

Petit guide à destination des administrations souhaitant basculer du payant à l'Open Data

Regards Citoyens - juillet 2014



[Extrait de L'Open Data à la Loupe – 2011 LAL](#)
[A2BProduction – LiberTIC – Nantes Métropole](#)



Ça y est ! Votre hiérarchie, ou mieux encore le gouvernement, a tranché : ces données supposées publiques que votre administration vend depuis des décennies à un oligopole d'acteurs établis doivent désormais rejoindre le monde de l'Open Data et vont enfin devenir librement réutilisables par tous !

Il n'est jamais inutile de le répéter : l'Open Data est clairement défini suivant des standards reconnus dans le monde entier et reposant notamment sur les [10 principes](#) de la [Sunlight Foundation](#) et l'[OpenDefinition](#) de l'[Open Knowledge Foundation \(OKFN\)](#). Ces principes assurent simplement que chacun soit mis sur un pied d'égalité lorsqu'il s'agit de réutiliser des données publiques : aucune discrimination d'ordre juridique, financier ou technique ne doit pouvoir entraver l'accès de chaque réutilisateur potentiel. Le passage à la gratuité permet donc déjà de lever toute contrainte financière, mais les aspects juridiques et techniques restent encore à sécuriser.

1) Licence : faites le bon choix

En termes juridiques, vous avez un choix simple à effectuer entre les deux licences compatibles avec l'Open Data respectueuses du cadre juridique français imposé par la loi de 1978 et reconnues internationalement : un modèle le plus libéral avec la [Licence Ouverte \(LO\)](#) rédigée par Étalab, qui n'impose aux réutilisateurs que de citer la source des données, et l'[Open Database Licence \(ODBL\)](#) d'OKFN, [traduite en français par la communauté](#) en 2010 et adoptée par de nombreuses collectivités locales, imposant en sus de redistribuer les données éventuellement enrichies et préserver ainsi ces biens communs.

Si vos données disposaient déjà d'acheteurs et que vous pensez qu'au vu de leur place prépondérante, la gratuité ne résoudra pas seule le problème de concurrence, vous pouvez mettre en place un mécanisme dit de « double licence » : mise en Open Data d'une part sous licence ODBL permettant à tous de contribuer à l'enrichissement de vos données, et possibilité d'autre part de contracter une licence payante non Open Data pour les acteurs souhaitant « propriétériser » les données en refusant de republier les données mélangées (Google, Microsoft, marchands de données juridiques...) ¹.

Dans tous les cas, écrire une nouvelle licence Open Data « maison » comme [a essayé de le faire la DILA](#) n'est en aucun cas un choix envisageable : un long travail a été réalisé depuis 4 ans par l'ensemble des acteurs — public, privé et société civile — pour converger vers cet écosystème simple autour de deux licences compatibles à l'international et laissant le choix entre [deux modèles politiques](#). Toute contrainte complémentaire ajoutée au sein d'une licence causerait inéluctablement la sortie du cadre de l'Open Data, promettant votre initiative d'ouverture au rejet et à un échec cuisant.

¹ La préservation des biens communs que sont ces données est un motif d'intérêt général, une telle politique rentre donc parfaitement dans le cadre des dispositions prévues à l'article 16 de la loi CADA.

2) Identifiez les contraintes légales spécifiques

Comme c'est le cas avec le [décret 2002-1064](#) pour les données juridiques de la DILA, d'anciens textes officiels anachroniques pourront parfois vous imposer des conditions ubuesques totalement incompatibles avec ces principes, comme par exemple interdire de concéder de sous-licences et donc limiter la redistribution des données, élément pourtant au cœur de la démarche d'élargissement des biens communs inhérente à l'Open Data.

