



HAL
open science

Ouverture des données publiques : le mythe de l'abondance et la réalité du travail

Paul Durliat, Amélie Latreille

► **To cite this version:**

Paul Durliat, Amélie Latreille. Ouverture des données publiques : le mythe de l'abondance et la réalité du travail. Sciences de l'Homme et Société. 2024. hal-04674768

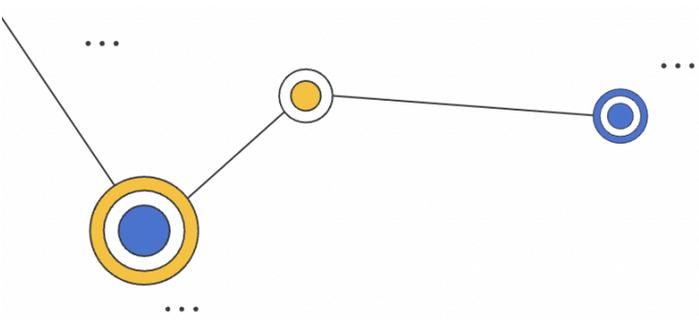
HAL Id: hal-04674768

<https://minesparis-psl.hal.science/hal-04674768v1>

Submitted on 21 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Paul Durliat ◉ Amélie Latreille

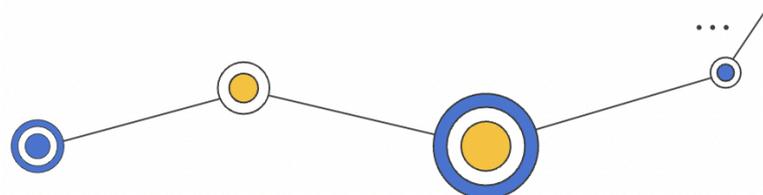
OUVERTURE DES DONNÉES PUBLIQUES

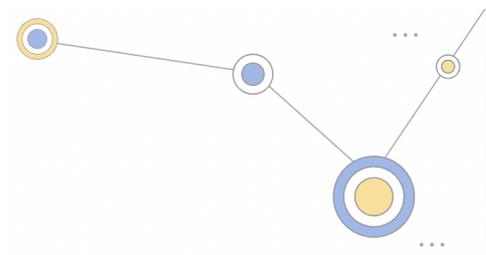
Le mythe de l'abondance et la réalité du labeur

Mémoire de fin de formation des ingénieurs du Corps des mines

Pilote : Olivier Baujard

2024





Résumé exécutif

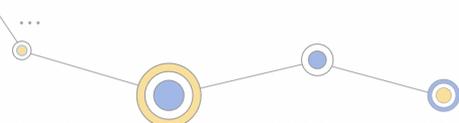
En 2016, la loi pour une République numérique introduit le principe d'**ouverture des données publiques par défaut**, obligeant les administrations à ouvrir automatiquement leurs données dès lors que celles-ci ne comportent pas d'informations sensibles.

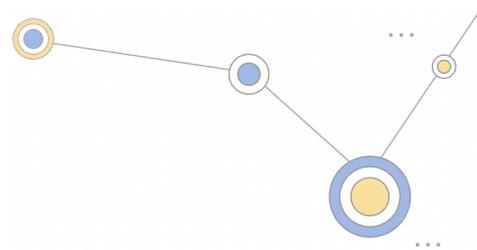
Cette politique ambitieuse porte la France en tête des classements internationaux de l'*open data*. **Pourtant, l'ouverture des données publiques est encore loin d'être une réalité ancrée dans les pratiques des administrations.** Comment est-ce possible ?

Si l'expression « par défaut » semble suggérer que les données publiques préexistent à leur ouverture sous une forme directement publiable, comme un pétrole brut prêt à être extrait, leur publication est tout sauf une simple formalité. Coûts d'identification, de structuration et de formatage, mobilisation des agents, culture du secret, incertitudes juridiques, etc. : les difficultés sont nombreuses et c'est parce qu'on les sous-estime que la loi ne peut être pleinement appliquée.

Il faudra avant tout convaincre les administrations de publier leurs données. Aujourd'hui, on agite surtout le potentiel de création de valeur des données ouvertes. Mais ce potentiel nous semble trop incertain et trop lointain pour mobiliser les administrations dans la durée. **Pour redonner du sens à l'ouverture des données, il faut assumer de ne pas poursuivre une logique économique et de porter la transparence comme objectif premier de l'ouverture des données publiques.**

Enfin, le volume considérable de données rend nécessaire une ouverture ciblée. Servir l'objectif de transparence implique de ne pas laisser le choix des données à ouvrir aux seules administrations. Nous sommes convaincus que la solution réside dans le renforcement du **droit d'accès aux documents administratifs** - qui n'est pas garanti aujourd'hui - et la publication des données ayant fait l'objet d'une demande des citoyens.

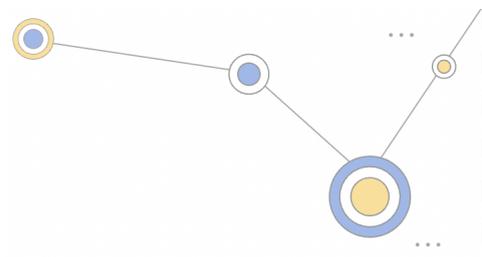




Sommaire

Liste des sigles utilisés.....	3
Introduction.....	5
1. Une ouverture incomplète et des opportunités manquées.....	7
1.1. Un cadre législatif ambitieux mais inégalement appliqué.....	8
1.1.1. Le cadre législatif de l'ouverture des données publiques.....	8
1.1.2. Les difficultés de la mise en oeuvre.....	13
1.1.3. Un droit d'accès aux documents administratifs défaillant.....	18
1.2. Les enjeux et les opportunités de l'ouverture des données publiques.....	21
1.2.1. Un enjeu démocratique.....	21
1.2.2. Un enjeu économique et de souveraineté.....	30
1.2.3. Un enjeu d'amélioration des services publics.....	37
2. Les difficultés sous-estimées de l'open data et de l'open data par défaut.....	45
2.1. Des efforts conséquents mais nécessaires pour ouvrir ses données.....	46
2.1.1. Sur le plan technique : des systèmes d'information qui n'ont pas été pensés pour l'ouverture des données.....	46
2.1.2. Sur le plan des ressources humaines : du temps et des compétences qu'il manque trop souvent aux administrations.....	50
2.1.3. Sur le plan financier : des coûts en plus mais des recettes en moins.....	55
2.2. Une clarification indispensable de la notion de « par défaut ».....	63
2.2.1. Une notion floue et inopérante.....	63
2.2.2. Un concept source d'incertitudes juridiques.....	65
3. Redonner un sens à l'ouverture pour relancer la dynamique.....	73
3.1. Une politique d'ouverture qui n'a pas toujours servi les mêmes objectifs.....	74
3.1.1. Aux origines de l'open data, une exigence de transparence.....	74
3.1.2. Un glissement vers la mise en avant d'objectifs économiques pour justifier l'open data.....	80
3.2. Redynamiser l'ouverture en réaffirmant son objectif premier : la transparence.....	86
3.2.1. Pour une politique d'ouverture des données guidée par l'objectif de transparence et reposant sur deux piliers : anticiper les besoins et répondre aux demandes.....	86
3.2.2. Pour une Cada renforcée.....	91
Conclusion.....	98
Table des figures.....	99
Table des tableaux.....	100
Bibliographie.....	101
ANNEXE : Synthèse des propositions.....	106

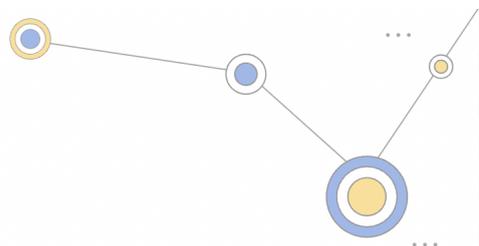




Liste des sigles utilisés

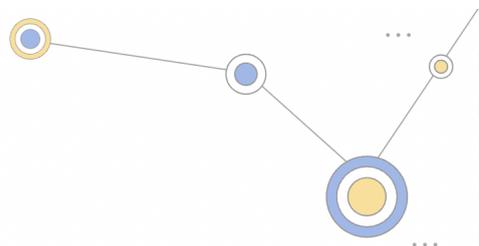
AMDAC	Administrateur ministériel des données, des algorithmes et des codes sources
ANCT	Agence nationale de la cohésion des territoires
API	<i>Application Programming Interface</i>
APIE	Agence du patrimoine immatériel de l'État
Arcep	Autorité de régulation des communications électroniques, des postes et de la distribution de la presse
Arcom	Autorité de régulation de la communication audiovisuelle et numérique
BAN / BAL	Base adresse nationale / Base adresse locale
Cada	Commission d'accès aux documents administratifs
CAPEX	<i>Capital expenditure</i>
CE	Commission européenne
Cerema	Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement
CKAN	<i>Comprehensive Knowledge Archive Network</i>
CNB	Conseil national des barreaux
CNCCFP	Commission nationale des comptes de campagne et des financements politiques
CNIL	Commission nationale de l'informatique et des libertés
CRPA	Code des relations entre le public et l'administration
CSV	<i>Comma separated values</i>
DGFIP	Direction générale des Finances publiques
DILA	Direction de l'information légale et administrative
DINUM	Direction interministérielle du numérique
ETP	Équivalent temps plein
EuroPAM	<i>European Public Accountability Mechanisms</i>





HATVP	Haute Autorité à la Transparence de la Vie Publique
IGN	Institut national de l'information géographique et forestière
IPS	Indices de position sociale
JORF	Journal officiel de la République française
LO 2.0	<i>Licence Ouverte / Open License</i>
Ma Dada	Ma Demande d'accès aux documents administratifs
MEPSIR	<i>Measuring European Public Sector Information</i>
OCDE	Organisation de coopération et de développement économiques
ODbL	<i>Open Database License</i>
OPEX	<i>Operating expenditure</i>
PAC	Politique agricole commune
PIB	Produit intérieur brut
PMSI	Programme de Médicalisation des Systèmes d'Information
PRADA	Personne responsable de l'accès aux documents administratifs
PSI	<i>Public Sector Information</i>
R.-U.	Royaume-Uni
RGPD	Règlement général sur la protection des données
RIP	Répertoire des informations publiques
RTI	<i>Right to Information</i>
SCSP	Subvention pour charge de service public
Shom	Service hydrographique et océanographique de la Marine
SQL	<i>Structured query language</i>
UE	Union européenne





Introduction

Du mythe de l'abondance à la réalité du labeur : une analyse de la politique française des données publiques

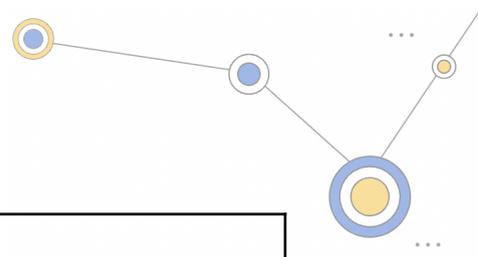
En 2016, la France franchit une étape majeure pour l'ouverture des données publiques avec l'adoption de la loi pour une République numérique, qui introduit l'obligation pour les administrations de publier leurs données « par défaut », dès lors qu'elles ne contiennent pas d'informations sensibles. Cette politique audacieuse a propulsé la France au sommet des classements internationaux en matière d'*open data*. Malgré cet élan, la réalité de l'ouverture des données publiques reste encore loin des ambitions affichées, avec de nombreux défis et résistances à surmonter.

Les ingénieurs-élèves des mines à la recherche de données pour nourrir leurs mémoires de fin d'études, comme tant d'autres étudiants, chercheurs, journalistes ou simples citoyens, sont les témoins malheureux d'un accès encore très imparfait aux données publiques. Prenons, à titre d'exemple, le cas de deux ingénieures-élèves de la promotion 2023 : travaillant sur le plastique à usage unique dans le domaine de la santé, elles ont cherché en vain des données sur les volumes de déchets produits par les hôpitaux. Ces données sont des informations environnementales qui, selon la loi, devraient être disponibles sur Internet. Forcées de constater que cette obligation n'était pas respectée, elles ont dû se résoudre à utiliser la seule donnée disponible, un résultat agrégé datant de 2010.

Constatant cet écart entre le cadre législatif très ambitieux et la persistance de difficultés d'accès, nous avons voulu interroger la pertinence et le caractère opérationnel d'une mise à disposition automatique des données publiques. ***Le principe du « par défaut » est-il possible et souhaitable ? N'existe-t-il pas une autre voie que le par défaut qui permette une ouverture plus efficace et plus complète des données publiques ?***

Nous avons décidé de sous-titrer ce travail : « du mythe de l'abondance à la réalité du labeur ». Dans les pages qui suivent, nous essaierons de vous convaincre que cette formule, aux accents de Rougon-Macquart, décrit au mieux la complexité de la politique française des données :

- (1) L'obligation d'ouverture des données par défaut, introduite par la loi pour une République numérique, est aujourd'hui très partiellement respectée.
- (2) Cette situation s'explique en partie par le fait que les difficultés associées n'ont pas été bien prises en compte dans l'élaboration de la politique française d'ouverture.
- (3) Pour surmonter ces obstacles, nous proposons de repenser les modalités concrètes de mise en œuvre de la politique d'ouverture, en gardant à l'esprit son objectif premier : la transparence.



Dans ce travail, nous entendons par :

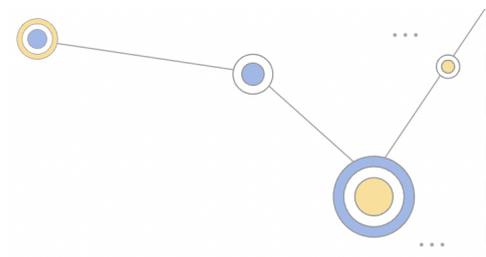
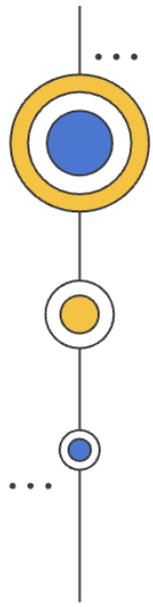
- **Administration :** L'État, les collectivités territoriales et toutes les autres personnes de droit public ou les personnes de droit privé chargées d'une mission de service public¹.
- **Information publique :** Toute information créée, collectée, traitée, conservée, entretenue, distribuée ou financée par ou pour un gouvernement ou une organisation publique².
- **Document administratif :** Tout document produit ou reçu par une administration dans le cadre de sa mission de service public³. Il peut s'agir de notes de service, de bases de données, de décisions, de codes sources de logiciel, de cartes, d'algorithmes, etc.
- **Donnée publique :** Représentation d'une information publique sous une forme conventionnelle destinée à faciliter son traitement.
- **Communication de document administratif :** Envoi par courrier (avec frais) ou courrier électronique (sans frais) du document demandé après une réponse favorable de l'administration concernée. Il s'agit de l'un des quatre modes d'accès aux documents administratifs prévu par l'article L. 311-9 du Code des relations entre le public et l'administration (CRPA).
- **Publication de document administratif :** Mise en ligne du document pour un accès généralisé et gratuit, bénéficiant à tous les citoyens. Cette mise en ligne peut être à l'initiative de l'administration ou à la demande d'un citoyen. Il s'agit de l'un des quatre modes d'accès aux documents administratifs prévu par l'article L. 311-9 du CRPA.
- **Ouverture des données publiques :** Mise à disposition généralisée des données publiques, en accès libre et gratuit, sous un format numérique facilement réutilisable. On parle alors de « **données ouvertes** » (ou **open data** en anglais).

¹ Article L300-2 du CRPA.

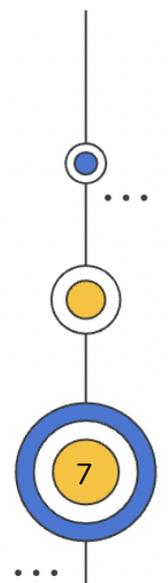
² OCDE, « Digital Broadband Content : Public Sector Information », Documents de travail de l'OCDE sur l'économie numérique, n° 112, 2006.

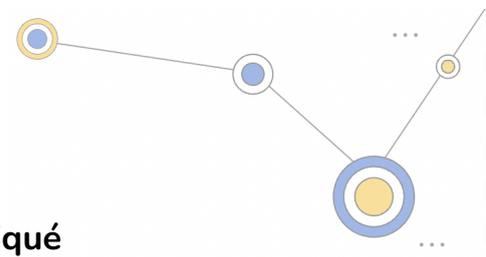
³ Article L300-2 du CRPA.





1. Une ouverture incomplète et des opportunités manquées





1.1. Un cadre législatif ambitieux mais inégalement appliqué

1.1.1. Le cadre législatif de l'ouverture des données publiques

Ce que dit la loi pour une République numérique

Le 10 avril 2012, alors candidat à l'élection présidentielle, François Hollande promet une diffusion libre et efficace des données produites par les établissements publics⁴. En 2014, il confie à la secrétaire d'État chargée du numérique, Axelle Lemaire, la mission de tenir cette promesse.

Pour la première fois en France, une consultation citoyenne en ligne sur un avant-projet de loi est organisée⁵. Elle recueille plus de 21 000 contributions, 147 000 votes, et aboutit à l'adoption de la **loi pour une République numérique, dite « loi Lemaire »**⁶, promulguée le 7 octobre 2016.

Cette loi présente plusieurs volets :

- **La circulation des données et du savoir**, en favorisant l'accès aux données, notamment publiques, et leur diffusion ;
- **La protection des droits dans la société numérique**, en affirmant le principe de neutralité du net et en renforçant la maîtrise des internautes sur leurs données personnelles ;
- **L'accès au numérique**, en favorisant l'accès des publics les plus vulnérables à internet et aux services en ligne.

Cette loi marque un tournant majeur dans la politique d'ouverture des données publiques française en introduisant le **principe d'ouverture par défaut des données publiques**⁷ : les documents administratifs présentant des données publiques doivent désormais être systématiquement mis en ligne par les administrations⁸. Il s'agit d'un changement de paradigme et de culture important pour les administrations puisqu'auparavant, la publication en ligne d'un document relevait de l'exception et non de la règle. Cette loi implique que les administrations rompent avec leurs habitudes et publient spontanément leurs données.

⁴ www.luiipresident.fr/francois-hollande/engagement/open-data-33934/

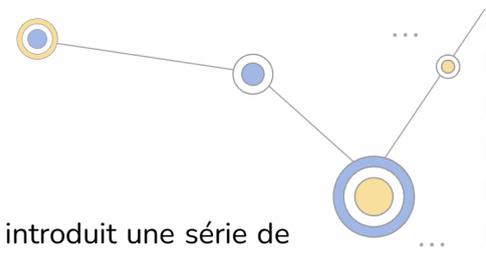
⁵ Dossier de presse « Loi pour une République numérique », octobre 2016. URL : www.economie.gouv.fr/files/files/PDF/DP_LoiNumerique.pdf

⁶ Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique.

⁷ Article L312-1-1 du CRPA.

⁸ Ne sont concernées que les administrations de plus de 50 agents ou salariés et de plus de 3500 administrés pour les collectivités territoriales.





Au-delà de ce nouveau principe de publication « par défaut », la loi introduit une série de mesures favorisant l'accès aux données publiques :

- Les administrations sont obligées de publier un **répertoire des informations publiques** (RIP) listant les principaux documents contenant des informations publiques⁹;
- Est créé un **service public de la donnée**, c'est-à-dire l'obligation de production et de diffusion de neuf jeux de données de référence¹⁰, qui permettent notamment d'identifier des produits, des services, des territoires ou des personnes ;
- Les documents doivent être communiqués dans un **standard ouvert**, aisément réutilisable et exploitable par un système de traitement automatisé¹¹;
- Les données doivent être publiées sous des **licences ouvertes**¹²;
- Les **données d'intérêt général** produites par des concessionnaires de réseaux publics d'électricité et de gaz naturel dans le cadre d'une délégation de service public, doivent être transmises aux administrations concédantes¹³.

En outre, la loi Lemaire **interdit les redevances** issues de la vente de données entre administrations, complétant ainsi la loi du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public, dite « **loi Valter** ». Celle-ci imposait déjà un principe de gratuité des données publiques en supprimant la plupart des redevances perçues par les administrations et en limitant les exceptions. La liste des informations publiques pouvant encore faire l'objet de redevances est très restreinte et inscrite dans le CRPA¹⁴.

Si la loi a permis des avancées considérables, rapidement des voix s'élèvent pour montrer les limites de son application. En décembre 2020, le rapport du député Eric Bothorel dresse un état des lieux de la politique publique de la donnée et rappelle la nécessité d'une plus grande ouverture des données publiques¹⁵. L'année suivante, dans une circulaire¹⁶, le Premier ministre Jean Castex prend acte des conclusions du rapport et rappelle aux administrations leur **obligation de supprimer les redevances** pour favoriser la circulation

⁹ Article L322-6 du CRPA.

¹⁰ Jeux de données listés à l'article [L321-5 du CRPA](#) : [BAN](#), [SIREN - SIRET](#), [COG](#), [PCI](#), [RPG](#), [ROAE](#), [RGE](#), [RNA](#), [ROME](#).

¹¹ Article L300-4 du CRPA.

¹² Par défaut, il s'agit des licences ODbL ou Licence Ouverte Etalab tel que le prévoit l'article D323-2-1 CRPA.

¹³ Valeurs foncières, décisions de justice, vitesses maximales autorisées sur le domaine public routier.

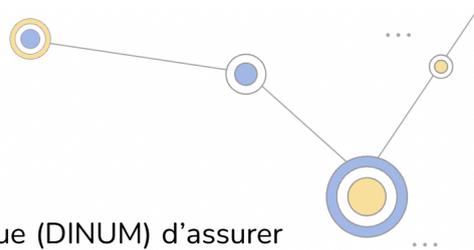
¹⁴ Article D324-5-1 du CRPA.

¹⁵ Eric Bothorel, Rapport de mission « Pour une politique publique de la donnée », 2020. URL :

www.ouvrirelascience.fr/wp-content/uploads/2021/02/Mission-Bothorel_rapport-pour-une-politique-publique-de-la-donnee_23.12.2020.pdf

¹⁶ Circulaire n°6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources.





des données. Il demande à la Direction interministérielle du numérique (DINUM) d'assurer la **coordination de la politique de la donnée à l'échelle interministérielle**.

La DINUM publie alors un **Cadre interministériel d'administration de la donnée**¹⁷, document dans lequel elle décrit les bonnes pratiques à adopter pour une ouverture efficace et utile des données : relations avec l'écosystème de réutilisateurs, réseau de correspondants données au sein des ministères, maintien au sein des administrations de profils experts, etc. Sont également présentés les outils proposés par la DINUM pour favoriser la mise à disposition et la réutilisation des données publiques : la plateforme *data.gouv.fr* pour le référencement des données, la plateforme *api.gouv.fr* pour l'accès aux données via des API¹⁸ ou encore la plateforme *schema.data.gouv.fr* pour le référencement des schémas de données¹⁹.

Le fer de lance de l'ouverture des données publiques : data.gouv.fr

De nombreuses administrations ont développé leurs propres portails *open data* pour rassembler et mettre en valeur leurs données. Certaines collectivités territoriales se sont ainsi dotées de **portails territoriaux** : *opendata.agglo-larochelle.fr* pour l'agglomération de La Rochelle ou *data.grandlyon.com* pour la métropole de Lyon. C'est aussi le cas d'administrations d'État, avec par exemple le portail *data.economie.gouv.fr* du Ministère de l'Économie.

Face à la multitude de portails et pour accéder plus facilement aux données publiques, la DINUM a reçu pour mission de répertorier ces données sur un portail unique, *data.gouv.fr*, et d'y intégrer les données des administrations sans portail propre. Le site *data.gouv.fr* est un **catalogue de données** en ligne qui rassemble **38 000 jeux de données**²⁰ publiés par diverses administrations : ministères, collectivités territoriales, opérateurs de l'État, etc. Ces données sont publiées directement sur *data.gouv.fr* ou moissonnées²¹ depuis les portails des administrations.

¹⁷ Cadre interministériel d'administration de la donnée, DINUM, 2021. URL :

www.numerique.gouv.fr/uploads/Cadre-interministeriel-donnee.pdf

¹⁸ Une API (Application Programming Interface), ou interface de programmation applicative, est une interface logicielle qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service afin d'échanger des données et des fonctionnalités.

¹⁹ Les schémas de données permettent de décrire des modèles de données : quels sont les différents champs, comment sont représentées les données, quelles sont les valeurs possibles, etc.

²⁰ Toutes les données et statistiques issues de *data.gouv.fr* ont été recueillies le 14 juin 2024.

²¹ Le moissonnage est l'extraction automatisée des données contenues dans un site internet.



Si le volume de données ouvertes est important, les **métadonnées**²² sont essentielles pour comprendre et interpréter les données, favorisant ainsi les réutilisations. Pour juger de leur qualité, l'équipe de *data.gouv.fr* a développé un score²³ basé sur une dizaine de critères : description des données, indication de la licence utilisée, formats de fichiers standards, etc. Cet indicateur révèle d'importantes différences entre les administrations et permet d'identifier les « bons élèves » de l'ouverture des données : en tête, on trouve ainsi la métropole de Lyon qui obtient un score moyen de 85% pour ses 596 jeux de données, ou encore le ministère de l'Intérieur avec 683 jeux de données et un score qualité de 68% ou le ministère de l'Économie avec 520 jeux de données et un score de 65%. Ces administrations se situent en effet bien au-delà de la moyenne : 7 jeux de données et un score qualité de 53%.

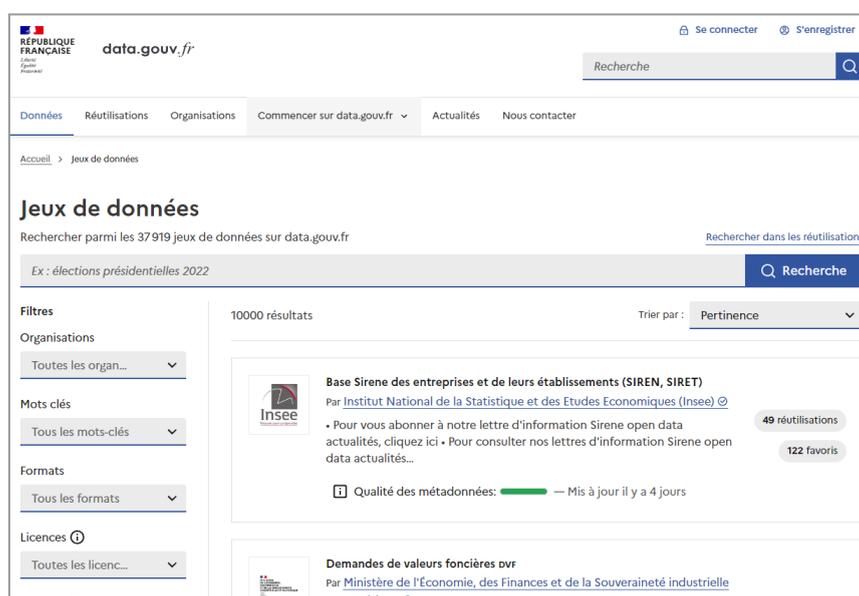


Figure 1 : Le catalogue des jeux de données sur *data.gouv.fr*, 2024.

Les administrations ne sont pas les seules utilisatrices de *data.gouv.fr* : tout individu peut créer un compte et **publier ses réutilisations**, comme des outils de visualisation des données ou des analyses issues de croisement de bases de données. Par exemple :

- La plateforme *Aux Alentours*²⁴ créée par la MAIF croise de nombreuses données pour identifier les risques auxquels est exposé un logement (inondation, recul du trait de côte, retrait gonflement des argiles, etc.).

²² Des métadonnées sont des données qui décrivent certaines autres données afin d'en faciliter la compréhension.

²³

guides.data.gouv.fr/guides-open-data/guide-qualite/ameliorer-la-qualite-dun-jeu-de-donnees-en-continu/ameliorer-le-score-de-qualite-des-metadonnees

²⁴ www.data.gouv.fr/fr/reuses/aux-alentours-par-maif/

- L'outil *CovidTracker*²⁵ créé par Guillaume Rozier a permis à des millions de Français de mieux visualiser les données du Covid publiées par l'Agence nationale de santé publique.
- La carte de *Visualisation des données sur les prix des carburants*²⁶ publiées par la DINUM a permis à des millions d'automobilistes d'identifier les stations d'essence les moins chères lors de la crise énergétique de 2021.

L'interaction entre producteurs et réutilisateurs de données est assurée sur *data.gouv.fr* par des **forums de discussion** associés à chaque jeu de données. Cette fonctionnalité simple est essentielle pour permettre des réutilisations de qualité.

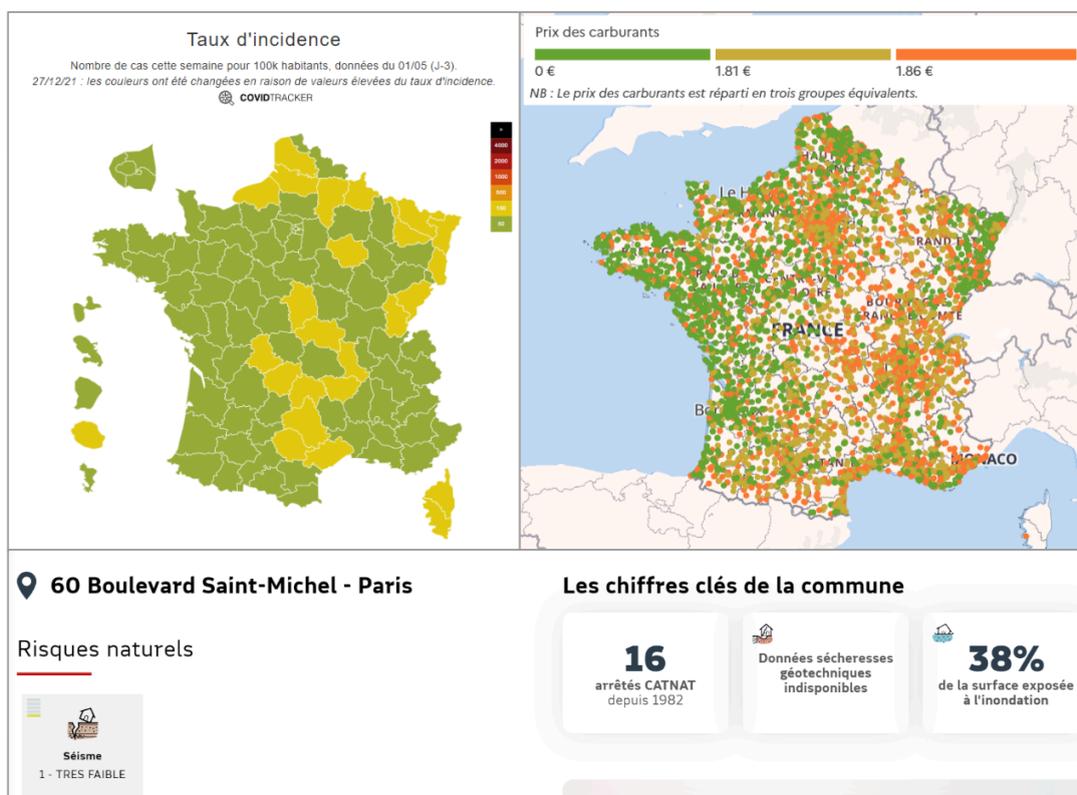


Figure 2 : Exemples de réutilisations de données publiques sur *data.gouv.fr* : CovidTracker en haut à gauche, la carte des prix des carburants en haut à droite, la plateforme Aux Alentours en bas. Source : *data.gouv.fr*, 2024.

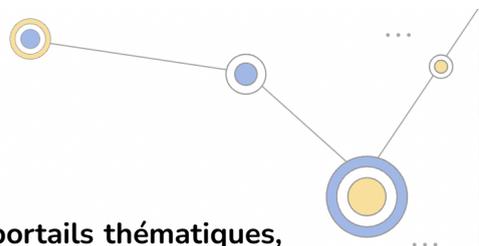
Avec plus de 2 millions de téléchargements uniques en moyenne chaque année²⁷, la plateforme *data.gouv.fr* semble être devenue la principale référence des réutilisateurs. Mais pour mobiliser davantage les producteurs de données et faciliter la découvrabilité des jeux

²⁵ www.data.gouv.fr/fr/reuses/covid-tracker-suivez-lepidemie-du-covid19-en-france-et-dans-le-monde/

²⁶ www.data.gouv.fr/fr/reuses/visualisation-des-donnees-sur-les-prix-des-carburants/

²⁷

stats.data.gouv.fr/index.php?module=CoreHome&action=index&idSite=109&period=range&date=previous30#?idSite=109&period=year&date=2024-06-15&category=General_Visitors&subcategory=General_Overview



de données, la DINUM s'attaque aujourd'hui au développement de **portails thématiques**, favorisant le développement d'un écosystème de producteurs et de réutilisateurs de données motivés par des intérêts communs : *meteo.data.gouv.fr* de Météo-France pour les données météorologiques, le projet Écosphères du Commissariat général au développement durable pour les données écologiques (en cours de développement), *geoportail.gouv.fr* de l'Institut national de l'information géographique et forestière (IGN) pour les données géographiques.

1.1.2. Les difficultés de la mise en oeuvre

La France championne de l'open data et cancre de l'accès aux documents administratifs ?

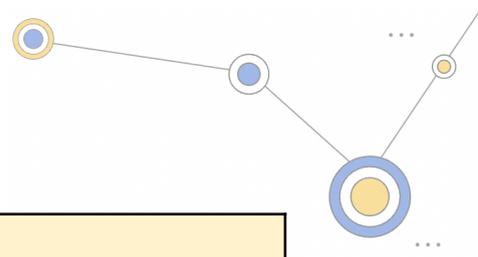
En janvier 2024, la DINUM annonce que la France occupe la **première place à l'échelle européenne et la deuxième place au niveau mondiale en matière d'open data**. Elle se réfère à deux classements, tous deux publiés en 2023 : l'*Open Data Maturity Report*²⁸ de la Commission européenne qui considère que la France est un meneur (« *trend settler* ») et le pays européen le plus « mature » pour la troisième année consécutive, et le *OURData Index*²⁹ de l'OCDE qui classe la France à la deuxième place mondiale, derrière la Corée du Sud.

Pourtant, un autre classement a attiré notre attention : le classement *Right to Information Ratings*³⁰ (RTI), établi chaque année par l'organisation de défense de droits de l'homme *Access Info Europe* et l'organisme à but non lucratif *Centre for Law and Democracy*. Il évalue la qualité du droit d'accès aux documents administratifs à travers le monde. En 2022, la France y occupe la 108^e position à l'échelle mondiale. Cela paraît très paradoxal : **comment peut-on être à la fois excellent pour publier spontanément ses données et très mauvais pour permettre aux gens d'accéder à ces mêmes données ?**

²⁸ Commission européenne, « Open Data Maturity Report », 2023. URL : data.europa.eu/sites/default/files/odm2023_report.pdf

²⁹ OCDE, « Open, Useful and Re-usable data (OURdata) Index », 2023. URL : www.oecd-ilibrary.org/docserver/a37f51c3-en.pdf?expires=1716483242&id=id&acname=guest&checksum=4F61536C5E771A4EED2BEA9F0AD85653

³⁰ www.rti-rating.org/



Le droit d'accès aux documents administratifs

En 1974, le journal Le Monde dénonce un projet du Ministère de l'Intérieur qui vise à croiser les fichiers administratifs des Français dans une seule grande base de données³¹. Cette nouvelle fait scandale et les Français réclament une plus grande transparence et un plus grand contrôle de l'administration.

À la suite de cette affaire, deux lois sont adoptées : la loi du 6 janvier 1978 dite « loi informatique et libertés »³² crée la Commission nationale de l'informatique et des libertés (CNIL), et la loi du 17 juillet 1978, dite « **loi Cada** »³³ crée la Cada, la **Commission d'accès aux documents administratifs**, et fonde le droit d'accès aux documents administratifs en France, c'est-à-dire le « *droit de toute personne à l'information* ». Il permet à tout citoyen de demander à une administration un document administratif, sauf s'il contient certains secrets. En 2020, le Conseil constitutionnel consacre ce droit comme un droit constitutionnel³⁴.

Désormais chaque administration nomme une **personne responsable de l'accès aux documents administratifs (PRADA)** chargée de répondre aux demandes. La Cada a quant-à-elle pour rôle de rendre des avis sur le caractère communicable ou non d'un document. **On ne peut demander que des « documents »**, l'administration n'a donc pas à fournir des informations ou tout autre forme de renseignements³⁵.

Le **site Ma Dada**³⁶ a été créé par des citoyens pour faciliter les demandes et publier en *open data* les documents transmis par les administrations. Il est une source importante de statistiques sur le droit d'accès aux documents administratifs.

Le principe d'**open data par défaut** introduit par la loi pour une République numérique renforce le droit d'accès aux documents administratifs dans la mesure où la publication doit désormais se faire en amont des demandes, par anticipation.

