



# Denis Berthault

## La réalité de l'open data

Selon Denis Berthault, nous sommes aux prémices de la libération des données publiques, un nouveau monde qui mettra au moins 20 ans à s'édifier. Nous sommes en période d'apprentissage, en train de découvrir la complexité technique, organisationnelle, économique, etc. de cette problématique. Il ne croît pas à un régime unique de l'ouverture des données publiques, car chaque type de données a ses particularités, ses contraintes, mais à des textes secteur par secteur. On en voit déjà l'illustration avec le projet de loi sur la santé. Il prône une vraie réflexion sur la gouvernance de l'open data et la prise en main du sujet par le pouvoir politique qui doit comprendre les enjeux de ce nouveau monde.

**Sylvie Rozenfeld : Les données publiques, c'est votre métier depuis fort longtemps, avec un parcours professionnel qui se situe entre le droit, la technique et le commerce. Vous avez suivi l'évolution de la diffusion des bases de données juridiques, en tant que chef des ventes à l'Européenne de données qui commercialisait des bases de données, puis en tant que directeur de l'information juridique chez ORT, concessionnaire de service public de la diffusion du droit pour Legifrance et aujourd'hui directeur du développement des contenus en ligne chez Nexis-Lexis. Vous êtes co-animateur du groupe « données publiques » au GFII. Vous avez donc particulièrement suivi et vécu cette évolution vers la libéralisation des données publiques. La pensiez-vous possible il y a seulement dix ans .**

**Denis Berthault :** Il y a deux étapes importantes. D'abord la directive de 2003 sur les données publiques qui a été transposée en France dans le cadre de la loi du 17 juillet 1978. On peut d'ailleurs s'interroger sur la qualité de cette transposition, qui a rajouté un volet réutilisation, ce qui ne facilite peut-être pas la compréhension d'un public nouveau, même s'il y a porosité entre accès et utilisation . Puis il y a eu l'aspect politique avec le discours de Barack Obama en 2008. Même si celui-ci ne fait que mettre des mots sur une situation qui existe, il exprime une volonté de rendre l'administration américaine plus efficace, plus rapide, moins coûteuse, plus transparente. Il met le coup de pied dans la fourmilière et oblige ainsi tous les Etats à se positionner.

#### **Pourquoi cela les oblige-t-il à se positionner .**

Car tout d'un coup les Etats-Unis portent un projet nouveau qui met en place des compétences qui jusque-là existaient mais qu'on ne rassemblait pas sous cet objectif. A partir de ce moment, Barack Obama ouvre la voie. Même si l'avancée du projet n'est pas aussi positive que le discours. Ce chemin va ensuite être emprunté par les Britanniques, les Scandinaves, etc. On s'aperçoit que la problématique des données publiques est indissociable de celle du gouvernement numérique dont l'open data fait partie. En revanche, l'open data ne se réduit pas à l'e-gouvernement. L'open data doit profiter à tous. Il ne peut pas y avoir qu'un seul gagnant : l'administration, la société civile ou le secteur privé.

#### **Pourquoi ?**

En France, on a abordé la libération des données publiques uniquement sous l'angle de la réutilisation commerciale des données publiques, avec le débat sur les données juridiques qui était en avant-garde. La démarche n'était pas liée à l'open data.

#### **L'idée était-elle de rémunérer l'administration ?**

Les données juridiques faussent le paradigme des données publiques. D'abord, les données juridiques sont consanguines au métier de l'Etat. Dans les années 70, les bases de données ont été créées par l'administration pour ses besoins. A la fin de la décennie, apparaît l'idée d'une diffusion qui va se faire auprès d'un cercle restreint,

principalement des juristes. De plus les données juridiques représentent des enjeux de pouvoirs, plus que des enjeux économiques. On ne voulait pas que le droit soit commercialisé par des entreprises américaines. On voulait une solution française, autonome, avec une démarche presque gaullienne. Ces grandes bases de données ont été conçues avec des moyens très importants, elles sont très puissantes d'un point de vue documentaire. Sans équivalent dans le monde. Je pense au fichier de législation consolidée Légi, à la base de données de jurisprudence de la Cour de cassation, du Conseil d'Etat, au fichier législatif du SGG, etc. Ce sont des bases de données conçues par des professionnels pour des professionnels. On est loin de l'open data.

