

Les enjeux stratégiques des data : le cas de l'open data



Sous la direction de Christian Harbulot

Table des matières

1. Contexte	5
2. Géopolitique des données et rapport de forces : quelles logiques d'influence ?	7
2.1. Géopolitique des données	7
2.2. La souveraineté numérique à la croisée des mondes matériel / immatériel	9
2.3. USA : les GAFAM et leur stratégie de conquête du monde immatériel	10
2.4. Chine, BATXH : quand les data servent le système	13
2.5. Où en est l'Europe ?	15
2.5.1 Les Etats-Unis et le Cloud Act	15
2.5.2. RGPD : l'arme européenne pour faire face au Cloud Act ?	16
2.6 La législation française et européenne : rétrospective	19
3. Les acteurs : comment mettent-ils en œuvre les projets d'open data ?	25
3.1 Les acteurs en France	25
3.1.1. Les collectivités	25
3.1.2. Les institutions publiques et administrations	27
3.1.3. Comment les entreprises exploitent-elles l'open data ?	28
3.1.4. Les Data brokers	29
3.1.5. Les associations & collectifs	29
3.2 Le cas des énergéticiens en Europe	31
3.3. Business model	32
3.4. Des exemples réussis de projets d'open data	36
3.4.1 En France	36
3.4.2 Aux Etats-Unis	37
3.5. Contrôle et vérification des données	37
4. L'impact de l'open data sur les questions de souveraineté	38
4.1. Souveraineté économique : risque d'attaque et dépeçage des data	38
4.2. La monétisation des données	39
4.3. L'IA	47
4.4. Self data	47
5. Conclusion	49
6. Bibliographie	51
7. Glossaire	53
8. Echiquiers	57

Introduction

Les "datas" ou données se posent comme l'un des challenges de la décennie à venir. Avec l'avènement du monde immatériel, le changement de paradigme conduit à une "dépendance invisible" entre états, via une domination économique, normative et culturelle, cela sans pouvoir être taxé d'ingérence ou de manœuvre coercitive visible : d'un Etat nation vers un Etat Plateforme. Les Etats-Unis et la Chine ne s'y sont pas trompés.

Le développement d'Internet et des outils numériques en cela y joue le rôle d'un accélérateur. C'est une source quasi inépuisable de données et le numérique offre de nouvelles opportunités pour les exploiter et faciliter leur circulation. La donnée n'est rien sans le partage et l'analyse qui en est faite. Plus qu'une tendance et véritable enjeu pour les institutions, les data ou données ouvertes sont considérées comme un vecteur de transparence, de lien social ou encore de croissance. La France a attendu 2016 pour s'engager dans la mise en œuvre d'une démarche open data sous la pression de l'Union Européenne. S'en est suivi un engouement dans certaines villes ou régions françaises, mais loin des ambitions affichées par les pouvoirs publics.

Ces dernières années, les entreprises ont eu à faire face aux nouvelles réglementations liées au numérique et à l'évolution rapide de leur environnement commercial. La collecte des métadonnées est une pratique stratégique. Il n'y a qu'à se pencher sur la lutte acharnée que se livrent la Chine et les Etats-Unis. L'enjeu pour ces deux nations est clairement l'acquisition d'une suprématie. Or, ces ressorts manquent cruellement en France.

Tout individu génère chaque jour de plus en plus de données, à mesure que les points de contact se multiplient. En 2019, la donnée s'est imposée comme la ressource principale. L'information exacte actualisée et précise est le nouvel eldorado. Son exploitation à travers l'Intelligence Artificielle (IA), la gouvernance ou le "deep learning" est devenue le levier de performance par excellence dont disposent les entreprises, tout secteur confondu. Créer de la valeur revient ici à placer la donnée au cœur du projet d'entreprise et des business models associés. Les données étant de plus en plus disponibles en temps réel, leur analyse a poussé les acteurs à être plus à l'écoute des changements rapides occasionnés par le Big Data et initier les évolutions en se montrant plus innovants et plus ingénieux.

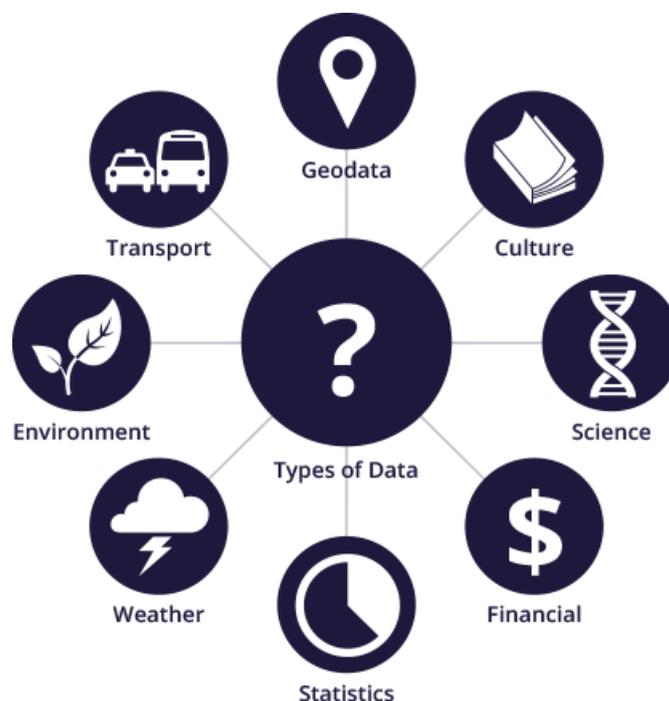
L'exploitation des données s'inscrit dans une continuité du rapport de force entre les GAFAM et BATX d'un côté et les Etats aussi bien que pour les entreprises du monde matériel. L'open data n'y échappe pas. L'open data apporte une vraie valeur ajoutée aux écosystèmes. Déjà de nombreux secteurs ont et/ou vont opérer leur transformation digitale. Les données présentent avantages et inconvénients. Le secteur de la santé en est un exemple concret : partager les données des patients tout au long de leur parcours santé a pour but d'améliorer la performance des soins et des traitements médicaux. Pourtant, l'émergence de ces nouvelles fonctionnalités suscitent des inquiétudes voire des oppositions, légitimes ou non.

Des résistances s'organisent. Des freins persistent. Certains acteurs demeurent réticents à l'adoption d'une stratégie orientée par la donnée ou data-driven. Ces volumes de données posent des questions de sécurité, dont on ressent davantage l'urgence dans certains domaines, tels que la finance ou la cybersécurité. Les déboires de certains grands groupes,

dont des données sensibles ont été exposées par inadvertance et d'autres piratées, ont souligné l'importance d'une structure qui encadre le stockage et l'exploitation des données. La data gouvernance, et la réglementation en matière de données (RGPD) se sont imposées comme un nécessaire bouclier défensif dans un environnement en pleine mutation et sous pression.

Des tendances de fond émergent comme le self data, l'usage de l'intelligence artificielle et du machine learning. L'open data est en train d'évoluer d'une démarche "ouverte/gratuite" vers un dispositif de "commercialisation/privée" sous impulsion de l'Europe. La monétisation pourrait être un contre-pouvoir des internautes comme des entreprises françaises et européennes. C'est une révolution dont les enjeux en termes de souveraineté et de stratégie de conquête sont évidents, même s'il est difficile aujourd'hui d'en dessiner les contours. Au cœur de cette nouvelle donne se trouve l'individu et le comportement humain. Les GAFAM ou autres BATX, géants du numérique américains ou chinois ayant marqué cette décennie que sont Google, Amazon, Facebook, Microsoft, Alibaba et à présent les NATU - Netflix, Airbnb, Tesla, Uber - l'ont bien compris. Ils contraignent par leur emprise tout l'écosystème (particuliers, entreprises et collectivités) à accélérer le rythme.

De vraies questions demeurent. Quels en sont les réels bénéfices ? Est-ce une stratégie efficace de partager les données ? Pourquoi l'open data s'est développé à plusieurs vitesses selon les pays ? Pourquoi les modèles économiques de l'open data existent-ils sur le papier alors qu'ils n'ont pas réellement fait leur preuve dans la réalité ? Est-ce un réel enjeu de souveraineté économique ou de puissance ? Une chose est sûre cependant demain, le succès reposera sur une meilleure exploitation de la donnée.



1. Contexte

L'open data est né aux Etats-Unis, en 1995, dans le domaine de la recherche scientifique. Le terme apparaît pour la première fois dans le National Research Council. L'idée est alors d'ouvrir les données relatives à l'environnement et à la géophysique. Les auteurs du rapport évoquent un échange ouvert d'informations scientifiques entre les pays. L'autre idée est la peur d'une privatisation des connaissances.

L'open data vise donc à mettre à disposition, au plus grand nombre, un bien commun : la connaissance. Les informations appartiennent donc à tout un chacun.

En décembre 2007, une trentaine de militants, entrepreneurs et chercheurs se réunissent en Californie. Leur but : définir le concept de donnée ouverte. Parmi eux se trouvent Tim O'Reilly et Lawrence Lessig. Tim O'Reilly est à l'origine du terme « web 2.0 ». C'est lui qui l'a défini puis rendu populaire. Il lance également le mouvement de l'open source. Lawrence Lessig est quant à lui à l'origine des licences Creative Commons, qui forment un des socles de l'open source.

Tim O'Reilly mentionne « l'open government », qui ouvrirait ses informations gouvernementales au public en open source. L'open data est né. Il concerne alors des données publiques issues du gouvernement.

Mais la volonté de ce groupe d'activistes est surtout de faire adopter cette définition et ce mouvement par le futur président des Etats-Unis. Dès son élection, Barack Obama met l'open government au centre de sa stratégie digitale.

Dans son Open Government Initiative, puis l'Affordable Care Act de 2010, les Etats-Unis autorisent l'ouverture des données de santé. Le gouvernement américain dispose d'ailleurs d'une plateforme en open data dédiée à la santé : healthdata.gov.

En 2011, les Etats-Unis ouvrent un portail national pour héberger les données publiées par les administrations. Quatre ans plus tard, Data.gov est menacé de fermeture.

La même année, le président américain lance l'Open Government Partnership en partenariat avec sept pays et des acteurs de la société civile. Parmi les plus gros donateurs, se trouvent le Royaume-Uni et les Etats-Unis, ainsi que diverses fondations comme Luminare Group, la Ford Foundation, The William and Flora Hewlett Foundation et Open Society Foundations. L'Open Government Partnership compte aujourd'hui 80 pays membres, dont la France.

En 2013, Barack Obama réaffirme la place de l'open data au cœur de sa politique et lance le concept de « smart government ». En mai, l'administration américaine annonce le partage de nouvelles données publiques sur data.gov. Près de 500 000 jeux de données, publiées par 180 administrations américaines, sont alors à la disposition des internautes.

Si l'open data est né à l'instar de l'open source d'une volonté de partage des connaissances et d'une protection face à une possible privatisation des informations dans le milieu scientifique, le mouvement bascule ensuite vers un outil de gestion en faveur d'une meilleure communication entre administrations et citoyens. Les données publiques peuvent être accessibles, librement, par tous. Certains y ont vu un souhait de transparence, d'autres un mouvement citoyen vers une démocratie plus participative.

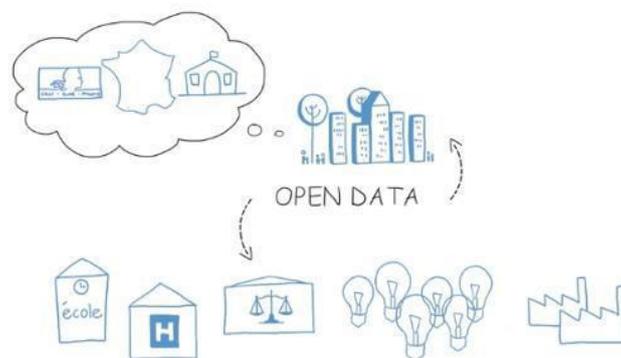
Au fil des années, l'open data s'est déplacé vers le monde des entreprises, tout en évoluant d'un modèle "ouvert/gratuit" vers une monétisation des données.

Le Support Center of Data Sharing financé par l'UE et mis en ligne fin juillet 2019, est dirigé par un consortium de trois entreprises : Capgemini, Fraunhofer Fokus (un institut allemand spécialisé en "open communications") and Time Lex (un cabinet d'avocats spécialisé dans les data et les données personnelles, basé à Bruxelles).

Derrière ce consortium, on peut se demander si le mouvement de l'open data n'est pas en réalité porté par des entreprises, et non des états, qui ont vu l'intérêt de collecter des données à grandes échelles sur des personnes. Qui peut mieux fournir des données de masse que des états et des collectivités ?

« Les efforts de la Commission européenne sur l'open data s'étaient jusqu'à maintenant concentrés sur les données publiques, les données collectées par les Etats. Ce sont des données qui peuvent avoir un impact fort sur l'économie, mais il existe encore un fort potentiel de développement si les entreprises partagent à leur tour leurs données dans un but B2B ou Business vers Gouvernement », explique Dinand Tinholt, vice-président en charge de l'Union Européenne chez Capgemini.

« Il y a bien évidemment des freins liés à la concurrence, avec des données considérées comme sensibles qu'elles ne souhaitent pas partager » souligne Dinand Tinholt. *« D'autre part, les entreprises sont encore mal informées sur la façon dont elles peuvent partager des données de manière sûre. De multiples questions se posent quant aux modalités techniques et légales du partage de données au-delà des frontières de l'entreprise et, dans beaucoup de cas, les entreprises ne connaissent pas la façon dont elles peuvent aller vers l'open data. »*



2. Géopolitique des données et rapport de forces : quelles logiques d'influence ?

2.1. Géopolitique des données

Nous produisons 2.5 quintillions de données par jour, dans le monde aujourd'hui. Cette nouvelle manne fait l'objet de convoitises et de stratégies de la part des entreprises qui cherchent à les exploiter. Extraire, identifier, croiser pour créer finalement de l'information, voire de la connaissance, donc un produit monétisable. Ces richesses deviennent un enjeu pour les entreprises qui réfléchissent à mettre en place une gouvernance de la data en s'équipant d'outils logiciels et de compétences afin de les valoriser au titre de la stratégie d'entreprise. Si les entreprises font en sorte de protéger au maximum leurs données craignant leur utilisation par la concurrence ou leur exploitation par des tiers, certains états (dont l'Etat français) affichent leur volonté d'ouvrir et de partager leurs données.