³¹ Philippe Boucher, « Une division de l'informatique est créée à la chancellerie " Safari " ou la chasse aux Français », Le Monde, 1974. URL :

www.lemonde.fr/archives/article/1974/03/21/une-division-de-l-informatique-est-creee-a-la-chancellerie-safari-ou-la-chasse-aux-francais_3086610_1819218.html

³² Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

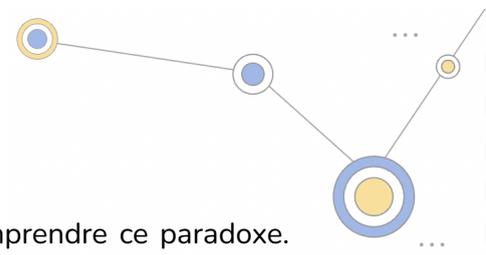
³³ Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal.

³⁴ Conseil d'État, Décision n° 2020-834, question prioritaire de constitutionnalité du 3 avril 2020.

³⁵ Une exception existe pour les données présentant un intérêt économique, social, sanitaire ou environnemental, et dont la publication peut être demandée.

³⁶ Pour « Ma Demande d'Accès aux Documents Administratifs ». URL : madada.fr/





L'analyse de la **méthodologie des classements** permettrait de comprendre ce paradoxe. Mais seul le classement RTI fournit la liste exhaustive des indicateurs mesurés pour évaluer les différents pays, ainsi que le détail des points accordés par indicateur et les raisons pour lesquelles les points sont accordés ou non. Pour les classements *OURData Index* et *Open Data Maturity Report*, il faut se contenter d'indicateurs agrégés qui ne permettent pas d'expliquer précisément pourquoi un pays est plus performant qu'un autre. Nous pouvons tout de même en tirer quelques enseignements.

En parcourant les indicateurs mesurés par le classement *OURData Index*, il apparaît que beaucoup de questions évaluent le contenu de la loi mais pas son application. Autrement dit, **il suffit que la théorie soit parfaite pour avoir une très bonne note, qu'importe la pratique !** Concernant le classement *Open Data Maturity Report*, impossible de juger si la méthodologie est la même puisque les critères de notation ne sont pas accessibles en *open data* (un comble !). Mais le fait que la France obtienne un score global de 98,3% pose tout de même question. Elle serait donc proche de la perfection en matière d'*open data*, pourtant nous verrons que cette affirmation peut largement être remise en question. Si le classement RTI peut lui-aussi être largement critiqué pour sa méthodologie, il a cependant le mérite de souligner des difficultés du droit d'accès aux documents administratifs qui sont bien réelles et que nous présenterons par la suite.

Loi pour une République numérique : les écarts entre théorie et pratique

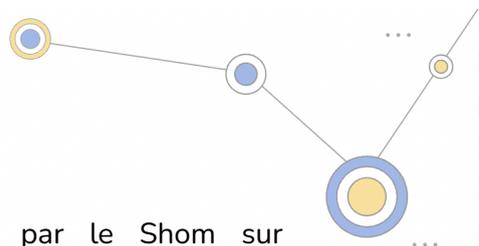
Il est difficile de mesurer de manière objective le niveau de progression de la mise en œuvre de l'*open data* en France. Nous nous sommes efforcés de rassembler quelques indicateurs et quelques éléments factuels pour appuyer ce constat. Parfois, nous avons dû constituer nos propres indicateurs à partir de diverses sources, comme *data.gouv.fr*, *Ma Dada* et *choisirleservicepublic.fr*. Ils montrent que la loi pour une République numérique est appliquée de façon très inégale.

Certaines mesures ont été très bien appliquées. Le **service public de la donnée** a bien été mis en place et les neuf jeux de données listés par la loi sont bien accessibles sur *data.gouv.fr*. La **suppression des redevances** a également été bien mise en œuvre, même si l'ouverture se fait de façon tardive pour certaines administrations³⁷ et aura nécessité un rappel à l'ordre du Premier ministre Jean Castex. Et même si la loi prévoit des exceptions pour permettre à quelques administrations³⁸ de percevoir des redevances sur une liste de données clairement identifiées, certaines de ces administrations ont choisi de ne pas percevoir ces redevances. On peut citer par exemple les cartes de l'IGN disponibles sur

³⁷ Météo-France n'a ouvert ses données qu'au début de l'année 2024.

³⁸ IGN, Météo-France, Service hydrographique et océanographique de la Marine (Shom) et bibliothèques.





geoportail.gouv.fr ou encore les horaires des marées publiés par le Shom sur *maree.shom.fr*. Cependant, certaines données comme les données de prévisions expertisées de Météo-France ou les cartes marines vectorielles du Shom n'ont pas été ouvertes comme le prévoit la loi.

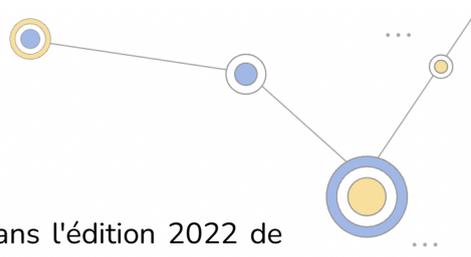
D'autres mesures sont appliquées de manière beaucoup plus imparfaite. Certaines **données d'intérêt général** identifiées par la loi ont bien été ouvertes et rencontrent un certain succès, comme la base des demandes de valeur foncière publiée par la Direction générale des Finances publiques (DGFIP)³⁹ ou la base des décisions judiciaires publiée par la Cour de Cassation⁴⁰. D'autres demeurent inaccessibles. Par exemple, les **données des concessionnaires d'eau ou d'électricité** tardent à être partagées aux collectivités concédantes par des délégataires qui craignent qu'elles ne leur servent de levier pour négocier la baisse des prix des redevances. La Fédération nationale des collectivités concédantes et régies estime que les collectivités ne sont pas suffisamment conscientes de l'importance de récupérer leurs données et ne sont donc pas prêtes à se battre pour les récupérer. Autre donnée d'intérêt général, la **base des vitesses maximales autorisées sur le domaine public routier** n'a pas été publiée faute de décret d'application. Une demande d'accès à ce document a été formulée auprès de Ministère de l'Intérieur, qui a répondu qu'un tel fichier n'existait pas. La Cada a rendu en 2021 un avis⁴¹ donnant raison au ministère : il est vrai que l'on ne peut pas communiquer un document qui n'existe pas ! Pourtant la loi demande explicitement que cette base de données soit constituée et publiée. D'après le ministère de l'Intérieur, certains éléments de cette base existent mais ne sont pas suffisamment fiables pour pouvoir être publiés. Pourtant, sans ces données, il est difficile d'évaluer l'impact de certaines politiques publiques comme le passage à 80 km/h voulu par le gouvernement d'Édouard Philippe.

Concernant les données publiques diffusées dans le cadre de l'*open data* par défaut, force est de constater que **leur qualité et leur format sont souvent loin de répondre aux standards prévus par la loi**. Les jeux de données publiés sur *data.gouv.fr* obtiennent un score de **qualité de leurs métadonnées de seulement 53% en moyenne**. Sur Ma Dada, une analyse des formats des documents transmis par les administrations montre que **70% sont au format PDF, contre seulement 10% au format Excel** pour permettre un traitement automatisé comme le prévoit la loi. Ainsi, pour les demandes de notes de frais des élus, très fréquentes sur Ma Dada, les administrations transmettent des fichiers PDF dans 60% des cas, et dans seulement 16% des cas, des fichiers de type Excel.

³⁹ www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/

⁴⁰ www.courdecassation.fr/acces-rapide-judilibre

⁴¹ cada.data.gouv.fr/20205263/



Bien souvent, les données ne sont même pas publiées en ligne. Dans l'édition 2022 de l'Observatoire open data des Territoires⁴², l'association OpenDataFrance estime que **seules 16% des collectivités territoriales** concernées par l'obligation de l'ouverture par défaut respectent cette obligation⁴³ dans la mesure où elles disposent d'un portail propre ou ont un compte à leur nom sur *data.gouv.fr*. En outre, ce résultat, déjà très faible, cache une grande diversité de situations avec des collectivités qui pour beaucoup ne font que « cocher la case » puisque **44% d'entre elles n'ont ouvert qu'un ou deux jeux de données**. Il faut également noter que 55% n'ont pas de portail dédié et publient uniquement sur *data.gouv.fr* 8 jeux de données en moyenne contre 71 pour celles qui ont à la fois également un portail territorial. Il y a donc bien un **open data à deux vitesses**, entre les collectivités qui se contentent du strict minimum et celles qui vont jusqu'au bout de la démarche.

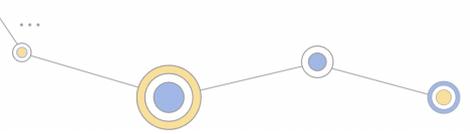
De façon générale, **la disponibilité des données concernant les marchés publics est bien en-dessous des standards européens**, comme le montre le classement *European Public Accountability Mechanisms (EuroPAM)*⁴⁴ financé par l'Union européenne. Il attribue à la France la note de 32/100, contre 52/100 pour la moyenne européenne. Cela est essentiellement dû à deux éléments. D'une part, les administrations françaises ont la possibilité de s'engager à passer des marchés auprès d'un prestataire sur une période donnée et pour des prestations déterminées⁴⁵. On appelle cela un accord-cadre, et chaque marché passé dans ce cadre s'appelle un marché subséquent. L'existence d'un accord-cadre doit être rendue publique, mais contrairement à 54% des pays européens, ce n'est pas le cas pour les marchés subséquents qui pourtant correspondent véritablement aux prestations délivrées. D'autre part, il n'est pas obligatoire en France de publier des informations sur les sous-traitants. Autrement dit, si une entreprise A remporte un marché public et sous-traite son exécution à une entreprise B, le fait que l'entreprise A ait le contrat est public, mais pas le fait que ce soit en réalité l'entreprise B qui l'exécute – cette information est collectée mais n'est pas nécessairement publiée.

⁴² OpenDataFrance, « Observatoire open data des Territoires », 2022. URL : www.observatoire-opendata.fr/wp-content/uploads/2023/03/Copie-de-Observatoire-Open-Data-Territoires-Edition2022.pptx-1.pdf

⁴³ Ce résultat donne une idée du faible avancement de l'ouverture des données dans les territoires mais il faut le nuancer. En effet, OpenDataFrance essaie d'identifier les portails couvrant plusieurs communes, mais certains lui échappent. Par exemple, les communes de l'agglomération de La Rochelle qui publient leurs données sur le portail de l'agglomération n'apparaissent pas dans le décompte d'OpenDataFrance, qui considère que le portail est celui de La Rochelle uniquement. Pour cette raison, on peut estimer que la part des collectivités territoriales respectant leur obligation d'ouverture serait en réalité un petit peu plus élevée.

⁴⁴ www.europam.eu/index.php?module=country-profile&country=France

⁴⁵ Article R2183-3 du Code de la commande publique.



Enfin, l'obligation pour un grand nombre d'administrations⁴⁶ de publier un répertoire des informations publiques (RIP) n'est pas du tout respectée. En témoigne cette statistique : **sur près de 40 000 demandes de RIP faites sur Ma Dada, seules 33 administrations ont transmis le document**⁴⁷. Pourtant le RIP est un élément essentiel de l'accès des citoyens à l'information publique pour deux raisons. D'une part, l'efficacité d'une politique d'*open data* repose notamment sur la découvrabilité des données, et avoir un annuaire des données est bien sûr très utile pour savoir si une donnée existe. D'autre part, pour les données qui ne sont pas publiées mais qui pourraient l'être, il est essentiel que les citoyens sachent qu'elles existent pour pouvoir les demander en vertu du droit d'accès aux documents administratifs.

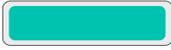
Classements mondiaux de l' <i>open data</i>	1 ^{er} en Europe (Open Data Maturity Report). 2 ^e dans le monde (OURData Index).	
Service public de la donnée	38 000 jeux de données, dont 9 jeux de données de référence, publiés sur <i>data.gouv.fr</i> .	
Fin des redevances	Appliquée, avec quelques exceptions à la marge.	
Format des données	Les documents transmis sur la plateforme Ma Dada sont à 70% des PDF et seulement 10% des tableurs.	
Open data par défaut	16% des collectivités concernées sont en open data. Score qualité sur <i>data.gouv.fr</i> : 53% en moyenne.	
Répertoire d'informations publiques	33 RIP transmis sur 40 000 requêtes via Ma Dada.	
Classement mondial de l'accès aux documents administratifs	108 ^e dans le monde (RTI)	

Tableau 1 : Quelques indicateurs de la mise en œuvre de l'ouverture des données publiques en France.

1.1.3. Un droit d'accès aux documents administratifs défaillant

La difficile application de la loi pour une République numérique s'expliquerait notamment par les faiblesses du droit d'accès aux documents administratifs en France.

⁴⁶ Ne concerne que les administrations de plus de 50 agents ou salariés et de plus de 3500 administrés pour les collectivités territoriales.

⁴⁷ Mesure faite le 22 mai 2024.

Demande d'accès aux documents administratif : une démarche qui peut devenir un véritable parcours du combattant



Figure 3 : Schéma des délais pour une demande d'accès à un document administratif.

Le droit d'accès aux documents administratifs permet au citoyen de demander des documents détenus par une administration dans le cadre de sa mission de service public et de solliciter, s'il en était besoin, l'avis de la Cada pour appuyer sa demande. Prenons l'exemple de Pierre, qui désire obtenir les notes de frais de son maire et envoie un mail à la mairie de sa commune pour demander les documents⁴⁸.

Au bout d'un délai légal d'un mois, la mairie n'a toujours pas transmis le document, cette **absence de réponse vaut refus implicite**. Pierre **saisit alors la Commission d'accès aux documents administratifs** qui rend un avis dans un délai d'une cinquantaine de jours⁴⁹. Cet avis est favorable : la Cada estime que Pierre a le droit de demander les notes de frais de son maire. Mais cet avis n'a qu'un caractère consultatif et la commune de Pierre n'est pas obligée de le suivre, elle persiste donc à refuser de communiquer le document.

Pierre **saisit donc le tribunal administratif** et écrit un mémoire pour expliquer pourquoi la décision de la commune est entachée d'illégalité. Le tribunal rend son jugement au bout d'un an et demi⁵⁰ : la commune doit communiquer les documents demandés. Mais cette dernière fait un **recours devant le Conseil d'État** qui va mettre deux années supplémentaires pour prendre une décision⁵¹ en faveur de Pierre : la commune finit par fournir à Pierre les notes de frais de l'élu pour ne pas payer l'astreinte imposée par le Conseil d'État. **Il aura donc fallu à Pierre près de quatre ans pour obtenir des**

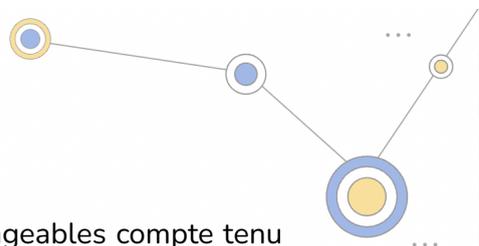
⁴⁸ Il peut demander la communication des documents, le document sera transmis à lui et lui-seul, mais aussi la publication (à moins que les documents ne soient communicables qu'à l'intéressé), et donc la mise en ligne des documents communiqués. À noter que du point de vue du demandeur, il est préférable de demander la publication, car elle implique aussi la mise à jour des documents.

⁴⁹ Cada, « Rapport d'activité 2022 - 2023 ». URL :

www.cada.fr/sites/default/files/CADA_RAPPORT-2022-2023-UK-vf.pdf

⁵⁰ D'après le Tribunal administratif de Paris, il faut entre sept mois et deux ans et demi selon la nature et la difficulté des dossiers pour rendre un jugement.

⁵¹ Sur un échantillon de décisions issu du [Conseil d'État](#) contenant le mot-clé « Cada », les délais pour rendre une décision sont de l'ordre de deux ans.



informations censées être publiques⁵². Ces délais sont très dommageables compte tenu de la vitesse à laquelle une information perd de la valeur : plus on attend et plus elle risque de devenir obsolète.

Notes de frais d'un maire : un cas loin d'être théorique

Le Conseil d'État a jugé le 8 février 2023 que **les notes de frais d'un maire devaient être communiquées à tout citoyen qui en ferait la demande**⁵³.

Cette décision est le fruit du combat d'un journaliste pour obtenir les notes de frais de la maire de Paris. Il a fait une première demande en janvier 2018 et la CADA a rendu un avis favorable en juillet 2018. En mars 2021, le Tribunal administratif de Paris a enjoint la Ville de Paris de communiquer les documents et l'a condamnée à verser 1000 € au journaliste. La Ville de Paris a fait un recours en Conseil d'État qu'elle a finalement perdu, ce qui a permis au Conseil d'État de faire jurisprudence sur le cas des notes de frais des maires.

Mais la défaite n'est pas totale pour la Ville de Paris qui a réussi à gagner **cinq ans avant de transmettre les documents**, permettant notamment de passer les élections municipales de 2020.

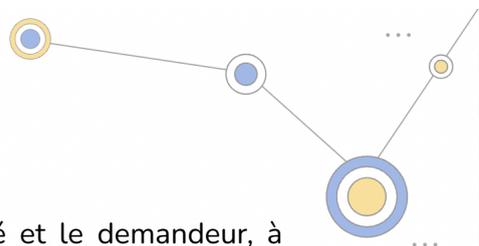
Si le cas de Pierre n'est pas une exception, tous les demandeurs ne rencontrent pas une telle résistance de la part des administrations. Certaines communiquent les documents dès la demande, d'autres à la suite de l'avis favorable de la CADA. Parfois les administrations attendent que le demandeur saisisse le tribunal administratif pour transmettre les documents et mettre fin aux poursuites. Ainsi, d'après l'association Ouvre-boîte, qui lutte pour obtenir la publication effective des documents administratifs, sur 18 demandes ayant abouti à la communication de documents, 3 ont abouti juste après la demande, 8 après l'avis de la CADA, 6 après le recours devant le tribunal administratif (mais avant le jugement) et une après la décision du tribunal administratif⁵⁴.

Par ailleurs, certaines demandes échouent non pas en raison d'une mauvaise volonté de l'administration mais parce qu'elles concernent **des documents couverts par des secrets qui n'ont donc pas vocation à être rendus publics**. La CADA rend alors un avis défavorable

⁵² Pierre a tout de même de la chance que le Conseil d'État ait jugé son affaire sur le fond et ne l'ait pas renvoyé devant le tribunal administratif pour un second jugement

⁵³ Conseil d'État, 10ème - 9ème chambres réunies, 08/02/2023, 452521.

⁵⁴ ouvre-boite.org/ils-ont-ouvert.html



en expliquant pourquoi le document ne peut pas être communiqué et le demandeur, à moins qu'il ne soit persuadé de son bon droit, ne va pas plus loin.

Une complexité qui ne permet pas de garantir ce droit fondamental

Les exemples de défaillance du droit d'accès aux documents administratifs sont nombreux et justifient que l'on s'inquiète sur la garantie de ce droit fondamental. Le sujet n'est pas anecdotique pour les administrations qui reçoivent environ **40 000 demandes de documents**⁵⁵ chaque année, qui aboutissent à près de **10 000 saisines de la Cada**⁵⁶.

Mais cette dernière ne rend qu'un avis consultatif et n'a pas de pouvoir de sanction. De nombreuses administrations choisissent donc d'ignorer les demandes puis les avis Cada associés. Ainsi, sur plus de 40 000 demandes de documents effectuées en cinq ans via Ma Dada, près de **86% n'ont pas même un accusé de réception** de la part de l'administration⁵⁷. Sur tous les avis rendus par la Cada, 57% d'avis sont favorables aux demandeurs. Pire, seules **63% des administrations tiennent au courant la Cada** des suites qu'elles donnent à un avis favorable. Parmi elles, **28% ne tiennent pas compte de cet avis**, elles persistent donc à refuser l'accès à un document que la Cada juge communicable⁵⁸.

Ainsi, non seulement la Cada est dans bien des cas aveugle quant aux suites données à ses avis, mais ses avis sont aussi trop rarement pris en compte par les administrations.

1.2. Les enjeux et les opportunités de l'ouverture des données publiques

1.2.1. Un enjeu démocratique

L'exemple américain : après l'Eldorado Obama, une vague de panique démocratique à l'arrivée de l'administration Trump.

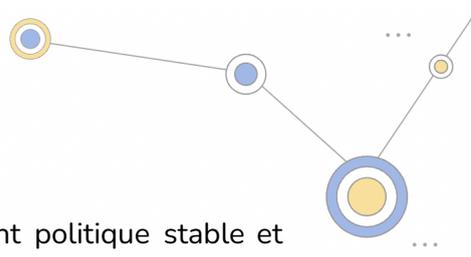
En 2008, la crise financière met en évidence le caractère très opaque des opérations réalisées par les grandes institutions financières. Les associations qui militent en faveur de l'ouverture des données se saisissent de l'occasion pour montrer l'importance de la

⁵⁵ Donnée difficile à estimer car personne ne l'agrège pour toutes les administrations, mais un PRADA interviewé pour le rapport d'activité 2022 - 2023 de la Cada estime à environ 25% le nombre de demandes reçues par les PRADA qui se sont ensuite poursuivies devant la Cada.

⁵⁶ Cada, « Rapport d'activité 2022 - 2023 ». URL : www.cada.fr/sites/default/files/CADA_RAPPORT-2022-2023-UK-vf.pdf

⁵⁷ Mesure à la date du 22 mai 2024

⁵⁸ Cada, « Rapport d'activité 2022 - 2023 ». URL : www.cada.fr/sites/default/files/CADA_RAPPORT-2022-2023-UK-vf.pdf



transparence pour prévenir les crises et préserver un environnement politique stable et soutenable⁵⁹. Barack Obama, alors candidat aux élections présidentielles américaines, est approché par les grandes figures du mouvement pour l'ouverture des données qui le convainquent de faire de la transparence une de ses principales promesses de campagne. Barack Obama sera le président de la transparence. Chose promise, chose due : **dès le lendemain de son investiture, le président Obama signe un mémorandum sur la transparence et le gouvernement ouvert**, dans lequel il s'engage à créer un niveau d'ouverture sans précédent au sein de son gouvernement. Il souhaite ainsi regagner la confiance des citoyens américains, abîmée par la crise, et favoriser la participation publique. Il présente cette ouverture comme une composante essentielle de la démocratie et un outil d'efficience et d'efficacité du gouvernement⁶⁰.

Huit ans plus tard, le candidat Trump porte un discours beaucoup moins enthousiaste sur l'ouverture des données publiques. Il laisse entendre à plusieurs moments qu'il pourrait revenir sur les décrets pris par Barack Obama, notamment le décret *Open Data* de 2013 qui oblige les agences fédérales à publier leurs données par défaut dans des formats lisibles par des machines. Et dans les semaines qui suivent son élection, Trump fait comprendre aux administrations qu'il veut mieux maîtriser les données publiées sur les sites gouvernementaux, en particulier les données environnementales⁶¹. Les agents anticipent le fait que ces portails *open data* ne soient plus alimentés par des données mises à jour, voire, dans le pire des cas, leur suppression pure et simple ainsi que celle des données associées. Certains d'entre eux prennent alors l'initiative de télécharger massivement les données pour les protéger. Ces efforts, non coordonnés et désorganisés, sont complétés par les actions de chercheurs et d'activistes, qui organisent ensemble de **grands hackathons pour réunir des volontaires qui téléchargent en masse les données gouvernementales**. Les chercheurs en climatologie de l'Université de Toronto lancent même une journée de « guérilla d'archivage » pour protéger les données climatiques des frasques du nouveau gouvernement américain^{62,63}. Ces archivistes bénévoles sont directement aidés par l'administration Obama, qui avant son départ partage une liste des sous-domaines du site web de la Maison Blanche et une liste de toutes les pages web affichées sur le site.

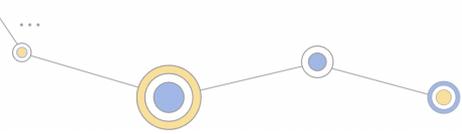
⁵⁹ Joel Gurin, *Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation*, 2014.

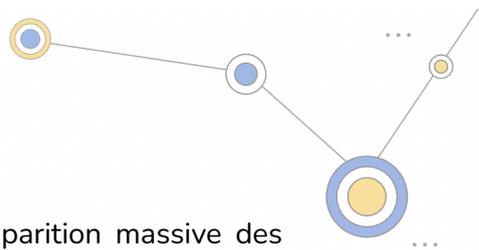
⁶⁰ Barack Obama, « Memorandum on Transparency and Open Government », 2009. URL : www.archives.gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf

⁶¹ Josh Guerstein, « Fears rise of Trump-era 'memory hole' in federal data », in *Politico*, 2016. URL : www.politico.com/story/2016/12/trump-federal-data-fears-232591

⁶² *Ibid.*

⁶³ En France, Christian Quest, ancien agent de la mission Etalab et fervent défenseur de l'open data et de l'open source, a lancé un projet d'archivage de l'ensemble de l'open data français : Opendatarchives.





Plus de peur que de mal ? Il n'est finalement pas constaté de disparition massive des données disponibles en ligne. Mais si elle ne semble pas nuire activement à la politique d'ouverture des données publiques, l'administration Trump ne l'encourage pas. La Maison Blanche supprime symboliquement son portail *open.whitehouse.gov*, considérant qu'il était redondant dans la mesure où toutes les données étaient déjà disponibles sur d'autres sites gouvernementaux et que les 70 000\$ de fonctionnement annuel de ce portail pourraient être mieux investis ailleurs⁶⁴. Par ailleurs, de nombreux experts du climat se plaignent de la soudaine perte de visibilité des données et des rapports relatifs aux questions climatiques sur le site de l'Environmental Protection Agency. Enfin, l'administration Trump baisse les dotations des agences productrices de données comme le Bureau of Economic Analysis, l'Economic Research Service et l'Energy Information Agency⁶⁵.

Dans son ouvrage *Les Données de la Démocratie*, le sociologue Samuel Goëta se base sur cet exemple pour poser la question de la survie du dispositif d'ouverture des données si un gouvernement extrême arrivait au pouvoir en France⁶⁶.

Les données ouvertes, un contre-pouvoir indispensable à la bonne marche démocratique ?

Pour le chercheur Yves Renaud, un contre-pouvoir se caractérise de la façon suivante : il s'agit d'une capacité de contre-expertise et de contre-proposition, qui offre aux citoyens un moyen de faire pression sur la décision et de contrôler la réalisation des politiques publiques, et qui permet *in fine* la mise en place d'une méthode démocratique⁶⁷. En ce sens, **les citoyens qui se saisissent des données publiques mises à disposition constituent un véritable contre-pouvoir**. En ayant accès aux chiffres précis, le citoyen peut refaire un calcul à l'origine du constat qui motive la politique publique, le contester et même faire une proposition alternative⁶⁸.

Ces dernières décennies, **les données publiques sont donc devenues des ressources essentielles au travail des « entrepreneurs de cause »**, qui identifient un dysfonctionnement et le documentent de façon à en faire un véritable problème public qui

⁶⁴ Danny Vinik, « What happened to Trump's war on data? », in Politico, 2017. URL : www.politico.com/agenda/story/2017/07/25/what-happened-trump-war-data-000481/

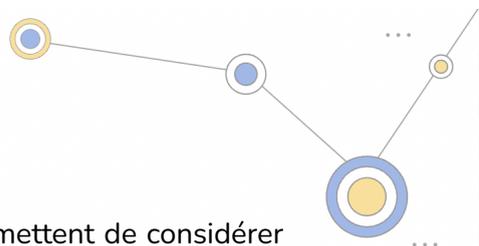
⁶⁵ *Ibid.*

⁶⁶ Samuel Goëta, *Les Données de la Démocratie*, 2024, p.108.

⁶⁷ Yves Renaud, « De la contestation à la concertation », *Annales de la recherche urbaine*, n° 89, 2001, p. 62-69. URL : www.persee.fr/doc/aru_0180-930x_2001_num_89_1_2380.

⁶⁸ Samuel Goëta, *Les Données de la Démocratie*, 2024, p.108.





soit ensuite ajouté à l'agenda politique⁶⁹. Ses récentes enquêtes permettent de considérer Alexandre Léchenet, journaliste et co-fondateur de l'Association des journalistes pour la transparence⁷⁰, comme un « entrepreneur de cause » actif dans le paysage journalistique français. En 2019, le journaliste lance une enquête sur les « écoles orphelines », des écoles primaires dont la composition sociale justifierait qu'elles soient classées en réseau d'éducation prioritaire mais qui ne le sont pas parce que les collèges de secteur ne sont pas suffisamment « défavorisés ». Dans ce cadre, il sollicite le ministère de l'Éducation nationale pour qu'il lui transmette les Indices de Position Sociale (IPS)⁷¹ des classes de CM2 et des collèges. Cet indice est notamment utilisé à Paris pour attribuer des points supplémentaires, appelés bonus IPS, aux collégiens de classe de troisième issus des collèges « défavorisés » pour augmenter leurs chances d'accéder au lycée de leur choix. Parce que l'IPS nourrit ainsi des décisions lourdes de conséquences pour l'avenir des élèves, le journaliste estime normal qu'il soit rendu public⁷². Pourtant, le ministère refuse cette demande, estimant que ces données pourraient contribuer à l'évitement des établissements les plus défavorisés⁷³ et persiste dans son refus malgré l'avis favorable de la CADA, saisie par le journaliste. En juillet 2022, l'administration est finalement condamnée par le Tribunal administratif de Paris à fournir les statistiques sociales de tous les collèges. Il estime en effet qu'il s'agit d'indicateurs agrégés qui ne sont pas susceptibles de porter atteinte à la protection de la vie privée des élèves. Et s'il reconnaît les craintes de renforcer le contournement de la carte scolaire comme légitimes, il juge que la situation ne relève pas de l'une des exceptions au droit d'accès prévues par la législation. Les IPS sont finalement publiés sur le site de l'Éducation nationale⁷⁴ à la fin de l'année 2022. Depuis leur publication, ils ont nourri les travaux de chercheurs et de journalistes, comme ceux des

⁶⁹ Gilbert Claude et Emmanuel Henry, « La définition des problèmes publics : entre publicité et discrétion », *Revue française de sociologie*, vol. 53, no. 1, 2012, pp. 35-59. URL :

www.cairn.info/revue-francaise-de-sociologie-1-2012-1-page-35.htm?contenu=article

⁷⁰ Association créée en décembre 2021 par Laura Motet (Le Monde), Alexandre Léchenet (Gazette des Communes) et Pierre Januel (journaliste indépendant) pour encourager les journalistes à user de leurs droits d'accès aux documents administratifs, et de les accompagner dans les procédures administratives et judiciaires.

⁷¹ L'indice est créé en 2016 par le ministère de l'Éducation nationale pour rendre compte du profil social des élèves. Il prend en compte les professions des deux parents et les combine à d'autres à d'autres critères, comme la taille du logement, les pratiques culturelles, le partage d'une chambre entre plusieurs enfants, etc.

⁷² Le Monde, « L'éducation nationale condamnée à révéler l'indice de position sociale des collèges et des CM2 », 15 juillet 2022. URL :

www.lemonde.fr/societe/article/2022/07/15/l-education-nationale-condamnee-a-reveler-l-indice-de-position-sociale-des-colleges-et-des-cm2_6134916_3224.html

⁷³ Le Monde, « L'éducation nationale condamnée à révéler l'indice de position sociale des collèges et des CM2 », 15 juillet 2022. URL :

www.lemonde.fr/societe/article/2022/07/15/l-education-nationale-condamnee-a-reveler-l-indice-de-position-sociale-des-colleges-et-des-cm2_6134916_3224.html

⁷⁴ data.education.gouv.fr/explore/?sort=modified&q=IPS

« Décodeurs » du Monde qui ont récemment démontré « l'ampleur du "tri social" entre les voies générale et professionnelle⁷⁵ ».

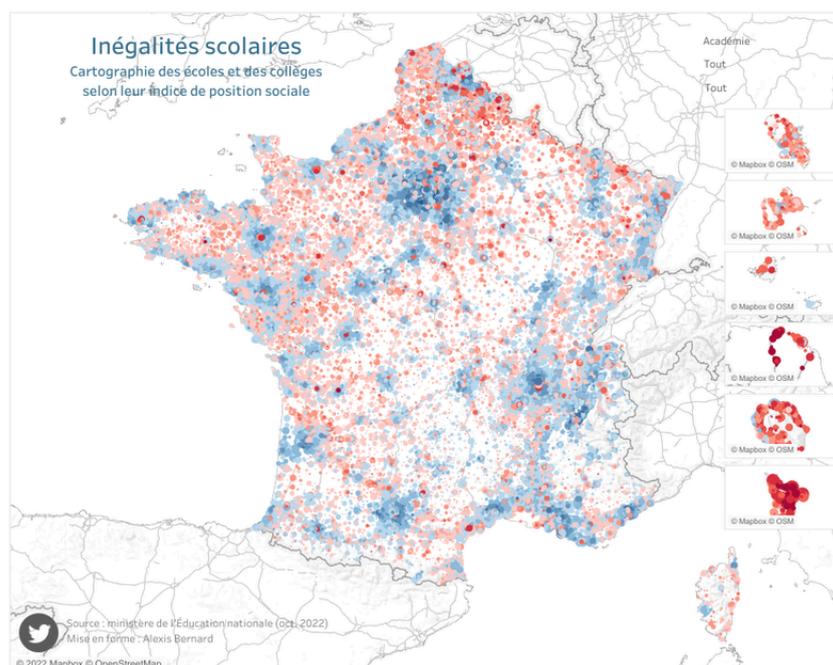


Figure 4 : Carte des inégalités scolaires. Source : Alexis Bernard, 2024.

Ouvrir les données publiques sensibles : garantie d'exemplarité ou moteur d'une « machine à scandales » ?

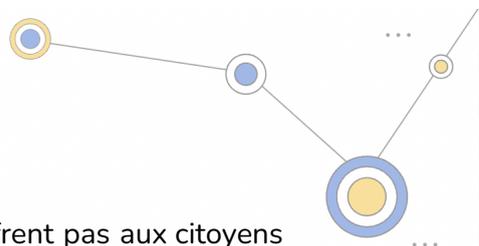
« En permettant à la société civile de constater par elle-même **l'intégrité des institutions** et la **probité des responsables publics**, l'open data peut permettre de franchir une nouvelle étape dans la définition de sociétés démocratiques exemplaires. »

Haute Autorité à la Transparence de la Vie Publique, 2016⁷⁶

Dans un rapport de 2016, la Haute Autorité à la Transparence de la Vie Publique (HATVP) reconnaît le rôle essentiel de l'ouverture des données publiques pour renforcer l'exemplarité des institutions démocratiques et des représentants politiques, sous le regard

⁷⁵ Le Monde, « Au lycée, de nouvelles données révèlent l'ampleur du « tri social » entre les voies générale et professionnelle », 11 janvier 2023. URL : www.lemonde.fr/les-decodeurs/article/2023/01/11/au-lycee-les-indices-de-position-sociale-revelent-l-ampleur-du-tri-social-entre-les-voies-generale-et-professionnelle_6157366_4355770.html#:~:text=L'%C3%A9ducation%20nationale%20a%20%C3%A9%20publier%20les%20IPS%20des%20%C3%A9%20A8ves.

⁷⁶ Haute Autorité à la Transparence de la Vie Publique, « Open data & intégrité publique - les technologies numériques au service d'une démocratie exemplaire », 2016. URL : hatvp.fr/wordpress/wp-content/uploads/2016/12/Open-data-integrite-publique.pdf.

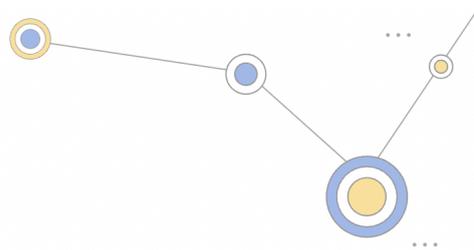


vigilant de « citoyens juges ». Mais toutes les données publiques n'offrent pas aux citoyens la capacité de juger de l'intégrité de leurs dirigeants politiques et administratifs. Il existe en effet des « **données inoffensives** » qui illustrent plutôt qu'elles ne révèlent⁷⁷. Ce sont par exemple des données neutres sur le fonctionnement du service public, comme la localisation des écoles ou les horaires des transports publics. Certains régimes non démocratiques comme la Russie, le Maroc ou Singapour les mettent à disposition gratuitement sous des formats ouverts et se prévalent de ce respect des règles de l'*open data* pour se présenter comme des régimes « ouverts ». Le sociologue Samuel Goëta alerte ainsi sur le caractère uniquement technique de la définition de l'*open data*, qui écarte la question - pourtant cruciale - de la nature des données publiées : « **La libération de données sur la propagande des leaders de Corée du Nord serait susceptible de répondre aux principaux critères du gouvernement ouvert et de l'*open data***⁷⁸ ». Plutôt que de féliciter certains régimes, sur la seule base du respect des critères techniques de l'*open data*, il faudrait donc être vigilant à toute forme de « *transparent washing* », soit une l'ouverture de données inoffensives pour mieux cacher l'absence de réformes réellement nécessaires à la transparence et à la démocratie⁷⁹.

