

Conclusion

Antoine Frachot (1989)

Directeur général du Groupe des Ecoles Nationales d'Économie et de Statistique (GENES)

Je vais commencer par une petite anecdote historique des années 70. En 1975-1976, un article du Monde annonce : « l'administration française est en train de faire la chasse aux Français ». Certains politiques avaient eu l'idée de constituer un grand fichier où seraient appariées des données administratives de toute sorte. En gros, il s'agissait de constituer une grande base de données quasi-exhaustive ; une idée absolument géniale fût d'appeler ceci « le projet Safari ».

Un psychanalyste pourrait expliquer, peut-être, ce qui a bien pu traverser l'inconscient des fonctionnaires et politiques à l'origine de ce projet et de sa dénomination, alors même qu'en 1975, la deuxième guerre mondiale ne s'est achevée que trente ans avant. Trente ans, c'est court, pour le temps historique, la sensibilité était encore extrêmement forte sur le mal que l'on pouvait faire à partir de données administratives.

Ce qui fait honneur à la France et aux politiques de cette époque, c'est d'avoir décidé d'arrêter le projet et de prendre le temps de réfléchir. Des débats à l'Assemblée Nationale, au Parlement, ont eu lieu pour essayer de définir une doctrine

des données à caractère personnel, et ce sont ces échanges qui ont donné naissance à la CNIL. Aujourd'hui encore, la doctrine de la CNIL est très marquée par cet événement. Le prix de thèse « Informatique et Libertés » décerné par la CNIL en 2010¹ raconte comment la doctrine des données à caractère personnel s'est forgée en France et comment la CNIL a été créée. Histoire passionnante.

Tout cela pour énoncer ma conviction que les thèmes abordés lors de ce colloque sont vraiment au cœur de la problématique du numérique. D'un côté, on veut « ouvrir » les données au maximum en s'appuyant sur le fait que les données, notamment administratives mais pas uniquement, sont des richesses extraordinaires et qu'elles permettent d'étayer et de mieux comprendre pour ensuite mieux agir dans l'intérêt général sur toutes sortes de problématiques, phénomènes économiques, sociaux ou relatifs à la santé ou à l'environnement. D'un autre côté, la question immédiate est : « comment protège-t-on ces données ? ».

Nous sommes donc face à un dilemme : en même temps ouvrir et fermer les données au

1 - « Les valeurs de la recherche : enquête sur la protection des données personnelles en épidémiologie », Nicolas Lechopier, Ed. Michalon, 2011.

maximum. Ma conviction est qu'il faut résoudre ce dilemme pour éviter une spirale dangereuse.

Ce point est très important et est au coeur de nos réflexions ces dernières années. Ce n'est pas forcément le cas dans tous les pays, mais au moins en France, le balancier, depuis cinq ans environ, penche vers l'ouverture des données. Il oscillera peut-être dans l'autre sens dans six mois. L'Etat et ses administrations ont lancé un grand nombre de débats dont on ne peut que se réjouir. Les données de l'Insee, par exemple, sur les ménages et les individus ont été ouvertes aux chercheurs et à la recherche scientifique dès 2008. Le GENES a bataillé pour que les données fiscales anonymisées soient ouvertes également aux chercheurs, les données individuelles telles que celles de nos déclarations d'impôt. Ceci a été acté par la loi en 2013.



Antoine Frachot

Aujourd'hui, nous sommes en train d'effectuer un nouveau mouvement d'ouverture pour les données de santé, données médico-administratives détenues par l'Assurance maladie qui constituent l'une des plus grosses bases de données santé dans le monde : toutes les prescriptions médicales, tous les actes en hôpital, etc.

Ce mouvement est tout à fait bienvenu puisque nous apercevons bien tous les enjeux possibles portant sur la santé publique, l'environnement, l'énergie, l'emploi... Mais se pose la question « comment faire pour sécuriser tout cela ? ».

Nous imaginons parfaitement qu'il suffirait, dans un proche avenir, que l'on apprenne que les données médicales détenues par l'Assurance maladie se soient retrouvées chez des assureurs ou des mutuelles pour qu'il y ait immédiatement une panique générale justifiée. Chaque jour, ou presque, nous apprenons qu'une entreprise s'est fait pirater ses données. Il y a quelques semaines, encore, l'Etat fédéral américain a annoncé que des données portant sur des fonctionnaires américains avaient été piratées. On n'ose pas penser à ce qui pourrait arriver dans nos démocraties si apparaissait une telle fuite à grande échelle.

La loi santé aujourd'hui, par exemple, programme une ouverture assez large. Si des données médicales fuyaient, ce n'est pas seulement la ministre de la Santé qui serait sur la sellette, très probablement obligée de démissionner, mais le gouvernement dans sa totalité serait en difficulté. Plus que ceci : c'est toute la confiance des citoyens vis-à-vis de l'administration qui serait anéantie. Quand nous remplissons nos déclarations fiscales ou quand nous répondons à un enquêteur de l'Insee, nous le faisons plus ou moins volontairement, d'une part parce que nous y sommes obligés, mais d'autre part parce que nous savons bien que les données de l'Insee, ou les données fiscales, ne vont pas aller errer dans la nature. Si demain une fuite, quelle qu'elle soit, se produit, cette confiance disparaît et la démocratie est minée.

