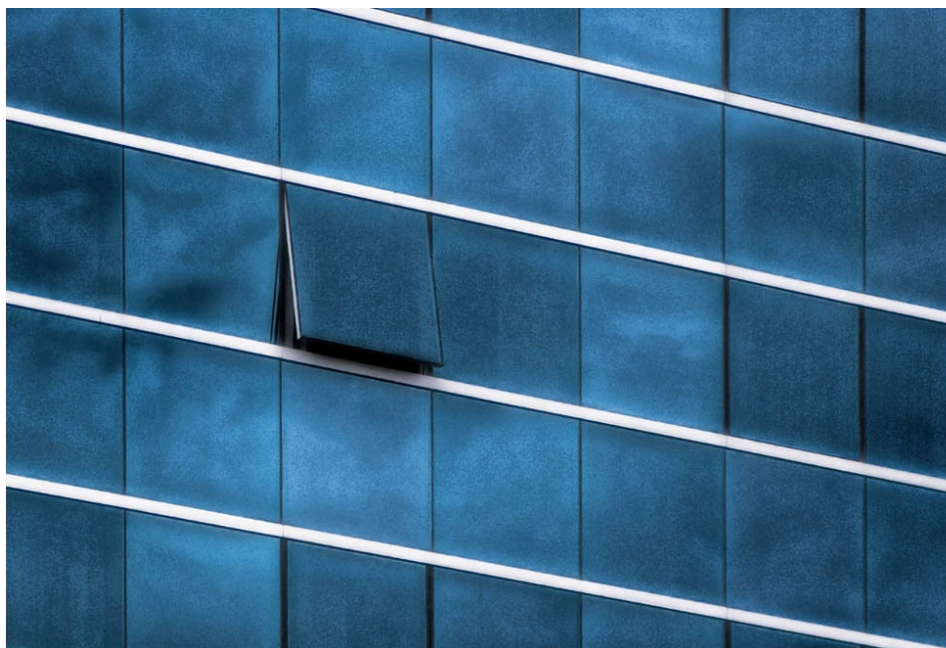


UNE MANNE DE DONNÉES ENCORE À LIBÉRER

LE 14 DÉCEMBRE 2011 AXEL BIZEL-BIZELLOT

Malgré un enthousiasme légitime pour la démarche d'ouverture des données, la France reste à la remorque des métropoles américaines et de l'Angleterre. De quoi motiver, selon le sociologue Sylvain Parasio, un lobbying pour un Open Data digne de ce nom.



Sylvain Parasio est sociologue, maître de conférence à l'Université Paris-Est / Marne-la-Vallée et chercheur au laboratoire techniques, territoires et société (**LATTS**) de l'École Nationale des Ponts et Chaussées. Ses recherches portent sur *“la publicité, le journalisme en ligne, les activités en ligne, et plus largement les implications sociales et politiques associées à l’usage des nouvelles technologies dans le monde des médias et de la communication”*.

Répondant aux questions de **journalismes.info**, il analyse les difficultés françaises quant à l'ouverture des données publiques.

Peut-on vraiment parler d'innovation à propos d'Etalab ?

Premièrement, il convient de rappeler que l'ouverture des données publiques n'est pas quelque chose de vraiment neuf, c'est un processus qui s'est développé aux États-Unis et en Angleterre depuis pas mal de temps, dans les grandes villes notamment. A ce titre, Washington et Londres peuvent être considérées comme des références, parce qu'elles sont en avance mais aussi parce qu'elles mettent énormément de données à disposition de tous. En France, il est clair que l'on a un retard à combler à ce niveau là.

Comment expliquer ce retard ?

Ce qui a posé le plus de problèmes, c'est la question des droits sur les données publiques. Quels usages en faire ? Doit-on permettre une récupération par les entreprises privées ? Ces questions ont longtemps fait débat, notamment dans le cadre de la licence. Certains souhaitaient que l'on puisse en faire une réutilisation commerciale, d'autres étaient plus réticents à l'idée qu'elles puissent être sources de profits. En même temps, il était compliqué d'en faire des services, de développer des applications sans intermédiaires privés et donc sans utilisation commerciale. Ce qui a été finalement tranché, c'est grossièrement que les organismes publics doivent s'assurer de la qualité de ces données, et laisser ensuite les intermédiaires les réutiliser, même si les débats perdurent.