⁷⁷ Samuel Goëta, *Les Données de la Démocratie*, 2024, p.147.

⁷⁸ Harlan Yu et David G. Robinson, « The New Ambiguity of 'Open' Government », *UCLA Law Review Discourse*, 2012. URL : papers.ssrn.com/sol3/papers.cfm?abstract_id=2012489

⁷⁹ N. Heller, « Is Open Data a Good Idea for the Open Government Partnership? », *Global Integrity blog*, 15 septembre 2015.



Pays	OD	D	Pays	OD	D
Royaume-Uni	1	18	Allemagne	16	14
Canada	2	12	Uruguay	17	11
France	3	22	Brésil	18	51
États-Unis	4	30	Suisse	19	7
Corée du Sud	5	24	Italie	20	34
Australie	6	15	Finlande	21	5
Nouvelle Zélande	7	2	Philippines	22	52
Japon	8	16	Singapour	23	70
Pays-Bas	9	9	Colombie	24	53
Norvège	10	1	Russie	25	146
Mexique	11	89	Irlande	26	8
Espagne	12	22	Chili	27	19
Danemark	13	6	Israël	28	29
Autriche	14	20	Belgique	29	36
Suède	15	4	Slovaquie	30	43

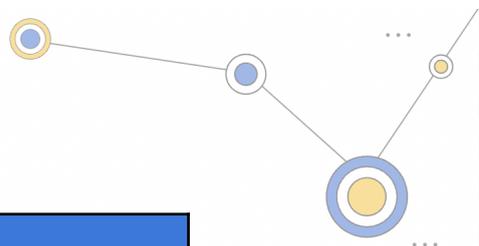
Tableau 2 : Comparaison du classement des pays selon leur indice open data et leur indice démocratique⁸⁰. En rouge, les rangs qui diffèrent de plus de 30 places entre les deux classements. Sources : Economist Intelligence; Open Data Barometer, 2022.

Les contrats publics font partie des données essentielles pour lutter contre la corruption. Dans la déclaration publiée à l'issue du sommet anticorruption tenu à Londres en mai 2016, les 50 pays signataires s'engagent donc à « garantir que l'attribution et l'exécution des marchés publics s'effectuent de manière ouverte, transparente et équitable, dans le respect des lois en vigueur – **l'ouverture des données de la commande publique étant désormais considérée comme l'option par défaut** – afin que les citoyens et les acteurs économiques disposent d'une information de qualité quant à l'emploi de l'argent public ».⁸¹

⁸⁰ Le *democracy index* est un indice créé en 2006 par le groupe de presse britannique *The Economist Group*. Calculé à partir de 60 critères, il permet de comparer le niveau de démocratie des différents pays dans le monde. URL : www.eiu.com/n/campaigns/democracy-index-2023/

⁸¹ Communiqué « Anti Corruption Summit: London 2016 ». URL : www.unodc.org/documents/treaties/UNCAC/WorkingGroups/ImplementationReviewGroup/20-24June2016/V1_603744e.pdf





Catégorie	Données
Intégrité dans le secteur public	Emplois, responsabilités, traitements publics
	Nominations
	Registres des intérêts
	Dons et cadeaux
Activités de lobbying et élaboration des politiques publiques	Registre des lobbyistes
	Rencontre des lobbyistes
	Dons aux partis politiques
	Consultations publiques
Utilisations des ressources publiques	Dépenses budgétaires et attribution des subventions
	Exécution budgétaire et audits financiers
	Historique des dépenses publiques
	Commandes publiques
	Performance des services publics

Tableau 3 : Données politiques fondamentales. Source : HATVP, 2016.

	Part des gouvernements qui publient ces données	Score ⁸² de la France sur 100
Budget	10%	80
Registre des sociétés	5%	95
Dépenses	3%	5
Contrats publics	3%	90
Propriété foncière	1%	5

Tableau 4 : Pourcentage des gouvernements qui publient des données politiquement sensibles et score de la France pour ces mêmes données. Source : Open Data Barometer, 2017.

⁸² Score construit à partir des critères suivants : existence de la donnée, disponibilité en ligne, caractère réutilisable, gratuité, licence libre, fréquence des mises à jour, découvrabilité, présence d'identifiants.



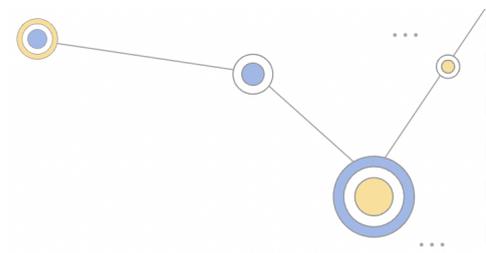
Lutter contre la corruption grâce à l'open data : l'exemple ukrainien

En 2014, la population ukrainienne dénonce la corruption généralisée du régime en place. À l'issue de cette Révolution de la Dignité, un groupe d'activistes civiques et d'experts en données se mobilisent pour proposer une réforme profonde du système de passation des marchés publics et permettre aux petites et moyennes entreprises d'enfin concourir de manière équitable au sein d'un système dominé par des proches du pouvoir et des intérêts puissants. Aidés par la Banque européenne pour la reconstruction et le développement et par l'Agence américaine pour le développement international, ils créent la plateforme *open source* ProZorro (« transparence » en ukrainien) sur laquelle sont déposées les données relatives aux contrats. Ces données sont ensuite analysées par une communauté de citoyens et d'acheteurs gouvernementaux, les Dozorro (« chiens de garde » en ukrainien), qui peuvent ainsi signaler les transactions à haut risque et les irrégularités aux autorités gouvernementales.

Les progrès réalisés grâce à cette plateforme sont considérables : 12 mois après son lancement, plus de 46 000 offres relatives à des marchés représentant un montant total de 300 millions de dollars ont été publiées. Le nombre moyen d'entreprises répondant aux appels d'offre passe de 2,1 à plus de 2,8⁸³. En octobre 2017, le pays aurait ainsi économisé près de 6 milliards de dollars de fonds publics⁸⁴.

⁸³ Haute Autorité à la Transparence de la Vie Publique, « Open data & intégrité publique - les technologies numériques au service d'une démocratie exemplaire », 2016. URL : hatvp.fr/wordpress/wp-content/uploads/2016/12/Open-data-integrite-publique.pdf.

⁸⁴ Harvard Kennedy School, « Overcoming Corruption and War: Lessons from Ukraine's ProZorro Procurement System », 2022. URL : www.hks.harvard.edu/publications/overcoming-corruption-and-war-lessons-ukraines-prozorro-procurement-system



1.2.2. Un enjeu économique et de souveraineté

« Aujourd'hui, j'annonce que nous mettons à disposition des citoyens encore plus de données gouvernementales et que nous les rendons plus faciles à trouver et à utiliser. Cela va permettre de lancer **plus de start-ups**. Cela va permettre de lancer **plus d'entreprises**. Cela va permettre de créer **plus d'emplois**. Cela va aider plus d'entrepreneurs à **créer des produits et des services** que nous n'avons pas encore imaginés. Ce type d'**innovation** et d'**ingéniosité** a le potentiel de transformer la façon dont nous faisons presque tout. »

Barack Obama, annonçant le décret « *Making open and machine readable the new default for government information* » le 9 mai 2013⁸⁵ (traduction des auteurs).

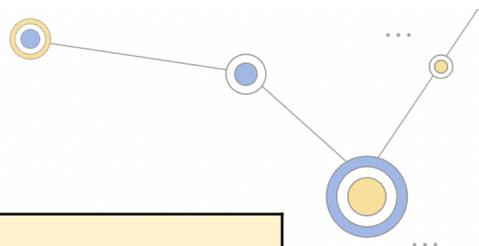
Des entreprises qui ont rapidement perçu le potentiel de création de valeur associé à l'ouverture des données publiques

Les entreprises américaines s'impliquent très tôt dans les mouvements appelant à l'ouverture des données publiques. Sous l'administration Obama, elles joueront un rôle actif dans la promotion d'une politique d'ouverture des données publiques. Le jour de l'annonce de l'ordonnance Obama sur l'ouverture des données publiques par défaut, le lobbyiste Kevin Richards, représentant les intérêts de l'organisation professionnelle TechAmerica, déclare ainsi : « *L'accès à la quantité monumentale de données gouvernementales alimentera un nombre incalculable de nouvelles idées innovantes dans ce pays. En faisant de l'ouverture des données la politique par défaut de l'ensemble du gouvernement fédéral au lieu d'une politique discrétionnaire, le président Obama a donné à l'industrie technologique américaine une clé pour développer notre leadership mondial à l'ère du Big Data.* »⁸⁶

⁸⁵ obamawhitehouse.archives.gov/blog/2013/05/10/recap-big-day-open-data

⁸⁶ *Ibid.*





La rencontre de Sebastopol, Californie

En décembre 2007 se rencontrent à Sebastopol (Californie) la Sunlight Foundation, Google et Yahoo! dans les locaux de la maison d'édition d'O'Reilly. L'objectif de cette rencontre est de lister les grands principes d'un *Open Government* afin qu'il améliore la participation civile tout en créant de nouvelles opportunités pour les entrepreneurs. Un gouvernement ouvert doit fournir des :

1. Données complètes
2. Données primaires (telles que collectées à la source)
3. Données fraîches (rendues disponibles dès leur création)
4. Données accessibles
5. Données lisibles par les machines
6. Données non discriminatoires (sans enregistrement préalable)
7. Données dans un format ouvert
8. Données sous licence ouverte (placées dans le domaine public)

Les acteurs privés comprennent en effet très rapidement le potentiel économique des données publiques. En 2014, Joel Gurin, président et fondateur du Centre pour les entreprises *open data* (CODE⁸⁷), publie *Open Data Now*⁸⁸. Dans ce guide destiné aux entrepreneurs pour les aider à créer de la valeur à partir de cette nouvelle matière qu'est la donnée publique, il liste 500 entreprises américaines dont l'activité repose en grande partie sur les données publiques. Il analyse plusieurs exemples parmi lesquels le cas de la startup Enigma⁸⁹ qui propose des études de marché aux petites entreprises. En utilisant les bases de données ouvertes de la Commission fédérale des communications, elle récupère les demandes d'interphones pour les vitrines ; elle en déduit les futurs emplacements des nouveaux restaurants McDonald's et les risques ou opportunités que cela peut créer pour d'autres types d'implantations commerciales⁹⁰. On peut également citer l'ouverture des données du *National Weather Service* aux États-Unis, qui a notamment permis la création de la startup *Climate Corporation*, qui propose des assurances aux agriculteurs.

⁸⁷ Center for Open Data Enterprise. URL : www.opendataenterprise.org/

⁸⁸ Joel Gurin, *Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation*, 2014.

⁸⁹ www.enigma.com

⁹⁰ www.strategy-business.com/article/14102

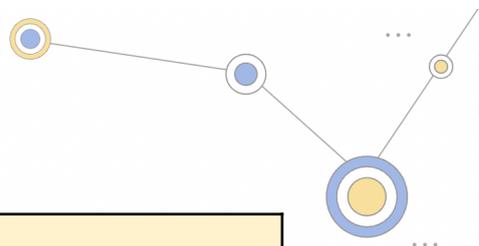


***Améliorer les transports publics grâce à l'ouverture des données :
l'exemple londonien***

En 2007, l'entreprise publique *Transport for London*, en charge des transports publics londoniens, est une des premières entreprises de transport à mettre à disposition ses données gratuitement sur une plateforme Internet.

Dix ans plus tard, une étude du cabinet Deloitte dresse le bilan de cette ouverture. En une petite décennie, 600 applications ont été créées grâce à ces données et sont utilisées par 42% des Londoniens⁹¹. Le gain ainsi créé pour les voyageurs est estimé à 95M£/an : ces derniers gagnent du temps en optimisant mieux leurs trajets avec des transports plus fiables et en ne payant plus l'information (suppression du service payant d'alerte SMS en cas de retard ou d'annulation). Les trajets facilités encouragent les usagers à voyager davantage, ce qui représenterait un gain en billetterie d'environ 20M£/an pour l'entreprise. Enfin, l'ouverture des données bénéficie à la ville de Londres, dont le meilleur système de transport améliore l'attractivité, favorisant l'arrivée de nouvelles populations et d'activités économiques. Par ailleurs, la création de startups autour de ces données de transport nourrit le dynamisme de l'écosystème tech local.

⁹¹ Deloitte, « Assessing the value of TfL's open data and digital partnerships », 2017. URL : content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf



Doctrine

Faciliter l'accès à l'information juridique : l'exemple de Doctrine

En 2016, l'entreprise française Doctrine crée une plateforme qui centralise l'information juridique et légale et propose une solution basée sur l'intelligence artificielle pour retrouver l'information de façon plus ciblée, plus efficace et contextualisée. Dans une étude de 2019⁹², le portail européen des données estime que si les données publiques ouvertes ne représentent que 20% des données utilisées sur Doctrine, elles génèrent plus de 80% des revenus de l'entreprise.

Au-delà de l'intuition : des estimations chiffrées très optimistes

« L'open data est la clé vers une meilleure gouvernance, des marchés plus efficaces et des décisions commerciales plus judicieuses ».

Tom Lee, fondateur de la Sunlight Foundation (traduction des auteurs)⁹³

Une fois ouvertes, les données publiques permettent de réduire les asymétries d'information, ce qui favorise un meilleur appariement de l'offre et de la demande ainsi qu'une meilleure allocation des ressources. De cette façon, les données ouvertes :

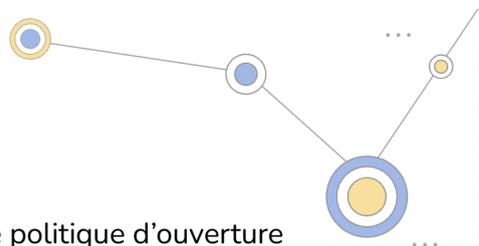
- Nourrissent et éclairent les décisions des administrations, des entreprises et des citoyens.
- Permettent l'émergence de nouveaux produits, services et marchés.
- Favorisent des gains d'efficacité pour les administrations et les entreprises⁹⁴.

Les données ouvertes favorisent ainsi la croissance économique de façon directe et indirecte. De façon directe dans la mesure où les entreprises qui fournissent des biens et services basés sur des données ouvertes génèrent des bénéfices et créent des emplois. De façon indirecte puisque l'utilisation de ces nouveaux biens et services offre des bénéfices à leurs utilisateurs : gains de temps, croissance de l'économie de la connaissance, efficacité des services publics, croissance des marchés connexes.

⁹²Commission européenne, « The Economic Impact of Open Data - Opportunities for value creation in Europe », 2019. URL : data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf

⁹³ *Ibid.*

⁹⁴ Artur Kovalchuk, Viktor Khanzhyn, Yaroslav Kudratskiy, « Economic potential of Open data for Ukraine », 2018. URL : tapas.org.ua/wp-content/uploads/2019/02/TAPAS_PDF_1.pdf



Plusieurs études cherchent à chiffrer l'impact direct que peut avoir une politique d'ouverture des données publiques sur l'économie. En 2013, le cabinet McKinsey estime le marché mondial de l'*open data* proche des 3,2 trillions de dollars⁹⁵. En 2015, c'est au tour de Capgemini de se prêter à l'exercice pour évaluer l'impact quantitatif de la réutilisation des données ouvertes publiques dans les secteurs public et privé en Europe⁹⁶. Il s'agit de documenter les preuves économiques des avantages de la réutilisation des données publiques dans l'Union européenne en fonction du stade de maturité *open data* des pays membres (débutant, avancé, suiveur et précurseur). Quatre métriques sont considérées dans cette étude :

- **Le volume du marché total des données ouvertes**, soit le marché des produits, services et contenus permis ou améliorés par des données ouvertes. Il est estimé entre 193 et 209 milliards d'euros pour l'année 2016.
- **Les emplois créés** directement (*data analysts*) et indirectement (les effectifs des ressources humaines nécessaires pour recruter ces nouveaux *data analysts*). Selon l'étude Capgemini, 75 000 emplois directs pourraient ainsi être créés en 2016.
- **Les économies réalisées sur les coûts des entreprises et des administrations**. Capgemini estime qu'ils pourraient atteindre 1,7 milliards d'euros d'ici 2020.
- **Les gains d'efficacité réalisés par la société dans son ensemble** : Comment les données ouvertes peuvent-elles sauver des vies sur les routes ? Comment les utiliser pour gagner du temps ? De quelle façon contribuent-elles au développement durable et à la préservation de l'environnement ?

Pour déterminer la valeur de chacune de ses métriques, les équipes de Capgemini appliquent la « logique des avantages ». Les auteurs n'entrent pas dans les détails de la méthode mais en résumant le principe de la façon suivante : il s'agit de déterminer quels indicateurs économiques influencent chaque métrique pour dessiner des solutions en partant du résultat souhaité. Dans l'étude de 2015 menée pour le compte de la Commission européenne, il a par exemple été choisi de partir de l'objectif de résultat « croissance de la taille du marché » (Figure 5).

⁹⁵ James Manyika et al., « Open data: Unlocking innovation and performance with liquid information », McKinsey, 2013. URL : www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20Data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx

⁹⁶ Capgemini Consulting, « Creating Value through data », 2015. URL : data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

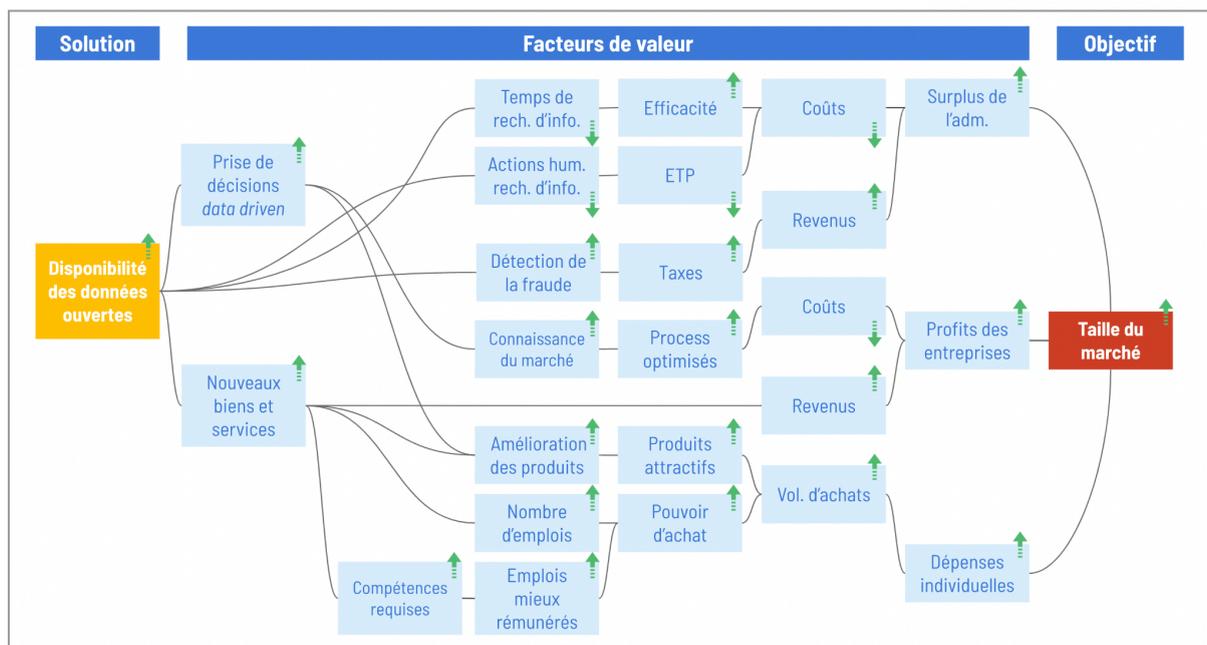
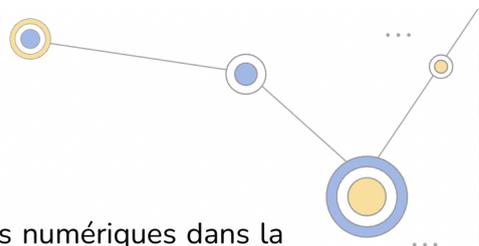


Figure 5 : Enchaînement des mécanismes qui aboutissent à l'augmentation de la taille du marché grâce à l'ouverture de données publiques selon la méthode « Logique des avantages » conçue par Capgemini. Source : Capgemini, 2015.

En 2019, la Commission européenne commande une nouvelle étude sur l'impact de la réutilisation des données ouvertes publiques et privées⁹⁷. Elle affirme qu'en 2019, le marché de la réutilisation de la données publique frôle les 200 milliards d'euros et représente plus d'un million d'emplois directs. Pour parvenir à ce résultat, les auteurs de l'étude applique au PIB européen de 2019 un pourcentage de 1,19% correspondant à la médiane des estimations de la valeur ajoutée associée à l'ouverture des données de différentes études. Cette estimation est donc moins optimiste que l'estimation de 2015, qui faisait l'hypothèse que la valeur ajoutée associée à l'ouverture des données représentait 1,35% du PIB européen⁹⁸. En projetant ces estimations de valeur ajoutée sur les prévisions de croissance du PIB européen, l'étude peut proposer des scénarios sur la taille du marché européen de l'open data à horizon 2025 : celle-ci pourrait alors représenter entre 199 et 334 milliards d'euros. Le dernier scénario, très optimiste, considère que la part du PIB généré par l'ouverture des données publiques devrait considérablement augmenter entre 2019 et 2025 grâce aux directives européennes, à la maîtrise croissante des sujets liés au Règlement général sur la protection des données (RGPD), à la montée en compétences des

⁹⁷ Commission européenne, « The Economic Impact of Open Data - Opportunities for value creation in Europe », 2019. URL : data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf

⁹⁸ L'étude Capgemini réalisée en 2015 pour la Commission européenne ne précise pas qu'elle applique ce pourcentage de 1,35% et n'explique donc pas pour quelles raisons ce ratio a été choisi.



secteurs public et privé et à la meilleure pénétration des applications numériques dans la société⁹⁹.

Un nouvel enjeu économique fort associé à l'ouverture des données publiques : le développement de modèles d'intelligence artificielle générative francophones performants

Ces dernières années, la course au développement de modèles d'intelligence artificielle et d'intelligence artificielle générative s'est emballée, renouvelant la question de l'accès aux données dans la mesure où la création, l'entraînement et l'adaptation de ces modèles requièrent un volume considérable de données. Pour cette raison, l'ouverture des données publiques et des données publiques culturelles peut offrir une ressource essentielle et précieuse aux chercheurs et développeurs travaillant sur des modèles d'intelligence artificielle.

L'ouverture des données publiques devient donc un enjeu économique fort, dans le cadre d'une intense compétition internationale, mais aussi un enjeu important de souveraineté. Ouvrir ses données publiques est une façon pour le pays de maîtriser davantage la nature et la qualité des données utilisées pour entraîner les modèles et de permettre le développement de capacités d'interaction spécifiques pour des applications nationales. Il y a un intérêt réel à entraîner les modèles sur le corpus patrimonial français, et notamment sur les archives numérisées, car les modèles d'intelligence artificielle sont encore quasiment exclusivement entraînés sur corpus anglophones, expliquant qu'ils soient bien moins performants et moins pertinents en français qu'en anglais¹⁰⁰.

Considérant ces nouveaux enjeux, la commission Intelligence artificielle¹⁰¹ regrette « l'ouverture encore trop limitée des données publiques » et déplore même un ralentissement du mouvement d'ouverture ces dernières années. Elle exhorte donc l'administration française à relancer la dynamique pour mieux soutenir les acteurs français de l'intelligence artificielle.

⁹⁹ Lateral Economic pour l'Open Data Institute, « Permission granted: The economic value of data assets under alternative policy regimes », 2016. URL :

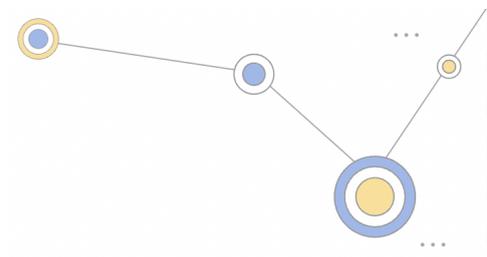
theodi.org/insights/reports/research-the-economic-value-of-open-versus-paid-data/

¹⁰⁰ Commission IA, « IA : Notre ambition pour la France », 2024, p.103. URL :

www.info.gouv.fr/upload/media/content/0001/09/4d3cc456dd2f5b9d79ee75f6ea63b47f10d75158.pdf

¹⁰¹ Commission IA, « IA : Notre ambition pour la France », 2024, p.103. URL :

www.info.gouv.fr/upload/media/content/0001/09/4d3cc456dd2f5b9d79ee75f6ea63b47f10d75158.pdf



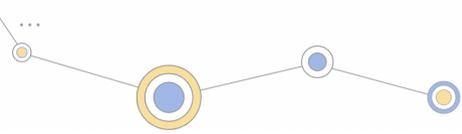
1.2.3. Un enjeu d'amélioration des services publics

En 2015, dans son rapport sur le potentiel de création de valeur associé à l'ouverture des données publiques¹⁰², le cabinet Capgemini insiste sur les différences de bénéfices selon les secteurs considérés et conclut que le secteur public est le secteur qui profite et profitera le plus largement de la réutilisation des données ouvertes, avec un gain estimé à 22 milliards d'euros pour l'Union européenne en 2020. La loi pour une République numérique, qui oblige les administrations à ouvrir leurs données, représenterait donc moins une contrainte supplémentaire qu'une nouvelle opportunité pour les acteurs publics.

Amélioration de la qualité des données

Lorsqu'elles publient leurs données, les administrations s'exposent aux regards et aux critiques des citoyens. L'ouverture des données devient alors un puissant levier pour améliorer la qualité des jeux de données. En sortant de la confidentialité, les données sont confrontées aux yeux neufs et exigeants de la société civile, des journalistes et des chercheurs qui peuvent signaler à l'administration concernée une erreur ou un oubli dans un jeu de données. Les forums de discussions associées aux jeux de données publiés sur data.gouv.fr regorgent ainsi d'exemples de réutilisateurs qui ont identifié un problème et demandent à l'administration de le corriger (Figure 6).

¹⁰² Capgemini Consulting, « Creating Value through data », 2015. URL : data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf



Pb dans le fichier Excel "maires-au-17-06-2020.xlsx" Copier le lien vers la discussion

 Jean-François
9 juin 2023

Bonjour,
A toutes fins utiles, je signale plusieurs problèmes dans le fichier Excel "maires-au-17-06-2020.xlsx"
Tout d'abord le format qui transforme les champs "code_departement" et "code_commune" en numérique. Le format CSV serait plus approprié.
Ensuite et surtout il y a quatre doublons pour 4 communes :
/*
4 lignes ne sont pas insérées car elles ne respectent pas la contrainte d'intégrité
Champ0 Champ1 Champ2 Champ3 Champ4 Champ5 Champ6 Champ7 Champ8 Champ9 Champ10
32 GERS 324 Ponsan-Soubiran CAZES Philippe M10/02/1958 71 Anciens agriculteurs exploitants 18/05/2020
57 MOSELLE 214 Fixem SCHREIBER Aurore F 14/02/1976 37 Cadres administratifs et commerciaux d'entreprise 18/05/2020

[Lire plus](#)

[Répondre >](#)

Problème de type de champs Copier le lien vers la discussion

 Jean-Michel
9 juin 2020

Bonjour,
Je relève des formats de champs mal définis.
Les champs "code dpt" et "code commune" sont définis en "standard" faisant sauter les 0 non significatifs
ex : commune 001 apparait en "1" dans les fichiers TXT et XLSX.
Pouvez-vous corriger cela ?
Merci d'avance

Figure 6 : Erreurs relevées par les réutilisateurs sur la discussion du jeu de données « Élections municipales 2020 - Liste des candidats élus au T1 et liste des communes entièrement pourvues » publiées par le ministère de l'Intérieur sur data.gouv.fr¹⁰³.

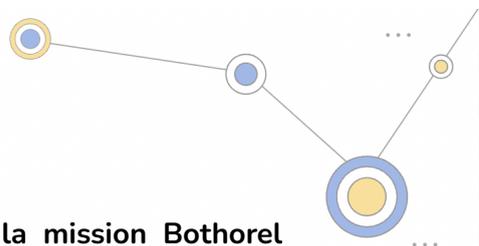
Mais surtout, l'ouverture des données participe de façon plus indirecte à l'amélioration de la qualité : dans la mesure où les administrations savent que les données qu'elles publient forgent aussi leur image, elles voudront anticiper les éventuelles critiques et apporter toutes les corrections nécessaires en amont de la publication. Toutefois, si la crainte de la critique peut jouer un rôle positif, elle peut également ralentir la dynamique d'ouverture des données, retenues par des administrations qui attendront d'avoir des « jeux de données parfaits » pour les publier.

Meilleure circulation des données au sein des administrations : partager pour gagner en efficacité et offrir un meilleur service public

Si l'objectif premier d'une politique d'ouverture des données publiques est la création de flux de données vers l'extérieur de l'administration, il s'est rapidement avéré que la publication de ces données présentait un autre avantage : celui d'améliorer, voire de créer,

103

www.data.gouv.fr/fr/datasets/elections-municipales-2020-liste-des-candidats-elus-au-t1-et-liste-des-communes-entierement-pourvues/#discussions



des circulations de données internes à l'administration. En 2020, **la mission Bothorel dénonce en effet un partage des données « scandaleusement faible » entre administrations**, conduisant parfois certains agents à devoir ressaisir des données disponibles dans des bureaux voisins. L'ouverture des données est alors le seul moyen pour ces agents de prendre connaissance de l'existence de données produites au sein même de leur administration, alors même qu'elles peuvent s'avérer essentielles pour la bonne conduite de leurs missions¹⁰⁴. Le rapport souligne que le partage est particulièrement difficile entre les administrations centrales et les collectivités territoriales, entre lesquelles les voies officielles de transferts de données sont quasiment inexistantes.

En échangeant davantage leurs données, par l'intermédiaire de leurs portails *open data* ou via des canaux non publics pour les données les plus sensibles, les agents peuvent exploiter le travail déjà accompli. Ils gagnent ainsi du temps et sont plus efficaces dans leurs missions. Ils peuvent surtout réutiliser ses données en interne pour les croiser avec des données issues de leurs propres services. Ils prennent ainsi de meilleures décisions, basées sur l'analyse d'informations sourcées et fiables¹⁰⁵. **Par ailleurs, la publication des données publiques représente un gain de temps non négligeable pour les agents, qui n'ont alors plus besoin de répondre aux demandes d'accès aux documents administratifs.**

¹⁰⁴ Éric Bothorel, Rapport de mission « Pour une politique publique de la donnée », 2020, p.9. URL : www.ouvrirlascience.fr/wp-content/uploads/2021/02/Mission-Bothorel_rapport-pour-une-politique-publique-d-e-la-donnee_23.12.2020.pdf

¹⁰⁵ Cadre interministériel d'administration de la donnée, DINUM, 2021. URL : www.numerique.gouv.fr/uploads/Cadre-interministeriel-donnee.pdf



base adresse **nationale**

Partager les adresses pour une base plus uniformisée et de meilleurs services publics

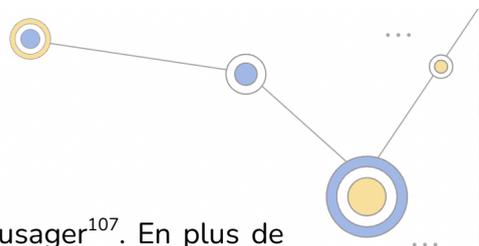
La Base Adresse Nationale est l'un des neuf jeux de données de référence du service public de la données. Elle rassemble plus de 25 millions d'adresses officiellement reconnues par l'administration. Depuis janvier 2020, elle est disponible en licence ouverte Etalab 2.0 sur adresse.data.gouv.fr. L'ouverture de cette base garantit à tous les citoyens de disposer de l'adresse qui permettra l'intervention des services d'urgence, le raccordement aux réseaux d'eau, d'énergie et de fibre optique, la distribution du courrier, la collecte des déchets ménagers, etc.

Avant 2019, l'IGN produisait des adresses à partir des informations collectées auprès d'une centaine de collecteurs dans les territoires, en lisant la presse régionale, en interagissant avec les offices du tourisme et les collectivités territoriales. Ce dispositif était considéré comme inefficace et trop coûteux.

En 2019, la DINUM, l'Agence nationale de la cohésion des territoires (ANCT) et l'IGN ont lancé le projet Base Adresse Nationale (BAN)/Bases Adresses Locales (BAL) : les collectivités territoriales sont désormais responsables de la la collecte des données au niveau local¹⁰⁶ et l'État, à travers l'IGN, conserve la responsabilité d'agrèger les BAL dans une base de données nationale.

La politique d'*open data* a donc permis un meilleur partage de données publiques entre administrations. S'en est rapidement suivie une réflexion plus large sur la possibilité de mettre en place de canaux d'échanges de données sécurisés pour permettre le transfert de données qui n'ont pas vocation à être rendues publiques mais dont le partage entre administrations est crucial pour l'efficacité du service public. La loi relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale, dite loi 3DS, adoptée en février 2022, réaffirme l'importance de mettre en place des canaux de partage de données entre administration pour renforcer le dispositif « **Dites-le-nous une fois** ». La loi modifie l'article L. 113-12 du CRPA pour que les citoyens ne soient plus tenus de produire des informations qu'ils ont déjà fournies. Les administrations sont désormais responsables de s'échanger entre elles

¹⁰⁶ Article 169 de la loi n° 2022-217 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale.



les documents dont elles ont besoin pour traiter la demande de l'utilisateur¹⁰⁷. En plus de simplifier les démarches administratives des usagers, **ce principe de partage « par défaut » entre administrations permet la circulation de données** nécessaires pour informer les personnes proactivement sur leurs droits sociaux et éventuellement leur attribuer automatiquement, diminuant ainsi les risques de fraude et de non-recours.



DIRECTION GÉNÉRALE DES
FINANCES PUBLIQUES

Simplification et lutte contre le non-recours aux droits : l'API Impôt Particulier

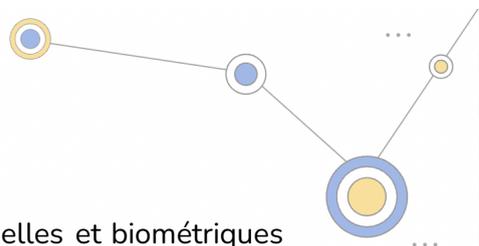
L'API Impôt particulier est une API mise en place par la DGFiP pour répondre au principe du « Dites-le-nous une fois ». Elle permet aux collectivités territoriales d'échanger avec le système d'information de la direction de façon sécurisée et en temps réel afin de récupérer des données ciblées issues de la déclaration de revenus de leurs administrés. Il suffit à l'utilisateur de se connecter sur le site sur lequel il effectue sa démarche, d'y indiquer son identifiant fiscal ou son état civil complet. La collectivité territoriale récupère alors immédiatement ses informations.

De cette façon, les collectivités peuvent traiter directement les demandes de ces derniers en leur proposant les prix correspondant à leurs niveaux de revenus. Ce service est particulièrement utilisé par les mairies qui souhaitent savoir quel doit être le prix d'une inscription en crèche ou d'un repas à la cantine scolaire en fonction des foyers.

Si le principe du « Dites-le-nous une fois » simplifie réellement les démarches des usagers, les échanges de données entre administrations ne suscitent pas l'adhésion de tous les Français. Certains dénoncent le risque de constituer un « grand fichier national » obtenu en rassemblant et en croisant toutes les données collectées par les différentes administrations sur chacun des administrés. En 1974, le projet Safari, qui prévoit la centralisation et l'uniformisation des fichiers de police, fait scandale. Au point que le premier ministre, Pierre Messmer, doit y mettre fin et crée une commission « informatique et libertés » chargée d'établir un cadre juridique à la constitution de ce genre de fichier. Il jette ainsi les bases de la loi « informatique et libertés » qui sera adoptée quatre ans plus tard. Cette crainte n'a aujourd'hui pas disparu. Elle se pose par exemple de nouveau en 2016¹⁰⁸ : la Ligue des droits de l'homme proteste alors contre la création par décret du fichier « titres

¹⁰⁷ La loi précise que cette obligation pour les administrations ne vaut que si elles sont techniquement capables de s'échanger les documents. Dans le cas contraire, elles en sont exemptées.

¹⁰⁸ theconversation.com/le-mega-fichier-de-la-discorde-69107



électroniques sécurisés », qui centraliserait des informations personnelles et biométriques sur la quasi-totalité de la population française¹⁰⁹.