De l'existence physique des données

Depuis l'arrivée du web dans les années 90, le réseau Internet est un enjeu de conflits géopolitiques qui donnent lieu à des stratégies de domination de la part des nations aux intérêts divergents qui cherchent à en contrôler le contenu, le fonctionnement et le développement économique.

Les data ne sont pas des objets virtuels qui circuleraient dans le cyber espace ou les « clouds » ; elles sont stockées dans des datacenters, des lieux physiques, regroupant des équipements constituant du système d'information d'une ou plusieurs entreprises avec des ordinateurs centraux, des serveurs, des baies de stockage, des équipements réseaux et de télécommunications, reliés aux opérateurs. Ces infrastructures existent physiquement et géographiquement.

Il en va de même pour les câbles sous-marins en fibre optique qui permettent d'interconnecter tous les pays de la planète ; ils font transiter 95 % des communications mondiales et 99 % des flux intercontinentaux.

L'affaire Snowden, qui lève le voile sur les faits de surveillance mondiale d'Internet par les services secrets américains et britanniques illustre le basculement en matière de cybersurveillance. Si l'Agence de sécurité nationale américaine (NSA) parvient à espionner la quasi-totalité de nos communications, ce n'est ni grâce aux satellites, ni même grâce au programme Prism, qui lui donnait accès aux serveurs des Facebook, Microsoft et autres Google. C'est en s'attaquant directement au "backbone", la colonne vertébrale de l'Internet.

Les Etats-Unis et le Royaume-Uni tirent nettement leur épingle du jeu dans cette cartographie du réseau de câbles sous-marin. Sur le territoire du Royaume-Uni, transite la

quasi-totalité des échanges Europe-Amérique. En 2013, Snowden révéla que les services secrets britanniques, captaient toutes les conversations et données qui passaient par le câble transatlantique, dans le cadre du programme Tempora. Les opérations d'espionnage ou de section des câbles sont fréquentes.

En 2015, par exemple, la Russie est suspectée par l'armée américaine de laisser dériver un navire à proximité de la côte Est, vers Cuba, surnageant ainsi le câble reliant les Etats-Unis à leur base de Guantanamo. En août dernier, le département de justice américain s'est opposé au projet du Pacific Light Cable Network qui avait pour but de relier San Francisco à Hong-Kong. Porté par une alliance entre Google, Facebook et leur partenaire chinois, Dr Peng Telecom & Media Group, le ministère de la justice soupçonne des risques d'espionnage du gouvernement chinois en raison de ses liens avec le groupe Dr Peng Telecom & Media.

Au-delà de la valeur intrinsèque des datas qui représente un enjeu, leur création, leur échange et leur exploitation sont liées à des infrastructures physiques et, à ce titre les datas, sont doublement des enjeux géostratégiques.

A l'heure du big data et de l'open data, de nouvelles menaces liées à la cybercriminalité ou l'utilisation des réseaux informatiques dans le cadre de conflits politiques, de combats militaires, de guerre économique, de renseignement ou de politique d'influence diplomatique et culturelle, émergent. Afin de se préparer au mieux à tout risque de « cyber-conflit », la France a créé par décret du 4 mai 2017 le COMCYBER, le Commandement de la cyberdéfense, placé sous l'autorité du Chef d'État-Major des armées qui rassemble l'ensemble des forces de cyberdéfense des armées françaises sous une même autorité opérationnelle, permanente et interarmées.

Quand des législations tel le Cloud Act (cf. infra) permettent à des états d'accéder à des données potentiellement stratégiques ou secret défense, il est vital d'avoir une stratégie offensive et non pas seulement de réponse.

La donnée, qu'elle soit ouverte ou non, est une richesse. On parle du « nouvel or noir de l'économie numérique » ; elle devient donc un enjeu des conflits géopolitiques.

La géopolitique des données à l'échelle locale

D'après l'ONU, près de 2,5 milliards d'êtres humains vont s'urbaniser d'ici 2050, faisant de la ville un territoire dense, avec des problématiques connues : habitat, transport, liaison domicile - lieu de travail, socialisation.

En France, un tiers du PIB est dédié à la construction et à la maintenance de nos villes sous tous leurs aspects : immeubles, maisons, réseaux de transports individuels et collectifs, etc. Ce marché juteux attire les acteurs du numérique pour qui la smart city offre des réponses aux problèmes de la cité et à ses multiples acteurs.

Dans son livre *Comment les géants du numérique veulent gouverner nos villes. La cité face aux algorithmes*, Jean Haëntjens, rapporte des propos tenus par les dirigeants d'IBM

lorsqu'ils ont lancé le concept de smart city, "les villes sont la synthèse des problèmes de l'humanité". En disant cela ils laissaient sous-entendre que leur offre de smart city pourrait prochainement apporter les solutions et en filigrane qu'elle allait leur permettre de se positionner sur un marché gigantesque. En réalité, la réalisation et la gestion des espaces urbains recouvre une multitude de sujets. La ville est le reflet des choix qui caractérisent une société - la vie sociale, les modes de vie, la dépendance énergétique, la régulation politique.



Les possibilités offertes, par la maîtrise et le partage des données sont nombreuses : la ville pourrait s'en servir comme outil planificateur ou encore pour améliorer la transition énergétique en utilisant un pilotage fin des systèmes de production et de distribution, de stockage et de consommation d'énergie. Elle pourrait revoir le parcours de soin et proposer une offre culturelle diversifiée.

Les métropoles fournissent des volumes massifs de données. Elles en collectent tout autant pour accomplir leurs missions de service public. L'ouverture des données permet potentiellement d'accompagner les ambitions économiques, sociales, environnementales, politiques, sous réserve que les pouvoirs locaux aient les ressources humaines, techniques et stratégiques pour mener à bien ces ambitions.

La récolte et l'exploitation de toutes ces données impose un nouveau mode d'organisation appuyé par des moyens techniques (une plateforme de diffusion), des moyens humains. Des postes de DPO, Data Protection Officer, ou CDO, Chief Data Officer se créent pour définir une politique de gestion, diffusion, exploitation de cette masse de données, des postes d'informaticien pour maintenir la plateforme et de plus en plus des postes d'analystes (Datascientists / Data analysts) pour traiter ce volume exponentiel de données.

2.2. La souveraineté numérique à la croisée des mondes matériel / immatériel

La souveraineté numérique est l'épicentre des nouveaux rapports de forces. Ce sont les Etats-Unis qui les premiers ont pris la mesure des enjeux monde matériel versus monde immatériel, qu'ils soient militaires ou technologiques et des rapports de force entre puissances. A l'origine, ce qui allait devenir Internet est né d'un besoin de survie informationnelle en cas de confrontation nucléaire. Les grilles de lecture de la guerre économique montrent que c'est d'autant plus vrai en matière d'open data. Les raisons principales sont l'absence de vision industrielle et d'anticipation politique.

Véritable or noir du monde immatériel, l'accès aux données est le point d'entrée de la conquête des marchés du monde immatériel. Déjà sous De Gaulle, les entreprises n'avaient pas mesuré la nécessité d'une politique de recherche de puissance dans le domaine de l'immatériel. Aujourd'hui, le cas de figure assez similaire. L'influence de la France s'est fortement affaiblie dans le domaine du monde immatériel. La France a tardé dans ce domaine à agir et à se positionner, refusant d'assumer officiellement une stratégie de conquête face aux Etats-Unis ou encore à la Chine. Ces supers puissances n'ont pas les mêmes états d'âmes. La suprématie technologique permet de créer des situations de dépendance durable. Les conséquences ne sont pas négligeables. Les GAFAM sont l'expression économique privée d'une volonté de détenir une suprématie dans les activités marchandes de l'économie numérique. C'est pourquoi la Chine a refusé d'être tributaire des GAFAM et a construit son propre écosystème. La grande majorité des entreprises françaises ont des difficultés à identifier l'intérêt de construire une politique de souveraineté dans l'économie numérique. Les GAFAM formatent et verrouillent le marché mondial, créant une perte de souveraineté fortement préjudiciable à nos intérêts économiques. Les entreprises françaises acceptent de dépendre des technologies américaines et bientôt chinoises. Cette situation est-elle souhaitable à moyen et long terme ? Il est évident que la guerre se joue à une plus large échelle. Si la définition des normes et de la réglementation constitue un premier signal d'alerte, l'offensive s'est élargie à de nombreux domaines : financier, sécuritaire et données de santé. L'enjeu majeur est l'usage commercial des données. Il n'est pas trop tard pour agir.

2.3. USA : les GAFAM et leur stratégie de conquête du monde immatériel

De plus en plus, les GAFAM (Google, Amazon, Facebook, Apple et Microsoft) sont pointés du doigt : appétit vorace pour les données, rétention des connaissances et des données privées, perte de souveraineté numérique, pour les particuliers aussi bien que les états.

Est-ce que les GAFAM s'intéressent à l'open data comme ils ont pu le faire en leur temps à l'open source ? Alphabet, la maison-mère de Google détient la place de premier contributeur sur la plateforme de code collaboratif GitHub, d'ailleurs rachetée par Microsoft en 2018. Tandis qu'Amazon commence à contribuer à des projets comme Kubernetes et Facebook met en accès libre l'un de ses data centers en enjoignant d'autres entreprises à faire de même. Les GAFAM semblent de plus en plus mener la danse de l'open source sur le plan mondial.

D'un point vue technique d'abord, l'open source est devenu la référence en matière de conception IT, en particulier pour les solutions cloud. En effet, elles ont l'avantage indéniable d'apporter un triptyque, coûts optimisés, interopérabilité des solutions et sécurité, et c'est bien ça qui intéresse les GAFAM. C'est sans compter sur l'implication des employés dans des projets open source qui est un facteur déterminant pour le rayonnement des entreprises : figurer parmi les premiers contributeurs de projets d'envergure (le noyau Linux par exemple), développer des modules en intégralité et être présent sur les communautés de codeurs majeures sont autant de gages de qualité et d'expertise pour les GAFAM. Cette

omniprésence permet donc d'être visible auprès de l'écosystème tech, d'attirer les meilleurs talents et de les retenir en leur garantissant une forte autonomie sur ces projets open source. Loin d'être cantonné à une communauté obscure et lointaine, l'open source est désormais une norme culturelle à laquelle se conforme naturellement beaucoup d'entreprises de la Silicon Valley, mais pas seulement.

Cette multiplication des forces y compris les GAFAM permet d'avoir davantage de contributions et d'accélérer ainsi l'innovation. Ces nouveaux arrivants apportent des cas d'utilisations inconnus au préalable qui apportent donc une perspective différente qui enrichira la technologie développée dans le projet.

Au-delà de ces contributions, les géants de la Tech développent par leurs initiatives des modules open source majeurs. Prenons par exemple le module Kubernetes, clef de voûte des solutions de conteneurisation développé par Google ou encore le fait que la communauté de développeurs mettant à jour le noyau de Linux se compose aussi en grande partie d'employés de ces grands groupes.

Mais, depuis ces dix dernières années, la communauté open source devient de plus en plus vigilante quant à l'indépendance de ses projets vis-à-vis de ces grands acteurs. En effet, ces acteurs sont si importants sur le marché, par leurs ressources financières, qu'apparaît le risque que le projet leur soit commercialement favorable.

Un autre principe fondateur de l'open source réside dans l'adage « celui qui fait, décide ». Dans l'absolu, c'est un principe tout à fait louable, mais grâce à la taille de leurs effectifs et ressources humaines, les Facebook, Google, Microsoft et consorts peuvent se permettre d'allouer plus de monde pour « faire », prenant ainsi de l'influence.

Si toute l'innovation d'aujourd'hui naît dans l'open source, c'est aussi grâce aux start-ups. Mais pour survivre, il leur faut monétiser leurs innovations. Ces start-ups n'en tirent par conséquent pas forcément tous les bénéfices commerciaux. Leurs innovations technologiques se retrouvent par la suite disponibles sur des clouds publics appartenant aux GAFAM. Alors que pour le moment les GAFAM s'accaparent ces données, leur monétisation paraît de plus en plus probable.

La présence des GAFAM dans la communauté open source constitue un point de vigilance, c'est aussi une opportunité indéniable. Les GAFAM investissent le champ urbain, ce qui impacte la souveraineté des villes. Pour faire en sorte à ce que les territoires stimulent les potentialités et maîtrisent, en même temps leur destin, l'administration locale devrait ajuster la réglementation et veiller à son application par tous, sans imposer des règles trop restrictives allant à l'encontre des objectifs initiaux. Pour plus d'efficacité, l'Open Innovation soutenue par les géants du numérique, permet l'acculturation sur la démarche d'open data ce qui fait progresser peu à peu la qualité de la data pour assurer une meilleure circulation.

Selon plusieurs experts de l'Open Data Institute (ODI), les GAFAM empêcheraient l'accès à de nombreuses données cartographiques qui pourraient bénéficier à d'autres entreprises. Apple, Uber, ou encore Google monopolisent les données numériques pour l'amélioration des cartes. Ils "verrouillent et étouffent" l'innovation. Ainsi, ils pourraient bientôt être obligés de partager leurs données avec les entreprises publiques et même leurs concurrents.

Dans le domaine de l'utilisation des données de géolocalisation, une enquête du site Quartz menée en 2017 révélait que Google avait récolté les données de localisation des utilisateurs via son système d'exploitation mobile : Android. Depuis 2017, Google collecte les informations sur les antennes-relais dont se servent les utilisateurs pour se connecter à Internet depuis leur smartphone et s'en servent pour développer des services associés. Certains services comme les livraisons de marchandises à domicile par drone ou véhicule autonome pourraient bénéficier de ces données pour se développer.

