

Open data : les mauvaises raisons de ne pas mettre à disposition des données

En dehors des cas prévus par la loi, les raisons apparentes de ne pas partager sont nombreuses. L'expérience sur le terrain conduit à penser qu'elles sont souvent mal fondées, voire qu'elles servent de prétexte à une réticence de fond à partager l'information. Florilège de ces arguments.

L'open data ? Je ne sais pas ce que c'est.

À partir de 2018 et l'application de la Loi pour une république numérique (LPRN), tout agent devrait savoir ce dont il s'agit, d'autant que le principe est simple : la mise à disposition gratuite, en ligne, des données du service public, en dehors de quelques exceptions de bon sens. Depuis la première constitution de la République française, en passant par la loi CADA de 1978, jusqu'aux récents textes comme la LPRN, il s'agit d'un mouvement de fond visant à plus de transparence et d'efficacité de l'action publique. Les entreprises peuvent y être aussi obligées et peuvent y trouver un grand intérêt.

La qualité comme argument

Mes données ne sont pas de bonne qualité, elles comportent des erreurs.

Qui prétend ne *jamais* en faire ? Et même avec de nombreuses erreurs, les données peuvent tout de même intéresser les réutilisateurs : certaines sont très faciles à corriger, elles peuvent intéresser à titre de comparaison, etc. En outre, on constate que la publication de données peut, dans certains cas, conduire les utilisateurs eux-mêmes à les corriger, au plus grand bénéfice du producteur. Ce dernier aurait en revanche intérêt, pour tout jeu de données, à indiquer un taux d'erreur, même grossier.

Mes données sont incomplètes.

Il est très probable qu'elles soient en réalité suffisantes sous un certain angle, par exemple restreintes à une zone géographique bien définie. Mêmes "incomplètes", ces données peuvent être utiles aux réutilisateurs pour initier un travail. Là encore, le producteur doit documenter clairement la couverture de ses données.

Mes données ne sont pas assez précises.

La précision recherchée ne dépend que de l'usage. Une donnée géolocalisée à une précision de plusieurs km près, sera par exemple amplement suffisante pour cartographier un phénomène national.

Mais le format de ces données est mal documenté !

Un réutilisateur vraiment intéressé trouvera toujours moyen de le faire a posteriori. Peut-être une bonne occasion pour vous de dialoguer avec ces réutilisateurs pour produire une documentation de qualité.

Le format de ces données n'est pas standard.

Si le format est documenté, ce n'est pas un problème. Si les réutilisateurs y voient un intérêt, convertir des données bien documentées d'un format à un autre n'est pas si difficile. Si le format est peu ou pas documenté, la simplicité de certaines données peut suffire à en comprendre le sens.

Des données pas adaptées à des usages supposés

Ces données sont trop anciennes.

Des données périmées pour leur usage initial peuvent néanmoins intéresser de nombreux acteurs – chercheurs, étudiants, consultants... – désireux d'effectuer par exemple des travaux de synthèse, d'analyse historique, etc.

La mise à jour de ces données n'est pas assez régulière.

Là encore, elles peuvent présenter un intérêt pour des usages d'analyse comparative, de travaux de synthèses ou autres.

Ces données n'intéressent personne !

Chiche ! Dans tous les cas, ne pas préjuger à l'avance qu'elles n'intéressent personne.

Les réserves juridiques et concernant la sécurité

Rien ne m'y oblige, pourquoi partager des données ?

Si vous êtes un acteur public de plus 3500 habitants, la loi pour une République numérique vous l'impose. Si vous êtes un acteur privé, le législateur impose d'ouvrir certaines données si vous êtes en délégation de service public, ou votre entreprise fait plus de 500 salariés, ou vous êtes un acteur de l'énergie, des transports publics ou des assurances. Les lois françaises et européennes vont vers une plus grande ouverture des données, qu'elles soient issues d'acteurs publics ou privés.

