

Publication des Annales des Mines
avec le concours
du Conseil général de l'Économie
et de l'École de Paris
du management

La Gazette

N° 129 • BIMESTRIEL
NOVEMBRE 2024

DE LA SOCIÉTÉ
ET DES TECHNIQUES

« Se défier
du ton d'assurance
qu'il est si facile
de prendre
et si dangereux
d'écouter »

Charles Coquebert
Journal des mines n°1
Vendémiaire An III (1794)

Ouverture des données publiques *Le mythe de l'abondance et la réalité du labour*

En 2016, la loi pour une République numérique impose aux administrations de publier leurs données « par défaut », si elles ne contiennent pas d'informations sensibles. Cette politique ambitieuse a propulsé la France au sommet des classements internationaux de l'*open data*. Pourtant, chercheurs, journalistes, étudiants ou simples citoyens regrettent encore souvent l'absence ou la mauvaise qualité des données publiques. Constatant cet écart entre le cadre législatif très ambitieux et la persistance de difficultés d'accès, nous avons voulu interroger le caractère opérationnel d'une mise à disposition automatique des données publiques. Ce principe du « par défaut » est-il possible et souhaitable ? N'existe-t-il pas une autre voie qui permette une ouverture plus efficace des données publiques ? C'est la question que nous avons étudiée dans le cadre de notre mémoire de fin de formation du Corps des Mines.



L'ouverture par défaut des données publiques signifie que les documents administratifs contenant des données publiques doivent être systématiquement mis en ligne. Il s'agit d'un réel changement de paradigme pour les administrations car auparavant la publication relevait de l'exception et non de la règle. Depuis 2016, elles doivent produire et publier des données dites « de référence » pour l'identification de services, de territoires ou de personnes comme le cadastre ou la base Sirene. Elles doivent recenser les documents dans lesquels figurent des informations publiques dans des répertoires des informations publiques (RIP). Enfin, les données publiées doivent être gratuites, dans des formats standards et aisément réutilisables¹.

Si certaines administrations créent leurs propres portails de données, la clé de voûte de l'ouverture est le site data.gouv.fr de la Direction interministérielle du numérique (DINUM). Cette plateforme centralise près de 38 000 jeux de données publiques et permet aux citoyens de poser des questions aux administrations productrices et de publier leurs réutilisations de données, comme le très fameux

CovidTracker qui agrégeait les données relatives à l'épidémie de Covid-19.

Une loi inégalement appliquée

La France fait la course en tête des classements mondiaux de l'*open data* : 1ère en Europe et 2e au monde en 2023. Pourtant, elle figure seulement à la 108^e place du classement mondial du droit d'accès aux documents administratifs. Comment peut-on être à la fois excellent pour publier spontanément ses données et très mauvais pour permettre aux gens d'y accéder ?

Cela s'explique notamment par la méthodologie des classements qui n'évalue que le contenu des lois et non leur application. Et dans la pratique, la loi pour une République numérique n'est que très inégalement appliquée. Si certaines mesures ont été mises en œuvre, comme la publication des jeux de données de référence, d'autres tardent à l'être. Sur 40 000 demandes de RIP effectuées par le collectif citoyen Ma Dada, seules 33 administrations ont transmis

leur répertoire. Seules 16 % des collectivités territoriales appliquent le principe d'*open data* par défaut et les documents publiés sont difficilement réutilisables avec par exemple des tableaux de chiffres transmis au format PDF.

Une ouverture adossée à un droit d'accès aux documents administratifs défaillant

En 1978, la France crée le droit d'accès aux documents administratifs. Ce droit diffère de l'*open data* dans la mesure où il ne s'agit pas d'une ouverture spontanée : le citoyen doit faire une demande auprès de l'Administration. Mais cette démarche peut être longue : les administrations répondent rarement, il faut alors passer devant la Commission d'accès aux documents administratifs (Cada) qui rend un avis non contraignant, peu pris en compte par les administrations. Si l'avis n'est pas suivi, le demandeur peut saisir le tribunal administratif, voire le Conseil d'État. Il semble donc prématuré d'imposer un principe d'ouverture par défaut alors même que l'accès aux documents administratifs s'apparente encore souvent à un parcours du combattant.

Les enjeux de l'*open data*

En ne respectant pas pleinement ce principe du « par défaut », les administrations se privent et privent les citoyens et les entreprises de nombreuses opportunités.