Google et l'Université de Pittsburgh défendent le fait de pouvoir utiliser des images libres de droits. Elles pourraient être mises à disposition par Facebook et Microsoft par le biais de leur fondation respective. Cette approche financièrement et énergétiquement gourmande n'est pourtant pas la meilleure méthode d'évolution du « machine learning ».

Autre cas d'école : le « Tomato gate » est représentatif des dérives actuelles. La tomate est un fruit en botanique et en droit européen. Au début des années 20, aux Etats-Unis, pour des raisons fiscales, il y eu un lobbying pour la ranger dans les légumes.

Aujourd'hui, les IA de Google, Facebook... apprennent le monde dans des bases de données américaines, et « pensent » donc que la tomate est un légume. Et les gens créent des services grâce à ces IA, qui disent que la tomate est un légume.

En cela, nous serons tous obligés de dire que la tomate est un légume, car les données de référence pour le développement de l'IA auront été uniquement américaines. Tous ces nouveaux sujets doivent s'intégrer dans une stratégie de souveraineté. Penser la donnée comme une infrastructure serait en ce sens une question de souveraineté économique.

Le groupe Google, représente 31 milliards de dollars de résultat net en 2018 – 84 % de son chiffre d'affaires est réalisé grâce à la publicité et, par conséquent, aux données personnelles de ses utilisateurs. Son moteur de recherche génère 3,3 milliards de requêtes par jour, sa plateforme de vidéo (YouTube) cumule 1 milliard d'heures visionnées quotidiennement, et son système d'exploitation mobile (Android) représente 86,7 % du marché. Les activités de la firme de Mountain View consistent aujourd'hui davantage à un gigantesque stockage de la connaissance humaine qu'à une simple entreprise du Web. Aux Etats-Unis, cette situation commence d'ailleurs à agacer les entrepreneurs et les politiques, mais également la justice : Google s'est vu attaquer par 50 procureurs américains en septembre dernier.

La France pas plus que l'Union européenne (UE) n'échappe à l'«emprise» de l'entreprise américaine. En octobre dernier, plusieurs dizaines de journalistes européens ont tiré la sonnette d'alarme contre la volonté hégémonique de Google, qui empêche toutes les recettes publicitaires générées par les contenus en open data de la presse en ligne sans

laisser trop de choix aux médias, sous peine d'être déréférencés du moteur de recherche star, avec les conséquences économiques qui s'en suivent.

Malgré les nombreuses administrations publiques qui utilisent aujourd'hui Qwant – comme la Mairie de Paris et ses milliers de fonctionnaires ou encore l'Assemblée Nationale – le groupe américain reste seul maître à bord concernant les collectes de données.

Cette compétition entre titans de la Silicon Valley est bien partie pour se dérouler à l'insu des consommateurs et à leur détriment. Ce véritable « scandale économique et démocratique » voit les internautes « offrant » littéralement leurs informations au géant américain. Surtout, en laissant agir les géants américains, la France s'enfoncé dans un statut de colonie numérique alors même que le temps presse. La bataille de la souveraineté se joue en effet aujourd'hui avec le futur de nos startups françaises. Celles-ci sont menacées dans tous les domaines, même les plus régaliens comme celui de la Défense ou encore celui de la justice, qui attise tous les appétits. Il est de plus en plus nécessaire que les gouvernements prennent des mesures pour imposer une transparence accrue à ces entreprises colossales.

«Nous sommes encore dans une société pré-capitalistique de la data, car d'un côté il y a une monétisation de la donnée, alors que de l'autre, l'auteur de l'information fait un don » explique Alain Bensoussan, avocat à la Cour d'appel de Paris, spécialiste en droit des nouvelles technologies de l'informatique et de la communication. *«Nous sommes dans un pseudo marché, totalement asymétrique, puisque nous donnons notre data en contrepartie de services gratuits, mais qui ne le sont absolument pas.»* L'apparente gratuité est un leurre, car l'internaute est devenu depuis longtemps le produit.

2.4. Chine, BATXH : quand les data servent le système

Comme les Etats-Unis, la Chine est partie à la conquête du monde immatériel et s'affronte sur ce terrain avec l'Inde et surtout les Etats-Unis. A travers cette conquête, elle compte devenir la première puissance mondiale, devant les USA.

Pour cela, elle peut notamment s'appuyer sur ses géants du numérique : les BATXH. Baidu, Alibaba, Tencent, Xiaomi, Huawei sont les concurrents directs des GAFAM (Google, Amazon, Facebook, Apple, Microsoft).

Les BATX sont en pleine croissance. Rien que sur l'année 2018, le chiffre d'affaires d'Alibaba a doublé. Méconnus en dehors de la Chine, les BATX étendent leur influence et se développent dans de nouveaux territoires physiques, comme aux Etats-Unis ou en Europe. La croissance du site Alibaba ou encore du producteur de smartphones Huawei en sont des exemples frappants.

Si les GAFAM prônent la liberté sur Internet (malgré la bulle informationnelle qui enferme chaque internaute), les BATX s'affichent comme un outil de contrôle démographique et de surveillance de masse pour la Chine. Depuis 2014, le gouvernement chinois a développé un "Plan de planification de la construction du système de crédit social (2014-2020)". Ce

programme est basé sur la collecte d'informations massive issues des services de police, administrations, réseaux sociaux, caméras de surveillance.

Ce « permis à points de bonne conduite » évalue le comportement des citoyens, sur Internet ou dans la vie réelle. Le score individuel, allant de 350 à 950 points, est calculé en fonction des « bonnes » ou des « mauvaises » actions en ligne ou hors ligne, faisant grimper ou descendre la note.

L'Etat chinois distribue de mauvais points : mauvaises fréquentations sur les réseaux sociaux, non-paiement des amendes, cigarette allumée malgré une interdiction, achat de jeux vidéo, diffusion de fausses informations, utilisation de tickets de transport périmés, etc.

S'il s'agit d'un fichier informatique et donc de data, ses conséquences, elles, sont bien réelles. En cas de mauvais score, les citoyens sont interdits d'exercer certains métiers, ne peuvent accéder au crédit ni monter une entreprise ou inscrire leurs enfants dans une école privée. Ils n'ont plus accès à l'avion, au TGV. Certains ont même interdiction de quitter le pays. A contrario, un bon score permet de voyager en première classe, d'obtenir un prêt, d'avoir une promotion... Pour avoir une bonne note, le citoyen doit s'occuper de ses parents, élever ses enfants, rembourser ses crédits, etc.

Ce dispositif modèle les habitants et les façonne selon la volonté du gouvernement. Une manière pour le régime chinois de contrôler ses 1,42 milliards de citoyens.

Les BATX participent aussi à la notation des Chinois et aident le gouvernement dans sa surveillance de masse. Ce système de crédit social est développé par Ant Financial Services Group, filiale du groupe Alibaba, en association avec le gouvernement chinois. Par ailleurs, au gré de leurs transactions sur Alipay Wallet, service de paiement d'Alibaba, le score des utilisateurs fluctue. Ceux qui voient leurs points augmenter obtiennent alors des avantages. De son côté, Tencent Credit récolte les données collectées sur des applications comme WeChat. Là encore, certains utilisateurs sont récompensés et d'autres punis.

Plusieurs applications permettent également aux citoyens de voir leur note et celle de leurs amis. D'autres, comme des sites de rencontres, incitent les Chinois à rendre publique leur score afin de rendre leur profil plus attractif.

Si ce « permis de bonne conduite » ou « fichier de réputation » est pour l'instant expérimenté dans certaines provinces (Jiangsu, Fujian et Guangdong), l'objectif est bien de le généraliser à l'ensemble du pays.

En Chine, la loi sur la cybersécurité (2017) protège et respecte en théorie la vie privée de ses usagers. Mais le gouvernement préfère invoquer ses droits et brandit la sécurité des données dans le cadre de la sécurité nationale.

Pour Pékin, les données font partie intégrante du développement technologique et économique du pays, et participent à son accroissement de puissance, y compris sur la scène internationale.

Car son système de crédit social s'applique aussi aux entreprises étrangères. A l'instar des citoyens, les sociétés sont elles aussi notées et reçoivent des bons ou mauvais points. Une manière, là encore, de façonner un certain mode de pensée et de gouvernance. Comme le

résume un article paru dans France 24 en août 2019, « *ce dispositif peut aussi être utilisé comme arme dans la guerre commerciale sino-américaine. Une nouvelle réglementation, adoptée en juillet 2019, prévoit de mettre en place un système de liste noire spécifique pour les entreprises qui “mettent en danger l’intérêt national” ou “violent les droits légitimes des consommateurs et du public”. Des notions suffisamment vagues pour être brandies par les autorités chinoises aux premières crises commerciales venues.* »

Les BATX ne sont pas encore des contrepouvoirs aux GAFAM, mais peuvent constituer une puissance contre l’Europe et la France dans la surveillance des populations en plus de la commercialisation des données. L’OD risque finalement de donner sans « effort » à des BATX, et aux GAFAM des informations via l’OD.

Celle-ci se voulant être un mouvement démocratique initialement, pour le bienfait des citoyens, semble se transformer en « piège » si les BATX se positionnent sur le marché de la data. Ils auront toute la liberté d’accéder à des données libres pour en faire un usage à la fois commercial et de contrôle de la population. Si ce constat est probable, quel est le risque à poursuivre dans cette voie, et ce, même si le RGPD peut peut-être nous protéger comme citoyen ?

2.5. Où en est l’Europe ?

2.5.1 Les Etats-Unis et le Cloud Act

En 2018, le « Clarifying Lawful Overseas Use of Data » Act (dit Cloud Act) clarifie l’usage des données hors des Etats-Unis. Le Cloud Act renforce les pouvoirs des agences de renseignement américaines, en facilitant leur accès aux données stockées dans des datacenters appartenant à des sociétés d’origine américaine.

Cette loi fédérale promulguée le 23 mars 2018 oblige les entreprises sises aux États-Unis à fournir les données stockées sur leurs serveurs, y compris ceux situés à l’étranger, en cas de mandat ou d’assignation en justice. Il s’agit pour les Etats-Unis de pouvoir mener des investigations sans passer par les règles de coopération judiciaire internationale.

Les forces de l’ordre ou agences de renseignement peuvent ainsi obtenir les données stockées sur les serveurs des opérateurs télécoms et des fournisseurs de services américains, partout dans le monde. Les prestataires de service doivent communiquer les « contenus de communications électroniques et tout enregistrement ou autre information relatifs à un client ou abonné, qui sont en leur possession ou dont ils ont la garde ou le contrôle, que ces communications, enregistrements ou autres informations soient localisés à l’intérieur ou à l’extérieur des Etats-Unis ».

2.5.2. RGPD : l'arme européenne pour faire face au Cloud Act ?

Le nécessaire encadrement de l'utilisation des données

Extraire et exploiter le potentiel de données est un défi majeur pour toute entreprise opérant au 21ème siècle.

Le monde du transport, est l'un des secteurs qui s'est le plus impliqué dans l'open data.

Sébastien Fraisse, directeur général d'Indigo, leader français du stationnement avec 430.000 places répartis dans plus de 150 villes, témoigne : « *Pour des raisons purement commerciales et parfois même contractuelles, nous devons aux villes un certain nombre de données que nous essayons progressivement d'enrichir* ». Dans cette veine, le projet de loi d'orientation des mobilités, dite LOM, ambitionne de favoriser l'émergence du covoiturage et de l'auto-partage. Elle prévoit aussi la mise en place d'applications de MaaS (mobility as a service) et instaure l'ouverture de certaines données.

Ces applications multimodales devront regrouper l'ensemble des moyens de transport disponibles sur le territoire de la collectivité et les informations en temps réel sur leur situation. La LOM prévoit que les entreprises devront partager leurs données notamment pour que les systèmes de billettique des services de transport puissent être intégrés dans d'autres plates-formes. Dans une interview à l'Usine Digitale, en juin 2019, Pascal Melet, directeur général du salon Autonomy indique que les entreprises de VTC vont bénéficier de cette ouverture sans pour autant donner de l'information de leur côté. En effet, n'ayant ni horaires ni prix fixes, ces plates-formes ont réussi à ne pas se voir imposer l'ouverture de leurs données.

La plateforme transport.data.gouv.fr offre un point d'entrée unique de dépôt et consultation des données d'usage sur le transport et la mobilité.

Néanmoins, le chemin reste long avant que l'offre des données ne soit encore complète et mise à disposition en temps réel. Pour une entreprise, cette démarche nécessite des ressources humaines dédiées, des infrastructures à créer et à maintenir et la mise en cohésion de la démarche. Comme dans tout projet transverse, il faut savoir valoriser le projet auprès du reste des collaborateurs et les solliciter en plus de leurs activités quotidiennes, sur un sujet dont la valeur ajoutée n'est pas toujours démontrée.

On distingue les données d'offre, des données d'usage.

- Données d'offre : elles permettent (dans le secteur du transport par exemple) d'informer les utilisateurs des différents modes de déplacements, mais aussi de créer les outils multimodaux de demain (ex : le MaaS : Mobility as a Service).
- Données d'usage : provenant des différentes applications, elles permettent d'analyser les comportements des utilisateurs et surtout d'aider la puissance publique à mieux piloter ses offres et/ou les offres qui sont présentes dans leur "ressort territorial".

Cependant, si l'ouverture des données a été rendue obligatoire par la législation pour les entreprises publiques et la plupart des collectivités, il n'en va pas de même pour les applications : Uber, Waze ou Citymapper n'ouvrent pas leurs données mais pourtant se servent des données publiques existantes pour fonder leur modèle économique. En effet, ces entreprises qui ne sont pas européennes ne sont pas soumises aux mêmes contraintes législatives.

Citymapper a dû « batailler » plusieurs années pour avoir accès aux données de la RATP, qui les a mis à disposition en 2017. Cette application britannique qui permet aux citoyens d'optimiser leurs déplacements en transports en commun, a lancé son offensive en 2012 via une pétition sur le site Change.org afin d'obtenir le soutien des usagers. Objectif : obliger la RATP à donner libre-accès à ses données des transports en temps réel. La RATP a résisté et développé sa propre application de trajets en temps réel.