Enfin, quel que vous soyez, les raisons de publier vos données sont nombreuses, même si rien ne vous y oblige. L'argument le plus important est qu'à l'heure où les données se multiplient, les données "monopoles" se font rares. La rareté se fait rare. Partager ses données permet donc de prendre une place stratégique dans un écosystème et vous laisse une chance de devenir ou rester l'acteur de référence.

Mes données contiennent des données à caractère personnel.

Le premier cas de figure est celui d'élus dans le cadre de l'exercice de leur mandat : ces données relèvent de la sphère publique et non de la sphère privée. L'agenda public d'un élu, par exemple, peut donc tout à fait être publié en open data.

Pour les autres cas, plus classiques, la publication de données à caractère personnel n'est absolument pas interdite en soi mais soumise à la loi CNIL. Si vous avez recueilli le consentement des utilisateurs et que vous êtes bien certains qu'ils en ont compris la portée, il est donc possible de publier ces données.

Mes données permettent d'atteindre à la sécurité des biens et des personnes.

Ces cas sont extrêmement rares. Ils concernent quelques jeux de données sur des centaines : les plans des centrales nucléaires, les réseaux d'eau potable, l'organisation du système d'information, etc.

De même, s'il convient que vos usagers puissent avoir connaissance de tous vos canaux de diffusion et de vos sites pour les accueillir, il n'est pas opportun qu'ils connaissent les adresses de vos entrepôts, de vos réserves... L'ouverture des données, ne doit en aucun cas venir perturber le bon fonctionnement de l'organisation.

Mes données contiennent des secrets commerciaux, de fabrication... ou leur publication va à l'encontre de droits d'auteurs, de propriétés intellectuelles, de clauses contractuelles...

Dans de telles situations, l'étude au cas par cas est nécessaire. Il est recommandé de ne pas aller au-delà des attendus des lois. Pour cela il est judicieux de prendre connaissance précisément de ces derniers. La Cnil préconise de caviarder le document ou de l'anonymiser.

Les données open data permettent des attaques cyber !

Nous avons entendu cet argument au plus haut niveau de l'État. Soyons extrêmement clairs : à moins de publier les clés d'accès à vos systèmes d'information (clés privées, comptes et mots de passe) (!), les données open data ne permettent pas des "attaques cyber". Depuis les débuts de l'open data en 2007, aucun cas de ce genre n'a été remonté.

Les réserves concernant les concurrents

Publier les données de l'entreprise favorisent mes concurrents

Cet argument peut s'appliquer à tous les jeux de données des acteurs privés. A vous de voir si cet argument n'est pas aussi "Publier les données de l'entreprise défavorisent mes

clients” ou bien s’il ne constitue pas un frein à l’innovation.

Des réserves relatives aux modalités

Je le fais déjà ! Toutes mes données sont exportées dans la base XXX.

Les collectivités et acteurs publics locaux participent au quotidien à l’alimentation de bases locales ou nationales : bases de l’IGN, base Sit@del2 relative aux permis de construire, base de l’Observatoire national des services d’eau et d’assainissement, Mérimée et Joconde pour les musées, etc. D’où cette légitime et désagréable impression de faire le travail plusieurs fois.

On n’a pas le temps. Et qui va mettre à jour ?

Le producteur de données se sent déjà dépassé par sa charge de travail et fait part d’un problème organisationnel qu’il ne peut résoudre seul. Ce sera au chef de projet en charge de l’open data d’identifier la cause réelle du problème et de proposer des solutions. Dans bien des cas, l’organisation de la production de données n’est pas en phase avec la stratégie de la structure.

L’open data c’est compliqué, on ne sait pas faire !

Certains interlocuteurs sont mal à l’aise et peu acculturés sur les différentes démarches menant au partage des données.

Fondamentalement, il ne s’agit que de publier des fichiers en ligne sous une des deux licences possibles. Publier un fichier en open data peut prendre 10 minutes. Deux ressources clé pour aller plus loin : l’ouvrage *Open Data* de Vincent Kober et la riche *Documentation* d’OpenDataLocale.