On a vu récemment plusieurs initiatives de ce genre se développer dans des villes françaises, comme Rennes ou Paris, qui ont commencé à mettre en ligne des catalogues de données. Mais quel est vraiment le contenu de ces catalogues ?

Pour l'instant, il est très pauvre. On a accès à des informations sur la liste des jardins, des kiosques... Il n'y a pas vraiment de cohérence. Rien à voir avec ce qui se fait à Chicago par exemple. Là-bas, les bases de données sont vraiment impressionnantes. On a par exemple accès aux noms et aux salaires de tous les employés des entreprises de transport public, et ces données ne sont même pas anonymisées ! Il y avait d'ailleurs eu un petit scandale récemment, puisque un "1" avait été rajouté malencontreusement devant un chiffre... La personne concernée a dû vivre un véritable cauchemar pendant quelques jours ! On trouve aussi tout ce qui concerne l'inspection alimentaire, les permis de construire, les stations de polices ou de pompiers, les crimes quartier par quartier... Rien à voir donc.

Comment expliquer un tel fossé ?

Pour être honnête, il convient de rappeler que le pouvoir des villes en France et aux États-Unis n'est pas du tout le même. La mairie de Paris ne dispose pas forcément d'autant de données que celle de Chicago. Pour les transports par exemple, la RATP n'est pas directement reliée à la mairie, contrairement à celle de Chicago. Deuxièmement, il faut aussi dire que Chicago a été une ville laboratoire de l'ouverture des données publiques. La ville a commencé à diffuser massivement, bien avant l'arrivée du web, dès les années 1960 à propos des données criminelles. En France, ces données sont toutes centralisées par le ministère de l'Intérieur ; on a accès aux statistiques département par département mais pas à l'échelle du quartier. À Chicago, on peut savoir précisément quel crime a été commis et où depuis les années 1990. Comme le taux de criminalité était particulièrement élevé, le maire avait mis en place une police de proximité. Chaque policier était responsable d'un bloc de la ville et devait tenir des réunions publiques régulières pour informer les habitants des quartiers. La diffusion par le web s'est faite dès 1996. L'administration n'a donc pas du tout le même rapport aux données, il n'y a pas de centralisation comme en France ; là bas, ces données sont des outils pour la police municipale. La comparaison est donc un peu difficile. Les mairies ont en France plus de mal à avoir ces données ; il y a un réel problème d'accessibilité. Mais au-delà, on peut se demander s'il n'y a pas une différence de rapport au citoyen. Publier les données nominatives avec les salaires des fonctionnaires serait par exemple absolument inouï ici.

Et par rapport au contenu de la future plate-forme data.gouv.fr ? Le fait que l'État ait la mainmise sur la publication des données publiques ne peut-il pas représenter un risque ?

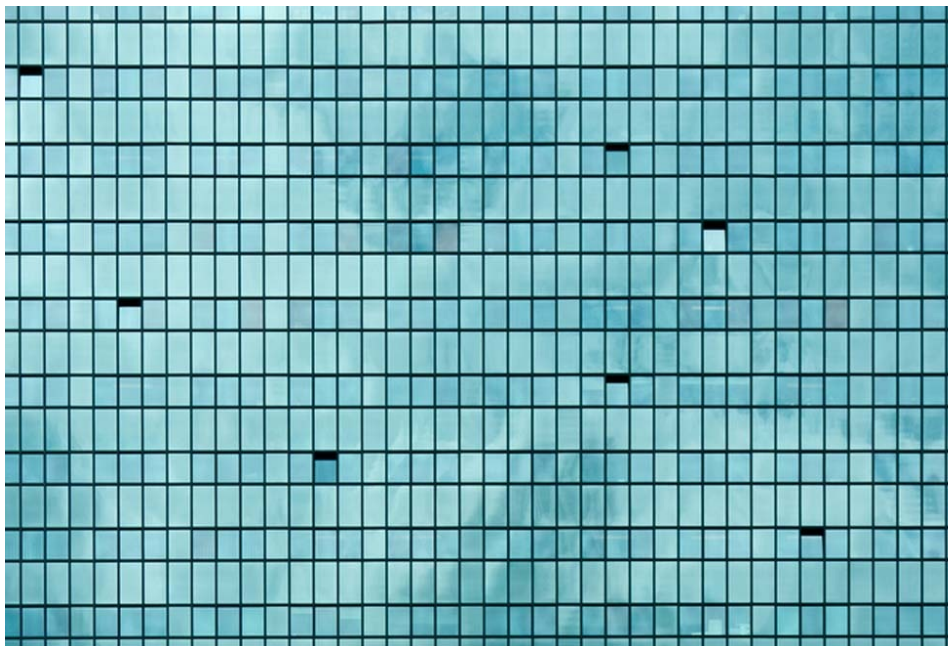
Pour l'instant, au niveau du contenu on ne sait pas vraiment... Mais vous avez raison de

