

Samuel Goëta

Les données de la démocratie

Open data, pouvoirs et contre-pouvoirs



C&F éditions

Dans la collection Société numérique

Adrian Daub

La pensée selon la tech. Le paysage intellectuel de la Silicon Valley.

Traduit de l'anglais (États-Unis) par Anne Lemoine

2022, ISBN 978-2-37662-034-1

Tim Hwang

Le grand krach de l'attention. La publicité, une bombe au cœur de l'internet.

Traduit de l'anglais (États-Unis) par Anne Lemoine

2022, ISBN 978-2-37662-035-8

Simone Pieranni

Red Mirror. L'avenir s'écrit en Chine.

Traduit de l'italien par Fausto Giudice

2020, ISBN 978-2-37662-021-1

Christophe Masutti

Affaires privées. Aux sources du capitalisme de surveillance.

2020, ISBN 978-2-37662-004-4

Zeynep Tufekci

Twitter & les gaz lacrymogènes. Forces et fragilités de la contestation connectée.

Traduit de l'anglais (États-Unis) par Anne Lemoine

2019, ISBN 978-2-915825-95-4

Catalogue complet : <https://cfeditions.com>

Ouvrage publié sous licence édition équitable

<https://edition-equitable.org>.

Texte diffusé sous licence Creative Commons BY-SA

<https://creativecommons.org/licenses/by-sa/4.0/deed.fr>.

ISBN 978-2-37662-071-6

Collection **Société numérique** – ISSN 2647-1493

C&F éditions, février 2024

35 C rue des Rosiers, 14000 Caen.

Samuel Goëta

Les données de la démocratie

Open data, pouvoirs et contre-pouvoirs

Préface d'Axelle Lemaire

Société numérique

C&F éditions

Table des matières

Préface d’Axelle Lemaire	
Open data : l’entrée dans la maturité	7
Introduction	17
I. Naissance et émergence de l’open data	27
Les racines multiples de l’open data	28
Les grandes étapes de la définition des principes de l’open data	59
Le long chemin en France vers l’ouverture des données par défaut	81
La pandémie de Covid-19, un moment charnière	109
II. Un bilan critique de l’open data	131
Un renouvellement de la transparence en demi-teinte	132
Des promesses démesurées de croissance et d’innovation	157
Une transformation des administrations en cours	177
III. Alternatives et propositions	193
Cinq stratégies alternatives pour obtenir des données	194
Chantiers et pistes d’action pour réduire les asymétries d’information et renforcer la démocratie	221
Conclusion	255

Open data : l'entrée dans la maturité

Préface par Axelle Lemaire

Le premier quart du ^{xxi}e siècle touche bientôt à sa fin. Jamais dans l'histoire de l'humanité la technologie n'a joué aussi rapidement un rôle aussi déterminant. En l'espace de 25 ans, des géants numériques sont apparus et des pratiques nouvelles qui ont bouleversé les pyramides économiques, les pouvoirs politiques, l'ordre géostratégique, et la vie sociale et intime de plusieurs milliards d'êtres humains. Le livre de Samuel Goëta paraît à un moment bien particulier, celui de l'entrée dans la maturité numérique : les enjeux sont désormais mieux appréhendés, le jeu des acteurs et les défis mieux cernés. Et si la vitesse des développements techniques ne ralentit pas, celle-ci croise une autre course contre la montre, cette fois réglementaire. *L'EU Artificial Intelligence Act*, ou *Règlement européen sur l'intelligence artificielle*, est intensément discuté à Bruxelles à l'heure où ces lignes sont écrites.

Dans le même temps, les fondements des régimes démocratiques sont secoués : la crise climatique révèle l'impuissance décisionnelle pour limiter l'effondrement du vivant ; les sources de tensions sociales se multiplient ; la désillusion politique s'installe ; les risques de manipulation informationnelle et les tentations autoritaristes se renforcent. Alors que les écrits se multiplient pour signer la fin de

l'utopie numérique, c'est dans ce contexte que l'auteur choisit de défendre l'ouverture des données publiques comme un mécanisme apte à revitaliser la démocratie.

La politique d'*open data*, ou ouverture des données publiques, a servi à alimenter la machine numérique, en fournissant la matière première nécessaire à l'entrée dans l'ère informationnelle dématérialisée, au partage et à l'analyse quantitative et qualitative de données, à l'alimentation des modèles algorithmiques. J'ai activement contribué à cette politique, par le truchement du mandat gouvernemental qui m'avait été confié en tant que secrétaire d'État chargée du numérique et de l'innovation entre 2014 et 2017. La loi pour une République numérique promulguée le 7 octobre 2016 a instauré l'obligation, pour les administrations de plus de 50 agents et les collectivités territoriales de plus de 3 500 habitants, de rendre accessibles en ligne les données dont la publication présente un intérêt économique, social, sanitaire ou environnemental. La loi pousse également la dimension d'ouverture vers d'autres rivages, ou d'autres branches composant l'arbre de la transparence de l'action publique : la transparence des algorithmes publics, l'accès aux publications scientifiques (la science ouverte), le développement des logiciels libres, l'accès des chercheurs aux bases de données administratives et l'appariement de données dans des conditions encadrées, le *text and data mining*, l'échanges de données entre administrations...

Dès lors, le livre de Samuel Goëta est venu m'interpeller, et me rappeler les questions que je m'étais posées à l'époque de l'élaboration de la loi. Pourquoi une ministre choisit-elle un combat plutôt qu'un autre? Pourquoi cette thématique a-t-elle résonné en moi au regard de mon histoire personnelle et de mes engagements politiques? Les choix assumés à l'époque ont-ils été les bons? Comment la suite devrait-elle se construire? L'entrée en matière de l'auteur suit d'ailleurs une démarche identique à celle qui fut la mienne : il commence par le récit personnel de sa découverte du sujet qui, depuis un séjour aux États-Unis, lui a fait comprendre au sein de quel mouvement de pensée son action de chercheur et d'entrepreneur s'inscrivait.

D'où l'envie de partager ici mon propre témoignage : je n'aurais peut-être pas été élue députée pour représenter les Français établis en Europe du Nord sans l'ouverture des données électorales opérée par plusieurs des pays de la circonscription. Pour mener campagne, avec les jeunes geeks qui me prêtaient main-forte (l'un est aujourd'hui lui-même député), nous avons décortiqué les résultats des bureaux de vote de Londres, Leeds, Edinburgh, Copenhague ou encore Stockholm pour comprendre les tendances politiques des électeurs et cibler nos actions de campagne. Ainsi ai-je découvert la puissance de la donnée lorsque celle-ci est bien comprise et bien utilisée. Auparavant, comme chercheuse en droit international des droits humains à University College London, je m'étais cassé les dents sur les informations classifiées par le FBI, en essayant de mettre à nu la réalité des traitements de torture infligés aux prisonniers de Guantánamo Bay. Et comme secrétaire nationale aux droits humains du parti politique auquel j'appartenais, je me désolais de constater que les débats relatifs à l'immigration étaient exclusivement guidés par l'émotion et la peur, sans être objectivés par la réalité des flux migratoires en France et le calcul de l'impact économique et social des migrations, toutes informations que le ministère référent refusait de publier. Et puis, comme collaboratrice parlementaire à Londres, j'avais vécu de l'intérieur l'affaire des dépenses parlementaires qui avait secoué comme jamais la Chambre des communes, temple de la démocratie britannique. Je démissionnerais la veille du scandale qui allait éclabousser mon propre patron, jusqu'à le mettre en prison. Et m'impliquerais ensuite avec ardeur, en tant que députée, dans la création de la Haute Autorité pour la transparence de la vie publique. Au fond, ce petit bout d'histoire individuelle s'inscrit dans un contexte historique, que Samuel Goëta rappelle d'ailleurs : ce n'est pas une coïncidence si les premiers projets d'open data ont été impulsés par des majorités progressistes, notamment suite à l'élection de Barack Obama à la présidence des États-Unis.

Pourquoi, somme toute, avoir fait le choix de l'open data pour la France? Pour emprunter le chemin, tracé par d'autres avant moi et qui

se poursuit encore, ancré dans un mouvement international très bien décrit par l'auteur et construit autour d'une idée forte : l'open data fait vivre la démocratie. Si les principes démocratiques qui animent l'action des communautés engagées dans l'open data sont à la fois simples et puissants, ceux-ci ont parfois été relégués au second rang des priorités. Il s'agit pourtant de concrétiser l'idée selon laquelle la transparence est gage de confiance, et seule la confiance est susceptible d'améliorer la qualité du débat démocratique. Avec l'open data, le pari est fait de l'intelligence collective, de la mise en lumière sans filtre des vérités objectivées par les données, afin de permettre aux acteurs du débat de décider en pleine conscience et d'assumer ensuite la redevabilité de leurs choix. L'open data, c'est le partage de l'information produite par les administrations, donc le partage du pouvoir avec les citoyens. L'affirmation que l'accès à l'information est une condition de la participation démocratique, qui trouve corps dans l'article 15 de la Déclaration du 26 août 1789 des droits de l'homme et du citoyen comme fondement de notre pacte social républicain : « *La société a le droit de demander compte à tout agent public de son administration* ».

Le présent ouvrage propose un décryptage détaillé des origines et des fondements du mouvement de l'open data, jusque dans ses applications les plus contemporaines. Au point de nous faire douter : lorsque nous posons un regard froid sur les débats qui animent les acteurs du numérique aujourd'hui, ce principe serait-il devenu obsolète ? Pire, pourrait-il se retourner contre les objectifs de redevabilité qu'il poursuivait à l'origine ? Le livre de Samuel Goëta pose des questions incontournables à travers des études de cas concrets, apporte des repères pour réussir les projets d'open data, et dessine des perspectives pour la suite. Finalement, les pages agissent comme un retour aux sources pour éclairer les enjeux du moment et nous guider dans la saine réaffirmation du sens démocratique de l'open data.

Le choix de l'open data, celui de partager l'information publique et d'autoriser sa réutilisation, n'est pas neutre. C'est tout l'intérêt de cette lecture que de révéler les vraies démarcations politiques qui se

cachent derrière des choix d'apparence technique, selon que l'on prône la gratuité ou les redevances, la protection de la propriété intellectuelle et du secret des affaires ou la transparence des jeux de données... Ce faisant, le texte navigue entre plusieurs disciplines, dans un affranchissement particulièrement original : histoire, sociologie et science politique, technique numérique et design de politique publique, témoignage engagé et guide des meilleurs usages. Oscillant entre la théorie et la pratique, l'idéal et la *realpolitik*, éclectique, inclassable, le livre est à l'image des personnes qui agissent au sein des communautés de l'open data, toujours mues par une aspiration commune.

On y comprend par exemple que ce n'est pas une loi, ni moins encore une ministre, qui aura permis de diffuser largement la culture de l'open data dans notre pays, mais plutôt l'épidémie de coronavirus. Le présent ouvrage raconte par le détail l'importance du suivi des données pour affronter la pandémie. Grâce justement au travail de collecte et de visualisation des données épidémiologiques fourni par des développeurs férus d'open data, les Français ont pu suivre les évolutions de la crise sanitaire en quasi temps réel. La prise de conscience de ce rôle joué par l'open data pour anticiper, suivre, gérer, et répondre aux problèmes soulevés par la crise sanitaire a été aussi soudaine que spectaculaire. Avec la pandémie, l'open data est entré dans le quotidien du plus grand nombre.

Au-delà de ce cas d'espèce, le livre restitue l'émergence des principes de l'open data en France, de l'inspiration trouvée dans la déclaration fondatrice de 2007 à Sebastopol en Californie, à la loi pour une République numérique de 2016, en passant par les différents décrets d'application et rapports publiés depuis. La deuxième partie du livre dresse un bilan critique de la politique d'open data en France : fallait-il inscrire dans la loi un principe absolu, celui de l'open data par défaut, ne souffrant d'exceptions que celles entrant dans les catégories recensées ? Et créer un nouveau service public de la donnée ? Ou aurait-on dû plus sobrement se contenter de définir des types de jeux de données à ouvrir, et laisser aux seules administrations le soin de choisir

souverainement de la mise en œuvre de leurs obligations? Je n'ignore pas que l'exécution de la loi pour une République numérique a pu décevoir. Aujourd'hui encore, les principes qui doivent guider les services publics – continuité, égalité, adaptabilité – ne trouvent pas à s'appliquer systématiquement à l'open data. Les décrets d'application ont tardé à venir; le portage politique n'a pas été au rendez-vous; l'agenda a été dépriorisé par les administrations; les collectivités locales ont été insuffisamment accompagnées et les ressources affectées sont restées trop insuffisantes. L'exemple de l'open data des décisions de justice est à ce titre représentatif des hésitations administratives et des enjeux corporatistes, qui ont pris le pas sur les questions légitimes à poser autour de l'anonymisation des informations personnelles. Pour autant, le sujet de l'open data a continué de tracer un sillon, et il me semble qu'il s'est désormais érigé au rang d'acquis législatif, à tel point qu'il serait difficile de le déboulonner. S'il fallait refaire un tel pari, je ne reculerais pas, car l'open data a contribué à apporter plus de transparence dans les décisions publiques, a aidé les journalistes à enquêter, a éclairé le débat public, a renforcé la maîtrise technique du numérique par les administrations, et a beaucoup nourri l'émergence d'un écosystème d'acteurs technologiques innovants.

