LA PROTECTION DES DONNÉES PERSONNELLES DANS L'*OPEN DATA* : UNE EXIGENCE ET UNE OPPORTUNITÉ

Mission d'information de la commission des lois du Sénat

Rapporteurs :

M. Gaëtan GORCE
Sénateur (socialiste) de la Nièvre

M. François PILLET
Sénateur (ratt. UMP) du Cher

16 avril 2014

SOMMAIRE

Pages

LISTE DES RECOMMANDATIONS

- *Accélérer le déploiement d'un* open data *respectueux de la protection des données personnelles*

Recommandation n° 1

Poser le principe que l'administration est tenue de mettre en ligne progressivement, en les anonymisant si nécessaire, toutes les bases de données qu'elle détient et qui seraient susceptibles d'être communiquées à un citoyen s'il en fait la demande ou qui font l'objet d'une diffusion publique sur un autre support

L'administration ne pourrait s'y opposer qu'en raison des coûts déraisonnables de gestion que cette mise en ligne imposerait (notamment les coûts d'anonymisation éventuelle), ou du risque avéré, qu'en dépit des précautions prises, des informations personnelles puissent être ré-identifiées

Recommandation n° 2

Afin de permettre aux administrations de satisfaire à l'obligation précédente, mettre en place une phase transitoire, pendant laquelle elles :

- opèreraient une recension complète des jeux de données qu'elles détiennent et décideraient de leur mise en ligne ;

- publieraient un calendrier pluriannuel des mises à dispositions programmées

Recommandation n° 3

Imposer aux administrations d'indiquer, pour chaque jeu de données, en marge du registre publié sur leur site internet les énumérant, s'il fera ou non l'objet d'une mise en ligne et, dans ce dernier cas, la raison pour laquelle elles s'y opposent

Recommandation n° 4

Le cas échéant, examiner l'opportunité d'étendre les cas, définis par la loi, dans lesquels, compte tenu de l'intérêt général qui s'y attache, des jeux de données incluant des données personnelles peuvent, par exception, être diffusés en ligne et ouverts aux réutilisations

• *Mettre en œuvre une doctrine de protection des données personnelles en matière d'*open data

** Anticiper et évaluer*

Recommandation n° 5

Prévoir, dès la conception de la base, dans la perspective de sa possible ouverture :

- les modalités de son anonymisation éventuelle ;

- le cas échéant, le marquage des jeux de données afin d'être en mesure de suivre les réutilisations éventuelles et dénoncer les mésusages

Recommandation n° 6

Procéder, préalablement à tout examen de l'opportunité d'ouvrir une base de données, ainsi, le cas échéant, qu'à intervalles réguliers, à une analyse du risque de ré-identification et des conséquences possibles d'une telle ré-identification

** Adapter la diffusion en fonction du risque*

Recommandation n° 7

En cas de risque avéré sur les données personnelles, impossible à éliminer par des procédés d'anonymisation, refuser l'ouverture des données ou, si le bénéfice social attendu de cette ouverture est jugé trop important, procéder à une ouverture restreinte de cette base

Recommandation n° 8

Concevoir à cette fin un *continuum* de solutions d'accès aux données, allant de l'*open data*, jusqu'aux modes d'accès les plus sélectifs

** Assurer une veille sur la diffusion et les réutilisations des données mises en ligne*

Recommandation n° 9

Assurer une veille sur la diffusion et les réutilisations des données publiques, en facilitant notamment les procédures par lesquelles un réutilisateur peut alerter l'administration compétente

Recommandation n° 10

Assurer aussi cette veille sur les données publiées par des tiers sur les sites publics

Recommandation n° 11

Prévoir que l'administration définisse une stratégie de rapatriement ou de suppression des jeux de données compromis, afin de remédier rapidement à la diffusion accidentelle d'informations personnelles

** Renforcer la protection offerte par la licence de réutilisation*

Recommandation n° 12

Exclure expressément les données personnelles du champ d'application de la licence ouverte utilisée par les administrations pour la réutilisation des données publiques

Recommandation n° 13

Interdire expressément dans le contrat de licence toute réutilisation abusive qui aboutirait à lever l'anonymisation des données

Recommandation n° 14

Intégrer au contrat de licence, une clause de suspension légitime du droit de réutilisation, ainsi que de suppression ou de rapatriement des jeux données compromis lorsqu'un risque de ré-identification est apparu

• *Adapter la gouvernance de l'*open data *aux exigences de la protection des données personnelles*

** Renforcer l'assistance aux acteurs de l'*open data

Recommandation n° 15

Mettre en place, auprès de la mission *Etalab*, une structure dédiée à la protection des données à caractère personnel et chargée d'assister les administrations :

- dans l'élaboration de l'étude d'impact préalable à la mise à disposition des données ;

- dans l'anonymisation éventuelle de la base ;

- dans la mise en place d'un mode d'accès restreint

Recommandation n° 16

Confier, à cette même structure, un rôle de veille sur les réutilisations abusives au regard de la protection des données personnelles, en la chargeant de recueillir les alertes éventuelles, d'en informer la CNIL, et de coordonner, le cas échéant, le retrait ou la reconfiguration de la base de donnée litigieuse

Recommandation n° 17

Rassembler et diffuser les bonnes pratiques et les recommandations en matière de protection des données personnelles dans l'*open data*

Recommandation n° 18

Investir les CIL et les PRADA d'attribution de coordination et de veille en matière de protection des données à caractère personnel dans le cadre de l'*open data*

** Garantir le financement des mesures d'anonymisation*

Recommandation n° 19

Garantir le financement par l'État des mesures d'anonymisation des données personnelles contenues dans des jeux de données publiques

Ne pas renoncer par principe au prélèvement d'une redevance en présence de coûts d'anonymisation élevés

Encourager le financement coopératif de l'anonymisation

** Clarifier le droit applicable*

Recommandation n° 20

Préciser que, lorsque des données personnelles sont mises en ligne en vertu de la loi, cette publication doit se limiter à la stricte mesure nécessaire au respect de l'objet visé par cette loi

AVANT-PROPOS

Mesdames, Messieurs,

Notre pays s'est résolument engagé sur la voie de l'ouverture et du partage des données publiques, plus connue sous le nom d'*open data*.

Deux idées animent cette politique. Comptables de leur gestion auprès des citoyens, les administrations leur ouvrent leurs fichiers. Elles leur donnent ainsi le moyen de mieux les contrôler. Par ailleurs, à l'ère du numérique, où l'information est source de richesse, elles leur offrent l'opportunité d'exploiter le formidable gisement que constituent ces données.

En créant une mission d'information chargée d'étudier l'*open data* et la protection de la vie privée de nos concitoyens, votre commission des lois a souhaité poursuivre une réflexion déjà engagée par la précédente mission d'information confiée à M. Yves Détraigne et Mme Anne-Marie Escoffier consacrée à « *La vie privée à l'heure des mémoires numériques* »[1] . Cette réflexion porte sur les nouveaux usages numériques et la façon dont ils peuvent se concilier avec les principes fondamentaux que le législateur a posés dès la fin des années 1970.

L'*open data* pose à cet égard une question spécifique : en principe, il exclut toute diffusion de données à caractère personnel, mais bien souvent, les données détenues par les administrations ont été élaborées à partir d'informations individuelles, qui peuvent être retrouvées grâce aux formidables capacités de traitement que permet l'informatique moderne. L'impératif de protection de la vie privée est-il en mesure de toujours prévaloir ? Comment s'en assurer ?

Telles sont les interrogations qui ont guidé vos deux rapporteurs, MM. Gaëtan Gorce et François Pillet, retenant l'angle exclusif de la protection des données personnelles, et renvoyant les questions plus générales à la mission commune d'information, présidée par notre collègue

[1] *La vie privée à l'heure des mémoires numériques. Pour une confiance renforcée entre citoyens et société de l'information, rapport d'information n° 441, (2008-2009), de M. Yves Détraigne et Mme Anne-Marie Escoffier, fait au nom de la commission des lois (disponible à l'adresse suivante : http://www.senat.fr/notice-rapport/2008/r08-441-notice.html).*

Jean-Jacques Hyest et rapportée par notre collègue Corinne Bouchoux, consacrée à l'accès aux documents administratifs et aux données publiques[1].

Après avoir recueilli, au cours de plus d'une trentaine d'auditions, les avis et les analyses d'experts et de représentants de l'ensemble des acteurs de l'*open data* –administration, associations de citoyens, entreprises, autorité de régulation–, vos rapporteurs jugent aujourd'hui nécessaire d'ouvrir la voie à un déploiement raisonné de l'*open data*, protecteur de la vie privée de nos concitoyens et conforme aux ambitions de transparence et d'intérêt général qui l'animent.

[1] Cf. *le site dédié à cette mission commune d'information à l'adresse suivante : http://www.senat.fr/commission/missions/acces_aux_documents_administratifs_et_aux_donnees_publiques/index.html*

TITRE LIMINAIRE – UNE RÉFLEXION NÉCESSAIRE

*A. L'*OPEN DATA *: UN MOUVEMENT OFFENSIF*

Depuis quelques années maintenant, l'*open data* fait l'objet d'un engouement tel que l'on a tendance à ne plus même s'interroger sur ce que recouvre cette expression anglo-saxonne, au risque de quiproquos.

Aussi est-il apparu nécessaire à vos rapporteurs de commencer par tenter de mieux cerner ce concept en remontant d'abord son cours jusqu'à ses origines (1), avant de s'attacher à décrire sa traduction en France (2). Les nombreuses auditions qu'ils ont menées leur ont enfin permis de mesurer les attentes que porte ce mouvement dans notre société (3).

1. Les origines de l'*open data*

Ainsi que l'expression même d'*open data* le laisse supposer, le mouvement de l'*open data* puise ses racines dans le monde anglo-saxon (a). Il a cependant largement essaimé depuis lors, se chargeant de sens successifs, ce qui rend nécessaire une clarification du concept (b).

a) Un concept venu du monde anglo-saxon

Le concept d'*open data* est né dans le milieu de la recherche scientifique publique. L'expression « *open data* » elle-même apparaît pour la première fois, en 1995, dans une publication du *National Research Council* américain relative à l'ouverture des données géophysiques et environnementales[1]. Selon ses auteurs, deux arguments plaident pour l'ouverture et le partage du résultat de leurs travaux : la nature transfrontière des phénomènes observés et des enjeux, ainsi que la crainte d'une privatisation des connaissances. En cela, l'*open data* rejoint la théorie économique des biens communs, ces biens non appropriables de manière exclusive qui appartiennent donc à tous.

La transposition de ce concept aux données issues non plus de la recherche mais de la gestion par les administrations s'opère dans les années 2000 sous l'impulsion de la théorie libérale anglo-saxonne, qui établit un continuum entre le politique et l'économique. Ainsi, deux types de bénéfices sont attendus de l'*open data* :

[1] *Cité dans l'ouvrage de Simon Chignard,* Open data – Comprendre l'ouverture des données publiques*, FYP éditions, France, 2012. Les développements de cette partie s'appuient sur cet ouvrage ainsi que sur l'audition par vos rapporteurs de son auteur.*

- dans le champ politique, l'*open data* rejoint l'exigence de transparence et de responsabilité, il est donc conçu dans une optique de revitalisation de la démocratie *via* la participation citoyenne ; l'*open data* est ainsi un instrument de l'*open government* mis en avant par le Président Obama dès le premier jour de son mandat en 2009 ;

- dans le champ économique, l'*open data* est envisagé comme facteur d'innovation, de création de nouveaux services, y compris publics, de contribution à la croissance, d'amélioration de la vie quotidienne.

Bien que l'exemple américain soit souvent mis en avant, l'Europe n'est pas restée à l'écart de ce mouvement. La mise en ligne, en janvier 2010, du site *data.gov.uk* par le gouvernement britannique de David Cameron, qui mettait ainsi ses pas dans ceux de son prédécesseur Gordon Brown, a suivi de peu celle du site américain *data.gov*. L'Union européenne elle-même s'était d'ores et déjà emparée de cette problématique des données des administrations avec l'adoption dès le 17 novembre 2003 de la directive 2003/98/CE du Parlement européen et du Conseil concernant la réutilisation des informations du secteur public. Il convient toutefois de noter que cette directive renvoie essentiellement à la seconde dimension économique de l'*open data*. En outre, la directive n'évoque pas à proprement parler l'*open data*, mais la réutilisation des données.

b) Essai de définition

La principale difficulté avec le concept d'*open data* est en effet sa polysémie, qui explique les hésitations sur sa traduction en français. À la fois mouvement à l'œuvre dans les politiques actuelles et injonction adressée aux acteurs publics pour davantage de transparence et de responsabilité, l'expression renvoie également de manière plus statique aux caractéristiques propres aux données ainsi « libérées ».

En s'inspirant de la liste des « huit principes des données du gouvernement ouvert » énumérés par une trentaine de « penseurs du web » en septembre 2007[1], on peut caractériser les données publiques car issues des administrations et ouvertes au partage, par un ensemble de critères techniques, juridiques et économiques :

- la mise à disposition dans un format technique le plus ouvert possible, qui facilite la réutilisation et n'impose pas l'utilisation d'un logiciel propriétaire ;

- l'utilisation de licences juridiques ouvertes, qui ne restreignent pas ou peu les utilisations possibles des données ;

- la limitation des redevances susceptibles de constituer des freins économiques pour les personnes réutilisant ces données.

[1] Cf. « Eight Principles of Open Governement Data », *8 décembre 2007, disponible à l'adresse suivante :* www.opengovdata.org

L'*open data* ne se résume donc pas à la simple publication des données des administrations publiques, il nécessite la mise en place d'outils techniques et juridiques adéquats par une politique volontariste.

2. La stratégie française d'ouverture et de partage des données publiques

À l'issue d'une gestation de plusieurs années, la stratégie française d'ouverture et de partage des données publiques ou *open data* a pris définitivement corps avec le lancement du portail « *data.gouv.fr* ». Initialement pensée dans un objectif essentiellement économique, cette stratégie a peu à peu été également intégrée à la modernisation de l'État et à l'amélioration des relations entre les usagers et les administrations.

*a) Les étapes de l'élaboration de la stratégie française d'*open data

Dès les années 1970, la France a institué un droit d'accès aux documents administratifs – érigé par le Conseil d'État en 2002 en liberté publique[1]. Elle a peu à peu diffusé les données détenues par ses administrations, notamment en identifiant les « *données publiques essentielles [devant] être accessibles à tous gratuitement sur internet* », pour reprendre la formule de M. Lionel Jospin, alors Premier ministre, dans son discours d'Hourtin, le 25 août 1997. Cependant, la valeur de ses données et l'intérêt de leur réutilisation ne se sont réellement révélées à l'administration qu'avec l'adoption de la directive européenne de 2003 et la parution en 2006 du rapport de MM. Maurice Lévy et Jean-Pierre Jouyet sur l'économie de l'immatériel[2]. Cette prise de conscience est à l'origine de la création d'un service à compétence nationale dénommé « Agence du patrimoine immatériel de l'État » (APIE), par un arrêté du 23 avril 2007.

Les débuts de la définition d'une stratégie étatique d'*open data* datent ainsi de la présentation, le 20 octobre 2008, du plan « France Numérique 2012 » par M. Éric Besson, secrétaire d'État chargé de la prospective, de l'évaluation des politiques publiques et du développement de l'économie numérique. Ce plan promouvait la diffusion des contenus publics et patrimoniaux et entendait « *favoriser la réutilisation des informations publiques par les agents économiques* », l'objectif étant de « *développer de nouveaux produits et services, contribuant ainsi à la croissance de l'économie numérique* »[3]. L'action n° 39 de ce plan consistait ainsi en la création d'un portail unique d'accès aux données publiques, conçu comme réponse aux attentes exprimées par différents acteurs en faveur de l'accès aux informations publiques. Ce portail

[1] *CE, 29 avril 2002, U., n° 228830.*

[2] L'économie de l'immatériel, la croissance de demain, *rapport remis par MM. Maurice Lévy et Jean-Pierre Jouyet au ministre de l'économie, des finances et de l'industrie, novembre 2006.*

[3] Cf. *Secrétariat d'État à la prospective, à l'évaluation des politiques publiques et au développement de l'économie numérique,* France numérique 2012 - Plan de développement de l'économique numérique, *La Documentation française, octobre 2008.*

unique donnant accès aux sites ministériels et aux informations qui y sont proposées devait s'accompagner de l'adoption de systèmes de recherche standardisés et de la définition de métadonnées et de référentiels communs. L'étude de pré-configuration de ce portail était confiée à l'APIE. Celle-ci se voyait également chargée de la rédaction de licences types de réutilisation des données publiques par l'action n° 41 dudit plan.

Décidées par le conseil de modernisation des politiques publiques le 30 juin 2010, la création et la mise en ligne du portail unique *data.gouv.fr* furent annoncées à l'issue du conseil des ministres du 24 novembre 2010. La mise en œuvre concrète prit la forme du décret du 21 février 2011 portant création de la mission « *Etalab* », placée auprès du Secrétariat général du gouvernement, puis d'une circulaire du Premier ministre, alors M. François Fillon, en date du 26 mai 2011. Cette dernière rappelait en ces termes le double objectif assigné à la stratégie d'ouverture des données publiques :

- « *encourager l'innovation par toute la communauté des développeurs et des entrepreneurs pour soutenir le développement de l'économie numérique* » ;

- « *contribuer à renforcer la transparence de l'action de l'État, mettre en valeur le travail des administrations et éclairer le débat public* ».

Pour ce faire, le Premier ministre insistait sur l'importance de « *permettre la réutilisation des informations publiques la plus facile et la plus large possible* ».[1]

Le site *data.gouv.fr* a été lancé le 5 décembre 2011.

La démarche d'ouverture et de partage des données publiques a été confirmée par le nouveau Gouvernement en octobre 2012 à l'occasion de l'intégration de la mission *Etalab* au sein du nouveau Secrétariat général pour la modernisation de l'action publique (SGMAP), créé par un décret en date du 30 octobre 2012. L'o*pen data*, désormais perçu comme « *vecteur de renouvellement démocratique, d'innovation pour l'économie et la société, et levier de transformation pour les administrations* », est à ce titre érigé en « *l'un des axes de la nouvelle modernisation de l'action publique* »[2]. À l'issue du séminaire gouvernemental sur le numérique du 28 février 2013, la politique d'ouverture des données publiques est déclinée en une série de mesures, anticipant l'adoption par le G 8 d'une charte sur l'ouverture des données publiques lors du sommet des 17 et 18 juin 2013.

[1] *Circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'État* data.gouv.fr *par la mission* Etalab *et l'application des dispositions régissant le droit de réutilisation des informations publiques.*

[2] Cf. *le communiqué de presse du Premier ministre en date du 31 octobre 2012.*

*b) Les instruments de l'*open data *: la mission* Etalab*, le portail* data.gouv.fr *et les licences*

L'*open data* traduit le passage de l'obligation de communication de documents administratifs sur demande des usagers à la mise à disposition de tous de certaines données des administrations – d'une logique de demande à une logique d'offre. Du fait du changement de modèle qu'il implique, l'*open data* a rendu indispensable la mise en place de nouveaux instruments.