Le nouveau paradigme de données a également permis l'émergence de nouveaux modèles commerciaux : les plates-formes numériques. Ces dernières se sont positionnées avec succès en tant que parties prenantes uniques et inévitables entre l'offre et la demande. Leur modèle de développement est basé sur les effets de réseau, ce qui signifie que la valeur de leur offre de service augmente de manière exponentielle avec la croissance du nombre d'utilisateurs. Les plates-formes numériques ont largement contribué à favoriser l'innovation et le développement économique dans de nombreux pays. Cependant, le marché doit maintenant être régulé afin d'éviter toute distorsion qui nuirait à tous les avantages liés à l'utilisation de telles plates-formes. L'UE a un rôle clé à jouer ici et doit retrouver la capacité d'agir maintenant : dans cinq ans, il sera trop tard.

RGPD et open data : deux mouvements antinomiques ?

L'open data s'inscrit dans une tendance d'ouverture et de partage qui considère que la donnée publique est d'intérêt général. L'objectif est de rendre la donnée disponible, réutilisable et de permettre la création de valeur nouvelle. Promulguée en 2016, la loi Lemaire pour une république numérique (cf. infra), consacre l'obligation pour les administrations et collectivités de plus de 3500 habitants de rendre publique les données qu'elles détiennent.

Le Règlement Général pour la Protection des Données, RGPD, entré en application en mai 2018, responsabilise les organismes publics et privés qui collectent et traitent de données à caractère personnel, dans un cadre juridique européen. Il renforce et unifie la protection des données pour les personnes.

Il a été conçu autour de 3 objectifs :

- Renforcer les droits des personnes
- Responsabiliser les acteurs traitant des données
- Crédibiliser la régulation grâce à une coopération renforcée entre les autorités de protection des données

La notion de donnée personnelle est extrêmement large puisqu'elle est définie par la CNIL comme « *toute information se rapportant à une personne physique identifiée ou identifiable* ». Sont également considérées comme personnelles les données qui, par le recoupement de plusieurs informations (date de naissance, sexe, ville, diplôme, ...) permettent d'identifier une personne.

Les personnes, dont les données sont traitées, sont protégées par une série de droits :

- Droit d'accès : droit pour la personne concernée d'obtenir du responsable de traitement la confirmation que des données personnelles sont ou ne sont pas traitées et lorsqu'elles le sont, l'accès aux données ainsi qu'aux informations la concernant. L'entreprise dispose ensuite d'un délai d'un mois pour accéder à la demande de la personne concernée.
- Droit de rectification : droit pour la personne concernée d'obtenir du responsable du traitement, dans les meilleurs délais, la rectification des données inexacts la concernant.
- Droit d'opposition : droit pour la personne concernée de s'opposer, pour des raisons tenant à sa situation particulière, à un traitement des données personnelles la concernant.
- Droit d'effacement (droit à l'oubli) : droit pour la personne concernée d'obtenir du responsable du traitement l'effacement, dans les meilleurs délais, des données personnelles la concernant.
- Droit à la portabilité : droit pour la personne concernée d'obtenir et de réutiliser les données la concernant pour ses besoins personnels.
- Droit à la limitation du traitement : droit pour la personne concernée d'interdire au responsable du traitement de se servir de certaines données collectées.

Une fois ces points précisés, on reconnaît facilement la difficulté de faire cohabiter ces deux règlements.

Néanmoins, la loi Lemaire, et l'article L.312-1-2 du code des relations entre le public et l'administration imposent que les données individuelles doivent, avant leur mise à disposition, être traitées de manière à occulter les mentions à caractère personnel. La question se pose alors de l'intérêt de disposer de données dont les informations ont été occultées ?

La mise à disposition des données publiques provoque des coûts importants (détail infra). Or, la sélection des données ainsi que le nettoyage total ou partiel des données ajoute des coûts supplémentaires. Quel choix doit faire le producteur de données : réduire le spectre

des données déjà diffusées ou les anonymiser au risque de les rendre inexploitable, pour les ré-utilisateurs ?

Face à cette impasse, il est apparu qu'en traitant les ré-utilisateurs différemment selon l'usage des données qu'ils faisaient, il était possible d'être conforme au RGPD.

Ainsi, la loi de modernisation du 26 janvier 2016, dont le nouveau Titre IV organise la mise à disposition de certaines données de santé au profit de la communauté des chercheurs ainsi que du grand public, permet d'utiliser des données personnelles, dans le respect du RGPD. Un autre exemple, est celui des données juridictionnelles ; les décisions de justice sont supposées être rendues publiquement (art. 451 du Code de procédure civile en matière civile et art. 10 du Code de justice administrative en droit public). Les décisions des cours suprêmes sont déjà disponibles, sur Légifrance. La loi du 7 octobre 2016 pour une république numérique prévoit la mise à disposition du public, à titre gratuit de l'ensemble des décisions de justice, « dans le respect de la personne concernée ». Cette mise à disposition doit être précédée d'une analyse du risque de ré-identification des personnes.

Au-delà des risques de divulgation criminelle ou accidentelle des données personnelles, se pose la question de l'accès des entreprises et des gouvernements.

Le RGPD permet de préserver les libertés individuelles, mais participe aussi à leur assurer plus de légitimité. L'enjeu est de taille. Notre dossier médical, nos opinions politiques, nos résultats scolaires, notre orientation sexuelle, nos achats, nos déplacements, nos amis, les amis de nos amis, les amis des amis de nos amis... Toutes ces informations et bien d'autres peuvent faire l'objet d'investigations, de recoupements et permettre d'établir un profil d'une précision redoutable.

Ainsi, les algorithmes d'Amazon prévoient avec une grande précision les livres ou les disques qui ont des chances de plaire à ses clients. En 2016, 30 % des livres vendus par le géant américain et 50 % des films loués par Netflix ont été choisis par des algorithmes prédictifs. Progressivement le client est enfermé dans une "bulle filtrante". La probabilité qu'il puisse rencontrer des idées nouvelles, qui l'amènerait éventuellement à changer d'opinion ou de vie, se réduit donc à mesure que la performance des algorithmes s'accroît. Les algorithmes de l'IA utilisent les données des internautes.

Beaucoup de ces informations sont déjà accessibles. Souvent, elles sont même mises à disposition volontairement par les utilisateurs eux-mêmes sur les réseaux sociaux et les blogs. Les entreprises qui ont accès aux données, les exploitent et analysent ainsi les goûts, mémorisent les choix et proposent des services sur mesure. A ce jour, le constat montre que les data sont essentielles dans le commerce du futur, et sans doute dans la vie citoyenne à venir.

2.6 La législation française et européenne : rétrospective

« *L'information doit être au service de chaque citoyen* ». C'est par ces mots que débute l'article 1 de la loi n°78-17 du 6 janvier 1978, créant la Commission nationale de l'informatique et des libertés (CNIL). Le texte poursuit : « *Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques* ». La même année, la loi CADA (loi n°78-753 du 17 juillet 1978) apporte un

complément, en garantissant la liberté d'accès aux documents administratifs aux citoyens, tout en évoquant la réutilisation des informations publiques.

Au fur et à mesure des années et des évolutions technologiques, les lois françaises ont été modifiées, complétées. Mais comme on le verra ci-après, la France va très vite accuser un certain retard, en matière de gestion des données.

En Europe, les grandes orientations de l'open data ont été données par deux grandes directives.

- La directive européenne 2003/98/CE (PSI 2003) fixe un ensemble de règles destinées à faciliter la réutilisation des données du secteur public. Elle encourage les pays européens à publier et mettre à disposition des documents dans tout format sous forme électronique. Dès 2003, la Commission inclut une possible réutilisation commerciale des données. Chaque pays doit dorénavant publier les données de son secteur public. La France connaît cependant un retard au démarrage. En effet, elle ne publie véritablement ses données qu'à partir de 2016, soit 13 ans après la directive européenne. Du côté européen, la commercialisation est mentionnée dans PSI 2003 à travers la réutilisation des données. L'Europe comprend donc très vite l'intérêt économique des données, ce qui n'est pas le cas de la France, qui tarde à appliquer et mettre en œuvre les directives.
- En 2007, la directive européenne 2007/2/EC dite « Inspire », vise à établir une infrastructure d'information géographique permettant la diffusion et le partage de données géographiques au sein de l'Europe dans le cadre de politiques environnementales. « Inspire » n'impose pas la collecte de nouvelles données géographiques mais coordonne la diffusion et le partage de celles déjà recueillies. La directive définit également des règles pour optimiser les recherches : mots-clés, classification des services et des séries de données, qualité et validité des données géographiques.

« Inspire » a notamment donné naissance au portail « Inspire Geoportal ». Celui-ci publie les données des états-membres en open data. Mais comme on peut le constater, tous les pays ne jouent pas le jeu. La Bulgarie et le Portugal sont les plus gros pourvoyeurs de données, là où le Royaume-Uni, la Suisse et le Liechtenstein n'en communiquent quasiment pas, ce qui traduit une mobilisation encore hésitante voire disparate des pays européens.

Les règles de l'open data ont été mises à jour en 2011, complétant les directives 2003/98/EC sur la réutilisation des informations du secteur public, et « Inspire » sur l'infrastructure d'information géographique.

En 2013, la directive 2013/37/UE (dite PSI 2013) modifie la directive 2003/98/CE sur l'ouverture et la réutilisation des informations du secteur public. Cette modification étend le champ d'application aux bibliothèques, musées et archives. Pour la première fois, l'Europe évoque la création de valeur à travers les données. « *L'un des principaux objectifs de l'établissement d'un marché intérieur est de créer les conditions qui permettront de*

développer des services à l'échelle de l'Union. (...) L'élargissement des possibilités de réutilisation du matériel culturel public devrait entre autres permettre aux entreprises de l'Union d'exploiter le potentiel de ce matériel et contribuer à la croissance économique et à la création d'emplois ». Cette directive mentionne aussi la possible mise en place de redevances.

En 2015, la Commission européenne va plus loin et évoque la construction d'une économie des données. La « Stratégie pour un marché unique numérique en Europe » décrit en introduction : « Dans moins de dix ans, la plus grande partie de l'activité économique dépendra d'écosystèmes numériques intégrant les infrastructures numériques, le matériel et les logiciels, les applications et les données ». « Les mégadonnées (...) sont indispensables à la compétitivité de l'UE ». Avec une augmentation de 40 % par an, leur croissance est « sept fois supérieure à celle du marché de l'informatique ». « Les données sont souvent considérées comme un catalyseur de croissance économique, d'innovation et de conversion au numérique. (...) L'absence de systèmes et de services ouverts et interopérables et de portabilité des données entre les services représente un obstacle supplémentaire à la circulation transfrontière des données et au développement de nouveaux services ».

L'interprétation française

Alors que l'Europe mentionne la monétisation possible des données dès 2003, la création de valeur par les données en 2013, elle évoque en 2015 « le marché unique numérique » consciente de la croissance économique et de l'innovation que toute entreprise peut tirer de ces données, nouvel or noir invisible.

Depuis la première directive européenne, la France n'a pas réagi. Il faut en effet attendre 2015 pour voir le gouvernement mettre en place les premières lois sur l'ouverture des données.

En 2015, l'article 4 de la loi n°2015-990 du 6 août 2015 pour la croissance, l'activité et l'égalité des chances économiques (dite loi « Macron ») modifie le Code des transports et rend obligatoire l'ouverture des données des services de transport et de mobilité. Ceux-ci doivent diffuser « librement, immédiatement et gratuitement » les données relatives « aux arrêts, aux tarifs publics, aux horaires planifiés et en temps réel, à l'accessibilité aux personnes handicapées, à la disponibilité des services, ainsi qu'aux incidents constatés sur le réseau et à la fourniture des services de mobilité et de transport ». Cependant, si la loi Macron rend obligatoire l'ouverture des données dans le transport, le décret d'application ne voit pas le jour et sera même suspendu en 2017.

Le même mois, la loi n°2015-991 du 7 août 2015 (Loi « NOTRe »), portant sur la nouvelle organisation territoriale de la République, oblige les communes de plus de 3 500 habitants et les établissements publics de coopération intercommunale, à rendre leurs données publiques accessibles en ligne, et ce, sous forme électronique. L'ouverture des données ne concerne alors plus un secteur d'activité mais la majorité des communes françaises, pour qui cette loi annonce une révolution numérique.

La France transpose la directive européenne 2013/37/UE dans sa loi n°2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public. L'article 5 modifie la loi CADA (loi n°78-753) et fixe le principe de gratuité pour la réutilisation d'informations publiques, mais mentionne l'idée d'une contribution financière.

L'ordonnance n°2015-1341 du 23 octobre 2015 du Code des relations entre le public et l'administration, qui entre en vigueur le 1^{er} janvier 2016, précise la nature des données à ouvrir : dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions et décisions.

Le décret du 25 mars 2016 relatif aux marchés publics oblige également dans son article 107, les acheteurs publics à donner accès aux données essentielles des marchés. L'acheteur offre un accès libre, direct et complet aux données essentielles du marché public (identification, nature et objet du marché, procédure, lieu d'exécution, durée...).

En France, la loi n°2016-1321 du 7 octobre 2016 pour une République numérique régleme l'usage et la mise à disposition des données ouvertes. Les collectivités de plus de 3500 habitants doivent ainsi mettre en ligne « *les bases de données, mises à jour de façon régulière, qu'elles produisent ou qu'elles reçoivent* » ainsi que « *les données, mises à jour de façon régulière, dont la publication présente un intérêt économique, social, sanitaire ou environnemental* ». Le texte insiste pour la première fois sur la notion de service public : « *La mise à disposition des données de référence en vue de faciliter leur réutilisation constitue une mission de service public relevant de l'Etat.* »

Le gouvernement oblige aussi les acteurs privés d'électricité, de gaz, de traitement des déchets à mettre leurs « *données à disposition du public par voie électronique, dans un format ouvert, aisément réutilisable et exploitable* ».