Un enjeu démocratique

Dès le lendemain de son investiture, le président Obama s'engage à ouvrir largement les données publiques. Il souhaite ainsi regagner la confiance des citoyens américains, abîmée par la crise financière de 2008. Il présente cette ouverture comme essentielle à la démocratie et source d'efficacité pour le gouvernement. En 2016, le candidat Trump souhaite revenir sur les décrets de son prédécesseur. Un vent de panique se lève parmi les chercheurs et les activistes qui téléchargent en masse les données publiques disponibles, craignant qu'elles ne soient supprimées. Ces derniers sont conscients que l'accès à ces données leur confère une capacité de contre-expertise, un contre-pouvoir essentiel à la démocratie. En France, le travail du journaliste Alexandre Léchenet qui a remis en question la cartographie de l'éducation prioritaire grâce aux données IPS (indice de position sociale) des établissements scolaires, est un exemple de cette contre-expertise citoyenne permise par les données publiques.

Un enjeu économique

En 2019, une étude de la Commission européenne affirme que le marché de la réutilisation de la donnée publique frôle les 200 milliards d'euros. Les exemples concrets de développements économiques alimentés par la donnée publique sont nombreux. En ouvrant ses données, l'entreprise publique qui gère les transports londoniens a permis la création de plus de 600 applications utilisées par 42 % des Londoniens. L'entreprise française Doctrine, qui propose un moteur de recherche juridique, génère quant à elle 80 % de ses revenus grâce aux données publiques. Les données publiques sont aussi essentielles au développement de modèles d'intelligence artificielle

francophones, aujourd'hui peu adaptés car entraînés sur un corpus anglo-saxon.

Un enjeu d'amélioration des services publics

Une politique d'*open data* sert aussi les administrations. Sur les forums de data.gouv.fr, les retours des réutilisateurs permettent d'améliorer la qualité des données. L'*open data* encourage aussi une meilleure circulation des données en interne : en publiant ses données, une administration les met aussi à disposition des autres administrations. Il permet également de valoriser l'action des services, à travers des initiatives comme le « baromètre des résultats de l'action publique ». Enfin, il crée des liens entre les agents et les citoyens, offrant à travers les données un terrain d'échange propice à l'amélioration de la qualité des services publics.

Les difficultés sous-estimées de l'*open data*

Les difficultés présentées par l'ouverture des données ont été sous-estimées par les législateurs, expliquant en large partie la faible application du principe d'ouverture par défaut.

L'ouverture des données n'est pas donnée !

Ouvrir ses données ne se fait pas d'un claquement de doigts. La création d'un portail *open data* ou la publication sur data.gouv.fr impliquent un certain nombre d'étapes techniques : identification des données à publier, extraction des données de logiciels métier rarement prévus pour faire de l'*open data*, choix d'une solution de portail et d'hébergement, procédure de diffusion et de mise à jour des données, etc. Autant de tâches et de sommes engagées pour les administrations ! Pour créer un portail *open data*, une commune de 50 000 habitants devra y consacrer six mois de travail et entre 50 000 et 100 000 euros. Si l'on prend en compte l'effort de mobilisation des équipes en interne, la facture double ou triple. Faire tourner le portail représente aussi un coût, qui peut aller de quelques milliers d'euros par an pour une petite administration, à quelques dizaines, voire centaines, de milliers d'euros pour les portails les plus gros. Si cela représente généralement moins de 0,1 % du budget d'une administration, il s'agit d'un arbitrage budgétaire auquel toutes les administrations ne sont pas forcément prêtes à consentir.

Oublions la métaphore du pétrole

Axelle Lemaire, ancienne secrétaire d'État chargée de l'élabo-ration de la loi pour une République numérique, affirmait : « Il faut penser données comme auparavant on pensait pétrole ». L'expression « par défaut » impliquerait donc que les données publiques préexistent à leur ouverture sous une forme directement réutilisable, comme des ressources naturelles gisant dans les administrations, un nouveau « pétrole » qu'il s'agirait simplement d'extraire et de présenter aux citoyens. Mais cette analogie est trompeuse. Pour ouvrir les données, il faut d'abord établir un inventaire pour identifier celles qu'il est possible et souhaitable d'ouvrir. Une fois cette étape réalisée, ces données ne sont toujours pas prêtes à l'emploi. Elles exigent encore un travail de « brutification », c'est-à-dire de transformation pour les rendre exploitables par le public. Les données s'apparentent bien davantage à un biocarburant

« Comment peut-on être à la fois excellent pour publier spontanément ses données et très mauvais pour permettre aux gens d'y accéder ? »



nécessitant un travail laborieux de culture, qu'à un pétrole prêt à jaillir au premier coup de pioche.