Par ailleurs, l'échange de données entre administrations peut aussi susciter des craintes et des réticences en interne, lorsque l'administration « fournisseuse de données » estime que le niveau de sécurité du système d'information de l'administration « demandeuse » est insuffisant et susceptible de mener à des fuites de données. L'administration « fournisseuse » peut alors craindre que cette fuite ne dissuade ses administrés de renseigner leurs informations.

Mise en valeur de l'action publique

L'ouverture des données publiques est un outil puissant de valorisation du travail des administrations. Les réutilisations des données publiques, qu'elles soient internes ou externes aux administrations, mettent en valeur le travail des agents, ce qui constitue pour eux une motivation supplémentaire pour devenir proactifs dans la publication des données de leurs services. Par ailleurs, les administrations peuvent ajouter sur leurs portails *open data* des tableaux de bord avec des indicateurs clairs et factuels qui témoignent de la bonne conduite des politiques publiques. Ces indicateurs et ces visualisations éclairent le citoyen sur la nature de l'action publique et illustrent l'impact concret des réformes et des mesures gouvernementales. *In fine*, on peut espérer que le partage de données favorise ainsi le consentement à l'impôt.

Sur son portail *open data*, le ministère de l'Économie propose par exemple des cartes qui permettent de visualiser la façon dont se répartissent les aides au développement distribuées par l'Agence française pour le développement dans le monde et entre les différents secteurs (Figure 7). Mais il propose également au citoyen de construire ses propres cartes et graphiques à partir des jeux de données ouverts par le ministère (Figure 8). De cette façon, le citoyen peut réellement s'approprier les informations qui sont mises à sa disposition.

¹⁰⁹ www.ldh-france.org/toussaint-2016-retour-du-fichier-gens-honnetes/

L'aide publique au développement par secteur

Le comité d'aide au développement de l'OCDE est le garant de la comptabilisation de l'aide publique au développement des principaux bailleurs. Chaque année, il rassemble, valide et publie les données de l'aide publique au développement. Il définit par ailleurs les activités pouvant être comptabilisées comme relevant de l'APD, qui comptent pour plus de 340 secteurs. Les flux présentés ci-dessous ont été agrégés au sein de 26 secteurs.

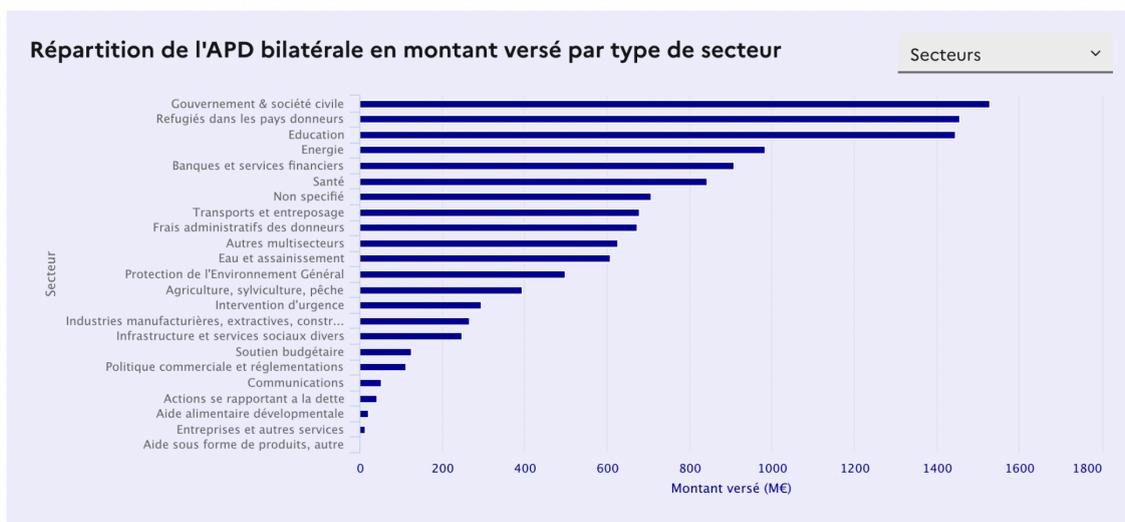


Figure 7 : Visualisation de la répartition par secteur des aides publiques au développement. Source : data.economie.gouv.fr¹¹⁰, 2022.

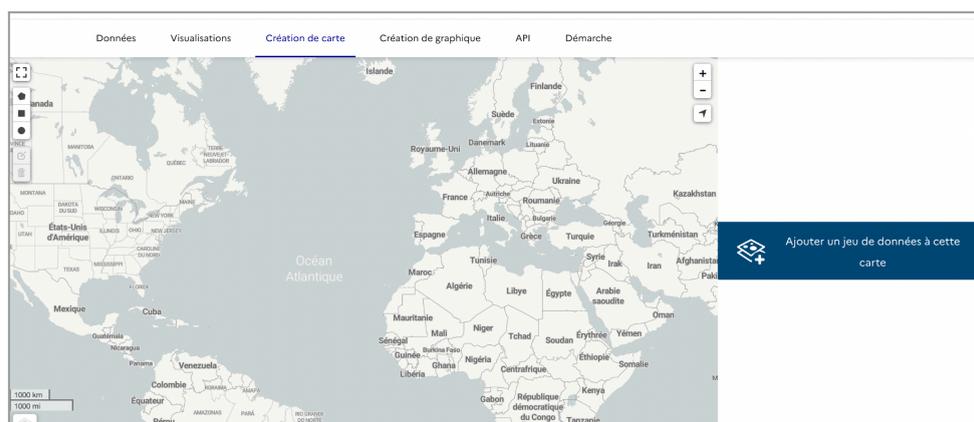


Figure 8 : Outil de construction de visualisations cartographiques et graphiques du ministère de l'Économie et des Finances. Source : data.economie.gouv.fr¹¹¹.

Le gouvernement français semble avoir saisi le potentiel de communication de la mise à disposition de données sur son action. Pour cette raison, il lance en janvier 2021 le « baromètre des résultats de l'action publique ». Nourri par les administrations centrales et par les préfetures, cet outil permet au citoyen de suivre la progression des mesures, politique par politique, territoire par territoire, grâce à des tableaux d'indicateurs de suivi (Figure 9).

¹¹⁰ data.aide-developpement.gouv.fr/pages/chiffres_cles/?q.tabs=sectorielle

¹¹¹ data.economie.gouv.fr/map/+9c0e2e4b5d864cde/edit/

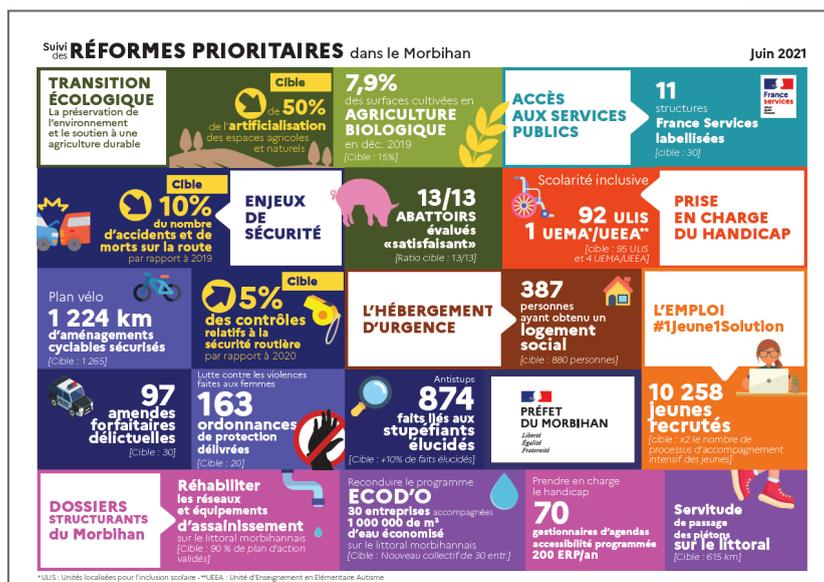
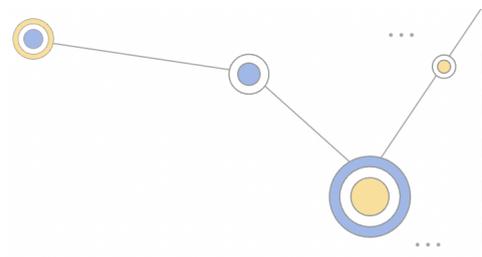
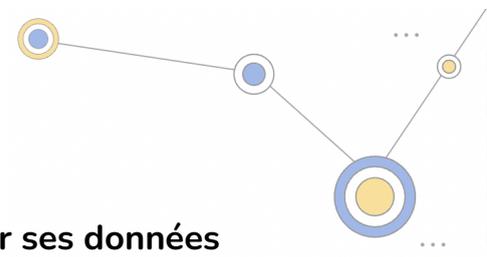


Figure 9 : Suivi des réformes prioritaires, Baromètre des résultats de l'action publique de la préfecture du Morbihan¹¹², juin 2021.



2. Les difficultés sous-estimées de l'*open data* et de l'*open data* par défaut





2.1. Des efforts conséquents mais nécessaires pour ouvrir ses données

2.1.1. Sur le plan technique : des systèmes d'information qui n'ont pas été pensés pour l'ouverture des données

« Sans une prise au sérieux des difficultés rencontrées par les agents, ces fractures risquent de s'approfondir et l'open data restera un projet annexe dans les politiques publiques numériques. »¹¹³

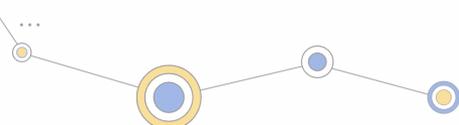
Samuel Goëta, 2024

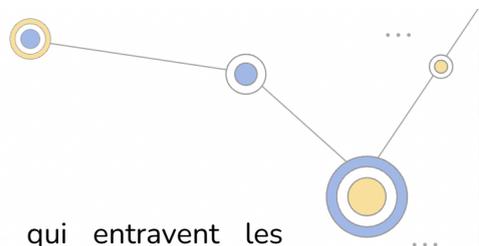
Au fur et à mesure de la numérisation de leurs usages et de leurs organisations, les administrations ont construit leurs systèmes d'information. Il s'agit d'ensembles organisés de ressources, de procédures et de technologies qui collectent, stockent, traitent et distribuent des informations. Ils comprennent :

- **Des ressources humaines** : les personnes qui utilisent, gèrent et administrent le système d'information.
- **Des ressources matérielles** : les équipements physiques utilisés pour traiter, stocker et communiquer des informations. Il s'agit notamment des serveurs, des ordinateurs, des périphériques de stockage et des réseaux informatiques.
- **Des données** : les informations brutes collectées, stockées et traitées par le système d'information sous forme de textes, de chiffres, d'images, de vidéos, etc.
- **Des ressources logicielles** : les programmes informatiques utilisés pour collecter, stocker, traiter et présenter des informations. Cela peut inclure des applications métier, des systèmes de gestion de bases de données, des logiciels de visualisation de données, des systèmes d'exploitation, etc.
- **Des procédures** : les règles et les processus définis pour gérer et utiliser le système d'information de manière efficace et sécurisée. Il s'agit par exemple de politiques de sécurité, des procédures de sauvegarde, du traitement des données, etc.
- **Des technologies de communication** : les outils et les infrastructures utilisés pour échanger des informations entre les différentes parties du système d'information et avec des systèmes externes. Cela peut inclure des réseaux informatiques, des protocoles de communication, des services web, etc.

Les systèmes d'information des administrations tels qu'ils existent aujourd'hui ont été pensés et construits pour une circulation de données essentiellement interne aux systèmes de l'organisation. L'extraction et le partage de ces données à l'extérieur n'étant pas toujours anticipés, ils donnent alors lieu à des « **frictions de données** ». Ce concept, développé par le

¹¹³ Samuel Goëta, *Les Données de la Démocratie*, 2024, p.191.





chercheur Paul Edwards¹¹⁴, décrit les forces socio-matérielles qui entravent les mouvements de données et l'ensemble des efforts qu'il faut fournir pour surmonter ces difficultés : il faut du temps, de l'énergie et de l'attention pour collecter, vérifier, stocker, déplacer, recevoir et accéder aux données.

De la difficulté d'extraire les données

Les données des administrations sont stockées dans des systèmes d'information qui sont d'abord et avant tout les outils de travail quotidien des agents. L'usage métier quotidien prime donc pour leur conception, souvent au détriment de l'export des bases de données rarement prévu dans les fonctionnalités des applications achetées par les administrations. Les utilisateurs sont dépendants de ces logiciels, qui produisent des interfaces métiers, des « vues utilisateurs » construisant des liens entre des tables hétérogènes dans la base de données. Ainsi, ces interfaces cachent l'organisation physique des données aux utilisateurs, or ces derniers ont besoin de cette information s'ils souhaitent extraire les données.

Pour cela, ils ont donc besoin de disposer du **schéma de la base**. Mais lorsque le système d'information a été conçu par un prestataire externe, le schéma lui appartient et il n'est pas obligé de le mettre à disposition des gestionnaires de bases de données. L'extraction peut alors se transformer en un véritable travail d'archéologie consistant à fouiller les entrailles des serveurs pour atteindre la racine des bases de données, au-delà des interfaces de visualisation. Ce passage de la « vue utilisateur » à la « vue physique » des données peut prendre des semaines voire des mois de travail¹¹⁵. Se pose ainsi la question de la relation entre les administrations et leurs prestataires, qui ne facilitent pas toujours l'accès aux données à proprement parler. Les sociologues Jérôme Denis et Samuel Goëta résument alors la situation de la façon suivante : « *Certains prestataires sont propriétaires des chemins d'accès et des systèmes de stockage de leurs bases de données : l'inaccessibilité des données est au cœur de leur modèle économique.* »¹¹⁶

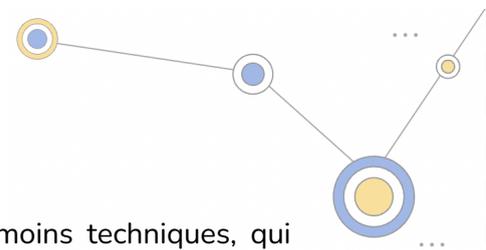
De la lourdeur de la publication

Une fois les données extraites, l'administration a le choix de les publier sur son propre portail *open data* ou simplement sur le portail *data.gouv.fr* mis à disposition par la DINUM.

¹¹⁴ Paul Edwards et al., « Science friction: Data, metadata, and collaboration », *Social Studies of Science*, 41(5), 2011, p.667-690.

¹¹⁵ Jérôme Denis et Samuel Goëta, « La fabrique des données brutes », in Ouvrir, partager, réutiliser édité par Clément Mabi et al., Éditions de la Maison des sciences de l'homme, 2017. URL : doi.org/10.4000/books.editionsmsh.9050.

¹¹⁶ *Ibid.*



Les deux options requièrent un certain nombre d'étapes plus ou moins techniques, qui représentent autant de nouvelles charges pour les agents de l'administration.

Les administrations, qu'il s'agisse de collectivités territoriales, de ministères ou d'opérateurs de l'État, sont nombreuses à souhaiter posséder leurs propres portails dans la mesure où la politique *open data* est d'abord conçue pour servir leur communication ou leur image. La construction de cette « vitrine » implique de :

1. **Choisir et configurer la plateforme.** Il peut s'agir de simples catalogues, listant les données déjà disponibles sur Internet, ou de plateformes de gestion de données, sur lesquelles les données peuvent être mises à jour et manipulées par les utilisateurs sans qu'ils aient à les télécharger¹¹⁷. Il existe des solutions propriétaires « clés en main » comme OpenDataSoft, utilisée en 2020 par 13 des 16 ministères¹¹⁸, mais aussi des options *open source*, comme CKAN¹¹⁹ ou DKAN¹²⁰, qui impliquent un travail beaucoup plus important pour les agents qui doivent eux-même configurer la plateforme (personnalisation de l'interface utilisateur, configuration des droits d'accès, mise en place de procédures de validation des données, etc.).
2. **Choisir une solution d'hébergement.** L'administration peut investir directement dans des serveurs ou recourir à des services d'hébergement pour stocker les données et héberger le portail open data.
3. **Sécuriser la plateforme.** L'administration doit investir dans des mesures de sécurité pour parer à d'éventuelles attaques (injection SQL, attaque sur l'API, attaque par déni de service et déni de service distribué DoS et DDoS, etc.). Il peut s'agir d'un chiffrement des données en transit ou au repos, de mécanismes d'authentification, de dispositifs de surveillance et d'alerte, etc.
4. **Choisir le format des données.** Les données partagées doivent être dans un format ouvert et standardisé, qui facilite l'accès et l'utilisation des données et les rend interopérables avec d'autres systèmes d'information. En fonction de la structure des données, les formats diffèrent. Pour une structure tabulaire par exemple, le CSV (*comma separated values*) est privilégié¹²¹. Il s'agit d'un format ouvert et lisible par des machines, qui se télécharge facilement et peut être enregistré sous des applications comme Excel. Par ailleurs, le format CSV est le format le plus simple permettant une grande diversité de réutilisations. Pour ces raisons, la Commission

¹¹⁷ data.europa.eu/elearning/fr/module9/#/id/co-01

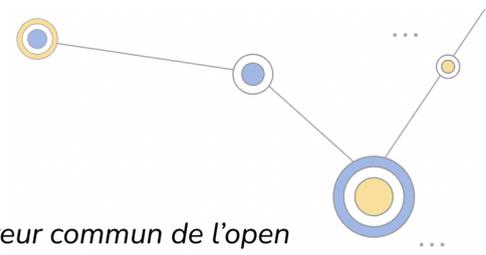
¹¹⁸ www.opendatasoft.com/fr/blog/ameliorer-laction-publique-grace-au-partage-de-donnees/

¹¹⁹ ckan.org/

¹²⁰ demo.getdkan.org/

¹²¹ data.europa.eu/elearning/fr/module9/#/id/co-01





européenne présente le CSV comme « le plus petit dénominateur commun de l'open data devant être utilisé en priorité autant que possible »¹²².

5. **Mettre à disposition les données.** Les administrations choisissent de proposer la possibilité de télécharger les bases de données dans leur intégralité ou de mettre à disposition des API qui permettent à l'utilisateur d'extraire de la base un nombre limité d'items.
6. **Maintenir le système opérationnel.** Une fois que le portail *open data* est mis en service, il faut continuer à investir dans sa maintenance pour garantir qu'il reste sécurisé, fiable et conforme aux normes en vigueur.
7. **Anticiper d'éventuels pics de fréquentation.** L'ouverture des données peut entraîner une augmentation de la demande d'accès aux données. Il faut alors faire évoluer le système de façon à ce qu'il puisse répondre à cette demande tout en maintenant des performances optimales.

Toutefois, il n'est pas nécessaire de créer son propre portail pour se conformer aux obligations de la loi pour une République numérique puisqu'une administration peut aussi choisir d'utiliser uniquement *data.gouv.fr*. Si l'investissement est moindre pour les équipes, il n'est pas nul. En sus de leurs missions habituelles, les agents doivent en effet :

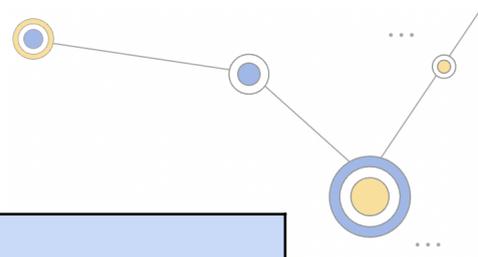
1. **Décrire le jeu de données** avec un titre précis qui correspond au vocabulaire employé par les utilisateurs et une description du contenu et de la structure des données, du contexte de production, etc.
2. **Ajouter les ressources**, c'est-à-dire des fichiers téléchargeables qui contiennent des informations à propos du jeu de données : données mises à jour, données historisées, documentation, code source, API, lien, etc. L'administration a alors la possibilité de publier des données issues d'un catalogue existant en utilisant le service de moissonnage de *data.gouv.fr*¹²³.
3. **Mettre à jour les jeux de données** puisque le principe d'*open data* par défaut s'accompagne d'une obligation pour les administrations de mettre à jour régulièrement les documents et données qu'elles publient¹²⁴.
4. **Animer la communauté des utilisateurs**, par exemple en annonçant la publication de nouveaux jeux de données sur les réseaux sociaux de l'administration et en répondant aux questions posées sur *data.gouv.fr*.

¹²² *Ibid.*

¹²³ doc.data.gouv.fr/moissonnage/intro/#principe

¹²⁴ Article L312-1-1 du CRPA.





Data.gouv.fr versus OpenDataSoft

Alors que de nombreuses administrations, collectivités territoriales comme ministères, choisissent de se tourner vers la solution privée OpenDataSoft pour créer leurs portails de données, certaines voix s'élèvent parmi les militants de l'*open data* pour dénoncer un recours paradoxal à une solution propriétaire pour mener une politique *open data* et dénoncer les dépenses inutiles d'argent public. En effet, ces administrations pourraient tout aussi bien utiliser une solution *open source*, ou le cas échéant, se reposer uniquement sur la solution de l'État, *data.gouv.fr*.

Faut-il craindre que les administrations ne se détournent de *data.gouv.fr* car elles lui préféreraient OpenDataSoft ? Il faut ici rappeler la différence entre ces deux solutions, qui nous semblent davantage complémentaires que réellement concurrentes.

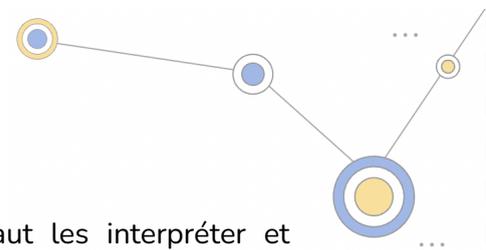
- **Un catalogue pour l'une, un service d'accompagnement pour l'autre.** *Data.gouv.fr* répertorie l'ensemble des données disponibles sur tous les portails des administrations. Il ne sera jamais en concurrence avec OpenDataSoft, qui propose une solution clé en main avec un important travail d'accompagnement. Ce n'est pas l'objectif de la DINUM, qui pourrait proposer des outils pour créer facilement des portails mais n'aura jamais la force de frappe suffisante pour accompagner leur déploiement.
- **Des clients différents.** OpenDataSoft sert l'administration productrice de données, qui achète une solution pour publier et présenter ses données. *Data.gouv.fr* sert l'usager du service public, le citoyen qui souhaite réutiliser les données.
- **Des exigences différentes.** OpenDataSoft est une base structurée. Si elle veut utiliser cette solution, l'administration doit préalablement structurer ses données conformément. Au contraire, *data.gouv.fr* accepte les données mal structurées, estimant que cela vaudra toujours mieux que pas de données du tout.

2.1.2. Sur le plan des ressources humaines : du temps et des compétences qu'il manque trop souvent aux administrations

Une mobilisation nécessaire des équipes

Au-delà de l'adaptation des systèmes d'information, l'ouverture des données publiques implique l'étroite collaboration de toutes les équipes en interne : direction des systèmes d'information, direction juridique, services de communication, documentalistes, mais aussi et surtout équipes métier, car ce sont elles qui savent quelles données sont produites par





les administrations, comment elles sont utilisées, comment il faut les interpréter et comment elles sont mises à jour. Pour ouvrir ses données de façon efficace et pérenne, il faut donc **impliquer l'ensemble des équipes et repenser la place de la donnée dans les processus métier**. Si un portail de Néogéo ou d'Opendatasoft peut être mis en service en moins d'une semaine, il faut plusieurs mois de travail en lien avec les équipes métier pour identifier les données à ouvrir et les intégrer aux portails. La création du portail est donc l'occasion de repenser les processus métier pour que la production et la mise à jour de données y soient plus automatisées, et pour que ces données soient mieux utilisées au quotidien par les agents de l'administration.

Au-delà de la mobilisation et de la collaboration des équipes, l'ouverture nécessite des **compétences techniques, qui manquent encore trop souvent au secteur public**. Pour y remédier, la grille salariale interministérielle des 55 métiers de la filière numérique a été fortement revalorisée en 2024¹²⁵. Il s'agit d'attirer des profils de techniciens, mais aussi des candidats qui, sans être experts, ont une appétence pour les sujets données et numériques. En effet, l'ouverture des données nécessite de comprendre des sujets techniques mais pas nécessairement de les maîtriser puisque des solutions comme celle d'Opendatasoft proposent une interface simple et intuitive, facilement manipulable par des agents sans compétence de développeur.

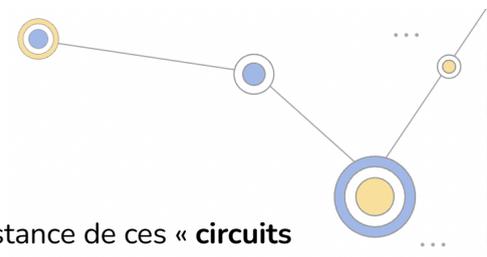
Un débat s'ouvre donc ici qui n'est pas propre au domaine du numérique : les administrations doivent-elles se doter de compétences en interne, ou au contraire profiter d'un écosystème privé qui propose des solutions très performantes ? Si elles profitent des meilleures solutions du marché offertes par des prestataires, il semble cependant indispensable que les administrations conservent les compétences suffisantes en interne pour être en mesure de négocier le cahier des charges et les prix.

L'épineuse question de la responsabilité

La mobilisation des équipes nécessite d'identifier clairement les personnes chargées de la production et de la mise en ligne des données. Mais la **dilution des responsabilités**, constatée aujourd'hui dans les administrations, explique en effet leur inertie sur le sujet. Lorsque les données sont la mission de tous, elles finissent par n'être la mission de personne. Sans mandat clair, les agents ont alors tendance à vouloir faire approuver la publication de leurs jeux de données par leurs responsables, qui peuvent eux-mêmes être

¹²⁵ Circulaire n° 6434/SG du 3 janvier 2024 relative à la politique salariale interministérielle des métiers de la filière numérique.





tentés de faire remonter la question à leur propre hiérarchie. La persistance de ces « **circuits de validation** » ralentit considérablement le rythme de libération des données publiques.¹²⁶

Dans une circulaire de 2021 adressée aux ministres et préfets de région¹²⁷, le Premier ministre Jean Castex prend acte de cette dilution des responsabilités et propose de structurer le paysage de la donnée publique en créant la fonction d'**administrateur ministériel des données, des algorithmes et des codes sources**. Les AMDAC sont chargés d'élaborer la feuille de route des différents ministères en matière de données et doivent constituer et animer un réseau de correspondants au sein des directions métier de son périmètre. Le directeur interministériel du numérique, en sa qualité d'administrateur général des données, des algorithmes et des codes sources, accompagne et coordonne cette communauté des AMDAC dans la mise en œuvre de leurs feuilles de route.

Par ailleurs, depuis 2016, la loi oblige les services ministériels et préfectoraux, les communes de plus de 10 000 habitants, les départements, les régions, les établissements publics nationaux et locaux, les personnes de droit privé chargées de la gestion d'un service public à désigner un **agent responsable de l'accès aux documents et des questions relatives à la réutilisation des informations publiques (PRADA)**¹²⁸. Les missions du PRADA sont de réceptionner les demandes d'accès aux documents administratifs et les éventuelles réclamations, et d'assurer la liaison avec la CADA¹²⁹. S'il n'est pas directement responsable de la publication des données publiques de son administration, il est donc sensibilisé à la question de la communicabilité des données. Il pourrait donc jouer un rôle moteur et structurant dans la politique d'ouverture des données de son administration.

Mais de façon générale, les administrations explicitent encore trop peu les missions liées à la production et à la publication des données publiques dans les fiches de poste des agents. Si l'on analyse les offres diffusées sur Choisir le Service Public¹³⁰, déposées directement sur le site par les employeurs publics ou transmises par des sites partenaires pour diffusion, 1% seulement des annonces mentionnent la question des données dans les descriptions des missions. Ces dernières sont concentrées dans la fonction publique d'État (73%), où elles sont en grande majorité des offres de catégorie A (75%) (Figure 10), alors que les collectivités territoriales et les établissements hospitaliers collectent l'essentiel des données de terrain, ensuite agrégées par les administrations centrales.

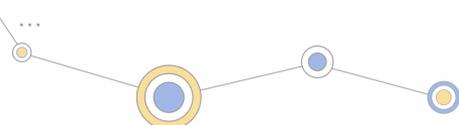
¹²⁶ Samuel Goëta, *Les Données de la Démocratie*, 2024, p.151.

¹²⁷ Circulaire n°6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources.

¹²⁸ Article L.330-1 du CRPA.

¹²⁹ Article R. 330-4 du CRPA.

¹³⁰ www.data.gouv.fr/fr/datasets/les-offres-diffusees-sur-choisir-le-service-public/



Par ailleurs, la part de ces offres semble rester stable dans le temps (environ 1% des offres en 2024 comme en 2022¹³¹) alors que dans le même temps les législations française et européenne ont renforcé la place des données dans les fonctions des administrations (Data Governance Act en 2022, Data Act en 2023), laissant plutôt présager une augmentation de la proportion des postes intégrant ces questions dans leurs missions.

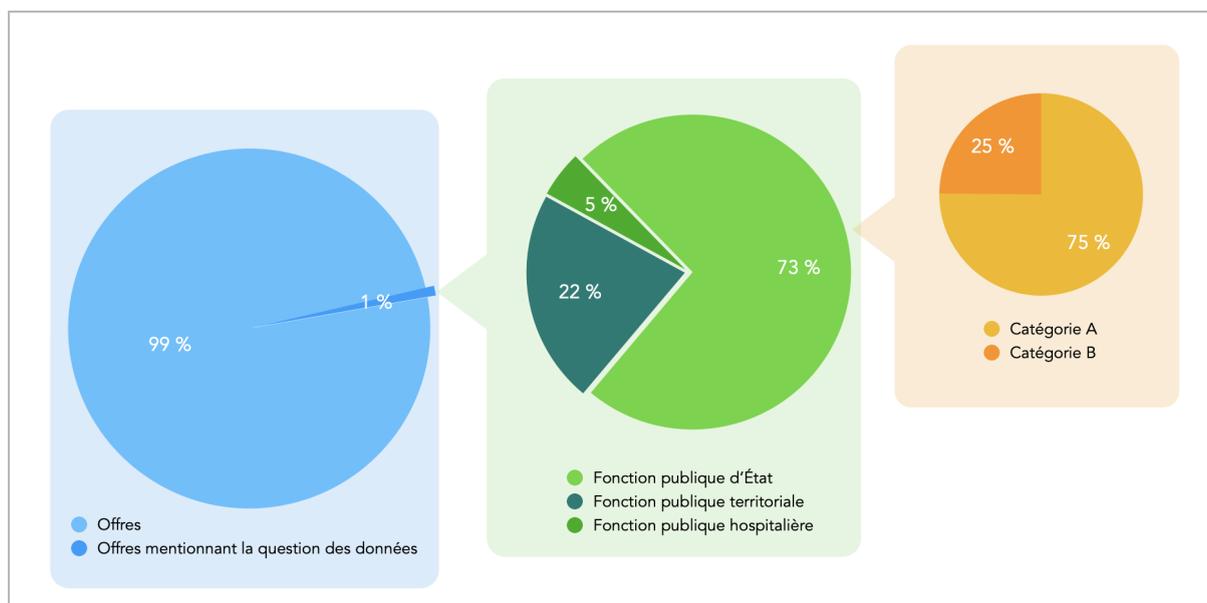
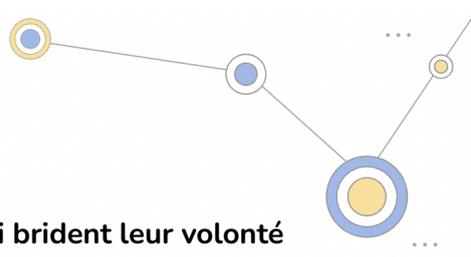


Figure 10 : Distribution des offres d'emplois publics qui mentionnent les termes « données », « data » ou « RGPD » entre les fonctions publiques d'État, territoriale et hospitalière et les catégories A et B en 2024. Source : data.gouv.fr.

Avec l'open data, un changement culturel majeur pour les administrations

L'exigence de transparence des administrations est finalement assez récente. Si l'article 15 de la Déclaration des Droits de l'Homme et du Citoyen de 1789 dispose que « *La Société a le droit de demander compte à tout Agent public de son administration* », il faut attendre la loi Cada de 1978 pour que ce droit d'accès aux documents administratifs devienne une réalité en France. D'autres pays l'ont précédée, la Suède reconnaît le droit d'accès depuis 1776 et les États-Unis ont adopté en 1966 le « *Freedom of Information Act* ». Quant à la publication spontanée de documents à l'initiative de l'administration, le véritable tournant n'intervient en France qu'avec la loi Lemaire de 2016. Il s'agit d'un véritable changement de paradigme pour les administrations, expliquant son adoption relativement lente.

¹³¹ Les offres mentionnant les données représentent (2348/304 178 = 0,77%) en décembre 2022 et (1241/141297 = 0,88%) en mai 2023.



À l'échelle individuelle, les agents font face à des contradictions qui brident leur volonté de partager des données. Si la loi Cada et la loi pour une République numérique exigent un plus grand partage des données, elles s'opposent, au moins en apparence, à d'autres textes juridiques. D'une part, les agents publics sont soumis à une obligation de secret professionnel et peuvent être pénalement poursuivis s'ils ne la respectent pas. La révélation d'un secret peut ainsi être punie d'un an d'emprisonnement et de 15 000 euros d'amende¹³². D'autre part, un agent public doit « *faire preuve de discrétion professionnelle pour tous les faits, informations ou documents dont il a connaissance dans l'exercice ou à l'occasion de l'exercice de ses fonctions* »¹³³. La démarche de transparence prévue par la loi Cada et la loi pour une République numérique détonne dans un corpus juridique où le secret demeure la règle. La frilosité juridique des agents se manifeste également quand ces derniers décident de communiquer le document au demandeur ou de le publier en ligne. Il est plus sûr juridiquement pour l'agent de communiquer le document uniquement à un tiers, qui devient alors responsable du traitement des données au sens donné par le RGPD, et donc responsable des conséquences de la diffusion de ces données.

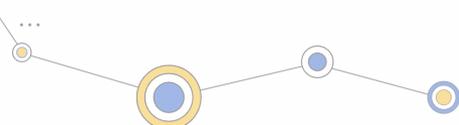
L'ouverture des données publiques bute par ailleurs sur un phénomène de « **database hugging** » (« câlins de bases de données »). Ce concept, introduit par le médecin et statisticien suédois Hans Rosling¹³⁴, décrit la rétention de la donnée par celui qui l'a produite. En effet, la donnée peut être perçue comme une source de pouvoir pour celui qui la « possède ». C'est le cas précédemment cité des données d'exploitation des entreprises qui gèrent les réseaux d'eau, d'énergie ou de gestion des déchets pour le compte des collectivités. La connaissance de ces données est un atout important pour les collectivités dans la négociation des contrats de délégation de service public, ce qui explique que les entreprises délégataires rechignent à partager ces données.

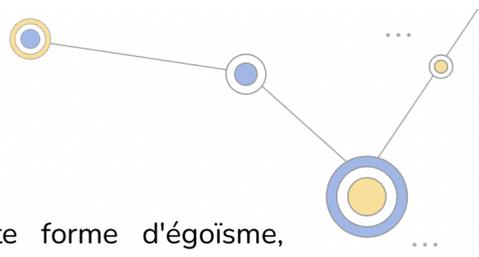
Mais souvent, la donnée comme source de pouvoir relève plus du fantasme que de la réalité, au détriment de l'intérêt général. C'est pourquoi le physicien et informaticien britannique Tim Berners-Lee a appelé en 2009 à lutter contre ce phénomène de *database hugging*. Les services administratifs français n'y échappent pas. L'exemple souvent cité est celui de deux bureaux d'une même direction d'un même ministère s'échangeant des tableaux de chiffres au format PDF pour éviter que l'autre ne s'approprie trop facilement les données, ce qui oblige à reporter les données à la main dans un tableur. La persistance de ce genre de comportement, décrite dans les travaux des chercheurs spécialistes de la question, comme Samuel Goëta, et a été confirmée par les acteurs des administrations centrales et

¹³² Article 226-13 du Code pénal.

¹³³ Article L121-7 du Code général de la fonction publique.

¹³⁴ blogs.worldbank.org/en/opendata/hugs-and-databases-memory-hans-rosling





territoriales que nous avons rencontrés. Tous déplorait cette forme d'égoïsme, contre-productive pour l'efficacité et la qualité de l'action publique.

2.1.3. Sur le plan financier : des coûts en plus mais des recettes en moins

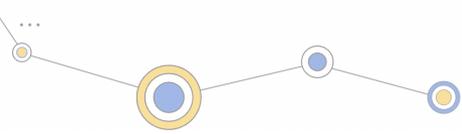
Combien cela coûte d'ouvrir ses données ?