En pratique, la loi oblige les collectivités de plus de 3500 habitants à publier leurs bases de données, sous réserve d'anonymisation et de protection de la propriété intellectuelle et du secret industriel et commercial. Ces données peuvent ensuite être exploitées et réutilisées par chacun, particulier ou entreprise.

Avec le décret 2016-1922 du 28 décembre 2016 relatif à la publication en ligne des documents administratifs, les administrations de 50 agents ou salariés doivent, à présent, elles aussi, ouvrir leurs données, s'ajoutant aux communes de plus de 3500 habitants. Lorsque les documents administratifs sont disponibles sous forme électronique, les collectivités territoriales doivent les publier en ligne, ainsi que leurs versions mises à jour.

Chaque année, décrets, lois, ordonnances complètent la liste des données à ouvrir. Le décret n°2017-331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence détaille dans l'article 1 les « *données de référence* » : répertoire des entreprises, répertoire national des associations, plan cadastral, registre parcellaire graphique, référentiel à grande échelle relatif à l'Institut national de l'information géographique et forestière, base adresse nationale, base de données de l'organisation administrative de l'Etat, le répertoire opérationnel des métiers et des emplois produit par Pôle emploi ainsi que le code officiel géographique.

En 2018, un pas est franchi dans la « *non-anonymisation* » de certaines informations. Dorénavant, avec le décret n°2018-1117 du 10 décembre 2018, certains documents peuvent être publiés sans anonymisation. C'est le cas des organigrammes, annuaires, répertoire national des associations et le répertoire des entreprises et de leurs établissements, résultats obtenus par les candidats aux examens et concours... Quant aux documents contenant des données personnelles et ne rentrant pas dans le cadre du décret du 10 décembre 2018, leur publication est possible à condition qu'elles aient fait l'objet d'un traitement rendant l'identification impossible. Cependant, si l'individu a donné son consentement préalable, aucune anonymisation n'est nécessaire. Cette notion de consentement n'est pas sans rappeler le RGPD, dont il s'agit d'un pilier fondateur.

En Europe, en 2019, la directive 2019/1024 mentionne l'importance des données pour favoriser le développement économique, la création et la protection d'emplois, « *au bénéfice des communautés locales* » tout en favorisant « *la responsabilité et la transparence* ». Cette ouverture des données doit en particulier booster la croissance et la compétitivité des TPE et PME. L'Europe aborde clairement la tarification. Si les données ouvertes étaient jusqu'ici gratuites, on a pu constater, au fil des directives, un glissement vers une monétisation des données. Dorénavant, les organismes du secteur public, bibliothèques, musées, archives et entreprises publiques peuvent faire payer la reproduction, la mise à disposition et la diffusion des documents à travers une redevance.

Par ailleurs, la directive étend une nouvelle fois le champ des données à celles issues de la recherche, qui doivent à présent être mises à disposition d'autrui. L'Europe accroît le type de données à ouvrir et incite ses pays membres à ouvrir ses nouvelles données. Dans le même temps, elle va dans le sens d'une monétisation de ces nouvelles informations ouvertes au public.

Enfin, la directive définit les « *données de forte valeur* » comme « *des documents dont la réutilisation est associée à d'importantes retombées positives au niveau de la société, de l'environnement et de l'économie, en particulier parce qu'ils se prêtent à la création de services possédant une valeur ajoutée, d'applications et de nouveaux emplois décents et de grande qualité, ainsi qu'en raison du nombre de bénéficiaires potentiels des services et applications à valeur ajoutée fondés sur ces ensembles de données* ». Ces données stratégiques doivent ainsi pouvoir faire l'objet d'un téléchargement massif.

Une mise en œuvre à plusieurs vitesses

Depuis la promulgation de la loi pour une République numérique que s'est-il passé ? Trois ans plus tard, ces objectifs restent encore bien loin devant, notamment pour le volet open data avec le constat de données publiques encore peu ouvertes à des formats facilement réutilisables par tout un chacun.

Les premiers constats montrent des avancées significatives pour mettre le projecteur sur la donnée, améliorer les services aux usagers (ex. transports), rendre plus efficace l'action des services publics (ex. offres d'emplois) et même créer de la valeur, comme avec l'ouverture

des textes de jurisprudence utilisés comme base d'apprentissage IA par les startups de la Legal Tech. Les données de références sont maintenant bien installées, et réutilisées, dans un service public de la donnée à l'échelon national.

L'open data tarde à changer d'échelle et à passer à la vitesse supérieure. On note une effervescence pour les échanges de données au sein d'écosystèmes fermés et les premières applications opérationnelles des villes intelligentes.

Le constat général est que l'open data ouvert et accessible à tous reste difficile à mettre en œuvre. Les plus optimistes diront que c'est devenu un usage courant, que l'accès aux données en open data se fait si naturellement qu'on ne s'en rend pas compte. C'est l'avis de l'association Open Data France, dans son ouvrage publié en juin 2019. Il y aurait un "*iceberg des réutilisations*" avec une partie émergée visible très faible et de nombreux cas de réutilisations invisibles. Mais cet ouvrage reprend de multiples réutilisations de données, y compris par les entreprises opérateurs de services publics, qui ne passent pas par des plateformes open data et n'en adoptent pas les licences, jugées parfois contraignantes.

Les plus pessimistes constateront que 90 % des collectivités ne publient pas leurs données. Celles qui les publient, les mettent rarement à jour ce qui limite les envies à réutiliser au-delà d'une analyse ponctuelle.

Mais compte tenu d'autres développements récents comme le RGPD ou celui des plateformes numériques qui valorisent les données et plus particulièrement de l'intelligence artificielle, on peut quand même se poser la question si les objectifs fixés pour avoir une longueur d'avance dans le numérique par les pouvoirs publics en France sont encore d'actualité.

La France n'est pas un cas isolé. Le ralentissement de la locomotive anglo-saxonne lancée en 2008, depuis le départ de l'administration Obama (Open Government Initiative) et l'enlisement du Royaume-Uni dans le Brexit, deux moteurs de l'open data de la première heure, n'est certainement pas étranger à cet essoufflement français. L'Europe a aussi cherché à insuffler une nouvelle dynamique pour l'open data mais son projet majeur reste plus le RGPD que l'OGI. Enfin, les changements à la DINSIC (départ d'Henri Verdier) et le remplacement de Mounir Majoubi par Cédric O au secrétariat du numérique, n'ont certainement pas non plus facilité la continuité des projets.

Le partage de données au sein d'écosystèmes (produits alimentaires, qualité de l'environnement, agriculture...) développe des initiatives positives pour tous et stimule l'innovation et les nouveaux usages. Mais on voit que la valeur est dans celui qui a la capacité à avoir les données les plus complètes et de qualité, quitte à les collecter avec son propre réseau d'ambassadeurs.

Si de nombreux acteurs se sont lancés dans l'ouverture de leurs données, d'autres - chercheurs ou institutions - se refusent à utiliser la donnée issue des sciences participatives et citoyennes. Ils remettent notamment en cause l'objectivité des participants, qui sont pour certains des militants engagés. D'autres chercheurs refusent aussi de confier à des tiers bénévoles des tâches de collecte dont ils estiment qu'elles devraient être financées par la recherche publique. À l'inverse, le manque de retour vers les contributeurs est un facteur de démotivation très fort. Les contributeurs amateurs souhaitent savoir comment (et pourquoi) sont utilisées les données qu'ils ont collectés.

3. Les acteurs : comment mettent-ils en œuvre les projets d'open data ?

3.1 Les acteurs en France

L'action de l'État en matière d'ouverture des données publiques est animée depuis fin 2011 par la mission Etalab qui dépend de la direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC).

Ces données sont diffusées de manière structurée selon une méthode et une licence ouverte garantissant leur libre accès et leur réutilisation par tous, sans restriction technique, juridique ou financière. En France, <https://www.data.gouv.fr> permet à tout un chacun d'accéder librement aux données publiques pour les partager, les améliorer et les réutiliser.

Cette plateforme officielle permet de répartir les données ouvertes dans plusieurs catégories : l'agriculture et l'alimentation, la culture, l'économie et l'emploi, l'éducation et la recherche, l'international et l'Europe, le logement, le développement durable et l'énergie, la santé et le social, la société, et les transports, tourisme et territoires. Le site tient un répertoire thématique des données ouvertes, chacun pouvant librement y accéder, l'utiliser, la modifier et la rediffuser, quel que soit son but. Grâce à cette catégorisation, les utilisateurs peuvent facilement sélectionner le domaine qui les intéresse.

Pour les entreprises, il s'agit surtout de faire œuvre de transparence, d'améliorer leur image et combattre les fake news par le biais de la mise à disposition de données claires et validées.

3.1.1. Les collectivités

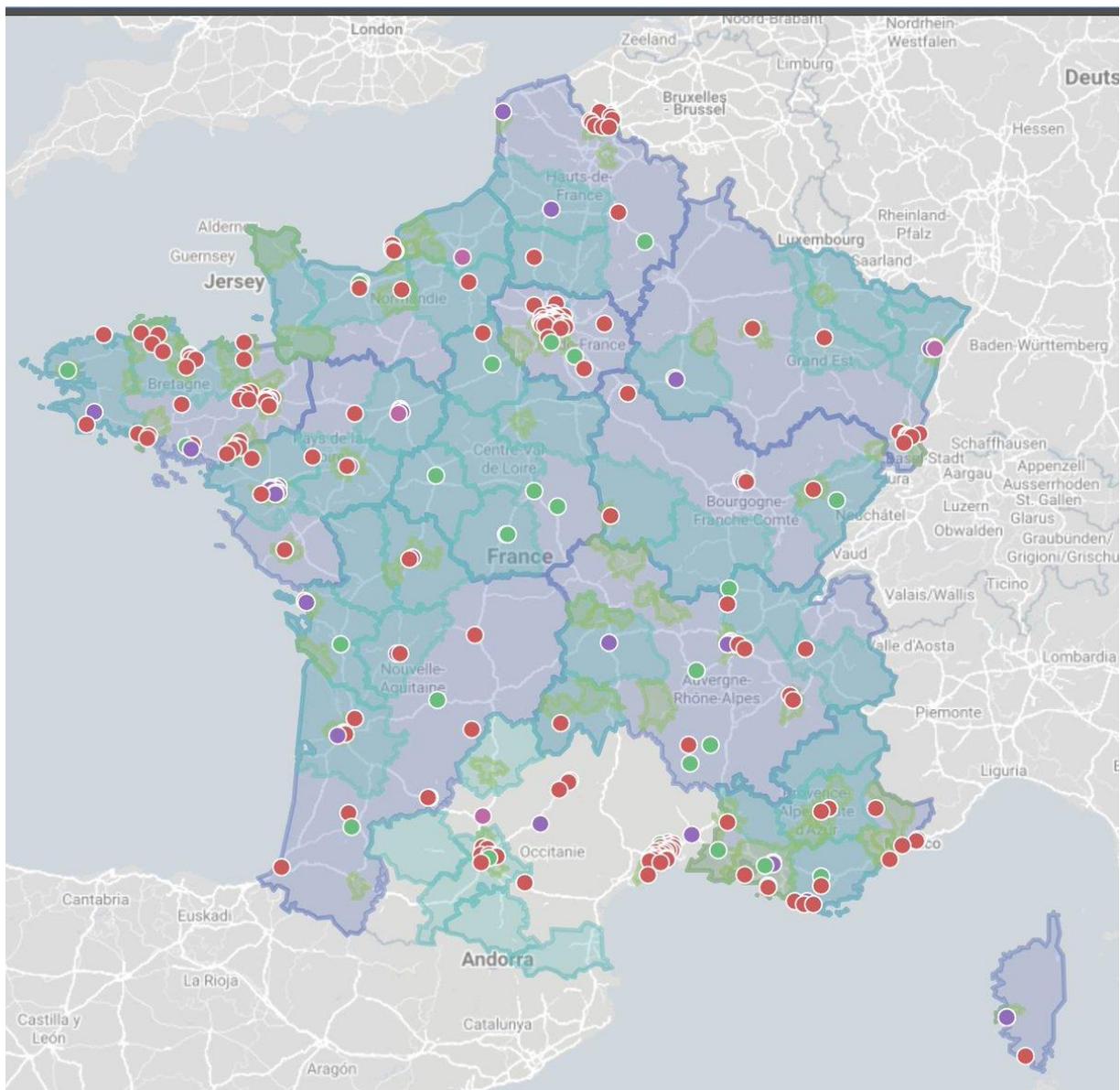
Toutes les collectivités territoriales de plus de 3500 habitants ou de 50 agents, l'Etat, et d'autres acteurs exerçant une mission de service public, ont l'obligation de publier leurs données et les mettre à disposition du public. Par exemple :

- La ville d'Angers publie des données comme les aménagements des voies cyclables, les zones de stationnement, la nomenclature des déchets ou encore les résultats des élections européennes. Ces données ont permis la création ou l'enrichissement d'applications ou de sites web déjà existant.
- Montpellier met à disposition différentes données comme le plan local d'urbanisme, la disponibilité des places de parking, les offres de transport en temps réel... Certaines données donnent ensuite lieu à des applications ou complètent des bases de données d'applications existantes. L'application Montpellier historique permet de se balader dans les rues de la ville et de connaître leur histoire, Handimap permet de calculer les itinéraires accessibles aux personnes à mobilité réduite, Freetaps recense les fontaines...
- A Combs-la-Ville, les internautes peuvent accéder aux marchés publics, aux délibérations des conseils municipaux, au répertoire national des élus, etc.
- En juillet 2019, la ville de Montreuil a publié ses données en open data. Mais, comme on peut le lire dans un article du Parisien, « *d'après l'Observatoire Opendata des*

territoires, seules trois autres municipalités du département [Seine-Saint-Denis] ont publié des jeux de données sur le site data.gouv.fr, une plate-forme globale : Rosny-sous-Bois (6 seulement), Villemomble (10) et Les Lilas (71) ».

- Le musée Gustave Courbet à Ornans dans le Doubs publie des données gratuitement comme la liste des œuvres de l'artiste. Ces données sont ensuite reprises sur Doubs Data, la plateforme de données du département.

