la Commission nationale de l'informatique et des libertés (Cnil), et de ses maintes modifications depuis lors.