INTELLIGENT CITY ET USAGES INNOVANTS DES DONNÉES PERSONNELLES :

des scénarios pour engager un rééquilibrage privé/public par les données

Geoffrey Delcroix
Responsable innovation, études et prospective,
Direction des technologies et de l'innovation - CNIL

Geoffrey Delcroix est responsable innovation, études et prospective au sein de la Direction des technologies et de l'innovation de la CNIL (Commission Nationale de l'Informatique et des Libertés).

Diplômé en sciences politiques, géopolitique et défense, Geoffrey a démarré sa carrière dans l'équipe Futuribles, un centre indépendant effectuant des recherches sur le monde contemporain, en tant que consultant et chercheur. Il est ensuite devenu responsable de la prospective au sein de la Délégation pour la prospective et la stratégie du ministère français de l'Intérieur avant d'entrer à la CNIL en 2011.

MOTS CLÉS

- OPEN DATA
- DONNÉES À CARACTÈRE PERSONNEL
- COMMUN INFORMATIONNEL
- LIBRE CIRCULATION DE LA DONNÉE
- PARTENARIAT PUBLIC-PRIVÉ

L'équipe se concentre sur trois missions :
- Explorer les tendances émergentes à la croisée des technologies numériques, de l'éthique et des données.
- Échanger des idées et être un point de contact et de dialogue avec les écosystèmes d'innovation du numérique (l'équipe travaille avec des start-ups, des laboratoires et des chercheurs sur ces sujets).
- Expérimenter des méthodes d'innovation et produire ou co-produire des présentations, des démonstrations de faisabilité ou des prototypes autour des questions de protection de la vie privée.

L'équipe publie sur différents sujets (véhicules connectés, Chatbots, robotique, IA, objets connectés, drones, santé numérique, algorithmes…). Tous les articles sont disponibles sur LINC (https://linc.cnil.fr/), CNIL Innovation & prospective et les médias portant sur l'innovation.

Le cinquième numéro des « Cahiers IP », intitulé La plateforme d'une ville, explore les questions associées à la ville intelligente et à l'utilisation des données dans la planification et les services urbains. Il contient des recommandations, notamment sur les différents outils qui pourraient nous permettre, à l'avenir, de mettre en place des utilisations pertinentes et contrôlées des données personnelles, dans l'intérêt de tous.

Face aux injonctions contradictoires de la smart city – personnaliser tout en respectant la vie privée, optimiser sans rejeter – et pour répondre au bouleversement du jeu des acteurs, notamment avec l'arrivée des industriels de la donnée, il convient de produire de nouvelles formes de régulation de la donnée urbaine, dans le respect des individus et de leurs libertés.

INTRODUCTION

Comment permettre le partage avec des acteurs publics de données collectées et exploitées par des acteurs privés, mais qui auraient une forte valeur ajoutée pour des finalités d'intérêt général, dans le respect des droits des entreprises en question, ainsi que des droits et libertés des personnes concernées? C'est une question à laquelle le droit et les politiques publiques essayent aujourd'hui de répondre. Comme décrit dans les autres parties du cahier « La plateforme d'une ville » de l'équipe d'innovation et de prospective de la CNIL, autorité française de protection des données, disponible en ligne (en français[1]), les nouveaux services de la ville numérique s'appuient de plus en plus sur des données personnelles, collectées et traitées pour un service commercial par des acteurs privés.

Ces données qui n'entrent pas dans le périmètre organique du Service public (régie directe, concession…) ont cependant une interaction forte avec les enjeux de service public, voire sont précieuses pour remplir des missions de service public.

Aujourd'hui, différents outils sont envisagés par les parties prenantes à ce débat. Tous présentent de sérieuses limites, tous offrent de vraies opportunités. Tous impliquent de trouver une adéquate balance des droits et devoirs entre les différents acteurs concernés.

Ces outils se distinguent selon deux axes. D'abord les obligations légales qu'ils feraient peser sur les acteurs privés : parmi les quatre propositions développées plus bas, certaines pourraient être mises en œuvre dans le cadre législatif existant, quand d'autres devraient faire l'objet de nouvelles dispositions légales pour être applicables. Ensuite la granularité des données :

1 https://linc.cnil.fr/la-plateforme-dune-ville-explore-les-enjeux-de-la-smart-city

dans certains cas, des données très fines seraient fournies à l'acteur public (dont des données personnelles), dans d'autres, l'acteur public aurait accès à des données agrégées et déjà anonymisées.