Ce ne sont pas uniquement les collectivités stricto sensu qui sont concernées par cette obligation. Toutes les activités parapubliques comme les archives, bibliothèques, musées ont aussi l'obligation de publier leurs données en open data. Il leur est dorénavant possible de faire payer une cotisation ou une redevance.



3.1.2. Les institutions publiques et administrations

Les institutions publiques telles que les ministères, les établissements scolaires ou les collectivités territoriales fournissent la plupart des données. Mais il est aussi très facile pour des particuliers de mettre en ligne des données en open data ou de partager leurs réutilisations.

Sur la plateforme open data du ministère de l'Éducation nationale, les internautes peuvent accéder à la liste des établissements fermés, aux adresses des établissements, aux résultats du brevet, aux programmes d'enseignement... Les réutilisations des données ont permis la mise en place de CartoPédago, un outil cartographique permettant aux enseignants de visualiser les lieux et activités culturels et sportifs en lien avec leur programme pédagogique et à proximité de leurs établissements.

Les établissements publics administratifs comme l'Institut national de l'information géographique et forestière (IGN), Météo France et le Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (Cerema) produisent et détiennent un grand volume de données.

- L'**IGN** met à disposition une offre gratuite et utilisable sous licences ouvertes. Dès 2009, l'IGN s'est engagé à faciliter le libre accès à ses données en proposant leur gratuité pour l'enseignement et la recherche, puis la gratuité de ses photos aériennes historiques. En janvier 2011, l'IGN a franchi un pas significatif en mettant en place la gratuité du référentiel à grande échelle (RGE®), pour tout usage relevant d'une mission de service public ne revêtant pas de caractère industriel ou commercial, et en offrant une gratuité d'utilisation d'une partie de ses services en ligne. Depuis 2012, l'institut poursuit sa démarche d'ouverture en diffusant gratuitement un nombre croissant de ses données et en enrichissant les contenus des services en ligne utilisables gratuitement.
- Le **Cerema** pratiquait jusqu'à présent une vente de données qui lui rapportait moins de 500.000 € par an, mais cette activité devrait prendre fin car le Cerema n'est pas listé comme établissement habilité à percevoir des redevances pour la réutilisation de ses données.

La Cour des comptes note cependant des recettes en baisse. « *En 2013, les ventes de données brutes représentaient encore 36 % (14,7 M€) du chiffre d'affaires de l'IGN : elles n'en représentent plus que 12 % (6,6 M€) en 2017.* » Aussi, « *les recettes grand public de Météo-France sont passées de 17,8 M€ en 2012 à 9 M€ en 2017* ». A noter qu'en 2019, la Cour des comptes pointe leurs « *difficultés récurrentes et multiples pour se conformer au droit : l'obligation de publication de répertoires des principales données détenues n'est pas respectée ; de nombreux logiciels et certaines de leurs bases de données, qui pourraient être qualifiés d'informations publiques, ne sont pas mis à disposition du public* ».

3.1.3. Comment les entreprises exploitent-elles l'open data ?

La mise en œuvre de l'open data diffère selon le type d'entreprise, sa volonté de s'engager dans ce processus et le secteur sur lequel elle opère.

- Pour le **groupe EDF**, l'open data consiste à publier, sur une plateforme ouverte et centralisée des informations, des statistiques, des cartographies, des données économiques, financières, environnementales ou sociales, et de rendre accessible ses données numériques à tous ses salariés et partenaires, collectivités, start-up... Le projet est en passe de devenir réalité. « *Ces données, détaille Catherine Fraissenon, chef de projet data et transformation numérique, sont celles que l'on publie déjà mais qui sont aujourd'hui éparpillées sur nos sites.* » Il est, par exemple, assez aisé de trouver le nombre de salariés du Groupe dans le monde mais beaucoup moins simple pour la France.
- **Okeenea**, PME française, qui équipe la ville de New York avec ses feux sonores sécurisant la traversée des piétons malvoyants, utilise des données en open data de la métropole (chantiers en cours, voies cyclables, localisation de l'arrêt de bus, etc.).
Rappelons qu'en 2012, Michael Bloomberg, alors maire de New York, a imposé à toutes les institutions publiques de la ville d'ouvrir leurs données et de les mettre à disposition des citoyens sur un portail unique. Les jeux de données sont depuis regroupés en diverses catégories : santé, transport, environnement, éducation, entreprise, administration, logement, sécurité publique, services sociaux et loisirs. Cette libération des données s'est aussi accompagnée de la mise à disposition de plusieurs incubateurs et de pôles de recherche universitaire, encourageant l'émergence de nouveaux services et start-ups. Le prix NYC BigApps récompense les meilleurs projets issus de ce mouvement d'ouverture des données.

Des hackathons pour encourager l'innovation et la création d'entreprises

Pour faciliter l'émergence de nouveaux services ou de start-ups, collectivités, fédérations et grandes sociétés n'hésitent pas à lancer des « hackathons ». L'idée consiste à convier développeurs, chercheurs, étudiants et volontaires à réfléchir à des applications, projets d'entreprises sur une thématique (immobilier, tourisme, RH...), comme par exemple :

- En 2017, l'Assurance maladie a organisé un hackathon en partenariat avec la Fédération française des diabétiques. Le thème était de créer un service ou une application contribuant à améliorer la prévention et la prise en charge du diabète. Pour ce faire, les participants pouvaient utiliser les données de l'Assurance maladie et celles rendues publiques par d'autres organisations. Les sélectionnés ont ensuite pu mettre au point leur projet en bénéficiant d'un accompagnement.

- A l'occasion de la Journée internationale des données ouvertes qui s'est tenue en mars 2018, l'Ardèche a organisé un hackathon "Tourisme et mobilité", ouvert à tous. L'objectif était de faire éclore des idées novatrices à partir de données touristiques disponibles en open data fournies par le département.
- Un hackathon du cinéma est organisé tous les ans permettant d'utiliser les données du portail Data cinéma. Une vingtaine de candidats (designers, développeurs, porteurs d'idées, professionnels ou étudiants) ont été invités aux Arcs pour élaborer et présenter leurs idées devant un jury.

3.1.4. Les Data brokers

Les courtiers (data brokers) comme dans toutes les « commodités » sont des acteurs essentiels de ce marché. On peut citer les plateformes Acxiom, Experian et Epsilon. Elles récupèrent les données publiées sur tous les sites des administrations, achètent des data (banque, organisme de crédit, entreprise de grande distribution. Ces sociétés spécialisées dans l'intermédiation ont pour vocation d'acheter le maximum de données personnelles et les revendre en réalisant des bénéfices records.

Ces data brokers, qui captent les données laissées par les utilisateurs sur la toile, les croisent et les analysent pour les revendre à de riches clients ou d'éminentes organisations publiques. Celles-ci s'en servent ensuite pour cibler leurs clients et ainsi leur proposer des produits ou des offres mieux adaptées à leurs besoins, envies ou habitudes. Il s'agit donc d'un troisième niveau de captation de la donnée. Par exemple, sur Facebook, un utilisateur renseigne son intérêt pour les yaourts nature. Facebook, plateforme d'intermédiation, va capter ces données et pouvoir lui proposer des publicités adaptées. Mais il va pouvoir également les vendre aux sociétés de data-brokers. Une nouvelle strate commerciale en plein développement. La société américaine Datalogix, par exemple, traque les data issues des transactions bancaires transitant par les sites de commerce en ligne ou chez les commerçants et magasins physiques. Ces données amassées sont ensuite vendues à des groupes comme Facebook ou Google pour les aider à mieux cibler leurs offres publicitaires. C'est donc un véritable écosystème de la donnée où dialoguent plateformes d'intermédiation, data brokers et autres clients publics ou privés. Mais contrairement aux plateformes d'intermédiation, les data brokers ne dispensent aucun service pour leurs utilisateurs.

3.1.5. Les associations & collectifs

Dans l'écosystème de la data, les associations et collectifs sont actifs.

- Fondée en 2012, **Open Food Facts**, association à but non lucratif, s'appuie sur les contributions de 9000 volontaires et référence à ce jour 600 000 produits dans 200 pays. C'est un projet collaboratif en ligne et mobile dont le but est de constituer une

base de données libre et ouverte sur les produits alimentaires commercialisés dans le monde entier.

Chaque contributeur peut ajouter et modifier des fiches de produits alimentaires, en basant les ajouts effectués sur les données inscrites sur les emballages. De ce fait, le GTIN (Global Trade Item Number) encapsulé dans le code-barres présent sur le packaging constitue l'identifiant de la fiche du produit alimentaire. Des applications pour téléphones mobiles permettent la capture de photos et d'informations qui sont ensuite retraitées manuellement par les volontaires, assistés par des algorithmes de vérification, ainsi que des systèmes basés sur l'intelligence artificielle.

- **Yuka**, lancée en janvier 2017, est une application qui s'appuie sur la base de données Open Food Facts, qui répertorie ingrédients, allergènes, composition nutritionnelle. Yuka représente en France plus de 11 millions d'utilisateurs en France.

Yuka a également ouvert la voie à de multiples concurrents qui cherchent à se différencier par les usages de ces données. Les derniers en date sont, la marque "C qui le patron" qui lance "C'est quoi ce produit ?" et l'association "60 millions de consommateurs". Face à cette concurrence, Yuka a cessé d'enrichir Open Food Facts avec ses propres retours de consommateurs via son application mobile, car elle alimente aussi ses concurrents. Yuka représente en France plus de 11 millions d'utilisateurs en France et un réseau d'ambassadeurs/utilisateurs. Cette bataille montre que les données ont de la valeur pour ceux qui savent les exploiter, et que l'open data n'est que le début d'une longue histoire de valorisation.

- **PartiCitaE** a pour vocation d'associer la société civile à la recherche. Cet observatoire scientifique est situé au sein de l'Université Paris Sorbonne. Il s'appuie sur un dispositif participatif qui assure la prise en compte des besoins sociétaux dans la recherche sur l'environnement. Mesure de la qualité de l'air, observation de la nature en ville (lichens) figurent parmi les thèmes proposés par les chercheurs qui poussent assez loin la notion de participation des citoyens. Ces derniers sont ainsi invités à construire le protocole de collecte des données et les outils permettant cette collecte.
- **Doctrine.fr** permet, grâce à son moteur de recherche, de trouver en quelques clics des décisions de justice en open data, à partir de plusieurs jeux de données différents.

Cela se traduit aussi par l'implication d'associations comme Regards Citoyens, Libertic, la Fondation internet nouvelle génération mais également le chapitre français de l'Open Knowledge Foundation soutiennent activement le mouvement open data. Dans le secteur privé, les entreprises peuvent aisément ouvrir leurs données grâce à la plateforme clé en main créée par OpenDataSoft.

3.2 Le cas des énergéticiens en Europe

Sur le seul secteur de l'énergie, la mise en œuvre de l'open data se caractérise est abordée de manière hétérogène par les acteurs.

Les différents énergéticiens européens sont divisés en quatre catégories (cf : tableau ci-dessous) :

- Certains énergéticiens (en rouge) ne proposent pas d'open data.
- Les énergéticiens en jaune proposent uniquement des données brutes sans traitement préalable sous un format Excel.
- Les énergéticiens en vert sont ceux qui sont les plus en avance en Europe. En plus de proposer des fichiers à télécharger, il est possible d'avoir accès à de services comme le suivi instantané des flux, des prévisions sur la consommation ou sur la production.
- Dans certains pays (en noir) comme la France, la Finlande ou la Suisse, les énergéticiens se démarquent en proposant un large panel de services open data.

Pays	Énergéticiens	Position vis-à-vis de l'open data
Allemagne	Tennet	Offre professionnelle
Allemagne	TransnetBW	Open data visuel
Autriche	Verbund	Absence de services open data
Belgique	Elia	Open data visuel
Espagne	Red Eléctrica de España	Open data visuel
Finlande	Fingrid	Services open data les plus avancés
France	Enedis	Services open data les plus avancés
France	RTE	Services open data les plus avancés
Italie	Enel	Services open data les plus avancés
Norvège	Statnett	Open data visuel
Pays Bas	Tennet	Offre professionnelle
République Tchèque	?EPS	Offre professionnelle
Royaume Uni	National Grid	Open data visuel
Suède	Svenska Kraftnät	Services open data les plus avancés

Suisse	Swissgrid	Services open data les plus avancés
--------	-----------	-------------------------------------

En Autriche, il existe plusieurs gestionnaires des réseaux. Diffuser ses données peut être ressenti comme la perte d'un avantage concurrentiel.

En Allemagne, les énergéticiens, comme TransnetBW, ont fait le choix de l'open data et pourraient bien perfectionner leur offre.

Et en Europe, la majorité des énergéticiens a fait le même choix en offrant à minima des données à télécharger.

Reste le cas particulier du **Royaume-Uni**. Le Brexit remet en question l'investissement des énergéticiens britanniques dans l'open data.

En conclusion, l'approche open data pour les énergéticiens européens ne semble pas leur apporter une réelle valeur ajoutée.

3.3. Business model

La donnée est une matière première qui en fonction de la stratégie adoptée par l'entreprise peut être considérée comme un actif stratégique ou un levier d'opportunités. En soi, elle n'a pas grande valeur, hormis pour celui qui la collecte. C'est la mise en relation avec d'autres données qui crée la valeur et leur donne une nouvelle nature en créant une information.

La donnée acquiert donc de la valeur par son partage, au moment où elle circule.

La partie émergée de l'iceberg, ce sont ces startups qui utilisent des données afin de créer de nouveaux services "out of nowhere". Ces services proposés se basent sur des données extrêmement détaillées, récupérées via l'open data. Ces sociétés sont des pures players de la data.

La partie immergée de l'iceberg, ce sont toutes ces entreprises qui utilisent de la donnée sans que cela soit leur cœur de métier. Elles l'utilisent pour améliorer leurs processus, pour décider mieux, plus vite, pour prendre des décisions sur le terrain et plus uniquement en central, etc.