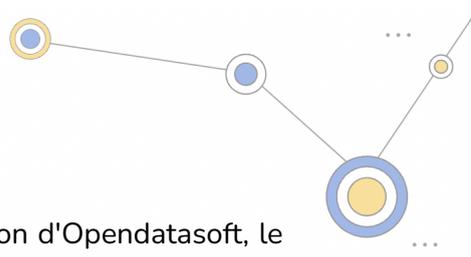
L'adaptation des systèmes d'information, comme la formation et la mobilisation des équipes, sont autant de tâches supplémentaires et de coûts pour les administrations qui s'engagent dans une démarche d'*open data*. Il y a d'abord les **coûts initiaux d'ouverture et de déploiement de solutions**, auxquels viennent ensuite s'ajouter les **coûts de fonctionnement** liés à l'hébergement, à la maintenance du portail et à la diffusion des données en « régime permanent », qu'il faut payer chaque année. Tous ces coûts varient suivant les solutions adoptées, le type de données diffusées ou encore leur volume, mais nous pouvons essayer de déduire à partir d'exemples¹³⁵ quelques ordres de grandeur des coûts que représenterait une politique d'*open data* pour une administration.

L'exemple de La Rochelle donne une idée de ce que peut représenter les **coûts d'ouverture et de déploiement** pour une collectivité. En 2012, la ville de 75 000 habitants a dépensé plus de 20 000 euros pour développer son propre portail à partir de la solution *open source* CKAN et y intégrer quelque 250 jeux de données. La collectivité aurait pu se tourner vers d'autres solutions, comme les portails *open source* de Néogéo qui offrent également des outils performants de visualisation des données géographiques. De façon générale, il est raisonnable d'estimer que pour une collectivité d'environ 50 000 habitants, l'ouverture d'un portail nécessitera environ **six mois de travail pour la préparation des données et entre 50 000 et 100 000 euros** pour la personnalisation du portail et l'accompagnement des équipes. Pour des portails plus importants, comme ceux de la métropole du Grand Lyon ou de la région Nouvelle-Aquitaine, qui comptent respectivement 640 et 8500 jeux de données, il faut compter un an de travail de préparation des données et 200 000 à 400 000 euros. Quelle que soit la solution choisie, **la facture double ou triple si l'on prend en compte la mobilisation en interne des équipes** pour identifier les données, les préparer et les intégrer dans la solution de diffusion.

Pour les coûts de fonctionnement, à savoir la maintenance de son logiciel, la Ville de La Rochelle a dépensé 10 000 euros par an entre 2012 et 2017, et un peu plus à partir de 2017 pour développer de nouvelles fonctionnalités de visualisation des données sur son portail. Pour une solution de Néogéo, il faut compter de **10 000 à 100 000 euros par an**

¹³⁵ Les chiffres sur les coûts des portails sont issus d'entretiens avec des représentants de la Ville de La Rochelle, de la DINUM et des entreprises Néogéo et Opendatasoft, sauf quand une autre source est précisée.





suivant la taille du portail. Et si l'administration a opté pour la solution d'Opendatasoft, le prix payé sera proportionnel au volume de données diffusées, pouvant représenter **quelques milliers d'euros** pour une collectivité de 5 000 habitants et **quelques dizaines de milliers d'euros** pour une métropole. Le prix est beaucoup plus important pour les gros diffuseurs de données : le ministère de l'Économie dépense ainsi 400 000 à 500 000 euros par an. On peut légitimement penser que la Direction de l'information légale et administrative (DILA), qui fait également partie des plus gros clients publics d'Opendatasoft, consacre un budget équivalent¹³⁶.

Dans son ensemble, une politique d'ouverture des données représenterait donc de quelques milliers à quelques centaines de milliers d'euros par an pour une administration. Est-ce beaucoup par rapport au budget global d'une administration ? Le portail *open data* du ministère de l'Économie représente seulement **0,0001% d'un budget total** de 435 milliards d'euros¹³⁷. Le portail du ministère de la Culture coûte quant à lui 24 000 euros par an¹³⁸, soit **0,001 % d'un budget total** de 4 milliards d'euros¹³⁹. Pour les collectivités, on estime leurs dépenses de fonctionnement en moyenne à 4 millions d'euros par an pour une collectivité de 5 000 habitants et à 400 millions d'euros par an pour une métropole. Ainsi, un portail *open data* ne devrait représenter que **0,01 % ou 0,1 % de leurs dépenses**. Si cela semble peu, allouer ces sommes à l'ouverture des données est un choix, qui se fait nécessairement au détriment d'autres politiques publiques. À titre d'exemple, le développement et le déploiement du portail du Grand Lyon a coûté autant que la rénovation thermique d'une école¹⁴⁰. Faire de l'*open data* implique donc pour les administrations de faire des **arbitrages budgétaires qui ne sont pas toujours simples**.

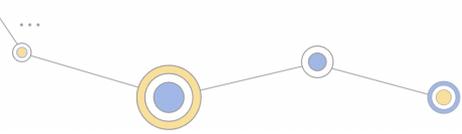
¹³⁶ Raphaëlle Aubert et Léa Sanchez, « Comment Opendatasoft est devenue l'acteur incontournable de l'ouverture des données publiques », Le Monde, 2024. URL : www.lemonde.fr/les-decodeurs/article/2024/03/05/comment-opendatasoft-est-devenue-l-acteur-incontournable-de-l-ouverture-des-donnees-publiques_6220195_4355770.html

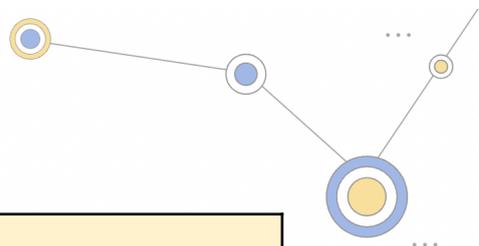
¹³⁷ www.budget.gouv.fr/budget-État/ministere?annee=135&loi_finances=50&type_budget=all&type_donnee_budg_et=ae&op=Valider

¹³⁸ Raphaëlle Aubert et Léa Sanchez, « Comment Opendatasoft est devenue l'acteur incontournable de l'ouverture des données publiques », Le Monde, 2024. URL : www.lemonde.fr/les-decodeurs/article/2024/03/05/comment-opendatasoft-est-devenue-l-acteur-incontournable-de-l-ouverture-des-donnees-publiques_6220195_4355770.html

¹³⁹ www.budget.gouv.fr/budget-État/ministere?annee=135&loi_finances=50&type_budget=all&type_donnee_budg_et=ae&op=Valider

¹⁴⁰ Référence : 500€ / m² selon le rapport d'information du Sénat « Transition écologique du bâti scolaire : mieux accompagner les élus locaux » (2023).





Quel budget pour data.gouv.fr ?

Quand les administrations n'ouvrent pas leurs propres portails et publient uniquement sur *data.gouv.fr*, les coûts sont essentiellement assumés par la DINUM, responsable de la plateforme.

Le déploiement de la politique d'*open data*, considéré comme un acquis, ne fait pas partie des priorités listées dans la feuille de route de la DINUM pour l'année 2024¹⁴¹. Mais cela ne veut pas dire que le sujet est délaissé : **le budget du pôle en charge de *data.gouv.fr* a doublé entre 2023 et 2024**, passant de un à deux millions d'euros, pour permettre le développement de nouvelles fonctionnalités et participer aux déploiements des portails thématiques des ministères.

En parallèle de ces nouveaux projets, ce budget assure le fonctionnement de *data.gouv.fr* : financement de l'hébergement des données sur des serveurs d'OVH et d'une équipe d'environ dix ETP comportant un administrateur système, trois développeurs, un chargé de déploiement pour contacter les producteurs de données, un *product owner* qui établit la feuille de retour des évolutions du produit, un *UX designer*, un expert des données et un expert des données géographiques, notamment pour mettre à jour les données du cadastre.

Des recettes en moins pour les gros producteurs de données

Non seulement la loi pour une République numérique n'a pas prévu de moyens supplémentaires pour les administrations qui doivent ouvrir leurs données, mais certaines ont aussi vu une source de revenus, les redevances, disparaître. En effet, en 2013, le Premier Ministre Jean-Marc Ayrault commande au magistrat de la Cour des comptes Adnène Trojette un rapport pour évaluer la pertinence des redevances demandées par certaines administrations aux réutilisateurs de leurs données. Le rapport Trojette¹⁴² note le fait que personne n'ait établi avec certitude les revenus totaux tirés des redevances, estimés **entre 35 millions et 100 millions d'euros**. Il propose alors de nouveaux calculs, estimant le montant des redevances à 34,7 millions d'euros, en rappelant que les administrations s'achetaient entre elles des données pour un montant de 5 millions d'euros. Il ne s'agit ici que des redevances directement liées à la réutilisation de données, mais même en considérant les estimations les plus hautes qui prennent aussi en compte les

¹⁴¹ DINUM, « Une stratégie numérique au service de l'efficacité de l'action publique - Feuille de route de la DINUM », 2023. URL : www.numerique.gouv.fr/uploads/Feuille-de-route-DINUM.pdf

¹⁴² Adnène Trojette, « Ouverture des données publiques - Les exceptions au principe de gratuité sont-elles légitimes ? », 2013. URL : www.numerique.gouv.fr/uploads/20131105-rapporttrojetteannexes.pdf

ventes de produits commerciaux éventuellement conçus à partir de données publiques, le **montant des redevances représente rarement plus de 1% du budget de l'administration**, et jamais plus de 6 %. La perte de ressources liée à la suppression des redevances est donc à relativiser. Mais le rapport Trojette souligne aussi que ces revenus peuvent offrir une « marge de manœuvre appréciable » si on les rapporte aux crédits de fonctionnement ou à la capacité d'autofinancement des administrations.

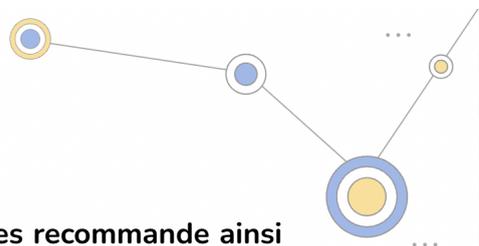
Service bénéficiaire	Recettes 2010	Recettes 2011	Recettes 2012
Service hydrographique et océanographique de la marine	1 320 000 €	2 102 000 €	1 300 000 €
Agence de services et de paiement	69 750 €	70 800 €	53 480 €
Office national d'information sur les enseignements et les professions	210 510 €	199 367 €	155 143 €
Institut national de la propriété intellectuelle	2 955 730 €	3 206 627 €	2 744 054 €
Institut national de l'information géographique et forestière	32 472 901 €	11 565 908 €	9 940 748 €
Cour de cassation	339 547 €	289 713 €	264 120 €
Institut national de la statistique et des études économiques	8 823 000 €	10 588 000 €	9 981 000 €
Météo-France (hors recettes refacturées à sa branche commerciale)	2 026 919 €	1 591 563 €	1 585 000 €
Service de l'observation et des statistiques	594 000 €	582 000 €	580 000 €
Commission d'accès aux documents administratifs	-	5 000 €	5 000 €
FranceAgriMer	300 000 €	300 000 €	300 000 €
Direction de l'information légale et administrative	817 971 €	812 541 €	892 326 €
Institut national de l'origine et de la qualité	31 610 €	72 195 €	79 265 €
Ministères économiques et financiers	1 288 472 €	1 604 050 €	1 955 234 €
Ministère de l'intérieur	18 294 €	3 087 267 €	3 865 282 €
Institut français du cheval et de l'équitation	72 384 €	64 905 €	81 671 €
Agence technique de l'information sur l'hospitalisation	432 983 €	420 623 €	543 719 €
Ministère de l'éducation nationale	99 006 €	84 913 €	131 091 €
Conseil d'État	226 779 €	149 052 €	231 508 €
Ministère de l'agriculture, de l'alimentation et de la forêt	7 010 €	7 100 €	16 700 €
TOTAL	52 106 866 €	36 803 623 €	34 705 341 €

Tableau 5 : Recettes des redevances sur la réutilisation des données publiques et part des acheteurs publics. Source : Rapport Trojette, 2013.

La fin des redevances implique donc de trouver de nouvelles ressources pour conserver cette marge de manœuvre. Il faut rappeler que la loi prévoit des exceptions pour certains jeux de données¹⁴³ produits par des administrations pour lesquelles le financement de la production de données représente plus de 25% de leurs ressources propres¹⁴⁴. Malgré cela, la Cour des comptes alerte en 2019 sur la situation de l'IGN, de Météo-France et du Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (Cerema) qui doivent faire face à des injonctions contradictoires : supprimer les redevances sur la vente de données tout en développant leurs ressources propres par la

¹⁴³ Liste fixée par décret, voir l'Article D324-5-1 du CRPA.

¹⁴⁴ Article R324-4-1 du CRPA.



commercialisation de produits et des services¹⁴⁵. **La Cour des comptes recommande ainsi de revoir le modèle économique de ces administrations.** C'est en particulier ce qu'a dû faire l'IGN, dont les revenus issus des redevances diminuaient depuis plusieurs années. **L'IGN a donc engagé une transformation profonde de son modèle.** D'une part, il s'est engagé dans des projets collaboratifs, notamment avec la communauté d'OpenStreetMap, avec qui il a développé Panoramax, un outil de vues immersives qui se veut l'alternative publique de Google Street View. D'autre part, l'IGN finance une partie de sa mission historique de producteur de données grâce à des **partenariats dans de grands projets financés par leurs commanditaires.** En 2022, l'IGN a ainsi reçu 16 millions de France Relance et du fonds pour la transformation de l'action publique pour le projet Lidar HD de cartographie 3D de l'intégralité du sol français¹⁴⁶. La même année, il a reçu 12 millions d'euros de la part du ministère de l'Agriculture pour la mise à jour du référentiel parcellaire graphique qui sert de base pour le calcul des aides de la PAC¹⁴⁷. Mais ces revenus sont conditionnés à l'existence de grands fonds publics ou de clients prêts à financer des projets. **Il y a donc lieu de se questionner sur la viabilité du modèle économique de l'IGN :** aujourd'hui, l'institut est en mesure de financer sa production de données avec ses ressources propres, mais sera-ce toujours le cas demain ?

La question se pose dans les mêmes termes pour Météo-France ou le Shom. Même si certaines données produites par ces administrations pourraient théoriquement déroger à l'obligation de gratuité des réutilisations¹⁴⁸, cela irait à l'encontre de la dynamique actuelle : la plupart font partie d'une liste de jeux de données de haute valeur (*High Value Datasets*) identifiés par l'Europe et qui doivent être mis à disposition gratuitement dès juin 2024¹⁴⁹. Pourtant, ces opérateurs produisent des données hautement stratégiques. C'est un **enjeu de souveraineté** pour la France que de continuer à les produire. Les subventions pour charge de service public (SCSP) de l'IGN et de Météo-France ont pourtant fortement diminué entre 2015, date de la loi Valter qui met fin aux redevances, et 2022¹⁵⁰. Si bien qu'en 2021, le montant des ressources propres de l'IGN a dépassé celui de sa SCSP¹⁵¹. Si l'on considère, dans l'esprit de la loi Lemaire, que la production de ces données relève d'un

¹⁴⁵ Cour des Comptes, Référé S2018-3287 sur la valorisation des données de l'IGN, de Météo-France et du Cerema.

¹⁴⁶ geoservices.ign.fr/lidarhd

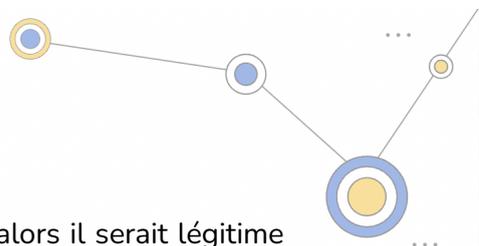
¹⁴⁷ Projet de loi de finances pour 2023 « Écologie, développement et mobilité durables », rapport général n° 115 (2022-2023), tome III, annexe 11, volume 3, déposé le 17 novembre 2022.

¹⁴⁸ Article D324-5-1 du CRPA

¹⁴⁹ Règlement d'exécution 2023/138 établissant une liste d'ensembles de données de forte valeur spécifiques et les modalités de leur publication et de leur réutilisation.

¹⁵⁰ Projet de loi de finances pour 2024 « Écologie, développement et mobilité durables », rapport général n° 128 (2023-2024), tome III, annexe 11, volume 3, déposé le 23 novembre 2023.

¹⁵¹ Projet de loi de finances pour 2023 « Écologie, développement et mobilité durables », rapport général n° 115 (2022-2023), tome III, annexe 11, volume 3, déposé le 17 novembre 2022.



véritable service public, et pas d'un simple service supplémentaire, alors il serait légitime de le financer entièrement par la SCSP.

Repenser la répartition de la valeur créée par les données publiques ?

S'il s'avère que les administrations engagées dans une démarche d'*open data* poursuivent ces efforts dans le temps car elles en perçoivent l'intérêt, certains agents publics peuvent regretter que les administrations produisent et mettent à disposition gratuitement des données ensuite réutilisées par des entreprises à des fins commerciales. Et parfois, ces entreprises ne paient pas d'impôts en France pour financer, au moins indirectement, les services qui produisent les données.

Cette critique n'est pas forcément légitime : les administrations n'ont pas vocation à gagner de l'argent avec leurs activités. Mais nous pouvons nous demander si une meilleure répartition de la valeur est possible entre les administrations qui supportent les coûts de diffusion des données et ceux qui en tirent les bénéfices économiques, et si le financement de l'*open data* public pourrait reposer davantage sur ceux qui en profitent très directement. Cela pose la question des licences qui pourraient être mises en place pour encadrer la réutilisation des données.

Tout d'abord, il ne faut pas confondre **licence et redevance**. Si la réutilisation d'une donnée publique est soumise à une redevance, alors elle doit obligatoirement s'accompagner d'une licence¹⁵², mais la réciproque est fautive. **Une licence permet de contrôler l'usage qui est fait des données**, sans pour autant aller jusqu'à faire payer le réutilisateur. Par exemple, l'association *Creative Commons* propose des licences moins restrictives que les droits d'auteurs classiques qui peuvent combiner différentes obligations associées à des sigles¹⁵³ :

- BY (attribution) : obligation de citer l'auteur initial.
- NC (non commercial) : interdiction de tirer un profit commercial de l'œuvre sans autorisation de l'auteur.
- ND (non derivative work) : interdiction d'intégrer tout ou partie des données dans une œuvre composite.
- SA (share alike) : obligation de rediffuser selon la même licence ou une licence similaire la réutilisation, on dit que cette licence a un caractère « viral ».

¹⁵² Article L323-1 du CRPA.

¹⁵³ creativecommons.org/share-your-work/licenses/

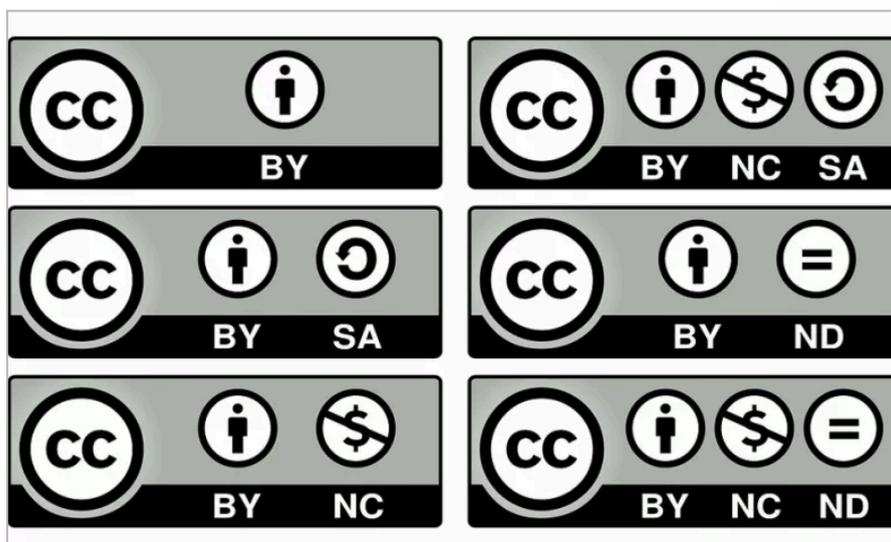


Figure 11 : Exemples de licences Creative Commons, repérables par des sigles.

Par exemple, la licence CC-NY-NC permet de réutiliser les données à condition que la finalité ne soit pas commerciale et que le nom du créateur des données soit bien cité. La loi pour une République numérique désigne **deux licences qui peuvent être utilisées par les administrations**¹⁵⁴ : la *Licence Ouverte / Open License (LO 2.0)* d'Etalab¹⁵⁵ et la licence *Open Database License (ODbL)*¹⁵⁶. La première se rapproche d'une licence CC-BY et la seconde d'une licence CC-BY-SA. Une administration a la possibilité de choisir une autre licence, mais elle doit alors justifier pourquoi les deux licences de référence ne conviendraient pas, et la licence proposée doit être homologuée par la DINUM. Les licences permettent donc aux administrations de valoriser leur travail (BY) ou d'enrichir leurs bases de données en demandant aux réutilisateurs de partager leur travail (SA). Mais en général, elles ne peuvent pas restreindre les utilisations, notamment commerciales, sauf « pour des motifs d'intérêt général et de façon proportionnée ». Autrement dit, **la loi ne permet pas aujourd'hui de faire payer les entreprises qui s'enrichissent grâce aux données publiques**.

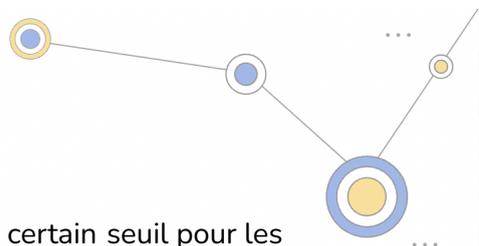
Serait-ce une bonne idée de réintroduire un caractère payant¹⁵⁷ pour couvrir les coûts de production et de diffusion des données, au moins pour les gros consommateurs de données qui en tirent profit ? On pourrait faire payer a priori les gens qui consomment beaucoup, ou a posteriori ceux qui ont une utilisation commerciale. Dans le premier cas, on pourrait penser à la mise en place de modèles de **tarification dits « freemium »**, tels qu'identifiés dans le rapport Trojette. Ces modèles favorisent l'innovation avec des données gratuites

¹⁵⁴ Article D323-2-1 du CRPA.

¹⁵⁵ etalab.gouv.fr/licence-ouverte-open-licence/

¹⁵⁶ opendatacommons.org/licenses/odbl/

¹⁵⁷ Pour se faire, on pourrait modifier par décret l'article D324-5-1 du CRPA.

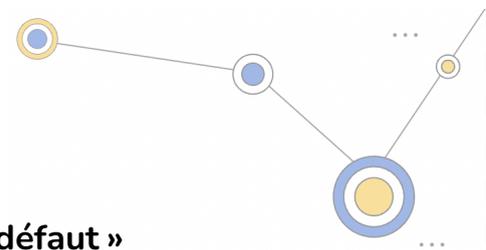


pour les petits consommateurs de données, et payantes à partir d'un certain seuil pour les grosses consommations. Le problème d'une telle tarification est la nécessité de contrôler tous ceux qui accèdent aux données pour mesurer leur volume de consommation, ce qui va à l'encontre des principes de l'*open data*. Elle priverait également de l'accès aux données des usagers qui auraient besoin de gros volumes mais pas nécessairement dans un but lucratif, par exemple des associations ou des chercheurs. En ce qui concerne le contrôle a posteriori, cela nécessite de posséder une capacité de veille des réutilisations et de contrôle de leur caractère commercial, ce qui semble là-aussi demander des moyens importants.

On pourrait également s'inspirer des **contrats de type *Material Transfer Agreement***¹⁵⁸ pratiqués notamment dans l'industrie pharmaceutique. Ce sont des accords dans lesquels la personne qui possède des droits de propriété intellectuelle sur des données, par exemple la formule d'une molécule, accepte de les mettre à disposition de quelqu'un d'autre pour un usage et un temps déterminés, par exemple un laboratoire qui produit des médicaments. Tout ceci en échange de rémunérations à différents jalons du projet et de *royalties* indexés sur les profits générés, par exemple par la vente du médicament. On pourrait imaginer de tels contrats pour la vente de données publiques, mais ici aussi cela pose de nombreuses questions. Comment dimensionner les modalités de l'accord (durée, jalons, *royalties*) ? Faudra-t-il des équipes dédiées pour établir ces contrats avec chaque réutilisateur ? Comment déterminer la part que représentent les données publiques dans la valeur générée par la réutilisation, et donc comment déterminer la valeur à partager à l'administration ? Pour une molécule de principe actif, c'est assez simple : le médicament n'existerait pas sans elle, mais pour une donnée en général, c'est beaucoup plus difficile. Ce juste partage de la valeur est-il possible dans le cas général sans passer par de lourdes tractations et négociations ?

Enfin, réintroduire un caractère payant nous semble contre-productif. Quelles que soient les modalités de mise en œuvre, cela entraînerait des entraves au libre-accès aux données publiques pour ceux qui n'en tirent aucun profit économique ou une charge supplémentaire pour l'administration devant contrôler des usages ou dimensionner des contrats. Par ailleurs, tout ceci irait à l'encontre de la dynamique actuelle de l'*open data* et en brouillerait le message.

¹⁵⁸ Quelques modèles de ces contrats sont proposés par l'université de Havard : otd.harvard.edu/industry-investors/sample-agreements/licensing/



2.2. Une clarification indispensable de la notion de « par défaut »

2.2.1. Une notion floue et inopérante

« Il faut penser données comme auparavant on pensait pétrole. »

Axelle Lemaire, Le Monde, 2015.

L'expression « par défaut » implique 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. Cette analogie pétrolière renvoie ainsi à un idéal d'objectivité et d'immédiateté, qui cache en réalité l'ensemble des « dispositifs par lesquels la transparence se réalise¹⁵⁹ ».

Les dispositifs de production des données : identifier les données qu'il est possible et qu'il est souhaitable d'ouvrir.

Il faut d'abord identifier les données qu'il est possible d'ouvrir. Pour cela, certaines administrations entament le recensement de toutes les données produites par leurs services. Dans le Cadre interministériel d'administration de la donnée, le directeur interministériel du numérique rappelle que ce travail est une étape préalable nécessaire à l'ouverture des données : « Les administrations doivent avoir connaissance du patrimoine de données produites ou collectées au sein de leur structure. Leur inventaire facilite l'identification des données qui ont vocation à être ouvertes, partagées ou exploitées dans le cadre de projets numériques »¹⁶⁰. Pourtant, cet exercice fastidieux est généralement vain, car les données sont souvent éparpillées entre plusieurs tables, voire plusieurs serveurs¹⁶¹.

L'idée qu'il soit possible de lister l'ensemble des données existantes dans les systèmes d'information d'une organisation s'inscrit tout à fait dans la métaphore pétrolière : la donnée est une ressource naturelle, encore trop mal connue, il faudrait donc la cartographier pour permettre son exploitation. En réalité, les agents qui explorent les systèmes d'information de leurs organisations à la recherche de jeux de données ne se retrouvent jamais confrontés à des données toutes prêtes qu'il suffirait de recenser. En amont, ils doivent identifier et stabiliser les circuits de diffusion des données, ce qui

¹⁵⁹ Jérôme Denis et Samuel Goëta, « La fabrique des données brutes », in Ouvrir, partager, réutiliser édité par Clément Mabi et al., Éditions de la Maison des sciences de l'homme, 2017. URL : doi.org/10.4000/books.editionsmsmh.9050.

¹⁶⁰ Cadre interministériel d'administration de la donnée, DINUM, 2021. URL : www.numerique.gouv.fr/uploads/Cadre-interministeriel-donnee.pdf

¹⁶¹ Samuel Goëta, *Les Données de la Démocratie*, 2024, p.184.



implique une véritable transformation de l'organisation administrative et ne peut donc pas s'envisager comme un processus de court terme, mené dans le cadre d'un inventaire.

Parmi les données qu'il est possible d'ouvrir, il faut ensuite déterminer les données qu'il est souhaitable d'ouvrir. Ces données dépendent du contexte politique : une collectivité territoriale voudra ainsi ouvrir celles qui ont déjà été ouvertes par les collectivités voisines pour ne pas sembler à la traîne et, si elle en a la possibilité, en publier qui n'ont encore été ouvertes par personne pour se démarquer. Les administrations ouvrent ainsi en priorité les données les moins difficiles à ouvrir et celles qui sont jugées les moins sensibles, sans préjuger des réutilisations qui pourront en être faites. La Ville de Paris par exemple publie une grande quantité de listes (listes des associations, listes des points de collecte de déchets, etc.) qui ne sont que très rarement téléchargées (Figure 12). Ce type de jeux de données permet ainsi à la Ville de se distinguer des autres métropoles françaises, sans risquer d'exposer des données politiquement sensibles.

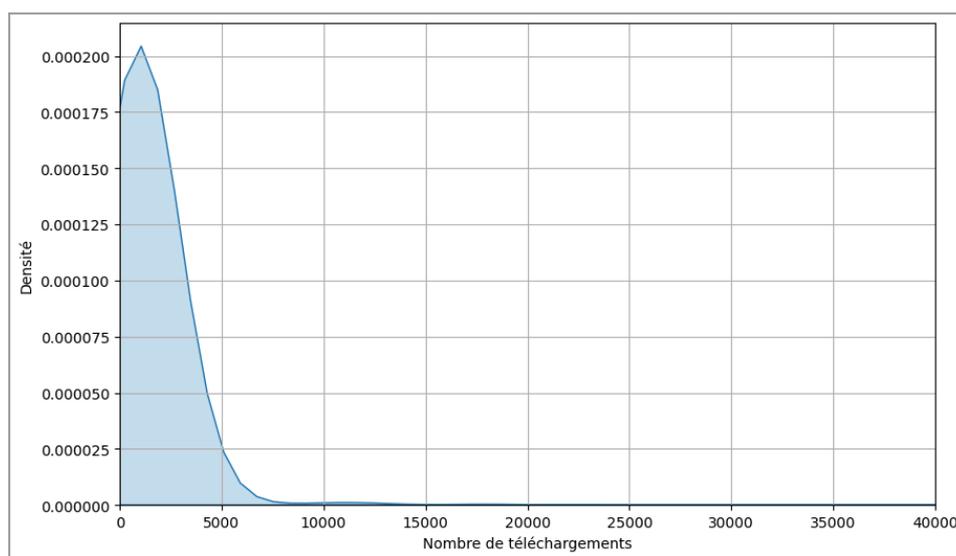
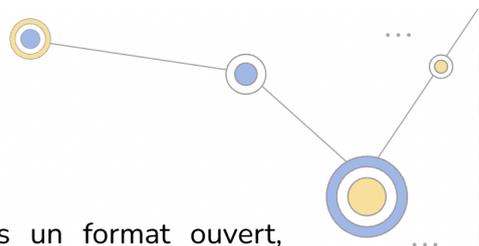


Figure 12 : Distribution du nombre total de téléchargements des jeux de données de la ville de Paris publiés entre 2010 et 2024. Source : data.gouv.fr.

Les dispositifs de transformation des données : assurer leur qualité de données « brutes »

Dans ses travaux, comme pour mieux mettre en évidence le caractère paradoxal des discours présentant les données comme un pétrole « brut », le sociologue Jérôme Denis parle d'un travail de « **brutification** » des données. Si la qualité des données telles qu'elles circulent entre les services des administrations était suffisante pour des usages internes, elle ne serait pas adaptée à d'éventuelles réutilisations externes. Il serait donc nécessaire de retravailler les données extraites des systèmes d'information avant de les ouvrir :

- 
- **Il faut reformater les données** afin qu'elles soient dans un format ouvert, facilement réutilisables par les citoyens. Or le format préconisé pour l'ouverture, le CSV, est rarement le format initial des données, généralement produites sur des applications propriétaires.
 - Une fois les données reformatées, il faut encore **les nettoyer pour leur donner un vernis d'objectivité et de vérité**. En effet, les données « métiers » telles qu'elles existent au sein des organisations ne sont pas des vérités objectives par essence, dans la mesure où elles sont toujours le fruit d'une construction et donc de décisions. Si le fait qu'elles soient peu précises n'a généralement aucun impact sur leur efficacité en interne, leur ouverture requiert un effort de correction et d'harmonisation pour tendre vers des « données génériques ».
 - Enfin, **les données doivent être désindexicalisées** : il s'agit alors d'effacer toutes les traces des usages précédents, de gommer complètement leur ancrage professionnel pour être compréhensibles par des citoyens non experts.

La présentation des données comme le « pétrole du 21^e siècle » est donc trompeuse lorsqu'elle est utilisée pour souligner une supposée immédiateté et facilité dans l'exploitation du potentiel des données et ainsi mettre en avant « l'incroyable gâchis » que serait la non réutilisation de ces données. En revanche, la comparaison peut être fructueuse si elle est utilisée au contraire pour rappeler que les données, comme les ressources du sous-sol, n'ont pas de valeur intrinsèque. Elles n'ont de valeur que dans la mesure où des acteurs, publics ou privés, y perçoivent un potentiel de création après un important travail d'extraction, de raffinage et de transformation.

2.2.2. Un concept source d'incertitudes juridiques

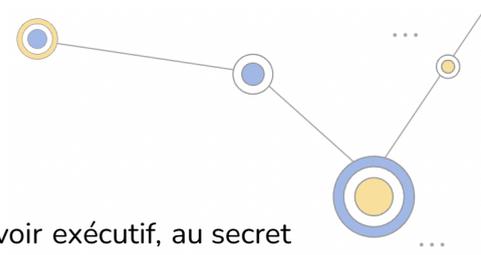
Le principe de « tout publier sauf... » : un imbroglio d'exceptions

Le périmètre des documents à ouvrir peut sembler clair. Il s'agit de tout publier sauf les **deux catégories de documents couverts par des secrets** :

- **Les documents qui ne sont communicables qu'à l'intéressé¹⁶²**, c'est-à-dire la personne morale ou physique qui est en cause dans les documents, et dont la communication porterait atteinte à la protection de la vie privée, au secret médical et au secret des affaires.
- **Les documents qui ne sont pas communicables¹⁶³** car leur consultation ou communication porterait atteinte notamment au secret des délibérations du

¹⁶² Article L311-6 du CRPA.

¹⁶³ Article L311-5 du CRPA.



gouvernement et des autorités responsables relevant du pouvoir exécutif, au secret de la défense nationale, à la conduite de la politique extérieure de la France, à la sûreté de l'État, à la sécurité publique, à la sécurité des personnes ou à la sécurité des systèmes d'information des administrations, à la monnaie et au crédit public, et à la recherche et la prévention d'infractions. En dehors de ces secrets, la loi établit une liste de documents spécifiques qui ne sont pas obligatoirement communicables, par exemple certains documents établis et détenus par l'Autorité de la concurrence dans le cadre de ses missions d'enquêtes, les informations couvertes par des droits de propriété littéraire et artistique¹⁶⁴, ou encore les avis du Conseil d'État¹⁶⁵.

On pourrait se dire que la règle est simple et rend le principe d'*open data* par défaut performant. Mais en réalité, **ce principe de « tout publier sauf ... » ne permet pas de distinguer clairement ce qui peut ou doit être publié de ce qui ne doit surtout pas l'être.**

Tout d'abord, la présence d'un secret dans un document ne suffit pas pour empêcher sa publication. En effet, l'article L311-7 du CRPA précise qu'**un document dont on peut occulter ou disjointre les secrets est communicable** après occultation ou disjonction de ces secrets¹⁶⁶. Si cet article élargit l'étendue des documents susceptibles d'être communiqués, il fait peser sur l'administration une double charge. D'une part, caviarder un document est une tâche chronophage. D'autre part, l'administration engage sa responsabilité si elle laisse par mégarde un secret dans un document publié.

Pour éviter que les administrations ne soient submergées par des demandes, la loi prévoit que **les administrations ne sont pas tenues de répondre à des demandes abusives**¹⁶⁷. **Mais ces cas sont définis de façon stricte** : il s'agit de demandes à caractère répétitif ou réclamant un volume trop important de documents et qui visent, de façon délibérée, à perturber le fonctionnement d'une administration¹⁶⁸. Le Conseil d'État a dû ajouter dans sa jurisprudence des cas où la demande peut être rejetée. C'est le cas si la charge liée au caviardage est trop importante pour l'administration, mais il convient ici de mettre en balance l'intérêt de cette communication pour le demandeur et la charge de travail qui pèse sur l'administration au regard des moyens dont elle dispose¹⁶⁹. C'est également le cas si le caviardage dénature trop le document¹⁷⁰.

¹⁶⁴ Article L311-4 du CRPA.

¹⁶⁵ Cela n'empêche pas le Gouvernement de décider de publier certains avis, en opportunité.

¹⁶⁶ Article L311-7 du CRPA.