Un concept source d'incertitudes juridiques

La loi liste les données sensibles qui doivent être protégées, notamment en matière de vie privée, de secret médical, de Défense nationale, etc. L'ouverture par défaut concerne tous les documents administratifs sauf ceux qui contiennent de telles données. Mais cette règle est moins simple et claire qu'elle n'en a l'air. Il n'est pas toujours aisé de juger du caractère sensible d'une donnée, qui fait parfois l'objet de débats entre des juridictions comme la CADA et les tribunaux administratifs. Certaines données qui devraient être publiques peuvent ne pas être publiées car les producteurs craignent des réutilisations malveillantes. C'est par exemple le cas des tracés des réseaux d'eau. *A contrario*, un document qui contient des données sensibles peut être publié si l'administration les occulte. Mais ce travail laborieux n'est pas sans risques : quelles conséquences si l'administration oublie d'occulter une donnée ? Jusqu'où occulter les données quand la richesse des informations disponibles en ligne permet facilement à partir de croisements de réidentifier des individus ? Pour ne pas se confronter à ces difficultés, les agents préfèrent souvent « oublier » de publier leurs données.

La transparence, boussole de l'ouverture des données

Au début des années 2000, une partie de la société civile se mobilise aux États-Unis et au Royaume-Uni pour réclamer une ouverture des données publiques comme un outil

démocratique. Elle rappelle que ces données lui reviennent de droit dans la mesure où elles ont été produites avec de l'argent public. Au Royaume-Uni, le journal *The Guardian* alimente le débat en lançant la campagne médiatique "Free our data" demandant la gratuité de ces « joyaux de la couronne ».

Cette demande de transparence est rapidement prise en compte par les gouvernements américain et britannique. En 2009, l'administration Obama lance le premier portail national de données publiques, suivi par le gouvernement Cameron et son portail data.gov.uk. Washington et Londres tentent ensuite d'exporter ce modèle en créant l'Open Government Partnership en 2011, marquant un tournant vers une plus grande transparence au niveau mondial.

Alors que les politiques d'*open data* se structurent outre-Atlantique et outre-Manche, la société civile française se saisit de la question, avec la création d'associations comme Regards citoyens, à l'origine du site NosDéputés.fr qui publie les données relatives aux activités parlementaires. L'appel est entendu par le gouvernement Sarkozy qui crée le portail national data.gouv.fr en 2011.

Cependant, cette exigence de transparence s'est progressivement effacée au profit des enjeux économiques de l'ouverture des données. En 2020, dans sa stratégie pour les données, l'Union européenne affirme son ambition de devenir *leader* dans un monde "data-driven". Elle décline cette stratégie à travers une série de directives, le Data Governance Act en 2022 et le Data Act en 2024, qui créent le cadre d'un

marché unique de la donnée. Il s'agit désormais d'encourager les administrations à ouvrir leurs données en leur faisant miroiter des retombées économiques en cascade.

Mais cette nouvelle approche, faisant de l'économie la carotte de l'*open data*, suscite des interrogations. Les études sur les retombées économiques de l'*open data* manquent souvent de rigueur, avec des méthodologies opaques et des résultats divergents. Or si les attentes en termes de croissance ne sont pas satisfaites, le risque de déception et de démobilisation est réel. Pour éviter cela, il est crucial de rappeler que l'ouverture des données publiques trouve avant tout son sens dans l'exigence de transparence, un fondement indispensable à toute politique d'*open data*.

Pour une politique de la données renouvelée

Pour relancer la dynamique d'ouverture, il faut la recentrer sur son objectif premier : la transparence. Cette réorientation est nécessaire pour que l'ouverture des données devienne un véritable service public, et pour redonner un sens à l'action des agents publics. La transparence doit guider chaque décision d'ouverture, en mettant l'accent sur deux piliers : l'anticipation des besoins et la réponse aux demandes des citoyens.