On ne sait pas produire un fichier un CSV

Certains agents sont mal à l’aise et peu acculturés sur les formats traditionnels de partage des données. Il est possible de commencer par publier des formats qu’ils maîtrisent mieux, comme le .xls.

Des raisons cachées ?

Toutes ces raisons de ne pas vouloir partager des données peuvent masquer des raisons inavouables ou inavouées.

La première concerne le **sentiment de dépossession des données**. Les producteurs le traduisent maladroitement par une remarque fréquente : “*on ne sait pas ce qui va être fait de ces données !*”. Il ne faut pas négliger ce sentiment qui constitue un des freins les plus forts. Les solutions sont connues : créer et multiplier les rencontres entre réutilisateurs et producteurs permet à ces derniers de comprendre les usages et les bénéfices d’un partage sans condition. De même, sur votre portail, il est utile de demander aux réutilisateurs ce qu’ils font des données, par exemple au moment du téléchargement.

Une deuxième raison réside en la **difficulté et l’appréhension à devoir répondre à des demandes**, naturellement suscitées par la publication : “*je n’ai pas les moyens de faire du support*” entend-on parfois. Ces demandes sont souvent peu nombreuses et une occasion de comprendre les bénéfices du partage de données. Le responsable de l’éditorialisation se doit idéalement d’organiser cet aspect : éventuellement effectuer un tri des questions posées, répondant lui-même aux questions simples, signalant explicitement les questions techniques auxquelles il ne peut répondre, etc.

Enfin certains voient l’obligation de publier des données comme une contrainte qui alourdit un peu plus leur travail quotidien, déjà bien **trop réglementé**.

Paradoxalement, l’open data c’était plus excitant lorsqu’on y était pas obligé (!). Avoir converti l’open data comme contrainte réglementaire, a peut-être tué un certain esprit pionnier, une belle idée en routine administrative...

Animer un atelier “les fausses barbes de l’open data”

Une animation simple permettant l’appropriation rapide de ces contenus.

Durée : environ 20-30 minutes

Public : des personnes qui pilotent ou animent une démarche open data.

Déroulé

5’ L’animateur propose aux participants de noter avec des gommettes ou de cocher les questions qu’ils ont déjà rencontrées, dans le cadre de leur mission ou dans d’autres cadres. Pour un petit groupe, il peut aussi faire un sondage à main levée.

15-25’ À chaque question, l’animateur interroge ensuite un participant par question pour lui demander quelques éléments de contexte et son attitude par rapport à la question. Une discussion peut s’ensuivre.

Des bonnes raisons de ne pas mettre à disposition des données ?

Ces raisons sont tout d’abord **réglementaires**. La loi indique clairement les données exclues de l’open data : [à compléter]

Mais il pourrait exister aussi des **raisons politiques** à ne pas publier certaines données, ou tout du moins, sous certaines formes. Les statistiques de délinquances offrent un exemple intéressant : alors qu’un acteur public s’engage dans une politique de revalorisation de certains quartiers, il peut ainsi jouer contre son camp en publiant des statistiques qui pourraient montrer du doigt ces mêmes quartiers.

Des raisons objectivables

Pourquoi moi ? Quel est le processus de décision ? Je ne suis pas décisionnaire.

Certaines données sont produites par plusieurs services ou plusieurs organisations et il est difficile d’identifier qui en a la propriété ou la gouvernance. Les agents peuvent ainsi se sentir illégitimes à prendre la décision d’ouvrir un jeu de données, même si la loi les y invite. Ce cas peut être difficile à traiter. Il nécessite en amont de clarifier la gouvernance des données, ce qui peut représenter beaucoup de temps et de discussions. L’idéal est d’obtenir un message clair et partagé du plus haut niveau, évitant aux agents de se poser 1000 questions sur le processus de décision.

On se tire une balle dans le pied !

Il arrive que les données publiées viennent frontalement contredire une politique publique ou une valeur portée par l’organisation. Dans certains cas, il faudra veiller au calendrier de publication des données ou expliquer les incompréhensions qui peuvent être produites par les démarches de communication et de publication de données. Ou encore ajuster ou revoir les politiques conduites.