(1) Une mission animant un réseau de correspondants

Parce que l'*open data* représente une révolution au sein de l'administration, il a été nécessaire de mettre en place une structure dédiée afin de préparer et accompagner cette stratégie : la mission *Etalab*.

Le décret du 21 février 2011 a confié à *Etalab* deux missions.

En premier lieu, *Etalab* a pris la suite de l'APIE pour la création du portail unique interministériel *data.gouv.fr*. Elle en est le développeur ainsi que l'hébergeur.

En second lieu, l'article 3 du décret a chargé *Etalab* d'une mission de coordination de l'action des administrations de l'État et d'appui à ses établissements publics administratifs.

Dans ce cadre, la mission *Etalab* a tout d'abord piloté la rédaction de la « Licence ouverte » (voir *infra*). Elle s'est ensuite vu confier l'organisation de débats thématiques dont l'objectif est d'identifier les jeux de données les plus pertinents au regard des bénéfices attendus. Ces débats associent donc la société civile et les différentes parties prenantes. En 2013, six débats ont été organisés, relatifs à l'ouverture des données de santé, d'éducation, de dépenses publiques, du logement, de l'environnement et des transports. L'assistance aux administrations prend enfin la forme de la parution de différents documents, parmi lesquels le *Vade-mecum sur l'ouverture et le partage des données publiques*, adressé par voie de circulaire du Premier ministre en septembre 2013[1] et destiné à faciliter l'appropriation concrète de la politique d'*open data* par chaque administration.

Par ailleurs, la mission *Etalab* est chargée de stimuler la recherche et l'innovation. Elle a ainsi instauré depuis 2012 le programme « DataConnexions » et organise dans ce cadre, à intervalles réguliers, des concours récompensant les meilleures réutilisations de données publiques. Le programme « Datalift », également mis en place par la mission, est davantage orienté vers la recherche afin d'encourager l'usage des données publiques pour la recherche ainsi que la recherche en sciences des données.

[1] *Circulaire n° 5677/SG du 17 septembre 2013 et vade-mecum disponible à l'adresse suivante :* http://www.modernisation.gouv.fr/laction-publique-se-transforme/en-ouvrant-les-donnees-publiques/lopen-data-son-vade-mecum

Structure légère composée de sept personnes, la mission *Etalab* s'appuie sur un réseau de coordinateurs et correspondants dans chaque administration. En effet, chaque ministère est responsable de la mise en œuvre de la stratégie d'*open data* pour ses propres données et a désigné un interlocuteur unique pour *Etalab*. La coordination de la stratégie d'ensemble est donc assurée par un comité de pilotage réunissant les coordinateurs *open data* nommés auprès de chaque secrétaire général de chaque ministère.

(2) Un portail unique

En application de la feuille de route fixée dès le plan « France numérique 2012 », a été créé « *un portail unique interministériel destiné à rassembler et à mettre à disposition librement l'ensemble des informations publiques de l'État, de ses établissements publics administratifs et, si elles le souhaitent, des collectivités territoriales et des personnes de droit public ou de droit privé chargées d'une mission de service public* ».[1]

Cette plateforme a été dénommée *data.gouv.fr*. Elle a été lancée dans une première version le 5 décembre 2011. Une nouvelle version a été mise en ligne le 18 décembre 2013 à l'occasion du comité interministériel pour la modernisation de l'action publique (CIMAP). Désormais, le portail est ouvert aux « *forces vives de la société* » : la nouvelle version du site accueille aussi bien les données publiques issues des autorités administratives que les données produites par « *la société civile (citoyens, associations,…), les innovateurs, les chercheurs, les journalistes, etc* ». L'objectif du portail n'est donc plus la simple mise à disposition des données publiques, mais la « *coproduction de données d'intérêt public* ».

(3) Un jeu de licences

Conformément à l'action n° 41 du plan « France numérique 2012 », des licences ont été élaborées par la mission *Etalab* et l'APIE, avec l'assistance du Conseil d'orientation de l'édition publique et de l'information administrative (COEPIA) et des administrations concernées, afin de favoriser la réutilisation libre et gratuite des données publiques, notamment celles mises en ligne sur le site *data.gouv.fr*.

L'ensemble des données mises à disposition sur le portail le sont sous le régime de la « Licence ouverte ». Cette licence permet au réutilisateur de :

- reproduire, copier, publier et transmettre l'information ;

- la diffuser et redistribuer ;

- l'adapter, la modifier, procéder à des extractions, la transformer ;

- l'exploiter à titre commercial ;

[1] Cf. *article 2 du décret n° 2011-194 du 21 février 2011 portant création d'une mission Etalab chargée de la création d'un portail unique interministériel des données publiques.*

sous réserve de la mention de sa « paternité » : source et date de mise à jour.

Des licences spécifiques permettent de subordonner la réutilisation de certaines données à des conditions particulières. Ces licences, élaborées par les administrations avec l'assistance de l'APIE qui a mis à leur disposition des modèles, sont validées par la mission *Etalab* qui les publie sur le site *data.gouv.fr*. De telles licences peuvent notamment être utilisées dans le cas où la réutilisation des données serait soumise à redevance. Cependant, il convient de noter que le comité interministériel pour la modernisation de l'action publique (CIMAP) du 18 décembre 2013 a réaffirmé le principe de la gratuité de la mise à disposition et du partage des données publiques. Il a décidé à ce titre de ne plus autoriser la création de nouvelle redevance et de supprimer plusieurs redevances existantes. Il a en outre précisé sa doctrine en matière d'exceptions au principe de gratuité en disposant qu'« *aucune redevance ne saurait être exigée sur les données résultant des missions de service public des administrations générales* » et que « *les opérateurs dont la mission même est de produire des données doivent rechercher des modèles économiques leur permettant de faire face à un paysage économique en profonde reconstitution* »[1].

c) Vers une obligation d'ouverture et de partage des données publiques pour les collectivités territoriales ?

Comme on a pu le voir, le mouvement de l'*open data* a été initié au niveau national sous l'impulsion de circulaires des Premiers ministres successifs, sans qu'il ait été besoin de recourir à la loi. Au niveau local, il est revenu à chaque collectivité de suivre, voire de devancer, l'État dans la mise en place de stratégies d'*open data* –la ville de Rennes faisant figure de précurseur en la matière– si bien qu'aujourd'hui une cinquantaine de collectivités, à tous les échelons, s'est aujourd'hui dotée de portails *open data*. Il semble cependant qu'une étape doive être franchie avec l'introduction dans le droit d'une obligation pour les collectivités territoriales de diffusion et de mise à disposition des données.

L'article 29 du projet de loi de développement des solidarités territoriales et de la démocratie locale (n° 497, 2012-2013), déposé le 10 avril 2013 sur le Bureau du Sénat, tend en effet à rendre obligatoire pour les collectivités territoriales de 3 500 habitants et plus, ainsi que pour les établissements publics de coopération intercommunale à fiscalité propre auxquels elles appartiennent, la mise à disposition des informations publiques se rapportant à leur territoire et dont ils disposent au format électronique, par une mise en ligne sur internet. L'étude d'impact qui accompagne le projet de loi précise que seraient concernées par cette obligation les « *données économiques, sociales, démographiques et territoriales* », et « *notamment visés les rapports accompagnant les documents budgétaires (budget*

[1] Cf. *le relevé de décisions du comité interministériel pour la modernisation de l'action publique du 18 décembre 2013.*

primitif, compte administratif) ou servant de base aux débats sur les orientations budgétaires ».

Selon l'étude d'impact, l'objectif poursuivi par cette disposition est multiple : il serait d'abord démocratique pour « *permettre une plus grande diffusion de l'information publique concernant le fonctionnement des collectivités territoriales au profit des citoyens* » et « *renforcer la confiance placée par le public dans les élus locaux* », mais également de bonne administration en « *simplifi[ant] l'accès des administrés à l'information* » et en « *amélior[ant] le fonctionnement des administrations qui sont les premières utilisatrices des données publiques* », enfin économique en visant à « *faciliter le développement économique et notamment la création de services innovants* » et à « *contribuer au rayonnement et à l'attractivité économique des collectivités territoriales* ». On retrouve ici les trois bénéfices attendus de l'*open data*.

3. Une forte demande de la part de la société civile et des entreprises

Les auditions conduites par vos rapporteurs ont démontré que la mise en œuvre par l'État et les collectivités territoriales de stratégies d'*open data* répond effectivement à une demande exprimée par la société civile et les entreprises, qu'elle soit davantage orientée vers un objectif politique (a) ou économique (b).

a) Un enjeu démocratique de transparence et de bonne administration

Certaines associations telles Regards Citoyens mettent en avant « *qu'en démocratie, les prises de décisions publiques sont transparentes* ». Mettre à disposition les données publiques mais également autoriser leur libre réutilisation permettrait au citoyen d'exercer un meilleur contrôle de l'action de leurs représentants mais aussi de l'administration, en adéquation avec l'article 15 de la Déclaration des droits de l'homme et du citoyen qui proclame : « *la Société a le droit de demander compte à tout Agent public de son administration*. » Dans la lignée des lois de transparence adoptées dans les années 1970, l'*open data* contribuerait à rendre l'administration moins opaque et plus accessible.

En outre, les promoteurs de l'*open data* voient en celui-ci un moyen d'améliorer l'information des citoyens et, partant, leur participation aux processus décisionnels. La réutilisation des données publiques par les usagers du service public peut en effet aller de leur simple enrichissement au développement de nouveaux services. Cette dernière modalité de participation des citoyens serait d'ailleurs, davantage que la transparence, le principal objectif poursuivi par les collectivités territoriales, selon M. Simon Chignard.

Ainsi, les données mises à disposition les plus utilisées et les plus prometteuses en termes d'applications innovantes à l'heure actuelle seraient

les données de mobilité ou encore les données événementielles en temps réel, comme l'a indiqué l'association *Open Data France* qui regroupe les collectivités territoriales engagées dans une démarche d'ouverture des données publiques. Si les données budgétaires font l'objet d'une forte demande, les associations font état de la difficulté de les exploiter du fait d'un défaut de normalisation des documents budgétaires qui empêcheraient les comparaisons entre différentes collectivités.

Enfin, *l'open data* permettrait de mieux évaluer l'efficacité des politiques publiques pour proposer des mesures d'amélioration, ainsi que l'a expliqué la Fondation iFRAP.

b) Un enjeu de valorisation économique et sociale des données

Bien que les modèles mettent en évidence un bénéfice socio-économique différé et évalué de manière incertaine de l'*open data* pour un coût immédiat pour l'administration[1], les acteurs économiques reçus par vos rapporteurs soulignent le gisement potentiel de valeur lié à la mise à disposition des données publiques.

Ainsi, pour la Fédération française des sociétés d'assurances (FFSA), le principal bénéfice à attendre de l'*open data* est une meilleure connaissance des risques assurables, qu'il s'agisse de la matière assurable – les biens à assurer – ou des caractéristiques des risques assurés. Davantage encore que des données personnelles anonymisées et agrégées, ce sont donc les données météorologiques, géographiques ou géologiques qui permettent d'améliorer la gestion des risques et la prévention : « *plus l'information sur l'exposition au risque d'un individu, d'une entreprise ou d'une collectivité est grande et fine, plus la gestion de ce risque est simple et moins les conséquences sont graves et coûteuses* ».

L'économie des données n'est cependant pas que potentielle, elle a suscité la création de plusieurs « start-up ». C'est ce qu'ont confirmé à vos rapporteurs MM. François Bancilhon et Jean-Marc Lazard, présidents directeurs généraux de deux entreprises. L'une, *Data Publica* travaille à produire des jeux de données à partir de sources diverses, dont les données publiques mises à disposition *via* l'*open data*. L'autre, *OpenDataSoft*, assiste des collectivités publiques pour automatiser le traitement de données en vue de leur ouverture et mise à disposition ; elle met également son expertise au service d'entreprises désireuses d'innover.

[1] Cf. Ouverture des données publiques – Les exceptions au principe de gratuité sont-elles toutes légitimes ?, *rapport remis au Premier ministre par M. Mohammed Adnène Trojette, juillet 2013.*

En dehors du secteur marchand, M. Mathieu Escot, représentant de l'UFC-Que Choisir, a fait valoir que l'*open data* présente l'avantage, pour sa propre activité, de faciliter l'accès à l'information pour assurer une meilleure défense des consommateurs. Il a souligné en outre que si la mise à disposition gratuite de données publiques peut permettre la création de nouveaux services marchands, elle également le développement de services non-marchands, encouragé par l'absence de barrière tarifaire pour l'accès aux données.

La principale vertu de l'*open data* serait donc la garantie d'un égal accès de chacun, particulier ou entreprise, aux mêmes données.

*B. L'*OPEN DATA *ET LA PROTECTION DES DONNÉES PERSONNELLES : UNE INTERROGATION LÉGITIME*

La création par votre commission de la présente mission d'information ainsi que la question posée ont suscité des réserves, voire des craintes, parmi les défenseurs de l'*open data*. Ceux-ci s'en sont ouverts à vos rapporteurs lors de leurs auditions : l'insistance sur la protection des données personnelles n'aurait-elle pas caché une opposition au principe-même de l'*open data* ?

Sans doute, l'enthousiasme légitime qu'a suscité l'engagement du Gouvernement en faveur de l'ouverture des données publiques, a-t-il nourri une méfiance *a priori* contre tout ce qui pouvait sembler freiner ou contrecarrer ce mouvement.

Toutefois, un tel procès d'intention était bien entendu sans fondement : l'objet de la mission d'information est justement de promouvoir un *open data* respectueux de la protection des données personnelles, et d'inciter à son déploiement. D'ailleurs, vos rapporteurs constatent, à l'issue de leurs travaux, que les arguments parfois employés pour dénoncer l'intérêt d'une telle interrogation, sont peu pertinents, quoiqu'énoncés de bonne foi (1). La question dont s'est saisie votre commission par le biais de cette mission d'information est pleinement légitime, ce qu'illustre l'attention dont elle fait l'objet en France, comme ailleurs en Europe (2).

1. Une nécessité : se garder de certains simplismes

Ceux qui contestent la pertinence de la question posée par la mission d'information ont développé deux arguments à l'appui de cette appréciation. Ceux-ci paraissent toutefois réducteurs et n'emportent pas la conviction.

a) Un problème résolu, avant même de le poser ?

Pour certains, la question de la mise en danger de données personnelles par l'*open data* ne se pose tout simplement pas.

Les représentants de Regards citoyens ont ainsi fait valoir que l'*open data* ne concerne que les données publiques, lesquelles excluent, par définition, les données personnelles ou portent seulement sur des informations personnelles que la loi fait obligation de publier. Nulle atteinte ne peut dès lors être portée en principe à la vie privée par l'ouverture des données publiques si celle-ci est accomplie correctement. Par ailleurs, la ré-identification par un tiers de données personnelles anonymisées, diffusées dans une base de données publiques, signalerait moins un problème propre de l'*open data*, qu'une infraction, par ce tiers, de la législation relative aux données personnelles.

Le directeur de la mission *Etalab*, M. Henri Verdier a partagé ce raisonnement, considérant que par essence, l'*open data* devrait exclure toute diffusion de données personnelles.

L'examen des bases de données aujourd'hui mises en ligne sur les sites internet des grandes administrations corrobore largement cette impression : la très grande majorité des bases portent sur des statistiques ou des données agrégées, qui ne présentent pas de lien avec des informations personnelles : chiffres de l'INSEE ou d'Eurostat, information géographique, renseignements d'ordre général sur les procédures, l'organisation ou le budget des administrations *etc*.

Pour autant, comme l'ont observé les représentantes de l'association pour le développement de l'informatique juridique, Mmes Élise Debiès et Nathalie Metallinos, l'argument ne saurait convaincre, pour au moins deux raisons.

Tout d'abord, la loi prévoit parfois expressément que des informations personnelles soient publiées. Il en va ainsi, par exemple, des délibérations et autres actes des collectivités territoriales, même si elles contiennent des informations nominatives[1]. S'assurer que cette diffusion s'accomplit dans des conditions satisfaisantes pour le respect de la vie privée des intéressés est pertinent.

Surtout, le fait qu'un document publié contienne des données personnelles ou identifiantes qui ne devraient pas être diffusées peut avoir échappé à l'administration qui l'a mis en ligne, que le procédé d'anonymisation auquel elle a recouru ait été déficient, ou que le caractère personnel de ces données ne lui soit pas apparu.

[1] *Art. L. 2121-26, L. 3121-17n L. 4132-16, L. 5211-46, L. 5421-5, L. 5621-9 et L. 5721-9 du code général des collectivités territoriales. Les documents nominatifs correspondent notamment à des arrêtés individuels relatifs aux agents. Si les informations personnelles doivent en principe être occultées (CE, 10 mars 2010,* Commune de Sète*, n° 308314), la CADA rappelle que cette occultation concerne seulement les appréciations portées sur l'agent public (avis 20101311 du 25 mars 2010).*

Le cas de la publication des aides reçues par les agriculteurs dans le cadre de la politique agricole commune en fournit une illustration. D'autres exemples seront développés dans les développements qui suivent[1].

Deux règlements européens font obligation[2] aux États membres de publier annuellement la liste des bénéficiaires des aides européennes versées aux agriculteurs par le fonds européen agricole de garantie (FEAGA) et le fonds européen agricole pour le développement rural (FEADER).

Saisie d'une question préjudicielle sur la conformité de cette obligation avec les articles 7 et 8 de la charte des droits fondamentaux, qui protègent la vie privée, la Cour de justice de l'Union européenne a toutefois estimé que l'atteinte portée aux personnes physiques excédait ce qui était acceptable au nom de l'impératif de transparence[3]. En revanche, elle a considéré que tel n'était pas le cas pour les personnes morales, compte tenu à la fois des obligations déclaratives qui étaient les leurs et de la contrainte que représenterait, pour les administrations, la charge de trier, parmi ces sociétés, celles dont le nom permettrait ou pas de déterminer l'identité de leurs sociétaires.

Conformément à cette décision, les données aujourd'hui publiées sur le site *data.gouv.fr* ne mentionnent plus que les subventions versées à des personnes morales. Pourtant, vos rapporteurs ont pu constater que le fichier publié incluait à la fois la localisation et le montant des subventions versées non pas à une personne morale, mais à une indivision successorale nommément désignée, qui rassemble les différents héritiers personnes physiques ou bien à plusieurs particuliers, propriétaires d'un terrain en indivision. La référence à une telle indivision, ainsi qu'à son adresse, est susceptible de permettre l'identification des personnes concernées. La conformité d'une telle mention avec la règle posée par la Cour de justice de l'Union européenne pourrait être discutée.