(1) Anticiper les besoins implique que les administrations ouvrent spontanément les données jugées utiles. Certaines, comme l'agglomération de La Rochelle, adoptent déjà cette approche proactive. Avant même que la loi les y contraigne, elles ont identifié les besoins des utilisateurs, par exemple la disponibilité de vélos en libre-service, pour ouvrir les données pertinentes. La Rochelle va plus loin en prêtant des données non publiques lorsqu'un intérêt général est identifié, comme ce fut le cas pour une start-up développant un outil de stationnement. Les attentes locales sont ainsi mieux satisfaites et un climat de confiance se crée entre producteurs de données et réutilisateurs potentiels.

Pour généraliser cette approche, nous pensons souhaitable que les administrations s'appuient davantage sur le réseau des Animateurs Territoriaux des Données, qui jouent un rôle clé dans la formation, la normalisation et la mise en commun des bonnes pratiques en matière d'*open data*. Le financement de ces initiatives doit être assuré, idéalement par des budgets dédiés ou, à défaut, par des fonds de relance et de transformation comme le plan « France 2030 ». Enfin, il est indispensable de maintenir dans la loi le principe d'*open data* par défaut, qui, en faisant de l'ouverture la règle, décomplexifie l'ouverture pour les agents.

(2) Le second pilier repose sur la capacité à répondre aux demandes d'accès aux documents administratifs. Il s'agit d'une obligation légale, mais aussi d'un moyen d'identifier les besoins et d'améliorer les services publics en conséquence. Une administration ouverte est une administration qui ne perçoit pas ces demandes comme des attaques, mais comme des opportunités d'amélioration, et entretient des liens avec des réutilisateurs de données en créant et animant de véritables communautés autour de leurs données, comme le fait l'IGN autour de ses « géocommuns ».

Pour que cela soit efficace, il est nécessaire de former les citoyens, et en particulier les journalistes, sur leurs droits en matière d'accès aux documents administratifs, trop souvent méconnus alors qu'ils constituent une ressource précieuse pour l'investigation et la participation citoyenne. Des modules de formation en école de journalisme pourraient pallier ce manque d'information et permettre aux journalistes d'exploiter pleinement les possibilités offertes par la loi. Il est également crucial de renforcer la CADA, en lui donnant le pouvoir de sanctionner les administrations qui refusent de communiquer des documents sans raison valable, tout en maintenant son rôle d'accompagnement et de conseil.

*

Malgré l'ambition de la loi pour une République numérique de 2016, l'*open data* en France peine à s'imposer dans les pratiques administratives. Loin du simple coup de pioche dans un gisement de pétrole, l'ouverture des données publiques est un défi technique, juridique et démocratique, dont la difficulté a été sous-estimée. Pour surmonter ces difficultés, nous sommes convaincus de l'importance de remettre les objectifs de transparence et d'amélioration des services publics au cœur de la politique d'ouverture. La France pourra alors faire des données publiques un véritable moteur de progrès et de confiance citoyenne².

Paul Durliat et Amélie Latreille,
ingénieurs des mines

NOTES

1. Le terme « réutilisateur » traduit l'idée que la donnée est utilisée une première fois par l'administration avant d'être réutilisée par la société civile.
2. Pour en savoir plus : DURLIAT P. & LATREILLE A. (2024), *Ouverture des données publiques : le mythe de l'abondance et la réalité du labeur*, mémoire de troisième année du Corps des Mines, <https://minesparis-psl.hal.science/hal-04674768>

La Gazette de la Société et des Techniques

La Gazette de la Société et des Techniques a pour ambition de faire connaître des travaux qui peuvent éclairer l'opinion, sans prendre parti dans les débats politiques et sans être l'expression d'un point de vue officiel. Elle est diffusée par abonnements gratuits. Vous pouvez en demander des exemplaires ou suggérer des noms de personnes que vous estimez bon d'abonner.

Vous pouvez consulter tous les numéros sur le web à l'adresse :

<https://www.annales.org/gazette.html>

RENSEIGNEMENTS ADMINISTRATIFS

Dépôt légal Novembre 2024

La Gazette de la Société et des Techniques

est éditée par les *Annales des Mines*
120, rue de Bercy – télédéc 797 – 75012 Paris
<https://www.annales.org/gazette.html>
Tél. : 01 42 79 40 84 – Mél. : michel.berry@ecole.org
N° ISSN 1621-2231

Directeur de la publication : Grégoire Postel-Vinay

Rédacteur en chef : Michel Berry

Illustrations : Véronique Deiss

Réalisation : Alexia Kappellmann - Annales des Mines

Impression : service de reprographie du ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique