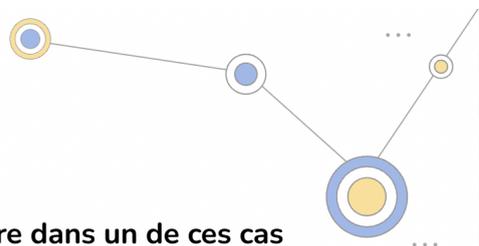
¹⁶⁷ Article L311-2 du CRPA.

¹⁶⁸ Avis 20170534 de la Cada.

¹⁶⁹ Décision n° 449620 du Conseil d'État.

¹⁷⁰ Décision n° 117750 du Conseil d'État.





Mais il n'est pas facile pour une administration de juger si elle rentre dans un de ces cas de figure qui la dispense de communiquer le document. Elle a donc le choix d'allouer des moyens pour caviarder les documents ou de refuser la communication et d'attendre que la Cada rende un avis, ce qui est généralement le plus avantageux pour elle. On comprend aussi qu'aucune administration n'est pas incitée à publier spontanément des documents qu'il faudrait préalablement caviarder. **Dans ces cas-là, loin d'être anecdotiques, l'open data par défaut est donc utopique.**

Occulter les mentions d'un document : oui, mais lesquelles ?

Par ailleurs, **il n'est pas toujours facile de juger si une information est couverte ou non par un secret.** La protection des données personnelles offre de multiples exemples de cette complexité et est un sujet régulièrement rencontré par les administrations. Ainsi, la CNIL a montré dans une consultation auprès d'administrations diverses que 56% d'entre elles ont déjà été confrontées à une question qui relevait de la protection des données personnelles dans le cadre de l'*open data*¹⁷¹.

On appelle donnée personnelle « toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement »¹⁷². Les données personnelles sont donc sensibles dans la mesure où elles peuvent permettre d'identifier une personne et porter atteinte à la protection de la vie privée ou au secret médical. Avec le RGPD, qui encadre fermement leur traitement, on passe d'une logique d'autorisation préalable auprès de la CNIL à une logique de responsabilité du responsable de traitement, avec un contrôle *a posteriori* et parfois un accompagnement *a priori* par la CNIL. Les administrations peuvent donc être sanctionnées par la CNIL si elles ne respectent pas le RGPD. **Pour sortir du cadre du RGPD, les données personnelles doivent être anonymisées**, c'est-à-dire qu'il faut rendre impossible l'identification d'une personne à partir de ces données. Mais anonymiser des données est loin d'être une formalité car il est très difficile de se couvrir contre le risque de réidentification. En effet, **en croisant plusieurs informations, il est souvent possible de réidentifier une personne.**

¹⁷¹ CNIL, « Rapport de résultats : Consultation Open data et données personnelles », 2014. URL : www.data.gouv.fr/storage/f/2014-04-16T09-44-53/rapport-questionnaire-cnil-open-data-et-donnees-personnelles-vdef.pdf

¹⁷² www.cnil.fr/fr/cnil-direct/question/une-donnee-caractere-personnel-cest-quoi

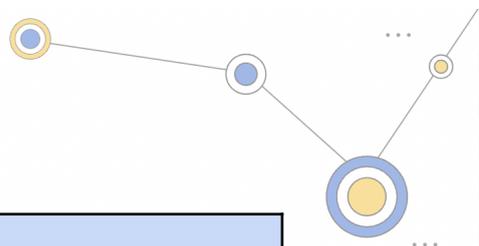
La K-anonymisation : principe et limites

Pour des données anonymisées et structurées dans un tableau, on peut définir une propriété, appelée k-anonymisation, qui mesure à quel point il est difficile de réidentifier une personne à partir de ces données. Plus précisément, les données sont dites k-anonymes si les informations relatives à une personne ne peuvent être distinguées d'au moins $k - 1$ autres personnes du même jeu de données. Prenons l'exemple suivant :

Nom	Âge	Genre	Taille	Poids	Ville	Date de dernière consultation	Maladie
Jeanne	29	Femme	162 cm	73 kg	Nantes	12 mars 2024	Pas de maladie
Samira	24	Femme	164 cm	71 kg	Paris	30 mai 2024	Infection virale
Hélène	27	Femme	172 cm	67 kg	Nantes	28 avril 2024	Cancer
Sarah	30	Femme	162 cm	72 kg	Paris	30 mai 2024	Lié au cœur
Mehdi	17	Homme	180 cm	86 kg	Paris	21 juin 2024	Cancer
Rémi	19	Homme	175 cm	80 kg	Paris	16 mai 2024	Lié au cœur
Quentin	18	Homme	187 cm	83 kg	Paris	27 avril 2024	Infection virale

Que l'on peut anonymiser ainsi :

Nom	Âge	Genre	Taille	Poids	Ville	Date de dernière consultation	Maladie
*	> 20	Femme	162 cm	73 kg	Nantes	mars 2024	Pas de maladie
*	> 20	Femme	164 cm	71 kg	Paris	mai 2024	Infection virale
*	> 20	Femme	172 cm	67 kg	Nantes	avril 2024	Cancer
*	> 20	Femme	162 cm	72 kg	Paris	mai 2024	Lié au cœur
*	≤ 20	Homme	180 cm	86 kg	Paris	juin 2024	Cancer
*	≤ 20	Homme	175 cm	80 kg	Paris	mai 2024	Lié au cœur
*	≤ 20	Homme	187 cm	83 kg	Paris	avril 2024	Infection virale



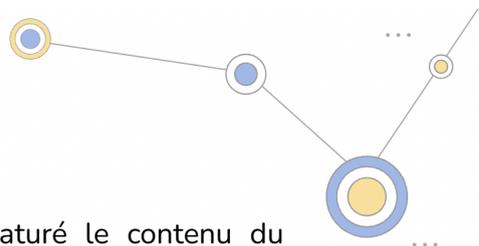
Ces données sont 2-anonymes pour l'âge, le genre ou la ville : il existe toujours au moins deux lignes contenant les mêmes valeurs pour ces attributs. Il n'est donc pas possible, connaissant ces attributs pour une personne, d'isoler sa ligne et d'en déduire par exemple sa maladie. En revanche, si Jeanne avait un cancer, alors on pourrait déduire que les deux femmes qui habitent à Nantes ont un cancer, et ainsi découvrir la maladie de Jeanne et Hélène.

Même si la k-anonymisation est correctement mise en place dans le jeu de données, un apport de données extérieures peut lui aussi rompre l'anonymat. Par exemple, si l'on a croisé Hélène sortant de l'hôpital en avril 2024, on peut savoir qu'elle est atteinte d'un cancer. Cela montre qu'il est généralement impossible de garantir mathématiquement que la réidentification d'une personne dans un jeu de données anonymisées est impossible. **Toutes les données sont potentiellement identifiantes**, il faut donc se demander quels risques sont encourus en diffusant ce genre de données et les comparer aux bénéfices qu'apporte cette diffusion.

Il est possible d'illustrer les difficultés de l'anonymisation et le risque de réidentification en s'appuyant sur les décisions des tribunaux judiciaires sur le portail Judilibre de la Cour de Cassation¹⁷³. De tels documents contiennent généralement un grand nombre de données à caractère personnel dont la divulgation irait à l'encontre de la protection de la vie privée. Ces données, notamment les prénoms, noms et adresses, sont donc anonymisées. Si on cherche des décisions qui contiennent la mention « directeur général de la société », on trouve notamment une décision dont voici quelques extraits :

- « 31^o à M. [BY] [UA], domicilié [Adresse 3], » : le nom d'une personne, M. [BY] [UA], a bien été anonymisé, ainsi que son adresse.
- « I - La Société [UA] frères, société anonyme, dont le siège est [Adresse 1], » : normalement, un nom de société n'est pas une donnée à caractère personnel, mais ici la Cour de Cassation a estimé, à raison, qu'il valait mieux l'occulter car le divulguer aurait permis de connaître le nom de M. [BY] [UA].
- « 2^o à M. [BY] [UA], en son nom personnel et en sa qualité de directeur général et administrateur de la société Sager, » : ici, on apprend que M. [BY] [UA] est directeur général d'une société. Or le nom des dirigeants de société est une information publique que l'on trouve facilement en ligne. Nous n'avons donc eu aucun mal à réidentifier M. [BY] [UA].

¹⁷³ www.courdecassation.fr/acces-rapide-judilibre



Aurait-il fallu supprimer cette mention ? Cela n'aurait-il pas dénaturé le contenu du document ? Comment savoir si une information peut permettre de réidentifier quelqu'un ? Cet exemple montre qu'il est parfois difficile pour une administration de juger si le contenu d'un document est bien communicable ou pas, et de le traiter pour protéger les secrets qu'il renferme.

Les risques réels ou supposés de la diffusion

Parfois, certaines données personnelles ont vocation à être publiées, c'est notamment le cas des nominations au sein de la fonction publique. La DILA et les ministères sont chargés de publier ces données sur le Journal Officiel ou dans différents bulletins : Bulletin Officiel du Ministère de l'Intérieur, Bulletin Officiel du Ministère de la Justice, etc. Ces données, bien que publiques, ne sont volontairement pas agrégées par la DILA au sein d'une seule et même base de données pour les rendre difficiles à compiler. L'idée est d'empêcher le croisement d'informations pour les protéger les individus, notamment ceux qui exercent des métiers sensibles comme les policiers ou les gendarmes.

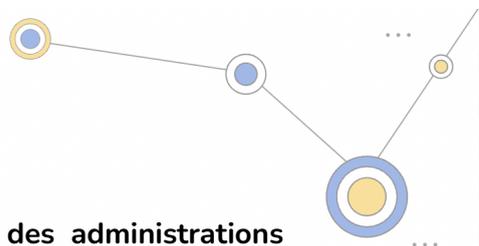
Pour les besoins de l'association Regards Citoyens, qui souhaitait plus de transparence sur les nominations en cabinets ministériels, un citoyen a voulu pallier l'absence d'annuaire central recensant toutes les nominations en créant son propre site : **JORFSearch**¹⁷⁴. Cet annuaire est consulté par près de 20 000 personnes tous les jours¹⁷⁵. Bon nombre d'entre elles sont convaincues qu'il s'agit d'un site officiel géré par l'administration et ne soupçonnent pas qu'un citoyen en est à l'initiative et réalise seul ce travail laborieux de croisement et de nettoyage des données, lui attirant régulièrement les foudres de la CNIL et de la DILA. Deux visions de l'*open data* s'opposent donc ici. Même lorsque des données sont publiées en ligne, la manière de les publier peut avoir une importance sur les risques induits par ces données.

Mais il n'y a pas que les données personnelles dont la diffusion peut présenter des risques. On peut citer ainsi les plans des réseaux d'eau qui non seulement ne contiennent a priori aucun secret, mais peuvent en plus être considérés comme des informations environnementales et donc bénéficier d'un régime juridique favorisant l'ouverture. Pourtant, il serait très dangereux pour la société qu'une telle information tombe dans les mains d'acteurs malveillants. C'est pourquoi la Cada, dans plusieurs avis, est favorable à la communication de ces données sous réserve qu'elle ne soit pas susceptible de porter atteinte à la sécurité publique¹⁷⁶.

¹⁷⁴ www.steinertriples.ch/ncohen/data/nominations_JORF/

¹⁷⁵ Interview du fondateur de JORFSearch

¹⁷⁶ Voir par exemple les avis 20212742, 20182902 et 20212398 de la Cada.



Même si certains risques sont bien réels, **il n'est pas rare que des administrations brandissent abusivement la protection de certains secrets**, comme la protection de la vie privée ou du secret des affaires, pour empêcher ou retarder la publication de certains documents. Ce fut le cas en 2021 lors de du « scandale McKinsey » qui a mis en évidence un recours massif de l'État aux prestations de cabinets de conseil, largement critiqué par un rapport du Sénat en mars 2022¹⁷⁷. Le journal Le Monde¹⁷⁸ et l'association Anticor¹⁷⁹ ont essayé d'obtenir plus de détails sur le contenu de ces prestations, mais ils ont déploré un manque de transparence de la part de certaines administrations faisant valoir la protection du secret des affaires pour éviter de communiquer des documents. La Cada a finalement rendu un avis favorable à Anticor et a demandé au ministère de la Santé de lui transmettre des documents tout en occultant les mentions relevant du secret des affaires, comme les prix unitaires¹⁸⁰. Cette affaire soulève aussi le manque de transparence des « accords-cadres » et l'opacité entourant le contenu des marchés subséquents, problème soulevé également par le classement EuroPAM mentionné précédemment. Dans l'affaire des IPS, le ministère de l'Éducation nationale avait, elle aussi, utilisé à tort la notion de secret, en l'occurrence la protection de la vie privée. Pourtant, ces données n'avaient aucun caractère personnel, comme l'ont souligné la Cada puis le Tribunal administratif de Paris.

Communiquer or not communiquer, that is the question ?

Savoir si l'on peut communiquer ou non un document administratif est donc loin d'être simple pour les administrations, peu encouragées à prendre l'initiative des publications. Sans compter que **même des administrations expertes de ces sujets** (la Cada, les tribunaux administratifs, le Conseil d'État), **se contredisent parfois entre elles**.

C'est le cas notamment au sujet de l'annuaire des avocats tenu par le Conseil national des barreaux (CNB). Celui-ci le publiait en ligne via un moteur de recherche qui ne permettait pas techniquement de télécharger la totalité de la base de données. L'association Ouvre-boîte a donc demandé au CNB de publier en ligne la totalité de la base, ce que le CNB a refusé de faire. La Cada a donné raison à Ouvre-boîte¹⁸¹, mais le Tribunal administratif de Paris a donné raison au CNB en se fondant sur un loi de 1971 qui laisse au CNB le soin de définir les modalités de publication de l'annuaire¹⁸². C'est finalement le

¹⁷⁷ Sénat, Rapport de commission d'enquête « Un phénomène tentaculaire : l'influence croissante des cabinets de conseil sur les politiques publiques », n° 578 (2021-2022), tome I, déposé le 16 mars 2022.

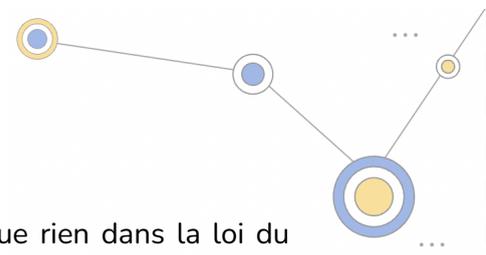
¹⁷⁸ Luc Martinon et al., « Recours aux consultants privés : la grande opacité de l'État », Le Monde, 2022. URL : www.lemonde.fr/les-decodeurs/article/2022/03/25/recours-aux-consultants-privés-la-grande-opacité-de-l-État_t_6119189_4355770.html

¹⁷⁹ www.anticor.org/2022/03/20/anticor-saisit-la-cada-dans-le-dossier-mckinsey/

¹⁸⁰ www.anticor.org/wp-content/uploads/2021/07/Avis-CADA-Mc-Kinsey.pdf

¹⁸¹ Avis 20191272 de la Cada.

¹⁸² Article 21-1 de la loi du 31 décembre 1971 portant réforme de certaines professions judiciaires et juridiques



Conseil d'État qui a tranché en faveur d'Ouvre-boîte¹⁸³, estimant que rien dans la loi du 31 décembre 1971 n'indique que l'on pourrait déroger aux principes de la loi Lemaire, et notamment que les documents administratifs doivent être publiés dans leur intégralité, dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé.

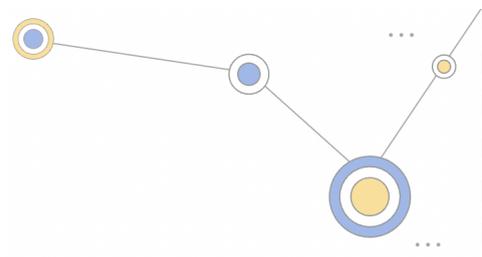
Autre exemple, celui de la convention de prêt conclue en 2014 entre le Front national et la *First Czech Russian Bank*. Une journaliste de Médiapart a demandé en 2016 à la Commission nationale des comptes de campagne et des financements politiques (CNCCFP) de lui communiquer ce document. La CNCCFP a refusé en brandissant la protection du secret en matière industrielle et commerciale. La Cada a rendu un avis favorable à Mediapart, arguant qu'il n'y avait pas de secret à protéger mises à part les coordonnées bancaires du Front national qu'il convenait d'occulter¹⁸⁴. Le Tribunal administratif de Paris est allé dans le sens de la Cada et a enjoint la CNCCFP de communiquer le document, tout en occultant les coordonnées bancaires du parti. Finalement, le Conseil d'État a cassé le jugement du tribunal administratif : la durée du prêt et le taux d'intérêt relèvent effectivement du secret industriel et commercial car ils reflètent la stratégie commerciale du prêteur¹⁸⁵. Le Conseil d'État a donc demandé à la CNCCFP de communiquer les documents à Mediapart en prenant soin d'occulter à la fois les coordonnées bancaires du Rassemblement national et les mentions relatives à la durée du prêt et au taux d'intérêt.

¹⁸³ Décision n° 450739 du Conseil d'État

¹⁸⁴ Avis 20161117 de la Cada.

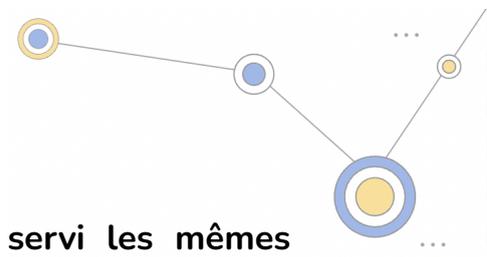
¹⁸⁵ Décision n° 420467 du Conseil d'État





3. Redonner un sens à l'ouverture pour relancer la dynamique





3.1. Une politique d'ouverture qui n'a pas toujours servi les mêmes objectifs

3.1.1. Aux origines de l'*open data*, une exigence de transparence

De la société aux gouvernements : l'inscription de l'open data à l'agenda politique des pays anglosaxons

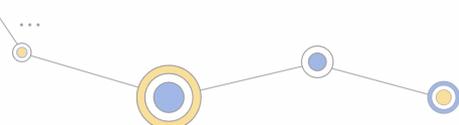
Le mouvement *open data* naît et se structure autour des enjeux de transparence dans les sociétés civiles américaine et britannique du début des années 2000.

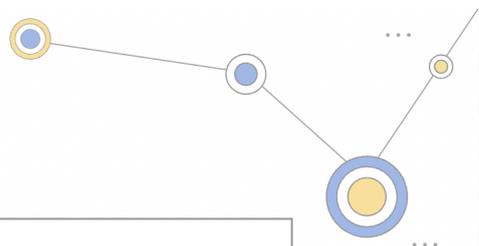
En 2004, au Royaume-Uni, l'économiste et activiste Rufus Pollock crée l'association Open Knowledge Foundation pour rassembler des acteurs militants convaincus de la nécessité de rendre des comptes aux administrés. L'association joue un rôle crucial dans la structuration de la pensée de l'*open data* en théorisant par exemple le « savoir ouvert ». Elle travaille aussi au développement d'outils *open source*, comme le logiciel CKAN qui permet de créer un catalogue de données, pour outiller les administrations et les entreprises qui souhaiteraient partager leurs données. En parallèle, elle mène des actions de sensibilisation en lançant par exemple le site *OpenSpending*¹⁸⁶ ou *Wheredoesmymoneygo* qui centralise et présente tous les documents relatifs aux dépenses des différents gouvernements disponibles sur Internet.

Deux ans plus tard, le journal *The Guardian* jette un pavé dans la mare en publiant l'article « Rendez-nous nos joyaux de la couronne » ("*Give us back our crown jewels*")¹⁸⁷ (Figure 13). Dans cet article, les journalistes expliquent que dans la mesure où les données publiques sont produites avec de l'argent public, issu des impôts, elles reviennent de droit aux citoyens. Dans les mois qui suivent la parution de l'article, le journal coordonne ensuite la campagne « Libérez nos données » ("*Free our data*") pour réclamer la gratuité des données publiques. L'objectif est alors de susciter un élan citoyen en faveur de l'ouverture de données.

¹⁸⁶ www.openspending.org/

¹⁸⁷ www.freeourdata.org.uk/





The Guardian

Give us back our crown jewels

Our taxes fund the collection of public data - yet we pay again to access it. Make the data freely available to stimulate innovation, argue Charles Arthur and Michael Cross

**Charles Arthur and
Michael Cross**

Thu 9 Mar 2006 11.47 CET

Figure 13 : Article du Guardian réclamant le partage des données que les administrations produisent avec de l'argent issu des impôts. Source : The Guardian, 2006.

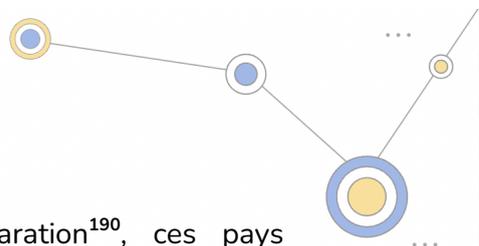
Les gouvernements américain et britannique prennent note de la croissance et de la structuration de ce mouvement civil et intègrent cette exigence de transparence dans leur agenda politique. En janvier 2009, quelques semaines après son élection, Obama signe ainsi deux mémorandums sur l'*open government* et reconnaît les principes de Sebastopol dans le contexte de l'après Afghanistan et Irak. En mars 2009, le portail *open data* américain, *data.gov*, est lancé par le directeur de l'innovation de la Maison Blanche ; et en décembre 2009, la directive *Open Government* contraint toutes les agences américaines à fournir un plan d'ouverture de leurs données.

Le portail *open data* britannique, *data.gov.uk*, est lancé l'année suivante alors que le tout nouveau premier ministre, David Cameron, présente son « agenda de transparence » (*Transparency Agenda*). Il met ainsi l'ouverture des données au cœur de son programme politique, *Big Society*¹⁸⁸, par lequel il souhaite transférer et redonner plus de pouvoir et d'indépendance aux gouvernements locaux et aux citoyens¹⁸⁹. Pour surveiller la bonne conduite de cette politique d'ouverture, il crée le *Public Sector Transparency Board*, composé de représentants de l'administration gouvernementale et de personnalités de l'*open data* chargé de mener à bien l'agenda du gouvernement sur les questions de transparence, en particulier la mise à disposition des données sur *data.gov.uk*.

Les États-Unis et le Royaume-Uni n'envisagent pas le sujet de l'*open data* comme un sujet de politique interne uniquement, **mais font de leur nouvelle exigence de transparence un enjeu international**. En 2011, ils s'associent ainsi à 6 autres pays dans le cadre de l'*Open*

¹⁸⁸ Il faut opposer *Big Society* à *Big State*, soit la centralisation de l'État.

¹⁸⁹ uk.ambafrance.org/data-gov-uk-l-ouverture-des



Government Partnership. En signant l'Open Government Declaration¹⁹⁰, ces pays s'engagent à la transparence, la participation et la collaboration avec la société civile. Ce partenariat pour un gouvernement ouvert compte aujourd'hui 70 pays.

L'année 2013 marque un tournant pour le mouvement *open data* qui fait officiellement, et pour la première fois, partie des sujets discutés à la table des grandes puissances mondiales. Le premier ministre David Cameron accueille en effet le sommet du G8 qu'il décide de dédier aux enjeux de transparence. À l'occasion de ce sommet, les pays membres signent une charte¹⁹¹ dans laquelle ils reconnaissent le potentiel des données pour renouveler la croissance économique et rendre les gouvernements plus efficaces¹⁹². Ils s'engagent aussi à :

- Faire de l'ouverture des données publiques la nouvelle norme ;
- Publier des données de qualité, sous leur forme originale ;
- Créer des portails uniques pour publier ces données (gratuitement et dans un format ouvert) ;
- Partager son expertise de l'ouverture des données avec les autres pays et identifier les jeux de données à ouvrir en priorité ;
- Développer la culture des données et encourager la société civile à se saisir des données publiées.

Une exigence de transparence aux racines de l'open data à la française

Si le mouvement *open data* n'est pas né en France, il rencontre rapidement un écho dans la société civile française. En 2009, l'association LiberTIC est créée à Nantes pour promouvoir la démocratie en ligne à travers l'ouverture des données publiques. La même année, est également fondée l'association Regards citoyens, qui milite pour le partage et la diffusion de l'information politique. Elle lance par exemple le site *NosDéputés.fr*, qui synthétise les différentes activités législatives et de contrôle du gouvernement des élus de la nation pour permettre aux citoyens de mieux comprendre le travail de leurs représentants (Figure 14).

¹⁹⁰ www.opengovpartnership.org/fr/process/joining-ogp/open-government-declaration/

¹⁹¹ G8, « Charte pour l'ouverture des données publiques », 2013. URL : fr.scribd.com/document/148580461/Charte-du-G8-pour-l-Ouverture-des-Donnees-Publiques-Francais

¹⁹² *Ibid.*

NOS DÉPUTÉS.FR
OBSERVATOIRE CITOYEN DE L'ACTIVITÉ PARLEMENTAIRE

Les Députés | Les Dossiers | **Les Citoyens** | FAQ

Tous les citoyens | Les derniers commentaires | Assister aux débats

Ce site présente les travaux des députés de la précédente législature. NosDéputés.fr reviendra d'ici quelques mois avec une nouvelle version pour les députés élus en 2024.

Bienvenue sur NosDéputés.fr

NosDéputés.fr est un site qui cherche à mettre en valeur l'activité parlementaire des députés de l'Assemblée nationale Française. En synthétisant les différentes activités législatives et de contrôle du gouvernement des élus de la nation, ce site essaie de donner aux citoyens de nouveaux outils pour comprendre et analyser le travail de leurs représentants.

Conçu comme une plateforme de médiation entre citoyens et députés, le site propose à chacun de participer et de s'exprimer sur les débats parlementaires. Au travers de leurs commentaires, les utilisateurs sont invités à créer le débat en partageant leur expertise lorsque cela leur semble utile. Peut-être pourront-ils ainsi nourrir le travail de leurs élus ?

Vous pouvez consulter l'activité de leurs collègues du **Sénat** sur notre autre initiative **Nos Sénateurs**.

Vous pouvez retrouver le bilan complet des députés de la précédente législature sur 2017-2022.nosdeputes.fr.

Toute l'équipe du collectif RegardsCitoyens.org.

La députée du jour


Claudia Rouaux
Un autre député au hasard

Trouver son député

Pour retrouver votre député sur le site, vous pouvez saisir son nom.

Si vous ne le connaissez pas, indiquez votre code postal ou le nom de votre commune, et nous essaierons de le trouver pour vous :

Exemples : patrick, 77840, saint-herblain, trois rivières, ...

Les principaux mots clés de la législature

AME amende ASN assurance audiovisuel autonomie cotisation culture demandeur dette directive disponible en vidéo dispositif domicile déficit eau EDF EHPAD européen exploitation éducation élevé ferroviaire forêt handicap immigration incendie index industriel installation intervention intervention uniquement IRSN jeux judiciaire juge locataire logement loyer LPM magistrat maladie mer militaire mineur mobilité motion mourir médecin

Activité parlementaire de la législature

Activité	Députés	Interventions	Amendements	Propositions	Quest. Écrites
50 %	~15%	~15%	~15%	~15%	~15%
75 %	~25%	~25%	~25%	~25%	~25%
100 %	~40%	~40%	~40%	~40%	~40%

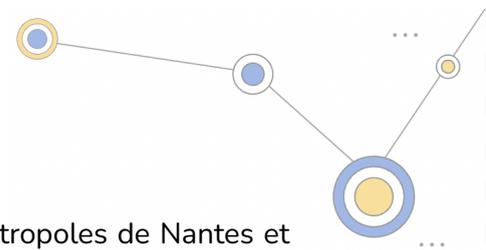
Figure 14 : Page d'accueil du site Internet NosDéputés.fr créé et alimenté par l'association Regards Citoyens (le « député du jour » présenté ici n'est pas un choix des auteurs, mais le résultat d'une sélection aléatoire par le site le jour de la capture d'écran) . Source : nosdéputés.fr.

Le mouvement se structure avec la création d'Open Knowledge France en 2012¹⁹³ : l'objectif est de rassembler les communautés existantes et de les associer aux activités de l'Open Knowledge Foundation pour développer des projets et activités spécifiques et valoriser les initiatives françaises à l'international. C'est dans ce cadre qu'est créé le site *madada.fr* en 2019, devenu depuis un site de référence pour les demandes de documents administratifs.

Mais la promotion de l'ouverture des données ne se confine pas uniquement au seul monde associatif. **Les collectivités territoriales ne tardent pas à se saisir du sujet.** Dès 2010, la métropole de Rennes lance l'opération « Data Breizh - La Bretagne en données »¹⁹⁴ et crée

¹⁹³ fr.okfn.org/2012/12/12/houssommeslokfn/

¹⁹⁴ data-bzh.fr/rennes-ville-open-data/



un premier portail *open data*. Elle sera rapidement suivie par les métropoles de Nantes et de La Rochelle. L'idée de publier gratuitement des données publiques vient donc d'abord des collectivités territoriales alors que l'État conserve encore une vision de l'ouverture des données « payante » portée par l'Agence pour le patrimoine immatériel de l'État¹⁹⁵.

L'ouverture des données publiques s'inscrit finalement à l'agenda politique national en 2011. Dans la perspective des élections présidentielles de l'année suivante, le candidat et président sortant Nicolas Sarkozy souhaite faire des données ouvertes un outil de promotion de son bilan. Est alors lancée la mission Etalab, placée sous l'autorité du premier ministre, pour créer le portail national *data.gouv.fr*. Au-delà de cette plateforme, inaugurée en décembre 2011, la mission crée la Licence Ouverte sous laquelle les données publiques seront publiées pour garantir leur gratuité et la possibilité de réutilisations commerciales à condition de citer la source¹⁹⁶. Cette initiative est saluée par l'association Regards citoyens.

La France rejoint officiellement les pays « ouverts » en intégrant l'*Open Government Partnership*. L'année suivante, à l'occasion du sommet mondial du partenariat, elle fait partie des 16 premiers signataires de l'*Open Data Charter* et s'engage à ouvrir les données publiques « par défaut ».

Le contexte international et les engagements pris par la France devant ses partenaires la poussent à adopter une série de lois sur l'ouverture des données publiques. En 2014, le Conseil national du numérique organise une grande concertation sur le numérique à l'issue de laquelle il remet un rapport dans lequel il préconise d'inscrire dans la loi le principe d'ouverture par défaut et de gratuité des données publiques, de mettre en place des dispositifs d'accompagnement des collectivités pour les aider à s'approprier leurs obligations légales et de renforcer la médiation numérique sur les sujets données¹⁹⁷. Dans les mois qui suivent la publication de ce rapport, pendant l'été 2015, une nouvelle consultation est alors organisée en ligne pour réunir les avis citoyens sur l'élaboration de la future loi pour une République numérique¹⁹⁸ (Figure 15). Plus de 20 000 citoyens, en grande majorité issus du monde associatif, ont donc la possibilité de s'exprimer sur les modalités et le degré d'ouverture qu'ils estiment souhaitables. La loi, finalement adoptée en septembre 2016, s'inscrit ainsi dès sa conception dans un esprit inédit d'ouverture et de collaboration avec la société civile.

¹⁹⁵ Samuel Goëta, *Les Données de la Démocratie*, 2024, p.85.

¹⁹⁶ www.etalab.gouv.fr/wp-content/uploads/2014/05/Licence_Ouverte.pdf

¹⁹⁷ Conseil national du numérique, « Ambition numérique : pour une politique française et européenne de la transition numérique », 2015. URL : cnnumerique.fr/files/2017-10/CNNNum--rapport-ambition-numerique.pdf

¹⁹⁸ www.republique-numerique.fr/project/projet-de-loi-numerique/consultation/consultation



Projet de loi pour une République numérique

8 488 Contributions 147 521 Votes 21 469 Participants

Consultation Terminé Classements des contributions Terminé Synthèse Réponses du Gouvernement Projet de loi transmis au Conseil d'Etat

Consultation terminée. Merci à tous d'avoir contribué.

Consultation

Du 26 sept. 2015 à 09:00 au 18 oct. 2015 à 23:59

Accéder au [texte intégral de l'avant-projet de loi](#) pour une République Numérique

TITRE Ier
La circulation des données et du savoir

Chapitre Ier
Economie de la donnée

Section 1
Ouverture des données publiques

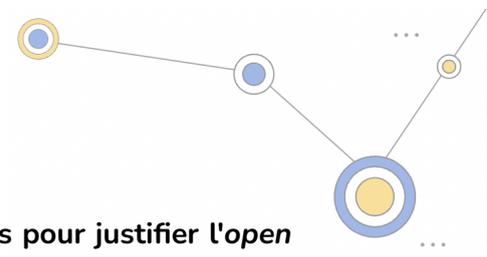
160 propositions Tri aléatoire + Nouvelle proposition

Proposition	Votes	Amendements	Arguments	Source
Gouvernement • 26 sept. 2015 Épinglé Article 1 - Open data par défaut (obligation de diffuser en ligne les principaux documents et données des organismes publics)	2 440	117	95	8
Timothé Poulain • 6 oct. 2015 Généralisation de l'utilisation des logiciels Open source par les instances publiques	64	0	8	0
Remtola • 28 sept. 2015 Non exclusion des citoyens réticents à l'égare du numérique	85	2	4	0
Armelle Gilliard - La reine Merlin • 18 oct. 2015 Continuer d'accompagner le mouvement open data et la montée en compétences de tous sur ces sujets	4	0	0	0
Regards Citoyens • 15 oct. 2015 Informers les citoyens de l'existence des données publiées	88	0	0	0

Figure 15 : Plateforme de la consultation en ligne organisée dans le cadre de l'élaboration de la loi pour une République numérique. Source : republique-numerique.fr.

L'analyse du contexte d'adoption de la loi pour une République numérique nous permet ainsi de comprendre que la mise à l'agenda politique de la question de l'ouverture des données en France est essentiellement la conséquence des « pressions » des partenaires internationaux et d'une « coalition de réformateurs numériques gravitant dans l'entourage du [Parti socialiste] »¹⁹⁹.

¹⁹⁹ Samuel Goëta, Les Données de la Démocratie, 2024, p.92.



3.1.2. Un glissement vers la mise en avant d'objectifs économiques pour justifier l'open data

« L'impact attendu des politiques de données ouvertes et la mise en place de portails de données est bien entendu économique. »²⁰⁰

European Data Portal, « Creating Value through data », 2015.

Vers la création d'un marché unique des données

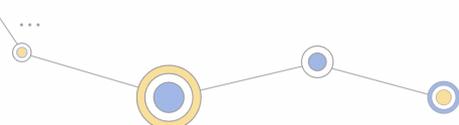
En 2003, les pays membres de l'Union européenne se dote d'un cadre commun sur la réutilisation des informations publiques en adoptant la Directive PSI (« Public Sector Information directive »), qui veut garantir l'égalité de traitement de tous les réutilisateurs des informations publiques publiées pour être réutilisées, qu'il s'agisse de réutilisateurs particuliers ou d'entreprises. Dix ans plus tard, l'Union européenne souhaite favoriser encore davantage les réutilisations des informations publiques : désormais, toute information publique rendue accessible par les États membres est « par principe » réutilisable. Dans le cas où l'accès à cette information ne serait pas gratuit, son prix ne doit pas dépasser le coût marginal de la mise à disposition pour réutilisation²⁰¹.

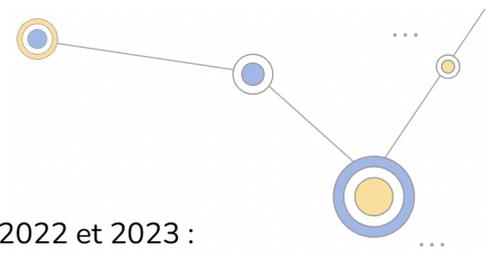
Ainsi, dès le début des années 2000, l'Union européenne associe à la question des informations publiques, puis des données publiques, des enjeux et des objectifs d'abord économiques : il s'agit de créer un marché européen de l'information publique. Cette ambition est réaffirmée avec force dans la Stratégie européenne pour les données²⁰², document stratégique publié par la Commission européenne en 2020 : il faut faire de l'Europe un leader dans un monde « *data driven* » et créer le cadre d'un véritable marché unique des données qui, à horizon 2025, permettrait la circulation des données à travers les frontières et les secteurs d'une manière sûre et sécurisée - ce qui implique de repenser le cadre européen des données, qui reposait historiquement sur la protection des données personnelles (RGPD). L'objectif est donc de stimuler les performances économiques des pays membres, en s'appuyant sur les trois piliers du marché intérieur : libre circulation des données, transparence et juste compétition.

²⁰⁰ Capgemini Consulting, « Creating Value through data », 2015. URL : data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

²⁰¹ Directive 2013/37/UE modifiant la directive 2003/98/CE.

