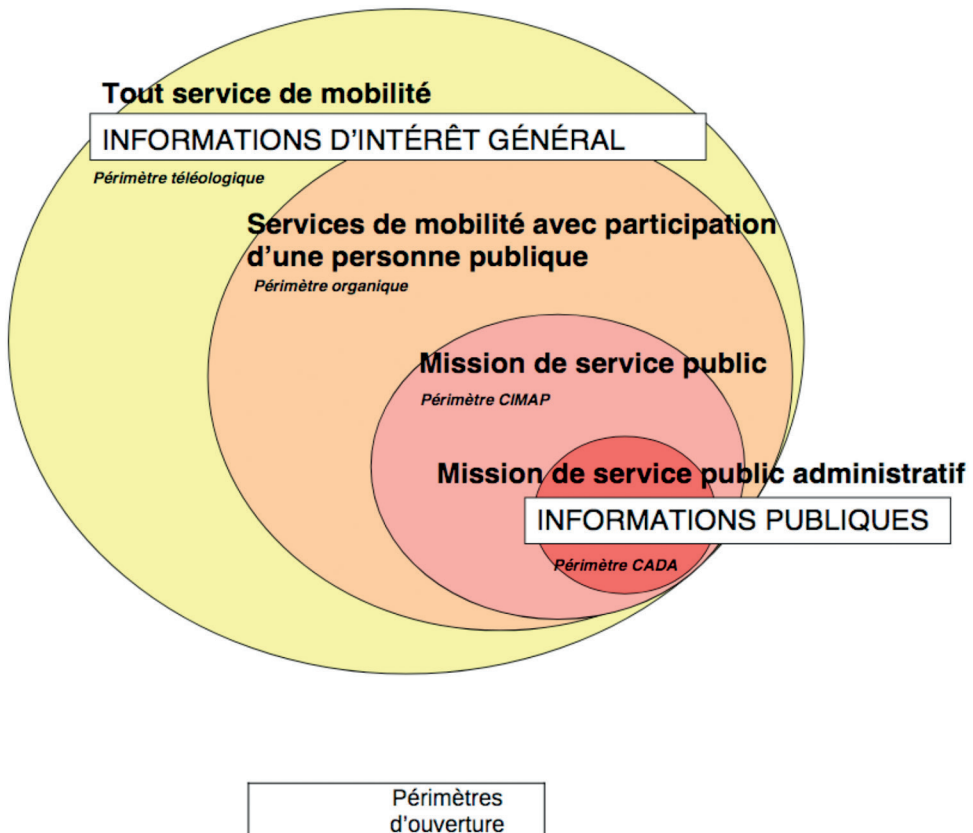


Comment définir et réguler les « données d'intérêt général » ?

Par Bertrand PAILHÈS

En 2015, les données de transport font l'objet d'un débat intense entre les entreprises de transport public (RATP, SNCF, Transdev, etc.), les nouveaux acteurs venus du numérique (Google, Uber mais aussi Citymapper) et les pouvoirs publics, qui souhaitent alors développer l'ouverture des données publiques.

En mars 2015, Francis Jutand remet son rapport sur l'ouverture des données de transport ⁽¹⁾, ouvrant la voie à des dispositions législatives qui seront intégrées dans la loi Macron promulguée en août de la même année. Il y définit les « informations d'intérêt général », qui rassemblent toutes « données (ou classes de données) des services de mobilité dont l'ouverture serait jugée opportune ». Y sont incluses des données de services publics mais également d'entreprises privées fournissant un service de mobilité.



Si les dispositions de la loi se sont finalement concentrées sur les acteurs délégataires d'une mission de service public, ce rapport fut le premier, en France, à élargir le débat de l'ouverture des données

(1) <https://cdn.nextinpact.com/medias/rapportjutand.pdf>

publiques (*l'open data*), en promouvant une vision complète des enjeux autour des données, dans un domaine où l'organisation des acteurs publics et privés est essentielle.

Par la suite, le rapport sur les « données d'intérêt général (DIG) » rendu en septembre 2015⁽²⁾ a étendu le cadre d'analyse juridique et économique à tous les domaines en identifiant les leviers et obstacles permettant l'accès à ces données. La communication de la Commission européenne sur l'économie de la donnée publiée en janvier 2017⁽³⁾ soulève également la question de l'accès aux données entre acteurs publics et privés. Enfin, la question du partage des données publiques et privées a figuré au centre du rapport Villani et de la stratégie française sur l'intelligence artificielle⁽⁴⁾.

Les DIG, pour quoi faire ?