#### **Mais Legifrance est constitué de ces bases de données.**

Pour diffuser ses données, l'Etat avait créé un dispositif qui associait secteurs public et privé, avec des concessions de service public, très décrites à l'époque à cause du monopole institué au profit du concessionnaire pour la réutilisation des données. Ce qui ne serait plus possible aujourd'hui. L'objectif était de trouver un diffuseur privé qui se rémunère sur le marché sans que l'Etat ait à le payer. Ce montage juridique a permis que Legifrance soit créé sans que l'Etat ait à le financer. Ce n'est qu'en 2002 que l'Etat a payé pour Legifrance. Avant, seul le concessionnaire supportait les coûts d'hébergement et de développement et ce n'est pas neutre, en 2014 Legifrance coûte 19 millions d'euros par an. C'est le concessionnaire, ORT qui a véritablement bâti Legifrance. Je ne dis pas que c'est une solution idyllique, mais elle a constitué un levier pour l'action de l'Etat qu'on pourrait considérer plus positivement aujourd'hui et qu'on pourrait mettre en œuvre, sans évidemment repartir dans des notions d'exclusivité. Mais les données juridiques faussent le paradigme des données publiques car ce sont de vieilles bases de données payantes qui sont devenues gratuites. Donc, cela ne correspond pas à une vraie politique d'ouverture des données. Comme le projet de gratuité des données du registre du commerce avec Infogreffe : la base de données existe déjà.

#### **Mais elle est payante. Alors qu'est-ce qu'une vraie politique d'open data ?**

Barack Obama a fait rupture en créant un chemin, forçant les administrations à rendre publiques leurs données sur des plateformes, avec certains formats, etc. Cela ne correspond pas à la politique actuelle de la France qui demande à l'administration d'ouvrir ses données, mais sans contrainte. Ce n'est pas ça l'open data.

#### **Pourtant, une volonté politique d'ouvrir ces données se manifeste, même si les résultats ne sont pas encore au rendez-vous. D'ailleurs pourquoi cette ouverture peine-t-elle à s'imposer en France ?**

Mais aussi ailleurs. Il faut souligner la création d'Etalab, en 2011. Le pouvoir politique prend la main, dans le sillage du discours d'Obama. Etalab, tel un missionnaire, va évangéliser les administrations qui sont très rigides, structurées. C'est le passage de la RGPP (revue générale des politiques

---

*« Après avoir fait cette  
transposition, sans doute a minima,  
on verra apparaître  
des textes plus verticaux. »*

---

publiques créée par Nicolas Sarkozy pour rendre l'administration plus efficace) en SGMap (secrétariat général à la modernisation de l'action publique). A partir du moment où Etalab et le SGMap se mettent ensemble, on voit apparaître des instruments intéressants. Un troisième vecteur me paraît novateur dans ce paysage de l'open data, c'est la Disc (direction interministérielle de la sécurité des systèmes d'information) compétente pour les gros systèmes. La question se pose de savoir comment faire sortir ces données, comment modifier les systèmes existants ? Entre les bases de données juridiques et les nouvelles données qui arrivent sur le marché, notamment grâce aux capteurs, où l'on prévoit tout dès le départ (format, export, etc.), il y a 98% des données qui restent à explorer. Là il faut rendre grâce à la création d'un chief data officer, avec la nomination d'Henri Verdier. La France est le premier Etat à avoir créé ce poste. Rattaché au Premier ministre, il a pour fonction la détection des gisements de données mais l'explication auprès de l'administration.