Ainsi, même s'il est acquis qu'en théorie, l'*open data* ne doit pas porter atteinte à la vie privée des administrés, cette assertion ne se vérifie pas toujours en pratique. Toute la question est justement de s'assurer qu'aucune donnée personnelle ne pourra être accidentellement diffusée à l'occasion de la mise en œuvre de l'*open data*, ni faire l'objet d'une ré-identification facilitée par un tiers. Quelle procédure mettre en place pour éviter ce type de

[1] Cf. infra, *II. A.*

[2] *Art. 42, point 8* ter, *et 44* bis *du règlement (CE) n° 1290/2005 du Conseil, du 21 juin 2005, relatif au financement de la politique agricole commune, tel que modifié par le règlement (CE) n° 1437/2007 du Conseil, du 26 novembre, ainsi que, d'autre part, le règlement (CE) n° 259/2008 de la Commission, du 18 mars 2008, portant modalités d'application du règlement n° 1290/2005 en ce qui concerne la publication des informations relatives aux bénéficiaires de fonds en provenance du Fonds européen agricole de garantie (FEAGA) et du Fonds européen agricole pour le développement rural (FEADER).*

[3] *CJUE, 9 novembre 2010,* Volker und Markus Schecke GbR et Hartmut Eifert c./ Land Hessen*, (req. n° C-92/09 et C-93/09).*

défaillance et quelles garanties apporter aux citoyens pour les protéger de tels dommages ?

b) Une interrogation secondaire ou encore prématurée ?

Sans nier la légitimité de la question posée, d'autres intervenants se sont interrogés sur sa priorité.

M. François Bancilhon, et M. Jean-Marc Lazard, tous les deux présidents directeurs généraux de deux entreprises, respectivement *Datapublica* et *Opendatasoft*, spécialisées dans le traitement des bases de données publiques, ont estimé que le problème de la divulgation de données personnelles par les administrations était secondaire, voire inexistant dans leur pratique professionnelle. Le premier s'est en outre inquiété du risque que représenterait, pour une économie des données encore émergente, une réglementation *a priori* de l'ouverture des données publiques, privilégiant plutôt la voie d'une sanction *a posteriori* des usages illégaux.

À plusieurs reprises au cours des auditions les intervenants ont jugé que les risques liés au « *big data* », c'est-à-dire à la collecte, au traitement et à la diffusion massifs des données personnelles par les entreprises privées étaient bien supérieurs à ceux suscités par l'*open data*.

Les représentants du conseil national du numérique ont ainsi fait valoir que se focaliser sur la question de la protection des droits fondamentaux détournait du problème clé de la régulation internationale des échanges de données et de l'activité économique qui en découle.

Constatant que l'ouverture des données publiques n'avait causé aucun scandale majeur jusqu'à présent au sein des grandes administrations, M. Charles Népote, chef de projet au sein de la fondation internet nouvelle génération (FING), a quant à lui estimé que l'interrogation sur sa compatibilité avec la protection des données personnelles était peut-être prématurée et qu'il convenait d'adopter pour l'heure une démarche plus pragmatique et poursuivre le chantier engagé sans le freiner *a priori*.

Vos rapporteurs partagent le souci de pragmatisme revendiqué par ces différents intervenants. Ils reconnaissent aussi que les questions que posent les nouveaux usages des données personnelles dépassent très largement la problématique de l'o*pen data*.

Toutefois, ils observent que, sans qu'il soit besoin de classer par ordre de priorité les multiples interrogations que suscitent les bouleversements de l'économie numérique, chacune est légitime si la réflexion qu'elle engage permet d'assurer une meilleure conciliation entre le bénéfice social qui en est retiré et la protection des libertés individuelles.

En outre, ils constatent -ce que la consultation publique organisée par la CNIL sur l'*open data* permet de confirmer (*cf.* encadré)– que les administrations sont régulièrement confrontées à des interrogations sur l'ouverture de jeux de données contenant des données personnelles.

Les résultats de la consultation publique organisée par la CNIL, du 7 janvier au 7 février 2014, sur le thème de l'*open data*

Cette consultation était ouverte aux différents acteurs de l'ouverture des données publiques : services producteurs, diffuseurs, réutilisateurs ou correspondants « *Informatique et libertés* ».

391 personnes ont répondu au questionnaire, ce qui constitue un total important, même s'il n'est pas forcément représentatif de l'ensemble des acteurs concernés.

Sur la question de la protection des données personnelles, la consultation livre plusieurs enseignements :

- 55% des répondants « *responsables* open data » et « *gestionnaires de données publiques* » se sont déjà demandés si certains jeux de données dont l'ouverture était envisagée, pouvaient contenir des données personnelles (44% des réutilisateurs répondants se sont également posé cette question) ;

- 50% des gestionnaires de données publiques répondants ont indiqué avoir déjà fait part de leur opposition à l'ouverture de certaines informations détenues ou produites par leur organisme au motif d'un risque d'identification des personnes concernées par les données.

Source : CNIL

Surtout, l'examen de cette question est aujourd'hui plus que jamais opportune. En effet, l'*open data* procède d'un double mouvement, pour le passé et pour l'avenir.

Il s'agit, en effet, d'une part d'ouvrir maintenant des bases de données constituées précédemment selon des modalités qui n'anticipaient pas leur divulgation future. S'attacher à la méthode qui doit être suivie pour éviter toute atteinte accidentelle à la vie privée n'est pas illégitime.

Il s'agit, d'autre part de concevoir, pour le futur, les nouvelles bases de données d'une manière telle que leur mise à disposition du public soit facilitée. Anticiper les difficultés susceptibles de se poser favoriserait le développement de l'*open data*.

De telles considérations sont d'ailleurs largement partagées, en France comme en Europe.

2. Concilier le développement de l'*open data* et la protection des données personnelles : une préoccupation partagée en France et au niveau européen

Comme on l'a vu précédemment, l'État et les administrations publiques réfléchissent depuis longtemps à l'ouverture des données qu'ils détiennent. Or, conformément aux prescriptions du cadre législatif et réglementaire français et européen[1], le souci de la protection des données personnelles est au cœur de cette réflexion.

Plusieurs éléments en témoignent.

[1] Cf., *sur ce point, infra, I.*

Si le *vade-mecum* publié par *Etalab* sur l'ouverture et le partage des données publiques, et diffusé auprès de l'ensemble des ministères en vertu d'une circulaire du Premier ministre[1], est assez peu disert sur le sujet[2], il peut s'appuyer sur un travail antérieur, beaucoup plus approfondi, le mémento sur « *la protection des informations à caractère personnel dans le cadre de l'ouverture et du partage des données publiques* »[3], publié précédemment par le conseil d'orientation de l'édition publique et de l'information administrative (COEPIA), qui constituait, jusqu'à la création d'*Etalab*, l'une des instances de préfiguration de la mise en place de l'*open data* français.

Le rapport d'activité de la commission d'accès aux documents administratifs (CADA) pour 2011, a pour sa part consacré de longs développements à la notion de données à caractère personnel, et notamment à la question de la mise en ligne et des réutilisations des données publiques.

Récemment encore, afin de répondre à la forte demande d'un accès plus important aux données détenues par les administrations de santé, la ministre des affaires sociales et de la santé a confié le soin à MM. Pierre-Louis Bras et André Loth de conduire une réflexion sur une plus large ouverture de ces données, sous la condition d'une sécurité suffisante pour les données personnelles. Les conclusions du rapport remis à l'issue de ces travaux[4] nourrissent aujourd'hui le projet que préfigure le groupe de travail sur l'*open data* en santé, mis en place le 21 novembre dernier.

Manifestant toute l'importance que revêtait le sujet, la commission nationale de l'informatique et des libertés (CNIL) a souhaité dresser un état des lieux des pratiques et des questions posées par l'*open data* et la protection des données personnelles, afin d'y apporter des réponses concrètes et opérationnelles. Elle a organisé à cette fin, le 9 juillet 2013, un premier séminaire sur ce thème[5] et lancé, en janvier 2014 une vaste consultation en ligne.

D'autres initiatives de la société civile illustrent l'attention portée à ces questions[6].

La France ne se distingue pas, de ce point de vue, des autres pays européens qui se sont engagés sur la voie de l'ouverture des données. Ainsi, au Royaume-Uni, pays qui fait figure de précurseur en la matière,

---

[1] *Circulaire du Premier ministre n° 5677/SG en date du 17 septembre 2013, sur l'ouverture et le partage des données publiques*

[2] http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/vademecum-ouverture.pdf

[3] http://www.gouvernement.fr/sites/default/files/fichiers_joints/coepia_memento_donnees_personnelles.pdf

[4] http://www.drees.sante.gouv.fr/IMG/pdf/rapport-donnees-de-sante-2013.pdf

[5] *Séminaire* « Open data, quels enjeux pour la protection des données personnelles » *du 9 juillet 2013 dont un compte-rendu peut être consulté à l'adresse suivante :* http://www.cnil.fr/fileadmin/documents/approfondir/dossier/OpenData/CR_Workshop_Open_Data_9_juillet_2013.pdf.

[6] *Colloque interdisciplinaire du 12 novembre 2013 l'open data et les données personnelles, organisé par l'université Panthéon-Assas Paris II, le CNRS et le CERSA.*

l'*Information Commissionner's Office* (ICO), qui rassemble les compétences de la CNIL et de la CADA, a publié dès 2012, un guide des bonnes pratiques destinée à promouvoir des procédés efficaces d'anonymisation afin de concilier la protection des données personnelles et l'ouverture des données publiques.

La même réflexion est en cours en Europe, notamment au sein du groupe dit « *de l'article 29* », qui réunit les CNIL européennes et devrait prochainement publier une opinion sur les techniques d'anonymisation en faveur de l'*open data*, après avoir adopté, le 5 juin 2013, une première opinion, sur l'*open data* et la réutilisation des informations publiques[1].

Tant au vu des travaux conduits en France et en Europe que du principe même de la question posée, la légitimité de l'interrogation suivie par la mission d'information de votre commission apparaît ainsi hors de doute. Contrairement à ce que craignent ceux qui s'en défient, y répondre de manière satisfaisante pourrait d'ailleurs être le meilleur moyen de garantir la promotion rapide et efficace de l'*open data*, au bénéfice de tous.

I. UN CADRE RÉGLEMENTAIRE PROTECTEUR, QUI DEVRAIT EN PRINCIPE GARANTIR LA PROTECTION DES DONNÉES PERSONNELLES

Bien que la stratégie d'*open data* ne repose à strictement parler sur aucun texte législatif, tous les textes réglementaires et circulaires renvoient à la loi du 17 juillet 1978, dite loi « *CADA* »[2] (A).

Cependant, il apparaît à l'examen que d'autres législations viennent s'articuler avec la loi « *CADA* » pour définir le cadre juridique dans lequel intervient l'*open data*, de manière à assurer le plein respect de la protection des données à caractère personnel : la loi du 6 janvier 1978, dite loi « *Informatique et libertés* »[3], les dispositions du code du patrimoine organisant le régime d'accès aux archives publiques ainsi que celles du code pénal relatives aux atteintes aux droits de la personne résultant des fichiers ou des traitements informatiques (B).

[1] http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf

[2] *Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal*

[3] *Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*

*A. LE FONDEMENT JURIDIQUE DE L'*OPEN DATA *: LES DROITS D'ACCÈS ET DE RÉUTILISATION CONSACRÉS PAR LA LOI DU 17 JUILLET 1978*

Conçue en 1978 pour créer un droit d'accès individuel aux documents administratifs, la loi « *CADA* » a été profondément remaniée en 2005 afin d'intégrer le droit à réutilisation créé par une directive européenne de 2003 (1). Grâce à cette révision, la loi « *CADA* » peut aujourd'hui servir de fondement juridique à l'*open data* dans chacun de ses deux volets de mise à disposition et de réutilisation des informations publiques (2).

1. La loi du 17 juillet 1978, réceptacle de la transposition de la directive sur la réutilisation des informations publiques

Ignoré jusqu'alors de la législation française, le droit à la réutilisation des informations publiques y a été introduit à l'occasion de la transposition, par ordonnance, de la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public.

À l'initiative de la commission des lois du Sénat, le législateur a choisi de transposer cette directive en modifiant et complétant la loi n° 78-753 du 17 juillet 1978 précitée[1]. Comme l'indiquait le rapport de notre collègue Bernard Saugey, l'introduction dans le droit français de ce principe de réutilisation nécessitait en effet l'ajustement de certaines dispositions contenues dans la loi « *CADA* ». Ainsi de son article 10 dans sa rédaction antérieure, qui disposait que le droit à communication « *exclu[ai]t, pour ses bénéficiaires ou pour les tiers, la possibilité de reproduire, de diffuser ou d'utiliser à des fins commerciales les documents communiqués* », ce qui ne le rendait pas compatible avec le principe de réutilisation des documents, y compris à des fins commerciales, établi par la directive. Le rapport notait également que la transposition de la directive posait le problème de la réutilisation des fichiers contenant des données à caractère personnel dans la mesure où la loi « *CADA* » permettait l'accès à des informations nominatives. Il proposait alors que « *la compétence qui pourrait être confiée à la CADA pour connaître des contentieux relatifs à la réutilisation des documents [prenne] en compte les règles de protection établies par la législation relative à l'informatique, aux fichiers et aux libertés (loi du 6 janvier 1978)* »[2].

L'ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques a procédé à des ajustements de la loi « *CADA* ». Comme l'indiquait le rapport au Président de la République relatif à l'ordonnance, publié au

---

[1] Cf. *l'article 1er de la loi n° 2004-1343 du 9 décembre 2004 de simplification du droit.*

[2] *Rapport de M. Bernard SAUGEY, fait au nom de la commission des lois, sur le projet de loi habilitant le Gouvernement à simplifier le droit (n° 5, 2004-2005)* http://www.senat.fr/rap/l04-005/l04-005.html#toc130

Journal officiel du 7 juin 2005, celle-ci a créé un chapitre consacré à la réutilisation des informations publiques sans remettre en cause l'économie générale du régime d'accès aux documents administratifs. Ainsi, deux régimes distincts demeurent juxtaposés : celui de l'accès et celui de la réutilisation. Un pouvoir de sanction est confié à la Commission d'accès aux documents administratifs (CADA) par l'article 18 de la loi pour garantir en particulier le respect du principe de séparation de ces deux régimes.

2. L'articulation entre deux régimes juridiques distincts

Or, d'un point de vue juridique, l'*open data* s'analyse en deux étapes trouvant leur fondement dans chacun de ces régimes : la mise à disposition par les administrations, en vertu du régime d'accès aux documents administratifs du chapitre Ier, d'une information publique d'abord (a), son éventuelle réutilisation par les internautes conformément au chapitre II ensuite (b).

a) La publication des documents administratifs

Lors de son adoption en 1978, et bien que visant pour l'essentiel à instaurer un droit d'accès individuel aux documents administratifs, l'article 9 de la loi « *CADA* » prévoyait la « *publication régulière* » des « *directives, instructions, circulaires, notes et réponses ministérielles qui comportent une interprétation du droit positif ou une description des procédures administratives* », ainsi que « *la signalisation des documents administratifs* » afin de rendre ce nouveau droit effectif.

L'ordonnance de 2005 a déplacé cette disposition à l'article 7 de la loi et introduit la faculté pour les administrations de publier également les autres documents administratifs qu'elles produisent ou reçoivent.

L'article 7 de la loi « *CADA* » distingue ainsi clairement deux catégories de documents : ceux dont la publication est impérative et ceux dont la publication est facultative. Dans les deux cas, le respect de la confidentialité des données à caractère personnel s'impose.

La mise à disposition de données dont la publication n'est pas prévue par une disposition légale ou réglementaire dans le cadre de l'*open data* repose sur cette faculté.

Article 7 de la loi du 17 juillet 1978, dite loi « *CADA* »

« *Font l'objet d'une publication les directives, les instructions, les circulaires, ainsi que les notes et réponses ministérielles qui comportent une interprétation du droit positif ou une description des procédures administratives.*

« ***Les administrations mentionnées à l'article 1er peuvent en outre rendre publics les autres documents administratifs qu'elles produisent ou reçoivent***[1]*.*

« *Toutefois, sauf dispositions législatives contraires, les documents administratifs qui comportent des mentions entrant dans le champ d'application de l'article 6 ou, sans préjudice de l'article 13, des données à caractère personnel ne peuvent être rendus publics qu'après avoir fait l'objet d'un traitement afin d'occulter ces mentions ou de rendre impossible l'identification des personnes qui y sont nommées.*

« *Un décret en Conseil d'État pris après avis de la commission mentionnée au chapitre III précise les modalités d'application du premier alinéa du présent article.* »

b) La réutilisation des informations publiques

La principale novation de l'ordonnance de 2005 sous l'influence du droit européen demeure toutefois l'inversion de la règle d'interdiction de réutilisation à des fins commerciales qui valait jusqu'alors.

L'article 10 de la loi « *CADA* » pose en effet le principe selon lequel les informations publiques, c'est-à-dire les informations figurant dans des documents produits ou reçus par les administrations, peuvent être utilisées par toute personne à d'autres fins que celles de la mission de service public pour laquelle elle avait été produite ou reçue. Ce même article 10, ainsi que le suivant, viennent toutefois immédiatement tempérer ce principe.

Il résulte ainsi de la combinaison des articles 10 et 11 que sont exclues du champ de la réutilisation telle que prévue par le chapitre II de la loi « *CADA* » certaines informations publiques à raison soit de leur nature, soit de leur auteur, soit enfin d'un droit entrant en concurrence avec le droit à réutilisation.

Le premier motif conduit à écarter les informations considérées comme n'étant pas des informations publiques car figurant dans des documents dont la communication ne constitue pas un droit en application du régime d'accès aux documents administratifs, à moins qu'ils ne fassent l'objet d'une diffusion publique.

Cette exception renvoie en premier lieu aux documents préparatoires puisque l'article 2 de la loi « *CADA* » précise : « *le droit à communication ne s'applique qu'à des documents achevés. Il ne concerne pas les documents préparatoires à une décision administrative tant qu'elle est en cours d'élaboration* ».

[1] *Ces administrations sont : l'État, les collectivités territoriales et les autres personnes de droit public ou les personnes de droit privé chargées d'une mission de service public.*

Cette exception fait, en second lieu, référence à l'article 6 de cette même loi qui énumère, d'une part, les documents qui ne sont pas communicables en raison soit de leur nature, soit de l'atteinte que cette communication porterait par exemple à un secret, et d'autre part, les documents qui ne sont communicables qu'à l'intéressé, ces derniers visant en particulier les documents administratifs dont la communication porterait atteinte à la protection de la vie privée. Dans son avis *Maire de Chelles* du 31 juillet 2008, la CADA a précisé ce point en indiquant que n'entrent dans le champ de la réutilisation que les documents communicables à tous[1].