L'exemple des transports est le meilleur exemple illustrant la manière dont les données doivent être collectées et partagées entre acteurs publics et privés. En effet, dans ce secteur :

- les données sont abondantes ;
- l'utilité de leur collecte massive est directement compréhensible pour traiter des problèmes quotidiens que sont les embouteillages, la pollution de l'air ou le fonctionnement des transports publics ;
- elles sont collectées par une grande variété d'acteurs, du plus privé (la voiture individuelle) au plus public (l'infrastructure en site propre sur domaine public, comme le métro), chaque acteur de la chaîne étant relié à ces deux dimensions ;
- le secteur est l'objet de l'attention de nouveaux acteurs perçus comme puissants, notamment les grandes plateformes numériques qui ont compris que la gestion du transport quotidien est une application plébiscitée par les utilisateurs, qu'ils soient en voiture (Waze) ou en transport en commun (Citymapper, Google Maps).

L'enjeu de l'accès aux données est donc double : d'une part développer une activité économique en soutenant l'innovation et en évitant la constitution de rentes et, d'autre part, améliorer la conduite des politiques publiques.

Du point de vue économique, la donnée est en effet un bien non rival, dont l'usage par une personne n'en prive pas les autres : le coût marginal de reproduction étant quasi nul, il est possible, à partir d'une même donnée, d'en multiplier les usages et les utilisateurs et, partant, de maximiser la valeur potentielle qu'il est possible de tirer de cette donnée. Encourager le partage de données entre acteurs est donc d'abord un moyen de développer l'innovation et de garantir une exploitation économiquement efficace de cette ressource que forment les données.

Par ailleurs, la maîtrise des données importantes pour l'économie et la société est un enjeu croissant identifié par les pouvoirs publics pour constituer les entreprises et services leaders de demain, dans l'ensemble des secteurs économiques. Ainsi, vouloir développer une industrie européenne de l'intelligence artificielle⁽⁵⁾, qui repose nécessairement sur de larges quantités de données, nécessite de traiter la question de l'accès aux données détenues par les entreprises ou les administrations quand celles-ci en ont le monopole⁽⁶⁾.

Du point de vue politique, le débat sur les données d'intérêt général rejoint la vision du numérique comme « commun », c'est-à-dire comme un ensemble de ressources n'appartenant à personne et

(2) <https://cdn2.nextinpact.com/medias/rapport-cytermann.pdf>

(3) Communication de la Commission européenne du 10 janvier 2017 « Créer une économie européenne fondée sur les données », http://europa.eu/rapid/press-release_IP-17-5_fr.htm

(4) <https://www.aiforhumanity.fr/>

(5) Voir notamment le rapport Villani.

(6) Par exemple, dans le cas des entreprises privées, les données de navigation Internet détenues par Google ou Facebook et, dans le cas des administrations françaises, les données de santé.

utilisable par tous, à condition d'en fixer des règles de gouvernance partagées : le logiciel libre, les protocoles Internet ou Wikipédia en sont les exemples les plus emblématiques. La donnée constitue un des enjeux les plus importants des prochaines années pour porter cette vision, qui se rapproche de celle portée par la communauté scientifique pour le savoir scientifique. L'objectif est d'accroître la transparence et ainsi le contrôle des entités détentrices des données pour, *in fine*, garantir la confiance des utilisateurs dans les outils numériques.

Les DIG, quelle définition ?

Les données d'intérêt général (DIG) sont donc des données, publiques ou privées, dont le partage et l'ouverture sont « d'intérêt général », c'est-à-dire d'un intérêt plus large que le seul intérêt du détenteur des données.

Il est essentiel de distinguer cette notion de celle de « *l'open data* » où le législateur poursuit un double objectif de transparence et d'innovation en ouvrant l'accès aux données publiques et en limitant au maximum les conditions de réutilisation (licence, prix, etc.), sans vérifier *a priori* la qualité de la réutilisation. *L'open data* trouve ainsi sa source dans le droit constitutionnel du citoyen à connaître de l'administration ; ce n'est pas le cas des DIG, dont l'accessibilité doit pouvoir être soumise à des conditions particulières et ne constitue pas un droit fondamental.

Il convient également de distinguer les DIG de l'accès administratif aux informations détenues par les personnes publiques ou privées : pour collecter l'impôt, l'administration doit accéder à certaines informations, tout en garantissant la confidentialité. Les données constituent un nouveau champ d'application de cette prérogative de puissance publique.

Dans les deux cas précédents, une raison impérative, d'intérêt général et imposée par le législateur, définit le régime d'accès aux données, et il ne semble pas nécessaire de développer de nouveaux concepts pour mettre en œuvre cet accès.