soulever le problème, il s'agit d'une volonté politique, qui peut être remise en question dans la mesure où il existe des réticences au sein du pouvoir. Après il convient de ne pas oublier que l'on ne part pas de rien non plus, que l'Insee existe depuis 1946 et propose des données de très bonne qualité. Il est possible effectivement que certaines données soient sujettes à caution, notamment sur les postes supprimés dans l'enseignement, ou concernant les élections : ces données sont diffusées par le ministère de l'Intérieur, mais juste à un moment, et il faut souvent se tourner vers les travaux des chercheurs pour les obtenir après. Alors oui, je pense globalement que l'État fait preuve de bonnes volontés avec Etalab, mais il reste en France beaucoup d'efforts à faire à propos de la transparence. Et je suis également persuadé qu'il faut continuer l'activité de lobbying pour l'ouverture des données publiques, comme le font remarquablement Regards Citoyens auprès du Sénat ou de l'Assemblée Nationale par exemple.

■ Au final, peut-on vraiment parler d'une avancée pour le citoyen ?

Oui, mais se pose le problème de l'accessibilité. Et pour rendre ces données brutes accessibles, il faut des intermédiaires. C'est là qu'intervient le journaliste. L'activité de lobbying des militants de l'Open Data a besoin derrière de personnes pour rendre les données obtenues accessibles à tous. Et c'est sûrement là que se trouve le principal problème en France.

Encore une fois, si l'on compare aux États-Unis, on tient compte de l'utilisation de bases de données dans le journalisme depuis beaucoup plus longtemps. Dès les années 1960, on trouvait là bas ce que l'on appelait les CAR (*computer associated reporter*). On peut situer l'origine au moment des émeutes de Détroit, en 1967. On a commencé à utiliser les premiers ordinateurs pour faire des questionnaires automatisés. Les gens dans la rue étaient principalement des noirs, et personne ne comprenait vraiment leurs revendications. Grâce à ce travail, on a pu montrer que contrairement aux idées reçues, la majorité des émeutiers étaient des gens qui avaient un assez bon niveau d'étude. Ensuite, cela s'est développé vraiment à partir de la fin des années 1980, où beaucoup de journalistes étaient aussi un peu statisticiens, et traitaient de sujets comme l'éducation, la démographie ou la criminalité, basés quand même essentiellement sur des chiffres.



En France, la culture est encore une fois très différente, le journalisme est beaucoup plus littéraire, engagé. Il est vraiment nécessaire de trouver des personnes ayant l'habitude de manipuler des bases de données, capables d'en faire des applications. Et pour l'instant, à part OWNI, on ne trouve pas grand monde... Il existe des initiatives comme **nosdeputes.fr**, mais ce sont des initiatives purement citoyennes ; elles n'émanent pas directement de la presse comme aux États-Unis. Même en Angleterre, on trouve beaucoup de journalistes qui s'intéressent par exemple aux comptes-rendus du Parlement et se spécialisent dans l'administration et l'État ; et ce même avant l'arrivée de l'informatique. Ce genre de travaux est peu développé en France, peut être que la défiance vis-à-vis du pouvoir est plus forte là-bas, que les citoyens sont plus demandeurs de transparence. Avec les quantités de données disponibles, il faut avoir la bonne idée journalistique à la base pour pouvoir en extraire une information ; il faut une interprétation objective des données, un peu comme le font les chercheurs en sciences sociales. On voit par exemple se développer des

partenariats, des passerelles entre le journalisme et les sciences sociales aux États-Unis, mais pas en France pour l'instant.