Article 6 de la loi du 17 juillet 1978, dite loi « *CADA* »

« I. - Ne sont pas communicables :

« 1° Les avis du Conseil d'État et des juridictions administratives, les documents de la Cour des comptes mentionnés à l'article L. 141-10 du code des juridictions financières et les documents des chambres régionales des comptes mentionnés à l'article L. 241-6 du même code, les documents élaborés ou détenus par l'Autorité de la concurrence dans le cadre de l'exercice de ses pouvoirs d'enquête, d'instruction et de décision, les documents élaborés ou détenus par la Haute Autorité pour la transparence de la vie publique dans le cadre des missions prévues à l'article 20 de la loi n° 2013-907 du 11 octobre 2013 relative à la transparence de la vie publique, les documents préalables à l'élaboration du rapport d'accréditation des établissements de santé prévu à l'article L. 6113-6 du code de la santé publique, les documents préalables à l'accréditation des personnels de santé prévue à l'article L. 1414-3-3 du code de la santé publique, les rapports d'audit des établissements de santé mentionnés à l'article 40 de la loi n° 2000-1257 du 23 décembre 2000 de financement de la sécurité sociale pour 2001 et les documents réalisés en exécution d'un contrat de prestation de services exécuté pour le compte d'une ou de plusieurs personnes déterminées ;

« 2° Les autres documents administratifs dont la consultation ou la communication porterait atteinte :

« a) Au secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif ;

« b) Au secret de la défense nationale ;

« c) A la conduite de la politique extérieure de la France ;

« d) A la sûreté de l'État, à la sécurité publique ou à la sécurité des personnes ;

« e) A la monnaie et au crédit public ;

[1] *CADA, avis n° n° 20082716 du 31 juillet 2008,* Maire de Chelles. *La CADA note d'abord que la directive 2003/98/CE du 17 novembre 2003 précitée, dont la loi «* CADA *» assure la transposition, exclut de son champ d'application non seulement les «* documents qui, conformément aux règles d'accès en vigueur dans les États membres, ne sont pas accessibles *», mais également les «* cas, dans lesquels, conformément aux règles d'accès, les citoyens ou les entreprises doivent démontrer un intérêt particulier pour obtenir l'accès aux documents *». Elle en déduit donc «* que les règles prévues au chapitre II du titre Ier de [la] loi ne s'appliquent qu'aux informations dont la communication constitue un droit pour toute personne, en application d'une disposition législative, et non à celles qui ne sont accessibles qu'à certaines personnes à raison de leur qualité ou de leur intérêt ».

« f) Au déroulement des procédures engagées devant les juridictions ou d'opérations préliminaires à de telles procédures, sauf autorisation donnée par l'autorité compétente ;

« g) A la recherche, par les services compétents, des infractions fiscales et douanières ;

« h) Ou, sous réserve de l'article L. 124-4 du code de l'environnement, aux autres secrets protégés par la loi ;

*« II. - **Ne sont communicables qu'à l'intéressé les documents administratifs** :*

*« - **dont la communication porterait atteinte à la protection de la vie privée**, au secret médical et au secret en matière commerciale et industrielle ;*

« - portant une appréciation ou un jugement de valeur sur une personne physique, nommément désignée ou facilement identifiable ;

« - faisant apparaître le comportement d'une personne, dès lors que la divulgation de ce comportement pourrait lui porter préjudice.

« Les informations à caractère médical sont communiquées à l'intéressé, selon son choix, directement ou par l'intermédiaire d'un médecin qu'il désigne à cet effet, dans le respect des dispositions de l'article L. 1111-7 du code de la santé publique.

« III. - Lorsque la demande porte sur un document comportant des mentions qui ne sont pas communicables en application du présent article mais qu'il est possible d'occulter ou de disjoindre, le document est communiqué au demandeur après occultation ou disjonction de ces mentions.

« Les documents administratifs non communicables au sens du présent chapitre deviennent consultables au terme des délais et dans les conditions fixés par les articles L. 213-1 et L. 213-2 du code du patrimoine. Avant l'expiration de ces délais et par dérogation aux dispositions du présent article, la consultation de ces documents peut être autorisée dans les conditions prévues par l'article L. 213-3 du même code. »

Sont ensuite exclues du champ du droit à réutilisation tel qu'autorisé par la loi « *CADA* » les informations produites ou reçues par deux types d'administrations : celles intervenant dans le cadre d'une mission de service public à caractère industriel ou commercial, ainsi que les établissements culturels, lesquels obéissent en la matière à leurs propres règles : organismes d'enseignement et de recherche – écoles, universités et instituts de recherche – et établissements, organismes ou services culturels – musées, bibliothèques, orchestres, opéras, ballets et théâtres[1].

Enfin, ne peuvent non plus faire l'objet d'une réutilisation les informations sur lesquelles des tiers détiennent des droits de propriété intellectuelle.

[1] *L'exclusion du champ d'application de la loi «* CADA *» conduit à ne pas appliquer à ces administrations et organismes le régime de réutilisation qu'elle institue, mais n'interdit pas la réutilisation de l'ensemble des informations produites ou reçues par ceux-ci à moins qu'ils ne s'y opposent ou l'encadrent, sous le contrôle du juge* (cf. *conseil de la CADA, n° 20062674 du 29 juin 2006,* Président du conseil général de l'Isère *et CAA Lyon, 3ème chambre, 4 juillet 2012,* Département du Cantal).

Articles 10 et 11 de la loi du 17 juillet 1978, dite loi « *CADA* »

« *Art. 10. –* ***Les informations figurant dans des documents produits ou reçus par les administrations mentionnées à l'article 1er, quel que soit le support, peuvent être utilisées par toute personne qui le souhaite à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus****. Les limites et conditions de cette réutilisation sont régies par le présent chapitre, même si ces informations ont été obtenues dans le cadre de l'exercice du droit d'accès aux documents administratifs régi par le chapitre Ier.*

« ***Ne sont pas considérées comme des informations publiques, pour l'application du présent chapitre, les informations contenues dans des documents*** *:*

« ***a) Dont la communication ne constitue pas un droit en application du chapitre Ier ou d'autres dispositions législatives, sauf si ces informations font l'objet d'une diffusion publique*** *;*

« *b) Ou produits ou reçus par les administrations mentionnées à l'article 1er dans l'exercice d'une mission de service public à caractère industriel ou commercial ;*

« *c) Ou sur lesquels des tiers détiennent des droits de propriété intellectuelle.*

« *L'échange d'informations publiques entre les autorités mentionnées à l'article 1er, aux fins de l'exercice de leur mission de service public, ne constitue pas une réutilisation au sens du présent chapitre.* »

« *Art. 11. – Par dérogation au présent chapitre, les conditions dans lesquelles les informations peuvent être réutilisées sont fixées, le cas échéant, par les administrations mentionnées aux a et b du présent article lorsqu'elles figurent dans des documents produits ou reçus par :*

« *a) Des établissements et institutions d'enseignement et de recherche ;*

« *b) Des établissements, organismes ou services culturels.*

B. LA TRIPLE GARANTIE APPORTÉE À LA PROTECTION DES DONNÉES PERSONNELLES

Comme on a pu le constater, la protection de la vie privée est l'une des préoccupations de la loi « *CADA* » et ce, depuis son origine[1]. Pourtant, à l'occasion de la modification de la loi par ordonnance en 2005, il a été jugé nécessaire de mieux prendre en compte les enjeux de protection des données à caractère personnel en tant que tels, comme la directive de 2003 y incitait le législateur, que ce soit par l'introduction dans la loi de garanties spécifiques (1) ou par renvoi aux dispositions de la loi « *Informatique et libertés* » (2). Par ailleurs, la protection des données à caractère personnel est également assurée par les dispositions du code pénal relatives aux atteintes aux droits de la personne résultant des fichiers ou des traitements informatiques (3).

[1] *L'article 6 dans sa version initiale excluait ainsi du droit d'accès les documents dont la consultation ou la communication porterait atteinte au « secret de la vie privée ».*

1. Les garanties prévues par la loi du 17 juillet 1978

Lors de la transposition de la directive 2003/98/CE du 17 novembre 2003 précitée, plusieurs dispositions ont été insérées dans la loi du 17 juillet 1978 afin de garantir la protection des données à caractère personnel, aussi bien au stade de la diffusion des informations publiques (a) qu'à celui de leur réutilisation (b).

a) Les garanties prévues dans le cadre du droit d'accès

Le troisième alinéa de l'article 7 de la loi « *CADA* » prévoit explicitement le cas où des documents administratifs contiendraient des données à caractère personnel. Dans cette hypothèse, il subordonne la publication du document à un traitement préalable afin « *d'occulter ces mentions [entrant dans le champ d'application de l'article 6] ou de rendre impossible l'identification des personnes qui y sont nommées* ».

b) Les garanties prévues dans le cadre du droit à réutilisation

L'article 13 de la loi « *CADA* » est dédié à la réutilisation d'informations publiques comportant des données à caractère personnel. Il crée en son premier alinéa un « *embryon de régime particulier* », ainsi que le désignait le rapport au Président de la République relatif à l'ordonnance n° 2005-650 du 6 juin 2005 précitée, publié au *Journal officiel* du 7 juin 2005. Ce régime spécifique a été conçu afin de couvrir les cas de réutilisations qui, n'étant constitutives ni d'un traitement automatisé ni d'un traitement portant sur les données à caractère personnel contenues ou appelées à figurer dans des fichiers, échapperait au régime prévu par la loi « *Informatique et libertés* ».

Article 13 de la loi du 17 juillet 1978, dite loi « *CADA* »

« Les informations publiques comportant des données à caractère personnel peuvent faire l'objet d'une réutilisation soit lorsque la personne intéressée y a consenti, soit si l'autorité détentrice est en mesure de les rendre anonymes ou, à défaut d'anonymisation, si une disposition législative ou réglementaire le permet.

« La réutilisation d'informations publiques comportant des données à caractère personnel est subordonnée au respect des dispositions de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. »

Ce régime soumet la réutilisation de telles informations à trois conditions alternatives.

- **Le recueil du consentement de la personne concernée**

Le fait qu'une information publique contenant des données à caractère personnel ait été diffusée en application de l'article 7 de la loi

« *CADA* », et notamment de son premier alinéa, ou d'une disposition légale n'emporte pas le consentement des personnes concernées pour la réutilisation de leurs données. Il est donc nécessaire de s'assurer de ce consentement avant toute réutilisation, notamment à des fins commerciales. Telle est l'interprétation de la CADA dans son avis du 19 avril 2012, *Directeur de la CNAMTS*[1], dans lequel elle a estimé que les dispositions du code de la santé publique prévoyant l'obligation d'inscription au tableau de l'ordre des médecins et l'affichage des tarifs des honoraires pratiqués ne pouvaient être regardées comme permettant la réutilisation, à moins de recueillir préalablement l'accord des médecins concernés.

- **L'anonymisation par l'autorité détentrice**

Cette anonymisation est à la charge de l'autorité détentrice. C'est pourquoi l'article 15 de la loi « *CADA* » prévoit la possibilité pour les administrations de percevoir des redevances pour l'établissement desquelles « *l'administration qui a produit ou reçu les documents contenant des informations publiques susceptibles d'être réutilisées tient compte des coûts de mise à disposition des informations, notamment, le cas échéant, du coût d'un traitement permettant de les rendre anonymes* ».

L'article 40 du décret d'application de la loi[2] prévoit cependant que si l'opération d'anonymisation représente un effort disproportionné, l'administration peut refuser d'y procéder, donc de communiquer le document en cause : « l*orsque la réutilisation n'est possible qu'après anonymisation des données à caractère personnel, l'autorité détentrice y procède sous réserve que cette opération n'entraîne pas des efforts disproportionnés* ».

- **L'autorisation par une disposition législative ou réglementaire spécifique**

Ce dernier cas d'autorisation ouvre potentiellement très largement le champ de la réutilisation d'informations publiques comportant des données à caractère personnel puisqu'il appartient non seulement au législateur mais également au pouvoir réglementaire de prévoir des exceptions au principe de non réutilisation de ces informations. Il convient d'observer que l'article 7 de la loi « *Informatique et libertés* » autorise quant à lui un traitement de données à caractère personnel dès lors qu'il vise à satisfaire le respect d'une obligation légale.

[1] *CADA, avis n° 20121581 du 19 avril 2012,* Directeur de la CNAMTS.

[2] *Décret n° 2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, pris pour l'application de la loi n° 78-753 du 17 juillet 1978.*

Article 7 de la loi du 6 janvier 1978, dite loi « *Informatique et libertés* »

« Un traitement de données à caractère personnel doit avoir reçu le consentement de la personne concernée ou satisfaire à l'une des conditions suivantes :

« 1° Le respect d'une obligation légale incombant au responsable du traitement ;

« 2° La sauvegarde de la vie de la personne concernée ;

« 3° L'exécution d'une mission de service public dont est investi le responsable ou le destinataire du traitement ;

« 4° L'exécution, soit d'un contrat auquel la personne concernée est partie, soit de mesures précontractuelles prises à la demande de celle-ci ;

« 5° La réalisation de l'intérêt légitime poursuivi par le responsable du traitement ou par le destinataire, sous réserve de ne pas méconnaître l'intérêt ou les droits et libertés fondamentaux de la personne concernée. »

Ce régime spécifique de l'article 13 de la loi « *CADA* » a été jugé par Mme Nathalie Mallet-Poujol, lors de son audition par vos rapporteurs, encore plus protecteur que celui garanti par la loi « *Informatique et libertés* ».

2. Le renvoi aux dispositions de la loi « *Informatique et libertés* »

En sus de ce régime spécifique, le second alinéa de l'article 13 de la loi « *CADA* » opère un renvoi vers la loi « *Informatique et libertés* ».

En effet, conformément à l'article 2 de cette même loi, dès lors qu'une réutilisation consiste en un traitement automatisé de données à caractère personnel ou en un traitement non automatisé de données à caractère personnel contenues ou appelées à figurer dans des fichiers, elle entre dans le champ d'application de la loi « *Informatique et libertés* », à moins qu'elle ne soit mise en œuvre que « *pour l'exercice d'activités exclusivement personnelles* ».

Outre les conditions de licéité prévues aux articles 6 à 10 de la loi « *Informatique et libertés* », en particulier l'obligation de traitement loyal des données et de recueil du consentement, la loi fait obligation à tout responsable de traitement de respecter des formalités préalables[1], en particulier l'obligation de déclaration auprès de la Commission nationale de l'informatique et des libertés (CNIL). Il appartient donc au réutilisateur de procéder, le cas échéant, à ces formalités.

La loi « *Informatique et libertés* » impose par ailleurs aux responsables de traitement certaines obligations[2]. Parmi celles-ci figure le respect des droits d'opposition, d'information et de rectification reconnus aux personnes physiques par les articles 38, 39 et 40. Tout réutilisateur doit donc mettre en capacité les personnes concernées d'exercer leurs droits.

[1] Cf. *chapitre IV de la loi « Informatique et libertés »*.
[2] Cf. *chapitre V de la loi « Informatique et libertés »*.

À titre incident, il convient d'observer que si la CADA distingue la diffusion de la réutilisation, considérant que « *la simple mise en ligne intégrale de documents, sans aucun commentaire ni ajout, en accès libre et gratuit ne permettant pas leur modification, ne constitue par une « réutilisation » au sens des dispositions du chapitre II* » de la loi « *CADA* »[1], cette mise en ligne constitue en revanche un traitement de données à caractère personnel, tombant sous le coup de la loi « *Informatique et libertés* ».

Enfin, la jurisprudence a permis de clarifier le régime applicable aux archives publiques.

Les articles L. 213-2 et L. 213-3 du code du patrimoine organisent le régime de communicabilité des archives publiques en appliquant des délais différenciés selon la nature de ces archives. Ils ne précisent cependant nullement le régime de réutilisation des informations publiques figurant dans ces archives.

À défaut d'interdiction ou d'encadrement de la réutilisation des informations publiques contenues dans les archives publiques, la cour administrative d'appel de Lyon a jugé que « *les informations publiques communicables de plein droit, figurant dans les documents détenus par les services d'archives publics, qui constituent des services culturels au sens des dispositions de l'article 11 de la loi du 17 juillet 1978, relèvent de la liberté de réutilisation consacrée de façon générale par cette loi, dans sa rédaction issue de l'ordonnance du 29 avril 2009* »[2]. La cour administrative d'appel a toutefois estimé qu'il appartenait « *à l'autorité compétente, saisie d'une demande de réutilisation de ces documents, de s'assurer que cette réutilisation satisfai[sai]t aux exigences qu'imposent les dispositions de l'article 13 de cette loi qui, s'agissant d'informations publiques comportant des données à caractère personnel, renvoient aux dispositions de la loi n° 78-17 du 6 janvier 1978* ».

Ainsi, bien que le chapitre II de la loi « *CADA* » ne s'applique pas aux informations publiques détenues par les services départementaux d'archives, il leur est fait obligation de s'assurer du respect des dispositions de son article 13 dès lors que des données personnelles sont en jeu, donc de s'assurer du respect par le réutilisateur des dispositions de la loi « *Informatique et libertés* ».

3. La répression administrative et pénale

La CADA n'est dotée d'aucun moyen pour assurer le respect de l'article 13 de la loi « *CADA* », l'article 18 de cette même loi ne lui confiant un

[1] Cf. *avis n° 20082716 du 31 juillet 2008* Maire de Chelles. *La CADA a en revanche précisé que* « le fait d'insérer [d]es documents accompagnés de commentaires ou sur un site invitant des tiers à émettre de tels commentaires, ou encore de subordonner leur accès au paiement d'une somme ou la publication de simples extraits constituent des formes de réutilisation au sens de l'article 10 de la loi. »

[2] *CAA Lyon, 3ème chambre, 4 juillet 2012,* Département du Cantal

pouvoir de sanction qu'en cas de réutilisation d'informations publiques en méconnaissance des dispositions de l'article 12 relatif à l'intégrité des données, ou des conditions de réutilisation prévues par une licence ou en violation de l'obligation d'obtention d'une licence.

La CNIL, en revanche, dès lors qu'elle est compétente, dispose d'un pouvoir de sanction en cas de réutilisation contraire à une disposition de la loi « *Informatique et libertés* », conformément au chapitre VII de cette dernière. Les décisions prononçant une sanction peuvent faire l'objet d'un recours de pleine juridiction devant le Conseil d'État, en vertu du dernier alinéa de l'article 46 de cette même loi.

La loi « *Informatique et libertés* » renvoie enfin, pour les sanctions pénales, aux dispositions des articles 226-16 à 226-24 du code pénal, relatifs aux atteintes aux droits de la personne résultant des fichiers ou des traitements informatiques. Dans le cadre de l'*open data*, le réutilisateur qui n'aurait pas pris les précautions nécessaires serait en particulier passible de la peine de trois ans d'emprisonnement et de 100 000 euros d'amende prévue au deuxième alinéa de l'article 226-22 en cas de divulgation, commise par imprudence ou négligence, de données à caractère personnel ayant pour effet de porter atteinte à la considération de l'intéressé ou à l'intimité de sa vie privée. Les personnes morales encourent, quant à elles, outre une amende égale au quintuple de celle prévue pour les personnes physiques, certaines peines prévues à l'article 131-9 du code pénal, en application de l'article 226-24 du même code.

Article 226-22 du code pénal

« Le fait, par toute personne qui a recueilli, à l'occasion de leur enregistrement, de leur classement, de leur transmission ou d'une autre forme de traitement, des données à caractère personnel dont la divulgation aurait pour effet de porter atteinte à la considération de l'intéressé ou à l'intimité de sa vie privée, de porter, sans autorisation de l'intéressé, ces données à la connaissance d'un tiers qui n'a pas qualité pour les recevoir est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende.

« La divulgation prévue à l'alinéa précédent est punie de trois ans d'emprisonnement et de 100 000 euros d'amende lorsqu'elle a été commise par imprudence ou négligence.