Les DIG, à l'inverse, peuvent constituer un outil pour faciliter le partage de données dans des cas plus complexes, où la prise en compte des spécificités du cas est nécessaire, sous réserve de lever les obstacles techniques, juridiques ou économiques.

Les DIG, quels obstacles ?

Obliger une personne détenant des données à les rendre accessibles à des tiers suppose d'abord de tenir compte des contraintes juridiques qui peuvent empêcher cet accès.

Contrairement à une idée répandue, il n'existe pas aujourd'hui de droits de propriété sur les données détenues par une personne : des droits de propriété intellectuelle s'appliquent aux bases de données (droit *sui generis*) mais ils ne s'étendent pas aux données elles-mêmes. Pour autant, le détenteur des données peut faire valoir des secrets légaux comme la protection des données personnelles, ou des obligations comme la sécurité informatique pour justifier une limitation d'accès aux données. Ces considérations, légitimes, doivent être prises en compte dans les modalités d'accès aux données, au travers par exemple des API⁽⁷⁾ et des procédures de sécurité comme la traçabilité des accès.

De telles règles d'ouverture pourraient également limiter la liberté d'entreprendre de l'entreprise détentrice. Si ce principe constitutionnel est important, il est toutefois possible, en France, d'en limiter les effets à condition de justifier d'un motif d'intérêt général qui surpasse la protection de cette liberté. La question de la proportionnalité est ainsi centrale quand il s'agit de définir l'accès aux DIG.

(7) Application Programming Interface : interface standardisée d'accès à des informations ou des fonctionnalités.

Sur un plan économique, malgré le caractère non rival de la donnée, l'obligation de rendre des données accessibles à des tiers peut réduire l'incitation du détenteur de la donnée à investir, dès lors qu'il sait qu'il sera contraint de partager la donnée et n'en aura pas le monopole. C'est pourquoi l'accès aux DIG ne doit pas nécessairement être gratuit mais fixé à un prix qui maintient l'incitation du détenteur à investir, voire constitue même un revenu complémentaire pour celui-ci. Bien évidemment, ce prix peut également dépendre de la nature du demandeur, un laboratoire de recherche publique n'ayant pas la démarche commerciale d'une *start-up*.

L'ouverture des DIG s'entend donc uniquement dans un environnement technico-économique maîtrisé par le détenteur, qui doit être en mesure de spécifier les conditions techniques d'accès aux données (format, interfaces ou API, et mesures de sécurité), mais également les conditions d'utilisation des données ainsi récupérées, leur utilisation étant, de toute façon, soumise aux règles de droit commun applicables (données personnelles, réglementations sectorielles).

Dès lors, comment organiser l'accès aux données d'intérêt général ?

Écartons d'emblée la fausse piste d'un droit de propriété intellectuelle associée à la donnée. Cette idée est séduisante par son analogie avec d'autres biens immatériels (brevets, œuvres) et semble permettre d'envisager une régulation « économique » de l'accès aux DIG : si un « propriétaire » de données a un intérêt économique à les ouvrir en les valorisant, il le fera par le jeu du marché. Mais de même que certaines infrastructures sont plus efficacement construites et entretenues par la puissance publique sans appropriation par un acteur particulier, certaines données, de nature stratégique ou indispensables pour comprendre et résoudre certains problèmes sociaux ou économiques, ne doivent pas être exclusivement détenues par un acteur.

L'exemple de la voiture connectée illustre cette impasse de la propriété : si le constructeur est « propriétaire », assureurs, sociétés d'autoroute, réparateurs, calculateurs de mobilité mettront en place de multiples dispositifs pour collecter directement les données plutôt que de risquer un prix trop élevé de la part du constructeur. Il semble ainsi beaucoup plus pertinent de prévoir que les données de l'automobile seront accessibles à tous ces acteurs, dans des conditions différenciées selon la nature du besoin de chacun.

Se pose ensuite la question de la définition d'un régime général : peut-on construire juridiquement cette notion de DIG sans s'accrocher à des considérations sectorielles (transport, logement, écologie) pour lesquelles il semble plus simple de qualifier à la fois les objectifs poursuivis et les données visées ? Le premier rapport de septembre 2015 sur les DIG invitait les pouvoirs publics à suivre cette voie, hormis dans deux cas : l'accès de la statistique publique aux données des entreprises dans le cadre d'enquêtes⁽⁸⁾ et l'extension de l'ouverture des données publiques aux entreprises privées ayant une mission de service public⁽⁹⁾. À la suite de la loi Macron de 2015, la loi pour une République numérique d'octobre 2016 a ainsi poursuivi l'accès à certaines données spécifiques, dans les domaines de l'énergie ou des vitesses maximales autorisées. La loi Santé a également engagé le mouvement de partage et d'ouverture des données de santé avec la création du système national des données de santé.