Dans un article précédent, intitulé « Partage ![2] *», nous soulignons qu'un modèle de régulation classique utilisé isolément a peu de chances d'être efficace et qu'une régulation adaptée à ces plateformes requiert un équilibre nouveau, plus dynamique, s'appuyant sur divers outils de régulation, comme autant de leviers à actionner : l'action sur les rapports de forces entre les acteurs (par le marché), l'action sur les systèmes et architectures techniques (par la technologie et le design), l'action sur des règles du jeu (par l'autorité et les normes), enfin l'action par l'autodétermination et le pouvoir redonné aux individus (empowerment).*

En croisant ces deux axes (obligations légales et agrégation des données) avec les quatre leviers de régulation, on obtient une matrice de quatre scénarios distincts, comme autant de futurs possibles, alternatifs ou combinables, pour de nouvelles formes de partage des données.

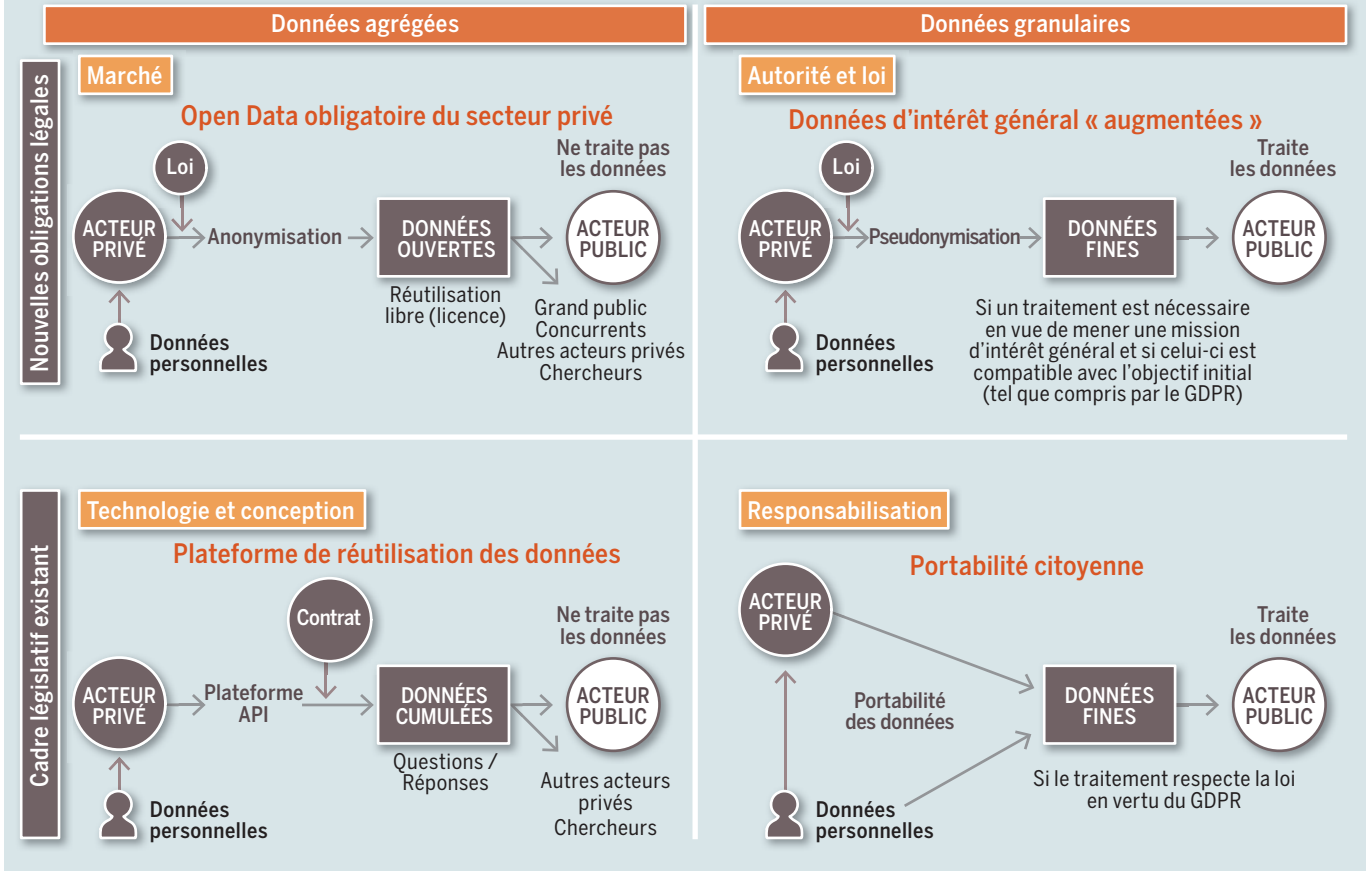
Ces scénarios proposent différentes formes de répartition des enjeux quant à la valorisation du capital en données fines et à la redistribution de la capacité à agir au profit de l'intérêt général, par la redéfinition de l'équilibre des rapports de force entre les acteurs publics et privés dans le cadre de finalités de service public.