Ce dernier point fait l'objet de controverses. D'ailleurs, le présent ouvrage interroge dans des termes assez critiques la réalité des retombées économiques de l'open data. Le discours pro-économique, tenu en particulier par les institutions européennes à travers de savants calculs modélisés, a pu faire rêver plus d'un élu local désireux de développer l'emploi et l'activité économique dans son territoire. Pour ma part, j'ai toujours été assez mal à l'aise à l'idée de vanter des retombées économiques directes, dans un raccourci «open data = emploi» un peu trop rapide. En revanche, la comparaison avec l'air, ou avec la lumière, comme élément essentiel à la vie économique des entreprises technologiques innovantes me paraît être une image plus conforme à ce que l'on constate : des effets indirects indéniables. Encore aujourd'hui, il m'arrive souvent de croiser des entrepreneurs de la French Tech ou des CTO (*chief technology officers*) travaillant pour

de grandes entreprises, qui témoignent de l'avance prise grâce à la possibilité de manipuler les grandes bases de données d'intérêt général, sur des thématiques liées à la mobilité, l'énergie, la construction, la liste des professions médicales... Ces bases sont désormais libres d'utilisation et leur usage crée une valeur économique nouvelle. Dans cet esprit, c'est bien à l'État et aux acteurs publics qu'il appartient de fournir et garantir la maintenance des infrastructures essentielles dont font partie les bases de données publiques.

Il n'empêche, Samuel Goëta dresse un bilan critique, et un peu déçu, des promesses de l'open data. Dès lors, l'ouvrage amène à s'interroger sur les raisons du retard observé dans la mise en œuvre du principe d'open data par défaut en France, en dehors des changements de majorité politique et de la volonté de se distinguer des pré-décesseurs. On peut penser que le caractère très transversal d'une politique d'open data rend compliqué l'amorçage d'une dynamique collective interministérielle, et les objectifs poursuivis par un ministère économique comme Bercy ne rejoignent pas forcément ceux des ministères de la Justice, du Logement ou des Affaires sociales. Surtout, il m'est apparu particulièrement compliqué de lancer une nouvelle politique publique, et de dégager les moyens budgétaires nécessaires pour en assurer l'exécution de manière pérenne en passant par les étapes budgétaires classiques. Ce fut vrai pour l'open data, mais aussi pour d'autres initiatives comme le programme French Tech, le plan France Mobile, la stratégie d'intelligence artificielle, la Grande école du numérique... Le temps des processus administratifs n'est pas celui du politique, et les obstacles propres à la transformation des organisations peuvent barrer la route aux volontés les plus farouches, lorsque le temps manque. Face à ce temps considéré comme trop long, il n'est pas étonnant que l'impatience des acteurs impliqués dans l'écosystème de l'open data se soit exprimée à plusieurs reprises.

Il reste que, depuis le lancement de la politique d'open data en France, les progrès accomplis dans la compréhension et la maîtrise de la culture de la donnée par la machine institutionnelle sont réels, et

fulgurants à l'échelle de l'histoire administrative française. On sous-estime l'impact positif de l'open data sur la maîtrise technique interne des jeux et enjeux de données par les agents publics. En ce sens, l'auteur a raison de démontrer le lien direct entre l'open data et les problématiques de souveraineté numérique, en expliquant comment le premier peut servir le second. Le raisonnement, un peu contre-intuitif (l'ouverture au service de la protection) est rarement mis en avant lorsqu'il s'agit de démontrer l'utilité de l'open data.

Toujours, derrière ce type de considérations stratégiques, le livre de Samuel Goëta nous renvoie à la question démocratique sous ses multiples facettes. On y découvre des illustrations très concrètes du pouvoir des données, notamment dans les évocations faites du droit d'accès aux documents administratifs, le *dada* de l'auteur. J'avais observé, pendant mes années londoniennes, l'importance prise par le Freedom of Information Act de 2000 au Royaume-Uni, et l'émergence du dispositif WhatDoTheyKnow créé par mySociety pour recueillir les requêtes soumises par des citoyens et publier les réponses des administrations. Une fois ministre, j'avais cherché à pousser une piste similaire, en confiant à la CADA la gestion d'une plateforme pour faciliter l'exercice du droit d'accès. En vain. C'est à l'initiative de l'écosystème des personnes engagées sur ces sujets, parmi lesquelles l'auteur du présent ouvrage, qu'un tel service a pu voir le jour. Ma Dada était né, et on ne peut que souhaiter longue vie à ce dispositif citoyen qui vient si utilement compléter les mécanismes mis en œuvre à travers l'open data.

La troisième partie de l'ouvrage propose des alternatives lorsque les données ne sont pas disponibles ou que leur qualité n'est pas au rendez-vous, et identifie des pistes pour que l'open data renforce plus encore la vie démocratique. Une sorte d'acte II de l'open data en France... À ce titre, le livre montre bien comment le « prochain combat » devra se trouver dans la question des moyens et des manières d'assurer la mise en œuvre des politiques d'open data. Cet acte II pourrait passer par la structuration territoriale de l'écosystème de l'open data. L'occasion, peut-être, de faire émerger un concept qui

m'est cher, celui du municipalisme numérique (selon lequel l'échelon local est le plus à même de mobiliser les citoyens autour de projets concrets liés à la vie en collectivité, en ayant au recours au numérique comme outil de renforcement du lien social). Les collectivités territoriales, et en particulier les régions en tant que chefs de file, seront amenées à jouer un rôle de plus en plus incontournable face aux crises et défis de notre temps; elles apportent cohésion, résilience, maillage local, et peuvent décliner le concept de solidarité dans et entre les territoires. L'acte II de l'open data pourrait ainsi s'insérer dans un agenda plus large de transferts de pouvoirs et de refondation de la vie démocratique locale.

L'on sent bien, au fur et à mesure que l'on avance dans les pages de l'ouvrage, qu'un certain nombre de questions vont être réouvertes, à l'aune des avancées technologiques, et en particulier de l'intelligence artificielle générative. La donnée étant devenue la nouvelle source de contrôle et de pouvoir, ne risque-t-on pas en prônant son ouverture et son accès librement réutilisable de déshabiller la sphère publique pour la donner en pâture aux intérêts économiques les plus puissants? Dans un monde de textes et d'images générés artificiellement, c'est-à-dire de manière automatisée et répondant à de simples demandes textuelles sous la forme de *prompts*, les données qui servent à entraîner les grands modèles de langage (les *LLM*, dont ChatGPT est l'application la plus connue à l'heure actuelle) deviennent des référentiels informationnels « authentiques ». L'information publique brute acquiert dès lors une valeur (de marché? d'éthique?) plus forte encore, puisqu'elle peut aider à fournir des bases de données non erronées, traçables, anonymisées, non biaisées, représentatives, respectueuses des cadres réglementaires. Face à la prolifération des contenus faux mais crédibles, et au risque de chaos informationnel, l'open data devient ainsi garante de la génération de contenus fiables et socle de la confiance numérique. Dans un mouvement doublement bénéfique, l'IA peut être employée pour analyser, traiter et interpréter les données ouvertes, et les données publiques peuvent être utilisées pour valider et vérifier les résultats produits par

des modèles d'IA. Mais dans ce monde du numérique « artificiel », le statut et la place des données publiques ouvertes restent incertains : communs numériques en partage, ou objet de prédation ? Les pages qui suivent nourrissent ces questions par un aller-retour entre l'histoire passée et l'actualité la plus récente. Et c'est là le grand intérêt de cet ouvrage : éclairer les enjeux qui traversent aujourd'hui l'open data par le rappel de ses principes fondateurs. On en ressort à la fois plus savant et renforcé dans nos convictions.

Et c'est heureux, car l'histoire n'est pas terminée. L'ouverture des données publiques – et ses extensions à travers les concepts de données altruistes, données d'intérêt général, données essentielles – devra jouer un rôle déterminant dans le plus grand défi qui se pose à notre siècle et à notre espèce : la lutte contre le changement climatique. Pour connaître les trajectoires de transition des pays vers des économies bas carbone, pour comprendre les choix de durabilité opérés par les entreprises, pour donner aux prochaines générations les clés de leur avenir, et éclairer dès à présent les décisions des responsables politiques et des investisseurs qui orientent les flux de financements, l'accès aux données environnementales devra demeurer libre. Les principes qui régissent l'open data devront pouvoir s'imposer, face aux appétits qui s'aiguisent sur le marché des données relatives à l'impact environnemental, et à l'opacité qui règne sur la réalité de cet impact. La transparence environnementale sera, là aussi, une exigence démocratique.

Introduction

L'enjeu démocratique

Je me suis intéressé à l'ouverture des données en 2008. Étudiant à l'université du Michigan cette année-là, je lisais un article expliquant les principes de l'*Open Government Data*, un mouvement prônant le partage et le libre usage des données produites par les acteurs publics¹. Au même moment, un de mes enseignants, Robert Frost, passionné de culture libre², me montrait combien les données étaient le fondement de l'information et du savoir. Potentiellement, nous pourrions refaire les calculs, proposer de nouvelles manières de voir le monde et réduire les asymétries d'information. Quelques semaines plus tard, Barack Obama était élu : la signature de trois mémorandums sur l'*open government* fut son premier acte en tant que président, ce qui lançait la création de *data.gov*, le premier portail *open data* national. J'avais le sentiment d'être aux premières loges de la naissance d'un mouvement d'ampleur. J'étais enthousiaste.

De retour en France, j'ai quelque peu déchanté. J'étais incapable de faire quoi que ce soit avec les premiers fichiers mis en ligne par plusieurs collectivités territoriales en 2010. Les acteurs publics étaient rares à ouvrir leurs données alors que les portails *open data* se multipliaient partout dans le monde. À part quelques collectivités territoriales pionnières comme Rennes, Paris, Nantes, Montpellier ou la région Provence-Alpes-Côte d'Azur, l'État semblait hermétique à ce mouvement et s'orientait même dans le sens de la commercialisation

1. Hubert Guillaud, «Principes pour des données publiques ouvertes», *Internet Actus*, 2007. <https://www.internetactu.net/2007/12/20/principes-pour-des-donnees-publiques-ouvertes/>.

2. Décédé en 2011, Robert Frost était professeur associé à la School of Information de l'université du Michigan. Arrière-petit fils du poète éponyme, farouchement opposé au copyright, il a créé un fonds dédié à l'accès ouvert alimenté par les recettes en droits d'auteur des œuvres de son arrière-grand-père. Sa passion et sa joie de vivre m'ont beaucoup inspiré.

des données publiques, y voyant une source potentiellement importante de revenus, là où l'open data implique généralement la gratuité et le libre usage. Les premiers exemples d'utilisation de données ouvertes relevaient de la prospective. On parlait de prototypes, d'expérimentations, de maquettes d'applications ou de services bricolés disponibles uniquement dans certains territoires. Je n'avais aucune expérience avec des données aussi complexes et je n'arrivais jamais à les transformer pour y donner un minimum de sens. J'ai tiré de cette expérience malheureuse l'idée que les données devaient devenir un objet de médiation afin de les rendre accessibles à un plus grand nombre de personnes.

Plus de dix ans après, certains sont encore restés sur ce souvenir d'un open data peu accessible avec des données décevantes et des usages marginaux. Certaines personnes sont rapidement passées à autre chose et se sont penchées sur d'autres tendances du numérique comme la blockchain, le métavers, les NFT ou l'intelligence artificielle générative. D'autres ont progressivement vu tout le potentiel de l'ouverture des données pour renouveler la transparence, faire foisonner l'innovation et transformer l'administration. C'était mon cas : après un mémoire et une première expérience professionnelle, je me suis lancé en 2012 dans un doctorat en sociologie pour étudier l'émergence du mouvement de l'open data, sa diffusion en France et la mise en œuvre concrète des projets d'ouverture de données dans les administrations françaises³. En parallèle, j'ai co-fondé l'association Open Knowledge France pour diffuser l'ouverture des données et faire que le plus grand nombre puisse s'en saisir. À l'issue de cette thèse en 2016, nous avons fondé avec Joël Gombin la société coopérative et participative Dataactivist qui accompagne des projets d'ouverture de données et de leur réutilisation. Son nom témoigne de l'engagement

3. Samuel Goëta, *Instaurer des données, instaurer des publics : une enquête sociologique dans les coulisses de l'open data*, thèse de doctorat en sociologie sous la direction de Jérôme Denis, Télécom ParisTech, 2016. <https://pastel.hal.science/tel-01458098>. Les chapitres 2 et 3 sont tirés en grande partie de ma thèse.

en faveur de l'ouverture des données de la coopérative, qui compte aujourd'hui vingt-deux salariés.