Si votre programme d'ouverture résulte d'une décision gouvernementale, vous n'aurez aucun mal à obtenir auprès de votre ministère de tutelle un décret corrigeant ces incongruités pour vous permettre d'adopter effectivement l'une des deux licences précitées. Faites donc cette démarche, plutôt que de chercher à tordre et réinventer les licences pour les plier à des contraintes de toute manière inconciliables avec l'Open Data. L'abandon de ces dispositions réduira de plus les risques juridiques pesant sur votre institution : le [rapport Trojette](#) a pu le démontrer, il sera extrêmement compliqué de chercher à concilier de telles contraintes avec les [dispositions européennes en matière de droit de la concurrence](#) ainsi qu'avec l'[article 16 de la loi CADA](#), qui indique que seules des dispositions d'intérêt général peuvent justifier des restrictions à la réutilisation.

3) Identifiez les contraintes liées aux données personnelles

Si vos données contiennent des informations à caractère personnel, elles sont a priori hors du cadre de l'Open Data. Certaines options restent cependant envisageables notamment si ces données ne sont pas liées à la vie privée des individus. Celles-ci seront alors qualifiées de données nominatives. Comme le récent avis de la [Cour de Justice de l'Union Européenne l'a illustré](#), les informations concernant des personnes publiques publiées pour des raisons légales, comme par exemple au sein du Journal Officiel, ont par nature vocation à être des données publiques et donc librement réutilisables.

Si aucune base légale n'autorise explicitement la publication de telles données à caractère personnel, la loi de 1978 prévoit deux cas de figure :

- retirer les informations personnelles de ces données : c'est une opération très simple si vos données sont structurées. Il s'agira dans ce cas d'une part de supprimer les données liées à la vie privée (adresse postale, téléphone, IP, religion ou appartenance syndicale...) pour ne conserver que des éléments à granularité statistique (code postal, département...), et d'autre part de pseudonymiser les noms et prénoms ². S'il subsiste un risque de violation de la vie privée après avoir effectué ces traitements et que vos données sont des données numériques, il vous faudra assurer l'anonymisation par le respect du [secret statistique](#), en agrégeant les données par paquets ³, par exemple en regroupant les informations par commune ou département.

² La CNIL recommande pour cette étape le recours à des [empreintes sha2](#) des noms et prénoms associés à une constante choisie de manière aléatoire pour le jeu de donnée.

³ Pour des données sans gros enjeu, les agrégats classiques sont de 5 personnes, pour sa base des ménages, l'INSEE utilise des agrégats de 11 ménages.

- demander l'autorisation aux intéressés : notamment si vous avez une relation contractuelle avec les personnes concernées par les données que vous souhaitez libérer, il ne vous sera pas très compliqué de recueillir leur consentement à la publication des informations les concernant, alors qu'il serait impossible aux réutilisateurs d'endosser la responsabilité juridique et technique d'une telle tâche ⁴.

Si l'anonymisation des informations fait peser un risque trop fort en termes de violation de la vie privée ou qu'il n'est pas possible de recueillir le consentement des personnes intéressées, votre jeu de données ne pourra pas être publié en Open Data. Il contiendra non plus des données nominatives mais des données personnelles liées à la vie privée, données non seulement totalement exclues du champ de l'Open Data mais dont la réutilisation est strictement interdite par la loi CADA de 1978 ⁵. Si vous vendez déjà ces données, il est plus que temps d'interroger la légalité sinon l'éthique de ce service...

4) Employez des formats ouverts

Pour être réutilisables librement par tous sans discrimination, il est indispensable que vos données soient mises à disposition sous une forme technique correspondant à des standards dont la définition est publique et non soumise aux conditions d'un éventuel propriétaire. Ces formats sont appelés [formats ouverts](#). Il en existe pour tout type de données. Les plus populaires sont : CSV, JSON, XML, SQL, OpenDocumentFormat (ODS).

Puisque vous vendiez auparavant ces données, vous avez de fortes chances qu'elles se trouvent déjà sous l'un de ces formats, comme c'est le cas par exemple pour la DILA. Autrement un petit effort sera probablement nécessaire pour adopter ces formats. Les informaticiens de votre service ou votre Direction des Systèmes d'Information seront certainement d'une grande aide : réaliser de telles conversions à partir de vos systèmes d'informations est dans les cordes de tout développeur informatique. C'est la clé du succès pour la réutilisation de vos données.