Un modèle économique doit concourir à la création de valeur. Pour ce qui est de l'open data, l'approche n'est pas si simple. La France forme des experts de la data : ingénieurs, mathématiciens, statisticiens, ce qui est un atout. Le secteur public est entré dans le mouvement de l'open data, par la force de la loi. Une interrogation persiste pour le privé.

Dans son article « Open data et business models », Étienne Roché définit et détaille les différents coûts et gains liés à la mise en place d'un projet open data.

Les **coûts fixes** s'appliquent à tout type de structure, publique ou privée, qui porte le projet d'open data. Il s'agit de coûts incompressibles de plateforme de distribution, généralement proportionnels au volume des données stockées et aux traitements associés.

Ils s'appliquent à :

- La nature des infrastructures (acquisition, location, exploitation, maintenance d'une plateforme informatique)
- La collecte de données
- Le nettoyage est aussi important pour en assurer la cohérence, la qualité, le respect de référentiel donné. Ces données sont donc traitées pour être utilisables.
- Le retraitement, qui peut concerner des données additionnelles à partir de données collectées et nettoyées comme par exemple une position GPS et un code postal du territoire concerné.
- L'intégration, il s'agit de la mise en relation des données à d'autres données sur la plateforme informatique, comme l'arrivée d'un bus, d'un RER avec celle d'une personne.
- La diffusion qui se révèle être des coûts techniques (sécurité, serveurs, bande passante...) pour garantir la mise à disposition et l'accessibilité, au regard du service proposé.
- Le juridique relatif à la création de licences.
- L'administration : coûts relatifs à l'organisation pour gérer les acteurs autour de la plateforme.
- La promotion enfin, peut aussi faire partie des coûts de marketing, de vente ou de recrutement de personnels compétents.

Qui participe au financement ?

Plusieurs acteurs peuvent intervenir tout au long de la chaîne de financement, comme :

- **l'Etat ou un opérateur public**, c'est alors l'impôt ou la taxe qui prend en charge ces coûts. L'open data est alors assimilable à un bien commun, avec sa réglementation associée.
- **Un ensemble de partenaires** : le financement est alors collectif et réparti en fonction de ce qui aura été contractualisé entre chaque participant.
- **Un mix privé-public** : le maintien de l'offre publique est assuré par les revenus de la partie privée.
- **Le participatif**, seule l'infrastructure est mise à disposition. Le temps et la force de travail des volontaires ne comptent pas. Le modèle peut être fragile selon les territoires.
- **Un acteur privé**, toutes les formes de partages peuvent exister, et sans doute nécessitent une contractualisation engageant les parties. Il y a sans doute des calculs faits par les différents acteurs sur leurs gains à le faire.
- **La contribution** : les biens communs peuvent être maintenus : tout acteur voulant œuvrer dans ce domaine doit participer au maintien de ce bien (donation par ex : Wikipédia ou Khan Academy).

Trois types de business models associés aux revenus

- **Modèle freemium**, classique en e-business. Ce modèle (contraction de free et premium), récupère des données ouvertes gratuites fournies par une plateforme

d'open data pour les utiliser dans une offre de service à plus haute valeur ajoutée qui sera payante (=premium). Dans le cas du modèle freemium, la valeur se décline de différentes façons, non exclusives :

- l'accès : les données peuvent être explorées avec plus ou moins de facilité. Des outils optionnels peuvent être mis à disposition pour extraire, manipuler, rapprocher les données présentes
 - la mise en forme : par défaut, les données sont mises à disposition sous une forme et/ou un format normalisé. Des formats additionnels peuvent être apportés (génération de rapports, extraction vers un support...)
 - la fraîcheur de l'information : il s'agit de la période de temps écoulée depuis la mise à disposition de la donnée sur la plateforme. Par défaut, l'accès à la donnée est gratuit mais il peut être rendu payant à la suite d'une certaine durée d'embargo
 - le retraitement de l'information en une nouvelle information : de façon optionnelle, la plateforme peut mettre à profit une expertise et une capacité de calcul particulière pour fournir une information dérivée de la première ;
 - l'activation d'une information supplémentaire à l'information disponible : il est possible d'ajouter une donnée additionnelle également disponible sur la plateforme mais de façon automatique
 - la capacité à récupérer l'information in extenso : selon les conditions techniques d'exploitation de la plateforme, il est possible de ne pas limiter le débit d'accès à l'information. Si cette soudaine absence de limitation semble aller à l'encontre du principe de non-rivalité du bien commun, il faut comprendre que la limitation initiale était plutôt là pour garantir l'accès à tous les acteurs de façon équitable
 - le nombre de sollicitations horaires : il s'agit d'une précision du point précédent garantissant la capacité à accéder à la plateforme avec une certaine intensité
 - l'accessibilité et les niveaux de services (SLA) : dispositif contractuel classique chez les hébergeurs et les fournisseurs de solutions en ligne pour s'engager sur garantie de disponibilité du système avec éventuellement une priorité d'accès à l'information en termes de vitesse
 - la licence payante : pour une exploitation privée des données, il est possible de créer une licence payante en marge de la licence en vigueur sur la plateforme (principes de « duallicense » et « share-alike » licences).
- **Financement croisé (Cross Subsidy)** : il s'agit d'un modèle plus évolué que le précédent et qui suppose une certaine attractivité de la plateforme. Le principe est de pouvoir faire cohabiter d'autres business models de façon symbiotique ; dit autrement, un business model ne peut in fine se passer de l'autre. Certains services peuvent être alors envisagés :
 - le conseil associé à la donnée : schéma de consulting similaire à celui de l'open source
 - la personnalisation de la donnée : il s'agit d'ajouter un étage supplémentaire de personnalisation par client de la plateforme agrégeant services et autres

informations dont seul ce dernier a besoin et indépendamment des besoins des autres

- l'outsourcing de la R&D (recherche et développement) : on peut citer, par exemple, la plateforme Kaggle, en Data Science, qui met à disposition des données issues d'acteurs industriels (sur leur propre initiative) qui souhaitent faire résoudre un problème par la communauté ou a minima récupérer de nouveaux points de vue (très précieux)
 - la publicité et l'image : schéma classique de partenariat d'image.
- **Modèles de réseaux et collaboratifs** : ce dernier type de business models porte sur les modèles les plus ambitieux en termes de mise en relation d'acteurs :
 - le crowdsourcing et les efforts distribués : le principe consiste à mettre à disposition d'une communauté la plus large possible une plateforme pour collecter une donnée normalement coûteuse, longue à acquérir ou nécessitant une fréquence de mise à jour élevée si elle était faite par un petit nombre d'acteurs. Il s'agit de s'allier à la multitude comme le décrivent parfaitement Henri Verdier et Nicolas Colin dans « L'âge de la multitude »³. Le meilleur exemple est la plateforme mondiale de cartographie Open Street Map (<http://openstreetmap.fr>) dont les intégrations dans d'autres plateformes industrielles sont de plus en plus nombreuses. Les données open data peuvent être d'origine publique ou privée. Cet écosystème comprend des acteurs de la société civile comme Open Data Institute ou l'Open Knowledge Foundation ou encore OpenStreetMap, des administrations publiques, des entreprises privées, des acteurs du monde de la recherche, des acteurs du monde du journalisme, des financeurs et des organisations internationales.
 - la collaboration public/privé : legislation.co.uk est le site officiel de publication des textes de loi anglaise. Comme le parlement anglais ne publie pas de texte en entier mais uniquement des changements se rapportant à une version précédente, seules les Archives nationales s'occupaient de consolider l'ensemble. Deux sociétés dans ce secteur (LexisNexis et Westlaw) travaillent à présent de façon conjointe pour avoir une version des textes la plus proche qui soit de la réalité afin que tout le monde puisse en bénéficier ;
 - la co-conception : le principe est de partager les plans d'un bien commun en vue de fédérer les différentes expertises s'y rapportant et d'itérer de façon collégiale. On retrouve ici ce qui a fait le succès de l'open source qui, stricto sensu, est une forme d'open data mais pouvant être appliqué à de nombreux domaines dont le secteur en forte croissance de l'open hardware.

3.4. Des exemples réussis de projets d'open data

3.4.1 En France.

En France, de nombreux exemples ont vu le jour comme :

- **MapBox** : le freemium à fonctionnalités variées. Ce fournisseur de services cartographiques est construit sur les données ouvertes de OpenStreetMap.
- L'application mobile **Sharette** met en relation les passagers et les conducteurs qui se déclarent sur le départ au même moment, au même endroit. Pour vérifier que les itinéraires de chacun correspondent, l'application se reporte aux données Openstreetmap (un Google Maps libre de droit). Lorsqu'il est plus intéressant de commencer un bout de trajet en transport en commun avant de monter en voiture, Sharette propose un trajet personnalisé grâce aux informations open data de la SNCF et de la RATP.
- L'application **Tranquilien** développée en partenariat avec la SNCF permet aux utilisateurs franciliens de déterminer l'affluence de voyageurs sur le réseau ferré et d'opter pour un itinéraire alternatif.
- **VroomVroom** s'attaque au marché des auto-écoles en offrant à ce stade un comparateur.
- **Kel Quartier** dresse le portrait-robot des habitants de chaque quartier de France métropolitaine. La start-up rassemble et revend aux professionnels de l'immobilier, agences bancaires ou encore aux assureurs, des données géolocalisées à l'échelle d'un quartier, qu'elles soient publiées librement ou sous licence propriétaire.

Le contre-exemple : Fourmisanté, créée en 2010, est un comparateur de tarifs médicaux. La mise en demeure de la Caisse Nationale d'Assurance Maladie stoppa cette initiative. La Cnam a demandé de retirer de son site les données publiques concernant les prix des médecins de secteur 1 et secteur 2. Fourmisanté les complétait par des informations venues des laboratoires et des associations de patients. À présent, le site continue de travailler avec d'autres données publiques. Il aide par exemple les internautes à évaluer le taux de remboursement sur les médicaments qu'ils veulent acheter. Un nouveau marché s'est récemment ouvert : la vente de médicaments est autorisée sur Internet. Fourmisanté essaye donc de devenir le comparateur de prix des médicaments.

3.4.2 Aux Etats-Unis

De nombreux exemples ont vu le jour Outre-Atlantique :

- **OpenCorporates** : est une start-up britannique, incubée à l'Open Data Institute, centralisant les informations publiques de plus de 77 millions d'entreprises dans le monde. Elle s'appuie sur le freemium fondé sur les finalités d'usage.
- **Enigma.io** est une plate-forme de recherche et de consommation de données ouvertes. Enigma.io obtient des données auprès des agences fédérales américaines et des entreprises.
- **Airdna** (Airbnb Data & Analytics), lancé en 2014, par Airbnb, propose une offre de services autour de la data à destination des loueurs. Les données de taux d'occupation des pied-à-terre, de demande saisonnière, de pricing par rapport à la localisation, possédées par Airbnb, sont notamment mises à la disposition des particuliers et des entreprises.

Aux Etats-Unis, un mouvement de protection des data a vu le jour ces dernières années, notamment autour des questions sur le réchauffement climatique. L'Université de Toronto a récemment organisé un événement visant à copier les pages de données en ligne menacées par l'administration Trump. De nombreux groupes sont actuellement engagés dans cette démarche de sauvegarde.

3.5. Contrôle et vérification des données

Le contrôle et la vérification des données représentent un enjeu stratégique, et posent des questions de légitimité. Qui détient le pouvoir sur la data : l'utilisateur, la plateforme... ? Quelle instance peut se placer au-dessus du producteur de données et s'assurer de la véracité des informations mises en ligne ? Les données sont mises en ligne par obligation (administrations, collectivités...) ou librement par des internautes. Mais les fichiers sont-ils ensuite relus, vérifiés ? L'information est-elle recoupée ?

De nombreux jeux de données en France sont incomplets ou obsolètes. Ils perdent alors immédiatement en valeur et sont parfois inutilisables.

Dans d'autres cas, une boucle de vérification est mise en place. Mais ce sont les internautes eux-mêmes qui vérifient les jeux de données publiés par d'autres contributeurs. C'est le cas d'Open Food Facts, dont les données ont servi à enrichir et développer Yuka. Au fur et à mesure du succès grandissant de l'application, les marques sont devenues contributrices et ont publié des données sur la composition de leurs produits. Les informations sont ensuite vérifiées par les autres contributeurs. Cependant, quelle est la légitimité des producteurs de données et de ceux qui les vérifient ? Qui sont les mieux placés : industriels, internautes... ?

Cette absence de contrôle par une autorité compétente permet facilement le trucage, l'omission voire la suppression de certaines informations.

Un scénario de conflit économique peut être envisagé :

Concernant les données cadastrales publiées sur le site <https://app.dvf.etalab.gouv.fr/>, chacun peut consulter librement une carte interactive indiquant le prix des ventes immobilières réalisées ces dernières années. Si les données proviennent de registres officiels (la base est produite par la direction générale des Finances publiques), n'y a-t-il aucune erreur possible ? Dans le cas d'une récupération malveillante voire d'un scénario de guerre micro-économique, un trucage des données en vue de faire descendre les valeurs foncières réelles d'un quartier pourrait facilement être mis en place. Dans le cas contraire où les données sont exactes et exhaustives, quid d'une estimation réalisée par un professionnel ? A-t-elle encore de la valeur ?

4. L'impact de l'open data sur les questions de souveraineté

4.1. Souveraineté économique : risque d'attaque et dépeçage des data

Le problème soulevé par l'open data est la mise à disposition de données publiques à ceux qui ont la capacité de les réutiliser qui sont en général ceux qui déjà maîtrisent la capacité d'exploitation de données. Il y a un objectif vertueux en plus de celui d'une « démocratie transparente » à vouloir mettre à disposition ces données publiques voir des acteurs privés s'en saisir pour développer un autre écosystème plus ou moins local.

Les « data brokers » ou plateformes d'intermédiation vont pouvoir être l'interlocuteur direct des communes ou des collectivités. Ils vont se servir des données publiques, couplées à d'autres services, rendant plus riches la relation avec ces « clients ».