Ils se différencient par la répartition de la charge de la protection des données personnelles, qui porte tantôt sur l'acteur privé, tantôt sur l'acteur public. Il conviendra le cas échéant d'adopter les bonnes pratiques permettant de garantir le respect des droits et libertés des personnes concernées.

Sans privilégier l'un ou l'autre de ces mécanismes, présenter l'économie générale de chacun et souligner leurs potentialités permet de mettre en lumière les enjeux qu'ils soulèvent pour la protection des données à caractère personnel des citoyens.

2 Voir le cahier IP Partage ! Motivations et contreparties au partage de soi dans la société numérique. https://linc.cnil.fr/fr/dossier-partage

Matrice de scénarios possibles pour le partage des données à l'avenir



GÉNÉRALISER UN « OPEN DATA DU SECTEUR PRIVÉ »

Agir sur le rapport de forces et créer les conditions d'une autorégulation efficace peut passer par l'instauration obligatoire de politiques d'open data du privé, pour les données dont l'importance pour le fonctionnement efficace du marché ou de politiques publiques d'intérêt général est avérée.

L'acteur privé met à disposition en open data certaines données qu'il traite par l'effet d'une obligation légale (sur l'exemple de ce qui a été prévu par deux lois de en France, la loi dite Macron ou la loi dite de transition énergétique[3]). Pour que ce processus soit conforme à la protection des données à caractère personnel, l'ouverture passe dans la majorité des cas par l'anonymisation, par des méthodes qui devront être conformes à la certification des processus d'anonymisation[4].

Un tel mécanisme a l'avantage de permettre la réutilisation par tous (concurrents, acteurs publics, chercheurs, citoyens…). Ce scénario présente bien sûr des inconvénients : l'anonymisation a un coût, à la fois financier pour l'acteur privé et en termes de perte d'information dans les jeux de données pour les réutilisateurs : l'acteur public ne disposerait par exemple pas de données très fines, utiles pour mener à bien des missions d'intérêt général. L'acteur privé reste maître du jeu quant à la qualité du jeu de données restituées.

ÉTENDRE LES DONNÉES D'INTÉRÊT GÉNÉRAL AU-DELÀ DES CONCESSIONS DE SERVICE PUBLIC

Changer les règles du jeu, c'est considérer qu'un intérêt supérieur justifie d'incarner des frontières intangibles posées par la société sur des sujets éthiques et politiques. Dans ce scénario, il s'agirait de permettre et encadrer la réutilisation de données personnelles par l'acteur public, pour certaines finalités d'intérêt public, sans porter atteinte aux droits des personnes concernées. Cela passerait par l'extension de la notion émergente de « données d'intérêt général », dans son périmètre et ses modalités. Les « données d'intérêt général » sont aujourd'hui restreintes aux entreprises concessionnaires de services publics, elles seraient étendues à des acteurs privés hors relations contractuelles avec la collectivité.

Ces données sont aujourd'hui anonymisées par l'acteur privé avant ouverture en open data. Il s'agirait d'ouvrir la voie à la restitution de certaines données

CNIL – Five BY Five – ©Léa Chassagne

fines à l'acteur public pour des missions de service public, charge à lui d'anonymiser ces données en cas d'ouverture en open data.

La balance des droits devra permettre d'éviter de porter préjudice à un acteur privé qui a investi pour construire son traitement de données et aussi d'éviter l'atteinte au droit à la vie privée des individus, qui ont consenti à un traitement dans le cadre d'un service particulier. La collectivité publique devient responsable de traitement et devra respecter l'ensemble des règles applicables (base légale, compatibilité des finalités, respect des principes de protection des données, etc.).

Un tel mécanisme aurait l'avantage de redéfinir l'équilibre des pouvoirs entre certains acteurs privés et les collectivités, qui disposeraient d'un levier efficace pour mener à bien des missions d'intérêt public, sans que cela ne conduise à porter atteinte aux droits des personnes concernées. Ce scénario aurait l'inconvénient d'être contraignant, pour les entreprises privées concernées qui devraient restituer des données, et pour les réutilisateurs publics, qui porteraient la charge de la protection des données personnelles.