Là où il y a dix ans nous illustrions le concept d'open data par des cas d'usage hypothétiques ou marginaux, beaucoup de nos interventions chez Dataactivist débutent aujourd'hui par un petit panorama des usages des données ouvertes. Et il ne nous faut pas longtemps pour montrer que l'open data a désormais envahi notre quotidien : le jeu Pokémon Go s'appuie sur OpenStreetMap (le « *Wikipédia de la carte* ») pour identifier les zones où placer des Pokémon rares, l'application Yuka a longtemps utilisé les données collaboratives d'Open Food Facts comme source de données sur les produits alimentaires, CovidTracker a rendu lisibles au plus grand nombre les indicateurs de Santé publique France ou Meilleurs Agents aide les acheteurs à connaître les prix des biens immobiliers à partir des données de la Direction générale des finances publiques (DGFiP). Pour suivre l'air du temps, j'évoque parfois l'entraînement (controversé, nous y reviendrons) de ChatGPT à partir des données ouvertes de Common Crawl, une base de données massive comportant le texte de plusieurs milliards de pages web. Il m'arrive aussi souvent d'évoquer le cas du passage aux 80 km/h sur les routes départementales pour expliquer l'ouverture des données. C'était le cas il y a quelques années lors d'une discussion à Marseille avec un chauffeur de taxi curieux de mon métier. En 2011, le ministère de l'Intérieur a publié sur data.gouv.fr (le portail national des données publiques) la base de données des accidents corporels de la circulation. Celle-ci décrit les véhicules impliqués, les conditions de circulation, les conséquences et la localisation précise pour chaque accident de la route reporté aux services de secours. Alors qu'en 2018, le Premier ministre Édouard Philippe, avait justifié le passage aux 80 km/h par les résultats prétendument positifs d'une expérimentation sur trois tronçons test, les données diffusées sur data.gouv.fr ont été utilisées par l'association 40 millions d'automobilistes pour vérifier la baisse de l'accidentologie. Elle n'a constaté aucune baisse sur les trois tronçons et a alors diffusé l'idée que cette mesure n'avait pas le fondement objectif avancé par le

gouvernement pour justifier de sa décision. Bavard et sans réserve, le chauffeur de taxi s'est alors exclamé : « donc en fait, vous me dites que le gouvernement a créé un site qui permet de montrer que ce sont des gros cons ? »

Cette anecdote qui m'amuse encore beaucoup est bien plus intéressante qu'elle n'y paraît. Et si l'open data était véritablement un outil pour critiquer les puissants ? Pourrions-nous remettre en cause l'ordre établi avec les données ouvertes ? Ce livre part d'une question que je me pose depuis plusieurs années⁴ : l'open data est-il vraiment un contre-pouvoir ? Comment mettre les données ouvertes au service des citoyens et de la démocratie ? Marie-Hélène Bacqué a délimité trois caractéristiques qui désignent un contre-pouvoir : une capacité de contre-expertise et de contre-proposition, un pouvoir de pression sur la décision et de contrôle sur la réalisation des politiques publiques, la mise en place d'une méthode démocratique⁵. En défendant l'importance des contre-pouvoirs dans la vie publique, je cherche ici à souligner le rôle démocratique essentiel que peut jouer l'ouverture des données dans leur outillage. Pierre Rosanvallon nomme « *contre-démocratie* » les formes de contrôle, d'empêchement et de mise à l'épreuve du jugement des décisions démocratiques⁶. Il a montré que la « contre-démocratie » ne s'oppose pas à la démocratie. Au contraire, elle l'enrichit, la conforte et fait système avec les institutions démocratiques. Elle permet de « *veiller à ce que le pouvoir élu reste fidèle à ses engagements, de trouver les moyens permettant de maintenir l'exigence initiale d'un service du bien commun* ». Les données ouvertes représentent selon moi un outil primordial dans la réalisation de cette « contre-démocratie ».

4. Avec Clément Mabi, nous avons coécrit un article à ce sujet : Samuel Goëta & Clément Mabi, « L'open data peut-il (encore) servir les citoyens ? », *Mouvements*, 2014, vol. 79, n° 3, p. 81-91. <https://doi.org/10.3917/mouv.079.0081>.

5. Marie-Hélène Bacqué. « Dispositifs participatifs dans les quartiers populaires, héritage des mouvements sociaux ou néolibéralisme ? Empowerment zones aux États-Unis et politique de la ville en France » in Marie-Hélène Bacqué (éd.), *Gestion de proximité et démocratie participative. Une perspective comparative*, La Découverte, 2005, p. 81-99.

6. Pierre Rosanvallon, *La Contre-Démocratie. La politique à l'âge de la défiance*, Seuil, 2006.

L'exemple le plus saisissant du rôle démocratique de l'ouverture des données porte sur un article de science économique qui a largement influencé les politiques d'austérité dans le monde⁷. En 2010, Carmen Reinhart et Kenneth Rogoff, deux économistes américains influents de l'université Harvard, publiaient un article⁸. dans l'*American Economic Review*, qui analysait 60 ans de données économiques pour montrer un lien entre dette et croissance. L'article pointait en particulier un effet de seuil qui entravait durablement la croissance lorsque la dette accumulée dépassait 90% du PIB annuel. Largement reprise par les économistes et les décideurs politiques, cette publication a eu une influence considérable sur les politiques d'austérité mises en place, notamment en Europe, au début des années 2010. Or, en 2013, Thomas Herndon, un étudiant en économie à l'université du Massachusetts, a montré qu'elle était fondée sur une série d'erreurs de calcul. Dans le cadre d'un devoir étudiant, il a tenté de reproduire sans succès les résultats de Reinhart et Rogoff. Après vérification par ses enseignants, il a écrit aux auteurs de l'article qui lui ont fourni le fichier Excel avec lequel ils ont produit les calculs. En vérifiant le fichier avec ses enseignants, il a pu montrer que les auteurs avaient attribué des poids différents aux pays dans le calcul de la moyenne, que des données sur certains pays n'ont pas été utilisées et surtout que le calcul de la moyenne omettait l'Australie, l'Autriche, la Belgique, le Canada et le Danemark. Ces pays ont pour point commun d'être au début de la liste triée par ordre alphabétique. Ces erreurs corrigées, le seuil de 90%, si utile pour justifier des politiques d'austérité, n'était plus du tout significatif. Cet exemple pourrait nous mener à de plus amples débats sur l'éthique scientifique (absence de revue par les pairs de l'article, accès aux données à la demande, usage de méthodes statistiques douteuses, instrumentalisation de la recherche...) mais mon but ici est de montrer l'importance démocratique de pouvoir accéder et utiliser les données qui fondent la

7. David Louapre, «Les politiques d'austérité, à cause d'une erreur sous Excel?», *Science Étonnante*, 2020. <https://scienceetonnante.com/2020/04/17/austerite-excel/>.

8. Carmen M. Reinhart & Kenneth S. Rogoff. 2010. «Growth in a Time of Debt», *American Economic Review*, 100 (2): 573-78. <https://doi.org/10.1257/aer.100.2.573>.

décision démocratique. Pendant longtemps, nous nous sommes contentés d'informations quantitatives «prémâchées» pour animer le débat public, sans avoir la possibilité de les remettre en cause. En ayant accès aux données dans leur plus fort niveau de précision, celles qui sont à la base de l'information et du savoir, nous pouvons refaire les calculs, contester les décisions et proposer de nouvelles interprétations.

Dans l'édifice démocratique, les données peuvent aussi servir de moyen d'*empowerment* au sens d'une émancipation pour les groupes sociaux marginalisés ou défavorisés et d'une remise en cause du système qui entretient les inégalités et certaines formes d'oppression. Dans *Data Feminism*, Catherine D'Ignazio et Lauren Klein montrent que les données sont une «*épée à double tranchant*», une arme qui peut être au service du pouvoir comme des contre-pouvoirs⁹. Le cas du comptage des féminicides en France le montre bien. La notion de féminicide, définie par la sociologue Diana Russel en 1976 comme l'assassinat d'une ou plusieurs femmes par un ou plusieurs hommes parce qu'elles sont femmes, a pris de l'ampleur dans l'espace médiatique français entre 2018 et 2019¹⁰. Outre l'actualité liée à #metoo en octobre 2017, le comptage des féminicides par plusieurs associations a joué un rôle déterminant dans la mise en avant dans le débat public de cette notion. Le politiste Patrick Hassenteufel a montré que la quantification constitue un facteur clé dans la «mise à l'agenda», dans la publicisation des problèmes publics, en donnant une légitimité scientifique associée au comptage¹¹. Dès 2016, le collectif «féminicides par compagnons ou ex» a initié un comptage des féminicides

9. Catherine D'Ignazio & Lauren Klein, *Data Feminism*, The MIT Press, 2020.

10. Selon une étude de la fondation Jean Jaurès s'appuyant sur des données Europresse, le terme «féminicide» comptabilisait 180 occurrences en 2017, 400 occurrences en 2018 et 5000 en 2019. Source : Margot Giacinti, «Parler de féminicide pour mieux lutter contre les violences sexistes et sexuelles», *Fondation Jean Jaurès*, 2020. <https://www.jean-jaures.org/publication/parler-de-feminicide-pour-mieux-lutter-contre-les-violences-sexistes-et-sexuelles/>.

11. Patrick Hassenteufel, «Chapitre 3. Les processus de construction et de mise à l'agenda des problèmes publics», *Sociologie politique de l'action publique*, Armand Colin, 2021, p. 105-129.

intimes et conjugaux pour sensibiliser à la notion, provoquer une prise de conscience et susciter la réaction des pouvoirs publics¹². Le travail du collectif, qui a documenté *a minima* 112 cas de femmes tuées par leur (ex)compagnon en 2022 à partir de sources de presse, a permis de montrer que les féminicides sont l'expression d'un rapport de domination des hommes sur les femmes, là où la Délégation aux victimes (DAV) du ministère de l'Intérieur a longtemps négligé le genre comme un facteur pertinent pour analyser le crime. Le travail de comptage du collectif a provoqué un déclic dans le débat public et une meilleure prise en compte du genre comme facteur déterminant : la DAV met désormais en avant que les femmes représentent 85 % des victimes d'homicides au sein du couple¹³. L'Inter Orga Féminicides étend le comptage aux cas de meurtres et de suicides, quelles que soient les circonstances et pas uniquement au sein du couple, décomptant 147 femmes mortes en raison de leur genre¹⁴.

Ces quelques exemples montrent l'importance de l'ouverture des données pour la démocratie. Le présent livre cherche à aider les citoyens à s'emparer des données ouvertes. Je m'inscris dans le mouvement de la littératie des données (*data literacy*), qui considère que la compréhension et l'usage des données doivent devenir une compétence essentielle, comme lire, écrire ou compter. Mon but est rendre accessible au plus grand nombre le concept de données ouvertes et de sortir de la situation actuelle dans laquelle les données sont réservées à une petite élite. Ce livre vise à mieux connaître le potentiel des données

12. Collectif Féminicides par compagnons ou ex, « Notre méthodologie », 2023. <https://www.feminicides.fr/notre-m%C3%A9thodologie>.

13. La moitié des femmes autrices avaient déjà été victimes de violences de la part de leur partenaire. Mission interministérielle pour la protection des femmes contre les violences et la lutte contre la traite des êtres humains, « Lettre n° 18 - Les violences au sein du couple et les violences sexuelles en 2021 », 2021. <https://arretonslesviolences.gouv.fr/sites/default/files/2022-11/Lettre%20n%C2%B018%20-%20Les%20violences%20au%20sein%20du%20couple%20et%20les%20violences%20sexuelles%20en%202021.pdf>.

14. #NousToutes, « Comprendre les chiffres pour mieux défendre les femmes et les enfants victimes de violences sexistes et sexuelles », 2023. <https://www.noustoutes.org/comprendre-les-chiffres/>.

disponibles et à ouvrir des pistes d'action pour que l'open data renforce les contre-pouvoirs et réduise les asymétries d'information entre la société civile et l'administration. Il s'adresse aussi aux professionnels des données en dressant un bilan critique de l'ouverture au regard des promesses de transparence, d'innovation et de transformation de l'administration formulées par les défenseurs des données ouvertes depuis plus d'une décennie. Il cherche à mettre en avant de nouveaux cas d'usage, proposer un regard distancié sur les avancées de ce mouvement et formuler des propositions pour une nouvelle étape de l'ouverture des données.

Dans une première partie, le livre commence par poser les bases de l'open data en présentant les racines profondes de ce mouvement, puis propose un historique des grands principes de l'open data et de leur traduction en France. Le chapitre qui termine cette partie propose un éclairage sur le rôle central des données ouvertes dans la gestion de la pandémie de Covid-19 et montre comment cette crise a questionné la mise en œuvre effective du principe d'open data par défaut prévu par la loi en France. La deuxième partie de l'ouvrage dresse un bilan critique de l'ouverture des données en France par rapport aux grandes promesses d'un renouvellement en profondeur de la transparence de la vie publique, d'un foisonnement de l'innovation et d'une transformation de l'administration. Enfin, la dernière partie propose des pistes d'action. Dans un premier temps avec cinq stratégies alternatives pour obtenir des données en l'absence de données publiques, puis par une série de pistes d'action pour que les données ouvertes réduisent effectivement les asymétries d'information et renforcent les contre-pouvoirs.