Il semble toutefois possible de développer un modèle général de la donnée « d'intérêt général », en répondant aux quatre questions suivantes :

- Quelle est la donnée visée, ou le type de donnée, et qui est son détenteur ?
- Qui demande l'accès aux données ?
- Pour quelle finalité une personne souhaite-t-elle accéder aux données ?
- Selon quelles modalités l'accès est-il fourni ?

(8) Introduit dans l'article XX de la loi pour une République numérique.

(9) Déjà prévu dans la loi CADA régissant l'accès aux documents administratifs.

Les cas où il suffit de répondre à une seule de ces questions pour obtenir l'accès à des données sont rares⁽¹⁰⁾ et c'est en combinant ces critères que l'accès d'un tiers à des données pourra être permis. Ainsi, un laboratoire de recherche publique pourrait faire valoir sa qualité pour accéder à des données détenues par une entreprise mais il devra justifier *a minima* de modalités sécurisées d'accès. De même, des données de transport d'une compagnie de taxis pourraient être *a priori* considérées comme d'intérêt général mais l'étendue et les conditions d'accès pourraient en être différenciées en fonction de la nature du demandeur et de la finalité qu'il poursuit : l'accès de la Mairie de Paris n'est pas le même que celui d'un service de calculateur d'itinéraires.

Pour traiter cette complexité casuistique, un régime général pourrait définir deux principes de régulation :

- la nécessité de faire prévaloir le contrat entre les parties sur toute intervention publique ;
- l'obligation de faire droit aux demandes d'accès raisonnables pour les détenteurs de DIG⁽¹¹⁾.

Le premier principe est évident quand on constate que le partage des données n'est pas une idée nouvelle dont les pouvoirs publics auraient le monopole : de nombreuses entreprises ont l'idée de mettre en commun leurs données ou d'en organiser le partage afin de permettre à des tiers d'imaginer de nouvelles applications ou de conduire des recherches. Récemment, Transdev a lancé une plateforme, « Catalogue », dont l'objectif était de rassembler les données de tous les opérateurs de transport, de BlaBlaCar à la SNCF. Uber a lancé « Uber Movement » pour permettre à des tiers d'accéder aux données de ses taxis.

Ces exemples montrent la prise de conscience des acteurs économiques, qui utilisent également les données pour construire un écosystème d'autres acteurs économiques dépendant d'eux. Mais ils montrent aussi le besoin de régulation : l'initiative de Transdev n'a pas réussi à convaincre la SNCF et la RATP de fournir leurs données et Uber est aujourd'hui seul à choisir quelles sont les données qu'il ouvre.

La puissance publique a donc un premier rôle d'arbitre à endosser, pour permettre l'accès de tiers, « dans des conditions raisonnables », qui peuvent être techniques, juridiques et économiques. Ces conditions tiendraient également compte des quatre questions mentionnées ci-dessus et notamment de la nature du demandeur et de la finalité poursuivie.

Afin de renforcer les possibilités d'accès, notamment aux données des acteurs privés, les pouvoirs publics devront également établir *a priori* les critères de définition des DIG (en les déclinant éventuellement par secteur). Parmi ces critères généraux, on peut évoquer l'importance des données pour un enjeu de politique publique (sécurité, santé, environnement, etc.), le caractère unique ou universel du jeu de données (notamment pour le développement économique) ou encore la contribution des utilisateurs à la constitution du jeu de données (dans le cas de Waze par exemple). Sur la base de ces critères, l'autorité publique engagerait une procédure transparente et contradictoire visant à déclarer « d'intérêt général » un ensemble de données et ouvrant ainsi droit aux demandes d'accès.

Ces pistes, déjà explorées dans les travaux cités en introduction de cet article, gagneraient enfin à être harmonisées au niveau européen afin d'éviter tout risque d'évasion réglementaire des acteurs concernés.

Le besoin de régulation de l'accès aux données publiques et privées « d'intérêt général » est de plus en plus important et il semble donc possible de définir un cadre simple, flexible et générique qui préserve les intérêts des parties tout en garantissant une circulation et une ouverture accrue des données.

(10) C'est le cas de la statistique publique (finalité qui permet un accès généralisé), de l'autorité judiciaire (qui a un pouvoir général d'enquête) ou de l'ouverture des données publiques (la nature du détenteur permet l'accès inconditionnel).

(11) Cette proposition est directement inspirée de l'obligation d'accès imposée à l'opérateur historique dans le secteur des télécommunications.