Ce scénario a le vent en poupe : après la loi pour une République Numérique qui en a posé les prolégomènes, suite au rapport « relatif aux données d'intérêt général » de 2015[5], des hypothèses de ce type sont développées par exemple par la Commission européenne dans sa réflexion sur la libre circulation des données[6] ou dans le rapport du parlementaire Luc Belot[7], qui en appelle à la définition d'une catégorie de « données d'intérêt territorial » et à leur recensement.

3 Loi n°2015-990 du 6 août 2015 pour la croissance, l'activité et l'égalité des chances économiques et Loi n°2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance

4 Voir l'avis du groupe de travail « Article 29 » (Union européenne) 05/2014 sur les « techniques d'anonymisation »

5 CGEIET et IGF. Rapport relatif aux données d'intérêt général, septembre 2015. https://www.economie.gouv.fr/files/files/PDF/DIG-Rapport-final2015-09.pdf

6 Document de travail du personnel de la Commission sur la libre circulation des données et questions émergentes en matière d'économie européenne fondée sur les données accompagnant le document « Créer une économie européenne fondée sur les données », janvier 2017

7 Luc BELOT. De la smart city au territoire d'intelligence[s]. Rapport au Premier ministre sur l'avenir des smart cities, avril 2017

PERMETTRE LA RÉUTILISATION SOUS LE CONTRÔLE DES ACTEURS PRIVÉS

Agir sur les systèmes et les architectures revient pour la régulation à prendre la mesure de la transformation actuelle des modalités techniques de l'économie de la donnée. À ce titre, il peut s'agir d'encadrer l'émergence de plateformes d'accès et de partage des données en s'appuyant sur les outils juridiques et techniques. À des logiques d'open data, de « lacs de données » et d'anonymisation en bloc, pourrait répondre une logique d'API, de « robinets de données » et de confidentialité différentielle (differential privacy).

L'acteur privé met en place une plateforme de réutilisation de ses données par des outils techniques (APIs…) qui permettent au réutilisateur de tirer parti de certaines données, sans les traiter lui-même : le réutilisateur pose une question à la base détenue par l'acteur privé, celui-ci ne lui envoie pas le jeu de données, mais la réponse. Un tel système, bien conçu, permet une exploitation riche des données tout en minimisant les risques d'atteinte aux droits des personnes concernées. La plateforme peut alors mobiliser, en plus de l'anonymisation, deux types d'outils :

- Des outils juridiques : un contrat doit encadrer ce que les réutilisateurs peuvent faire ou non, par exemple, une clause interdisant au partenaire de tenter de réidentifier les personnes et de porter atteinte à leur anonymat, et des clauses traitant du partage de responsabilité ;
- Des outils techniques : d'audits, de contrôle, de vérification et d'analyses des logs en temps réel qui analysent dynamiquement les risques (par exemple pour limiter les possibilités d'attaque par inférences de la base).

Un tel mécanisme qui ne nécessiterait pas de nouvelles obligations légales aurait l'avantage pour l'acteur privé de ne pas être contraint à l'ouverture en bloc de données, l'acteur public n'aurait pour sa part pas à supporter la charge de la protection des données personnelles. Ce scénario aurait pour inconvénient le coût de développement et de maintenance de la plateforme par l'acteur privé, qui pourrait cependant lui offrir de nouveaux débouchés et de nouveaux revenus par la vente de données anonymisées.

ACTIONNER LA PORTABILITÉ CITOYENNE

Permettre à chacun de déterminer l'usage de ses propres données, donner les moyens de la participation citoyenne à la réalisation de missions d'intérêt général, ce sont là des opportunités offertes par le nouveau règlement sur la protection des données personnelles.

Le RGPD introduit un droit à la portabilité qui favorise la réutilisation de données personnelles par un nouveau responsable de traitement, sans que le responsable initial du traitement ne puisse y faire obstacle, et ce sous le contrôle exclusif de la personne concernée. Cette disposition qui permettra aux utilisateurs de migrer d'un écosystème de services à l'autre (concurrent ou non) avec leurs propres données pourrait leur permettre d'actionner une « portabilité citoyenne » au profit de missions d'intérêt général.

