

RENÉ PADIEU

Grandes bases de données et protection des personnes

Journal de la société française de statistique, tome 146, n° 3 (2005),
p. 89-93

http://www.numdam.org/item?id=JSFS_2005__146_3_89_0

© Société française de statistique, 2005, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

GRANDES BASES DE DONNÉES ET PROTECTION DES PERSONNES

René PADIEU *

RÉSUMÉ

La mobilisation de grands fichiers de données administratives pour établir des statistiques, d'une part, et la possibilité d'enchaîner ou apparier de tels fichiers aussi bien que ceux de collectes statistiques pour constituer des panels, d'autre part, présentent un grand intérêt. Mais elles supposent une grande rigueur et l'observation de règles strictes pour protéger la vie privée des personnes concernées. En effet, ces possibilités qui ont connu un fort développement depuis un demi-siècle ont suscité de vives controverses. Il semble que nous soyons aujourd'hui parvenus à un stade de maturité. Les questions techniques, morales et juridiques ont été grandement clarifiées. Tout, certes, n'est pas résolu; mais nous avons désormais les moyens de traiter ces problèmes de façon satisfaisante. Je propose de revenir ici, d'abord, sur la « crise de croissance » que nous avons traversée; puis, d'énoncer les enjeux tels qu'ils me semblent se présenter.

Mots clés : données administratives, confidentialité, législation, vie privée.

ABSTRACT

Large administrative databases are being exploited to gather statistics, and such files – as well as those compiled from surveys – can now be chained or linked in order to construct panels. While such processing is of great value, it requires considerable rigor and compliance with strict rules to protect the privacy of individuals concerned. The use of these procedures has expanded sharply in the past half-century, sparking heated debate. We now seem to have reached a stage of maturity. Technical, moral, and legal issues have been substantially clarified. Not all problems have been solved, but we now have the means to deal with them in a satisfactory manner. My article revisits the “growth crisis” that we have just been through, and then describes what I regard as today's challenges.

Keywords : administrative data, confidentiality, privacy, protection laws.

1. La crise de croissance

Cette expression veut exprimer le fait que l'existence et l'utilisation de grandes bases de données a suscité des difficultés qui n'ont pas été surmontées d'emblée. La seconde moitié du xx^e siècle a vu (même s'il y avait eu bien sûr des antécédents) trois évolutions majeures, qui jouaient sur des plans différents bien qu'elles n'aient pas été sans lien :

* 4, rue du Cloître Notre-Dame, F-75004 Paris. Courriel : rene-padieu@wanadoo.fr

- *l'affirmation du respect dû à la vie privée*, qui n'est apparu dans le droit qu'à cette époque ¹, récemment donc, bien qu'il nous paraisse aujourd'hui aller de soi ;
- *la création de fichiers considérables* induits à la fois par la tertiarisation de l'économie, avec des entreprises traitant de plus en plus d'information, et par l'accroissement du rôle de l'État, conduisant les administrations à manier des masses croissantes de données ;
- *l'informatique*, apparue dans les années cinquante et qui a permis l'énorme développement que nous connaissons des capacités de stockage, de la vitesse et du perfectionnement des traitements, de la commodité de transmission et de mise en relation.

La ressource que constituaient pour l'analyse statistique ces immenses gisements de données a été tôt repérée. Les puissants outils dont on allait disposer en permettaient l'exploitation, ainsi qu'un contrôle de qualité, un raffinement de codification et la production de résultats sans précédents. D'abord, les statisticiens y ont vu une aubaine et des projets ont surgi (Sirène, Safari, DAS (« salariés III »), SUSE, Enéide, Gamin, ...). La protection des personnes n'était pas explicitement prise en considération, mais elle était néanmoins très présente à la conscience des praticiens. C'est que le secret statistique était une valeur professionnelle depuis longtemps assumée : la première mention juridique, à ma connaissance, remonte à 1898 en Suisse, et la loi britannique qui l'établit est de 1920. En France, il est formulé par la loi du 7 juin 1951 sur la statistique publique. Mais les sociétés privées de sondage l'avaient posé dans leur charte de 1948 au niveau européen ². Pour les praticiens, prendre toutes précautions pour préserver la confidentialité des données était une nécessité évidente.

Or, dans les années 1970, l'inquiétude s'est répandue que ces beaux instruments et ces merveilleux projets pouvaient menacer la vie privée. Le public ne faisait du reste pas la différence entre traitements statistiques et autres traitements, administratifs, policiers ou commerciaux. Il est connu que c'est le projet « SAFARI » de l'Insee ³ qui a mis le feu aux poudres. Mais, pour cela, il fallait qu'il y ait de la poudre ! La même inquiétude se faisait jour dans

1. Le droit au respect de la vie privée est proclamé par la Convention européenne des droits de l'Homme (4 novembre 1950) alors qu'elle ne figurait pas, par exemple, dans la Déclaration de 1789. De son côté, le Code Civil garantit aussi ce droit, en son article 9, mais ceci ne remonte qu'à 1970 !

2. Code international CCI/ESOMAR de pratiques loyales en matière d'études de marché et d'opinion.

3. À cette époque, l'informatisation des grands systèmes administratifs était encore relativement nouvelle. L'Insee gérait depuis quarante ans de façon manuelle, sur des registres-papier, un répertoire de l'ensemble des personnes nées en France (et de travailleurs étrangers), tenu à jour par copie des actes d'état-civil. Ce répertoire a été informatisé, il s'appelle aujourd'hui « RNIPP » (répertoire national d'identification des personnes physiques). Le premier projet était désigné par l'acronyme « Safari » : les informaticiens se complaisaient à trouver des sigles formant un mot existant par ailleurs, facile à mémoriser. Ainsi avait-on « Sirène » pour « Système Informatique des Répertoires d'ENTreprises et Établissements », etc. L'alerte quant à un éventuel « fichage » des citoyens a été donnée par un article du journal *Le Monde* titré « Safari ou la chasse aux Français », qui a eu un grand retentissement.

d'autres pays, qui ne devait rien à SAFARI. Simultanément, dans plusieurs pays d'Europe et d'Amérique du Nord, des législations analogues à notre loi « Informatique et Libertés » ont vu le jour à la fin de la décennie. Puis une Convention du Conseil de l'Europe intervenait en 1981⁴.

D'abord, les choses se sont assez mal passées entre les statisticiens et les protecteurs de données (c'est-à-dire la Commission Nationale de l'Informatique et des Libertés, la CNIL). Les statisticiens ne voyaient pas l'intérêt de ces nouvelles règles. Ils ne les avaient pas attendues pour veiller très correctement au risque de dévoilement⁵. Selon eux, la nouvelle loi mettait des barrières là où ce n'était pas utile et gênait la recherche. Ils considéraient que les concepteurs de la loi et ses gardiens ne comprenaient rien à la façon dont les choses se présentaient pour la statistique... Mais, pour ces gardiens de la loi, ces statisticiens qui prétendaient s'en abstraire étaient suspects : ou bien d'être animés d'intentions perverses ; ou bien, au mieux, d'être de dangereux naïfs, inconscients du péril qu'ils faisaient courir aux gens.

Depuis, les choses se sont grandement arrangées. Les statisticiens ont compris que l'existence d'une loi était plutôt une bonne chose. Ils auraient pu en avoir l'idée plus tôt : lorsque la loi de 1951 avait été adoptée, ce n'est pas le législateur qui avait pris l'initiative de l'imposer aux statisticiens, c'était eux qui l'avaient réclamée. Une des premières tâches de l'Insee nouvellement créé avait justement été d'en reconnaître l'intérêt et de la proposer. De son côté, la CNIL a progressivement mieux pris la mesure de ce que le travail statistique impliquait. Contrairement à la perception de certains, elle a même défendu des enquêtes contre les protestations de quelques citoyens⁶. Et, elle a pu atténuer parfois la rigueur de la loi. Par exemple, l'exploitation des déclarations annuelles de salaires, qui s'était faite sans encombre de 1950 à 1978, était devenue illégale à cette dernière date : la CNIL a demandé que la loi soit aménagée (ce qui fut fait en décembre 1986) et s'est, entre-temps, abstenue de sanctionner. Certes, des discussions ont encore lieu, mais beaucoup plus sereines.