Première partie.
Naissance et émergence
de l'open data

Les racines multiples de l'open data

Lorsqu'on évoque le terme open data, la plupart des personnes l'associent à un phénomène nouveau dont la généalogie pourrait remonter à une dizaine d'années. L'ouverture des données a pourtant tout de la fausse nouveauté tant les concepts, les pratiques, les réseaux d'acteurs et le cadre réglementaire ont largement devancé l'avènement du terme. Avant de retracer l'émergence des principes essentiels de l'ouverture des données, je voudrais vous inviter à situer les différentes filiations dans lesquelles cette notion s'inscrit et à sortir de l'image d'une uniformité des acteurs qui se sont réunis autour du terme open data.

Dans ce chapitre, j'aborderai six facettes de l'ouverture des données en commençant par la transparence qui apparaît aujourd'hui comme une exigence essentielle en démocratie. Elle a donné lieu aux lois sur le droit d'accès à l'information publique, qui accordent aux citoyens un accès direct aux documents administratifs et constituent le socle juridique de l'ouverture des données. L'open data s'inscrit aussi dans la lignée de la pensée cybernétique pour qui la libre circulation de l'information constitue une condition essentielle de survie et d'équilibre des sociétés démocratiques. Cette filiation relie l'ouverture des données au mouvement du logiciel libre : ses principes ont servi de matrice à ceux de l'open data. La philosophie de l'ouverture et de la liberté prônée par le mouvement du logiciel libre s'est étendue pour réclamer la libre circulation des œuvres de l'esprit, au-delà du code informatique. Autre racine, le mouvement dit de la science ouverte a fait progressivement émerger le partage des données brutes et leur utilisation par d'autres comme une pratique scientifique légitime et courante, aujourd'hui devenue la norme dans de nombreuses

disciplines scientifiques. Le terme *open data* a connu ses premières apparitions dans les sciences pour désigner des pratiques de partage de données. Dans une tout autre lignée, je m'intéresserai aux industries de l'information qui ont réclaté dès les années 1980 aux États-Unis la mise à disposition gratuite des données du secteur public et le droit à leur libre utilisation. Je montrerai aussi comment certaines données ouvertes ont servi de données d'entraînement aux modèles d'intelligence artificielle qui font aujourd'hui tant débat. Enfin, je montrerai comment l'ouverture des données s'inscrit à la suite de nombreuses initiatives de transformation ou de modernisation de l'administration qui visent à casser les silos organisationnels, faire circuler l'information et favoriser la collaboration avec le public.

■ La transparence et le droit d'accès à l'information

Entendue comme divulgation de l'information au public, la demande de transparence a été popularisée principalement au XVIII^e siècle par des penseurs comme Kant, Rousseau ou Bentham¹⁵. La transparence administrative, permise par le droit d'accès à l'information publique, trouve son acte de naissance en Suède. En 1766, le Parlement de Suède a adopté un décret relatif à la liberté de la presse et au droit d'accès aux documents publics. Il prévoit que tout citoyen suédois est autorisé à consulter les actes officiels, un document étant « officiel » dès lors qu'il est conservé par une administration¹⁶. En France, avec la Déclaration des droits de l'homme et du citoyen (DDHC) de 1789, le droit d'accès a été revendiqué comme un élément indispensable à la souveraineté démocratique. L'article 15 de la DDHC qui stipule que « *la Société a le droit de demander compte à tout agent public de son administration* » pose les bases d'un devoir de reddition des comptes aux

15. Clare Birchall, « Introduction to "Secrecy and Transparency": The Politics of Opacity and Openness », *Theory, Culture & Society*, 2012, vol. 28, n° 7-8, p. 7-25.

16. Thorsten Ader & Max Schoenthal, « L'accès aux informations relatives aux activités de l'État, en particulier du point de vue des médias », *IRIS plus, Observations juridiques de l'Observatoire européen de l'audiovisuel*, 2005, n° 2. <https://rm.coe.int/16807833bd>.

citoyens. L'abbé Sieyès serait à l'origine de cet article¹⁷. Homme d'Église et essayiste, il a été l'un des principaux acteurs de la Révolution française et l'auteur de plusieurs pamphlets dans lesquels il a esquissé les contours d'une société démocratique. Selon Sieyès, la démocratie représentative ne peut s'exercer que si les représentants rendent des comptes aux citoyens qui leur ont délégué l'exercice du pouvoir. Pour s'assurer que la souveraineté nationale soit bien exécutée, il faut comprendre dans quelles conditions les représentants ont exercé le pouvoir qui leur a été délégué et donc pouvoir demander des comptes.

Pour rendre applicable le droit consacré par l'article 15 de la DDHC, la loi du 7 messidor an II (25 juin 1794) a établi le principe de publicité des archives et a affirmé le principe de libre consultation dans son article 37 qui stipule que «*tout citoyen pourra demander dans tous les dépôts, aux jours et heures qui seront fixés, communication des pièces qu'ils renferment*». Outre les archives, les bibliothèques et les musées ont contribué à populariser la notion de transparence en occupant une place centrale dans les villes et rendant les connaissances accessibles à tous¹⁸. L'idée de transparence s'est aussi développée avec l'importance croissante des États-Unis au xx^e siècle. On peut citer par exemple le juge Louis Brandeis, un des principaux inspirateurs du programme du président Wilson, qui a écrit dans son essai *Other People's Money* cette formule devenue emblématique au sujet de la transparence : «*la lumière du soleil est dit-on le meilleur des désinfectants ; l'éclairage électrique le plus efficace des policiers*¹⁹». Critiquant les banquiers de son époque accusés de conflits d'intérêts dans leurs investissements, Brandeis plaidait en faveur de la *publicité*, l'idée que les banquiers doivent porter à la connaissance du public les commissions ou profits qu'ils perçoivent²⁰. Symbolisant l'importance de la

17. Ben Kafka, *The Demon of Writing. Powers and Failures of Paperwork*, Princeton University Press, 2012.

18. Clare Birchall, *op. cit.*, 2012.

19. Louis Brandeis, *Other People's Money And How the Bankers Use It*, F.A. Stokes, 1914.

transparence dans la défense de l'intégrité publique et dans la lutte contre la corruption, sa métaphore sur la lumière est aujourd'hui devenue le slogan de nombreux militants de l'open data comme la Sunlight Foundation (qui en tire son nom) ou Joshua Tauberer²¹, un activiste de l'ouverture des données que j'évoquerai plus tard dans cet ouvrage.

Jusqu'au milieu du xx^e siècle, la transparence reposait sur le truchement d'institutions comme les archives ou la Cour des comptes qui assuraient la diffusion des informations publiques pour le compte des citoyens. Aux États-Unis, à la suite de la Seconde Guerre mondiale, le mouvement dit de l'*Open Government* a obtenu l'instauration d'un droit d'accès des citoyens aux documents et données produits par les acteurs publics²². Portée en grande partie par des journalistes, la revendication d'*Open Government* désignait la révélation des secrets du fonctionnement de l'État et s'est développée dans la critique de l'opacité de l'armée américaine lors de la guerre froide, en particulier pendant la guerre du Vietnam. Autour de la notion de gouvernement ouvert, les journalistes ont tenté d'obtenir un droit d'accès à l'information publique. En 1966, le *Freedom of Information Act* signé par le président Johnson a donné à chaque citoyen américain le droit d'exiger auprès de l'administration les informations publiques qu'elle détient. En France, la loi Cada du nom de la Commission d'accès aux documents administratifs, a accordé en 1978 le droit pour chaque citoyen d'obtenir les documents produits dans le cadre d'une mission de service public concernant le demandeur ou ne tombant pas sous le coup d'un des secrets protégés (vie privée des individus, délibérations du gouvernement, sûreté de l'État, secret industriel et commercial...) L'exercice de ce droit s'effectue par le dépôt d'une requête officielle

20. Lawrence Lessig, «Et si la démocratie se condamnait elle-même?», *Courrier International*, 2010. <https://www.courrierinternational.com/article/2010/02/25/et-si-la-democratie-se-condamnait-elle-meme>.

21. Joshua Tauberer, *Open Government Data: The Book*, 2014. <https://opengovdata.io/>.

22. Harlan Yu & David G. Robinson, «The New Ambiguity of "Open Government"», *UCLA Law Review*, 2012. <https://www.uclalawreview.org/pdf/discourse/59-11.pdf>.

auprès d'un agent public, la Cada pouvant arbitrer en cas de litige. Aujourd'hui, le « droit de savoir » est devenu la norme dans la plupart des démocraties : selon l'ONG Access Info, 134 pays ont adopté une législation de droit d'accès à l'information²³.

À la différence du droit d'accès, l'ouverture des données publiques pose un principe de diffusion proactive : là où le droit d'accès suppose une demande, l'ouverture consiste à publier de manière volontaire les données sans attendre une requête. Le mouvement de l'open data a aussi mis en lumière les données comme étant les éléments de base à partir desquels l'information et le savoir sont créés²⁴. La publication de données ouvertes est ainsi souvent considérée comme une forme supérieure de transparence, réputée plus objective que les formes de révélation narratives ou interprétatives telles que le scandale ou les rumeurs qui sont associées à l'illégalité et l'immoralité²⁵. En diffusant les données sous leur forme granulaire, c'est-à-dire dans leur plus grand niveau de précision, les asymétries d'information entre le public et l'administration devraient se réduire et des connaissances inédites devraient émerger.

Dispensées d'un biais narratif ou interprétatif, les données « brutes » sont associées à une objectivité nouvelle qui configure les politiques d'open data comme le renouveau, voire le paroxysme de la transparence de l'État²⁶. Evelyn Ruppert a montré que les politiques d'ouverture des données sont fondées sur le postulat de l'existence de « publics des données » (*data publics*), des collectifs d'individus qui pourraient potentiellement inspecter le fonctionnement de l'État à partir des données²⁷. S'intéressant au *Transparency Agenda* mis en

23. Centre for Law and Democracy & Access Info, « Global Right to Information Rating », 2020. <https://www.rti-rating.org/>.

24. Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*, Sage Publications, 2014.

25. Clare Birchall, « Radical Transparency? », *Cultural Studies ↔ Critical Methodologies*, 2014, vol. 14, n° 1, p. 77-88.

26. Samuel Goëta, « L'open data : une forme ultime de transparence? » in Andrea Catellani et al., *La communication transparente. L'impératif de la transparence dans le discours des organisations*, Presses Universitaires de Louvain, 2015.

place par le gouvernement britannique de David Cameron en 2010 suite au scandale des dépenses des parlementaires, Ruppert a montré la dimension performative de l'ouverture des données : les politiques d'open data réalisent ce qu'elles énoncent. Dans l'esprit du *Transparency Agenda*, le simple fait d'exposer les données positionne les militants associatifs, les chercheurs ou les journalistes comme des « auditeurs de salon » (*armchair auditors*), un peu comme si l'existence de citoyens motivés et compétents pouvait suffire à décupler le pouvoir d'inspection de la Cour des comptes. Peu importe pour les décideurs si ces différents publics s'emparent effectivement des données ou sont même en capacité de s'en saisir, la publication de ces données suffirait à atteindre un nouveau degré dans la transparence et le contrôle citoyen de l'action publique. L'ouverture des données marque ainsi une reconfiguration de l'expertise selon Evelyn Ruppert. Alors que le Premier ministre David Cameron avait annoncé la naissance d'une nouvelle ère *post-bureaucratique* où les agents publics ne sont plus les seuls à interpréter les données, elle met en garde sur le fait que « *le pouvoir sur les données va ostensiblement reposer sur un public imaginé* » alors qu'en réalité « *la plupart des gens manquent des outils ou de l'expertise pour donner du sens aux téraoctets de données diffusés* ». Retenons ces questions importantes posées par Evelyn Ruppert, j'y reviendrai à plusieurs reprises.

■ La liberté de l'information et le logiciel libre

Au-delà des revendications de transparence de la vie publique, la demande d'ouverture des données s'inscrit dans un mouvement plus large qui considère que la diffusion de l'information est impérative pour la survie des sociétés démocratiques. L'idée d'ouverture a émergé au milieu du xx^e siècle dans les écrits de philosophes et de

27. Evelyn Ruppert, «Doing the transparent state: open government data as performance indicators» in R. Rottenburg, S. E. Merry, S.-J. Park & J. Mugler, *A World of Indicators: The making of governmental knowledge through quantification*, Cambridge University Press, 2015, p. 51-78.

théoriciens comme Karl Popper²⁸ qui, en 1945, dénonçait les « sociétés fermées » fondées sur des vérités incontestables. La pensée cybernétique a aussi joué un rôle déterminant dans l'émergence de l'idéologie de l'ouverture en postulant que la circulation de l'information est indispensable à la perpétuation de la civilisation. Selon les théories de la cybernétique, dans un monde menacé par l'entropie, le chaos inéluctable annoncé par les principes physiques de la thermodynamique, les sociétés « ouvertes » où l'information circule librement parviennent à créer de l'ordre et à lutter contre l'effondrement²⁹.