« Dans les cas prévus aux deux alinéas précédents, la poursuite ne peut être exercée que sur plainte de la victime, de son représentant légal ou de ses ayants droit. »

II. UNE PROTECTION TOUTEFOIS FRAGILISÉE PAR UNE DOUBLE FAILLE

Notre droit parvient à une conciliation équilibrée entre les exigences qui fondent l'*open data* et la protection des données personnelles, donnant à cette dernière la priorité à moins que l'innocuité de la diffusion d'informations privées soit garantie, parce que la base a été anonymisée, que l'intéressé a donné son consentement ou que la loi exige cette diffusion.

Cependant, l'efficacité de cette réglementation dépend de la façon dont les administrations la mettent en œuvre. Or, des auditions et des travaux qu'ils ont conduits, vos rapporteurs tirent le constat que cette mise en œuvre est susceptible de présenter deux failles.

La première a trait à la qualité du procédé d'anonymisation mis en œuvre : tous ne présentent pas la même efficience et, selon la nature des données en cause, le risque de ré-identification n'est pas nul, ce qui pourrait aboutir à la diffusion indirecte et accidentelle d'informations personnelles (A).

La seconde faille concerne le pilotage et l'accompagnement des administrations dans le déploiement de l'*open data*. Ceux-ci n'apparaissent pas toujours suffisants, laissant ces administrations démunies face à des exigences de protection des données personnelles qu'elles maîtrisent mal (B).

A. PREMIÈRE FAILLE : LE RISQUE DE RÉ-IDENTIFICATION

Des trois possibilités de mise en ligne de documents administratifs contenant des données personnelles[1], l'anonymisation de la base de données est certainement le procédé le plus commode.

En effet, son champ d'application est général : contrairement aux données personnelles diffusées sur la base d'une exigence légale spéciale, il peut concerner toutes les bases de données administratives librement communicables. En outre, il permet de se dispenser du recueil préalable du consentement des intéressés, ce qui représenterait, pour la plupart des bases de données envisageables, initialement conçues pour un tout autre objectif, une contrainte insurmontable.

Pour autant, il n'est pas infaillible : imparfaitement mis en œuvre, il est susceptible de permettre la ré-identification des données, au détriment des administrés concernés.

[1] *Consentement de l'intéressé à la diffusion de ses données, anonymisation ou exigence légale de diffusion.*

1. Un risque avéré

a) Les techniques d'anonymisation

L'anonymisation peut être définie comme l'opération de suppression de l'ensemble des informations permettant d'identifier directement ou indirectement un individu, contenues dans un document ou une base de données.

Selon les informations fournies par la CNIL à vos rapporteurs, trois méthodes peuvent être employées pour parvenir à ce résultat : la substitution parfois appelée aussi « *pseudonymisation* », la suppression ou le « *masquage* », et, enfin, l'agrégation.

- ***La substitution ou la* « pseudonymisation »**

Cette méthode consiste à remplacer l'identifiant initial d'une personne par un autre identifiant arbitraire, un pseudonyme[1].

Pour garantir la traçabilité et la mise à jour des informations dans la base et éviter d'associer à un individu les données relatives à un autre, faute de disposer d'un identifiant pérenne, il est nécessaire que, pour chaque personne, ce pseudonyme soit unique.

À cette fin, il peut être établi de trois manières différentes.

Une table de correspondance secrète peut être générée, qui associe une fois pour toutes, l'ensemble des identifiants avec les pseudonymes qui leur ont été attribués. Le niveau de sécurité de cette technique d'anonymisation est faible. L'opération est en effet réversible, puisqu'on peut retrouver l'identifiant à partir du pseudonyme et que celui qui détient la table lit à livre ouvert dans la base donnée : l'anonymisation n'est garantie qu'autant que cette table reste secrète.

La seconde façon de procéder à la « *pseudonymisation* » est d'utiliser un algorithme de chiffrement : l'ensemble des identifiants initiaux sont transformés en des pseudonymes uniques. L'opération est, là encore, réversible, puisque l'on peut retrouver l'identifiant à partir du pseudonyme, pour peu que l'on sache quel algorithme de chiffrement a été utilisé.

La dernière façon de procéder à la substitution d'un pseudonyme à l'identifiant initial est de recourir à une fonction dite de « *hachage* », qui présente la particularité, par rapport aux algorithmes de chiffrement standards, de ne pas être réversible : il n'est pas possible de retrouver l'identifiant initial à partir du seul pseudonyme, même si l'on connaît la fonction de hachage utilisée.

[1] *Ce pseudonyme peut parfois être un nom de code, le risque étant alors qu'il y ait une correspondance implicite entre l'identifiant initial et ce nom de code. Le plus souvent, il s'agit d'une suite arbitraire de caractères alphanumériques.*

Toutefois, en dépit de cette irréversibilité de principe, cette technique peut être mise en échec en reconstituant, par réitération, une table de correspondance. Cette méthode pour casser l'anonymisation suppose d'importants moyens informatiques : elle consiste à appliquer la fonction de hachage à l'ensemble des identifiants possibles (par exemple, l'ensemble des noms et prénoms des individus susceptibles d'appartenir à la base de données). Ainsi, on retrouve, pour chacun, le pseudonyme unique qui lui est attribué par la fonction de hachage initialement utilisées.

Il est possible de renforcer la sécurité de l'anonymisation en ajoutant préalablement aux identifiants initiaux une clé secrète arbitraire : par exemple au nom « *Jean Dupont* », on associe la clé « *azerty* », pour donner un second identifiant « *Jean Dupontazerty* », qu'on soumet alors à la fonction de hachage. Celui qui souhaitera reconstituer la table de correspondance devra donc non plus seulement tester l'ensemble des noms et prénoms possibles, ce qui est relativement facile, mais aussi l'ensemble des modifications que ces identifiants sont susceptibles de connaître à partir de clés inconnues.

La sécurité du dispositif repose cependant encore une fois sur la confidentialité des outils utilisés : la clé secrète d'une part, la fonction de hachage utilisée d'autre part.

Il est encore possible de durcir l'anonymisation, en procédant à un double hachage avec clé secrète, qui consiste à réaliser une première fois l'opération, et à soumettre le pseudonyme obtenu à une seconde fonction de hachage avec clé secrète. Pour assurer une pleine confidentialité, les clés peuvent être renouvelées régulièrement. Toutefois, dans ce cas, il n'est plus possible de suivre dans le temps l'évolution des données relatives à un individu, puisqu'il n'y aura plus de moyen de mettre en relation son pseudonyme à un moment donné, avec un second pseudonyme généré plus tard.

- ***La suppression, le* « masquage » *ou l'ajout de bruit***

L'opération consiste à dégrader l'information initiale, en supprimant certaines données, ou, au contraire, en ajoutant des informations qui noient la donnée initiale identifiante.

Ainsi, dans le premier cas, plutôt que de retenir la date de naissance complète, seule sera conservée l'année de naissance.

Dans le second, les informations seront déformées selon un procédé qui n'en altèrera pas l'usage premier. Ainsi, une base de donnée rassemblant les salaires des employés d'une société, par classe d'âge, destinée à étudier le salaire moyen, sera publiée en ajoutant et en retranchant 1000 euros dans une proportion identique aux salaires d'une même classe. Le calcul du salaire moyen n'en sera pas affecté, puisque les additions et les soustractions se compenseront. Dans le même temps, aucun des salaires publiés ne correspond à celui d'un des employés. L'inconvénient d'une telle méthode

est de fausser, dès l'origine, les données, ce qui les rend moins pertinentes pour une utilisation autre que celle d'origine.

- ***L'agrégation***

Cette dernière méthode consiste à rassembler plusieurs données de même type, afin de produire une donnée agrégée qui conserve l'information dont on a besoin, mais rend impossible l'identification de la part d'information agrégée qui correspond à un individu donné.

Par exemple, plutôt que de retenir les deux informations selon lesquelles M. Dupont s'est présenté au guichet de telle administration à 10h et M. Durand à 11h, on retiendrait l'information agrégée selon laquelle deux personnes se sont présentées dans la matinée au guichet concerné, ce qui permet de connaître, avec une certaine précision, le taux de fréquentation de cette administration, sans poser de problème d'identification des administrés en cause.

Ce procédé est d'usage courant en matière statistique. Plus le niveau d'agrégation est élevé, moins il y a de risques de ré-identification, mais, dans le même temps, moins l'information est précise. L'autorité en charge de la base de données doit donc trouver un équilibre satisfaisant entre la robustesse de l'anonymisation garantie par le niveau d'agrégation et la précision des données qu'elle autorise. D'une manière générale, la CNIL recommande de ne pas faire figurer de statistiques conçues à partir d'une agrégation inférieure à dix éléments.

b) Des techniques qui ne sont pas infaillibles

Comme l'a indiqué le représentant de l'institut national de la recherche en informatique et en automatique (INRIA), M. Claude Kirchner, lors de son audition, aucune technique d'anonymisation n'est en théorie infaillible.

Ainsi, la sécurité offerte par les procédés de pseudonymisation dépend, comme on l'a vu précédemment, de la confidentialité des outils de codage utilisés (algorithme, clé secrète, table de correspondance). Il peut toutefois être facilement remédié à ce défaut, en élevant le niveau de confidentialité de ces instruments.

En réalité, la principale faille de l'ensemble de ces procédés d'anonymisation tient aux données mêmes auxquelles ils sont appliqués.

Les liens établis entre elles, qui constituent la raison d'être des bases informatiques qui les rassemblent sont parfois aussi identifiants que chacune de ces données prise isolément.

Prenons l'exemple d'une base de données rassemblant le nom, la date de naissance, le lieu de naissance et le niveau d'imposition. Elle pourrait être anonymisée en ne retenant que l'initiale du nom et l'année de naissance (anonymisation par suppression ou masquage). Cependant si l'on croise ces

informations dégradées avec le lieu de naissance, il est possible de ré-identifier certains individus, en particulier, ceux nés dans des communes comptant suffisamment peu de naissances par année, pour que l'initiale du nom de famille permette d'identifier l'intéressé à coup sûr et de connaître, par ricochet, le montant de son imposition[1].

Une illustration d'une telle ré-identification est fournie par l'erreur commise par le fournisseur d'accès à internet américain, AOL, en 2006 (*cf.* encadré).

Une première faille de confidentialité dans une démarche d'*open data* : l'exemple d'AOL

En 2006, l'entreprise américaine AOL, fournisseur d'accès internet, a publié en ligne une vaste base de données qui rassemblait 20 millions de recherches effectuées sur son site par 650 000 utilisateurs. L'objectif de l'opérateur était à la fois de montrer l'étendue des services qu'il proposait, et de proposer à l'attention du public et de la recherche, une ressource particulièrement riche.

La base avait été anonymisée selon un procédé de pseudonymisation : chaque identifiant (nom d'utilisateur sous AOL, adresse IP…) avait été remplacé par un nombre choisi aléatoirement. Ainsi les chercheurs conservaient la possibilité d'attribuer à une même personne l'ensemble des recherches qu'elle avait effectuées, sans que son identité leur soit connue.

L'opérateur a cependant négligé que l'historique de recherche d'un individu (c'est-à-dire l'ensemble des recherches internet qu'il a effectuées pendant une période donnée) est très identifiant : ainsi certains internautes vérifient-ils à intervalle régulier ce qui est publié sur leur compte, en effectuant une recherche sur leur nom. De la même manière, beaucoup de recherches portent sur des services offerts à proximité, ce qui permet, par recoupement, d'approcher l'adresse possible de l'intéressé. Les choix de recherche fournissent aussi des informations sur l'âge, la profession, les goûts ou préférences d'une personne : tous éléments indirectement identifiants.

Le résultat de cette opération est que certains internautes furent réellement identifiés[2].

D'une manière générale, il suffit de peu de données –et des données anodines en apparence– pour que l'empreinte laissée par celles-ci permette d'identifier une personne parmi d'autres, ce qui permet de la retrouver dans la base en dépit de son anonymisation et donc d'avoir accès, ensuite, à tous son dossier. Ainsi le rapport de Pierre-Louis Bras et André Loth sur la gouvernance et l'utilisation des données de santé, rappelle que 89 % des

[1] *À titre d'exemple, une étude américaine a montré que 97 % des électeurs de la ville de Cambridge dans le Massachusetts pouvaient être identifiés par le seul croisement de leur date de naissance et des neuf chiffres du code postal correspondant à leur adresse (Latanya Sweeney (1997),* « Weaving technology and policy together to maintain confidentiality », Journal of Law, Medicine and Ethics, *25, p. 98-110, cité par Kieron O'Hara (2011),* Transparent Government, not Transparent Citizens : A Report on Privacy and Transparency for the Cabinet Office, *disponible à l'adresse suivante :* *https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/61279/transparency-and-privacy-review-annex-a.pdf)*

[2] *Kieron O'Hara (2011),* Transparent Government, not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office, *préc., p. 39.*

patients ayant connu un séjour à l'hôpital en 2008 sont identifiables si l'on connaît les informations suivantes, relativement aisée à retrouver : l'hôpital d'accueil, le code postal du domicile, le mois et l'année de naissance, le sexe, le mois de sortie et la durée du séjour[1]. Ce chiffre atteint 100 % pour les patients hospitalisés deux fois la même année[2]. La robustesse des opérations d'anonymisation auxquelles sont soumis le système national d'information inter-régimes de l'assurance maladie (SNIIRAM) et le programme de médicalisation des systèmes d'information (PMSI), qui rassemblent les données de santé de nos concitoyens, n'y change rien : la richesse des informations contenues dans ces bases rend le risque de ré-identification très important, ce qui justifie les mesures de restriction d'accès aujourd'hui mises en œuvre.

La ré-identification par croisement d'informations qui subsistent dans la base, après anonymisation, se trouve d'ailleurs grandement facilitée par le recours à d'autres jeux de données publiés, qui peuvent rendre identifiant des liens entre plusieurs données qui jusqu'alors ne paraissaient pas permettre de caractériser une personne. Ce procédé a permis la ré-identification de nombre des clients de l'entreprise américaine de diffusion de DVD en ligne, Netflix, à partir de la base de données pourtant anonymisée qu'elle avait diffusée en ligne (*cf.* encadré ci-dessous).

Une seconde faille de confidentialité dans une démarche d'*open data* : l'exemple de Netflix

L'entreprise américaine Netflix offre un service de location en ligne de DVD et permet ensuite à ses utilisateurs de noter ou de recommander les films qu'ils ont visionnés. Ces appréciations lui donnent la possibilité de mieux cerner les goûts de ses clients et de leurs proposer, ce faisant, des films susceptibles de leur plaire.

Souhaitant affiner ses programmes d'analyse des préférences de ses utilisateurs, l'entreprise a publié en ligne, les recommandations de 500 000 d'entre eux, afin que des programmeurs indépendants développent des applications plus performantes que le logiciel utilisé par l'entreprise pour proposer à ses clients des films conformes à leurs goûts. Un prix d'un million de dollars était en jeu.

L'entreprise avait pris le soin d'anonymiser les données directement identifiantes et de modifier légèrement les autres. Deux informaticiens, Arvind Narayanan et Vitaly Shmatikov, sont toutefois parvenus à percer cette anonymisation et à ré-identifier plusieurs profils d'utilisateurs.

[1] *Ainsi, il suffirait, pour retrouver dans la base le dossier d'une personnalité, de savoir à quelles dates elle a été hospitalisée et où (les autres informations –domicile, âge et sexe, étant très facile à connaître), pour identifier le pseudonyme qui lui a été attribué et avoir potentiellement accès à l'ensemble de ses informations médicales.*

[2] *Pierre-Louis Bras, André Loth,* Rapport sur la gouvernance et l'utilisation des données de santé, *remis à la ministre des affaires sociales et de la santé, septembre 2013, p. 27. Le rapport est consultable à l'adresse suivante : http://www.drees.sante.gouv.fr/rapport-sur-la-gouvernance-et-l-utilisation-des-donnees-de,11202.html.*

En effet, ils se sont aperçus que la seule information donnée par le croisement entre l'appréciation portée sur trois films, et la date à laquelle ils ont été loués, était suffisante pour retrouver l'auteur de ces appréciations, s'il avait fait état d'appréciations identiques aux mêmes dates, sur un autre site ou dans un forum de discussion dans lequel il apparaissait sous sa véritable identité. L'identification pouvait même être accomplie avec moins d'informations si le film en question était relativement rare[1].

Même la technique dite de l'agrégation ne garantit pas contre cette ré-identification, si le degré d'agrégation choisi est mal calibré.

Comme M. Michel Isnard l'a indiqué à vos rapporteurs, l'INSEE, dont le travail est pourtant souvent exemplaire du point de vue de la protection des données personnelles, a été, au début de l'année 2013, à l'origine d'une fuite sur l'imposition de certains contribuables. En effet, recourant à la technique du « *carroyage* », l'institut a divisé la France en carrés de 200 mètres de côté, en associant à ces carreaux l'imposition moyenne des habitants concernés. Ces données furent publiées sur internet, dans le cadre d'une démarche d'*open data*. Il s'est cependant avéré que certains carrés, situés dans des territoires peu peuplés, ne comptaient qu'un seul foyer fiscal, dont il était aisé, ensuite, de retrouver l'adresse et donc l'identité. L'institut a depuis revu sa méthodologie[2].

2. Des risques jusqu'à présent limités, mais des conséquences susceptibles d'être graves pour les personnes concernées comme pour l'administration

Dans un contexte de diffusion exponentielle de données de tous ordres, publiques ou privées, sur tous les objets possibles, ce qu'on désigne parfois du terme de « *big data* », les techniques de recoupement d'informations constituent un moyen très efficace pour percer l'anonymat des bases de données.

Ce risque de ré-identification n'est pas nul. **Vos rapporteurs constatent cependant, qu'à ce jour, les dispositifs d'anonymisation utilisés ont été suffisamment robustes pour éviter toute diffusion accidentelle de données personnelles**. En dehors des quelques cas précités, les personnes entendues au cours des auditions n'ont pas fait état d'autres exemples de ré-identification avérée de données publiées.

Force est de reconnaître, toutefois, que l'immense majorité des jeux de données mis en ligne porte sur des informations impersonnelles, pour lesquelles le risque de ré-identification ne se pose pas. Nous sommes cependant au commencement de l'*open data* : le mouvement doit prendre de

[1] *Arvind Narayanan et Vitaly Shmatikov (2008),* « Robust de-anonymisation of large sparse datasets », Proceedings of the 2008 IEEE Symposium on Security and Privacy, *p. 11-125, cité par Kieron O'Hara, préc., p. 40-42.*

[2] Cf. http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees&page=donnees-detaillees/donnees-carroyees/donnees_carroyees_diffusion.htm.

l'ampleur et le nombre de bases anonymisées devrait progresser sensiblement, augmentant d'autant les situations dans lesquelles une fuite d'information personnelle pourrait survenir.

En outre, la probabilité d'une telle fuite doit être croisée avec sa gravité, pour décider s'il convient ou non, comme le pensent vos rapporteurs, de mettre en œuvre de manière plus systématique les moyens de la prévenir.

Or, la gravité de l'atteinte portée aux intéressés dépend de deux éléments : d'une part de la nature des informations en cause, d'autre part de la portée de la diffusion accidentelle.

Incontestablement certaines données sont plus sensibles que d'autres, comme celles relatives à l'état de santé, à la vie familiale, aux affaires judiciaires ou au patrimoine. Les fichiers concernés (état civil, base de données de santé, fichier fiscal, casier judiciaire ou base de décisions de justice…) font d'ailleurs l'objet de précautions particulières.