Longtemps, le dilemme que je mentionnais plus haut a été vraiment insoluble, ou plutôt souvent résolu par le principe : « il suffit d'interdire d'accéder à des données ». Il est vrai qu'aujourd'hui, même les chercheurs aux intentions les plus pures ont un mal fou à accéder à des données administratives. En restant sur les données de santé, nous avons beaucoup côtoyé des médecins, lors de ces dernières années, puisque nous travaillons à cette ouverture des données de santé avec le ministère de la Santé et d'autres organismes : les médecins nous citent de nombreuses situations où, avec plus de données, des vies humaines auraient pu être sauvées, ou des perspectives d'amélioration considérable du dispositif français de santé. Le Médiateur a énormément marqué le pouvoir politique car, si les données avaient été un peu plus ouvertes à la recherche scientifique, les effets secondaires de ce médicament auraient peut-être pu être détectés bien plus tôt.

Aujourd'hui, nous arrivons un peu plus à sortir par le haut de ce dilemme grâce à la métho-

dologie et à la technologie. La présentation de Philippe Tassi a montré que les techniques de cryptage et de chiffrement permettent de bien réaliser des actions que l'on faisait mal il y a encore dix ou vingt ans. En technologie, nous avons développé, au GENES, un instrument qui s'appelle le centre d'accès sécurisé aux données. Ce CASD permet aujourd'hui à plus d'un millier de chercheurs en France et en Europe de travailler sur des données individuelles françaises, comme celles de l'Insee. Chercheurs, sociologues, économistes peuvent travailler sur ces données grâce à la technologie que nous avons pu développer, sans que jamais ces données ne sortent des serveurs.

Nous sommes donc, heureusement, sur le bon chemin. Avec le mouvement d'open data qui a été initié ces dernières années, la France est en train de reprendre beaucoup d'avance et les chercheurs résidant en France vont pouvoir dans les années qui viennent faire des travaux innovants que d'autres pays ne peuvent pas lancer, le tout avec un niveau de sécurité très élevé.

Pour finir, deux points concernant l'actualité. Après examen en Conseil des Ministres, la loi sur le numérique sera diffusée à la rentrée de septembre. Et le 18 juin, le gouvernement présente son plan numérique, et un premier projet vise à faciliter les appariements de base de données.

Ainsi, pour l'instant, des études sur santé et travail sont considérablement difficiles à faire. Pourquoi ? Parce que, si l'Assurance maladie a des informations très précises sur notre état de santé, ou l'Education Nationale sur les parcours scolaires, et si l'Insee ou les services fiscaux disposent d'éléments sociodémographiques également très précis, marier deux bases est légalement impossible.

Mais « demain », si ces recoupements sont autorisés ou facilités, toujours avec des conditions très strictes, bien sûr, on pourra faire des études fondamentales sur les liens entre la vie professionnelle et le travail avec la santé, sur les inégalités sociales, sur les parcours scolaires.

Il ne s'agit pas d'une renaissance de Safari, mais de réaliser des études à forte valeur ajoutée pour les politiques publiques, et non de « ficher » ou « fliquer » les gens.

Un deuxième projet gouvernemental est également très important, en cours de maturation et

de réflexion : il s'agit de la notion de données d'intérêt général. Ce concept est apparu récemment, il y a un an ou deux. Il consiste à envisager un échange entre système public et système privé, car de nombreuses entreprises privées aimeraient accéder à des données administratives, alors que les administrations et services de l'Etat souhaiteraient utiliser des données privées pour améliorer la connaissance économique et sociale. Par exemple, les données de téléphonie mobile seraient bienvenues pour faire des études sur les déplacements ou le tourisme en France. Les données de la SNCF ou d'EDF seraient également sûrement très intéressantes. On peut aussi penser à des données sur l'efficacité énergétique, la consommation électrique. D'une manière générale, il y a une conviction partagée par beaucoup que des gisements de données d'intérêt général peuvent servir à l'administration et aussi au développement de petites entreprises, de start-up qui veulent créer de nouvelles applications innovantes.

Certes, ce chantier est difficile. Nous serons unanimement d'accord pour dire que les données d'intérêt général doivent être utilisées par le plus grand nombre. Si ce sont vraiment des données d'intérêt général, nous devons les utiliser pour le bien-être général. Cette idée est compliquée à mettre en œuvre tant que ce concept de données d'intérêt général ne sera pas défini très précisément. C'est pourquoi, récemment, le Premier ministre a lancé une mission avec le Conseil d'Etat, l'Inspection Générale des Finances et d'autres services de l'Etat pour essayer de donner de la chair et du sens à cette notion de D.I.G., qui peut aussi conduire à des innovations scientifiques et des progrès en matière de politique publique tout à fait considérables.

Ces chantiers sont passionnants. Et c'est pourquoi je remercie vraiment, du fond du cœur, Catherine Grandcoing, Philippe Tassi et tous les organisateurs, pour avoir lancé ce débat et cette réflexion en regardant les deux faces du même sujet, et pas uniquement l'aspect « sécurité des données ». Ces deux aspects sont intimement liés, nous devons absolument les affronter pour en tirer du bien. ■