

A S S E M B L É E N A T I O N A L E

X I V ^e L É G I S L A T U R E

Compte rendu

Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique

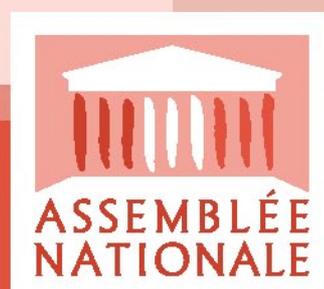
– Audition de M Henri Verdier, directeur d'Etalab, administrateur général des données sur l'*open data* et de M. Mohammed Adnène Trojette, conseiller référendaire de la Cour des Comptes, sur le principe de gratuité d'usage des données publiques2

Mercredi
1^{er} octobre 2014
Séance de 18 heures 30

Compte rendu n° 05

SESSION ORDINAIRE DE 2014-2015

**Présidence de
M. Christian Paul,
coprésident**

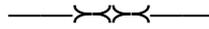


COMMISSION DE RÉFLEXION ET DE PROPOSITIONS SUR LE DROIT ET LES LIBERTÉS À L'ÂGE DU NUMÉRIQUE

Mercredi 1^{er} octobre 2014

La séance est ouverte à dix-huit heures quarante.

(Présidence de M. Christian Paul, co-président)



La Commission procède à l'audition de M. Henri Verdier, directeur d'Etalab, administrateur général des données, et de M. Mohammed Adnène Trojette, conseiller référendaire, auteur en 2013 d'un rapport au Premier ministre sur le modèle économique des redevances d'utilisation des données publiques, sur l'open data et le principe de gratuité d'usage des données publiques

M. le coprésident Christian Paul. Mme Christiane Féral-Schuhl, coprésidente de la commission, est retenue par une cérémonie publique et m'a prié de vous transmettre ses excuses.

Je signale la parution de l'étude annuelle du Conseil d'État, qui porte en 2014 sur « Le numérique et les droits fondamentaux ». Plusieurs d'entre nous ont été auditionnés dans le cadre de ce travail. Il s'agit d'une somme très utile à la fois par l'analyse détaillée de l'état du droit et par les propositions qu'elle contient. Nous recevrons ses auteurs très prochainement.

J'en viens à l'objet de notre séance : l'ouverture des données publiques, c'est-à-dire la publication, par des entités publiques – mais peut-être aussi des entités privées –, d'informations dans des formats qui en permettront la réutilisation. Nous entendrons M. Henri Verdier, membre de notre commission et directeur de la mission Etalab, qui pilote la politique d'ouverture des données publiques, administre le portail data.gouv.fr et accompagne les administrations, les collectivités et les organismes publics dans leur démarche d'ouverture des données. Nous entendrons également M. Mohammed Adnène Trojette, auteur en 2013 d'un rapport au Premier ministre très remarqué sur l'ouverture des données publiques et, plus particulièrement, sur le modèle économique des redevances d'utilisation des données publiques et sur la création d'activité économique résultant du processus d'ouverture.

Je propose d'articuler notre réflexion autour de trois axes.

Premièrement l'ouverture des données publiques. Quel bilan peut-on en dresser en France et dans d'autres pays ? De quelle nature sont les obligations juridiques ? Présentent-elles, le cas échéant, des faiblesses ? Faut-il aller plus loin et inscrire dans la loi le principe d'ouverture des données, ou en rester, comme l'estime le Conseil d'État, à un droit très souple ?

Deuxièmement la protection des droits individuels dans ce contexte d'ouverture massive. Quelles garanties apporter en matière d'anonymisation et, plus généralement, de protection des données personnelles ? La protection est-elle un frein à l'ouverture ? Quelle est l'articulation entre les pouvoirs de la CADA Commission d'accès aux documents administratifs (CADA), qui a un rôle précontentieux, ceux de la Commission nationale de

l'informatique et des libertés(CNIL), qui veille au respect de la loi « *informatique et libertés* », et Etalab ?

Troisièmement la dimension économique de l'ouverture des données publiques. Ce processus permet aux administrations, aux entreprises et, au fond, à tout un chacun, d'exploiter et de valoriser des contenus. Il a donc une utilité économique, provoquant création de richesse et créations d'entreprises ; il a également une utilité sociale. Comment, dès lors, valoriser ces données au départ. L'accès doit-il être gratuit ou non ?

M. Henri Verdier, directeur d'Etalab, administrateur général des données. Si je suis chargé, au sein des services du Premier ministre, de la politique d'ouverture des données publiques, je parle ce soir en mon nom propre. Nous sommes une commission de réflexion et de prospective : les questions naissantes que nous abordons sont loin d'être toutes tranchées !

Le mouvement de l'*open data* n'est pas qu'une expression, c'est une réalité que la démarche d'ouverture des données publiques et de transparence n'épuise pas. Il est l'occasion pour l'État de faire sa révolution numérique, c'est-à-dire d'entrer dans un ensemble de pratiques, de relations avec l'extérieur, de manières d'agir, d'intervenir dans la société et de créer de la valeur, qui sont elles-mêmes contemporaines du numérique.

Un premier paradoxe révèle à quel point ces questions sont neuves et complexes : on oscille sans arrêt, à leur sujet, entre un récit sur la transparence et un récit économique, en cherchant désespérément des solutions qui fonctionneraient pour les deux. Pour ma part, je ne suis pas certain que l'on puisse durablement fonder une démarche de création de biens communs numériques sur des concepts issus d'une réflexion sur la transparence.

En France, cependant, l'*open data* prend appui sur toute une tradition juridique de construction progressive de la transparence. La loi 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, dite loi CADA, en est une étape importante. Comme le souligne le Conseil d'État, il s'agit, somme toute, du seul levier en « droit dur » pour qui veut chercher les données et faire valoir ses droits.

L'*open data* ressortit également à une culture d'ingénieur. Le souhait d'obtenir des données brutes perturbe parfois les administrations : elles estiment qu'on peut faire dire tout et son contraire à ces données, qu'elles comportent des erreurs, etc. Mais, du point de vue d'un ingénieur qui sait jongler avec les données, il est insupportable de passer deux jours à toiler un fichier auquel quelqu'un aura ajouté des règles de format, des couleurs ou des cryptogrammes. Lorsque Tim Berners-Lee fait scander à l'assistance d'une conférence TED (*Technology, Entertainment and Design*) : « *We want raw data !* », « *Nous voulons des données brutes !* », ce n'est pas un postulat philosophique ou épistémologique. Cela signifie : « Laissez-nous travailler, nous pouvons faire plus vite et mieux que vous ! Comment libérer toute la puissance de la donnée si on ne laisse pas les gens se faire une opinion en quelques minutes ? » Combien de fois ne sommes-nous pas contraints d'aller chercher une information à la page 1445 de tel ou tel rapport alors que, si nous avions disposé du tableau brut initial, une simple règle de trois nous aurait permis de retrouver cette donnée, assortie de nombreux autres chiffres correspondant à notre recherche ?

La réflexion sur les biens communs est ancienne, mais le web 2.0 lui a donné une nouvelle actualité. Des personnalités comme Tim O'Reilly ou Beth Noveck, fascinées, au tournant des années 2000, par la capacité des gens à fabriquer Wikipédia, TripAdvisor,

OpenStreetMap, se sont demandé comment permettre à cette formidable énergie créatrice de s'investir également dans la sphère publique : les États doivent apprendre non seulement à délivrer des ressources pour aider à résoudre différents problèmes, mais à devenir eux-mêmes des plateformes pour que les gens prolongent et amplifient l'effort de la puissance publique. La donnée est souvent la première condition, mais il peut y en avoir d'autres : modèles, codes, puissance de calcul, etc. Cette deuxième tradition vient alimenter le récit de transparence.

En outre, la manipulation de données numériques conduit à travailler avec des codeurs, des développeurs, des personnes maîtrisant les « méthodes agiles », donc à injecter dans le fonctionnement de l'État les réflexes de l'internet et du web. C'est un puissant moyen de transformation de l'action publique.

C'est pourquoi la mission Etalab a été placée auprès du Premier ministre, avant de s'intégrer en 2012 au secrétariat général pour la modernisation de l'action publique (SGMAP), structure qui s'occupe de l'innovation de services et de la stratégie technologique de l'État.

L'ouverture des données comporte trois grandes dimensions : une dimension de démocratie, une dimension d'innovation économique et sociale et une dimension d'efficacité de l'État lui-même.

S'agissant de la démocratie, il est évident, comme le disait Edwy Plenel, que les informations les plus brûlantes et les plus gênantes ne vont pas se retrouver en ligne dans une démarche d'*open data* bienveillante et naïve. Il a fallu mettre en place une Commission nationale des comptes de campagne et des financements politiques, une Haute Autorité pour la transparence de la vie publique et diverses autres institutions pour construire les points de transparence et de redevabilité aux citoyens. La méthode n'est pas celle de l'ouverture passive. Cela dit, le mouvement de l'*open data* a permis des progrès en matière de transparence.

Outre cet aspect, l'*open data* est facteur de capacitation (*empowerment*). On peut donner aux citoyens des ressources pour qu'ils prennent des décisions plus informées et contribuent de manière plus éclairée au débat public. Nous avons des témoignages, par exemple, de discussions entre associations locales et municipalités au sujet d'aménagements routiers après qu'a été publiée en *open data* la carte géolocalisée des accidents de la route. Il ne s'agit pas seulement d'une transparence visant à « surveiller » un État supposé *a priori* fraudeur et paresseux : il s'agit aussi de livrer des ressources pour régler des problèmes et impliquer les personnes dans la décision publique. Pour certains, l'aboutissement de l'*open data* est l'*open government*, c'est-à-dire une pratique de pouvoir plus ouverte, plus incluante, plus collaborative. La dimension démocratique de l'*open data*, c'est l'addition de la transparence, de la concertation et de la codécision.