#### **Comment font les autres Etats ?**

Les Belges ont pris le problème par le biais de la simplification administrative. Les administrations doivent coopérer pour les échanges de fichiers. Ainsi, elles seront plus à jour, coûteront moins, etc. Les fichiers seront de meilleure qualité, ce qui leur permettra de faire de l'open data. J'ai retenu de cette expérience la nécessité d'un pouvoir normatif qui force les administrations à agir. Elles ont des intérêts divergents entre elles, par rapport au secteur privé, par rapport au secteur associatif ou aux citoyens. On ne voit pas encore apparaître un intérêt commun, d'un partage dans certaines conditions. Pour en revenir au chief data officer, je pense qu'il va bientôt terminer cet état des lieux, la cartographie de ce que l'administration possède et pourrait éventuellement sortir. Ensuite, on va s'apercevoir qu'extraire ces données va coûter cher car il va falloir modifier les systèmes d'information, diligenter des expertises, rédiger des cahiers des charges, etc. Si les données sont gratuites, comment l'Etat va-t-il se financer ?

#### **Est-ce pour ça que vous dites que le dogme de la gratuité freine la diffusion des données publiques .**

Je vois passer des projets industriels exploitant l'open data de 5 ou 10 millions d'euros. Le prix des données n'est pas forcément une question centrale. Ce n'est pas parce qu'elles sont gratuites ou payantes que le projet se fera ou non. Les porteurs de projet, surtout s'ils veulent pérenniser leurs investissements préféreront payer et avoir des données de qualité, à jour et structurées. Si c'est gratuit, c'est donc la personne publique (le citoyen) qui paie. Les budgets publics n'étant guère prospères en ce moment, j'en tire des conclusions. Je ne pense pas aux gros partenariats public/privé qui sont lourds et compliqués à mettre en œuvre mais plutôt à une concession de service public avec des obligations. Notons qu'il ne peut plus y avoir d'exclusivité, mais on pourrait imaginer que le concessionnaire n'ait pas le droit d'intervenir sur le secteur concurrentiel ou qu'il puisse le faire mais en payant le coût

de la réutilisation des données, pour en disposer dans des conditions strictement identiques aux concurrents actuels ou futurs. Je trouve cette solution intéressante.

#### **Il ne semble pas que ce soit l'option envisagée.**

Certes, car on occulte le débat. Ce sera peut-être envisagé en 2016 ou 2017. Il est encore trop tôt.

#### **Cela pourrait-il apparaître dans le futur projet de loi numérique, qui devrait comporter un chapitre sur l'ouverture des données publiques ?**

Je pense que les esprits ne sont pas encore assez mûrs pour que ce texte puisse l'envisager. Cela viendra peut-être par la suite quand on évoquera la question des investissements. Peut-être proposera-t-on la participation de l'Agence pour le patrimoine immatériel de l'Etat, créée dans la foulée du rapport Lévy et Jouyet.

Je n'arrive pas à comprendre qu'on trouve très logique que l'Etat valorise son patrimoine maritime, les trottoirs de Paris loués aux cafés, etc. et pas ses données. Je me demande si, dans certains cas, on ne pourrait pas avoir une vision patrimoniale de la donnée publique. Si l'administration a le sentiment que l'open data ne sert que le secteur privé, cela ne marchera pas. Elle pensera avoir affaire à une privatisation rampante et le risque est bien réel. Par conséquent, il faut que d'une façon ou d'une autre l'administration soit aussi gagnante. L'ouverture des données publiques pourraient être l'occasion de la rendre plus efficace et de meilleure qualité. Thierry Mandon, secrétaire d'Etat à la Réforme de l'Etat et à la Simplification, arrive avec un certain pragmatisme et fait bouger les lignes.

#### **Bientôt la directive européenne doit être transposée dans le droit français.**

Avant le 15 juillet, inutile de dire que nous serons hors délais. Mais la directive ne nous apporte pas grand-chose. Nous sommes déjà quasiment au niveau de ce texte. Le vrai débat n'est plus européen mais français. L'administration se dote-

elle des moyens nécessaires ? Je pense que le pouvoir politique, au sens large, a envie de moderniser, de simplifier, d'être plus rapide. Et les données publiques représentent un des moyens. Mais il n'y a pas qu'une seule grille de lecture. Cela dépend du type de données.