Parfois aussi, les associations de statisticiens ont senti l'intérêt d'énoncer par elles-mêmes des règles professionnelles là où n'existait qu'une culture, réelle certes mais implicite. En 1980, l'*American Statistical Association* avait donné l'exemple en rédigeant des « Ethical Guidelines for Statistical Practice ». En 1985 ont été adoptés, d'une part, en France, le code de déontologie de l'Association des administrateurs de l'Insee et anciens élèves de l'ENSAE et,

4. Convention pour la protection des personnes à l'égard du traitement automatique des données à caractère personnel, Traité n° 108, Conseil de l'Europe, Strasbourg, 1981.

5. Je me souviens avoir entendu Jacques Desabie, dans le département de qui s'élaborait safari, dire que bien sûr la confidentialité était assurée comme pour toute enquête ou traitement statistique, dans les conditions prescrites par la loi de 1951, mais encore qu'une procédure de sabotage devrait être prévue pour le cas où un gouvernement imposerait une utilisation non conforme. Nous avons encore à cette époque, trente ans après, la période de Vichy très présente à l'esprit.

6. Par exemple, l'enquête « actifs financiers ».

d'autre part, au niveau mondial, la Déclaration d'Éthique Professionnelle de l'IIS (Institut international de statistique)⁷.

Il y a dix ou quinze ans, la protection des données semblait relever de deux logiques distinctes. Pour les Anglo-Saxons, y compris scandinaves, nord-américains et australiens, c'était une affaire d'algorithmes de cryptage, brouillage, masquage, etc⁸. Pour les Allemands et les Latins (chez nous donc, en particulier) c'était une question d'obligations et d'interdits. D'un côté, la sécurité des personnes était assurée par l'ingénieur; de l'autre, par le juriste. Maintenant, l'autorité de protection (la CNIL) reconnaît les vertus des logiciels, tandis que la législation s'est harmonisée, au moins au plan européen. Le débat reste plus épineux avec les États-Unis lorsqu'il s'agit de transmission transfrontalière de données.

Notre loi Informatique et Libertés de 1978 vient d'être modifiée, en août 2004⁹, pour se conformer à une directive européenne de 1995 (la France a beaucoup tardé pour cette transposition). Elle réaffirme le principe d'une collecte et d'une utilisation loyales des données personnelles, recueillies pour une finalité explicite et exclusive. Parmi les évolutions qui intéressent les statisticiens, citons :

- un assouplissement du principe de finalité : une utilisation statistique ou scientifique est compatible avec une finalité de collecte initialement différente, ce qui autorise la mobilisation des bases de données administratives;
- un aménagement en ce qui concerne les données sensibles (convictions, appartenances politiques ou syndicales, origines ethniques, santé, comportements sexuels, données judiciaires). Leur collecte et utilisation restent très encadrées, l'interdiction de principe ne pouvant être levée que pour des motifs et à des conditions très strictement limités. L'une d'elles demeure le consentement exprès, c'est-à-dire préalable et écrit, ce qui est parfois rédhibitoire pour des recherches auprès de personnes mentalement perturbées ou en rupture avec la société. Or, ici, une porte s'est entrouverte : la CNIL peut dispenser du consentement exprès lorsque l'inscription du projet dans le programme public d'enquêtes visé par le CNIS atteste de l'intérêt public et de la pertinence d'un protocole de recueil particulier;
- la mention explicite de codes professionnels, que les institutions ou organisations peuvent soumettre à la CNIL afin qu'elle en vérifie et atteste la conformité à la loi. Même si nous ne savons pas encore quels seront les effets de cette reconnaissance, nous pouvons réfléchir à l'établissement de tels codes (par exemple, par la SFdS);