Des communautés d'utilisateurs pourraient exercer leur droit à la portabilité vis-à-vis d'un service pour mettre leurs données à disposition d'un acteur public, pour une finalité spécifique en lien avec une mission de service public. L'acteur public deviendrait responsable de traitement, et devrait donc respecter les principes de protection des données.

Un tel mécanisme aurait pour avantage de constituer des nouveaux jeux de données à usage de service public, sans imposer de nouvelles contraintes légales aux acteurs privés. Ce scénario aurait pour inconvénient la masse critique à atteindre, l'adhésion et la participation devant être conséquentes pour permettre la constitution de jeux de données pertinents. L'intégration de systèmes d'opt-in simplifiés, innovants et peu contraignants pourrait cependant favoriser la participation.

Un tel processus permettrait dans une vision plus prospective, d'aboutir à la création bottom-up d'un « commun » informationnel, construit par les individus au profit de l'intérêt général. Il s'agirait alors de construire les moyens de gouvernance de ce commun informationnel, par exemple par des « régies de données ».

Le rééquilibrage des forces entre les acteurs privés et publics sur la gestion de la ville, pour l'amélioration des politiques publiques, devrait s'accompagner pour la CNIL d'un encadrement renforcé de la collectivité publique, qui devra respecter le Règlement général à la protection des données (GDPR)[8] et notamment la notion de finalités légitimes dans la réutilisation des données qui lui seront restituées.

RÉGULER PAR LES COMMUNS ET UNE STRUCTURE DE GOUVERNANCE DÉDIÉE

Face aux injonctions contradictoires de la smart city – personnaliser tout en respectant la vie privée, optimiser sans rejeter – et pour répondre au bouleversement du jeu des acteurs, notamment avec l'arrivée des industriels de la donnée, il convient de produire de nouvelles formes de régulation de la donnée urbaine, dans le respect des individus et de leurs libertés.

"UN TEL PROCESSUS PERMETTRAIT DANS UNE VISION PLUS PROSPECTIVE, D'ABOUTIR À LA CRÉATION BOTTOM-UP D'UN 'COMMUN' INFORMATIONNEL, CONSTRUIT PAR LES INDIVIDUS AU PROFIT DE L'INTÉRÊT GÉNÉRAL. IL S'AGIRAIT ALORS DE CONSTRUIRE LES MOYENS DE GOUVERNANCE DE CE COMMUN INFORMATIONNEL, PAR EXEMPLE PAR DES « RÉGIES DE DONNÉES »."

8 Le Règlement général sur la protection des données (GDPR) est le nom couramment utilisé pour désigner le cadre réglementaire de l'Union européenne adopté en 2016 pour la protection des personnes physiques concernant le traitement des données personnelles et sur la libre circulation de ces données : http://data.europa.eu/eli/reg/2016/679/oj

Des propositions de modes de régulation innovants et efficaces sont intéressantes, par exemple la production et la gouvernance de communs de la ville, associées à la mise en place de structures nouvelles de gouvernance de ces données. L'adoption de mécanismes de cette nature apporterait en outre des outils intéressants pour la mise en conformité au règlement européen sur la protection des données (RGPD), par exemple par rapport à la notion centrale de consentement.

DÉFINIR DES COMMUNS

Dès 2014, Valérie Peugeot abordait la question des données de la smart city sous l'angle des communs, proposant de « déborder le cadre strict des données personnelles pour s'intéresser aux données numériques en général [...] en s'inspirant des travaux d'Elinor Ostrom [...] à développer une sphère de données en Communs, c'est-à-dire de données qui peuvent être considérées comme une ressource collective, et qui n'entrent ni dans le régime des biens gérés par la puissance publique stricto sensu, ni dans un régime de marché ». Ce régime de Communs repose sur une gestion par une communauté de la ressource considérée, qui organise ses règles de gouvernance, en s'appuyant sur un « faisceau de droits » (bundle of rights). Valérie Peugeot propose d'étendre ces communs aux données de la sphère publique, aux données produites en licence de partage (Wikipédia, Open Street Map, etc.), et à certaines données produites par des entreprises privées. Pour aller encore plus loin dans cette logique de production de communs, il faudrait in fine probablement y intégrer les données de référence de l'open data, les données d'intérêt général telles que définies par la loi République numérique et d'autres données d'intérêt général telles qu'elles pourraient être définies dans le futur par la loi. On peut par exemple penser à celles détenues par les industriels de la donnée, tel Waze, collectées dans le cadre d'un marché « données contre services » avec les utilisateurs.