Justement, en parlant de transparence, quels types de barrières pourraient demeurer ?

Obtenir la diffusion de certaines données peut rester compliqué, les administrations doivent aussi y trouver leur compte. Un des arguments du lobbying Open Data, c'est la transparence bien sûr, mais aussi la rationalisation, que l'administration soit plus efficace, qu'il faut améliorer le rapport entre l'État et le citoyen. La situation financière des médias en France peut aussi constituer une sorte de barrière. Je ne pense pas qu'il y ait de volonté de censure. Le problème pour obtenir une réelle situation de transparence est à mon sens à chercher du côté des intermédiaires. L'État bouge, doucement, mais bouge malgré tout, et c'est du côté des journalistes qu'on peut observer au final le plus de retard. Et enfin, il ne faut pas oublier le rôle de l'internaute non plus. On peut expliquer le succès d'un site comme Everyblock dans le sens où il est devenu de plus en plus communautaire. Les citoyens l'utilisent par eux-mêmes; ils ont su en dépasser le côté austère. Au final, le processus d'ouverture des données publiques est clairement un jeu à 3, entre l'État, le journaliste et l'internaute lui-même. Les principales difficultés de la France sont donc à mon sens à chercher à la fois du côté d'une administration trop centralisée, d'un journalisme trop littéraire et d'un manque d'intérêt de la part de l'internaute pour l'instant.

Publié initialement sur [Journalismes.info](#) sous le titre **L'ouverture des données publiques est un jeu à 3 entre l'Etat, le journaliste et l'internaute**

Photos par [Ecstaticist \[cc-byncsa\]](#) ; [Daniel*1977 \[cc-byncsa\]](#) ; [Victor Bezrukov \[cc-by\]](#)

MIKEY

le 14 décembre 2011 - 12:31 • [SIGNALER UN ABUS](#) - [PERMALINK](#)



Cool. Des tonnes de données. Pourquoi ne pas les mettent sur Internet? Sur que sa favorisera l'espionnage espionnage de notre contrée. Comme le dit le quantique de l'Archange St Michel, il faut battre le démon qui empêche l'élévation des ames. Je crois à la culture pour tous, accessible par tous. Mais la gestion de telles données, je laisse sa aux politiques, c'est leurs boulots !

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

[LUI RÉPONDRE](#)

JM BOURGOGNE

le 20 décembre 2011 - 9:03 • [SIGNALER UN ABUS](#) - [PERMALINK](#)



Très intéressant article. Il ne faudrait jouer plus loin le jeu de l'ouverture : il n'y a pas que Rennes, Paris et récemment l'état qui ont ouvert leur données. Il y a Montpellier (avril), Communauté de Bordeaux (juillet), CR Aquitaine(aout), Toulouse (sept.), CG71 (oct.), Nantes (nov.), CR PACA qui vient de passer une délibération avant hier, CG 41 en pilote depuis 3 mois, etc... Ca commence à faire pas mal de citoyens concernés.... et beaucoup de collectivités impliqués.

La question n'est plus tellement l'ouverture mais la réutilisation de ces données : pertinence, usages, production collective quand elles sont incomplètes (ce qui très souvent le cas).

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

[LUI RÉPONDRE](#)

BIDULETTE

le 11 janvier 2012 - 20:10 • [SIGNALER UN ABUS](#) - [PERMALINK](#)



Merci pour cet article très intéressant. Quelques précisions peuvent être utiles.

Les expressions "ouverture des données", "libération", ne me paraissent pas complètement appropriées : les données n'étaient pas enfermées, ou bien, si l'on considère qu'elles l'étaient, c'était avec permission de sortie. Quiconque pouvait déjà auparavant avoir accès aux mêmes informations auprès des administrations. La nouveauté (certes de grande importance) est la diffusion volontaire en ligne.

Même des informations nominatives sont communicables : il ne faut pas confondre cette notion avec celle de mise en cause de la vie privée (le nom d'un fonctionnaire est communicable, l'adresse de son domicile ne l'est pas).