²⁰² Commission européenne, « Stratégie européenne pour les données », 2020. URL : commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr





Cette stratégie repose sur trois textes votés respectivement en 2019, 2022 et 2023 :

- La directive sur les données ouvertes, l'*Open data Act*²⁰³, introduit la notion de « jeux de données de forte valeur » (*High Value Datasets*), des données stratégiques devant être rendues accessibles gratuitement par les États membres d'ici juin 2024²⁰⁴.
- Le règlement sur la gouvernance des données, le *Data Governance Act*²⁰⁵, définit un certain nombre d'exigences socles pour créer un cadre harmonisé d'échanges de données.
- Le règlement sur les données, le *Data Act*²⁰⁶, confère pour la première fois un pouvoir contraignant aux administrations publiques en leur permettant d'utiliser des données de détenteurs privés pour gérer des crises, les prévenir ou conduire des missions d'intérêt public. Applicable en 2025, il constituera une étape fondamentale dans la mobilisation de données privées à des fins d'intérêt général.

Cette vision essentiellement économique du potentiel des données, aujourd'hui portée par l'Union européenne, s'inscrit dans la continuité de la **vision patrimoniale des données adoptée par la France** dès le début des années 2000. En 2006, le rapport de la commission sur l'économie de l'immatériel²⁰⁷ présente les données publiques comme un patrimoine immatériel de l'État. À la suite de ce rapport est créée l'Agence du Patrimoine Immatériel de l'État (APIE), rattachée à la DGFIP et à la Direction générale du Trésor²⁰⁸. L'APIE évalue le montant des « actifs » immatériels de l'État (13,5 Md€ au bilan de l'État en 2008²⁰⁹) et accompagne les administrations dans la mise en place de portails sur lesquels elles peuvent vendre leurs données.

Si elle a évolué au fil du temps pour davantage intégrer les enjeux de transparence, **l'approche économique de l'ouverture des données est encore très présente dans les discours et justifications officiels**. En 2021, dans le Cadre interministériel d'administration de la donnée, on peut par exemple lire que la donnée est un « *actif stratégique de la*

²⁰³ Directive (UE) 2019/1024 du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

²⁰⁴ Règlement d'exécution (UE) 2023/138 du 21 déc. 2022 établissant une liste d'ensembles de données de forte valeur spécifiques et les modalités de leur publication et de leur réutilisation.

²⁰⁵ Règlement (UE) 2022/868 du 30 mai 2022 portant sur la gouvernance européenne des données.

²⁰⁶ Règlement (UE) 2023/2854 du 13 déc. 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données et modifiant le règlement (UE) 2017/2394 et la directive (UE) 2020/1828

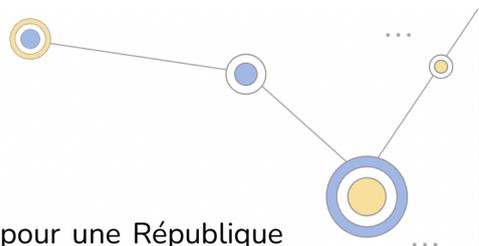
²⁰⁷ Maurice Lévy et Jean-Pierre Jouyet, « L'Économie de l'immatériel : la croissance de demain », novembre 2006.

²⁰⁸ En 2020, l'APIE devient la mission Appui au patrimoine immatériel de l'État et dépend désormais de la Direction des affaires juridiques (DAJ) du ministère de l'Économie.

²⁰⁹ Agence du Patrimoine Immatériel de l'État, « Rapport d'activité 2016 : Le patrimoine immatériel public, une ressource pour moderniser et créer de la valeur », 2016. URL :

www.economie.gouv.fr/files/files/directions_services/apie/APIE/Publications/Rapport_activite_APIE_2016.pdf?v=1561651393





révolution numérique »²¹⁰. Mais surtout, cette vision structure la loi pour une République numérique en guidant notamment les choix de titres de chapitres et de sections (Figure 16).

- 
- ▣ Titre Ier : LA CIRCULATION DES DONNÉES ET DU SAVOIR (Articles 1 à 39)
 - ▣ Chapitre Ier : Economie de la donnée (Articles 1 à 29)
 - + Section 1 : Ouverture de l'accès aux données publiques (Articles 1 à 16)
 - + Section 2 : Données d'intérêt général (Articles 17 à 24)
 - + Section 3 : Gouvernance (Articles 25 à 29)
 - ▣ Chapitre II : Economie du savoir (Articles 30 à 39)

Figure 16 : Capture d'écran Légifrance des premières en-têtes de la loi pour une République numérique. Source : Légifrance.

Un potentiel plus qu'incertain

« On sait très peu de choses sur les sous-jacents économiques et organisationnels et sur les implications de l'utilisation des données ouvertes pour l'organisation ou l'économie dans son ensemble. »

Heli Koski dans un rapport pour le ministère des finances finlandais²¹¹.

Si les politiques européennes et françaises justifient l'ouverture des données publiques surtout et avant tout par leur potentiel économique, il semblerait qu'aucune étude sérieuse n'ait encore réussi à démontrer la réalité et l'ampleur de ce potentiel. Les promesses de croissance faites aux administrations pour les inciter à ouvrir plus largement leurs données feraient ainsi courir le risque de susciter des déceptions et des frustrations chez ses agents, susceptibles de les décourager et *in fine* de ralentir le rythme de l'ouverture.

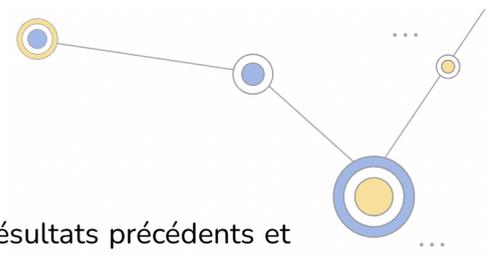
Le Portail européen des données, alors même qu'il multiplie les rapports d'évaluation de l'impact économique des données ouvertes, ne manque pas de souligner **les faiblesses des études menées ces dernières années à ce sujet**²¹² :

- Ces dernières manquent cruellement de transparence concernant les méthodologies utilisées, qui ne sont souvent pas décrites, voire même pas évoquées. Par ailleurs, les hypothèses de départ ne sont que très rarement explicitées.
- Les liens sont souvent peu clairs entre la littérature existante et la méthodologie. Certaines études se contentent par exemple de partir de résultats d'autres auteurs

²¹⁰ Cadre interministériel d'administration de la donnée, DINUM, 2021. URL : www.numerique.gouv.fr/uploads/Cadre-interministeriel-donnee.pdf

²¹¹ Heli Koski, « The Impact of open data – a preliminary study », 2015. URL : vm.fi/documents/10623/1107406/The+Impact+of+Open+Data/1c432b3a-a5e8-41ea-a5ea-135280a69ea3?version=1.0

²¹² Commission européenne, « The Economic Impact of Open Data - Opportunities for value creation in Europe », 2019. URL : data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf

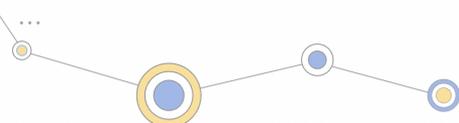


et d'y appliquer un ratio, ou alors prennent l'ensemble des résultats précédents et présentent leur moyenne comme étant le bon résultat.

- Les auteurs ont la fâcheuse tendance de sélectionner des nombres « arrangeants » issus de sources secondaires souvent obsolètes, plutôt que de collecter eux-mêmes de nouvelles données primaires plus fiables (à travers des questionnaires par exemple).
- Le périmètre est généralement peu clair et les résultats sont présentés sans que l'on sache s'il s'agit de l'impact de l'ouverture des seules données publiques ou de l'*open data* dans son ensemble (incluant alors les données ouvertes par les entreprises).
- Les approches sont souvent trop macroéconomiques (estimation à un niveau agrégé, exprimé en part de PIB) ou trop microéconomiques (estimation à un niveau local ou dans un secteur spécifique). Le risque est alors de surestimer ou au contraire de sous-estimer la valeur de l'*open data*.
- Les études sont toujours des appréciations *ex ante* de la création de valeur associée à l'ouverture des données. Il n'y a pas encore d'évaluation *ex post* sérieuse, basée sur des données réelles.

Pour toutes ces raisons, **il y a autant de résultats différents que d'évaluations**²¹³. Entre 2000 et 2020, huit différentes études ont été menées pour évaluer la taille du marché que pourrait créer l'ouverture des données publiques dans l'Union européenne. Les auteurs se reprennent et se citent entre eux, mais obtiennent pourtant - et sans que l'on comprenne réellement comment, des résultats très éloignés. Les évaluations, exprimées en pourcentage du PIB européen, font en effet le grand écart, ne représentant pour certains qu'à peine 1% et pour d'autres près de 5% (Figure 17).

²¹³ Commission européenne, « The Economic Impact of Open Data - Opportunities for value creation in Europe », 2019, p.22. URL : data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf



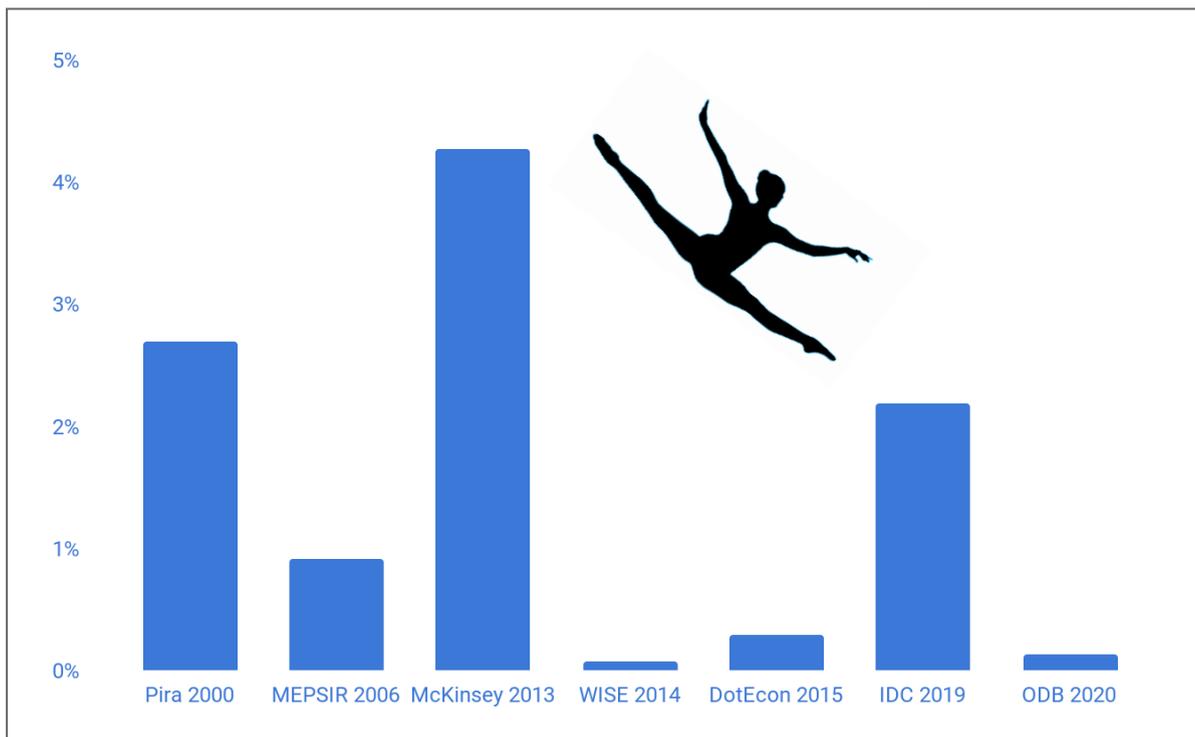


Figure 17 : Marché des données ouvertes en % du PIB européen. Source : European, 2019.

De la même façon, les études ne s'accordent pas sur les domaines stratégiques pour lesquels il faudrait ouvrir les données en priorité. La Charte de l'Open data adoptée à l'issue du sommet du G8 de 2013 liste 14 types de données à forte valeur. En 2014, la Commission européenne en identifie, quant à elle, seulement 5 types²¹⁴. Le rapport Capgemini de 2015 met en avant les 13 domaines de données qui présentent les plus fortes valeurs commerciales. Les débats persistent donc sur le caractère plus ou moins prioritaire de l'ouverture des données selon les domaines (Tableau 6).

²¹⁴ Commission européenne, « Guidelines on recommended standard licences, datasets and charging for the re-use of documents », 2014. URL : digital-strategy.ec.europa.eu/en/library/commission-notice-guidelines-recommended-standard-licences-datasets-and-charging-re-use-documents

Domaines	Exemples	CE ²¹⁵	G8 ²¹⁶	MEPSIR ²¹⁷	Cappgemini ²¹⁸
Informations géographiques	Codes postaux, cartes nationales et locales	OUI	OUI	OUI	OUI
Observation de la terre et de l'environnement	Données météo, qualité de l'eau, niveaux d'émissions, etc.	OUI	OUI	OUI	OUI
Transport	Horaires des transports publics	OUI	OUI	OUI	OUI
Statistiques	Indicateurs démographiques et économiques	OUI	OUI	OUI	NON
Entreprises	Liste des entreprises enregistrées, etc.	OUI	OUI	OUI	OUI
Justice et délinquance	Statistiques criminelles	NON	OUI	NON	NON
Education	Liste des écoles, etc.	NON	OUI	NON	OUI
Finance et contrats	Appels d'offres, budgets locaux et nationaux, etc.	NON	OUI	NON	NON
Développement mondial	Aides au développement, sécurité alimentaire, etc.	NON	OUI	NON	NON
Système légal, démocratie et politique	Contacts des administrations, résultats des élections, législation, salaires (échelles de rémunération), cadeaux	NON	OUI	OUI	OUI
Santé	Données relatives aux prescriptions médicales, aux performances des hôpitaux, etc.	NON	OUI	OUI	NON
Science et recherche	Données sur le génome, activités de recherche et d'éducation, expériences résultats	NON	OUI	NON	OUI
Mobilité sociale	Données sur le logement, assurance maladie et allocations de chômage	NON	OUI	NON	OUI
Tourisme et loisirs	Fréquentation des différents sites touristiques, nuitées, etc.	NON	NON	NON	OUI
Agriculture, foresterie, pêche	Production agricole	NON	NON	NON	OUI
Monde culturel	Pratiques artistiques et sportives	NON	NON	NON	OUI

Tableau 6 : Identification des catégories de données à haute valeur selon les différentes études. Sources : Commission européenne, 2014; Open Data Charter, 2013; MEPSIR, 2016; Cappgemini, 2015.

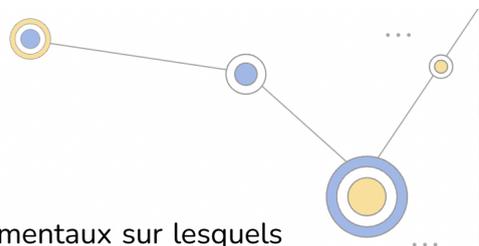
Au-delà du manque de rigueur de ces études, toute tentative d'évaluation chiffrée sérieuse des bénéfices économiques de l'*open data* n'est-elle pas condamnée d'avance ? Il est en

²¹⁵ *Ibid.*

²¹⁶ G8, « Open Data Charter », 2013. URL : assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf

²¹⁷ Marc Jacquinet, *Measuring European Public Sector Information Resources*, 2016. URL : www.researchgate.net/publication/297032105_Measuring_European_Public_Sector_Information_Resources

²¹⁸ Cappgemini Consulting, « Creating Value through data », 2015, p.32. URL : data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf



effet très difficile d'envisager pouvoir surmonter les obstacles fondamentaux sur lesquels butent les calculs²¹⁹ :

- Comment distinguer les bénéfices tirés de l'*open data* uniquement des bénéfices du traitement des données massives dans leur ensemble ?
- Comment séparer la valeur des données de la valeur de l'innovation ?
- Comment mesurer les bénéfices de l'*open data* pour le secteur public, qui est à la fois fournisseur et consommateur de services riches en données et qui rachète souvent ses propres données « enrichies » par d'autres ?

Sans réponse à ses questions, il nous semble qu'une politique d'ouverture des données fondée essentiellement sur des perspectives de croissance économique serait trop fragile. Pour cette raison, pour donner du sens à l'ouverture, réaffirmer l'objectif historique de transparence nous semble plus consistant.

3.2. Redynamiser l'ouverture en réaffirmant son objectif premier : la transparence

3.2.1. Pour une politique d'ouverture des données guidée par l'objectif de transparence et reposant sur deux piliers : anticiper les besoins et répondre aux demandes

Nous l'avons vu, la dynamique d'ouverture des données publiques peut servir des objectifs variés. À l'origine motivée par des exigences de transparence démocratique, elle s'en est peu à peu écartée pour se diriger vers des considérations beaucoup plus prosaïques de développement économique, avec son lot de promesses déçues pour les administrations. Nous pensons qu'il faut redonner du sens à la politique d'ouverture pour chaque agent public, et cela passe avant toute chose par l'affirmation d'un objectif clair : **la boussole de l'action publique doit être la transparence**. Ce n'est qu'en la suivant que l'ouverture des données publiques pourra intégrer la culture des administrations et constituer un véritable service public de la donnée.

Le cas de La Rochelle illustre concrètement ce à quoi pourrait ressembler cette politique d'ouverture renouvelée. Cette agglomération est l'une des plus en avance en matière d'*open data*. Pourtant, de l'aveu de ses propres agents, elle n'est pas capable d'appliquer pleinement le principe d'*open data* par défaut de la loi Lemaire. Incapable de tout ouvrir instantanément, elle a adopté une approche pragmatique qui s'appuie sur deux piliers :

²¹⁹ Capgemini Consulting, « Creating Value through data », 2015, p.55. URL : data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

anticiper les besoins et répondre aux demandes. Nous pensons que cette approche permettrait de réconcilier la loi et sa mise en œuvre.

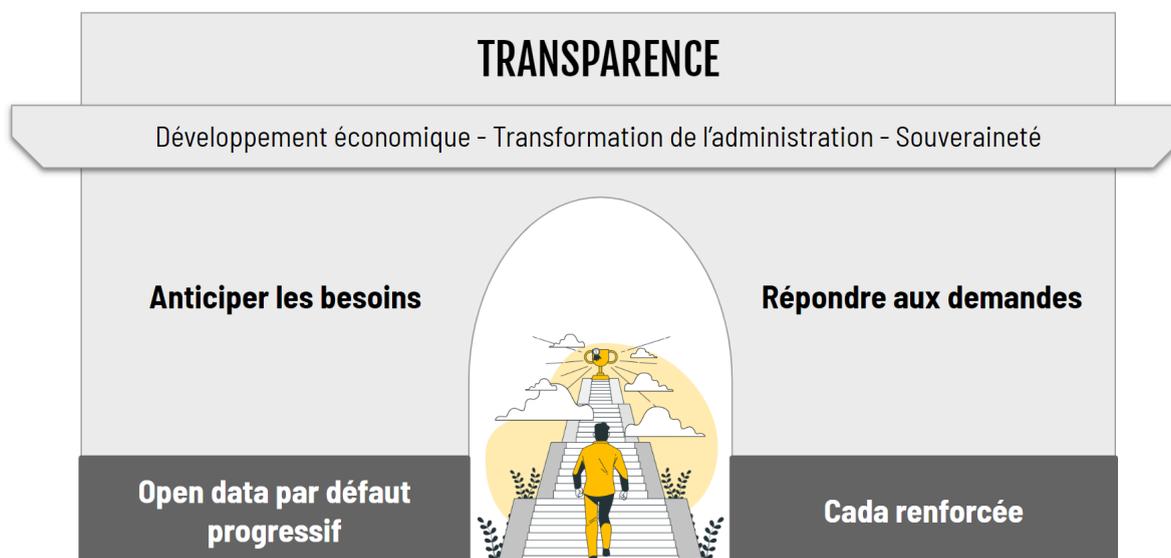


Figure 18 : L'organisation d'une nouvelle dynamique d'ouverture des données publiques.

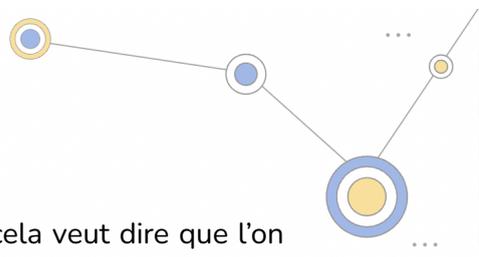
La transparence, oui, mais pour quoi faire ?

Si nous pensons que la transparence doit être réaffirmée comme l'objectif premier de la politique d'ouverture des données publiques, **il ne s'agit pas d'appliquer aveuglément ce principe**. Le professeur de droit Jean-Denis Bredin décrit les risques que peut représenter une transparence excessive²²⁰. Il montre qu'elle devient peu à peu une « Vertu supérieure » qui écrase toutes les autres, faisant la chasse aux secrets, à la discrétion et à l'intimité, accusés de faire obstacle au chemin de la vérité. Dans nos démocraties, une telle transparence s'opposerait au droit de la personne : « *La démocratie devrait-elle préserver des secrets, parce que le principe fondamental qui l'inspire serait non une loi morale mais le respect de l'homme, de sa liberté, de sa dignité et de son bonheur ?* ». Mazarine Pinget abonde dans ce sens dans un essai au titre explicite : *La Dictature de la transparence*²²¹. Elle décrit la transparence comme une idéologie qui ne se voit plus, qui est invisible, une fausse évidence dont on oublie le sens à force de la manier. Cette transparence présente un danger en donnant une illusion de savoir, de même qu'une photo ne montre qu'une part infime de la réalité quand c'est tout le hors-champ qui referme l'essentiel de l'information.

²²⁰ Jean-Denis Bredin, « Secret, transparence et démocratie », *Pouvoirs* 2001/2 (n° 97), p. 5-15.

DOI 10.3917/pouv.097.0005. URL : revue-pouvoirs.fr/secret-transparence-et-democratie/

²²¹ Mazarine Pinget, *La Dictature de la transparence*, Robert Laffont, Paris, 2016.



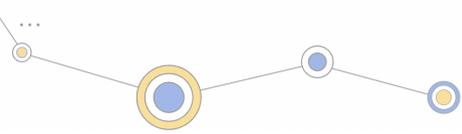
Elle est aussi une forme de terrorisme moral : si on ne montre pas, cela veut dire que l'on cache.

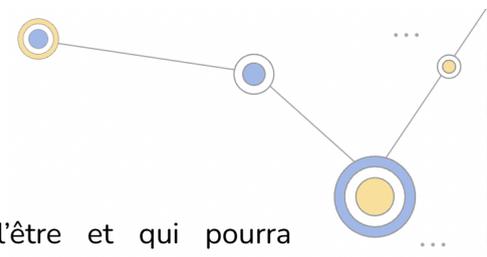
La transparence comme boussole de l'ouverture des données publiques n'est bénéfique, selon nous, que si elle sert les deux objectifs suivants. D'une part, elle doit permettre à chaque citoyen de demander des comptes à son administration conformément à l'article 15 de la Déclaration des Droits de l'Homme et du Citoyen de 1789. C'est le cas des demandes de notes de frais des élus ou des demandes effectuées dans le cadre de l'affaire McKinsey. D'autre part, la transparence doit permettre d'alimenter le débat public, comme dans l'affaire des IPS ou du financement du Rassemblement national. Ce sont ces deux finalités qui justifient l'intérêt public de l'ouverture des données publiques, c'est cette transparence qu'il faut porter tout en protégeant les secrets qui doivent l'être.

Mais ériger la transparence comme objectif principal ne dispense pas d'avoir des objectifs secondaires. En replaçant la transparence au centre de la politique d'ouverture, nous pensons lever des freins culturels, mettre fin à la déception face aux promesses économiques non tenues et redonner un sens pour les agents qui participent à l'ouverture des données. Nous pensons que cela relancera une dynamique qui servira aussi, bien sûr, l'innovation et le développement économique, la souveraineté numérique ou encore la transformation de l'administration, corollaires de l'ouverture des données publiques.

Premier pilier : anticiper les besoins

Anticiper les besoins, c'est l'administration qui ouvre spontanément les données dont qu'elle juge utiles pour les citoyens et les entreprises. Nous l'avons vu, l'*open data* au sein des administrations est à deux vitesses : si certaines ne se sont pas du tout saisies de ce sujet, d'autres sont au contraire très actives, voire proactives. C'est notamment le cas de l'agglomération de La Rochelle qui ouvre proactivement des données en fonction des projets en cours dans les écosystèmes locaux; et quand les données n'ont pas forcément vocation à être rendues publiques, les agents de l'agglomération vont quand même proposer des contrats de prêt de données alors que rien ne les oblige à le faire. Par exemple, une start-up voulait développer un outil pour savoir où les gens garent le plus souvent leur voiture et s'ils paient bien leur stationnement. Une telle application nécessitait des données de voirie non publiques que l'agglomération a prêtées dans le cadre d'un contrat *ad hoc*. Les agents de La Rochelle sont également en train de développer une plateforme appelée *Terreze* pour faciliter le partage de données sur la transition écologique, au service de l'objectif de neutralité carbone. Elle proposera notamment une matrice de croisement pour aider les réutilisateurs à savoir ce qu'ils peuvent faire ou non lorsqu'ils croisent des données couvertes par des licences différentes. C'est vers ce genre





d'initiatives qu'il faudrait tendre, en ouvrant ce qui peut l'être et qui pourra vraisemblablement être réutilisé, et en facilitant les réutilisations.

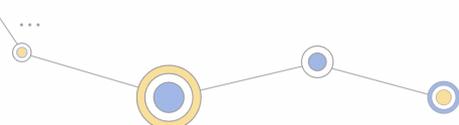
Dans cet *open data* à deux vitesses, les administrations les plus proactives pourraient aider les autres administrations à ouvrir leurs données. À l'échelle ministérielle, les AMDAC et leurs réseaux de correspondants constituent d'ores et déjà un relais pour porter la politique d'*open data* dans les ministères, que ce soit en administration centrale ou en administration déconcentrée. Au niveau des collectivités territoriales, il ne serait pas efficace d'avoir un portail *open data* par commune. C'est pourquoi la commune de La Rochelle essaie de couvrir toutes les communes de l'agglomération dans son portail *open data*. Elle fait également partie du réseau d'**Animateurs Territoriaux des Données**²²². Il s'agit d'une des composantes du programme OpenDataLocale²²³ animé par l'association OpenDataFrance et soutenu par l'ANCT, la DINUM et la Banque des Territoires. Ce programme a été créé à la suite de l'adoption de la loi Lemaire pour aider les collectivités locales à appliquer cette loi. Ce réseau de collectivités proactives et motivées produit donc des ressources pour soutenir les collectivités qui veulent se lancer dans l'*open data* : fiches, formations, outils de normalisation et de production de données, etc. Elle permet également de coordonner des actions, de mettre en commun des moyens et de partager des bonnes pratiques. Le développement d'un tel réseau doit être encouragé : ce n'est qu'avec des exemples de réussites que des collectivités pourront être encouragées à faire de l'*open data*. Des communautés de producteurs de données se constituent également autour des portails thématiques soutenus notamment par la DINUM. Si ces portails ne font pas beaucoup de visites car les réutilisateurs de données ont le réflexe d'aller sur *data.gouv.fr*, c'est parce que la stratégie adoptée par la DINUM est plutôt d'utiliser ces portails pour mobiliser les producteurs de données. Il faut donc poursuivre l'effort de développement de ces portails, notamment autour des données de haute valeur (*High Value Datasets*) identifiées par l'Union européenne.

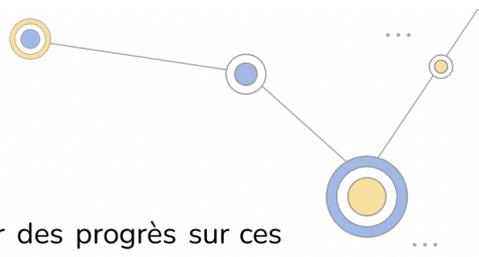
Sur le plan du financement, l'idéal serait de sanctuariser un budget pour financer des projets d'*open data* dans les différentes administrations. À défaut, le financement de l'*open data* peut passer par les grands fonds de relance ou de transformation. Par exemple, des projets de l'IGN ont bénéficié de crédits du plan « France 2030 » et du Fonds pour la transformation de l'Action publique, et les collectivités territoriales ont bénéficié d'un volet « Numérique / Data » dans le programme « Action Coeur de Ville »²²⁴ pour financer des projets de réutilisation des données publiques. Il s'agit de continuer sur cette lancée **en prévoyant systématiquement dans ce type de fonds un volet sur le développement du**

²²² opendatafrance.gitbook.io/opendatalocale/atd

²²³ opendatafrance.fr/projets/opendatalocale-maquette-open-data-des-territoires

²²⁴ opendatafrance.gitbook.io/experimentations-action-coeur-de-ville-data/contexte





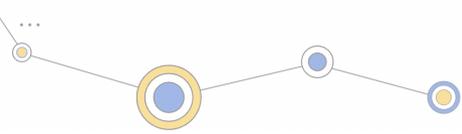
numérique et l'ouverture des données dans les administrations, car des progrès sur ces sujets contribuent à l'amélioration de toutes les autres politiques publiques.

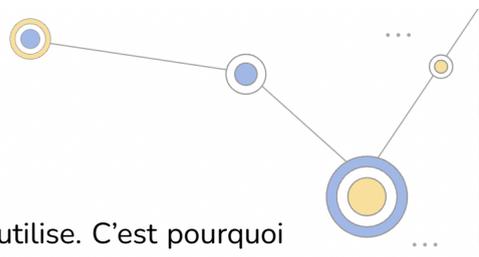
Enfin, il faut bien sûr conserver dans la loi le principe d'open data par défaut. Si les agents que nous avons rencontrés regrettent que ce principe ne soit pas bien appliqué, ils reconnaissent qu'il est très utile et qu'il a complètement changé la logique juridique autour de l'ouverture des données. La charge de la preuve repose désormais sur ceux qui ne veulent pas ouvrir leurs données et non l'inverse comme c'était le cas avant. Ainsi, il tue le débat sur l'opportunité de l'ouverture puisque la loi oblige à le faire. Cela laisse beaucoup plus de liberté aux agents sensibilisés à l'open data pour ouvrir les données qu'ils jugent utiles pour les citoyens et les entreprises. Il faut donc voir ce principe d'open data par défaut non pas comme une injonction à tout ouvrir tout de suite, mais plutôt comme un cap guidant une ouverture progressive et pragmatique.

Deuxième pilier : répondre aux demandes

Répondre aux demandes, c'est ouvrir en priorité les documents réclamés par les citoyens en vertu du droit d'accès, bien sûr lorsque la loi le permet. Il ne faut donc pas considérer ces demandes comme des attaques contre l'administration et les ignorer, mais plutôt les voir comme un guide qui indique les données qui peuvent présenter un intérêt. De même que dans une entreprise, il ne faut pas ignorer les critiques négatives des clients mais au contraire les considérer avec grand intérêt car elles sont une source importante d'amélioration. Si l'administration est réticente à ouvrir ses données car elle n'en a pas les moyens, il vaut mieux le dire au demandeur que de l'ignorer, et éventuellement voir avec lui si des délais plus longs ou une demande allégée peuvent convenir. Si l'administration est réticente car elle a peur de divulguer des secrets, elle peut demander conseil à la CADA, idéalement en tenant au courant le demandeur de cette démarche. Enfin, si l'administration est réticente car elle a peur que les documents puissent nourrir des critiques à son égard, alors il vaut mieux faire taire les critiques en s'améliorant plutôt que d'essayer de les éviter, ce qui ne peut que motiver les demandeurs à arracher ces documents des mains de l'administration et est dommageable pour la confiance entre citoyens et administrations.

Pour mieux répondre aux demandes, il est aussi souhaitable que les administrations tissent des liens avec les demandeurs de documents et les réutilisateurs de données. Un désavantage de l'open data, pointé par des nombreux producteurs de données comme le Shom ou la DILA qui ont perçu pendant longtemps des redevances, est qu'il empêche les producteurs de données de connaître le profil des réutilisateurs. Avec des redevances, il fallait s'abonner aux données et éventuellement donner quelques renseignements, les administrations savaient donc pour qui elles produisaient les données. Avec l'open data, on





verse les données sur son portail et on ne peut pas suivre qui les réutilise. C'est pourquoi certaines administrations essaient d'animer les communautés de réutilisateurs. C'est le cas de l'IGN qui appuie sa stratégie sur le concept de « géocommuns »²²⁵, des bases de données géographiques comme le projet *Panoramax* qui incluent dans leur gouvernance des réutilisateurs potentiellement externes à l'administration, par exemple des contributeurs d'OpenStreetMap. C'est aussi le cas du projet *Green Data for Health*²²⁶, plateforme de données environnementales portée par l'Ecolab du Commissariat général au développement durable, qui organise des événements et des challenges pour animer sa communauté de réutilisateurs et stimuler la valorisation de ses données. La Ville de La Rochelle avait quant à elle organisé un hackathon en 2017 autour de la réutilisation de ses données. Il lui a valu le deuxième prix dans la catégorie « Participation citoyenne » aux Prix européens de l'innovation urbaine organisés par Le Monde en 2018²²⁷. Tisser des liens entre administrations et réutilisateurs de données publics permet de créer un climat de confiance entre ces acteurs, les encourageant à collaborer autour d'un intérêt commun d'amélioration de la qualité des services publics.

Ce deuxième pilier de l'ouverture ne peut fonctionner que s'il existe **des citoyens avertis qui s'intéressent aux données de l'administration** et qui sont armés pour les demander et les exploiter. Aujourd'hui, ces « citoyens experts » sont pour beaucoup des journalistes. Pourtant, parmi les quelques journalistes que nous avons rencontrés, plusieurs n'étaient pas au courant de l'existence du droit d'accès aux documents administratifs. Cela est d'autant plus étonnant que ce droit est une source importante d'informations pour alimenter des enquêtes. Sans pour autant inciter à abuser de ce droit en faisant à tout va des demandes qui n'ont pas nécessairement d'intérêt, il nous semble essentiel que les journalistes aient *a minima* connaissance de l'existence du droit d'accès. Cela peut passer par un court module de formation en école de journalisme.

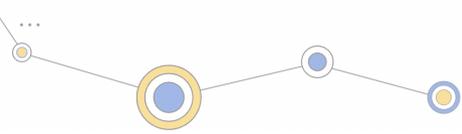
3.2.2. Pour une Cada renforcée

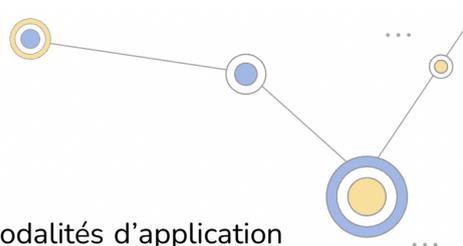
Mais cela doit aussi s'accompagner d'une Cada renforcée, capable de mieux accompagner les administrations face à une éventuelle augmentation des demandes de documents, mais aussi de veiller à ce qu'il n'y a pas d'abus, que ce soit de la part des administrations qui s'obstineraient à refuser d'ouvrir des documents publics, ou de la part de demandeurs qui accablent des administrations par des demandes abusives.

²²⁵ www.ign.fr/publications-de-l-ign/institut/evenements/synthese_consultation_publicque_geocommuns.pdf

²²⁶ gd4h.ecologie.gouv.fr/

²²⁷ Le Monde, « Palmarès 2018 des prix européens » Le Monde » Smart Cities », 2018. URL : www.lemonde.fr/smart-cities/article/2018/05/16/palmares-2018-des-prix-europeens-le-monde-smart-cities_5299809_4811534.html





Pour cela, il faut rénover les missions de la Cada. La diversité des modalités d'application du droit d'accès aux documents administratifs dans le monde nous montre que les possibilités d'évolution sont nombreuses.

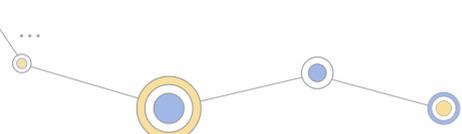
Caractéristique du droit d'accès	France	États-Unis	R.-U.	Suède	Estonie	Allemagne
Le droit d'accès aux documents administratifs est un droit constitutionnel ?	OUI	NON	NON	OUI	OUI	NON
Possibilité de faire appel devant une administration indépendante ²²⁸	OUI	NON	OUI	OUI	OUI	OUI
Cette administration a aussi pour mission la protection des données personnelles	NON	-	OUI	NON	OUI	OUI
Cette administration peut prendre des décisions juridiquement contraignantes concernant le droit d'accès	NON	-	OUI	NON	OUI	NON
Recours judiciaire possible	OUI	OUI	OUI	OUI	OUI	OUI
Obligation de désigner des PRADA (ou équivalents) dans les administrations	OUI	OUI	NON	NON	OUI	NON

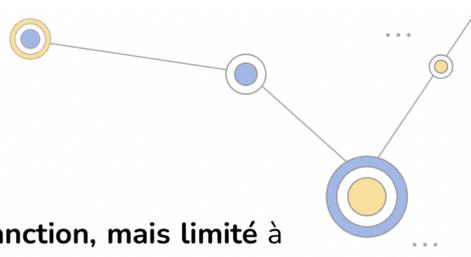
Tableau 7 : Le droit d'accès aux documents administratifs dans le monde.