---

***« Il y a tout un travail de normalisation, de structuration, de mutualisation, de format d'échanges, voire d'anonymisation qu'il va falloir concevoir dans un mode industriel. »***

---

#### **Qu'entendez-vous par différentes grilles de lecture ?**

On peut établir quelques principes généraux, mais on ne peut pas descendre très loin. Par exemple, les problématiques des données de santé n'ont rien à voir avec celles des transports. Les enjeux, les acteurs, les risques, etc. ne sont pas les mêmes. Après avoir fait cette transposition, sans doute a minima, on verra apparaître des textes plus verticaux. On intégrera un chapitre « données publiques » à un texte sectoriel. Par exemple, en matière d'environnement, la directive Inspire de 2007 prévoit un régime de données publiques open data en la matière.

Le magistrat de la Cour des comptes, Mohammed Adnène Trojette, a remis en question l'utilité des redevances,

dans un rapport de novembre 2013 sur l'ouverture des données publiques. En 2012, elles ne rapportaient « que » 34,7 millions d'euros à l'État, dont près de 5 millions provenaient des acheteurs publics eux-mêmes. Que pensez-vous d'une suppression de la redevance.

Le rapport Trojette est habile. Selon lui, il ne peut y avoir de redevance sur les données mais sur les services autour des données.

### **Pouvez-vous rappeler sa mission ?**

Il devait définir les conditions économiques dans lesquelles les données publiques pouvaient être réutilisées. Il a collecté les chiffres de la réutilisation des données de la Dila, de l'IGN, de l'Insee, de l'Inpi, de Bercy, de l'Intérieur, etc. Ce rapport propose un système assez souple de paiement du service de mise à disposition des données. Les données sont gratuites, mais les méta-données sont payantes. Il peut y avoir d'autres solutions.

De toute façon, on va verticaliser de plus en plus avec une approche par secteur : les données urbaines, sociales, de la vieillesse, du handicap, etc. Celles qui émanent des conseils généraux, des régions, des ministères, des hôpitaux, etc. Cela ne va pas être simple de les extraire. Suivant les particularités, on va prévoir des règles, des exceptions, des modèles d'anonymisation, des modèles économiques particuliers. Et puis, si nous voulons qu'il y ait des échanges de données, il faut prévoir des normes. Dans les années 90, nous avions l'EDI. Ce n'est plus « tendance », et pourtant, pas d'open data sans EDI, avec une définition des formats d'échanges, du responsable du fichier émetteur, de ses droits et obligations, de son droit à contracter avec d'autres administrations, etc. C'est un nouveau monde qui va mettre au moins 20 ans à s'édifier. Sur ce sujet, tout est à inventer.

### **Il y a la question de l'extraction des données, dans un format exploitable.**

Des données mises à disposition au format pdf sont inexploitables. Ce n'est pas de l'open data. Certaines administrations ne font encore que comme ça.

### **Alors que faudrait-il faire pour avoir des données exploitables et interoperables ?**

Il y a tout un travail de normalisation, de structuration, de mutualisation, de format d'échanges, voire d'anonymisation qu'il va falloir concevoir dans un mode industriel. Un organisme interministériel pourrait mutualiser au mieux les savoirs. Il faudrait donc éviter que chaque administration développe sa propre plateforme.

Il faut réfléchir à une gouvernance de l'open data. Henri Verdier, l'administrateur général des données y travaille. Il pose des éléments de réflexion. En France, on a une vision que j'appelle AtoC de l'open data : l'administration vers le citoyen. Cette dimension existe bien sûr mais elle n'est pas la seule. Il y a aussi AtoA, les échanges d'administrations entre elles et AtoB entre les administrations et le secteur marchand. Mais pour l'instant, on se concentre sur le AtoC et sur la gratuité. Je comprends très bien que dans le AtoC, les données soient gratuites. Pour le reste, c'est différent.

### **Vous qui suivez l'actualité des données publiques, vous devez être déçu par le projet de loi sur la santé, qui prévoit une libération à minima des données de santé ?**

Un des gros problèmes des données de santé est leur organisation au sein de l'administration. Il existe des conflits énormes entre certains établissements publics, qui ont un poids plus important que certains ministères, et le ministère de la Santé.

### **Pensez-vous à la Caisse primaire d'assurance maladie (CPAM) ?**

Oui. Tous ces organismes ont des logiques très cohérentes de données quand ils sont seuls, mais de là à les échanger. Les données de santé n'ont rien à voir avec les données de transport. C'est la raison pour laquelle on a créé un régime spécial pour les données de santé. Et ce sera un premier jet. Je parie que nous verrons un nouveau projet de loi d'ici trois ans.

### **Le projet de loi a mis en place une véritable usine à gaz, extrêmement complexe. Et avec un champ très restreint.**

Car nous ne sommes pas mûrs. Nous sommes en période d'apprentissage et l'on se pose des questions.

On a commencé à identifier les acteurs, leurs contraintes, leurs ambitions, etc. Il y a beaucoup d'acteurs que ça intéresse : les laboratoires pharmaceutiques, les banques...

On a connu les mêmes débats avec le dossier médical personnalisé. Il faudra un jour se demander pourquoi il a

coûté si cher pour si peu d'efficacité et pourquoi l'EDI santé s'est bloqué.

### **Pourquoi, donc ?**

Problème de gouvernance, antagonisme entre acteurs, volonté de chacun d'imposer sa norme, etc. Cela montre que le politique doit prendre la main. Pour cela, il doit comprendre les enjeux de ce nouveau monde.

### **Vous disiez justement au début que Barack Obama avait imposé de manière autoritaire l'ouverture des données publiques. Ne s'est-il pas retrouvé face aux mêmes contraintes ?**

Bien sûr, et tout le monde les a. L'ère numérique pose des problèmes d'organisation dans le secteur public comme dans le secteur privé. Les hiérarchies sont affaiblies. L'administration française n'est pas préparée à cela. Elle n'est pas très partante pour extraire ses données, qui permettent son contrôle.

### **Si je comprends bien, l'open data n'est pas pour demain.**

C'est un long chemin.

Peut-être qu'il y aura une progression rapide dans tel domaine et pas dans tel autre, mais l'évolution se fera secteur par secteur. Encore une fois, l'expérience qu'on a acquise avec les producteurs historiques comme la Dila ou l'IGN ne nous intéresse pas, car c'est trop spécifique.

Pour le reste, on commence tout juste à percevoir la complexité du chantier. L'open data est un révélateur. Comme quand la mer se retire à marée basse, c'est à ce moment qu'on découvre ce qu'elle recouvrait. Nous sommes

---

*« Avec l'open data, nous sommes face à de vrais enjeux de souveraineté. »*

---

dans une phase de découverte. Et il est donc encore trop tôt pour établir les bons constats. Donc, en ce qui concerne les données de santé, nous sommes dans une phase de découverte des problèmes et des enjeux. Il est donc normal que la loi de Marisol Touraine soit décevante. Dans trois ans, on l'amendera, après avoir appris de cette période de test.

**Vous dites que c'est un révélateur parce que nous ne sommes pas conscients de tous les enjeux. Il y a aussi les enjeux de pouvoir. Les Etats perdent de leurs pouvoirs au profit des marchés. Est-ce que cela ne conduit pas à une certaine dépossession ?**

Ca dépend.

Les applications sur les smartphones servent à capter beaucoup de données. Les personnes deviennent, à leur tour, productrices de données de santé. Elles font ainsi irruption sur le marché. Mais les données de santé représentent un tel enjeu que cela attire de gros acteurs tels que Google, etc. Si un opérateur propose une contrepartie intéressante en échange de nos données de santé, tout le monde va accepter. Doctissimo est pour l'instant la meilleure base de contenus de santé en France.