Ces entreprises qui revendiquent œuvrer pour l'intérêt général cesseraient alors de limiter l'intérêt général à la somme des intérêts particuliers de leurs clients, pour réellement rendre ré-exploitable par la collectivité les données dont elles se nourrissent. Les recommandations présentées plus haut (étendre la notion d'intérêt général et activer des systèmes de portabilité citoyenne) pourraient permettre d'aller en ce sens.

Cette approche de communs et de dépassement des logiques de l'open data prennent forme depuis quelques années. Le CNNum (French Digital Council), dans un avis d'avril 2017 relatif à la libre circulation des données dans l'Union européenne propose des modalités de partage des données[9] : « Les États membres pourraient encourager différents acteurs à mettre en commun leurs données sur la base du volontariat, afin de concourir à un programme de recherche, un projet industriel ou à une politique publique, ponctuellement ou durablement. Les données mises en commun pourraient être collectées par un organisme public puis agrégées avant d'être réutilisées ou redistribuées. » Le rapport relatif aux données d'intérêt général, propose, pour les données du secteur privé, que l'on puisse invoquer le motif d'intérêt général pour la transmission obligatoire des données, notamment pour la conduite de politiques publiques sectorielles, l'information des citoyens et le développement économique. Si la puissance publique en est la seule destinataire, ou que la réutilisation est non commerciale, le droit de propriété n'est pas atteint. En cas de réutilisation commerciale, le rapport voit l'indemnisation comme la seule solution, afin de ne pas porter structurellement atteinte aux acteurs privés. Car c'est bien là l'un des enjeux de l'approche, aujourd'hui relativement conceptuelle, de communs : s'il y a un intérêt pour la somme des parties, le gain pour les acteurs qui sont aujourd'hui en position de force en ce qui concerne les données est plus incertain. L'objectif est donc d'arriver à maximiser la valeur pour la société dans son ensemble sans dissuader les acteurs à l'origine de la création de ces nouvelles données.

GOUVERNER LES COMMUNS, POUR MIEUX PROTÉGER LES DONNÉES PERSONNELLES

Constituer des communs urbains ne pourrait aller sans organiser les modes de gouvernances de ces données. Le CNNum, dans son avis, donne l'exemple sectoriel du US Bureau of Transportation Statistics, qui agrège les données des compagnies aériennes américaines concernant le trafic aérien. Mais d'autres vont plus loin avec la proposition de véritables acteurs tiers de confiance à l'échelle territoriale, un outil de gestion à gouvernance partagée, en mesure de faire respecter la conformité, notamment à la Règlementation sur la protection des données personnelles. C'est ce genre de modèle que propose Datact avec la Régie de données[10], un tiers acteurs, personne morale à gouvernance partagée entre la ville acteur public et les différentes parties prenantes de la ville – un véritable commun de la ville -, mais aussi un système d'information et de traitement de données par lequel il serait possible d'ouvrir et fermer, à la demande, les flux de données pour les différents acteurs qui les nécessiteraient. Ce tiers acteur œuvrerait pour l'organisation des flux de données entre différentes parties-prenantes, à la fois un hub et un point de contrôle de la licéité des échanges, du respect des licences applicables et de la protection des données personnelles, par la mise à disposition de moyens de recueil du consentement.

Un tel dispositif permettrait en outre de sortir d'une logique d'anonymisation par défaut des jeux de données de la ville. Il serait aussi possible, comme proposé par exemple dans le cadre du projet Open Algorithms[11], de permettre à certains acteurs d'utiliser des données sans les récupérer et dans le respect des droits des personnes concernées. Un tel type d'outil de gestion offrirait l'avantage d'ouvrir la donnée urbaine et de rééquilibrer le rapport de forces entre l'acteur public et les acteurs privés non soumis aux contrats publics. Il offrirait aux petites entreprises, collectifs, citoyens et associations qui le souhaitent de se réapproprier ces communs urbains, il permettrait surtout, pour les ré-utilisateurs qui souhaiteraient traiter des données personnelles, de demander le consentement explicite et éclairé des individus concernés.

9 CNNum, Avis du Conseil national du numérique sur la libre circulation des données dans l'union européenne, avril 2017, https://cnnumerique.fr/wp-content/uploads/2017/04/AvisCNNum_FFoD_VFinale.pdf

10 Concevoir une régie de données territoriales - Vers une nouvelle fabrique de services urbains, Dossier produit par Le hub agence et Chronos, Datact, La gazette des communes, mai 2014

11 http://www.opalproject.org