Pour ce qui est de la dimension économique et sociale, il est exact que des start-up s'emparent des données et inventent des services innovants. La mission Etalab en connaît déjà plusieurs centaines grâce au concours Dataconnexions. Mais ce sujet a tendance à éclipser d'autres aspects de la création de valeur par l'*open data* : les grands groupes, eux, se gardent de dire qu'ils ont utilisé ces données pour améliorer leurs ABAC (*attribute based access controls*), leurs calculs d'actuariat, la distribution de leurs enseignes dans les zones de chalandise, etc. Nous avons néanmoins des retours discrets. La carte géolocalisée et horodatée de tous les accidents de la route que le ministère de l'intérieur a mis en accès public l'année

dernière intéresse beaucoup de gens qui veulent améliorer les modèles de décision. Pour autant, ce n'est pas un service que l'on trouverait dans un *app store* pour *smartphones*.

En abaissant la barrière à l'entrée – puisque l'on rend gratuites des données qu'il aurait été compliqué d'obtenir ou coûteux de produire –, on autorise aussi des innovations hors marché. Nombre de très belles histoires de l'*open data* sont des créations de services de proximité, culturels, sociaux. Handimap, par exemple, est un navigateur qui calcule les itinéraires pour les personnes en fauteuil roulant dans deux villes de France. Deux ingénieurs l'ont réalisé en cinq mois *pro bono*, pour rendre service. S'ils avaient dû d'abord élaborer un cadastre en trois dimensions, cela n'aurait pas été possible. C'est l'accessibilité de la donnée qui a permis l'innovation hors marché.

On voit aussi des citoyens éclairés qui ont envie d'en avoir le cœur net au sujet des interactions médicamenteuses, des accidents de la route près de chez eux, de la pollution des rivières, etc. Donner de l'autonomie à des citoyens qui ont des décisions à prendre, y compris par le moyen du vote, c'est aussi créer de la valeur.

Au sein de l'État lui-même, nous voyons les jeux de données numériques passer d'une administration à l'autre et donner lieu à des gestes d'innovation forts. C'est le cas du dispositif de marchés publics simplifiés, qui sera déployé prochainement. Il est désormais techniquement possible qu'une entreprise réponde à un appel d'offres public sans envoyer un seul papier à l'administration. Grâce au travail de labour de l'*open data*, nous avons pu puiser dans le *Bulletin officiel des annonces de marchés publics*, dans la base SIRENE (système national d'identification et du répertoire des entreprises et de leurs établissements), dans les registres des tribunaux de commerce, pour vérifier nous-même l'immatriculation des sociétés, l'habilitation des gérants, etc. On ne fait donc plus peser la charge de la preuve sur la partie qui répond à l'appel d'offres.

Plus généralement, notre combat quotidien pour l'efficacité de l'État consiste à casser les « silos ». De l'avis de certains, y compris dans cette commission, c'est de propos délibéré que l'on a construit en France un État très compartimenté. Cette rigidité de l'administration aurait été une condition pour conduire un pays jugé ingérable, ingouvernable, en proie à plusieurs révolutions au XIX^e siècle. Aujourd'hui, on atteint certaines limites. Le ministère chargé du logement, par exemple, doit acheter une fois par an la base de données des prix de l'immobilier pour asseoir sa politique d'aide au logement, alors que cette information existe dans plusieurs endroits dans l'État. De telles bizarreries se comptent par centaines. Ne serait-ce que pour cette raison, il y a beaucoup d'efficacité à gagner en brisant les silos.

Une autre dimension est l'*open innovation* : lorsqu'une administration se met à partager ses données et lorsque l'on s'organise, comme le fait data.gouv.fr, pour étudier les usages possibles et rencontrer les utilisateurs, on rencontre aussi des innovateurs, des gens qui pensent différemment, maîtrisent d'autres technologies, posent d'autres questions aux données... Cela devient une des principales motivations du partage de données : sortir de ses ornières, de ses habitudes, rencontrer d'autres communautés, devenir ainsi plus efficace.

L'*open innovation* s'inscrit dans un mouvement plus vaste de création d'autres formes de relations avec l'extérieur. Etalab s'occupe actuellement de la base de donnée des adresses géolocalisées, la BANO (base d'adresses nationale ouverte). Ces informations peuvent se révéler très importantes pour les SAMU, les pompiers, la police, les télécommunications, les services fiscaux. Quatre bases de ce type existent aujourd'hui dans

l'État, mais aucune n'est ouverte, aucune n'est complète et aucune n'est sans erreurs. Or « déboguer » quatre bases de données qui ont leurs propres standards, leur propre histoire, leur propre architecture, peut être l'affaire de plusieurs années et de plusieurs dizaines de millions d'euros si l'on procède « à l'ancienne ». Mais il se trouve que la puissante communauté *OpenStreetMap*, une sorte de Wikipédia de la cartographie forte de près de 10 000 contributeurs en France – la deuxième communauté au monde, juste après l'Allemagne –, avait engagé ce travail. Etalab a décidé de l'aider à achever la tâche. Trois mois après, et pour moins de 100 000 euros d'investissement public dans des projets que nous avons développés et apportés à la communauté, nous avons déjà refait 80 % des adresses en *open data*, c'est-à-dire à peu près autant que les bases existantes. Nous avons conjugué toutes les informations existantes : 14 millions de ces adresses viennent du cadastre, 2 millions sont « crowdsourcées » par les citoyens, 1,4 million sont issues de politiques d'*open data* de différentes villes. Il s'agit là d'un bien commun qui permet aux autres de « déboguer » leurs bases de données, mais qui ouvre aussi d'autres stratégies similaires dans l'État.

On voit aussi émerger puissamment, en particulier aux États-Unis, des stratégies fondées sur la donnée qui visent non seulement à mieux allouer ses ressources, à trouver des économies intelligentes, mais aussi à mieux organiser ses *process*, à donner plus de pouvoir aux acteurs, etc. C'est ainsi que beaucoup de villes américaines utilisent le logiciel Predpol pour faire passer les patrouilles de police aux heures et aux endroits où la probabilité de crimes et de délits est la plus forte. Là où ce dispositif est mis en place, on constate une baisse de la criminalité de 30 %.

Dans la mesure où il rend les données accessibles et maniables, l'*open data* permet à l'État de s'appliquer à lui-même des stratégies qui n'étaient pas envisageables quand aucun service ne disposait des données du service voisin. Il y a dix jours a été créée la fonction d'administrateur général des données – équivalent de « *chief data officer* » –, confiée elle aussi à Etalab avec pour mandat d'utiliser nous-mêmes nos propres données pour améliorer les politiques publiques. C'est une façon de boucler la boucle : je crois en effet que l'on ne peut mener durablement une politique d'ouverture des données si l'on n'utilise pas soi-même les données et si l'on n'a pas un regard exercé sur leur destination, leur format, etc.

Il me semble donc que notre réflexion doit aussi porter sur la dimension stratégique de l'*open data*. Il y a là une modernité numérique qui emporte une transformation des pratiques de l'action publique.

Pour ce qui est du cadre juridique français, je renvoie au tableau très complet qu'en dresse le Conseil d'État.

Nous avons d'abord la Déclaration des droits de l'homme et du citoyen. Nos ancêtres ont proclamé le 26 août 1789 que « *La société a le droit de demander compte à tout agent public de son administration* », que « *Tous les citoyens ont droit de concourir personnellement, ou par leurs représentants,* » à la formation de la loi, qu'ils « *ont le droit de constater, par eux-mêmes ou par leurs représentants, la nécessité de la contribution publique, de la consentir librement, d'en suivre l'emploi et d'en déterminer la quotité, l'assiette, le recouvrement et la durée.* » Aujourd'hui, tout cela est possible. Les gens savent lire et écrire, ils ont un ordinateur et ils sont connectés. Il faut tenir la promesse de la Déclaration des droits de l'homme et du citoyen !

Personnellement, j'ai découvert sur le tard que la France n'a pas attendu que l'*open data* soit un terme à la mode dans le monde anglo-saxon. Les rapports de la Cour des comptes

sont publics depuis 1850. Le Bureau de statistique générale, ancêtre de l'Institut national de la statistique et des études économiques (INSEE), remonte lui aussi au XIX^e siècle. La loi CADA fonctionne depuis quarante ans. La loi de 1951 a organisé un grand service public de la statistique publique. Mes collègues de l'*Open government partnership* en sont toujours ébahis car la France n'a pas raconté cette histoire sur la scène internationale.

Certes, nous pouvons faire beaucoup mieux, mais nous n'avons nullement à rougir devant nos voisins. Très souvent, nos données sont de bien meilleure qualité que chez eux. Aucun pays au monde, par exemple, ne dispose d'un outil tel que le SNIIRAM (système national d'information inter-régimes de l'assurance maladie), qui rassemble toutes les données des ordonnances remboursées. Aucun système étranger n'ouvre de données aussi précises, ciblées, granulaires et sensibles.

Nous avons une grande tradition statistique et de grands services publics. Certes, il existe des poches de résistance et certains s'exaspèrent de ne pas avoir accès à certaines données. Mais aucun pays ne pourrait nous regarder de haut en faisant valoir son avance sur nous.

Vous me demandez, monsieur le président, s'il faut changer la loi. Je crois que oui. Le cadre posé par la loi CADA est celui du droit individuel d'un citoyen de saisir l'État pour se faire transmettre un document. Comme le souligne le Conseil d'État, il ne contraint pas l'État à anticiper ces requêtes en construisant des référentiels de données complets, cohérents, utiles et créateurs de valeur. Nous sommes donc dans une ambiguïté permanente : nous nous efforçons de fabriquer des infrastructures de données avec pour tout cadre juridique la menace d'un éventuel recours individuel devant la CADA. Les points de vue et les finalités ne sont pas les mêmes. Il faudra donc, selon moi, un cadre positif.

D'ailleurs, peut-on durablement parler de « biens communs informationnels » ou d'« infrastructures critiques » d'une économie naissante ? Lorsque les Américains se félicitent de leur politique d'*open data*, ils commencent par parler du réseau GPS, signifiant par là qu'un système d'informations accessible à tous fait naître une filière industrielle entière, celle des services mobiles, mais – et ils se gardent de le dire – ils en conservent la maîtrise puisqu'ils sont au cœur de la plateforme qui diffuse la donnée. L'effet GPS, nous pourrions l'obtenir en matière d'énergie et de transports ou dans certains domaines de la santé. Le compteur électrique intelligent d'EDF pourrait servir de base arrière à des milliers de services à valeur ajoutée. Peut-on penser cela à travers la réflexion qui a donné naissance à la loi CADA ? Il y a, me semble-t-il, deux dimensions. Ni les principes ni les finalités ne sont tout à fait les mêmes.