Cette croyance dans la circulation de l'information comme vecteur de stabilité et d'harmonie se retrouve dans le schéma de la pyramide du savoir qui place les données comme le fondement de l'information et du savoir³⁰.

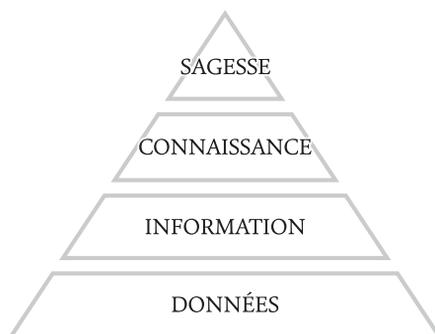


Figure 1. La pyramide du savoir de Russel Ackoff.

28. Karl Popper, *The Open Society and Its Enemies*, Routledge, 1945 (*La Société ouverte et ses ennemis*, Seuil, 1979).

29. Nathaniel Tkacz, « From open source to open government: a critique of open politics », *Ephemera: Theory and Politics in Organization*, 2012, vol. 12, n° 4, p. 386-405.

30. Luciano Floridi, « "Data" » in W. Darity, *International Encyclopedia of the Social Sciences*, Macmillan Social Science Library, 2008.

Martin Frické, « The knowledge pyramid: a critique of the DIKW hierarchy », *Journal of Information Science*, 2009, vol. 35, n° 2, p. 131-142.

Jennifer Rowley, « The wisdom hierarchy: representations of the DIKW hierarchy », *Journal of Information Science*, 2007, vol. 33, n° 2, p. 163-180.

L'auteur de ce schéma publié en 1989³¹ est Russel Ackoff, un théoricien des organisations et du management de l'information, largement influencé par la théorie cybernétique. La forme de la pyramide indique une progression : les données seraient le fondement de l'information dont découlent ensuite le savoir et puis la sagesse. Ce modèle est évidemment critiquable : la philosophie qui consiste à atteindre une forme de sagesse, doit-elle être nécessairement fondée sur des données ? La collecte massive et la circulation des données sur un phénomène comme le réchauffement climatique a-t-elle entraîné mécaniquement plus de sagesse dans nos comportements ? Malgré ses limites, ce modèle a beaucoup circulé au point de devenir un des fondements des sciences de l'information.

Apparus au début des années 1980, les mouvements liés au logiciel libre ont aussi largement milité pour la liberté de circulation de l'information. Sébastien Broca dans *L'Utopie du logiciel libre* revient sur les fondements intellectuels de ces mouvements et soutient que « *la libre circulation de l'information est la valeur la plus communément associée au combat du Libre, que ce soit par les universitaires ayant étudié le sujet ou par les hackers eux-mêmes*³² ». Des partisans du logiciel libre se sont en effet engagés pour que la connaissance et les œuvres bénéficient des mêmes libertés d'utilisation, de modification, de copie et de redistribution que le code informatique des logiciels qu'ils utilisent. Dans la continuité de luttes pour la circulation de l'information, des libristes ont participé à la définition des principes de l'open data en défendant notamment l'usage de formats aux spécifications ouvertes et de licences dites « *copyleft* » qui assurent que les créations dérivées héritent de la même licence et préservent le statut de bien commun de l'œuvre.

Dans *Open Standards and the Digital Age*, Andrew Russel invite à remonter plus loin que la filiation évidente du logiciel libre pour comprendre le foisonnement contemporain de mouvements qui se

31. Russel Ackoff, « From Data to Wisdom », *Journal of Applied System Analysis*, 1989, vol. 16, p. 3-9.

32. Sébastien Broca, *Utopie du logiciel libre. Du bricolage informatique à la réinvention sociale*, Le Passager Clandestin, 2013.

revendiquent de l'ouverture (*open access, open science, open government...*). Selon Russel, l'histoire des standards de télécommunications a fait émerger une philosophie voire une idéologie de l'ouverture³³. Dans les années 1970, des ingénieurs ont affirmé la supériorité des systèmes ouverts lors de la conception des standards mondiaux de télécommunication. En particulier, en août 1977, l'International Organisation for Standardization (ISO) a créé un groupe de travail intitulé Open Systems Interconnection (OSI) pour concevoir des standards ouverts d'interconnexion des réseaux et des terminaux. Bien que l'OSI ait échoué à définir les standards d'Internet face au modèle TCP/IP, la métaphore de l'ouverture s'est imposée auprès d'ingénieurs issus de domaines très variés comme un cri de ralliement pour décloisonner la société et les savoirs.

À travers ces multiples filiations (cybernétique, logiciel libre, standards ouverts), les militants de l'open data ont renouvelé la croyance en la libre circulation de l'information. En ayant accès aux données brutes, au matériau de l'information et du savoir, chacun serait en capacité de produire de nouvelles connaissances. Rob Kitchin présente l'accès aux données comme un facteur d'innovation et d'avance stratégique : « *celui qui a accès à de larges données de grande qualité a un avantage compétitif sur ceux qui sont exclus de pouvoir générer de la compréhension et de la sagesse*³⁴ ». Si le savoir est le pouvoir³⁵, ouvrir les données serait un moyen efficace de le distribuer.

33. Andrew Russel, *Open Standards and the Digital Age. History, Ideology, and Networks*, Cambridge University Press, 2014.

34. Rob Kitchin, *op. cit.*

35. L'expression anglo-saxonne « *knowledge is power* » est souvent employée pour justifier la circulation de l'information et l'ouverture des données. Savoir serait synonyme de pouvoir. Quand j'entends cet argument, je montre souvent un extrait du premier épisode de la saison 2 de la série *Game of Thrones* dans lequel Littlefinger (intrigant conseiller du roi) menace la reine Cersei de dévoiler sa relation incestueuse avec son frère en disant « *knowledge is power* ». Elle demande alors immédiatement à sa garde de l'arrêter et de lui couper la gorge, les interrompant au dernier moment dans leur geste. Elle lui répond alors « *power is power* ». Réduire le pouvoir au savoir, c'est oublier une des caractéristiques essentielles du pouvoir, le monopole de la violence physique légitime, que Max Weber avait théorisé.

■ Le partage des données de la recherche

Pour Bruno Strasser, l'émergence des sciences guidées par les données (*data-driven sciences*) trouve ses sources dans les méthodes comparatives issues des sciences du vivant³⁶. Les cabinets de curiosité où des espèces hétéroclites étaient entreposées, classifiées et exposées sont en quelque sorte les ancêtres des bases de données modernes. Cette tradition de collection a abouti au XIX^e siècle à l'essor des muséums de sciences naturelles et leurs immenses collections d'espèces. Au début du XX^e siècle, il montre que cette longue tradition a décliné sous l'assaut de la science expérimentale qui a consacré le laboratoire comme le lieu de la recherche scientifique légitime. Le projet de codifier et de numériser l'ADN dans les années 1960 a marqué le retour à la tradition comparative de collection et de comparaison des espèces. Strasser a montré qu'une des raisons du succès de Genbank, la première base massive de données en génétique, a reposé sur le partage des données et l'apparence d'ouverture du système. Dans d'autres *data-driven sciences* comme l'astronomie, la botanique ou la génétique, le partage et la réutilisation de grandes bases de données sont devenus des pratiques courantes. Avec la disponibilité de données de plus en plus importantes, de nouveaux métiers ont émergé comme les *computational scientists* qui analysent les données produites par d'autres. Certains ont vu d'un mauvais œil ces chercheurs qui publient en leur propre nom des résultats s'appuyant sur les données produites par d'autres, les qualifiant parfois de « chercheurs parasites³⁷ ».

Les politiques d'open data s'inscrivent dans la lignée de ces pratiques anciennes de partage de données scientifiques. Le terme même

36. Bruno J. Strasser, «Data-driven sciences: From wonder cabinets to electronic databases.», *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012, vol. 43, n° 1, p. 85-87.

37. L'expression vient de la tribune de 2016 : Dan Longo & Jeffrey Drazen, «Data Sharing», *New England Journal of Medicine*, 2016, vol. 374, n° 3, p. 276-277. Chaque année, le prix nommé ironiquement «Research Parasite Awards» consacre les meilleures recherches nées de la réutilisation de données produites par d'autres chercheurs.

d'open data est apparu dans les sciences : il est employé pour la première fois dans les années 1970 dans des accords que la Nasa a signés avec des pays partenaires en vue du partage de données satellitaires³⁸. En 1995, la National Academy of Sciences aux États-Unis a publié un rapport intitulé *On the Full and Open Exchange of Scientific Data* qui serait le premier usage public du terme. Ce rapport réclame le partage des données « brutes », une notion qui vient des sciences mais qui est centrale dans les principes de l'open data, désignant les mesures telles qu'elles sont produites par les instruments sans avoir fait l'objet d'un traitement pour les corriger ou les « nettoyer ». Les *Science and Technology Studies* nous ont appris à prendre des précautions avec la notion de données brutes, souvent associée à un idéal d'objectivité. Répandant l'idée que la production de l'information serait transparente et naturelle, la notion de données brutes est un « oxymore et une mauvaise idée³⁹ ». Car de fait, l'étude de la production de données scientifiques rend compte de l'impossibilité d'identifier un moment où la donnée aurait connu un état brut, sans traitement⁴⁰. Les « données » sont toujours des « obtenues », pour reprendre l'expression de Bruno Latour⁴¹. Elles sont le fruit de choix, de théories et d'interventions et ne sont jamais pré-interprétatives comme le laisse penser l'expression « données brutes ».

Devenu la norme dans de nombreuses disciplines, le partage des données doit permettre la reproductibilité des résultats et l'émergence de nouveaux savoirs. Les bases de données sont ainsi devenues des publications scientifiques à part entière⁴². Dans ce contexte, il est

38. Harlan Yu & David G. Robinson, 2012, *op. cit.*

39. Geoffrey C. Bowker, *Memory Practices in the Sciences*, The MIT Press, 2006.

Lisa Gitelman, *“Raw Data” Is an Oxymoron*, The MIT Press, 2013.

40. David Ribes & Steven J. Jackson, « Data Bite Man: The Work of Sustaining a Long-Term Study » in Lisa Gitelman, *“Raw Data” is an Oxymoron*, The MIT Press, 2013, p. 147-166.

41. Bruno Latour, « Le “pédofil” de Boa-Vista : montage photo-philosophique », *La clef de Berlin et autre leçons d'un amateur de sciences*, La Découverte, 1993, p. 171-225.

42. Geoffrey C. Bowker, « Biodiversity Datadiversity », *Social Studies of Science*, 2000, vol. 30, n° 5, n° 643.

courant d'entendre que les sciences comme les organisations feraient face à un « déluge de données ». Or, ces perceptions apparaissent régulièrement depuis la Renaissance en même temps que les technologies de gestion de l'abondance d'information⁴³. Les articles relatant le « déluge de données » ont aussi charrié l'idée que le caractère massif et exhaustif des bases de données pourrait bouleverser la science. Dans un nouveau paradigme scientifique⁴⁴, les données toujours plus massives n'auraient qu'à « parler » pour interpréter les phénomènes et générer du savoir. Pour Chris Anderson, éditeur de *Wired*, le déluge des données rend la méthode scientifique obsolète⁴⁵. Dans un article célèbre, il n'annonçait pas moins que la fin de la théorie. Plutôt que de partir d'une hypothèse et de la tester, il suffirait d'explorer les données pour identifier des corrélations, des relations statistiques entre des phénomènes, et comprendre la causalité. Même si cette approche a été largement critiquée par la suite au sein des communautés scientifiques, elle a eu un impact auprès du grand public. La demande d'ouverture des données est entrée dans le débat politique la même année que la publication de cet article, en 2008. La possibilité de réutiliser l'ensemble des données du secteur public a été perçue comme un moyen de générer des recherches inédites et de créer de nouvelles formes de connaissances.

Concomitamment au mouvement de l'open data, la notion de science ouverte a émergé pour désigner la « *diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique* »⁴⁶. S'inscrivant dans la lignée des principes de l'accès ouvert aux publications scientifiques⁴⁷, elle a été définie notamment à la suite de la

43. Daniel Rosenberg, « Early Modern Information Overload », *Journal of the History of Ideas*, 2003, n° 64, p. 1-9.

44. Tony Hey, Stewart Tansley & Kristin Tollé, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 2009.

45. Chris Anderson, « The End of Theory: the Data Deluge Makes the Scientific Method Obsolete », *Wired*, juin 2008. <https://www.wired.com/2008/06/pb-theory/>.

46. Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, « Initiez-vous à la science ouverte », *Ouvrir la Science*, 2023. <https://www.ouvrir.lascience.fr/initiez-vous-a-la-science-ouverte>.

déclaration de Berlin de 2003 sur le libre accès à la connaissance. Signée aujourd'hui par plus de 700 établissements, celle-ci réclame l'accès libre à toutes les œuvres et données issues d'un travail de recherche. La déclaration a été complétée par la publication en 2007 d'un rapport de l'OCDE sur l'accès aux données issues des recherches financées sur fonds public⁴⁸. Plus de trente pays se sont engagés en ce sens. Le concept de science ouverte s'est ensuite diversifié en plusieurs écoles qui ont porté leur attention sur des objets différents⁴⁹ :

- l'école « publique » qui soutient que la science doit être accessible à un public plus large en ouvrant le processus de recherche (sciences citoyennes) et en favorisant la diffusion des produits de la recherche aux non-initiés;
- l'école « démocratique » qui promeut le libre accès aux produits de la recherche (publications, données, codes sources, graphiques...);
- l'école « pragmatique » qui, inspirée par des concepts d'innovation ouverte, considère la science ouverte comme une méthode permettant de rendre la recherche et la diffusion des connaissances plus efficaces;
- l'école « de l'infrastructure » qui s'intéresse à l'infrastructure technique qui permet les pratiques de recherche émergentes (informatique répartie, portails de données...);
- l'école « de la mesure » qui s'intéresse aux normes alternatives pour déterminer l'impact scientifique, notamment en lien avec la valorisation du partage de données dans les carrières.

47. Les principes de l'accès libre aux publications scientifiques ont été définis par l'initiative de Budapest pour l'accès ouvert de 2002 (BOAI, <https://www.budapestopenaccessinitiative.org>) et la déclaration de Bethesda pour l'édition en libre accès en 2003 (<https://www.ouvri.lascience.fr/declaration-de-bethesda-pour-ledition-en-libre-acces/>).

48. OCDE, *Déclaration de l'OCDE sur l'accès aux données de la recherche financée par des fonds publics*, 2004. <https://www.ouvri.lascience.fr/declaration-de-locde-sur-lacces-aux-donnees-de-la-recherche-financee-par-des-fonds-publics/>.

49. Benedikt Fecher & Sascha Friesike, « Open Science: One Term, Five Schools of Thought », in Bartling, S., Friesike, S. (eds), *Opening Science*, Springer, 2014. https://doi.org/10.1007/978-3-319-00026-8_2.

Sur le sujet des données, les principes FAIR⁵⁰ se sont imposés à partir d'une publication de 2016 dans *Nature — Scientific data* pour désigner un ensemble de principes directeurs pour gérer les données de la recherche qui doivent être :

- Faciles à trouver pour les humains et les machines avec des identifiants uniques et des métadonnées complètes;
- Accessibles pour stocker durablement les données et faciliter leur accès;
- Interopérables pour favoriser les échanges entre systèmes informatiques et augmenter les capacités des métadonnées à être combinées avec des langages et des formats ouverts et largement partagés;
- Réutilisables pour augmenter l'impact des données avec une description précise des données et de leurs conditions de réutilisations.

Ces principes sont aujourd'hui promus sur l'ensemble du cycle des données de la recherche, par exemple par l'Agence nationale de la recherche (ANR) en France qui, depuis 2019, demande à tous les projets financés de décrire la façon dont les données sont produites, documentées, (ré)utilisées, gérées et partagées pendant et après le projet afin de les rendre FAIR (Faciles à trouver, Accessibles, Interopérables, Réutilisables)⁵¹.

■ La modernisation des administrations

Au-delà des promesses de croissance, les politiques d'open data reposent également sur leur capacité à provoquer des changements dans les rouages de l'administration et ainsi engendrer sa «modernisation». Les projets d'ouverture des données publiques sont souvent soutenus par les services en charge de la transformation du service

50. Mark D. Wilkinson *et al.*, «The FAIR Guiding Principles for scientific data management and stewardship», *Scientific Data*, vol. 3, n°1, 2016. <https://doi.org/10.1038/sdata.2016.18>.

51. «FAQ Plan de Gestion des Données (PGD)», *Agence nationale de la recherche*, 2023. <https://anr.fr/fr/lanr/engagements/faq-pgd/>.

public qui ont fait le constat d'une organisation manquant de transversalité et laissant peu de place à l'initiative des agents. La réutilisation des données produites dans d'autres divisions de l'organisation encouragerait la fin du travail « en silos », une expression désignant le cloisonnement des services et l'insuffisance des pratiques de partage des informations. L'ouverture des données publiques offrirait la possibilité de créer des services innovants en utilisant des données qui n'étaient jusqu'alors pas connues ou qui ne circulaient pas hors du service qui les produisait.

L'ouverture des données encouragerait aussi la participation des citoyens dans l'élaboration des politiques par la réduction des asymétries d'information. En proposant des données dans leur plus grande précision, les politiques d'open data faciliteraient l'apparition d'une citoyenneté informée au fondement de politiques publiques d'inspiration libérale⁵². Cette ouverture permettrait dans le même mouvement l'émergence de contre-pouvoirs à même de remettre en cause de manière factuelle les politiques publiques en cours d'élaboration ou leur mise en œuvre. Des actions collectives se mobilisent régulièrement pour que les données prennent en compte des situations jusqu'alors occultées ou ignorées⁵³. Plus largement, l'ouverture de l'information stratégique au public est perçue comme une source d'efficacité et d'innovation des organisations. Les politiques d'open data s'inscrivent dans la continuité des travaux des théoriciens des organisations qui encouragent à diffuser l'information. En particulier, les travaux de Henry Chesbrough, à l'origine de la notion d'innovation ouverte⁵⁴, ont répandu l'idée qu'ouvrir l'information stratégique peut constituer une source d'innovation bien supérieure au régime classique du confinement des ressources en interne.

52. Andrew Barry, *Political Machines: Governing a Technological Society*, The Athlone Press, 2001.

53. Isabelle Bruno, Emmanuel Didier & Julien Prévieux, *Statactivisme. Comment lutter avec des nombres*, Zones, 2014.

54. Henry Chesbrough, *Open Business Models: How to Thrive in the New Innovation Landscape*, Harvard Business School Publishing, 2006.

Ces bénéfiques internes et externes positionnent les politiques d'ouverture des données comme un objet d'intérêt pour la transformation du service public. La Banque mondiale encourage ainsi financièrement les pays en développement à ouvrir leurs données à la fois pour faire émerger une société civile informée, mais aussi pour « moderniser » le fonctionnement des administrations. L'open data est devenu en quelque sorte un « produit d'exportation » pour les pays occidentaux qui proposent des méthodologies « clé en main » pour transformer l'administration par l'ouverture des données⁵⁵. Cette « modernisation » s'inscrit aussi dans le contexte de réduction des dépenses de l'État dans lequel l'ouverture des données financières permettrait un contrôle de l'action publique qui dépasse les instances dédiées et l'étend à la société civile dans son ensemble. Des recherches critiques ont alors associé l'ouverture des données publiques à la privatisation du service public⁵⁶. Cette critique doit être prise au sérieux : l'ouverture des données peut aussi servir à un désinvestissement de l'État dans les services à destination des citoyens. L'État peut se contenter d'ouvrir les données et déléguer au secteur privé la création de services, abandonnant alors le principe d'égalité d'accès au service public. L'éditeur Tim O'Reilly qui a participé à l'organisation de la conférence fondatrice des principes de l'open data (abordée en détail dans le chapitre suivant) a écrit en 2010 l'article très commenté « Government as a Platform⁵⁷ ». Dans une rhétorique libertarienne tirée d'exemples où la société se substitue à l'État, l'article défend une conception de l'État comme une plateforme mettant des ressources à disposition de la société civile et des acteurs privés. Ces derniers ont la liberté de développer des biens et des services finaux à l'aide de ces ressources, une approche qui peut créer les

55. Clare Birchall, « "Data.gov-in-a-box": Delimiting transparency », *European Journal of Social Theory*, 2015, vol. 18, n° 2, p. 185-202.

56. Jo Bates, « "This is what modern deregulation looks like": co-optation and contestation in the shaping of the UK's Open Government Data Initiative », *Journal of Community Informatics*, 2012, vol. 8, n° 2.

57. Tim O'Reilly, « Government as a Platform », *Innovations: Technology, Governance, Globalization*, MIT Press, 2011, vol. 6, n° 1, p. 13-40.

conditions d'un désengagement de l'État, et l'expérience l'a montré, d'un accroissement des inégalités.

■ Les industries de l'information

L'émergence des politiques d'open data doit aussi beaucoup aux entreprises pour lesquelles la réutilisation des données constitue le cœur de l'activité. L'industrie de l'information est particulièrement développée aux États-Unis où toutes les données publiques sont dans le domaine public, sans redevances ni restrictions, depuis 1988⁵⁸. Cette situation est le fruit d'une intense campagne de lobbying menée par les industriels de l'information qui ont obtenu du gouvernement fédéral que les données soient systématiquement partagées et que les services à valeur ajoutée soient fournis principalement par le secteur privé plutôt que par les acteurs publics. En France, la situation a longtemps été différente avec le succès des services « télématiques » sur Minitel qui ont dégagé d'importantes sources de revenus dans les années 1980 pour les acteurs publics. Maurice Ronai parle d'une prolifération et d'un « activisme télématique » dans les administrations.

Selon lui, les administrations exploitaient leurs gisements naturels : le vaste domaine des droits et démarches. Les serveurs proliféraient parfois à l'intérieur d'un même ministère. Des préoccupations nouvelles s'affirmaient : faire du trafic, générer des recettes. La diversité des pratiques tarifaires, les bricolages contractuels débouchèrent alors sur des situations ambiguës : monopoles de commercialisation, accords d'exclusivité, relations de troc, recours à des associations...

Même si le Minitel a progressivement disparu des foyers dans les années 1990 avec le développement d'Internet, ces recettes ont créé un précédent et ont donné l'impression que les données produites par l'administration française pouvaient être une source de revenus pour le financement du service public.

58. Maurice Ronai, « Données publiques : accès, diffusion, commercialisation », *Problèmes politiques et sociaux*, Éditions La Documentation française, 1997, n° 773-774, p. 68.

Les politiques d'open data ont fortement incité à l'abandon de ces redevances. Dans les années 2000, la Commission européenne s'est inspirée du cadre réglementaire états-unien pour promouvoir l'abandon des redevances et faire émerger une industrie de l'information de taille comparable à celle des États-Unis. En 2003, l'Union européenne a adopté la Directive sur les informations du secteur public dite PSI (*Public Sector Information*) pour inciter les États membres à diffuser gratuitement leurs données. La Commission européenne a multiplié les études sur le potentiel économique de la libération des données, évaluant jusqu'à 200 milliards d'euros par an la valeur de leur circulation optimale dans les pays de l'Union⁵⁹. Ces promesses de croissance doivent justifier les efforts financiers et humains (suppression des redevances, développement de portails, travail sur les données, communication et événements, recherche et développement) consentis par les États pour ouvrir leurs données. Le pari de l'open data repose sur un retour sur investissement par la taxation des revenus de la réutilisation dont il est prévu que le montant dépasserait largement le coût de l'ouverture et l'abandon d'éventuelles redevances.

Ces prévisions de croissance se sont transformées en revendication à la suite d'une campagne du quotidien britannique *The Guardian* en 2006 intitulée «*Free Our Data*», lancée par l'article «*Rendez-nous les joyaux de la couronne*». Elle exigeait la gratuité des données «payées par les impôts». Parmi ces «joyaux», les données de l'institut géographique britannique Ordnance Survey ont été perçues comme les plus prometteuses en termes de croissance. La tribune insistait sur le fait qu'aucun des services de cartographie numérique comme Google Maps n'était britannique du fait des fortes redevances pour utiliser les données de l'Ordnance Survey. Le *Guardian* évoquait alors la différence avec la conception américaine.

Les États-Unis considèrent que les données collectées avec l'argent du contribuable devraient lui être restituées gratuitement. Et une étude détaillée montre que l'attitude fermée du Royaume-Uni envers ses données

59. Graham Vickery, *Review of recent studies on PSI re-use and related market developments*, 2011.

*signifie perte des opportunités commerciales et même frein à la recherche scientifique dans des domaines comme le changement climatique*⁶⁰.

Cette campagne, menée par un des principaux quotidiens du pays, a créé des réseaux et mobilisé des acteurs importants dans le développement de l'open data en Grande-Bretagne.

Rappelons enfin que l'open data est apparu dans le débat public aux alentours de 2007, en même temps que le développement de l'économie des applications mobiles avec le lancement des boutiques d'applications sur smartphone (Apple AppStore, Google PlayStore...). Une vague de création de services et d'emploi dans le développement informatique était espérée avec l'accès ubiquitaire aux services rendu possible par le smartphone. Dans ce contexte, les données ouvertes étaient souvent envisagées comme un carburant de cette nouvelle économie. Chaque jeu de données ouvertes pouvait potentiellement donner lieu à la création d'une jeune pousse. L'argument de la création d'emplois par ces nouvelles entreprises a joué un rôle important dans le contexte de la crise économique de 2008, l'open data étant perçu comme un instrument de relance.

■ L'entraînement des modèles d'intelligence artificielle

Même si elle fait l'objet d'une actualité brûlante, la notion d'intelligence artificielle (IA) est un terme flou dont la définition est débattue et qui reste marqué par l'imaginaire de la science-fiction, coloré de peur et de fascination. Les acteurs techniques parlent plutôt d'apprentissage automatique ou *machine learning* qui, selon le site *Elements of AI*, est un sous-domaine de l'informatique désignant des systèmes qui améliorent leur performance pour une tâche donnée au fur et à mesure qu'ils acquièrent de l'expérience et des données⁶¹. Cette définition est très

60. Charles Arthur & Michael Cross, «Give us back our crown jewels», *The Guardian*, 9 mars 2006. <https://www.theguardian.com/technology/2006/mar/09/education.epublic>.

61. Elements of AI, «Qu'est-ce que l'IA?», 2023, <https://course.elementsofai.com/fr/1>.

proche de celle d'intelligence artificielle qui désigne des systèmes ayant deux caractéristiques, toujours selon *Elements of AI* :

- l'autonomie, la capacité d'exécuter des tâches dans des environnements complexes sans être constamment guidé par l'utilisateur ;
- l'adaptabilité, la capacité d'améliorer leur performance grâce à l'apprentissage et l'expérience.

Cette adaptabilité est, en grande partie, due aux larges ensembles de données qui ont permis l'entraînement et l'adaptation des modèles d'IA. Pour comprendre le rôle des données ouvertes dans ce processus d'entraînement, le *Contre-Atlas de l'intelligence artificielle* de Kate Crawford⁶² explore dans son chapitre «*Data*» la capture massive et vorace des données par l'industrie de l'IA. L'autrice affirme que l'IA n'est ni artificielle ni intelligente et, pour contrer le voile d'opacité de cette industrie, explore les ressources naturelles ou humaines, les infrastructures, la logistique et les histoires derrière ces technologies. L'industrie de l'IA s'est construite sur la collecte massive de données, considérant que toute forme d'image, de texte, de son ou de vidéo pouvait représenter des données brutes pour le test et l'entraînement de ses algorithmes. Historiquement, l'IA était conçue dans les années 1970 comme la réalisation de systèmes experts qui imitent le raisonnement d'un professionnel. De manière déterministe, les applications répondent de manière prévisible à l'action de l'utilisateur. À partir des années 1980, l'approche probabiliste s'est développée en utilisant des probabilités pour modéliser et résoudre des problèmes, tolérant l'incertitude et la méconnaissance de certaines règles métiers⁶³. Parmi ces techniques, l'apprentissage profond ou *deep learning* s'est développé sous l'impulsion de trois chercheurs en informatique (Yoshua Bengio, Yann Le Cun et Geoffrey Hinton) qui sont arrivés à

62. Kate Crawford, *Contre-atlas de l'intelligence artificielle. Une cartographie politique, sociale et environnementale de l'IA*, Zulma, 2022.

63. Dominique Cardon, Jean-Philippe Cointet & Antoine Mazières, «*La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle*», *Réseaux*, 2018, vol. 211, n° 5, p. 173-220. <https://doi.org/10.3917/res.211.0173>.

perfectionner cette technique et à convaincre la communauté scientifique au milieu des années 2000, après une longue période de désintérêt⁶⁴. Les trois chercheurs ont reçu ensemble le prix Turing, considéré l'équivalent du prix Nobel en informatique, pour leur contribution «*d'une importance technique majeure et durable dans le domaine informatique*». Les réseaux de neurones s'inspirent du cerveau humain en utilisant une structure de réseaux neuronaux artificiels organisés par couches. Les données circulent de la couche d'entrée à travers plusieurs couches de réseaux neuronaux cachés, «profonds», avant d'arriver à la couche de sortie. L'organisation interne de ces réseaux se réalise lors d'une phase d'apprentissage au cours de laquelle la machine est exposée à de grands volumes de données qui reconfigurent la forme du réseau et modifient le poids de chaque neurone⁶⁵.

Kate Crawford explique que les données publiques ont constitué, avant l'arrivée d'Internet, un des principaux gisements de données pour l'industrie émergente de l'IA. Elle étudie en particulier les corpus de texte en Traitement automatique du langage (TAL) pour lesquels les chercheurs ont dû faire preuve d'inventivité pour identifier des sources de données textuelles massives. Elle donne l'exemple d'IBM qui, au milieu des années 1980, cherchait des données pour entraîner ses modèles de reconnaissance de la parole et de traduction. L'entreprise a fini par trouver une source de données inattendue dans le procès pour démantèlement *antitrust* d'IBM qui s'est déroulé à partir de 1969 et qui a duré 13 ans avec plus d'un millier de personnes auditionnées. Le procès Enron a constitué une autre source de données très exploitée par les chercheurs en TAL. La commission fédérale de régulation de l'énergie avait estimé que la saisie des courriels de 158 employés devait être diffusée à des fins de transparence. Près d'un demi-million d'échanges de messages en langage naturel, sans

64. CScience, «Les pères fondateurs du deep learning: Yoshua Bengio, Yann Le Cun et Geoffrey Hinton», 2020. <https://www.cscience.ca/2020/07/28/les-peres-fondateurs-du-deep-learning-yoshua-bengio-yann-le-cun-et-geoffrey-hinton/>.

65. Pablo Jensen, *Deep earnings. Le néolibéralisme au cœur des réseaux de neurones*, C&F éditions, 2021.

formalisme, étaient alors publiés et ces données ont servi de matériau d'entraînement aux outils d'IA. L'autrice souligne que l'industrie considère souvent ces corpus de texte comme interchangeables, ne se préoccupant généralement pas des biais et des manques notamment en termes de diversité linguistique. Sur ce point, le cadre juridique des États-Unis, évoqué précédemment, qui place les données dans le domaine public et se préoccupe bien peu de la protection des données à caractère personnel, explique en partie «l'avance» des États-Unis dans le domaine de l'IA⁶⁶.

L'arrivée d'Internet a tout changé en ce qui concerne l'accès aux données d'entraînement. Kate Crawford dénonce un «pillage» : tout ce qui est publié en ligne peut devenir donnée d'entraînement pour les systèmes d'IA. Elle donne l'exemple d'ImageNet, une base d'aujourd'hui 14 millions d'images annotées, qui a servi à l'entraînement des principaux modèles et travaux de recherche sur la vision par ordinateur et la reconnaissance des objets présents dans les images⁶⁷. Créée en 2006 par la chercheuse en informatique Fei-Fei Li, aujourd'hui professeure à Stanford, ImageNet s'appuie sur la base de données lexicale ouverte WordNet qu'elle complète de millions d'images collectées sur le Web et les réseaux sociaux sans préoccupation de la vie privée. Le processus d'annotation de ces images, l'association des images à des mots sélectionnés dans un vocabulaire pré-établi, a été réalisé par des «travailleurs du clic» payés quelques centimes de l'heure et recrutés sur la plateforme Amazon Mechanical Turk⁶⁸. Même si les limites du copyright sont souvent franchies, la

66. Dans le domaine de la reconnaissance faciale, Simone Pieranni explique le *leadership* de la Chine par la généralisation de la surveillance de la population chinoise, fournissant des données massives, et des conditions permissives d'exploitation des données des caméras de surveillance déployées partout dans le monde par les entreprises chinoises dans le cadre de l'initiative des nouvelles routes de la soie.

Simone Pieranni, *Red Mirror. L'avenir s'écrit en Chine*, C&F éditions, 2021.

67. Dave Gershgorn, «The data that transformed AI research—and possibly the world», *Quartz*, 2017. <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world>.

68. Antonio A. Casilli, *En attendant les robots : Enquête sur le travail du clic*, Seuil, 2019.

communauté des chercheurs en IA met à disposition de nombreuses bases de données d'entraînement en open data pour stimuler la recherche et optimiser les technologies d'apprentissage. Le site [kaggle.com](https://www.kaggle.com) propose plusieurs milliers de jeux de données, issus d'extractions de page web ou de données publiques ouvertes, pouvant être utilisés pour entraîner des modèles et participer à des compétitions de *data science*. Le *fair use* ou « usage légitime » est souvent invoqué comme exception aux droits d'auteur pour l'entraînement des modèles. Mais cet argument juridique est de plus en plus contesté alors que les modèles génératifs de texte, d'images, de sons ou de vidéos imitent les auteurs initiaux voire, dans certains cas, les concurrencent sur leurs marchés⁶⁹.

Parmi ces données ouvertes utilisées à des fins d'entraînement de modèles d'IA, Common Crawl a récemment été mis en lumière avec le lancement en novembre 2022 de ChatGPT. Le modèle de langage GPT, sous le capot du chatbot d'OpenAI, a été entraîné avec des corpus de texte issus à plus de 60% de Common Crawl, 22% de WebText2 comprenant les publications les mieux notées sur Reddit, des données issues de livres et à 3% issues de Wikipédia⁷⁰. OpenAI, qui n'a plus d'*open* que le nom, déclare aujourd'hui rejeter l'open source et ne publie plus le détail des données utilisées pour l'entraînement des nouvelles versions de ChatGPT⁷¹. Common Crawl a été constitué à partir de 2007 comme un projet à but non lucratif visant à rendre accessible aux chercheurs l'extraction, la transformation et l'analyse des données du Web. En 2023, les données ouvertes de Common Crawl contiennent le texte de 240 millions de pages web

69. James Vincent, « The scary truth about AI copyright is nobody knows what will happen next », *The Verge*, 2022. <https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data>.

70. Dennis Layton, « ChatGPT — Show me the Data Sources », *Medium*, 2023. <https://medium.com/@dlaytonj2/chatgpt-show-me-the-data-sources-11e9433d57e8>.

71. Julien Lausson, « “Nous avons tort” : en dévoilant GPT-4, OpenAI dit rejeter l'open source désormais », *Numerama*, 2023. <https://www.numerama.com/tech/1307322-nous-avons-tort-en-devoilant-gpt-4-openai-dit-rejeter-lopen-source-désormais.html>.

collectées sur 16 années. L'organisation, qui compte aujourd'hui huit salariés, vante les vertus de l'open data pour améliorer la prise de décision individuelle ou gouvernementale, favoriser les collaborations interdisciplinaires et résoudre des problèmes complexes⁷². Derrière ces objectifs on ne peut plus louables, Common Crawl est dirigée par un rassemblement d'acteurs influents de la Silicon Valley. Si l'on consulte les profils des membres de son conseil d'administration, on trouve des avocats en propriété intellectuelle, des consultants, des communicants et des *data scientists*, tous très insérés dans l'économie de la Silicon Valley. On trouve aussi trois représentants de Google comme Peter Norving, directeur de recherche, l'ingénieur de recherche Pete Warden⁷³, ou encore Praveen Paritosh, *senior researcher* en IA. Enfin, la présence de Kurt Bollacker confirme l'étonnante convergence entre cyberculture et contre-culture qui a animé la Silicon Valley à partir des années 1970, autour de la figure de Brand⁷⁴. Bollacker est *digital research director* de la Long Now Foundation fondée et présidée par Stewart Brand pour promouvoir une vision de long terme des problèmes mondiaux. Il est aussi conseiller stratégique de MLCommons où il anime le groupe de travail sur les jeux de données d'entraînement pour le *machine learning* (ML). Cette organisation qui se revendique du mouvement des communs de la connaissance compte parmi ses membres fondateurs les principaux géants de la technologie mondiale (Alibaba, Baidu, Google, Intel, Meta, Microsoft, Nvidia...) pour échanger sur les données d'entraînement et partager des benchmarks de performance des technologies d'IA.

Cette mention du mouvement des communs dans le nom MLCommons fait écho à l'avertissement de Kate Crawford sur la

72. Common Crawl, «Common Crawl - Mission», 2023. <https://commoncrawl.org/mission>.

73. Il est aussi auteur d'un manuel sur la recherche de données publiques, publié aux éditions O'Reilly, maison qui a accueilli la rencontre fondatrice des principes de l'open data comme nous le verrons dans le chapitre suivant. Pete Warden, *Data Source Handbook: A Guide to Public Data*, O'Reilly, 2011.

74. Fred Turner, *Aux sources de l'utopie numérique : De la contre-culture à la cyberculture*, Stewart Brand, un homme d'influence, C&F Éditions, 2013 (2^e ed. 2021).

«capture des communs» par l'industrie de l'IA. On pourrait débattre de la caractérisation de communs pour l'ensemble des données du Web. À la suite en particulier des travaux d'Elinor Ostrom⁷⁵, les communs désignent des ressources partagées, gérées, et maintenues collectivement par une communauté. David Bollier le résume bien : «*un commun, c'est : une ressource + une communauté + un ensemble de règles sociales. Ces trois éléments doivent être conçus comme formant un ensemble intégré et cohérent*⁷⁶». Le fait de partager une ressource ne suffit pas à la qualifier de communs mais toujours est-il que les acteurs industriels s'en sont saisis comme un slogan. Crawford dénonce un mouvement d'enclosure du fait qu'«*un jeu de données peut toujours être disponible mais la méta-valeur des données — le modèle créé avec — est détenue par le privé*». À propos de projets ouverts comme ImageNet ou Common Crawl, elle évoque un «mythe» de la collecte de données comme une pratique charitable en informatique qui obscurcit les rapports de pouvoirs et protège ceux qui en profitent le plus. Cet avertissement interroge le mouvement de l'open data. Certains acteurs (cela m'est arrivé) peuvent voir l'ouverture comme un «bien en soi» c'est-à-dire, au sens de Nicolas Dodier, une visée qui, en tant que telle, pourrait clore l'argumentation aux yeux de ses énonciateurs⁷⁷. Les propos de Kate Crawford résonnent aussi avec ceux du sociologue espagnol César Rendueles⁷⁸. Dans son dialogue sur les communs avec le sociologue et homme politique Joan Subirats, cofondateur du mouvement Barcelone en commun qui a codirigé la capitale catalane, Rendueles critique l'usage médiatique des communs comme «*un concept sympathique et consensuel, peu suspect de*

75. Elinor Ostrom, *Gouvernance des biens communs : Pour une nouvelle approche des ressources naturelles*, De Boeck Supérieur, 2010. Elinor Ostrom, *Discours de Stockholm, en réception du Nobel d'économie 2009*, C&F éditions, 2020.

76. David Bollier, *La renaissance des communs pour une société de coopération et de partage*, Charles Léopold Mayer, 2014.

77. Nicolas Dodier, «L'espace et le mouvement du sens critique», *Annales. Histoire, sciences sociales*, 60-1, p. 7-31, 2005.

78. César Rendueles & Joan Subirats, *La cité en communs. Des biens communs au municipalisme*, C&F éditions, 2020.

complicité avec la bureaucratie et le marché et qu'on accueille toujours chaleureusement». Il avertit que cette apparente sympathie du concept peut mener à des « *affinités monstrueuses* » et à coïncider avec certains intérêts privés. La récupération des communs n'est pas l'apanage de l'industrie de l'IA, elle est aussi courante dans la sphère politique qui, depuis quelques années, en fait un usage immodéré. Certains comme Lionel Maurel désignent cette récupération politique sous le terme de « *commons washing*⁷⁹ », un blanchiment qui peut parfois délayer le concept de sa substance.

L'avertissement de Kate Crawford sur le risque d'enclosure des communs du Web par l'industrie de l'IA prend tout son sens quand on s'intéresse à un acteur comme OpenAI qui a progressivement abandonné toute idée d'ouverture de ses modèles. Elle mérite néanmoins d'être nuancée car la communauté des chercheurs en IA continue de diffuser des modèles ouverts. Par exemple, des chercheurs de Google ont introduit en 2017 dans un article en accès ouvert le modèle des *transformers*. Cette méthode de traitement du langage, qui introduit des mécanismes dits d'« attention » pour prendre en compte le contexte précédent et suivant chaque mot d'une phrase, a constitué le fondement de l'émergence des grands modèles de langage (abrégé LLM de l'anglais « *large language model* »). Google a rendu public en 2018 BERT, un modèle de traitement du langage naturel (NLP) publié sous une licence ouverte autorisant les usages commerciaux, qui a été pré-entraîné sur de grandes quantités de textes issues de Wikipédia notamment et a longtemps été une référence pour traiter des données textuelles⁸⁰. Le modèle des *transformers* a servi de fondement à GPT, le modèle de langage derrière ChatGPT, dont la deuxième version beaucoup plus réduite était open source. Après la sortie de ChatGPT fin 2022, de nombreux LLM open source ont été publiés à un rythme

79. Calimaq, « Leurs Communs numériques ne sont (toujours) pas les nôtres! », *S.I.Lex*, 2020. <https://scinfolex.com/2020/04/12/leurs-communs-numeriques-ne-sont-toujours-pas-les-notres/>.

80. Pierre-Carl Langlais, « ChatGPT : comment ça marche? », *Sciences communes*, 2023. <https://scoms.hypotheses.org/1059>.

soutenu⁸¹. Parmi ceux-ci, on peut citer Llama publié par l'équipe de recherche en IA de Meta (précédemment Facebook) dirigée par le Français Yann Le Cun. La version 2 autorise les usages commerciaux mais contrevient aux principes de l'open source en interdisant Llama pour les entreprises de plus de 700 millions d'utilisateurs (Google ou Amazon sont ciblés) et l'entraînement d'autres IA⁸². Cette dernière condition rappelle la situation ironique d'OpenAI qui, après avoir largement exploité des données du Web sans se préoccuper du droit d'auteur, voit émerger de sérieux concurrents à GPT comme Vicuna. Ce modèle ouvert a été entraîné en partie avec les données de ShareGPT, un site qui permet à quiconque de partager les données de ses échanges avec ChatGPT. Sur le marché émergent des modèles d'IA ouverts, la start-up franco-américaine Hugging Face a récemment levé 235 millions de dollars pour développer sa plateforme et sa communauté facilitant le partage de modèles ouverts et de données d'entraînement. Après une levée de fonds de 100 millions d'euros en quelques jours à l'été 2023, suivie d'une deuxième levée de fonds record de 385 millions d'euros fin 2023⁸³, une équipe de chercheurs français issus de Google DeepMind et de Meta a lancé Mistral AI, pour devenir leader mondial des LLM open source⁸⁴. Ces modèles ouverts pourraient rapidement dépasser les capacités de ceux d'OpenAI et des autres acteurs majeurs de l'industrie, comme a pu le faire Stable Diffusion pour la génération d'images après avoir été entraîné avec les données d'images ouvertes par le projet LAION.

81. Pour suivre cette actualité, Dataactivist et Opsi ont lancé en 2023 un blog et une communauté sur le sujet des LLM ouverts : <https://www.lebonllm.fr/>.

82. Benj Edwards, «Meta launches Llama 2, a source-available AI model that allows commercial applications», *Ars Technica*, 2023. <https://arstechnica.com/information-technology/2023/07/meta-launches-llama-2-an-open-source-ai-model-that-allows-commercial-applications/>.

83. AFP, «La start-up française Mistral AI a levé 385 millions d'euros», *Le Monde*, 2023. https://www.lemonde.fr/economie/article/2023/12/11/la-start-up-francaise-mistral-ai-a-leve-385-millions-d-euros_6205065_3234.html.

84. «Mistral AI dégage sa première alternative à ChatGPT», *Maddynews*, 2023. <https://www.maddynews.com/2023/09/27/mistral-ai-premiere-alternative-chatgpt/>.

Cette situation inquiète au plus point des géants du secteur comme Google dont un mémo interne écrit par un chercheur a fuité en mai 2023⁸⁵ pour alerter sur l'avantage compétitif acquis par les modèles ouverts en quelques mois après la sortie de ChatGPT :

La vérité désagréable est que nous ne sommes pas en position de gagner cette course aux armements, et OpenAI non plus. Pendant que nous nous chamaillions, une troisième faction a tranquillement mangé notre déjeuner. Je parle, bien sûr, de l'open source. En clair, ils nous dépassent. [...] Si nos modèles conservent un léger avantage en termes de qualité, l'écart se réduit étonnamment vite. Les modèles open source sont plus rapides, plus personnalisables, plus privés et plus performants. Ils font des choses avec 100\$ et 13 milliards de paramètres que nous avons du mal à faire avec 10 millions de dollars et 540 milliards de paramètres. Et ils le font en quelques semaines, pas en quelques mois. Cela a de profondes implications pour nous. Nous n'avons pas de recette secrète. Notre meilleur espoir est d'apprendre et de collaborer avec ce que d'autres font en dehors de Google.

Ces modèles ouverts pourraient donc progressivement faire évoluer les positions dominantes dans l'industrie de l'IA tout en créant de nouveaux oligopoles.

■ Ouvrir des données : un projet consensuel ?

Tout au long de ce chapitre, nous avons pu voir que l'ouverture des données s'inscrit dans la lignée d'une diversité de filiations, chacune mobilisant des acteurs et arguments différents. Loin d'être un phénomène récent, l'ouverture des données s'inscrit dans un héritage historique et conceptuel varié, comprenant la transparence et le droit d'accès à l'information, la liberté de l'information et le logiciel libre, le partage des données de la recherche, la modernisation des administrations, les industries de l'information et l'entraînement des modèles

85. Google «We Have No Moat, And Neither Does OpenAI», *SemiAnalysis*, 2023, <https://www.semianalysis.com/p/google-we-have-no-moat-and-neither>.

d'intelligence artificielle. Cette pluralité des facettes de l'ouverture des données se retrouve tout au long de l'histoire des politiques d'open data que nous allons aborder dans les chapitres suivants, d'abord dans l'émergence des principes de l'open data et ensuite dans sa mise en œuvre en France. Du fait de ces multiples facettes, l'ouverture des données a pu créer une forme de consensus transpartisan comme l'a constaté en 2012 la chercheuse Valérie Peugeot en listant les affiliations politiques des élus ayant adopté une démarche d'ouverture des données publiques⁸⁶. Dix ans après la publication de cet article, le constat est toujours valable : les politiques d'ouverture des données traversent l'échiquier politique. D'après Valérie Peugeot, l'ouverture des données présente les caractéristiques d'une innovation démocratique remarquable du fait de la diversité d'échelles territoriales engagées dans la démarche (villes, départements, régions, états, organisations internationales) et de la rapidité de son adoption par les autorités politiques. Surtout, cette innovation démocratique se distingue selon elle par son caractère consensuel car *« la couleur politique du gouvernement en place au moment de son adoption n'apparaît pas comme un caractère discriminant »* et *« lorsqu'un changement politique est intervenu, aucun gouvernement n'a cherché à remettre en question ce choix politique »*. Cette « convergence en trompe-l'œil » masque selon elle une double lecture de l'ouverture des données :

- une approche libertarienne selon laquelle ouvrir les données constitue une opportunité pour *« rendre au marché des fonctions tenues par l'État et donc de contribuer à une réduction de la fiscalité, à une diminution globale du rôle de la puissance publique dans la société »* ;
- une approche issue des biens communs informationnels qui consiste *« non pas à défaire l'État mais à en compléter l'action par des formes de gouvernances auto-organisées »*.

86. Valérie Peugeot, « L'ouverture des données publiques : convergence ou malentendu politique ? » in Bernard Stiegler, *Confiance, croyance, crédit dans les mondes industriels*, FYP Éditions, 2012, p. 1-9.

Pour analyser une politique d'ouverture des données, il faudra donc regarder au-delà de la simple décision d'ouvrir les données et s'intéresser au détail des choix réalisés, comme les types de jeux de données mis à disposition, leur contenu ou les incitations à la réutilisation des données. La question du choix des licences encadrant la réutilisation des données constitue le premier de ces choix qui peuvent « teinter » politiquement une stratégie d'open data en faveur d'une lecture ou de l'autre. C'est cette question de l'encadrement juridique qui a d'abord été traitée dans la définition des grands principes de l'open data.

Colophon

Cet ouvrage est composé par Hervé le Crosnier et Nicolas Taffin en HTML et CSS selon la spécification pour les médias paginés, avec l'aide de PagedJS (<https://pagedjs.org>). Ils remercient Valérie Peugeot et André Sintzoff pour leur relecture, ainsi que Julien Taquet pour son aide.

Le caractère utilisé pour le texte est Cala de Dieter Hofrichter qui s'est inspiré de Nicolas Jenson et de la Renaissance vénitienne, pour la transposer au présent numérique, avec un rythme, une rondeur et une lisibilité contemporaines. Utilisé pour le code ou les liens, le caractère Attribute text de Viktor Nübel est une des plus belles typographies du mouvement « faux monospace » qui reprend les formes des lettres à espacement constant des terminaux informatiques en y ajoutant un peu de la proportionnalité typographique.

ISBN 978-2-37662-071-6

Achévé d'imprimer en janvier 2024

par Présence Graphique à Monts (37)

Dépôt légal janvier 2024

Samuel Goëta

Avec une préface d'Axelle Lemaire

LES DONNÉES DE LA DÉMOCRATIE

Open data, pouvoirs et contre-pouvoirs

Apparu il y a une vingtaine d'années dans les sciences, le mouvement de l'open data promeut le libre usage des données, matière première de l'information.

Gouverner, c'est disposer de données pour choisir les orientations et évaluer les résultats. Loin du culte du secret et du capitalisme de surveillance, l'ouverture des données contribue à créer de nouveaux services numériques, améliorer l'efficacité des services publics et alimenter le débat citoyen. En France, la loi pour une République numérique de 2016 a fait de cette ouverture une norme pour les administrations.

Quelles sont les motivations politiques, économiques et informationnelles de ce mouvement ? Comment médias, entreprises et société civile s'emparent-ils des données ? Comment la culture des données peut-elle devenir un outil au service de la démocratie et des contre-pouvoirs ?

Ponctué de cas d'usage et particulièrement accessible, cet ouvrage de Samuel Goëta présente les origines du mouvement open data, pose un bilan critique de ses réalisations et propose des pistes d'action pour que les données ouvertes renforcent la démocratie.



Société numérique



9 782376 620716 >

27 € – imprimé en France
ISBN 978-2-37662-071-6
<https://cfeditions.com>