Doter la Cada d'un pouvoir de sanction

Une Cada renforcée, c'est d'abord une Cada dotée d'un pouvoir de sanction. Ici, nous ne voulons pas stigmatiser les administrations. Le droit d'accès aux documents administratifs est exigeant et il faut bien sûr faire preuve de discernement : bon nombre d'administrations font de leur mieux pour communiquer les documents qui leur sont demandés, mais ne disposent pas d'assez de moyens et de temps pour accomplir toutes leurs missions. Cependant, nous pensons qu'étant données les performances médiocres du droit d'accès, il faut mener jusqu'au bout la logique de la loi pour une République numérique et prévoir **des sanctions contre les administrations qui feraient particulièrement preuve de mauvaise volonté.**

²²⁸ Cada en France, *Riksdagens Ombudsmän* en Suède (équivalent du Défenseur des droits en France), *Information Commissioner's Office* au Royaume-Uni, *Bundesbeauftragte für den Datenschutz und die Informationsfreiheit* en Allemagne, *Data Protection Inspectorate* en Estonie. Aux États-Unis, le recours se fait directement devant les tribunaux.





Il se trouve que **la Cada dispose d'ores et déjà d'un pouvoir de sanction, mais limité** à ceux qui feraient une mauvaise utilisation des données publiques²²⁹, soit en violant une obligation d'obtenir une licence, soit en ne respectant pas les termes prévus par la licence, soit en altérant les données, en dénaturant leur sens ou en ne mentionnant pas leurs sources et date de dernière mise à jour dans le cadre d'une réutilisation à des fins commerciales. Nous n'avons trouvé qu'**une seule sanction sur le site de la Cada**, datant de 2008, à l'encontre d'une entreprise du secteur l'agroalimentaire²³⁰. Celle-ci avait laissé penser, dans une campagne de publicité pour une nouvelle huile de friture, que l'Agence française de sécurité sanitaire des aliments avait émis des recommandations spécifiquement sur cette huile en extrapolant des recommandations plus générales sur les apports nutritionnels journaliers.

La Cada n'a en revanche pas de pouvoir de sanction dans le cadre de processus d'accès aux documents administratifs, ni à l'égard des demandeurs en cas de demandes abusives, ni à l'égard des administrations en cas de refus abusifs. On peut réfléchir à la pertinence d'un pouvoir de sanction dans les deux cas, dans la mesure où l'accès aux documents administratifs fait bien l'objet d'une « **double frustration** » que déplore Marc Dandelot, ancien président de la Cada : « *D'un côté, des gens veulent contrôler les finances d'une commune des quinze dernières années. De l'autre, les administrations sont excédées par ces demandes, car elles n'ont pas les moyens d'y répondre.* »²³¹

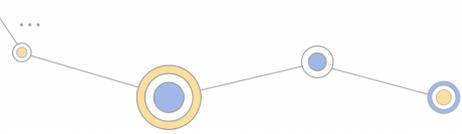
Concernant les sanctions à l'égard des demandeurs de documents, nous pensons qu'elles ne sont pas nécessaires. En effet, les administrations peuvent d'ores et déjà ignorer les demandes abusives²³². Par ailleurs, il semble difficile de pouvoir identifier formellement la personne à l'origine d'abus, dans la mesure où un simple mail suffit pour faire une demande de document. Certes, une éventuelle sanction à l'égard des demandeurs pourrait montrer aux administrations qu'on ne les accable pas et que la Cada peut aussi les protéger, mais elle pourrait aussi envoyer un signal très négatif aux demandeurs en ajoutant une dimension de risque juridique, même infime, à chaque demande. **La possibilité de sanctionner les demandeurs serait donc contre-productive.** Mais la Cada peut continuer d'accorder des délais de réponse plus longs pour les administrations qui ont peu de moyens, ou de mettre un terme aux demandes qu'elle juge abusives.

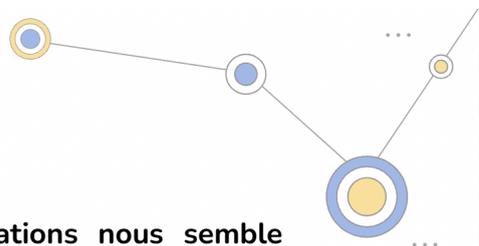
²²⁹ Article L326-1 du CRPA.

²³⁰ Avis 29983262 de la Cada.

²³¹ Maxime Ferrer, « La très difficile transparence des administrations en France », Le Monde, 2019. URL : www.lemonde.fr/les-decodeurs/article/2019/01/18/la-cada-la-transparence-au-rabais_5411293_4355770.htm

²³² Article L311-2 du CRPA.





En revanche, un pouvoir de sanction à l'égard des administrations nous semble nécessaire. C'est notamment ce que propose Raphaël Maurel, Maître de conférences en droit public, dans une tribune publiée dans le journal Le Monde en 2022²³³. Il rappelle les très mauvais chiffres publiés par la Cada sur le nombre important d'administrations qui ne suivent pas les avis de la Cada. Il souligne également que la page « Suivi des avis de la Cada »²³⁴ du site internet de la Cada demeure désespérément « en cours de construction ». C'était le cas en 2022 et c'est encore le cas en 2024. La Cada devrait être en mesure de **sanctionner les administrations qui ne la tiennent pas au courant des suites données à ses avis.** Elle pourrait également exiger que les échanges entre administrations et demandeurs soient publics, tout en les anonymisant, ou *a minima* publier les motifs invoqués par les administrations pour ne pas communiquer leurs documents. Cela permettrait de progresser en termes de transparence du processus d'accès aux documents administratifs, et ajouterait une forme de pression sur les administrations qui n'auraient pas de raisons légitimes pour ne pas suivre les avis.

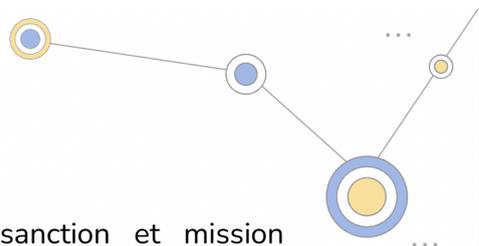
Une Cada renforcée, c'est aussi **une Cada qui rend non pas des avis consultatifs mais des décisions contraignantes.** Le parcours d'accès aux documents administratifs pourrait ainsi être raccourci et simplifié : la Cada rend un premier jugement et on peut imaginer un recours directement devant le Conseil d'État. Cette procédure serait ainsi similaire à d'autres procédures déjà existantes, par exemple l'accès au Programme de Médicalisation des Systèmes d'Information (PMSI). C'est la CNIL qui autorise l'accès aux données du PMSI, et sa décision peut être attaquée devant le Conseil d'État. Une telle procédure permet d'éviter le passage devant le tribunal administratif, ce qui permet de pousser jusqu'au bout la logique de la Cada qui était de rendre un premier avis pour limiter le recours au juge administratif. En faisant des décisions de la Cada des objets juridiques contraignants, **les administrations qui ne les respectent pas pourraient ainsi être sanctionnées.** Cela nécessite que la Cada fasse preuve de discernement envers les administrations en difficultés, mais fasse aussi preuve de fermeté envers les administrations qui refusent abusivement de communiquer leurs documents. Un refus abusif pourrait être un refus répété malgré les injonctions de la Cada, ou alors un refus qui va manifestement à l'encontre de la jurisprudence de la Cada, comme c'est le cas des communes qui persistent dans leur refus de communiquer les notes de frais des élus locaux²³⁵.

²³³ Raphaël Maurel, « Dotons la Commission d'accès aux documents administratifs d'un véritable pouvoir de sanction à l'encontre de l'administration », 2022. URL :

www.lemonde.fr/idees/article/2022/09/15/dotons-la-commission-d-acces-aux-documents-administratifs-d-un-veritable-pouvoir-de-sanction-a-l-encontre-de-l-administration_6141667_3232.html

²³⁴ www.cada.fr/lacada/suivi-des-avis-de-la-cada

²³⁵ Sur 1565 demandes de notes de frais créées depuis le 8 février 2013, seules 352 (22%) ont été suivies de la communication des documents.



À l'objection que l'on ne pourrait pas concilier pouvoir de sanction et mission d'accompagnement ou de conseil, on peut rétorquer que **plusieurs administrations cumulent de telles missions sans que cela ne pose de problème**. C'est par exemple le cas de la CNIL dont les missions comportent à la fois un volet « Accompagner / conseiller » et un volet « Contrôler et sanctionner »²³⁶. Elle propose ainsi sur son site un grand nombre de ressources pour aider tous ceux qui manipulent des données personnelles à se mettre en conformité avec le RGPD, par exemple avec un guide de la sécurité des données personnelles²³⁷ ou encore un guide de sensibilisation au RGPD pour les associations²³⁸. Son collège peut siéger en séance plénière ou en formation restreinte, ce qui lui permet dans un cas de rendre des avis, de prendre des décisions, d'analyser l'évolution des technologies, et dans l'autre de prononcer des sanctions à l'écart de ceux qui enfreignent la loi. On pourrait citer de nombreuses autres autorités indépendantes qui cumulent ces deux missions : l'Arcom, la HATVP, l'Arcep, etc.

Renforcer l'accompagnement des administrations par la Cada

Il s'agit aussi de mieux accompagner les administrations dans l'ouverture de leurs données. Cela passe d'abord par une clarification et une consolidation du corpus juridique. Nous l'avons vu, les principes généraux inscrits dans la loi ne permettent pas de trancher de nombreux cas d'ouverture litigieux, mais les avis de la Cada et la jurisprudence de la justice administrative permettent souvent d'éclairer les administrations sur les actions à mener. Il est donc essentiel que **ce corpus juridique soit facilement accessible, facilement compréhensible et consolidé** pour permettre à quiconque de comprendre rapidement si des données peuvent être ouvertes et dans quelles modalités. La Cada propose ainsi en ligne des fiches thématiques qui expliquent la jurisprudence autour de certains sujets²³⁹ : environnement, budget et comptes des collectivités, documents électoraux, etc.

Cet effort de clarification du droit doit être poursuivi, en essayant d'être le plus concret possible. Ainsi, certains guides proposés par la Cada sont exhaustifs quant à la présentation du cadre légal, mais ne permettent pas forcément d'agir sur des cas concrets. C'est notamment le cas de son guide à destination des PRADA²⁴⁰ qui se borne à énoncer les différents textes de loi sans les illustrer par des exemples concrets ou par une explication de la jurisprudence dans les cas les plus courants. Par ailleurs, il présente l'ouverture des données comme une décision algorithmique et complètement déterministe. Mais c'est oublier que dans de nombreux cas, la communication et ses modalités ne

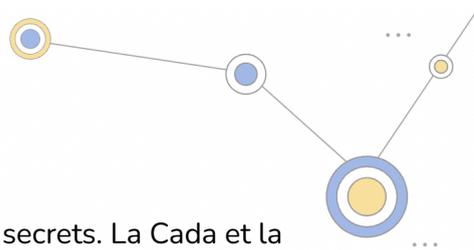
²³⁶ www.cnil.fr/fr/les-missions

²³⁷ www.cnil.fr/fr/guide-de-la-securite-des-donnees-personnelles

²³⁸ www.cnil.fr/sites/cnil/files/atoms/files/cnil-guide_association.pdf

²³⁹ www.cada.fr/administration/les-fiches-thematiques

²⁴⁰ www.cada.fr/sites/default/files/guide_des_prada.pdf



relèvent pas de l'évidence, notamment quand il faut occulter certains secrets. La Cada et la CNIL ont également établi conjointement un guide pratique de la publication en ligne et de la réutilisation des données publiques²⁴¹ qui manque lui aussi d'exemples concrets. La CNIL a en revanche présenté son propre projet de guide pratique²⁴² qui s'appuie sur des exemples très concrets et propose des fiches pratiques pour certains cas d'usage. Ce guide a vocation à évoluer en fonction des cas d'usage qui apparaissent. **La Cada pourrait s'inspirer d'un tel guide en éclairant les cas d'usage les plus fréquents dans ses fiches thématiques et en les illustrant avec des exemples très concrets.**

Enfin, **la Cada pourrait sensibiliser les administrations au-delà des seuls PRADA** pour diffuser plus largement la connaissance du droit d'accès aux documents administratifs, de l'*open data* et une certaine culture de la transparence. Cela devrait se faire *a minima* dans la formation initiale et/ou continue des agents et cadres de la fonction publique et pourrait inclure la participation d'autres acteurs comme la DINUM ou l'ANCT. Il ne s'agit pas ici de faire de chaque agent un expert, ni de demander à ouvrir sans discernement toutes les données publiques. Il s'agirait plutôt de **montrer aux agents que l'ouverture des données peut aussi les aider dans leurs missions quotidiennes, en illustrant ces bénéfices dans des cas très concrets.** Nous en avons vu quelques-uns : retours constructifs des réutilisateurs pour améliorer la qualité des données comme c'est le cas sur *data.gouv.fr*, mise en valeur avec le baromètre de l'action publique, échanges facilités des données entre administrations en passant directement par les portails *open data*, etc. Il faut aussi montrer que répondre à des demandes de documents renforce la confiance des citoyens. Une telle sensibilisation aurait donc pour but de convaincre chaque agent et chaque cadre qu'il ne faut pas prendre une posture défensive face à une demande de document et que **communiquer aux citoyens les données dont ils ont besoin fait aussi partie, pour toute administration, de ses missions élémentaires de service public.**

Une Cada ainsi renforcée, qui accompagne mieux et qui peut sanctionner, c'est aussi **une Cada avec des moyens supplémentaires pour instruire les dossiers.** Certains, notamment Axelle Lemaire en 2015 lors de l'élaboration de sa loi²⁴³, ont même proposé de fusionner la Cada et la CNIL pour les renforcer. L'argument principal est que les missions de ces deux institutions allaient de plus en plus se recouper, comme nous avons pu le voir sur le sujet des données à caractère personnel. Par ailleurs, cela se fait dans de nombreux pays comme au Royaume-Uni, en Allemagne ou en Estonie. Mais nous pensons que cela n'est pas nécessaire. Ces deux organismes ont des missions qui demeurent assez différentes, et pour

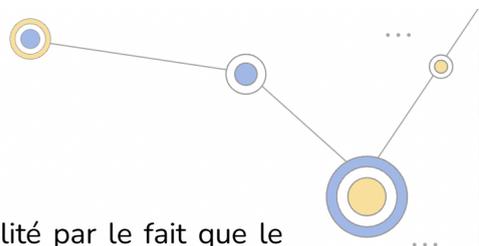
²⁴¹ www.cada.fr/sites/default/files/guide_pratique.pdf

²⁴² www.cnil.fr/sites/cnil/files/2023-08/projet_de_guide_ouverture_partage_et_reutilisation_de_donnees.pdf

²⁴³

www.lemonde.fr/pixels/article/2015/10/09/le-gouvernement-veut-fusionner-la-cnil-et-la-cada_4786269_4408996.html

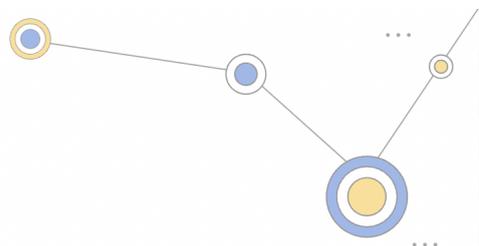




les sujets communs elles collaborent d'ores et déjà, ce qui est facilité par le fait que le président de la Cada siège de droit au Collège de la CNIL et réciproquement. Selon nous, il suffirait, dans un premier temps, de **renforcer les moyens de la Cada pour lui permettre d'accomplir pleinement ses missions**. Aujourd'hui, la Cada a un plafond d'emplois d'environ 20 ETP et un budget de près de 2 millions d'euros pour traiter plus de 10 000 saisines par an²⁴⁴. Même si les activités ne sont pas les mêmes, on peut comparer ces moyens à ceux de la CNIL qui sont sans commune mesure : 288 emplois et 26 millions d'euros pour traiter près de 16 000 plaintes, effectuer 340 contrôles et prononcer 42 sanctions en 2023²⁴⁵.

²⁴⁴ Cada, « Rapport d'activité 2023 ». URL : www.cada.fr/sites/default/files/CADA_RAPPORT_2022_2023.pdf

²⁴⁵ CNIL, « Rapport d'activité 2023 ». URL www.cnil.fr/fr/la-cnil-publie-son-rapport-annuel-2023



Conclusion

Dans ce travail, nous avons voulu analyser la politique française d'ouverture des données publiques, instaurée par la loi pour une République numérique de 2016, et examiner la pertinence de l'obligation de publication par défaut des données publiques. Il s'agissait de comprendre pourquoi cette obligation, malgré son ambition, n'est pas encore pleinement intégrée dans les pratiques des administrations françaises.

Nous avons montré que, bien que la France occupe une position de leader international en matière d'*open data*, l'application de la loi reste inégale. La publication par défaut des données publiques est confrontée à des obstacles significatifs qui en limitent l'efficacité. Les administrations rencontrent ainsi de nombreuses difficultés, telles que des coûts importants de structuration et de publication des données, une culture administrative encore marquée par le secret, et des incertitudes juridiques persistantes. De plus, l'approche actuelle, principalement axée sur le potentiel économique des données ouvertes, semble trop incertaine et lointaine pour mobiliser durablement les administrations.

Une solution réside dans la réaffirmation de l'objectif premier de transparence, plutôt que la seule valorisation économique. En encourageant les initiatives de publication de données et en renforçant le droit d'accès aux documents administratifs, il est possible de rendre cette politique plus opérationnelle et effective. Cela s'appuie en particulier sur un principe d'*open data* par défaut pragmatique et progressif, et sur une CADA renforcée, capable à la fois de mieux accompagner et de sanctionner.

Enfin, l'ouverture des données publiques est un chantier en constante évolution. À cet égard, il pourrait être intéressant dans de prochains travaux de comparer les politiques et législations de différents pays pour identifier les meilleures pratiques et l'apport éventuel de technologies émergentes, comme l'intelligence artificielle, pour faciliter l'ouverture et l'exploitation des données publiques.

Si l'enjeu aujourd'hui est de diffuser cette politique de la donnée rénovée dans toutes les administrations, les organisations les plus avancées nous permettent d'entrevoir l'étape suivante. À l'extérieur de l'administration, c'est une gouvernance partagée de certaines données, produites et partagées avec d'autres acteurs (citoyens ou entreprises) dans le cadre de communs numériques. Au sein des administrations, c'est un désilotage et un meilleur partage des données de manière générale, et pas uniquement des données publiques. L'ouverture des données publiques amorce ces transformations qui participent à l'amélioration générale des services publics.

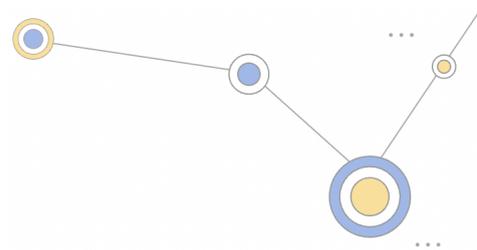


Table des figures

Figure 1 : Le catalogue des jeux de données sur data.gouv.fr, 2024.....	11
Figure 2 : Exemples de réutilisations de données publiques sur data.gouv.fr.....	12
Figure 3 : Schéma des délais pour une demande d'accès à un document administratif.....	19
Figure 4 : Carte des inégalités scolaires.....	25
Figure 5 : Enchaînement des mécanismes qui aboutissent à l'augmentation de la taille du marché grâce à l'ouverture de données publiques selon la méthode « Logique des avantages » conçue par Capgemini.....	35
Figure 6 : Erreurs relevées par les réutilisateurs sur la discussion du jeu de données « Élections municipales 2020 - Liste des candidats élus au T1 et liste des communes entièrement pourvues » publiées par le ministère de l'Intérieur sur data.gouv.fr.....	38
Figure 7 : Visualisation de la répartition par secteur des aides publiques au développement.	43
Figure 8 : Outil de construction de visualisations cartographiques et graphiques du ministère de l'Économie et des Finances.....	43
Figure 9 : Suivi des réformes prioritaires, Baromètre des résultats de l'action publique de la préfecture du Morbihan, juin 2021.....	44
Figure 10 : Distribution des offres d'emplois publics qui mentionnent les termes « données », « data » ou « RGPD » entre les fonctions publiques d'État, territoriale et hospitalière et les catégories A et B en 2024.....	53
Figure 11 : Exemples de licences Creative Commons, repérables par des sigles.....	61
Figure 12 : Distribution du nombre total de téléchargements des jeux de données de la ville de Paris publiés entre 2010 et 2024.....	64
Figure 13 : Article du Guardian réclamant le partage des données que les administrations produisent avec de l'argent issu des impôts.....	75
Figure 14 : Page d'accueil du site Internet NosDéputés.fr créé et alimenté par l'association Regards Citoyens.....	77
Figure 15 : Plateforme de la consultation en ligne organisée dans le cadre de l'élaboration de la loi pour une République numérique.....	79
Figure 16 : Capture d'écran Légifrance des premières en-têtes de la loi pour une République numérique.....	82
Figure 17 : Marché des données ouvertes en % du PIB européen.....	84
Figure 18 : L'organisation d'une nouvelle dynamique d'ouverture des données publiques..	87



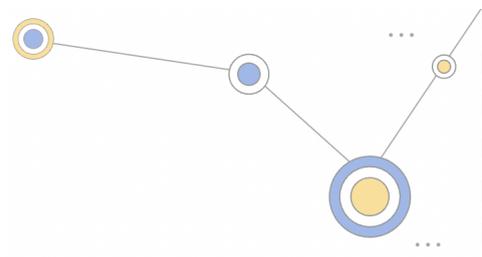
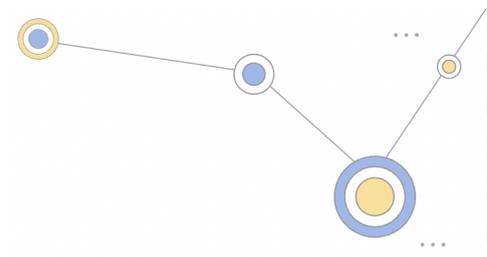


Table des tableaux

Tableau 1 : Quelques indicateurs de la mise en œuvre de l'ouverture des données publiques en France.....	18
Tableau 2 : Comparaison du classement des pays selon leur indice open data et leur indice démocratique.....	27
Tableau 3 : Données politiques fondamentales.....	28
Tableau 4 : Pourcentage des gouvernements qui publient des données politiquement sensibles et score de la France pour ces mêmes données.....	28
Tableau 5 : Recettes des redevances sur la réutilisation des données publiques et part des acheteurs publics.....	58
Tableau 6 : Identification des catégories de données à haute valeur selon les différentes études.....	85
Tableau 7 : Le droit d'accès aux documents administratifs dans le monde.....	92





Bibliographie

Adnène Trojette, « Ouverture des données publiques - Les exceptions au principe de gratuité sont-elles légitimes ? », 2013. URL : www.numerique.gouv.fr/uploads/20131105-rapporttrojetteannexes.pdf

Agence du Patrimoine Immatériel de l'État, « Rapport d'activité 2016 : Le patrimoine immatériel public, une ressource pour moderniser et créer de la valeur », 2016. URL : www.economie.gouv.fr/files/files/directions_services/apie/APIE/Publications/Rapport_activite_APIE_2016.pdf?v=1561651393

Artur Kovalchuk, Viktor Khanzhyn, Yaroslav Kudlatskiy, « Economic potential of Open data for Ukraine », 2018. URL : tapas.org.ua/wp-content/uploads/2019/02/TAPAS_PDF_1.pdf

Barack Obama, « Memorandum on Transparency and Open Government », 2009. URL : www.archives.gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf

Cada, « Rapport d'activité 2022 - 2023 ». URL : www.cada.fr/sites/default/files/CADA_RAPPORT-2022-2023-UK-vf.pdf

Capgemini Consulting, « Creating Value through data », 2015. URL : data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

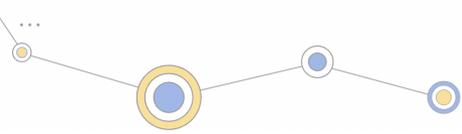
CNIL, « Rapport d'activité 2023 ». URL : www.cnil.fr/fr/la-cnil-publie-son-rapport-annuel-2023

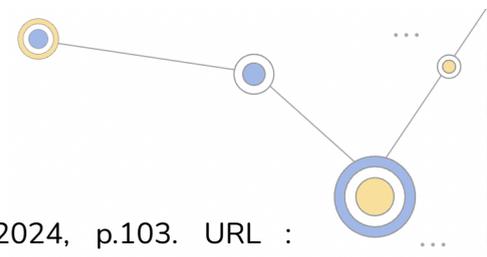
CNIL, « Rapport de résultats : Consultation Open data et données personnelles », 2014. URL : www.data.gouv.fr/storage/f/2014-04-16T09-44-53/rapport-questionnaire-cnil-open-data-et-donnees-personnelles-vdef.pdf

Commission européenne, « Open Data Maturity Report », 2023. URL : data.europa.eu/sites/default/files/odm2023_report.pdf

Commission européenne, « Stratégie européenne pour les données », 2020. URL : commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr

Commission européenne, « The Economic Impact of Open Data - Opportunities for value creation in Europe », 2019. URL : data.europa.eu/sites/default/files/the-economic-impact-of-open-data.pdf





Commission IA, « IA : Notre ambition pour la France », 2024, p.103. URL : www.info.gouv.fr/upload/media/content/0001/09/4d3cc456dd2f5b9d79ee75feea63b47f10d75158.pdf

Communiqué « Anti Corruption Summit: London 2016 ». URL : www.unodc.org/documents/treaties/UNCAC/WorkingGroups/ImplementationReviewGroup/20-24June2016/V1603744e.pdf

Conseil national du numérique, « Ambition numérique : pour une politique française et européenne de la transition numérique », 2015. URL : cnumerique.fr/files/2017-10/CNNum--rapport-ambition-numerique.pdf

Danny Vinik, « What happened to Trump's war on data? », in Politico, 2017. URL : www.politico.com/agenda/story/2017/07/25/what-happened-trump-war-data-000481/

Deloitte, « Assessing the value of Tfl's open data and digital partnerships », 2017. URL : content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf

DINUM, « Cadre interministériel d'administration de la donnée », 2021. URL : www.numerique.gouv.fr/uploads/Cadre-interministeriel-donnee.pdf

DINUM, « Une stratégie numérique au service de l'efficacité de l'action publique - Feuille de route de la DINUM », 2023. URL : www.numerique.gouv.fr/uploads/Feuille-de-route-DINUM.pdf

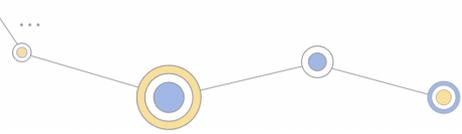
Éric Bothorel, Rapport de mission « Pour une politique publique de la donnée », 2020. URL : www.ouvrirlascience.fr/wp-content/uploads/2021/02/Mission-Bothorel_rapport-pour-une-politique-publique-de-la-donnee_23.12.2020.pdf

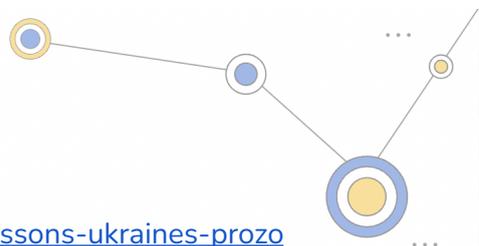
G8, « Open Data Charter », 2013. URL : assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf

Gilbert Claude et Emmanuel Henry, « La définition des problèmes publics : entre publicité et discrétion », Revue française de sociologie, vol. 53, no. 1, 2012, pp. 35-59. URL : www.cairn.info/revue-francaise-de-sociologie-1-2012-1-page-35.htm?contenu=article

Harlan Yu et David G. Robinson, « The New Ambiguity of 'Open' Government », UCLA Law Review Discourse, 2012. URL : papers.ssrn.com/sol3/papers.cfm?abstract_id=2012489

Harvard Kennedy School, « Overcoming Corruption and War: Lessons from Ukraine's ProZorro Procurement System », 2022. URL :





www.hks.harvard.edu/publications/overcoming-corruption-and-war-lessons-ukraines-prozorro-procurement-system

Haute Autorité à la Transparence de la Vie Publique, « Open data & intégrité publique - les technologies numériques au service d'une démocratie exemplaire », 2016. URL : hatvp.fr/wordpress/wp-content/uploads/2016/12/Open-data-integrite-publique.pdf.

Heli Koski, « The Impact of open data – a preliminary study », 2015. URL : vm.fi/documents/10623/1107406/The+Impact+of+Open+Data/1c432b3a-a5e8-41ea-a5ea-135280a69ea3?version=1.0

James Manyika et al., « Open data: Unlocking innovation and performance with liquid information », McKinsey, 2013. URL : www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx

Jean-Denis Bredin, « Secret, transparence et démocratie », *Pouvoirs* 2001/2 (n° 97), p. 5-15.

Jérôme Denis et Samuel Goëta, « La fabrique des données brutes », in *Ouvrir, partager, réutiliser* édité par Clément Mabi et al., Éditions de la Maison des sciences de l'homme, 2017. URL : doi.org/10.4000/books.editionsmsmh.9050.

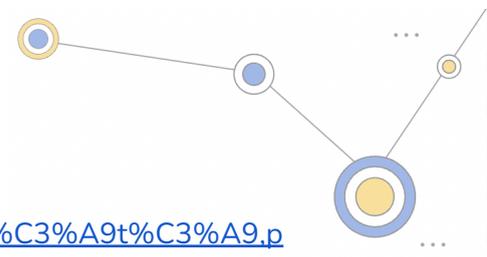
Joel Gurin, *Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation*, 2014.

Josh Guerstein, « Fears rise of Trump-era 'memory hole' in federal data », in *Politico*, 2016. URL : www.politico.com/story/2016/12/trump-federal-data-fears-232591

Lateral Economic pour l'Open Data Institute, « Permission granted: The economic value of data assets under alternative policy regimes », 2016. URL : theodi.org/insights/reports/research-the-economic-value-of-open-versus-paid-data/

Luc Martinon et al., « Recours aux consultants privés : la grande opacité de l'État », *Le Monde*, 2022. URL : www.lemonde.fr/les-decodeurs/article/2022/03/25/recours-aux-consultants-privés-la-grande-opacité-de-l-État_6119189_4355770.html

Manon Romain et al., « Au lycée, de nouvelles données révèlent l'ampleur du « tri social » entre les voies générale et professionnelle », *Le Monde*, 11 janvier 2023. URL : www.lemonde.fr/les-decodeurs/article/2023/01/11/au-lycée-les-indices-de-position-social-e-revelent-l-ampleur-du-tri-social-entre-les-voies-générale-et-professionnelle_6157366



[4355770.html#:~:text=L'%C3%A9ducation%20nationale%20a%20%C3%A9t%C3%A9,publicier%20les%20IPS%20des%20%C3%A9l%C3%A8ves.](#)

Marc Jacquinet, Measuring European Public Sector Information Resources, 2016. URL : www.researchgate.net/publication/297032105_Measuring_European_Public_Sector_Information_Resources

Martin Untersinger, « Le gouvernement veut fusionner la CNIL et la CADA », *Le Monde*, 9 octobre 2015. URL : www.lemonde.fr/pixels/article/2015/10/09/le-gouvernement-veut-fusionner-la-cnil-et-la-cada_4786269_4408996.html

Maurice Lévy et Jean-Pierre Jouyet, « L'Économie de l'immatériel : la croissance de demain », novembre 2006.. URL : <https://www.astrid-online.it/static/upload/protected/Rapp/Rapport-Jouyet-Levy---Economie-de-l-Immat-riel.pdf>

Maxime Ferrer, « La très difficile transparence des administrations en France », *Le Monde*, 2019. URL : www.lemonde.fr/les-decodeurs/article/2019/01/18/la-cada-la-transparence-au-rabais_5411293_4355770.html

Mazarine Pinget, *La Dictature de la transparence*, Robert Laffont, Paris, 2016.

N. Heller, « Is Open Data a Good Idea for the Open Government Partnership? », *Global Integrity blog*, 15 septembre 2015.

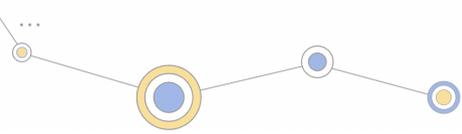
OCDE, « Digital Broadband Content : Public Sector Information », Documents de travail de l'OCDE sur l'économie numérique, n° 112, 2006.

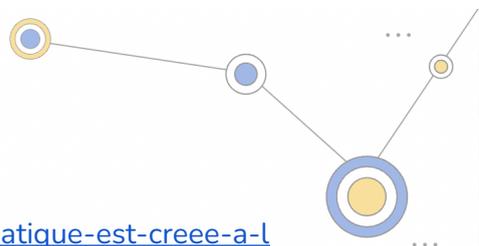
OCDE, « Open, Useful and Re-usable data (OURdata) Index », 2023. URL : www.oecd-ilibrary.org/docserver/a37f51c3-en.pdf?expires=1716483242&id=id&acname=guest&checksum=4F61536C5E771A4EED2BEA9F0AD85653

OpenDataFrance, « Observatoire open data des Territoires », 2022. URL : www.observatoire-opendata.fr/wp-content/uploads/2023/03/Copie-de-Observatoire_Open_Data_Territoires_Edition2022.pptx-1.pdf

Paul Edwards et al., « Science friction: Data, metadata, and collaboration », *Social Studies of Science*, 41(5), 2011, p.667-690.

Philippe Boucher, « Une division de l'informatique est créée à la chancellerie " Safari " ou la chasse aux Français », *Le Monde*, 1974. URL :





www.lemonde.fr/archives/article/1974/03/21/une-division-de-l-informatique-est-creee-a-l-a-chancellerie-safari-ou-la-chasse-aux-francais_3086610_1819218.html

Raphaël Maurel, « Dotons la Commission d'accès aux documents administratifs d'un véritable pouvoir de sanction à l'encontre de l'administration », *Le Monde*, 15 septembre 2022. URL :

www.lemonde.fr/idees/article/2022/09/15/dotons-la-commission-d-acces-aux-documents-administratifs-d-un-veritable-pouvoir-de-sanction-a-l-encontre-de-l-administration_6141667_3232.html

Raphaëlle Aubert et Léa Sanchez, « Comment Opendatasoft est devenue l'acteur incontournable de l'ouverture des données publiques », *Le Monde*, 2024. URL : www.lemonde.fr/les-decodeurs/article/2024/03/05/comment-opendatasoft-est-devenue-l-acteur-incontournable-de-l-ouverture-des-donnees-publiques_6220195_4355770.html

Samuel Goëta, *Les Données de la Démocratie*, 2024, p.108.

Sénat, Rapport de commission d'enquête « Un phénomène tentaculaire : l'influence croissante des cabinets de conseil sur les politiques publiques », n° 578 (2021-2022), tome I, déposé le 16 mars 2022.

Violaine Morin, « L'éducation nationale condamnée à révéler l'indice de position sociale des collèges et des CM2 », *Le Monde*, 15 juillet 2022. URL : www.lemonde.fr/societe/article/2022/07/15/l-education-nationale-condamnee-a-reveler-l-indice-de-position-sociale-des-colleges-et-des-cm2_6134916_3224.html

Yves Renaud, « De la contestation à la concertation », *Annales de la recherche urbaine*, n° 89, 2001, p. 62-69. URL : www.persee.fr/doc/aru_0180-930x_2001_num_89_1_2380.

ANNEXE : Synthèse des propositions

Propositions	Réponse au diagnostic
Formation des agents publics à l' <i>open data</i> et au droit d'accès aux documents administratifs, au service de la transparence.	Redonner du sens, encourager à la proactivité et diffuser une culture de la donnée.
Former les journalistes au droit d'accès aux documents administratifs.	Créer un intérêt de la société civile pour les données publiques et alimenter la confiance dans l'administration.
Une Cada renforcée avec plus de moyens d'accompagnement.	Clarifier la jurisprudence notamment sur la protection des secrets, protéger les administrations face aux demandes abusives.
Une Cada dotée d'un pouvoir de sanction.	Réduire les délais de réponse aux demandes de documents et sanctionner les administrations qui refusent d'ouvrir leurs données sans motif légitime.
Favoriser l'animation des communautés de réutilisateurs des données publiques.	Créer un lien de confiance et de coconstruction entre les administrations et les réutilisateurs, fédérer ces acteurs autour d'intérêts communs.
Soutenir et développer le réseau des animateurs Territoriaux des Données.	Augmenter le nombre de collectivités en <i>open data</i> , mettre en commun les moyens, diffuser une culture de la donnée.
Poursuivre le développement de portails thématiques.	Donner de la visibilité aux données et animer la communauté des producteurs de données.
Intégrer dans les grands fonds publics un volet numérique et ouverture des données.	Financer des projets d'ouverture ou de réutilisation des données publiques, en soutien des autres politiques publiques.