La gravité de l'atteinte portée à la vie privée des victimes est aussi fonction de l'étendue de la fuite, qui dépend elle-même du mode de publication retenue par l'administration : dans son principe, l'*open data* commande une publication qui autorise le téléchargement de la base de données et son appropriation par tous ceux qui le souhaitent. Or, à moins de ne s'être jamais dessaisie des données et d'en avoir toujours gardé le contrôle, l'administration ne sera pas en mesure de contenir la fuite et de rapatrier les bases compromises : la fuite sera irrémédiable.

À la gravité des conséquences d'une ré-identification pour les victimes répond celle des conséquences pour l'administration elle-même.

En principe, la ré-identification procédant d'un recoupement d'informations, et donc, d'un traitement de la base de données, la responsabilité de l'atteinte ainsi portée à la vie privée d'un administré devrait incomber à celui qui y a procédé volontairement.

Toutefois, on ne peut exclure, comme l'a relevé M. le professeur Gilles Gugliemi lors de son audition par vos rapporteurs, que la responsabilité de l'administration soit engagée à raison des négligences graves qu'elle aurait commises et qui auraient favorisé la diffusion d'informations personnelles, comme elle peut l'être à chaque fois qu'elle divulgue elle-même une information couverte par le secret administratif ou non communicable parce qu'elle porte atteinte à la vie privée d'un administré ou d'un fonctionnaire[1].

[1] *Sur ce point* cf. *en particulier, Maryse Deguergue,* « Promesses, renseignements, retards », in Répertoire de la responsabilité de la puissance publique, *Dalloz.*

B. SECONDE FAILLE : DES ADMINISTRATIONS PARFOIS DÉMUNIES FACE AU DÉFI DE L'OUVERTURE DES BASES DE DONNÉES

Même si, à travers *Etalab* et le secrétariat général à la modernisation de l'action publique (SGMAP), les gouvernements récents ont donné une forte impulsion en faveur de l'ouverture des données publiques, la conduite decette politique d'*open data* relève de chaque administration, pour ce qui concerne ces propres données.

Or, faute d'un pilotage ou d'un accompagnement suffisant, beaucoup se trouvent démunies face au défi que cela représente pour elles, s'agissant de la protection des données personnelles des administrés.

1. La nécessité, pour les administrations, de s'adapter à la nouvelle donne de l'*open data*

Pour les administrations, l'ouverture de bases de données établies à partir d'informations personnelles représente un double changement.

Il s'agit d'abord d'un changement d'usage, comme l'a souligné le président du conseil d'orientation de l'édition publique et de l'information administrative (COEPIA), M. Michel Pinault, lors de son audition. Une base de données conçue à l'origine pour une fin strictement gestionnaire doit être adaptée ou reconvertie, afin de pouvoir la publier en ligne. Or, la façon dont cette base a été conçue à l'origine est susceptible de rendre cette reconversion particulièrement difficile.

Ce premier changement se double d'un changement d'approche. Jusqu'à présent, sauf exception[1], la façon dont l'administration préservait le secret sur les informations personnelles contenues dans ses bases de données était d'en contrôler la diffusion. Cette diffusion, dans le respect de la vie privée des administrés, devenant la règle avec l'*open data*, l'administration doit s'efforcer de garantir la correcte anonymisation de la base de données, ce qui suppose d'une part de développer une compétence technique différente, et, d'autre part, de mobiliser des moyens humains ou financiers supplémentaires.

L'assistance qu'elles peuvent recevoir pour faire face à ce double changement est donc cruciale. Or, vos rapporteurs ne peuvent que dresser le constat, à cet égard, d'un défaut de pilotage et d'accompagnement.

[1] *Certaines bases de données particulièrement sensibles, comme le SNIIRAM, sont anonymisées dès l'origine, indépendamment de toute possibilité de diffusion. Ainsi le numéro de sécurité sociale de bénéficiaires est systématiquement codé.*

2. Un défaut de pilotage et d'accompagnement pour garantir la protection des données personnelles

S'attachant uniquement à la question de la protection de la vie privée, la mission d'information n'a envisagé le pilotage mis en place par le Gouvernement en matière d'*open data*, que sous cet angle. Or, force est de constater qu'en dépit de quelques initiatives bienvenues, celui-ci fait défaut et laisse certaines administrations plus démunies que d'autres face à l'ampleur de la tâche qui leur incombe.

a) Etalab *: un rôle d'impulsion plus que de direction*

Structure légère, comptant, comme on l'a vu, à peine une dizaine de collaborateurs, *Etalab* joue un rôle d'impulsion, comme son directeur, M. Henri Verdier, l'a précisé à vos rapporteurs lors de son audition : elle laisse les ministères libres de décider de la stratégie d'ouverture des données qu'ils souhaitent mettre en place.

Etalab échange avec les administrations, en particulier grâce à un réseau de correspondants appartenant aux secrétariats généraux des principaux ministères, mais elle ne paraît pas en mesure de leur assurer une assistance technique, notamment en matière d'anonymisation, au-delà de quelques recommandations générales.

Il n'entre pas non plus dans ses attributions de surveiller les jeux de données mis en ligne et de contrôler qu'ils ne présentent pas de risque vis-à-vis de la vie privée. D'ailleurs *Etalab* ne procède pas à un contrôle *a priori* des jeux de données qui lui sont transmis pour publication sur son site internet *data.gouv.fr*, limitant sa fonction à celle d'un hébergeur plutôt qu'un éditeur de contenus.

Ni instance de contrôle, ni instance de pilotage, *Etalab* se cantonne donc à un rôle d'animation, laissant aux administrations la charge de conduire elles-mêmes l'ouverture de leurs données.

b) Des administrations qui s'organisent empiriquement, faute d'accompagnement suffisant

Quelques initiatives ont été lancées pour aider les administrations à s'assurer que les données qu'elles envisageaient de publier ne posaient pas de problème vis-à-vis de la protection de la vie privée des administrés. Elles demeurent cependant isolées ou limitées.

Ponctuellement, la commission nationale de l'informatique et des libertés a publié des recommandations relatives à l'anonymisation de certains fichiers. Ainsi, dès le début des années 2000, elle s'est penchée sur la question de l'anonymisation des décisions de justice, qui étaient publiées sans que le nom des parties soit occulté, à l'exception des décisions rendues en matière pénale ou familiale. La recommandation adoptée le 23 novembre

2001, qui détaille les éléments à anonymiser, et préconise de bloquer l'indexation par les moteurs de recherche, fait aujourd'hui encore référence auprès des juridictions, des entreprises ou des sites publics qui éditent les décisions de justice.

Plus récemment, le COEPIA a, comme on l'a vu précédemment, rédigé un mémento spécialement dédié à cette question en juillet 2013. Ce document –qui n'est étonnamment pas mentionné sur le site d'*Etalab*– s'attache à clarifier le droit applicable en la matière et les obligations qui incombent aux administrations, renvoyant notamment à deux guides pratiques édités par la CNIL les questions plus techniques relatives à l'anonymisation. Il s'agit là d'un outil pertinent qui gagnerait à faire l'objet d'une diffusion plus systématique, notamment auprès des administrations locales.

À l'initiative de plusieurs collectivités territoriales, qui s'étaient engagés dans un projet d'*open data*, une association, *Open Data France*, a été créée pour favoriser les échanges d'expérience et la diffusion des bonnes pratiques. Lors de leur audition, les représentantes de cette association, Mmes Erwane Monthubert et Sandrine Mathon, ont toutefois indiqué que si un groupe de travail sur les enjeux relatifs aux données personnelles avait été constitué en son sein, cette réflexion ne faisait que commencer, ces données ne constituant selon elles qu'une minorité des données détenues par les collectivités territoriales.

Utiles aux administrations, de telles initiatives demeurent cependant circonscrites et ne lèvent pas toutes les difficultés que ces administrations peuvent rencontrer.

Vos rapporteurs constatent à cet égard un manque de cohérence dans les approches des différents ministères ou collectivités territoriales.

Certains paraissent mieux armés que d'autres pour faire face aux enjeux de l'anonymisation. Ainsi, comme l'a souligné M. Michel Isnard, chef de l'unité « *affaires juridiques et contentieuses* » de l'institut national de la statistique et des études économiques (INSEE), les ministères dotés d'un service statistique sont plus familiers des techniques employées par l'INSEE pour garantir le respect du secret statistique et permettre la diffusion de vastes jeux de données.

De la même manière, certaines administrations, comme celles des finances ou des affaires sociales, habituées à protéger le secret fiscal ou le secret médical, ont une attention plus vigilante à la protection des données personnelles, ainsi qu'une compétence particulière en la matière.

Les lacunes du réseau des correspondants informatiques et libertés (CIL)[1] ou des personnes responsables de l'accès aux documents administratifs (PRADA)[2], constituent une autre source d'inégalité entre les administrations, puisque les premiers comme les seconds pourraient utilement les accompagner dans leurs démarches d'*open data*.

Un exemple présenté à la fois par M. Simon Chignard et les représentantes d'*Open Data France*, Mmes Erwane Monthubert et Sandrine Mathon, illustre bien **le défaut d'accompagnement auquel sont confrontées les collectivités et les administrations, et l'incertitude qui s'ensuit pour la protection des données personnelles**.

De nombreuses communes publient sur leur site internet la liste, par année, des prénoms des enfants nés sur leur territoire. Cette liste est très incomplète, puisque n'y figurent, par précaution, que les prénoms attribués plus de cinq fois dans l'année. Toutes les communes respectent cette même règle, estimant qu'il s'agit d'une prescription de l'INSEE. Or, il n'en est rien, même si l'institut conseille en cas de risque d'identification de ne diffuser que les données qui présentent un nombre suffisant d'occurrences.

La règle suivie paraît donc arbitraire. En outre, sa légitimité, en l'espèce, pourrait être discutée, puisque l'identification précise des intéressés supposerait qu'on possède déjà l'information sur leur nom complet, leur année et leur commune de naissance, ce qui reviendrait à disposer de plus d'informations qu'on ne pourrait en apprendre par cette liste.

III. POURSUIVRE LE DÉVELOPPEMENT DE L'*OPEN DATA*, EN L'ASSORTISSANT DE GARANTIES PLUS SOLIDES POUR LA PROTECTION DES DONNÉES PERSONNELLES

Des travaux et des auditions qu'ils ont conduits, vos rapporteurs retiennent le constat d'une situation paradoxale.

Le cadre juridique de l'*open data* est relativement protecteur. Les premières données diffusées à ce stade ne paraissent pas, à de rares exceptions, présenter de danger pour la vie privée de nos concitoyens.

Pourtant, dans le même temps, le risque d'une ré-identification des données publiées existe et se trouve aggravé par la profusion de jeux de données mis en ligne par l'administration comme par les personnes privées elles-mêmes. La façon dont l'État et les collectivités territoriales conduisent l'ouverture de leurs données devrait pouvoir dissiper les inquiétudes. Il n'en

[1] *Au cours de son audition, M. Paul-Olivier Gibert, président de l'association française des correspondants à la protection des données à caractère personnel (AFCDP) a regretté le retard des administrations françaises en matière de désignation de CIL.*

[2] *Le rapport d'activité de la CADA pour 2012 recense 1 598 PRADA en 2013. S'il se félicite de compter un correspondant dans tous les grands ministères et les agglomérations les plus importantes, il juge en revanche la situation insatisfaisante dans les moyennes agglomérations.*

est rien : le sentiment prédomine d'un défaut de pilotage ou d'accompagnement qui laisse parfois les administrations démunies face à une tâche nouvelle qu'elles ne maîtrisent pas toujours.

Ces failles ne remettent pas en cause la pertinence de l'ouverture des données publiques, mais la façon dont elle est conduite. Vos rapporteurs sont convaincus que les réponses qui y seront apportées devraient permettre de franchir une nouvelle étape dans le déploiement de l'*open data* (A), ce qui suppose de concevoir, en la matière, une doctrine de protection des données personnelles (B), ainsi qu'une nouvelle gouvernance (C).

A. ACCÉLÉRER LE DÉPLOIEMENT D'UN OPEN DATA *RESPECTUEUX DE LA PROTECTION DES DONNÉES PERSONNELLES*

À plusieurs reprises au cours des auditions, des personnes entendues ont exprimé leur crainte que la réflexion engagée par la mission d'information ne soit qu'un prétexte pour freiner le mouvement d'ouverture des données publiques.

Vos rapporteurs y voient au contraire une opportunité de pousser plus loin ce mouvement.

En effet, ils observent, en premier lieu, que si le droit français connaît, comme on l'a vu précédemment, trois concepts qui s'approchent de l'*open data*, la communication des documents administratifs, la diffusion publique et la réutilisation, il ne prévoit pas, à proprement parler, d'obligation pour les administrations de mettre en ligne une base de données.

Une seule obligation pèse sur l'État et les collectivités territoriales : remettre à celui qui le demande le document qu'il souhaite, s'il appartient à ceux qui peuvent lui être communiqués. Cette obligation s'accompagne d'une autorisation faite à tout un chacun de réutiliser, sous quelques exceptions, les informations publiques correspondant aux documents communicables ou celles qui ont fait l'objet d'une diffusion publique.

Dans ce cadre, les informations personnelles apparaissent comme une limite, qui interdit, sauf consentement de l'intéressé, fondement légal ou anonymisation, la communication ou la réutilisation.

Vos rapporteurs constatent, en second lieu, que plusieurs des personnes qu'ils ont entendues, ont défendu l'intérêt qui s'attacherait à la mise en ligne plus systématique par l'administration des documents ou des jeux de données qui lui sont demandés.

M. Serge Daël, président de la CADA, a ainsi souligné que cette mise en ligne réglait par définition la question de la communication individuelle du document en cause.

Les représentants des associations de réutilisateurs des données publiques, comme ceux des entreprises privées ont partagé cette analyse, saluant l'avancée que constitue la création d'*Etalab*, mais appelant à poursuivre encore ce mouvement.

M. Jean-Baptiste Soufron, secrétaire général du conseil national du numérique a jugé souhaitable de consacrer une obligation, pour les administrations, de mettre en ligne l'ensemble de leurs jeux de données, proposant qu'elles puissent saisir la CADA pour être autorisées à ne pas le faire pour certains fichiers.

Sans retenir cette dernière option, qui présenterait l'inconvénient de submerger la CADA de demandes de dérogations nombreuses, et priverait les autorités administratives de la responsabilité qui est la leur d'assurer la protection des données personnelles contenues dans leurs fichiers, vos rapporteurs s'accordent sur l'intérêt d'imposer une obligation de mise en ligne sous condition, respectueuse de la vie privée des administrés.

Ils proposent donc de poser le principe que l'administration est tenue de mettre en ligne progressivement toutes les jeux de données qu'elle détient –en les anonymisant si nécessaire–, qui seraient soit déjà publiés sur un autre support, soit susceptibles d'être communiqués à tout citoyen s'il en fait la demande.

Le champ des informations ainsi publiables serait donc identique à celui couvert par la législation CADA. Il exclurait donc les informations personnelles, qui devraient être occultées, ou certaines informations couvertes par le secret, mais inclurait l'ensemble des documents qui font l'objet d'une diffusion publique, ou ceux dont la communication est prévue par la loi.

Le régime de réutilisation de ces données ne serait pas modifié par rapport au droit actuel.

L'avantage d'une telle obligation serait de **contraindre les administrations à envisager l'anonymisation, à des fins de publications, des bases de données qu'elles détiennent**.

Cependant, afin de tenir compte des contraintes légitimes de gestion que connaissent les administrations ainsi que des risques avérés de ré-identification des informations publiées, trois tempéraments seraient apportés à cette obligation.

Le premier serait le temps laissé à l'administration pour procéder progressivement à cette mise en ligne (*cf.* recommandation n° 2).

Le second serait la possibilité qui lui serait reconnue de s'opposer à la mise en ligne en raison de son coût déraisonnable ou des graves difficultés de gestion qu'elle serait susceptible d'occasionner. Il pourrait s'agir d'une impossibilité de garantir la mise à jour en temps réel des informations

publiées, ou du coût disproportionné de son anonymisation, compte tenu de la modicité du bénéfice social attendu.

Enfin, dernier tempérament, ce pouvoir d'opposition pourrait aussi être mis en œuvre si, en dépit des mesures d'anonymisation, un risque de ré-identification trop important persiste.

Recommandation n° 1

Poser le principe que l'administration est tenue de mettre en ligne progressivement, en les anonymisant si nécessaire, toutes les bases de données qu'elle détient et qui seraient susceptibles d'être communiquées à un citoyen s'il en fait la demande ou qui font l'objet d'une diffusion publique sur un autre support

L'administration ne pourrait s'y opposer qu'en raison des coûts déraisonnables de gestion que cette mise en ligne imposerait (notamment les coûts d'anonymisation éventuelle), ou du risque avéré, qu'en dépit des précautions prises, des informations personnelles puissent être ré-identifiées

La mise en œuvre de cette obligation exigera du temps, puisqu'il faudra, le cas échéant, reconfigurer ou adapter certaines bases de données – les nouvelles pourront, elles, être conçues, dès l'origine, dans cette perspective d'une diffusion en ligne. Vos rapporteurs jugent par conséquent nécessaire de poser le principe d'une phase transitoire pendant laquelle les administrations se conformeront progressivement à cette nouvelle obligation.

Pour être utile, cette phase transitoire doit être organisée : chaque service doit procéder, de manière plus systématique qu'aujourd'hui, à la recension des bases de données qu'il détient. La mise en ligne des jeux de données susceptibles de l'être doit être programmée, en tenant notamment compte du temps nécessaire à leur anonymisation éventuelle.

Ce chantier conduit par les administrations doit être placé sous la vigilance des citoyens : sauf exception, liées par exemple à des impératifs de sécurité publique, la liste de l'ensemble des bases de données élaborées ou détenues par chaque service doit être publiée sur le site du ministère, de l'établissement public ou de la collectivité en cause. Pour chaque base, mention devrait être portée de la date prévisible à laquelle elle sera mise en ligne, ou, si cette mise en ligne est refusée, de la raison de ce choix. Ainsi les citoyens pourront plus facilement contester devant la CADA ou le juge administratif la décision de l'administration.

Vos rapporteurs relèvent à cet égard, que d'ores et déjà, l'article 17 de la loi « *CADA* » fait obligation aux administrations de tenir à la disposition des usagers un répertoire des principaux documents dans

lesquelles figurent les informations publiques susceptibles de faire l'objet d'une réutilisation.

Force est toutefois de constater, que peu d'administration respectent cette règle et tiennent à jour une telle liste sur leur site internet. *Etalab* a en principe lancé un vaste mouvement de recension. Il est absolument nécessaire de le faire aboutir.