Je ne cite pas les textes réglementaires, mon commentaire est déjà trop long...

Par ailleurs, il faut tenir compte d'un aspect de la question qui est évoqué dans un autre commentaire... des sociétés commerciales trouvent là d'un coup une nouvelle source de revenus, elles vont pouvoir récupérer ces données gratuites, en proposer en accès payant, avec des stratégies marketing qui pourraient même faire un peu oublier à certains que ces infos sont ailleurs accessibles gratuitement. Elles n'auront peut-être pas toujours beaucoup de scrupules sur la protection de la vie privée (exemple : <http://www.archivistes.org/La-reutilisation-des-donnees>, communiqué "La réutilisation des données nominatives ? Gare au fichage des individus !" de l'Association des archivistes français, concernant une affaire de réutilisation qui n'est pas vraiment réglée à ce jour). Dernier point: la question, évoquée, de l'interprétation des données. Bien sûr, l'accès aux données brutes est mieux que pas d'accès du tout, mais leur utilisation dépend de la capacité de l'utilisateur à les critiquer en fonction de leur contexte de création. Encore faut-il qu'il dispose des informations sur ce contexte.

VOUS AIMEZ



0

VOUS N'AIMEZ PAS



0

LUI RÉPONDRE

8 pings

Open Data by Solucom - Pearltrees le 14 décembre 2011 - 10:35

[...] Et par rapport au contenu de la future plateforme data.gouv.fr ? Le fait que l'État ait la mainmise sur la publication des données publiques ne peut-il pas représenter un risque ? Pour l'instant, au niveau du contenu on ne sait pas vraiment... Mais vous avez raison de soulever le problème, il s'agit d'une volonté politique, qui peut être remise en question dans la mesure où il existe des réticences au sein du pouvoir. Une manne de données encore à libérer » OWNI, News, Augmented [...]

guizmo69007 | Pearltrees le 14 décembre 2011 - 11:06

[...] Une manne de données encore à libérer » OWNI, News, Augmented Comment expliquer ce retard ? Ce qui a posé le plus de problèmes, c'est la question des droits sur les données publiques. Quels usages en faire ? Doit-on permettre une récupération par les entreprises privées ? [...]

La démarche "open data" « Chez toi le 14 décembre 2011 - 23:00

[...] article intéressant sur l'ouverture des données et le retard français en la matière, sur [owni Share](#) [...]

OPENDATA - données accessibles aux citoyens by liozone - Pearltrees le 15 décembre 2011 - 15:37

[...] Une manne de données encore à libérer » OWNI, News, Augmented Et je suis également persuadé qu'il faut continuer l'activité de lobbying pour l'ouverture des données publiques, comme le font remarquablement Regards Citoyens auprès du Sénat ou de l'Assemblée Nationale par exemple. [...]

Open Data by showloulou - Pearltrees le 16 décembre 2011 - 10:35

[...] Une manne de données encore à libérer » OWNI, News, Augmented C'est là qu'intervient le journaliste. L'activité de lobbying des militants de l'Open Data a besoin derrière de personnes pour rendre les données obtenues accessibles à tous. Et c'est sûrement là que se trouve le principal problème en France. [...]

L'ouverture des données publiques « Livre, lecture et numérique le 17 décembre 2011 - 0:35

[...] font remarquer. Le site OWNI estime que les données ne sont pas encore totalement libérées. Un article en particulier mentionne cette difficulté. "Au final, peut-on vraiment parler d'une [...]"

Interventions by acourt - Peartrees le 22 décembre 2011 - 19:04

[...] Une manne de données encore à libérer » OWNI, News, Augmented Encore une fois, si l'on compare aux États-Unis, on tient compte de l'utilisation de bases de données dans le journalisme depuis beaucoup plus longtemps. [...]

Définitions/concepts/réflexions by josselinhenno - Peartrees le 2 janvier 2012 - 0:37

[...] Une manne de données encore à libérer » OWNI, News, Augmented Encore une fois, si l'on compare aux États-Unis, on tient compte de l'utilisation de bases de données dans le journalisme depuis beaucoup plus longtemps. [...]