5) Faites au plus simple pour la publication

Sur internet, la mise à disposition de données brutes représente des coûts informatiques très faibles contrairement à la création d'interfaces ou la réalisation de traitements. Une simple archive compressée contenant les données mises à disposition sur un site web est suffisante et de nature à endiguer les surcoûts éventuels en termes de bande passante. Disposer de l'ensemble d'un jeu de données est le plus souvent indispensable aux réutilisateurs, et suffisant pour pallier à la création d'une API dont le développement représenterait un coût important de développement et maintenance pour votre institution. Attachez-vous surtout à assurer la mise à disposition sur une adresse url stable au gré des mises à jour. Maintenir un accès aux archives des mise à jours précédentes peut être un plus : ils intéresseront très certainement les chercheurs.

4 C'est notamment la solution qu'aurait du conseiller la CNIL concernant la publication des avantages consentis par les laboratoires pharmaceutiques aux professionnels de santé : les conventions liants les laboratoires aux médecins peuvent tout à fait intégrer une clause informant les médecins que les informations nominatives relatives à ce contrat seront librement réutilisables. Libre à eux d'accepter ou non la convention

Si votre base de données est conséquente (plusieurs gigaoctets de données) c'est qu'elle contient vraisemblablement des images ou des documents PDF. Si ces documents sont déjà publiés sur Internet, ne les incluez pas à votre jeu de données Open Data et remplacez les par des url absolues pointant vers ces documents. Les réutilisateurs se concentreront certainement dans un premier temps sur la partie structurée de vos données avant de s'intéresser aux documents inclus. Vous pouvez également distribuer ces images ou documents sous la forme d'une archive téléchargeable via des protocoles de répartition pair à pair de bande passante. C'est le choix retenu par Wikipedia, OpenStreetMap, Archive.org mais également la NASA ou le gouvernement britannique pour la diffusion de ces types de données.

Quoi qu'il en soit, l'accès aux fichiers mis en Open Data doit être proposé sous la forme de liens directs de téléchargement, et non derrière une quelconque restriction technique, formulaire requérant une intervention humaine ou collecte de données à caractère personnel sur les réutilisateurs : tout citoyen français doit être capable de télécharger les données de manière anonyme à partir d'une simple url.

6) Oubliez la comm égocentrée

Référez sur data.gouv.fr vos jeux de données publiés. Disposer d'une vitrine les recensant sur le site de votre propre administration est utile pour multiplier les points d'accès aux données, mais il n'est nullement nécessaire de perdre du temps et de l'argent en investissant dans le développement d'un portail Open Data maison, comme a pu le faire le [Ministère de la Recherche](#) via un marché public. La mission Étalab a déjà réalisé ce travail pour vous, profitez-en !

Vous ne publiez pas ces données pour promouvoir qui que ce soit, vous-même ou votre administration, mais pour assurer une mission de service public. L'accès libre aux données publiques pour tous est l'objectif en soi. Si les données ont bien été libérées sous conditions Open Data, les réutilisations arriveront sans doute d'elles-mêmes. Ne perdez pas donc votre temps avant même l'ouverture à préparer des communications, hackathons, sites officiels de réutilisation... Ce n'est pas là que vous êtes attendus et vous risquez de susciter auprès de la communauté des attentes que vous ne seriez pas capable de combler. La tâche d'ouverture est claire et balisée, le reste ne peut et ne doit venir qu'ensuite !

Ce n'est qu'une fois ces actions réalisées que vous pouvez envisager de communiquer. Votre premier réflexe sera alors de prendre attache avec [Étalab](#), si vous ne l'aviez pas déjà fait. De par leurs missions, ils connaissent bien la communauté Open Data et seront à même de vous indiquer le bon moment et les bons moyens pour communiquer. Les jeux de données que vous avez rendus publics vont certainement intéresser des réutilisateurs : n'hésitez pas à valoriser leurs initiatives en les félicitant publiquement par exemple via les réseaux de microblogging ou en les invitant à rencontrer vos équipes en charge de la production des données. C'est par ces échanges avec les citoyens que l'Open Data pourra susciter le plus d'impacts positifs pour votre institution.

Bon courage et à bientôt dans le petit monde de l'Open Data français !