Si l'économie du futur doit reposer en partie sur de telles infrastructures, considère-t-on que seules les données mises à disposition par l'État constitueront les biens communs ? La société et le législateur – mais aussi la concertation qui s'engagera au sein du Conseil national du numérique – doivent faire porter leur réflexion sur les données essentielles qui appartiennent à tous, d'où qu'elles viennent, même si elles sont produites par des délégataires de service public ou par des entreprises privées. Leur circulation est à la fois un lien social et une condition du développement économique.

En matière de données personnelles, je considère que la France a tout à gagner à garantir une forte protection de la vie privée. Beaucoup d'entreprises se plaignent de ces garanties, certes, mais beaucoup d'autres affirment que c'est un atout puissant à l'international. Lorsqu'elles lèvent des fonds aux États-Unis, les *start-up* qui réussissent –

Blablacar, Criteo, etc. – sont heureuses de faire valoir qu’elles sont françaises et que la CNIL considère qu’elles font du bon travail.

Cela dit, le débat sur les données et la vie privée est en train de partir en vrille car on mélange tout. Les écoutes illicites ne sont pas l’*open data*, pas plus que la vie privée que tout un chacun met sur Facebook ou encore les *predictive analytics*. La question de savoir si le partage des données publiques fait peser des menaces sur la vie privée doit reposer sur des éléments sérieux et réels. Non qu’il n’y ait aucun risque, mais les administrations connaissent et respectent très bien deux grands cadres : la loi de 1951, qui définit le secret statistique, et la loi « Informatique et libertés » de 1978. À ma connaissance, l’ouverture des données publiques n’a donné lieu à ce jour à aucun accident touchant à la vie privée. Dans aucun pays au monde on ne réquisitionne les données pour les diffuser telles quelles. Ce sont ceux qui les produisent qui en ont la maîtrise. Nous faisons pression sur eux, ils résistent : cette dialectique est une bonne chose !

Il faut néanmoins prendre au sérieux le cas de données si précises et sensibles que l’on serait incapable d’en garantir l’anonymisation irréversible. Il n’y en a pas tant que cela et on voit à peu près où. Probablement dans la santé : pour garantir l’irréversibilité de l’anonymisation du SNIIRAM, il faut assurément prendre la précaution de s’entourer de quelques mathématiciens. On sait par ailleurs faire du floutage statistique sur les accidents de la route, la distribution des revenus, l’impôt sur la fortune. L’INSEE peut faire circuler des données très précises et néanmoins floutées.

En la matière, il n’est pas interdit de faire preuve d’inventivité. Dans le cadre du travail réalisé à la demande de Mme Marisol Touraine sur l’*open data* en santé, nous avons d’ailleurs accepté une ouverture permettant l’accès aux données mais où celles-ci restent sur notre serveur, de telle sorte que les erreurs sont réversibles.

Bref, la protection de la vie privée est une question sérieuse mais, selon moi, moins grave que ce que l’on a dit parfois. Il existe des contrefeux, des contrepouvoirs et des sécurités.

Tout en saluant le travail de la CNIL, je crois que le cadre intellectuel de la loi de 1978 – principe de finalité et principe de proportionnalité – ne peut plus tenir dans un monde de *big data*. On ne peut demander aux gens, avant qu’ils ne constituent ou ne partagent une base de données, de déclarer tout ce qu’ils en tireront : c’est en contradiction avec la manière dont on innove dans le numérique.

S’agissant enfin de l’impact économique, il y a de fortes présomptions. Nous disposons de *case studies* mais pas encore d’études macroéconomiques qui établiraient une corrélation entre l’*open data* et la croissance. On oscille entre ceux qui font la somme des chiffres d’affaires des cinq cents plus grosses *start-up* et McKinsey qui avance le chiffre de 3 000 milliards de dollars, soit deux fois le PIB de l’Allemagne. La vérité doit se situer quelque part entre les deux...

Cependant, dans certains secteurs – information géographique, information météorologique... –, on a des preuves sérieuses que les pays ayant fait de l’*open data* ont une activité économique plus intense et qui croît plus vite que les autres.

Reste l’éternelle question des administrations qui s’inquiètent du coût de la démarche. Là aussi, il faudra sans doute revenir sur la distinction entre la loi CADA et les

biens communs informationnels. Cette loi est très belle dans son esprit et il faut lui rester fidèle. Mais, lorsque l'on a construit progressivement le droit de réutilisation des informations contenues dans les documents administratifs, on a introduit l'idée que ces données avaient été rassemblées pour les besoins de la mission de service public et financées par elle. D'une certaine manière, elles étaient déjà amorties. Aussi ne faut-il pas les vendre. La loi concède que l'on peut éventuellement facturer le coût marginal, mais cela remonte à l'époque où l'on utilisait des dizaines de photocopieuses. Dans l'esprit, les choses étaient déjà claires : si les citoyens savent donner une deuxième vie et une deuxième valeur à ces données, tant mieux, mais ne nous mettons pas à financer le service public par la vente de produits dérivés !

Il y a là une véritable question. Peut-on ouvrir gratuitement l'information géographique, au risque de voir disparaître l'opérateur national ?

Je pense que l'on peut poser une première borne en indiquant que, en dehors de la sphère du service public administratif, il est légitime de rechercher un modèle économique pour assurer le financement durable des informations.

Mais il faut aussitôt poser une deuxième borne : essayer de vendre de la donnée brute est une manière stupide de rechercher un *business model*. Cela ne marche pas, cela freine l'innovation, cela empêche de révéler la valeur de la donnée, bref, cela revient à vendre son blé en herbe.

M. le coprésident Christian Paul. Il nous faut donc réfléchir à la nature de l'obligation susceptible d'être introduite dans le droit français pour aller un peu plus loin que la loi CADA.

M. Henri Verdier. En Grèce, depuis quelques semaines, un acte administratif n'est pas valide tant qu'il n'est pas en ligne. Un fonctionnaire n'est pas remboursé de ses notes de frais si elles ne sont pas en ligne. La formule est radicale mais intéressante !

M. Mohammed Adnène Trojette, conseiller référendaire à la Cour des comptes. Je suis heureux de pouvoir vous apporter un éclairage sur un sujet qui m'a mobilisé pendant plusieurs semaines.

L'ouverture des données publiques est un sujet sensible. Elle ne permet pas seulement d'apporter des réponses au citoyen : elle lui permet aussi de poser de nouvelles questions. C'est sans doute la raison principale des craintes qu'éprouvent certains.

Le sujet est également sensible en ce qui concerne la capacité de modernisation de nos administrations et l'émergence de filières industrielles d'avenir en France ou en Europe.

J'ai eu à connaître de questions liées à l'informatique et au numérique depuis plusieurs années, aussi bien dans mon cursus universitaire qu'en tant que magistrat à la Cour des comptes ou dans le cadre d'activités associatives. Je précise, du reste, que je m'exprime ici en mon nom propre et que mes propos n'engagent en rien la juridiction à laquelle j'appartiens.

Je précise aussi que mon travail a été rendu à l'automne 2013 et que des éléments d'actualisation doivent lui être apportés. Je me propose, dans un premier temps, d'en faire une rapide synthèse.

Ce rapport demandé par le Premier ministre portait sur les cas où les informations publiques reçues ou produites par des services publics administratifs sont mises à la disposition du public dans le cadre de la loi CADA de 1978 sous forme commercialisée. Mon travail intervenait dans un contexte où le Gouvernement souhaitait donner un second souffle à la politique d'ouverture des données publiques, à un moment où l'on pouvait légitimement s'interroger sur les raisons des résistances opposées par les administrations à l'ouverture de certains jeux de données publiques commercialisés.

Ma première mission était d'apporter des clarifications factuelles et dûment étayées sur plusieurs points. D'abord, quelles sont les motivations des administrations pour commercialiser les données ou pour mettre en place ou maintenir une redevance de réutilisation des données publiques, quand bien même certaines de ces administrations mettent à disposition des jeux de données à titre gratuit ?

Je devais également estimer le montant total des redevances perçues par l'administration et m'interroger sur la nature et la pertinence des modèles économiques retenus dans le cadre de cette commercialisation.

Il m'était aussi demandé de m'intéresser à l'impact socioéconomique de l'ouverture des données publiques, en m'appuyant sur des comparaisons internationales.

Je devais enfin explorer les perspectives d'évolution des modèles de mise à disposition des données publiques.

D'un point de vue méthodologique, je me suis appuyé sur un échange direct avec les administrations concernées en leur adressant un questionnaire et en les auditionnant systématiquement, de sorte que j'ai pu établir, pour chaque administration produisant ou recevant des données publiques, une fiche de synthèse qui rappelle le ou les jeux de données commercialisés, les montants ainsi perçus et d'éventuels éléments de précision concernant les coûts de production et de collecte. Ces fiches sont consignées dans la deuxième partie de mon rapport.

J'ai également adressé un questionnaire aux missions économiques de la France dans trente-six pays. Je tiens à la disposition de votre commission les réponses qui m'ont été faites et que la direction générale du Trésor a compilées.

J'ai enfin auditionné de nombreux experts issus d'administrations françaises et européennes, de la société civile, du monde universitaire et du tissu économique.

J'ai tiré cinq grands enseignements de ce travail.

Premièrement, les motivations invoquées par les administrations et les opérateurs pour mettre en place ou maintenir une redevance de réutilisation des données publiques ne sont pas toujours convaincantes.

Deuxièmement, les recettes perçues en 2012 étaient, non pas de 100 millions d'euros et en hausse comme certains – y compris au sein de l'administration – le pensaient, mais de 35 millions d'euros et en baisse d'un tiers par rapport à 2010. Précisons que des acteurs publics sont eux-mêmes acheteurs pour un montant de 4,5 millions d'euros. Le revenu de 35 millions en 2012 a été perçu essentiellement par deux services publics administratifs, l'INSEE et l'Institut national de l'information géographique et forestière (IGN).

Troisièmement, les modèles économiques retenus par les administrations, outre qu'ils portent presque toujours atteinte au principe de gratuité affirmé et réaffirmé par les plus hautes autorités de l'État, se comportent souvent comme des barrières à l'entrée pour les nouveaux acteurs économiques, en particulier les moins bien dotés en capital : citoyens, associations, *start-up*.

Quatrièmement, tout porte à croire que la politique d'ouverture des données publiques est source de très importants bénéfices pour la société dans son ensemble. Ces bénéfices peuvent se chiffrer en milliards d'euros d'effets directs et indirects, notamment en termes de bien-être social, de création de valeur et d'emploi, de gisements de productivité au sein des administrations et des entreprises, notamment du fait des externalités ainsi favorisées.

Cinquièmement, la mise à disposition des données publiques, en particulier des données brutes, doit se faire selon un modèle dit de « plateforme », attirant les utilisateurs les plus innovants et favorisant l'apparition d'écosystèmes autour de ces plateformes.

J'ai constaté que les motivations des administrations qui commercialisaient des données publiques étaient dans certains cas particulièrement fragiles.

La première à être mise en avant est le besoin budgétaire. Il y a quelques années, plusieurs opérateurs ont été amenés, que ce soit dans le cadre de leur contrat d'objectifs et de moyens ou dans le cadre de mandats liés à la révision générale des politiques publiques, à dégager des ressources propres pour se financer et permettre à l'État de limiter la subvention pour charge de service public qu'il leur allouait. La commercialisation de données publiques a constitué pour certains opérateurs une réponse.

D'autres motivations sont bien plus fragiles. Certaines administrations ont justifié leurs redevances de réutilisation par le fait que, en l'absence de telles redevances, elles ne pourraient pas continuer à recevoir, collecter ou produire ces informations publiques. Or, à mon sens, un opérateur ou un service public administratif ne peut s'exonérer de la réalisation de sa mission de service public. L'État a l'obligation de mener à bien les missions qui lui sont confiées dans le cadre des lois votées par le Parlement et des règlements adoptés par le Gouvernement. En particulier, la loi organique relative aux lois de finances (LOLF) n'autorise pas de couvrir des besoins inhérents à l'exercice de missions de service public par des recettes qui ne sont pas votées régulièrement par le Parlement.

Une autre motivation me semble également choquante : alors que les gouvernements successifs ne cessent d'affirmer le principe de gratuité de l'usage des données publiques afin que cette réutilisation soit la plus massive possible, certaines administrations ont décidé de mettre en place ou de maintenir une redevance de réutilisation afin de réguler la demande de données publiques, par crainte des questions pouvant être posées mais aussi, de façon plus compréhensible, par crainte de ne pouvoir faire face aux demandes supplémentaires, dans une période où les moyens humains et financiers sont limités.

En outre, beaucoup d'administrations vendent des jeux de données comme on vendrait des légumes au marché : les données seront beaucoup plus chères au « kilo » si on en achète de petites quantités. Cela a pour effet immédiat de réserver l'achat de l'ensemble de la base de données aux acteurs les mieux dotés en capital et de créer une barrière à l'entrée pour les plus petits acteurs.

J'attire l'attention de votre commission sur les risques qu'il y a, pour le décideur public, de maintenir durablement ces redevances de réutilisation suivant des modèles économiques sous-optimaux. Je pense en particulier au risque d'approche patrimoniale du jeu de données publiques et de la ressource qui est tirée de sa commercialisation, observable dans certaines administrations et dans certains ministères. C'est une conséquence de l'interprétation que l'on a pu faire du rapport de MM. Maurice Lévy et Jean Pierre Jouyet sur le patrimoine immatériel de l'État. Les producteurs de données ont été incités à trouver un moyen de rentabiliser leurs bases, si bien qu'aujourd'hui certaines administrations font obstacle à une approche interministérielle de la gestion des données publiques et des ressources qui peuvent en être tirées. La création d'un administrateur général des données publiques devrait néanmoins permettre de dépasser cette approche patrimoniale et ministérielle.

Il arrive parfois que l'administration s'appuie sur des dispositions législatives ou réglementaires pour justifier son inertie. De ce point de vue, il faudra apporter un soin particulier à la transposition de la directive du 26 juin 2013 modifiant la directive, dite « PSI », de 2003 relative à la réutilisation des informations du secteur public. Je pense également à l'article 15 de la loi CADA, qui autorise les administrations à commercialiser leurs jeux de données publiques et à tenir compte des coûts de collecte et de production, alors même que la LOLF, je l'ai dit, interdit le financement de cette activité par une recette qui ne serait pas votée régulièrement par le Parlement. L'ambiguïté de ces dispositions peut amener l'administration à camper sur ses positions.

Le même type de difficulté a pu apparaître s'agissant des données personnelles, comme Henri Verdier vient de l'indiquer.

Je veux enfin souligner le risque que constitue une approche défensive et protectionniste de la part des administrations en matière de mise à disposition des données publiques. Une des motivations des administrations est aussi d'éviter que certains grands groupes internationaux, dont plusieurs ne paient pas leurs impôts en France, accaparent les jeux de données publiques de notre pays, notamment les jeux de données publiques souverains. Mais, en réalité, le montant demandé pour les données publiques est dérisoire au regard de la force de frappe de ces groupes, qui seront toujours en mesure de payer le prix exigé, ou, si ce prix devient réellement très élevé, de payer leurs employés ou de motiver leurs utilisateurs les plus fidèles pour produire des jeux de données concurrents, ce qui risque de limiter la réutilisation des données du service public.

Les conclusions de mon rapport ont été reprises dans le cadre de décisions du Gouvernement. Après le comité interministériel pour la modernisation de l'action publique du 18 décembre 2013, a notamment été décidée la suppression de plusieurs redevances de réutilisation de données publiques. Cela a été l'occasion pour le Gouvernement de clarifier sa doctrine en matière de redevances, qui ont toutes vocation à disparaître, et de confier au SGMAP une mission d'accompagnement des opérateurs dans leur transition vers de nouveaux modèles économiques de mise à disposition de données publiques. Pour certains opérateurs, en effet, il est vital de trouver des recettes se substituant à celles qu'ils tirent de la commercialisation de données.

M. le coprésident Christian Paul. Je propose que la première série de questions porte sur le volet économique de notre réflexion.

M. Godefroy Beauvallet. Vous trouvez légitime, monsieur Verdier, que l'on s'attache à la valorisation des données et que l'on cherche un *business model*. Sachant que ce modèle ne peut pas reposer sur la vente de données brutes, que préconiseriez-vous ?

Il faut prendre en compte la compétition entre les plateformes privées, notamment les plus puissantes, et la donnée publique. Rendre publique une donnée, c'est rendre la tâche plus difficile aux plateformes qui cherchent à détenir un monopole sur tel ou tel aspect de l'identité ou de l'activité – cet effet de capture étant, on le sait, une des dérives de leur économie. En matière de production scientifique, par exemple, des effets de capture ont lieu faute de mise à disposition publique simple.

En somme, quels seraient des *business model* acceptables qui ne se résumeraient pas à de la capture plus intelligente ? Cette question doit être traitée en lien avec celle des API (*application programming interfaces*). On peut en effet imaginer un monde dans lequel les services publics ne mettraient pas à disposition leurs données mais des API permettant de faire des calculs sur ces données.

M. le coprésident Christian Paul. La question du modèle économique se pose pour les entités publiques qui produisent les données, mais aussi pour d'autres entités, publiques ou privées, qui ne sont pas productrices de données brutes mais dont l'activité naît d'innovations dans l'usage de ces données. Y a-t-il un partage de la valeur ou cette idée est-elle un peu taboue ?

Par ailleurs, l'environnement juridique existant constitue-t-il un frein ? Est-il au contraire un atout pour avancer dans de bonnes conditions ? Faut-il le faire évoluer ?

M. Philippe Aigrain. Si je suis entièrement d'accord avec ce qui vient d'être dit, j'ai tout de même un doute. À vous entendre, on a l'impression que tout le monde a pris conscience des coûts de transaction, de la faible valeur des contenus bruts et dès lors, de la non-validité des motivations budgétaires. Or, dans la pratique, on constate que des organismes publics tentent encore de se transformer en acteurs ayant vocation à valoriser directement les données – pour deux raisons que vous n'avez pas évoquées, à savoir les partenariats public-privé et les investissements d'avenir. J'ai pris part à des travaux dans le domaine de la valorisation de la recherche à financement public, qui ont mis en évidence que lorsqu'un cofinancement est mis en place, la plupart du temps ce ne sont pas les producteurs des résultats de recherches qui les exploitent – soit qu'ils n'y aient pas intérêt, soit qu'ils n'aient pas l'idée de le faire ou ne soient pas dotés d'une structure adaptée. Or le pouvoir d'exploiter les données, accordé à un cofinancier privé se révèle malgré tout assez rentable. Je me demande donc si cette question fait partie de vos réflexions. Pouvez nous suggérer une stratégie visant à éviter que cette activité, qui représente 40 milliards d'euros et va donc bien au-delà des petits ajustements budgétaires, échappe ainsi à la sphère publique. Cette question est loin d'être anecdotique, car nous sommes en présence d'un cheval de Troie capable de mettre à mal la politique des données ouvertes telle que vous la présentez.

M. Henri Verdier. Il s'agit là d'un problème complexe car, pour le moment, l'ouverture des données publiques s'ancre dans la loi CADA, et la seule chose qui existe dans ce cadre, c'est la vente par le service public administratif de sous-produits de son activité de service public. Un décret du Premier ministre affirme que seule une autorisation du Premier ministre peut autoriser la création d'une nouvelle redevance, ce que le Secrétariat général du Gouvernement sait parfaitement. Aucune n'a été créée depuis trois ans et on assiste actuellement à leur extinction progressive, ce qui n'a rien d'étonnant quand on sait que

certaines d'entre elles n'avaient aucune justification – je pense par exemple à la base de données des adresses des lycées de France, dont la vente ne présentait aucun intérêt pour le service public.

M. Philippe Aigrain. Prenons tout de même un exemple concret : lorsque la Bibliothèque nationale de France conclut un partenariat public-privé d'exploitation de documents relevant du domaine public et négocie d'obtenir 20 % du chiffre d'affaires de cette exploitation – par exemple, la vente sous forme d'EPUB de ce qui est fourni gratuitement par le biais du projet Gutenberg ou d'Internet Archive –, la CADA n'a pas vocation à s'appliquer. En revanche, nous souhaiterions vivement qu'elle puisse s'appliquer au contenu des accords négociés avec les acteurs privés, afin qu'au moins le cheval de Troie ne soit plus un secret pour le public.

M. Henri Verdier. C'est là, sans doute, l'une des questions que devra se poser la commission au moment du débat sur la loi numérique. Aujourd'hui, le seul instrument juridique dont nous disposons est la loi CADA, prévoyant les conditions d'accès aux documents administratifs et aux informations qu'ils contiennent, et la propriété intellectuelle constitue l'une des restrictions à l'application de cette loi, qui s'arrête à vie privée et à la propriété intellectuelle, donc aux œuvres. Ainsi dans le cadre de certains partenariats public-privé le contribuable se trouve-t-il dans l'impossibilité de savoir ce qu'il paye, le délégataire de service public lui opposant le secret des affaires, ce qui est assez paradoxal.

M. Philippe Aigrain. En l'occurrence, c'est uniquement le contrat d'exclusivité.

M. Henri Verdier. De toute façon, on ne peut pas activer la CADA dans ce cas.

Aujourd'hui, on réfléchit à la vente de données brutes par certains services publics « moins administratifs », si j'ose dire, tels Météo France ou l'IGN, au regard du principe de gratuité des documents administratifs, mais j'ai tendance à considérer que ce cas de figure n'était pas celui visé par le législateur. On a montré, de manière assez probante, qu'il valait mieux laisser circuler les données pour que les gens s'en emparent et que les innovateurs innover : c'est ainsi que l'on produit le plus de valeur économique et sociale. Cela dit, un tel raisonnement ne résulte pas d'une grande réflexion sur les infrastructures-clés et les biens communs, et plus l'économie va se fonder sur les données – en particulier celles créées par le service public, plus nous aurons de problèmes à résoudre.

Indiquer ce que devrait contenir une future loi n'est pas aisé, la principale difficulté consistant à déterminer la nature des données ayant vocation à être rendues publiques. Aujourd'hui, à partir des cartes de cantine électroniques des fonctionnaires, on est en mesure de déterminer le nombre de calories consommées en moyenne par chaque agent public ; cet exemple est révélateur du fait que les données les plus intéressantes sont souvent celles issues de l'informatique de gestion – donc les plus banales en apparence –, et non de l'informatique de décision. La Direction interministérielle des systèmes d'information et de communication (DISIC) a publié il y a quinze jours un nouveau plan stratégique qui me paraît très prometteur, ayant pour objet l'architecture des systèmes d'information, les interfaces de programmation (*Application Programming Interface*, ou API) et les plates-formes.

Pour la préservation des biens communs, on a le droit de jouer aussi sur les conditions juridiques elles-mêmes. Aujourd'hui, pour le service public administratif, Etalab encourage les licences les plus ouvertes possibles – une décision interministérielle est allée dans ce sens –, ce qui implique une mise à disposition sans conditions des données. La

communauté OpenStreetMap a exprimé la crainte qu'un opérateur géant ne s'approprie les données rassemblées grâce au travail de ses contributeurs, pour ensuite effectuer seul toutes les mises à jour. C'est ce qui est arrivé à l'IGN qui, après avoir vendu ses fonds de cartes pour 14 millions d'euros à Google, s'est entendu dire qu'on se passerait de lui pour les mises à jour, le réseau des téléphones Android étant parfaitement suffisant pour recueillir les données nécessaires – et désormais, les innovateurs délaissent l'IGN pour recourir à Google Maps. Pour éviter cela, nombre de collectivités locales, ainsi que certaines communautés telles OpenStreetMap, ont opté pour des licences *share-alike*, prévoyant que l'utilisateur d'une donnée ait l'obligation soit de remettre le fruit de son travail au pot commun, soit de payer.

Sans aller jusqu'à exposer des *business models*, il me semble qu'il convient de tracer dès maintenant des trajectoires de pérennisation du bien commun par ceux qui, parce qu'ils en tirent le plus de profit, ont intérêt à agir. Prévoir de tels mécanismes est utile dans certains cas – certainement pas dans ceux de la base de données des adresses des lycées de France ou encore de la base des prix de l'essence relevés par la DGCCRF, mais dans les situations où il est nécessaire d'investir durablement dans la mise à jour. J'ai mentionné tout à l'heure les compteurs électriques intelligents, qui ont une grande valeur potentielle, mais qui nécessiteront des investissements considérables pour anonymiser de manière irréversible les données qu'ils produiront – ce qui ne représente pas une nécessité de service public pour EDF.

Il existe plusieurs stratégies permettant de faire payer celui qui tire le plus de profit de l'exploitation des données, qu'il s'agisse de la mise en place de plates-formes, d'un système de *freemium*, ou encore d'API progressives, en ce qu'elles établissent une tarification différente en fonction de la fréquence à laquelle la base de données est sollicitée par l'utilisateur – il est possible, par exemple, de fixer un seuil à cinq appels par heure. Il est difficile d'être plus précis sur ce point, la plupart des cas ayant vocation à être examinés au cas par cas.

Tout ce que je viens d'évoquer est d'ordre technique et ne recouvre pas l'aspect juridique des choses. Comme l'a indiqué le Conseil d'État, il n'y a pas d'obligation légale pour une administration de mettre en œuvre des moyens de mettre à disposition des citoyens – et éventuellement de valoriser – les données se trouvant en sa possession. Quant à la loi sur les documents administratifs, elle se rapporte uniquement aux données de l'informatique de décision, et non à celles de l'informatique de gestion, qui ont pourtant souvent plus de valeur.

Sur ce point, je conclurai par une anecdote. Quand la Ville de Paris a voulu mettre en place son portail *open data*, elle a souhaité qu'y figurent tous les secteurs d'activité de la ville, y compris celui de la culture. Les responsables de ce secteur ont estimé qu'il était impossible de partager la moindre donnée, partant du principe que les œuvres étaient protégées par la propriété intellectuelle. Cela a duré un an ou deux, jusqu'à ce que quelqu'un s'avise du fait que les données recueillies par le biais des cartes de bibliothèque électroniques avaient en réalité une grande valeur en termes de sociologie culturelle. Comme vous le voyez, il ne serait pas facile de fixer dans une loi quelles sont les données ayant vocation à être partagées.

M. le coprésident Christian Paul. À l'heure où une loi est en préparation, c'est pourtant bien la question qui se pose à nous dans le cadre de la réflexion que nous menons. Nous commençons à cerner quelles sont les données dont la mise à disposition serait difficile ou injustifiée, mais ne peut-on vraiment pas aller plus loin dans la préconisation ?

Mme Corinne Erhel. M. Verdier a évoqué les licences *share-alike*, conçues sur le principe selon lequel l'utilisateur de données a l'obligation soit de remettre le fruit de son

travail au pot commun, soit de payer. Mais en pratique, comment limiter le recours à l'achat des données par des acteurs pour lesquels le prix n'est pas un obstacle ?

M. Henri Verdier. Ce qui nous gêne avec Google, Apple, Facebook et Amazon – les GAFAs –, ce n'est pas qu'ils réussissent à inventer de nouveaux services, mais qu'ils en viennent parfois à bloquer l'innovation après l'avoir fait. Tant que le bien commun est inaliénable et que tous les innovateurs potentiels peuvent s'en servir, il n'y a pas de problème : ce qui est anormal, c'est qu'un opérateur en situation de monopole puisse tout bloquer après s'être emparé des données.

Pour ce qui est du contenu de la future loi, je suis tenté de vous rappeler que c'est à vous, législateur, qu'il revient de le déterminer. Cependant, l'idée d'un *open data* par défaut, consistant à inverser la charge de la preuve – il faudrait tout mettre à disposition par défaut, et justifier du refus de le faire pour telle ou telle donnée – est intéressante. La semaine dernière, François Hollande a rejoint le sommet de l'*Open Government Partnership* (OGP) à l'issue de la réunion du Conseil de Sécurité de l'ONU – et hier, les pays membres de l'OGP se sont à nouveau retrouvés à Paris. Ces réunions ont été l'occasion de constater qu'à l'heure actuelle, la plupart des pays se tournent vers l'*open data* par défaut – ainsi les Irlandais et les Grecs viennent-ils de l'adopter.

Dans un tel système, tout est gratuit par principe, et c'est seulement quand il devient nécessaire de financer durablement l'infrastructure de mise à disposition d'un bien commun mobile que l'instauration d'une redevance se justifie. Certaines des redevances mises en place sont inacceptables en ce qu'elles servent uniquement à préserver un tout petit écosystème de réutilisateurs vivant de la rente qu'ils se sont constituée, à limiter – par la dissuasion – le nombre d'utilisateurs s'adressant à l'administration, ou encore à financer le site internet d'un ministère, par l'effet d'une mauvaise habitude dont on peine à se défaire.

Quand on dit que la CADA n'a pas assez de pouvoirs, cela ne signifie pas qu'elle n'a pas assez d'argent. Je rappelle ce qui a été dit à ce sujet lors de l'audition de son président, Serge Daël. Un citoyen désirant obtenir des données relatives à la réserve parlementaire pour l'exercice 2012 a dû, après s'être vu refuser l'accès à ces données par l'administration – la décision de refus n'étant acquise qu'au bout de deux mois –, saisir la CADA pour recueillir son avis avant de porter l'affaire devant le tribunal administratif qui, deux ans plus tard, lui a donné raison. Le problème, c'est que la procédure ne fait pas jurisprudence : pour obtenir les mêmes renseignements pour l'exercice 2013, il faudrait tout recommencer ! Il n'est pas concevable de maintenir un système aussi rigide, d'où la proposition de Serge Daël, formulée en son nom propre, d'offrir la possibilité de déposer un référé communication devant le juge administratif, dans le cas où l'administration refuse de communiquer des documents malgré un avis favorable de la CADA. Cette idée est très intéressante, de même que celle consistant à ce que les décisions rendues pour un exercice donné valent aussi pour les suivants.

Par ailleurs, certains pays tels que les Pays-Bas ou le Danemark ont défini un socle de données dites essentielles – correspondant à sept ou huit registres-clés, contenant des données se trouvant au cœur de tout, telles les adresses géolocalisées ou les codes SIREN des entreprises –, dont ils ont sanctuarisé le financement, considérant que cela faisait partie du service public de mettre les données en question à disposition de tous. C'est également une piste à retenir.

M. Mohammed Adnène Trojette. Sur la question des bénéficiaires et du modèle économique à mettre en place, il me semble que nous devons garder trois choses à l'esprit.

Premièrement, il est apparu dans certains pays que l'ouverture des données publiques a permis au secteur concerné de faire progresser son chiffre d'affaires de manière significative – on parle de taux à deux chiffres – par rapport à d'autres pays où cette ouverture n'a pas eu lieu. Le cas du Royaume-Uni a été examiné en détail dans le cadre d'un rapport indépendant remis au Premier ministre britannique, qui mettait en évidence un bénéfice de huit milliards d'euros, dont deux milliards d'euros d'effets directs sur la société.

Deuxièmement, le montant de 35 millions d'euros de redevances concerne une vingtaine d'organismes, dont deux concentrent à eux seuls environ 20 millions d'euros.

Troisièmement, nous sommes dans un contexte de concurrence entre les plates-formes publiques et privées. Plus les utilisateurs de données vont vers les plates-formes privées, moins ils vont vers les publiques, dont les données ont tendance à se périmérer rapidement. Dès lors, il importe d'établir des priorités, et de se demander quelles sont les urgences : à mon sens, la première est de faire en sorte que les utilisateurs les plus innovants continuent d'utiliser les données publiques, afin que les missions de service public continuent d'être exercées dans de bonnes conditions. Le premier principe qui doit nous guider dans le choix d'un modèle économique consiste donc à écarter les modèles portant atteinte au principe de gratuité affirmé depuis plusieurs années par les gouvernements successifs.

Le deuxième principe à poser en urgence, dans l'hypothèse de l'*open data* par défaut justifiant une commercialisation exceptionnelle pour certaines données, c'est que la tarification ne doit pas entraîner une impossibilité totale d'accéder gratuitement à la base de données brutes. L'idée – correspondant à celle de la redevance retenue dans le cadre de la LOLF – est que la tarification se justifie par la mise en place de services supplémentaires consistant en un traitement spécial administré au jeu de données, ou dans l'autorisation d'utiliser avec une intensité particulière les ressources obtenues de l'administration – il peut s'agir du nombre d'accès concurrents sur les bases de données, de la bande passante mobilisée, de la fréquence des mises à jour auxquelles il est procédé, ou encore de la mise en place d'une sorte d'embargo aux termes duquel les utilisateurs payants disposeraient des données quelques jours avant les autres, comme cela a été le cas avec la base de données des carburants mise à disposition sur data.gouv.fr par le ministère de l'économie et des finances.

En tout état de cause, certains modèles de tarification sont à proscrire. Je pense en particulier aux modèles basés sur le volume de données téléchargées, qui empêchent les acteurs incapables de payer d'accéder à l'ensemble de la base de données – alors que seul un accès total a une réelle valeur. Les modèles de tarification dégressive sont également à écarter, ainsi que ceux limitant la gratuité dans le temps et ceux basés, sans aucune nuance, sur l'existence d'une activité commerciale – si l'administration refuse parfois de mettre gratuitement des données à disposition de certaines sociétés, s'attendant à ce qu'elles en fassent une utilisation économique, une telle position ne peut se justifier dans la mesure où l'un des intérêts de la mise à disposition massive de jeux de données est précisément de permettre l'exercice d'activités économiques, *a fortiori* lorsque ces activités sont localisées en France et en Europe, où elles créent de l'emploi et améliorent le bien-être économique et social.

En ce qui concerne l'environnement juridique, je précise qu'à mon sens, l'article 15 de la loi CADA pose un problème dans la mesure où il introduit de l'ambiguïté par rapport au principe de gratuité et entretient une confusion quant aux motivations autorisées pour la mise en place et le maintien d'une redevance de réutilisation – le coût de collecte et de production des données ne pouvant, en tout état de cause, être couvert que par le budget de l'État. De

plus, cet article met en place une véritable usine à gaz destinée à vérifier que le montant total des redevances perçues ne dépasse pas un certain plafond, dont la détermination donne lieu à des calculs très compliqués.

M. Henri Verdier. De ce point de vue, il est certain que la directive dite PSI – *Public Sector Information* – nous simplifie la vie en introduisant la notion de fraction significative des revenus de l’opérateur et en posant quasiment l’exigence de tenir une comptabilité analytique détaillée permettant de justifier de la redevance.

M. Daniel Le Métayer. La notion d’*open data* par défaut me fait penser à celle de *privacy by default* utilisée au sujet de la protection de la vie privée, un domaine où la protection par conception – *privacy by design* – tend à s’imposer de plus en plus. J’aimerais savoir si la notion d’*open data* par conception aurait du sens – en d’autres termes, si l’on peut imaginer d’intégrer d’emblée aux futurs appels d’offres cette exigence, qui impliquerait de faire dès le départ les choix techniques, notamment en termes de format de données, de nature à faciliter la réutilisation ultérieure des données.

M. Mohammed Adnène Trojette. Si la mise en place d’une clause d’ouverture de données publiques par défaut est une question de principe, son adoption éventuelle se traduit par des conséquences techniques. Il est bon pour la productivité des administrations que l’ouverture par défaut soit prise en considération le plus en amont possible, c’est-à-dire dès la conception et la réalisation des systèmes d’information ou des applications informatiques, afin que la production des données, leur ouverture potentielle et les échanges avec le monde extérieur – par exemple au moyen d’interfaces de programmation (API) – soient gérés dans les meilleures conditions. La même problématique se pose au sujet de la protection des données personnelles et en particulier de la nécessité d’anonymiser ces données avant de les rendre publiques.

M. Henri Verdier. Poser cette question, c’est déjà un peu y répondre. Vous seriez effrayés si je vous disais combien de systèmes tournent sur des noyaux vieux de trente ans ; or, l’un des obstacles à l’ouverture des données réside souvent dans les difficultés que pose leur mise en œuvre sur des systèmes qui n’ont pas été conçus pour cela. C’est un fait, l’*open data* passe par le renouvellement de ces systèmes, qui prendra peut-être une dizaine d’années. La DISIC s’est saisie du problème, et le cadre d’architecture technologique prépare les conditions de l’interopérabilité et de la circulation des données. Par ailleurs, il reviendra à l’administrateur des données publiques de se prononcer sur la gouvernance de la donnée, et de produire un rapport annuel sur l’intelligence de sa circulation. Mais, je le répète, restructurer le système actuel pour le rendre simple, fluide, efficace et sûr prendra une dizaine d’années.

M. Godefroy Beauvallet. En matière de protection des données personnelles, il existe, me semble-t-il, une contradiction relative à la pertinence des modèles de données. Si la publicité se fait bien au moment de la conception, il faut qu’il en soit de même *ex post*, afin de permettre l’exercice d’un droit d’interpellation des administrations concernées – pas seulement individuel, auprès de la CADA, mais portant sur la structure même de la base. En d’autres termes, les usagers informés de l’existence de tel ou tel type de données doivent avoir la possibilité de réclamer leur ouverture ; de même, l’administrateur général des données et les juridictions de contrôle doivent pouvoir demander un audit. Ces trois étapes – publicité du modèle, droit d’interpellation, devoir d’audit – me semblent indispensables à la solution technique du problème, qu’une simple loi ne suffira pas à régler.

M. Henri Verdier. En préalable à toutes les questions qui viennent d'être évoquées, il me semble qu'il convient de mener une réflexion préalable sur les missions de service public. J'ai déjà fait mention de l'avis donné par la CADA, en décembre dernier, sur l'accès au système national d'information inter-régimes de l'assurance maladie (SNIIRAM), qui contient tous les remboursements d'ordonnances médicales. Certains citoyens s'étant demandé si la CNAM n'aurait pas eu là les moyens de détecter une surprescription du Mediator, la CADA a répondu que, s'agissant d'un traitement d'usage courant, relevant de l'informatique légère, il était légitime que les citoyens obtiennent une réponse à leur question.

Ce qui me fascine dans cette décision, c'est qu'elle semble ouvrir d'énormes perspectives en termes de gisements de données, dans la mesure où l'administration ne peut plus se retrancher derrière le fait que des données ne sont pas immédiatement disponibles : il suffit qu'un traitement minime de l'information les rende accessibles pour qu'elle ne puisse plus refuser de les communiquer – en l'occurrence, la CNAM a dû s'exécuter, étant précisé que cette affaire pourrait bien avoir des suites pénales. Il me paraît bon que le principe de partage du savoir se trouve en quelque sorte intégré aux missions de service public, plutôt que de ne sembler concerner que quelques passionnés qui s'efforcent de faire respecter ce principe, parfois au risque de sanctions hiérarchiques.

Mme Corinne Erhel. J'en reviens à l'idée du socle de données essentielles adoptée par certains pays, dont il a été question tout à l'heure. Pouvez-vous nous indiquer comment s'est faite cette évolution – de manière concertée ou autoritaire – et s'il existe un corpus européen en la matière ?

M. Henri Verdier. Pour ce que j'en sais – il faut rester prudent, car les choses changent très vite dans ce domaine –, seuls deux pays européens, à savoir les Pays-Bas et le Danemark, se sont engagés à fond dans cette démarche, accomplie de façon plutôt consensuelle. Il faut bien comprendre que toutes les données ne sont pas de même nature et que, même si cette distinction peut donner lieu à des controverses, on peut les classer selon qu'elles sont plus ou moins brutes : ainsi, un PIB ou un taux de croissance sont des données résultant du travail de statisticiens, tandis que des immatriculations d'entreprises au registre du commerce, des codes postaux ou le nombre d'habitants d'une commune sont considérés comme des données brutes. Certaines données telles que les codes postaux sont considérées comme centrales dans la mesure où elles permettent de fusionner un grand nombre de fichiers, et c'est sur ce type de données qu'un consensus s'est dégagé quant à la nécessité de les rendre accessibles et gratuites.

M. le coprésident Christian Paul. Nous allons arriver au terme de notre réunion, et j'aimerais que le restant de nos échanges soit consacré à un sujet que nous avons pour le moment peu évoqué, celui de la protection de la vie privée.

M. Philippe Aigrain. Je veux tout de même préciser que nous ne devons pas craindre de faire figurer dans la loi l'obligation incombant aux acteurs du secteur privé de contribuer à la mise en commun de données que je qualifierai de sociales ou sociétales. Il ne faut pas perdre de vue le fait que notre appareil statistique est entièrement né d'obligations de contribution mises à la charge des entreprises – à leurs frais –, qu'il s'agisse de données comptables alimentant les comptes de la Nation ou de données sociales, lesdites obligations ayant, à l'origine, été justifiées par la nécessité pour l'État d'être en mesure d'accomplir ses missions régaliennes en matière d'impôt et de mettre en place des politiques publiques justes. Aujourd'hui, ces contributions ne se font plus seulement à destination de l'État, mais aussi

des citoyens, et j'estime que les données recueillies au moyen des compteurs électriques intelligents devraient y donner lieu, au moins en partie.

Pour ce qui est de la protection des données, j'ai coordonné le projet relatif à la gestion de l'information personnelle de santé, ce qui me permet d'affirmer que la question de la sensibilité des données dépasse très largement le cadre d'une base de données, la plupart des informations réellement sensibles résultant du croisement entre plusieurs bases. Ainsi les bases d'adresses géolocalisées ne permettent-elles pas de déterminer, par elles-mêmes, qui habite à telle ou telle adresse : ce n'est qu'en recoupant ces bases avec d'autres fichiers que l'on peut établir des correspondances et obtenir des informations sensibles, du type de celles ayant servi à établir les listes de la rafle du Vel' d'Hiv.

M. Henri Verdier. Je perçois un sophisme dans le raisonnement consistant à affirmer qu'une donnée publique au sens de la loi CADA peut devenir personnelle si on lui adjoint d'autres informations glanées sur Internet. À mon sens, le problème se situe en amont, dans le fait même qu'une donnée personnelle soit disponible sur le web : il aurait fallu réfléchir avant d'autoriser qu'une telle chose ne soit possible. En raisonnant comme vous le faites, on peut en arriver à des aberrations : ainsi certains ont-ils cru pouvoir affirmer que la mise à disposition de la carte des arbres remarquables par communes constituait une violation de la vie privée, en ce qu'elle permettait de savoir quel arbre se trouvait dans le jardin de telle ou telle personne. Si l'on veut protéger la vie privée, ce n'est pas en restreignant la circulation des données publiques qu'il faut le faire, mais en s'interrogeant sur celle des données personnelles !

M. Philippe Aigrain. J'entends bien, mais je pense que si nous voulons convaincre de l'innocuité de l'ouverture des données publiques, nous devons prendre en compte la réalité du problème que je soulève. Le terme même de « données » recouvre des situations extrêmement hétérogènes, y compris en matière de protection : ainsi appelle-t-on parfois « données » des choses qui n'en sont pas, telles que des communications ou des documents couverts par la propriété intellectuelle. De même, les données brutes que vous évoquiez tout à l'heure peuvent être considérées comme des informations n'ayant pas encore subi le calibrage qui va en faire de véritables données. Quand on demande à connaître le texte de l'accord de partenariat public-privé entre la BNF et Believe Digital, ce n'est pas à une donnée que l'on souhaite accéder, mais à un document – d'ailleurs, le « D » de la CADA signifie bien « documents », et non « données ».

Fixer un cadre conceptuel rigoureux dans le cadre de la future loi sera d'une grande utilité pour la suite. Je comprends bien qu'il ne faille pas se focaliser sur un scénario catastrophe si l'on veut avancer, mais certains événements se sont déjà produits – je pense notamment aux bases de plaques d'immatriculation rendues publiques aux États-Unis –, qui montrent que le risque existe.

M. Henri Verdier. Je voudrais tout de même souligner deux points. Premièrement, notre administration est pétrie d'une culture marquée par l'instauration du secret statistique en 1951 et par la création de la CNIL en 1978. Deuxièmement, le partage de ces outils statistiques que sont les données publiques brutes est en fait très rare – personnellement, je n'ai jamais eu l'occasion d'observer un flux de données brutes émis dans l'exercice d'une mission de service public. Certes, nous devons être prudents – je n'irai pas jusqu'à dire que le danger est inexistant –, mais tout le monde crie au loup alors qu'on n'en a jamais vu la queue !

J'insiste également sur le fait qu'il ressort de la jurisprudence de la CNIL et de la CADA que toutes les données à caractère personnel ne relèvent pas forcément de la vie privée : il n'est donc pas justifié d'y voir un tabou. Oui, vous pouvez trouver en ligne le salaire des fonctionnaires – une décision a été rendue en ce sens –, ou aller consulter au centre des impôts le montant de la contribution acquittée par votre voisin, en vertu d'une décision datant de la Révolution française, fondée sur le principe selon lequel chacun doit pouvoir vérifier qu'il paye un impôt juste.

M. Philippe Aigrain. Je suis d'accord pour considérer que nous ne devons pas nous laisser brimer par une vision trop restrictive : mon intervention se voulait constructive et visait à souligner que nous ne devons négliger aucun des aspects du problème dont nous débattons.

M. le coprésident Christian Paul. Nous ne pouvons pas nier le fait que, si elle est parfois invoquée comme un prétexte à certaines résistances, la protection de la vie privée est aussi une vraie nécessité, reconnue par les partisans de l'*open data* eux-mêmes : il ne faut donc ni surprotéger, ni sous-protéger. Ainsi, en ce qui concerne les bases SNIIRAM de l'assurance maladie, qui constituent l'un des gros enjeux en matière d'*open data*, il est certain que si des incidents survenaient dans les années qui viennent, on oublierait les immenses services rendus par cette base en termes de prévention et de pharmacovigilance pour se focaliser sur les conséquences dommageables d'une anonymisation réalisée de manière imparfaite.

M. Henri Verdier. Il ressort des travaux d'une commission ayant réuni plus de quarante participants provenant de tous les secteurs du monde de la santé qu'il n'est pas envisageable d'ouvrir les bases SNIIRAM sans conditions, une telle opération présentant trop de risques – même si certains partisans radicaux du *big data* ont tenté de faire valoir que les gains que l'on pouvait attendre de l'ouverture des bases étaient supérieurs aux risques encourus. La CNAM sait depuis longtemps faire des *datamarts* – des bases de données relationnelles destinées aux usagers – protégés par des floutages statistiques très résistants, basés sur la méthode de l'échantillonnage. Toute la question est de savoir comment faire pour procéder à ces *datamarts* en grande quantité et à la volée : il est évident que la CNAM n'est pas du tout disposée à affecter à cette tâche les dizaines d'agents qu'elle nécessiterait.

M. Philippe Aigrain. Daniel Le Métayer a évoqué les rapports entre le *privacy by design* – la protection des données par conception – et ce que pourrait être une politique de données ouvertes reposant elle-même sur la conception. Alors que la règle de base du *privacy by design*, c'est la décentralisation, la France est sans doute le pays où les dossiers médicaux de toute nature – dossier médical personnel, dossier pharmaceutique, dossiers d'assurance maladie – sont le plus centralisés, ce qui peut se justifier par le fait que l'hébergement physique et l'exploitation logicielle de données ouvertes sont d'autant moins faciles que ces données sont décentralisées.

M. Henri Verdier. Si les données sont ouvertes, elles circuleront.

M. Daniel Le Métayer. Un mot capital a été prononcé, celui de « risque », au sujet duquel il me semble que nous ne devons pas être dogmatiques. Je ne suis pas aussi optimiste que M. Verdier qui, constatant qu'aucune catastrophe n'est survenue pour le moment, estime que la loi de 1951 suffit à nous protéger. En réalité, la situation a beaucoup évolué : aujourd'hui, on dispose de multiples sources d'informations et de techniques permettant de les recouper, ce qui multiplie les risques. Rien qu'avec les compteurs intelligents, les techniques les plus sophistiquées d'analyse des données recueillies permettent de savoir tout ce qui se

passer au domicile d'une personne, et même connaître la marque de sa machine à laver ou l'heure à laquelle elle se lève.

M. Henri Verdier. Sur ce point précis, je dois vous dire que ceux qui ont essayé de faire ce que vous dites – j'en connais plusieurs, notamment des *start-up* du *big data* – n'ont pas encore réussi, car cela implique de décoder une multitude de signaux : ce sera possible demain, mais ça ne l'est pas aujourd'hui.

M. Daniel Le Métayer. La loi va devoir évoluer. Ainsi, on considère à l'heure actuelle qu'une donnée est soit anonyme, soit personnelle, ce qui n'est pas vraiment le cas d'un point de vue technique. Pour autant, la question de l'anonymisation – et de ses limites – ne peut être éludée. Nous devons travailler dans un esprit d'analyse de risque : pour chaque situation donnée, il nous revient de peser les bénéfices et les risques avant de prendre la décision qui convient.

Pour conclure, j'aimerais savoir si les administrations qui ne possèdent pas la capacité technique d'anonymiser elles-mêmes leurs données peuvent s'adresser à Etalab, et si l'on ne pourrait pas envisager une sorte de mutualisation des compétences en la matière.

M. Henri Verdier. Pour ce qui est de l'optimisme, je suis moi aussi extrêmement préoccupé de voir que nous entrons dans un monde où la vie privée dans ses aspects les plus intimes, y compris le corps, peut se trouver étalée aux yeux de tous. Ce monde, c'est celui de Twitter et de Facebook – où les gens partagent beaucoup de choses de façon volontaire –, c'est celui où American Express est capable de prédire l'imminence d'un divorce et où les objets communicants recueillent quantité de données biométriques.