Recommandations n° 2 et 3

Afin de permettre aux administrations de satisfaire à l'obligation précédente, mettre en place une phase transitoire, pendant laquelle elles :

- opèreraient une recension complète des jeux de données qu'elles détiennent et décideraient de leur mise en ligne ;

- publieraient un calendrier pluriannuel des mises à dispositions programmées

Imposer aux administrations d'indiquer, pour chaque jeu de données, en marge du registre publié sur leur site internet les énumérant, s'il fera ou non l'objet d'une mise en ligne et, dans ce dernier cas, la raison pour laquelle elles s'y opposent

Enfin, cette nouvelle impulsion en faveur d'un *open data* respectueux de la vie privée doit aussi conduire, selon vos rapporteurs, à s'interroger sur l'opportunité d'étendre prudemment les cas prévus par la loi dans lesquels la diffusion et la réutilisation de bases de données contenant des informations personnelles peuvent être autorisées, compte tenu de l'intérêt général qui s'attache à une telle diffusion.

En effet, dans certains cas, l'exigence de transparence, ou la relative innocuité d'une réutilisation de données qui font déjà l'objet d'une publication, pourraient justifier une exception à l'interdiction de principe de diffusion, avec libre réutilisation, de données personnelles.

Chaque situation appelle un examen particulier, et vos rapporteurs se sont abstenus de trancher. Mais plusieurs cas qui leur ont été soumis au cours des auditions leur semblent mériter l'attention.

Il en va ainsi de l'interdiction de réutilisation qui frappe le fichier recensant, pour chaque médecin, les avantages qu'il a reçu d'un laboratoire pharmaceutique.

Ce fichier a été instauré par un décret du 21 mai 2013[1], conformément à l'article 2 de la loi du 29 décembre 2011 relative au renforcement de la sécurité sanitaire du médicament et des produits de

[1] *Décret n° 2013-414 du 21 mai 2013 relatif à la transparence des avantages accordés par les entreprises produisant ou commercialisant des produits à finalité sanitaire et cosmétique destinés à l'homme.*

santé. Il devrait permettre à un internaute se connectant sur le site de l'assurance maladie, d'interroger la base de données pour savoir quels avantages tel professionnel de santé a reçu de telle entreprise. En revanche, les données, considérées comme personnelles, ne seraient pas réutilisables.

De la même manière, la CADA s'est opposée à la réutilisation des tarifs déclarés par les médecins, qui sont pourtant disponibles et consultables eux aussi sur le site de l'assurance maladie[1].

Dans un cas comme dans l'autre, de solides arguments étayent le refus de réutilisation : risque d'opposition à la collecte des renseignements de la part des intéressés, mise en jeu de leur vie privée. Il est toutefois nécessaire, sur ces sujets, comme sur d'autres, que les administrations prennent le temps de reconsidérer leur poids par rapport à l'exigence d'une plus grande transparence.

Recommandation n° 4

Le cas échéant, examiner l'opportunité d'étendre les cas, définis par la loi, dans lesquels, compte tenu de l'intérêt général qui s'y attache, des jeux de données incluant des données personnelles peuvent, par exception, être diffusés en ligne et ouverts aux réutilisations

B. METTRE EN ŒUVRE UNE DOCTRINE DE PROTECTION DES DONNÉES PERSONNELLES

Pour vos rapporteurs, cette nouvelle impulsion donnée à l'*open data* est consubstantielle d'une meilleure prise en compte des dangers qui menacent les données privées. Pour prévenir ceux-ci et remédier aux faiblesses de la situation actuelle, une véritable doctrine de protection des données personnelles doit être mise en œuvre.

Cette doctrine se décline en quatre points.

1. Anticiper et évaluer

Les experts entendus par vos rapporteurs s'accordent sur un point : la protection la plus efficiente est celle conçue dès l'origine. Les bases de données créées par l'administration devraient dès leur origine garantir le respect de la vie privée, ce que l'on désigne parfois par l'expression anglaise de « *privacy by design* ». Les bases de l'assurance maladie, déjà évoquées (SNIIRAM et PMSI) ont été pensées sur ce modèle.

La démarche d'*open data* impose cependant d'adapter cette précaution : il ne faut plus seulement assurer la confidentialité de la base

[1] *Avis CADA, n° 20121581 du 19 avril 2012,* Directeur de la CNAMTS.

lorsqu'elle est utilisée par l'administration, il faut aussi garantir celle des données personnelles, si cette base devait être mise en ligne en tout ou partie.

D'un point de vue pratique, il convient donc d'une part, d'anticiper ses modalités d'anonymisation éventuelle et concevoir la structure de la base pour faciliter cette anonymisation, et, d'autre part, de veiller à ce que les jeux de données publiés puissent être tracés, comme l'a recommandé la représentante de l'association pour la diffusion de l'informatique juridique, Mme Nathalie Metallinos, afin de suivre les mésusages qui pourraient en être faits.

Recommandation n° 5

Prévoir, dès la conception de la base, dans la perspective de sa possible ouverture :

- les modalités de son anonymisation éventuelle ;

- le cas échéant, le marquage des jeux de données afin d'être en mesure de suivre les réutilisations éventuelles et dénoncer les mésusages

De telles précautions ne sont pas une assurance contre toute menace sur les données personnelles. Conformément à l'exception apportée au principe de l'*open data*, il est nécessaire d'évaluer le risque sur les données personnelles, en cas de mise en ligne, pour décider de s'y opposer ou non.

Il s'agit d'une démarche habituelle, promue notamment par la CNIL dans ses guides pratiques[1]. Elle est au cœur de la stratégie mise en œuvre par l'*Information Commissioner's Office* (ICO) –équivalent britannique de la CNIL et de la CADA réunies[2].

Elle consiste à s'interroger préalablement à l'ouverture de la base sur les risques de ré-identification ou de fuites de données personnelles, ainsi que sur leurs conséquences, et à déterminer s'il est souhaitable ou non de procéder à cette ouverture. Cette analyse est effectuée par l'administration concernée, qui peut la reconduire à intervalles réguliers, pour tenir notamment compte des nouvelles possibilités de ré-identification.

[1] Cf. *CNIL,* Guide : Gérer les risques sur les libertés et la vie privée*, disponible à l'adresse suivante :*
http://www.cnil.fr/fileadmin/documents/Guides_pratiques/CNIL-guide_Securite_avance_Methode.pdf

[2] *Sur cette stratégie,* cf. *le guide pratique de l'ICO, Anonymisation : managing data protection risk code of pratice, disponible à l'adresse suivante :*
http://ico.org.uk/~/media/documents/library/Data_Protection/Practical_application/anonymisation-codev2.pdf

Recommandation n° 6

Procéder, préalablement à tout examen de l'opportunité d'ouvrir une base de données, ainsi, le cas échéant, qu'à intervalles réguliers, à une analyse du risque de ré-identification et des conséquences possibles d'une telle ré-identification

2. Adapter la diffusion en fonction du risque

L'analyse du risque doit aboutir à une décision. Cependant, il faut éviter d'enfermer celle-ci dans un schéma binaire de refus complet ou d'acceptation totale.

En effet, le bénéfice pour la société d'une ouverture du jeu de données peut ne pas être négligeable. À cet égard, vos rapporteurs considèrent que, même s'il ne s'agit plus à proprement parler d'*open data*, puisqu'un contrôle s'exerce alors sur le réutilisateur et la finalité de sa réutilisation, des procédures d'accès restreint à certaines données sensibles participent du même mouvement, puisqu'elles permettent d'en tirer un bénéfice social.

La France dispose d'ores et déjà, avec l'accès aux données statistiques, contrôlé par le conseil national de l'information statistique et son comité du secret statistique ou l'accès aux données de santé de plusieurs modèles solides d'accès restreint, qui ont fait leurs preuves (*cf.* encadré)[1]. Ces modèles évoluent actuellement dans le sens d'une ouverture plus grande, mais toujours maîtrisée.

De telles solutions peuvent utilement inspirer le législateur ou les administrations, pour définir un continuum de modalités d'accès aux informations détenues par les administrations, afin d'adapter la diffusion au risque pesant sur les données personnelles.

Recommandations n° 7 et 8

En cas de risque avéré sur les données personnelles, impossible à éliminer par des procédés d'anonymisation, refuser l'ouverture des données ou, si le bénéfice social attendu de cette ouverture est jugé trop important, procéder à une ouverture restreinte de cette base

Concevoir à cette fin un *continuum* de solutions d'accès aux données, allant de l'*open data*, jusqu'aux modes d'accès les plus sélectifs

[1] *D'autres modèles existent, comme celui de l'accès aux informations fiscales, récemment étendu par la modification apportée par la loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche à l'article L. 135 D du livre des procédures fiscales, ou celui de l'accès au fichier national des immatriculations de véhicules (art. L. 330-1 à L. 330-5 du code de la route).*

Deux exemples d'accès restreint à des données sensibles

• ***L'accès aux données de santé enregistrées dans le SNIIRAM (système d'information inter-régime de l'assurance maladie)***

La réglementation en vigueur distingue deux types d'accès à cette base de données, l'un permanent, l'autre ponctuel.

L'accès permanent est régi principalement par un arrêté du ministre chargé de la sécurité sociale, pris après avis motivé de la CNIL, qui désigne les organismes ou administration bénéficiaire d'un tel accès.

L'étendue de cet accès varie en fonction de la nature de l'entité concernée. Ainsi, seuls les organismes gestionnaires de l'assurance maladie (régimes de base d'assurance maladie, caisse nationale de solidarité pour l'autonomie) et les agences publiques exerçant une mission de veille dans le domaine de la santé (institut de veille sanitaire, haute autorité de santé, agence nationale de la sécurité du médicament et des produits de santé, médecins des agences régionales de santé) ont accès à la totalité des données.

Ont accès aux données agrégées de la base, ainsi qu'à un échantillon des bénéficiaires, les services des ministères compétents et ceux des agences régionales de santé, certaines agences de santé (agence de la biomédecine, agence technique de l'information sur l'hospitalisation, institut des données de santé…), des centres de recherches (CNRS, institut national du cancer, institut national de la santé et de la recherche médicale…), ainsi que des fédérations professionnelles ou de patients (fédérations hospitalières, union des professions de santé, collectif interassociatif sur la santé).

Enfin, ont seulement accès aux données agrégées, les fédérations professionnelles régionales, ou les associations membres des collectifs associatifs précédemment évoqués[1]

Les accès ponctuels correspondent à l'accès, limité dans le temps, à une sous-base du SNIIRAM, ou à l'obtention d'une extraction des données, à des fins de recherche principalement. Seules les demandes à but non lucratif sont recevables.

Selon le cas, l'autorisation d'accès doit être validée par l'institut des données de santé (accès temporaire aux données agrégées ou à l'échantillon des bénéficiaires) ou par cet institut et la CNIL (extraction de données du SNIIRAM). Les demandes sont instruites selon une procédure complexe qui peut aussi faire intervenir, outre ces deux institutions, le comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé ou le conseil national de l'information statistique.

• ***L'accès aux données des enquêtes de l'INSEE***

Les fondements de la protection des données des enquêtes statistiques conduites par l'INSEE ont été posés dès les années 50, par la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, qui a créé, en contrepoint de l'obligation de répondre aux enquêtes statistiques, la garantie du secret sur les informations ainsi communiquées.

Cette loi a institué un comité du secret statistique, placé auprès du conseil national de l'information statistique, et chargé de se prononcer sur toute demande d'accès aux données individuelles collectées dans le cadre d'enquêtes de l'INSEE, formulée par une équipe de recherche.

[1] *Pour une liste complète* cf. *Pierre-Louis Bras, André Loth,* Rapport sur la gouvernance et l'utilisation des données de santé, *préc., p. 31-34.*

Afin de faciliter l'accès à ces données, en conservant le même degré de protection des données personnelles, l'INSEE a mis en place un dispositif technique, le centre d'accès sécurisé distant aux données (CASD), qui permet au service producteur des données de surveiller les opérations effectuées sur les données, d'éviter certains croisements ou extractions de données, tout en apportant au chercheur l'ensemble des éléments dont il a besoin.

Les équipes de recherche ont aussi la possibilité d'accéder à des données anonymisées, soit publiées directement sur le site de l'INSEE sous forme de statistiques agrégées, soit extraites à la demande et anonymisées, sous condition de licence d'usage, interdisant notamment la rediffusion à destination de tiers.

3. Assurer une veille sur la diffusion et les réutilisations des données mises en ligne

Comme on l'a vu précédemment, les risques pour la vie privée de nos concitoyens procèdent soit d'une ré-identification des données personnelles à la faveur de certaines réutilisations, soit d'un manque de vigilance des administrations qui, sans le savoir, auraient publié dès l'origine des informations privées.

La doctrine de protection des données personnelles de l'administration, dans le cadre de l'*open data*, ne serait donc pas complète, si elle n'incluait pas une veille sur la diffusion et les réutilisations des jeux de données publiés.

Certes, ce contrôle échoit aussi à la CNIL, qui peut sanctionner les réutilisateurs qui cherchent à lever l'anonymisation des jeux de données. Toutefois, il sera souvent plus expédient que l'administration, avertie de la fuite, s'efforce d'y parer.

Vos rapporteurs se sont interrogés sur la façon dont cette veille devrait s'exercer et elle leur a paru pouvoir emprunter deux canaux : le premier est celui de l'alerte citoyenne. Si le site *data.gouv.fr* permet déjà aux réutilisateurs d'avertir par mail l'hébergeur des données de ce qu'elles paraissent présenter un défaut, tous les sites des administrations, loin de là, ne prévoient pas une formule aussi commode, pourtant très habituelle sur internet. Le procédé pourrait donc être étendu.

Par ailleurs, il conviendrait de tirer parti de ce que les administrations sont elles-mêmes consommatrices des jeux de données qu'elles diffusent, *via*, par exemple leur intranet, ce qui leur offre l'occasion, dans leurs tâches habituelles de gestion de déceler les éventuels défauts des informations publiées. Ceci renforce leur capacité de veille.

Cette obligation de veille devrait être étendue aux jeux de données publiés par d'autres contributeurs que les administrations, et hébergés sur un site public, comme celui de *data.gouv.fr*, qui met en avant cette démarche collaborative.

En effet, si la loi pour la confiance dans l'économie numérique[1] a mis en place un régime de responsabilité limité pour l'hébergeur de contenus édités par un tiers, ce dernier régime ne concerne que la responsabilité civile de l'hébergeur. Or, les personnes publiques relèvent en principe d'un régime de responsabilité administrative. Il n'est donc pas certain que la limitation de responsabilité prévue par cette loi s'applique à elle, d'autant plus que l'activité d'hébergeur exercée par la puissance publique dans le cadre de l'*open data* pourrait tout à fait être conçue comme une activité d'intérêt général, dans la mesure où il s'agit, par ce biais, d'enrichir le bien commun des données mises à disposition. Cette incertitude sur le régime juridique applicable à l'administration a été confirmé à vos rapporteurs par les services de la CNIL. Elle pourrait être lourde de conséquence, puisqu'un particulier, victime de la diffusion de données personnelles sur un site public, pourrait ainsi être autorisé à poursuivre l'administration responsable de cette mise en ligne.

Il apparaît donc sage d'étendre le devoir de surveillance des administrations aux données publiées par des tiers sur leur propre site, qui peuvent parfois d'ailleurs être des données publiques retraitées par leurs soins.

Recommandations n° 9 et 10

Assurer une veille sur la diffusion et les réutilisations des données publiques, en facilitant notamment les procédures par lesquelles un réutilisateur peut alerter l'administration compétente

Assurer aussi cette veille sur les données publiées par des tiers sur les sites publics

En cas de ré-identification ou de diffusion accidentelle de données personnelles, l'administration doit y porter remède.

Ceci suppose bien entendu de mettre fin à la mise en ligne des informations compromises. D'autres solutions sont envisageables : reconfigurer la base de données, pour prévoir un degré d'agrégation supérieur ou supprimer la donnée à l'origine du problème.

Le rapatriement des jeux de données compromis se heurte au fait qu'une donnée ouverte est une donnée qui circule et est reprise par d'autres. Toutefois, dans les cas les plus graves, il peut être du devoir de l'administration de tenter malgré tout de rapatrier ces données ou d'en limiter la circulation. Le marquage des jeux de données trouve ici son utilité, comme l'archivage, par l'administration, des coordonnées des réutilisateurs si ceux-ci ont accédé au service par un compte dédié ou une adresse mail. L'administration peut aussi demander le déréférencement sur les moteurs de

[1] *Article 6 de la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique.*

recherches des jeux de données compromis, hébergés sur d'autres sites que le sien, ou proposer aux réutilisateurs de leur adresser une nouvelle base non défectueuse.

Vos rapporteurs s'étonnent que cette question importante ne soit pas traitée par le mémento précité du COEPIA, consacré à la protection des données personnelles dans le cadre de l'*open data*. Ils jugent ainsi nécessaire que l'administration se penche sur la stratégie de rapatriement ou de suppression des jeux de données compromis qu'elle pourrait mettre en œuvre, afin de réagir rapidement et efficacement en une telle occurrence.

Recommandation n° 11

Prévoir que l'administration définisse une stratégie de rapatriement ou de suppression des jeux de données compromis, afin de remédier rapidement à la diffusion accidentelle d'informations personnelles

4. Renforcer la protection offerte par la licence de réutilisation

Le recours aux licences en matière d'*open data* vise à garantir la libre réutilisation des données d'un réutilisateur à l'autre, ou, dans certains cas spécifiques, à encadrer les conditions de cette réutilisation (redevance, limitation des droits *etc.*).

Il semble à vos rapporteurs que cet instrument pourrait aussi être mobilisé pour renforcer la protection des données personnelles.

Vos rapporteurs se sont à cet égard étonnés que la licence ouverte publiée par *Etalab*, n'exclue pas expressément les données personnelles de son champ. L'expression « *données personnelles* » n'est d'ailleurs mentionnée qu'à l'occasion de la citation de l'article 10 de la loi « *CADA* », dans un « à propos » explicatif à la fin du contrat de licence.

Même si cette absence est de peu d'effet, puisque les dispositions légales s'imposent en tout état de cause et qu'elles excluent la réutilisation de données personnelles, en dehors du cadre fixé par la loi « *Informatique et libertés* », elle témoigne assurément d'un manque de pédagogie. Cette lacune est d'autant plus surprenante que la licence ouverte britannique[1], rappelle expressément que les données personnelles sont exclues de son champ d'application.

De la même manière vos rapporteurs jugent pertinent de préciser, au titre des prohibitions d'usage, l'interdiction de soumettre le jeu de données utilisé à un traitement destiné à permettre la ré-identification de personnes physiques.

[1] *Open Government Licence, disponible à l'adresse suivante :* http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/

Enfin, il leur semble nécessaire d'intégrer aux contrats de licence une clause prévoyant que le service producteur peut suspendre le droit de réutilisation, supprimer ou demander le rapatriement du jeu de données, s'il s'avère qu'il présente un risque pour le respect de la vie privée. L'intérêt d'une telle mention serait d'éviter tout recours contre l'administration, sans faute lourde, pour le préjudice éventuellement causé au réutilisateur à raison de la suppression de ce jeu d'informations.