7. Ces deux textes couvrent non seulement la protection des enquêtés, mais aussi l'obtention des résultats et leur diffusion, ainsi que l'indépendance des travaux. Par ailleurs, on peut mentionner l'adoption, en 1992 par la Conférence des statisticiens européens, puis en 1994 par l'ONU au niveau mondial, des dix « Principes de la Statistique Officielle », qui couvrent aussi bien la protection des enquêtés que la conception et l'indépendance du travail statistique.

8. Curieusement, alors que la Déclaration de l'IIS était une initiative anglo-américaine, c'est moi qui ai fait introduire dans le projet l'idée que l'ordinateur n'était pas seulement une menace mais aussi un moyen puissant pour protéger l'intimité.

9. Voir l'article de Michel Isnard dans ce même dossier.

- la mention que des outils informatiques concourent à la sécurité des données. La CNIL a désormais la possibilité de labelliser de tels produits. Elle avait du reste pris les devants : c'est par exemple en accord avec elle que la SFdS avait organisé en 2003 un cours pour promouvoir les logiciels d'appariement sécurisé d'abord développés par les épidémiologistes.

2. Les enjeux actuels

Si nous pouvons parler de maturité, c'est que les malentendus et controverses sont pour l'essentiel aplanis et que nous disposons des moyens techniques et juridiques pour développer des projets féconds. Cela ne veut pas dire que tout soit résolu. D'autres progrès techniques, d'autres ajustements juridiques seront nécessaires et, surtout, les esprits ont encore un travail à faire pour bien repérer les problèmes et les solutions.

Deux objectifs : d'une part exploiter au mieux la ressource informationnelle que constituent les gisements de données existants ou à créer et les puissants outils de traitement ; d'autre part, offrir les garanties les plus sérieuses quant au respect de la vie privée.

Pour assurer ce respect dans les faits, nous avons des règles à observer et des outils à mettre en œuvre. Simultanément, nous devons rassurer le public aussi bien que les autorités de protection. C'est donc un travail d'explication, de démonstration et de conviction. Nous devons en effet rassurer pour pouvoir travailler, mais il importe aussi que la protection soit réelle : autant pour que la confiance ne coure aucun risque, que par obligation éthique.

Certains de nos collègues, constatant que le risque de révélation des données personnelles est extrêmement minime, se demandent s'il est bien raisonnable de consacrer autant d'efforts et de moyens. Les risques objectifs sont en effet bien plus ténus que ce que nos concitoyens imaginent. (Du moins, tant que les statisticiens travaillent proprement : et il importe de s'assurer qu'il en est bien ainsi !) Il faut voir aussi que le risque n'est pas seulement à la mesure des probabilités d'accident et du dommage effectif, mais que l'on a affaire à une demande sociale à laquelle il y a lieu de déférer. Bien ou mal fondée, elle est légitime et, tant que nous n'avons pas le moyen de la raisonner, si même elle devait l'être, nous devons la prendre comme une contrainte nécessaire.

Il convient donc de faire un travail d'élucidation en direction de l'opinion en général et, particulièrement, des responsables politiques et administratifs : pour comprendre leurs attentes, pour leur faire connaître et comprendre nos activités et nos sauvegardes. Et de façon symétrique, ce travail externe doit aussi être conduit en adossement à un travail interne, pour sensibiliser la profession aux enjeux complémentaires de l'information et de la protection, et pour lui faire connaître, utiliser et développer encore, tant les règles que les outils.

Nous devons en effet jouer sur trois registres : celui de la règle, celui de la technique et, entre les deux, non moins indispensable, celui de la conscience des professionnels.