Nous avons en France un concept très précieux, celui d'information à caractère personnel. Tous les agents publics que j'ai rencontrés ont bien compris qu'une information à caractère personnel, ce n'est pas seulement une information nominative, mais une information désignant une personne. Aujourd'hui, sur les 13 000 séries de fichiers présentes sur data.gouv.fr, une petite cinquantaine seulement est relative à des informations à caractère personnel telles que la distribution des revenus. Toutes les autres sont des données d'essence statistique, provenant généralement des services statistiques des ministères, qui savent parfaitement mettre en place des anonymisations statistiques – en tout état de cause, un contrôle est effectué sur ce point par le Conseil national de l'information statistique, au sein duquel on trouve un comité du secret chargé d'émettre des avis. De fait, très peu d'informations à caractère personnel sont partagées par l'administration française, et le seul incident dont je me souviens, rapporté par la presse en 2013 – l'identité et l'imposition de certains contribuables avaient été retrouvées dans une base pourtant anonymisée, en appliquant la technique du carroyage –, a été immédiatement corrigé.

En résumé, le risque existe, mais ce n'est pas dans les données gérées par l'administration qu'il est le plus grand, car l'État évite autant qu'il le peut de manipuler des données à caractère personnel, et les agents publics ont le réflexe de le faire avec précaution, en consultant la CNIL ou le comité du secret à chaque fois que cela paraît nécessaire. Pour ce qui est d'Etalab, notre service n'est pas responsable en droit – nous sommes dotés d'un statut d'hébergeur, et c'est l'administration d'où provient la donnée qui choisit de la partager sur data.gouv.fr, où elle possède un compte. J'avoue que nous avons eu quelques alertes, et que nous avons parfois dû réanonymiser sans tarder certaines données dont la publicité semblait de nature à pouvoir poser problème, mais très franchement, je ne crois vraiment pas que le plus gros danger pour la vie privée de nos concitoyens provienne du partage de données par

les services statistiques des ministères. Régulièrement, certains évoquent des scénarios catastrophe du genre : « et si un fou s’amusait à publier la liste des personnes décédées à l’hôpital ainsi que la cause du décès, avec une granularité assez fine pour que quelqu’un croise cette liste avec la rubrique nécrologique du journal local et puisse ainsi déterminer qui est mort de quoi ? » Mais c’est de la science-fiction, personne n’a même jamais songé à faire ça !

M. Philippe Aigrain. En tant qu’auteur de nombreuses propositions de nouvelles formes de financement des activités culturelles, je suis un peu inquiet à l’idée qu’il soit possible, comme cela a été dit tout à l’heure, de déterminer quels sont les livres les plus empruntés dans telle ou telle bibliothèque. Une donnée sensible, ce n’est pas forcément une donnée personnelle, et lorsqu’on évoque des solutions reposant sur un traitement à l’échelle d’une société, l’hypothèse selon laquelle quiconque puisse avoir accès à un fichier global des consommations culturelles suscite une très forte réticence – les seuls cas où cela peut être envisagé sont ceux où l’État attribue des aides, notamment dans le secteur du cinéma.

J’ai l’impression que nous avons du mal à nous comprendre, dans la mesure où vous êtes centré sur le problème des données publiques produites par les administrations, alors que la question qui se pose aux citoyens est celle, plus globale, de l’ensemble des données. Si, contrairement à ce que l’on entend dire, il est rare que les gens mettent n’importe quoi sur Facebook, en revanche, une masse énorme de données extrêmement sensibles est capturée, à l’insu des personnes concernées, par des acteurs privés. Quand le moment sera venu de formuler des propositions sur le contenu d’une nouvelle loi relative à l’ouverture des données en général, nous devons être attentifs à tous les risques potentiels. En effet, l’histoire a montré que la dénégation des risques a fait plus de tort à l’innovation que leur prise en compte.

M. Daniel Le Métayer. Si je suis d’accord sur le fait que les sociétés commerciales sont responsables d’un grand nombre d’abus, j’estime que ce n’est pas une excuse pour l’administration, qui doit avoir un comportement exemplaire lorsqu’elle met en ligne des données publiques.

M. Henri Verdier. Pour moi, l’administration a toujours été exemplaire jusqu’à présent. Cela dit, dans la mesure où il n’existe pas à l’heure actuelle de verrous de sécurité d’ordre juridique, rien ne permet d’affirmer qu’aucun dérapage ne pourra jamais se produire. La seule disposition légale en matière de données, c’est le droit dont dispose le citoyen d’accéder à un document : rien ne régit le devoir de l’administration de produire des données. Je le répète, en l’état actuel des choses, cet aspect des choses n’est organisé que par l’existence de tabous, l’esprit de sérieux des agents, la culture des services statistiques et le fait que, matériellement, l’organisation hiérarchique des services nécessite de franchir sept ou huit niveaux avant de procéder au partage d’une donnée.

M. Daniel Le Métayer. Loin de moi l’idée de mettre en cause le sérieux de l’administration. Je veux simplement souligner qu’il est difficile de définir la granularité idéale des données : pour ce qui est du découpage géographique, par exemple, comment déterminer si la diffusion de données peut se faire à partir d’un carreau ou une cellule IRIS de l’INSEE comprenant huit, dix ou onze foyers ? Si aucun problème sérieux ne s’est jamais posé en France, il y en a eu ailleurs, notamment aux États-Unis, où le dossier médical du gouverneur du Massachusetts, rendu accessible sous forme anonymisée, a été identifié rien qu’en procédant au croisement de quelques fichiers – et les exemples de ce type ne vont pas manquer de se multiplier à mesure que l’ouverture des données va devenir plus fréquente.

M. Henri Verdier. Tout ce qui est d'une granularité fine sera désanonymisable, seules les données suffisamment agrégées ne le seront pas – dans ce domaine, il est difficile de fixer des limites très précises. J'ai dit tout à l'heure que sur les 13 000 séries de fichiers présentes sur data.gouv.fr, seule une cinquantaine comportait des informations à caractère personnel : en fait, je me demande s'il existe ne serait-ce qu'une seule série de fichiers contenant des données relatives à un individu précis. Et pour ce qui est de la carte géolocalisée et horodatée des accidents de la route mise en ligne par le ministère de l'intérieur, que j'ai évoquée tout à l'heure, je dois préciser que les données la composant avaient été préalablement floutées.

M. Daniel Le Métayer. J'en reviens à ma question relative à la mutualisation des compétences en matière d'anonymisation des données. Je sais qu'il existe en Grande-Bretagne un réseau appelé *UK Anonymisation Network* (UKAN) chargé de fédérer les compétences sur ce point. Pensez-vous qu'une telle structure ait vocation à être mise en place en France et, le cas échéant, qu'il soit temps d'engager une réflexion sur ce point ou qu'il soit encore trop tôt pour cela ?

M. Henri Verdier. Nous avons déjà la CNIL et le comité du secret, mais il est vrai qu'une structure capable d'intervenir à la volée, de manière plus rapide, aurait certainement son utilité pour répondre à des questions s'inscrivant dans un registre plus technique que juridique.

M. Mohammed Adnène Trojette. Pour répondre à la question qui m'a été posée par Godefroy Beauvallet au sujet des juridictions de contrôle, je me bornerai à souligner – en mon nom propre – que l'équivalent néerlandais de la Cour des comptes a publié récemment, sur son autosaisine, un rapport sur la situation de l'*open data* dans les administrations des Pays-Bas, qui aura vocation à être réactualisé dans les mois et les années qui viennent.

Une autre question, qui m'a été posée par écrit, était la suivante : « Comment surmonter les difficultés actuellement rencontrées en matière de libre accessibilité et de réutilisation gratuite et automatisée des données publiques (qualité des données brutes, interopérabilité, formats d'enregistrement, licences d'utilisation, *et cætera*). C'est une question importante, à laquelle on peut répondre de deux manières. La première, que je qualifierai d'ancienne, est d'attendre qu'une administration vue comme omnipotente et omnisciente résolve tous les problèmes cités, qui constituent autant d'obstacles à la diffusion d'informations publiques en interne à l'administration. Si je ne conteste pas les pouvoirs et la compétence de l'administration, je pense que cette approche crée une distance entre l'administration et les citoyens usagers, fondée sur une asymétrie d'informations et de compétences qui n'existe plus aujourd'hui.

La seconde approche, qui me semble être celle à privilégier, consiste à s'appuyer sur la capacité de tous les acteurs à travailler ensemble en mettant en commun les moyens, parfois limités, dont ils disposent. L'un des exemples de cette collaboration nous a été donné lors de l'ouverture des déclarations d'intérêts et d'activités des parlementaires, effectuée dans le cadre de l'application de la loi sur la transparence de la vie publique. La qualité des données brutes n'était pas excellente, la plupart des formulaires étant complétés à la main, et leur opérabilité était réduite, les documents ayant été scannés et enregistrés au format pdf. En revanche, une innovation est à noter en ce qui concerne la licence d'utilisation puisque, sur décision du Secrétaire général du Gouvernement par délégation du Premier ministre, c'est une licence *open data* qui a été retenue, ce qui a permis à de nombreux acteurs – associations, développeurs informatiques, journalistes et citoyens – d'intervenir sur ce contenu, et j'avoue

avoir fait partie de ceux qui ont contribué à la réalisation de l'interface destinée à permettre la réutilisation par tous des informations fournies par les élus. En moins d'une semaine, et sans faire appel à une prestation de service d'un coût de plusieurs dizaines de milliers d'euros, ce travail réalisé en commun a permis d'aboutir à un jeu de données de grande qualité, vérifié, disponible sous licence ouverte et distribué sur data.gouv.fr. Nous avons ainsi fait progresser collectivement notre démocratie, et je pense que c'est vers des exemples de ce genre qu'il faut s'orienter lorsqu'on parle d'*open government*.

M. le coprésident Christian Paul. Vous avez été charitable en omettant d'indiquer quels progrès pourrait accomplir l'Assemblée nationale en matière d'ouverture de ses données, mais nous comptons sur chacun de vous pour nous faire parvenir par courrier électronique vos propositions sur ce point – plus sérieusement, je vous confirme que toutes les contributions de nature à nourrir les préconisations qu'il nous revient de formuler seront les bienvenues.

Je vous remercie pour vos interventions, qui ont permis un débat très riche.

La séance est levée à vingt heures quarante-cinq.