Recommandations n° 12, 13 et 14

Exclure expressément les données personnelles du champ d'application de la licence ouverte utilisée par les administrations pour la réutilisation des données publiques

Interdire expressément dans le contrat de licence toute réutilisation abusive qui aboutirait à lever l'anonymisation des données

Intégrer au contrat de licence, une clause de suspension légitime du droit de réutilisation, ainsi que de suppression ou de rapatriement des jeux de données compromis lorsqu'un risque de ré-identification est apparu

C. *ADAPTER LA GOUVERNANCE DE L'*OPEN DATA *AUX EXIGENCES DE LA PROTECTION DES DONNÉES PERSONNELLES*

Les auditions conduites par vos rapporteurs ont mis en évidence le manque d'assistance apportée aux différentes administrations nationales et locales pour assurer de manière satisfaisante la protection des données à caractère personnel dans la mise en œuvre de l'*open data*. Un besoin d'informations et de conseils pratiques s'est exprimé en matière de protection des données personnelles, auquel seule une gouvernance structurée serait à même de répondre.

1. Organiser l'assistance aux acteurs de l'*open data*

Vos rapporteurs ont acquis la conviction que la gouvernance décentralisée, structurée en réseau, de la stratégie d'*open data* telle qu'elle a été conduite jusqu'à présent était la plus adaptée à un mouvement de cette nature. Il n'est en effet pas souhaitable de confier la politique d'*open data* à un organisme unique qui se substituerait aux producteurs et se chargerait seul de la mise en ligne des données de l'ensemble des administrations et établissements publics de l'État et des collectivités territoriales. Cela risquerait de déresponsabiliser les producteurs alors même qu'il est indispensable que chacun intègre la logique d'*open data* non seulement dans son action, mais également dans son travail quotidien de recueil et de traitement des données.

En revanche, il apparaît indispensable d'instaurer un référent unique à même de répondre aux différentes interrogations et besoins techniques des producteurs de données portant spécifiquement sur la problématique de la protection des données à caractère personnel dans l'*open data*. C'est pourquoi vos rapporteurs préconisent que soit instituée auprès de la mission *Etalab* une structure spécifiquement dédiée à cette problématique.

Recommandations n° 15 et 16

Mettre en place, auprès de la mission *Etalab*, une structure dédiée à la protection des données personnelles et chargée d'assister les administrations :

- dans l'élaboration de l'étude d'impact préalable à la mise à disposition des données ;

- dans l'anonymisation éventuelle de la base ;

- dans la mise en place d'un mode d'accès restreint

Confier à cette même structure un rôle de veille sur les réutilisations abusives au regard de la protection des données personnelles, en la chargeant de recueillir les alertes éventuelles, d'en informer la CNIL, et de coordonner, le cas échéant, le retrait ou la reconfiguration de la base de données litigieuse

Par ailleurs, vos rapporteurs ont pris connaissance avec intérêt de la plateforme britannique *UK Anonymisation Network (UKANON)*. Mise en place l'an passé, cette plateforme, accessible sur internet, a pour objectif de recenser les bonnes pratiques en matière d'anonymisation et d'offrir conseils et renseignements à toute personne souhaitant traiter et diffuser des données personnelles après anonymisation. Financée par l'*Information Commissioner's Office (ICO)*, cette initiative est coordonnée par un consortium réunissant l'Université de Manchester, l'Université de Southampton, l'*Open Data Institute (ODI)* et l'*Office for National Statistics*, homologue de l'INSEE.

Comme nombre de personnes entendues lors des auditions le faisaient remarquer à vos rapporteurs, la France dispose actuellement de différents organismes dont l'expérience en matière d'anonymisation des données personnelles pourrait être non seulement mutualisée, mais également mise à disposition de tous afin de sécuriser la diffusion de données publiques issues de données personnelles. Une initiative similaire à *UKANON* pourrait être conduite sous l'égide de la CNIL et avec le concours de la CADA et de la mission *Etalab*, des opérateurs publics spécialisés comme l'INSEE, ainsi que de laboratoires de recherche tels ceux de l'INRIA. Cela permettrait de faire se rencontrer les approches juridique et technique afin de mettre à disposition des usagers tant des recommandations d'usage que des solutions techniques.

Recommandation n° 17

Rassembler et diffuser les bonnes pratiques et les recommandations en matière de protection des données personnelles dans l'*open data*

Les « *correspondants Informatique et libertés* » (CIL) présents au sein des administrations et établissements publics et les quelques 1 600 PRADA peuvent par ailleurs constituer des relais efficaces pour faire pénétrer au sein des administrations les problématiques de protection des données à caractère personnel. Ils présentent en effet l'avantage d'assurer un contrôle de proximité des traitements de données.

Confier de nouvelles missions aux CIL dans le cadre de l'*open data* conduit cependant à poser de nouveau la question du renforcement de leur statut. À cet égard, vos rapporteurs souhaitent rappeler les travaux sur ce sujet du groupe de travail de votre commission[1] qui a débouché sur l'adoption par le Sénat, le 23 mars 2010, de la proposition de loi de nos collègues Anne-Marie Escoffier et Yves Détraigne. Cette proposition contient en particulier des dispositions visant à rendre obligatoire la désignation d'un CIL dans toutes les structures dans lesquelles plus de cent personnes mettent en œuvre des traitements de données personnelles ou y ont directement accès, ainsi que dans les structures recourant à des traitements de données soumis au régime d'autorisation préalable. Elle tend également à préciser le rôle du CIL ainsi que ses liens étroits avec la CNIL. Cette proposition de loi est toujours en instance devant l'Assemblée nationale.

Recommandation n° 18

Investir les CIL et les PRADA d'attribution de coordination et de veille en matière de protection des données personnelles dans le cadre de l'*open data*

2. Garantir le financement des mesures d'anonymisation

Vos rapporteurs ont pris acte de la décision du Gouvernement de réaffirmer le principe de gratuité de la réutilisation des données publiques, prise à la suite de la remise au Premier ministre du rapport de M. Mohamed Adnène Trojette sur la légitimité des exceptions au principe de gratuité dans le cadre de l'*open data*[2]. Ainsi, lors du comité interministériel pour la modernisation de l'action publique du 18 décembre 2013, le Gouvernement

[1] *Rapport d'information fait au nom de la commission des lois par le groupe de travail relatif au respect de la vie privée à l'heure des mémoires numériques (n° 441, 2008-2009).*

[2] Cf. Ouverture des données publiques – Les exceptions au principe de gratuité sont-elles toutes légitimes ?, *rapport remis au Premier ministre par M. Mohammed Adnène Trojette, juillet 2013.*

a-t-il décidé de ne plus autoriser la création de nouvelle redevance et de supprimer certaines des redevances existantes.

Vos rapporteurs considèrent cependant qu'eu égard à l'importance de l'enjeu de la protection de la vie privée, il est impératif que l'État assure l'anonymisation de certains jeux de données plutôt que de renoncer à leur ouverture. Dès lors, ils demandent au Gouvernement de s'engager sur le financement pérenne par le budget de l'État des mesures d'anonymisation indispensables à la mise en ligne de ces jeux de données.

Toutefois, conscients du coût de l'anonymisation de certains jeux de données dans un contexte de restriction budgétaire, vos rapporteurs estiment qu'au regard des bénéfices attendus de l'*open data* par certains acteurs économiques, il ne serait pas de bonne gestion de renoncer par principe au prélèvement de redevances visant spécifiquement le financement de l'anonymisation.

Par ailleurs, vos rapporteurs invitent le Gouvernement à expérimenter de nouvelles modalités de financement. Ils proposent ainsi d'encourager la contribution volontaire des différents acteurs intéressés à l'ouverture de certains jeux de données qu'ils estimeraient à fort potentiel et qui pourtant ne seraient pas mis à disposition, soit que les administrations n'en aient pas prévu la diffusion dans le cadre de leur programme pluriannuel, soit que ces acteurs souhaitent que cette diffusion soit accélérée par rapport au programme arrêté, soit enfin, que les administrations aient refusé de diffuser certains jeux de données du fait de l'effort disproportionné que représentait leur anonymisation, en application de l'article 40 du décret d'application de la loi « *CADA* ».

À cet égard, une première solution pourrait être de recourir au financement coopératif. Vos rapporteurs ont noté avec intérêt les annonces faites le 14 février dernier par Mme Fleur Pellerin, alors ministre déléguée chargée des petites et moyennes entreprises, de l'innovation et de l'économie numérique, concernant le renouvellement du cadre juridique relatif au financement participatif ou « *crowdfunding* », qui met en relation, via des plates-formes internet, des porteurs de projets en quête d'argent avec des particuliers désireux de donner, d'investir ou de prêter. Une autre voie serait de recourir à une forme de mécénat.

Recommandation n° 19

Garantir le financement par l'État des mesures d'anonymisation des données personnelles contenues dans des jeux de données publiques

Ne pas renoncer par principe au prélèvement d'une redevance en présence de coûts d'anonymisation élevés

Encourager le financement coopératif de l'anonymisation

3. Clarifier le droit applicable en matière de réutilisation de données publiques contenant des données personnelles

Un consensus s'est dégagé au cours des auditions conduites par vos rapporteurs pour dénoncer le manque de clarté du cadre juridique dans lequel s'inscrit l'*open data*. Si certains préconisent une simple clarification des concepts utilisés et une harmonisation des jurisprudences entre CADA et CNIL, d'autres vont plus loin et, soulignant le paradoxe d'une législation à la fois très protectrice et inadaptée car conçue à une époque où les développements actuels n'étaient pas envisagés, souhaitent une remise à plat de ce cadre juridique pour instaurer un véritable droit à la réutilisation.

Vos rapporteurs estiment quant à eux que la transposition, attendue avant la date-limite du 18 juillet 2015, de la directive 2013/37/UE du Parlement européen et du Conseil du 26 juin 2013 modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public, pourrait être l'occasion de remédier à certaines lacunes du cadre juridique actuel. Il leur a cependant semblé que cette question excédait le champ de la mission qui leur a été confiée par votre commission et relevait davantage de la mission commune d'information consacrée à l'accès aux documents administratifs et aux données publiques.

Néanmoins, ils ont souhaité faire état d'une difficulté qui leur est apparue s'agissant de la réutilisation d'informations publiques contenant des données à caractère personnel ayant fait l'objet d'une publicité légale. La conciliation de l'obligation de publication et de celle d'occultation, posée au troisième alinéa de l'article 7 de la loi « *CADA* », est en effet délicate, comme l'illustre le conseil CADA n° 20121488 du 7 juin 2012, *Président du conseil régional de Bourgogne*, à propos de l'obligation de publication des délibérations du conseil régional. L'occultation des données à caractère personnel, rendue nécessaire par les dispositions de la loi « *Informatique et libertés* » en cas de mise en ligne sur internet, risquait de vider de son sens la publication de l'information publique en cause. Aussi la CADA a-t-elle préconisé de s'en tenir à une publication sur supports traditionnels[1], à l'instar de la pratique du *Journal officiel* s'agissant des mesures de naturalisation par exemple.

[1] « Au regard des instruments servant de support à la publication, la commission estime qu'en prescrivant la publication, la loi autorise nécessairement le recours aux supports traditionnels que constituent l'affichage aux lieux habituels et les recueils des actes administratifs. La commission considère en revanche que l'obligation légale de publication n'autorise le recours à la mise en ligne de l'acte sur un site internet que pour autant que sont en outre satisfaites les dispositions de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Ces dernières, en faisant éventuellement obstacle à la mise en ligne de mentions nécessaires à une publication suffisante, peuvent imposer le recours aux supports traditionnels. »

Cette position de la CADA pourrait être reprise dans une disposition législative.

Recommandation n° 20

Préciser que, lorsque des données personnelles sont mises en ligne en vertu de la loi, cette publication doit se limiter à la stricte mesure nécessaire au respect de l'objet visé par cette loi

CONCLUSION GÉNÉRALE

L'ouverture et le partage des données publiques représentent, pour les administrations, un défi de taille, et pour les citoyens ou les entreprises, une opportunité considérable. Rien ne serait pire, cependant, que de manquer l'un ou l'autre, faute de s'être assuré que cette ouverture respecte bien l'intérêt de chacun.

Notre pays s'est honoré à placer l'exigence de protection de la vie privée au cœur de la révolution numérique. Les fondements solides de notre droit devraient, en principe, garantir que l'*open data* se déroule sans heurt, car les lois de 1978 excluent que les administrations puissent diffuser autre chose que des informations publiques.

Les premiers développements de l'*open data* français peuvent sembler rassurants, de ce point de vue : à de rares exceptions près, les données mises en ligne paraissent se conformer aux exigences juridiques.

Toutefois, ce cadre protecteur est fragilisé par deux failles : un risque avéré de ré-identification et un défaut de gouvernance qui laisse souvent les administrations démunies face à des problématiques qu'elles ne maîtrisent pas toutes avec la même aisance.

Loin de vouloir freiner un mouvement dont l'utilité sociale est acquise, votre mission d'information a plutôt vu l'opportunité de donner un nouvel élan à l'ouverture et au partage des données publiques, à la condition que soient mises en œuvre une doctrine et une méthode qui garantissent la meilleure protection des données personnelles possible. Car, une fois cette protection assurée, aucun obstacle au déploiement de l'*open data* n'est plus légitime.

La mission d'information de votre commission recommande donc de poser le principe d'une obligation de mise en ligne des données détenues par les administrations, à moins que le coût en soit trop important ou que les risques pour la vie privée ne puissent être levés par une anonymisation efficace.

Les administrations de l'État et des collectivités territoriales doivent s'investir dans cette voie et mettre en œuvre une doctrine de protection des données personnelles, en anticipant et évaluant les risques, en y adaptant les formats de diffusion des données et en exerçant une veille vigilante.

Elles doivent être secondées dans cette tâche par l'État, qui devra veiller à leur apporter, par une structure dédiée, une assistance technique, organisationnelle et juridique, et garantir le financement des chantiers qu'engage l'*open data*, notamment en matière d'anonymisation.

Trouver dans la protection des données personnelles le levier pour porter plus haut l'exigence de transparence de l'action publique : le pari est audacieux, mais il est conforme à l'ambition que notre pays a démontrée depuis les lois pionnières de 1978.

LISTE DES PERSONNES ENTENDUES

Conseil d'État

M. Christian Vigouroux, président de la section du rapport et des études

M. Jacky Richard, rapporteur général

M. Laurent Cytermann, rapporteur général adjoint

Direction générale des finances publiques

M. Bruno Rousselet, chef du service de la gestion fiscale

Mme Catherine Brigant, sous-directrice des missions foncières, de la fiscalité du patrimoine et des statistiques

Commission nationale de l'informatique et des libertés

Mme Isabelle Falque-Pierrotin, présidente

M. Édouard Geffray, secrétaire général

Mme Tiphaine Inglebert, conseillère pour les questions parlementaires et institutionnelles

M. Gwendal Legrand, chef du service de l'expertise informatique

M. Stéphane Petitcolas, ingénieur expert

Commission d'accès aux documents administratifs

M. Serge Daël, président

M. Nicolas Polge, rapporteur général

Mission Etalab

M. Henri Verdier, directeur de la mission

Mme Suzanne Vergnolle, chargée de mission Affaires publiques et juridiques

Agence du patrimoine immatériel de l'État

Mme Danielle Bourlange, directrice générale

Institut national de la statistique et des études économiques

M. Michel Isnard, chef de l'unité « Affaires juridiques et contentieuses »

Conseil national du numérique

M. Jean-Baptiste Soufron, secrétaire général

Mme Somalina Pa, rapporteur

M. Charly Berthet, rapporteur adjoint

Comité du secret statistique

M. Jean Gaeremynck, président

M. Michel Isnard, secrétaire

Institut national de l'information géographique et forestière

M. Pascal Berteaud, directeur général

M. François Baudet, Secrétaire général

Open Data France

Mme Erwane Monthubert, conseillère municipale de Toulouse, déléguée aux Technologies de l'information et de la communication

Mme Sandrine Mathon, chef du service Administration, Direction des Systèmes d'Information, Mairie de Toulouse et Toulouse Métropole

Association française des correspondants à la protection des données à caractère personnel

M. Paul-Olivier Gibert, président

Association des archivistes de France

M. Jean-Philippe Legois, président, directeur des Archives municipales de Sevran

Mme Katell Auguié, vice-présidente, responsable du service de l'information d'Orvault

Mme Cyril Longin, coordinateur du groupe Open data de l'association, directeur des Archives municipales de Saint Etienne

M. Hervé Bousquet, trésorier, archiviste chez Sanofi

Mme Charlotte Maday, présidente de la section Archives des universités, rectorat et organismes de recherche

Mme Alice Grippon, déléguée générale

Fédération française des sociétés d'assurance

M. Arnaud Chaput, directeur Prospective et innovation

M. François Rosier, directeur adjoint des affaires juridiques, fiscales et de la concurrence

Mme Cécile Malguid, responsable des études santé, direction des assurances de personnes

M. Jean-Paul Laborde, directeur des affaires parlementaires

Mme Viviana Mitrache, attachée parlementaire

Mouvement des entreprises de France

M. Marc Lolivier, président du Comité « Droit du numérique »

Mme Émilie Dumerain, juriste, chargée de mission, direction droit de l'entreprise

Mme Ophélie Dujarric, chargée de mission senior, direction des affaires publiques

UFC-Que choisir

M. Mathieu Escot, chargé d'études santé

M. Antoine Autier, chargé de mission « nouvelles technologies »

GrDF

M. Anthony Mazzenga, délégué du pôle Stratégie

M. Sylvain Chapon, délégué des Affaires publiques

Reflets.info

M. Antoine Champagne, directeur de la rédaction

iFRAP

Mme Agnès Verdier-Molinié, directeur

M. Samuel-Frédéric Servière, chercheur en finances publiques

Regards Citoyens

M. David Gayou, administrateur, ingénieur recherche et développement

M. Tangui Morlier, administrateur, consultant en informatique

Conseil d'orientation de l'édition publique et de l'information administrative

M. Michel Pinault, président

M. Olivier Garnier, secrétaire

M. Eric Gristi, secrétaire adjoint

Association pour le développement de l'informatique juridique

Mme Élise Debiès,

Mme Nathalie Metallinos,

co animatrices de l'atelier « Protection des Données Personnelles »

Institut national de la recherche informatique automatique

M. Claude Kirchner, délégué général à la recherche et au transfert pour l'innovation (DGRTI)

Fondation internet nouvelle génération

M. Charles Népote, chef de projet « Partage des données publiques »

Personnalités qualifiées

M. Mohammed Adnène Trojette, auteur d'un rapport sur la rémunération de l'open data

M. François Bancilhon, président-directeur général de Data Publica

M. Alain Bensoussan, avocat spécialiste du droit des technologies de l'information et de la communication

Mme Danièle Bourcier, professeur de droit public

M. Simon Chignard, auteur de « L'Open data, comprendre l'ouverture des données publiques »

M. Gilles Guglielmi, professeur de droit public, Université Paris-II Panthéon Assas

M. Nicolas Kayser-Bril, journaliste

M. Jean-Marc Lazard, président directeur général d'OpenDataSoft

M. André Loth, directeur de projet (DREES), co-auteur d'un rapport sur la gouvernance et l'utilisation des données de santé

M. Jean-Marc Manach, journaliste

Mme Nathalie Mallet-Poujol, directrice de recherche au CNRS

M. Benoît Tabaka, directeur des politiques publiques de Google France

Mme Véronique Tauziac, experte auprès du chef du service du pilotage et des politiques transversales à la direction générale de l'administration et de la fonction publique

M. Gilles Trouessin, consultant en sécurité informatique, auteur d'une étude sur la désanonymisation des fichiers de santé