Elle a des faiblesses et des inconvénients. Si l'administration ne fait rien, il risque de se constituer des bases de données de mauvaise qualité. L'administration ne peut pas rester immobile.

**Aux risques de se voir dépossée.**

Tout à fait. Et ça va particulièrement vite dans le domaine de la santé. Puis les banques ont besoin de fichiers, comme les assureurs. Ils chercheront ou créeront le moyen d'obtenir ces données. On peut imaginer une prime ou une réduction en contrepartie des données. Et sur leur parc, ils pourraient ainsi se créer une base de données. L'administration doit donc bouger, c'est particulièrement marquant dans le domaine de la santé.

Pour revenir à votre question sur les enjeux de pouvoirs, il existe un risque que l'administration ne soit plus la seule productrice officielle de la donnée. Prenons l'exemple de la météo. Pour acheter des données à moindre prix, il suffit d'aller en Norvège qui diffuse les données françaises obtenues gratuitement. Avec l'open data, nous sommes face à de vrais enjeux de souveraineté.

**Il y a également le risque de voir des données produites en France et exploitées par des grands acteurs étrangers.**

De quel étranger parlons-nous ? Il y a Google, mais que penser des Chinois ou des Indiens ? C'est un vrai challenge dans un domaine où la France se situe entre les 3ème et 4ème places des classements de l'open data. Nous sommes confrontés à une compétition au sein de l'Europe mais aussi au niveau international. La Grande-Bretagne a opté pour une orientation business de ses données publiques. Elles doivent permettre à l'administration de faire des économies et elles sont mises à disposition du secteur privé qui va les réutiliser. David Cameron les envisage comme des instruments de rayonnement, de transparence et de business.

Il existe aussi des enjeux de compétition entre régions ou communes. Par exemple, si elles veulent mettre en valeur leur patrimoine culturel, elles ont intérêt à mettre des données à disposition pour être mieux référencées ou que leur site soit plus visité.

**Vous parlez de transparence. N'est-on pas en passe de vivre la fin de l'anonymat ? L'anonymisation est-elle réellement possible ?**

Je voudrais d'abord souligner que la France a une certaine maturité sur le sujet. La Cnil est en avance sur la réflexion et semble réussir à fédérer ses homologues européens. L'anonymisation est une question complexe juridiquement et techniquement onéreuse. Et puis on ne peut pas anonymiser correctement l'information sans structuration préalable. C'est grâce aux formats qu'on va pouvoir anonymiser les données à un prix raisonnable. Grâce au taggage des données, on va pouvoir envisager des degrés d'anonymisation en fonction des usages. Il faut donner confiance, en passant par l'anonymisation.

**Et est-ce que l'open data va changer nos vies ?**

La problématique de l'open data ne concerne pas que les données publiques. Il y a aussi les données privées, celles détenues par les entreprises et celles que possèdent et produisent les personnes. Les deux n'ont rien à voir. Ces deux domaines sont en train de se dissocier. Les enjeux, les contraintes ne sont pas les mêmes. Si une entreprise désire échanger ses données, cela la concerne ainsi que ses clients. C'est donc du domaine de la liberté contractuelle. Quand vous me demandez si l'open data va changer nos vies, je vous répondrais oui, dans une certaine mesure pour l'ouverture des données publiques, mais énormément s'agissant des données privées.

Dans quelques années, ma voiture, ma maison, etc. recevront toutes les données émises par différentes sources, des secteurs privé et public. Renault annoncera peut-être un jour qu'il vend toutes les données sur ses véhicules, les pannes, les tests, etc. Cela donnerait une bonne vision sur la fiabilité du véhicule. Une entreprise peut avoir envie de diffuser ses données pour une question d'audience, de concurrence, de confort de ses clients, etc. On en est encore à l'âge de pierre.

**Propos recueillis par Sylvie ROZENFELD**

*« L'anonymisation est une question complexe juridiquement et techniquement onéreuse. »*