A terme le citoyen n'ira plus sur le site de la commune, mais utilisera la plateforme. Il ne sera plus en direct avec ses élus. Un intermédiaire sera entre lui et eux. Quelques plateformes plus modestes d'intermédiation pourraient faire payer ce qu'aujourd'hui, les citoyens ont gratuitement en étant en direct avec les élus ou le site de la commune, etc.

Il reste à établir comment conserver les avantages qu'auront les utilisateurs à accéder à des services démultipliés, fournis via une plateforme sans que la souveraineté « publique Etat, régions, communes » n'en soient affaiblie. L'émergence de champion numérique français est peut-être une solution, que les derniers gouvernements encouragent.

Opendatasoft, entreprise française intervenant à l'international, propose des services aux collectivités et aux entreprises pour mettre à disposition leurs données, en créant un standard technique, lui appartenant. Son objectif vise à créer une norme afin de soumettre et créer une dépendance cognitive.

4.2. La monétisation des données

La monétisation des données consiste à créer des revenus, en plus du chiffre d'affaires traditionnel de l'entreprise, en utilisant les données à des fins commerciales (vers l'externe) ou en les exploitant (en interne).

D'après Barbara H. Wixom, chercheuse au MIT, dans son livre *How to Monetize Your Data*, publié en 2017, il existe trois grandes façons de monétiser les données. La première consiste à vendre les données directement, de manière autonome ou en faisant appel à un Data Broker en guise d'intermédiaire. La deuxième est celle de l'échange d'informations. Il est possible d'échanger les données contre d'autres données, mais aussi contre des biens ou des services. Les données peuvent également permettre de négocier des arrangements avec ses partenaires. La troisième consiste à améliorer les performances de l'entreprise ou à créer de nouveaux produits en se basant sur les données.

De nombreuses entreprises cherchent à monétiser leurs données. Pour les géants du web comme Google et Facebook, la Data Monetization est même devenue une activité centrale. Mais elle est risquée à plus d'un titre et soumise à des réglementations strictes (le RGPD en Europe). Rappelons le cas de Facebook qui s'est enrichi en échangeant des données individuelles avec des partenaires, mais a perdu la confiance de ses utilisateurs.

Quelle valeur ajoutée possible pour une entreprise ?

Plusieurs options s'ouvrent : d'un côté une version open data a minima, de l'autre une version payante tarifée au volume, pour des fréquences de mises à jour plus importantes, pour un engagement de correction ou encore pour un support technique.

L'entreprise s'attire de nombreux prospects et préserve son capital en évitant que ses clients ne partent à la concurrence, les fidélisant en leur permettant de tester très tôt son produit. Elle s'assure en outre une image positive auprès des connaisseurs et notamment de ceux qui, de toutes façons, n'auront jamais les moyens d'acheter ces données.

Les entreprises se protègent et savent aujourd'hui chiffrer la perte occasionnée. Pour le marché de la monétisation, Meteosat a évalué la perte occasionnée à 120 000 € par client.

La monétisation des données personnelles

Depuis l'entrée en vigueur du RGPD, les internautes européens deviennent légalement propriétaires de leurs données personnelles sur Internet. Dans la pratique, l'usage des données personnelles a déjà été préempté par de nombreux acteurs dont les GAFAM. Les utilisateurs de Facebook par exemple, acceptent implicitement de donner leurs données personnelles, et d'en perdre le contrôle, afin de bénéficier pleinement des services du géant américain.

Avec la RGPD, la négociation de données personnelles a changé d'échelle. Des entreprises de courtage spécialisées, les **databrokers**, brassent des milliards de données provenant de millions d'internautes à travers le monde et les vendent à des entreprises de tous secteurs qui souhaitent identifier des prospects intéressés par leurs produits. Par exemple, une banque peut demander au courtier un fichier de 200 000 personnes susceptibles d'acquérir un bien immobilier dans l'année, au tarif de quelques centimes par contact. 6 millions de foyers français figureraient ainsi dans la base d'Acxiom, l'un des neuf courtiers les plus puissants, qui généraient en 2012 un chiffre d'affaires total estimé à 426 millions d'euros, selon l'Université de Caroline du Nord.

Et le filon du commerce de données n'est pas prêt de se tarir. L'essor des objets connectés (IoT) conjugué à l'utilisation de plus en plus intensive des réseaux sociaux multiplie les occasions de collecter des données. Le Boston Consulting Group prévoit ainsi que la valeur des données personnelles en Europe pourrait atteindre 1000 milliards d'euros d'ici 2020, soit 8 % du PIB européen.

Mais est-ce que le citoyen pourra monnayer les données personnelles ? Si tous les citoyens se mettent à contractualiser avec les entreprises intéressées, la valeur de leur data baissera, sur une simple projection de la loi de l'offre et de la demande. Par ailleurs, pour le cas de la France, si chacun contractualisait en direct, quels moyens de protection la CNIL pourrait-elle mettre en œuvre pour protéger les citoyens français ?

Les données personnelles ont été évaluées par un sondage auprès du public à 500 € : l'internaute est prêt à « vendre ses données personnelles » pour ce prix. La question est de savoir si les données personnelles individuelles ont une valeur alors que c'est la quantité qui en fait la valeur et l'interconnexion d'usage.

4.3. Bouclier défensif : vers la création d'un service public de la data ?

- **Prendre le leadership européen**

Dès le début des années 2000, l'Europe a mis en place une politique d'ouverture des données. Elle a rapidement mentionné une possible monétisation et, plus largement, fait part de sa vision en termes d'accroissement de valeurs à travers la mise en place de nouveaux services, la création d'emplois et le développement économique.

Quinze ans plus tard, l'Europe affine sa réflexion et évoque un « marché unique numérique ». La mise en place de l'open data est à ce stade encore très inégale. Certains pays membres publient de nombreuses informations, tandis que d'autres ne mettent rien en ligne. Il n'existe aucun organe de contrôle, ni de sanction pour les pays en cas de non-application des directives relatives à la publication des données en format ouvert.

Pourtant, pour faire face au développement des GAFAM, en particulier dans le domaine des données de santé, et des BATX en matière de contrôle social, l'Europe mène une réflexion sur une future politique européenne en matière de data. Aucun pays n'a ouvertement pris le leadership sur la question. La France a un rôle à jouer.

L'open data présente une opportunité pour la France de montrer la voie et de reprendre la main sur sa souveraineté numérique et ses informations stratégiques nécessaires à son

accroissement économique et social. Il s'agit de regagner en autonomie, mais aussi d'utiliser les données comme un rempart, une monnaie d'échange, un bouclier défensif.

Cependant, il faut une impulsion politique claire. Malheureusement, en France, l'open data est loin d'être une priorité. Il a fallu attendre 13 ans entre la première directive européenne et la mise à disposition des premières données publiques, dans un format libre et ouvert. Autant d'années perdues que la France aurait pu utiliser pour déployer systèmes et outils adaptés à cette transformation numérique.

Aujourd'hui, l'open data n'est ni gérée, ni pilotée, mais plutôt laissée à la portée de ceux qui veulent s'en occuper. Cette absence de pilotage fait perdre toute crédibilité à cette démarche, et laisse la porte ouverte aux intrusions, malveillances et récupérations. Les data représentent des failles dans notre sécurité nationale.

Le pays peut devenir fer de lance d'une Europe numérique forte, face à l'hégémonie numérique actuelle américaine, et à venir, indienne ou chinoise, et recouvrer sa souveraineté.

- **Un service public de la data**

Le service public de la data consisterait à aider les entreprises à bénéficier de données publiques, utiles à leur développement.

- Pour cela, la France doit **identifier les acteurs** relevant des intérêts économiques, industriels et scientifiques et centraliser les informations dont les groupes, sociétés, PME ont besoin pour créer de nouveaux services ou produits innovants.

Il est indispensable d'identifier les secteurs critiques liés à l'industrie française (mobilité autonome, hôpital & santé...), de demander à chaque acteur de définir les données nécessaires à son développement. Un service public pourrait leur fournir ces données dans un format facilement accessible et utilisable.

- Un **guichet unique des données** ou des délégations locales pourraient aller à la rencontre des acteurs économiques pour définir les data idoines. Des services « data » pourraient être mis en place dans les entreprises, afin de communiquer leurs besoins aux pouvoirs publics.

Les données pourraient ainsi être des outils au cœur du développement des entreprises, qui pourraient les intégrer dans leurs stratégies de knowledge management ou change management.

- Que ce soit pour les entreprises ou les administrations, des **formations** aux data sont à développer à grande échelle afin de former et sensibiliser chefs d'entreprise, responsables, salariés mais aussi agents administratifs sur les risques et les opportunités qu'elles représentent.
- Il est possible d'imaginer des **synergies**, notamment dans le transfert de données, pour limiter les coûts. Mais cela ne peut fonctionner que si les infrastructures techniques peuvent supporter un grand volume de données et si celles-ci sont pertinentes, mises à jour et à haute valeur ajoutée.

- **Améliorer la pertinence des données**

La qualification des données doit être au cœur des préoccupations du service public de la data. Pour être pertinentes, les données doivent être intéressantes, mises à jour, dans un format compréhensible, mais aussi être régionales. Rappelons que l'open data est un instrument de maîtrise de son espace. En effet, des données n'ont pas la même signification à Paris ou Limoges, par exemple, dans le cadre de développement de services liés à la mobilité. Il est nécessaire de connaître le contexte, pour pouvoir proposer des idées innovantes adaptées, avec l'aide d'experts locaux.

Chaque état a des données qualifiées qui correspondent à des problématiques propres à son espace géographique, aux cultures... Aujourd'hui, les GAFAM sont les seuls acteurs performants dans la collecte, le traitement et l'analyse des données géographiques à grande échelle. La France doit améliorer la pertinence de ses data et savoir leur donner du sens pour en exploiter tout le potentiel.

- **Vérifier et protéger les données**

L'open data pose des problèmes de sécurité et de fiabilité. Les données sont facilement manipulables et récupérables. Sans organe de vérification, toute erreur, volontaire ou non, peut subvenir et se glisser dans les informations partagées. Pour éviter tout risque d'intrusion ou manipulation subversive des fichiers, la France doit mettre en place une série d'actions préventives. A commencer par la traçabilité des données qui doit garantir leur fiabilité et leur sécurisation tout au long du « parcours technique » de la data (câbles, serveurs...). Pour s'assurer qu'aucune erreur ne s'est glissée, on peut imaginer la mise en place d'un outil de vérification des données ou de recoupage des informations. Il est également indispensable de sécuriser les données, serveurs, datacenters...

Dans le domaine de la santé, que penser d'une manipulation subversive des données de nos concitoyens ? Et ce, par un Etat ou une entreprise étrangère ? La France doit assurer la protection de son patrimoine immatériel. Les données sont des biens stratégiques et doivent être intégrées à toute politique de sécurité nationale. A titre d'exemple, les informations doivent être stockées sur des serveurs français, protégées par un droit adapté, afin de se prémunir de lois extraterritoriales comme le Cloud Act.

- **Vers une masse critique**

Les initiatives à petites échelles (notamment portées par des hackathons), développées par des PME, institutions et grands groupes ont été des échecs. Et si la faute était à imputer à la masse de données ? S'il suffisait en réalité d'une masse critique, de davantage de données, autrement dit de big data, pour mettre au point produits et services ?

La législation européenne incite états et collectivités à ouvrir toujours plus de données, qui sont aussi, de plus en plus personnelles.

- **Miser sur la technique et la formation**

En France, l'ouverture des données a été imposée aux communes de plus de 3500 habitants et administrations de 50 agents et plus, en 2015 puis 2016. La mise à disposition des données dans un format ouvert, nécessite des infrastructures techniques de qualité, qui représentent un coût pour les collectivités et administrations. Par ailleurs, de nombreuses infrastructures techniques étaient à l'origine de simples bases de données, qui n'étaient pas prévues pour le partage en format ouvert. A cela s'ajoute la nécessité d'avoir du personnel formé aux enjeux du big data. Mais à l'heure où les collectivités manquent d'argent et récupèrent régulièrement de nouveaux domaines de compétence, rien ne les incite à s'intéresser ni à s'investir dans l'open data, qui leur est pourtant imposé.

L'ouverture des données revient donc à des collectivités, qui ne sont ni expertes, ni formées, et n'ont pas le budget pour ouvrir leurs données correctement. Du côté des entreprises, même son de cloche. Selon la Fedma, Fédération européenne du marketing direct et interactif, « *certaines entreprises ont des bases de données où tout est imbriqué. Elles ne peuvent pas partager une partie et garder l'autre privée. Il y a des considérations techniques à prendre en compte* ». Face à l'aspect stratégique des données, il est important de donner aux administrations, collectivités et entreprises toutes les clés pour mettre en œuvre des politiques de données adaptées aux enjeux de la France.

4.4. Smart cities, encore l'influence des USA et de l'Asie



Ouvrir les données des villes... et après ? Pour quels usages d'utilité publique ? Comment les favoriser ? Comment partager les données d'intérêt général entre acteurs publics et privés ? Peut-on piloter les villes avec ces données ? Selon Mohammed El Kettani, urbaniste, l'open data constitue une nouvelle

arme puissante au profit de villes performantes. L'open data permet aux villes d'accompagner leurs ambitions économiques, sociales, politiques, énergétiques. En tant que productrice et récupératrice de données, la ville a toutes les cartes en main.

Ce pouvoir est à double sens puisque l'ouverture des données, si elle permet de gagner en connaissances est le garant de transparence de la vie politique en donnant au citoyen la possibilité de contrôler les actions politiques. L'open data est donc une brique importante des projets de smart cities. Elles représentent un capital numérique immense qui nécessite des expertises pour le traiter : urbanisme, aménagement, étude des systèmes urbains, architecture, environnement et énergies...

Aujourd'hui, le numérique envahit l'espace physique de la ville - indications de présence de Wifi, invitation à suivre les pages Facebook ou Twitter d'une institution, flash codes - mais ses usages sont en constante évolution. Chaque jour, 84 % des Américains possesseurs de smartphones utilisent une application de type open data. La ville connectée a une quinzaine d'années : le concept prend forme au début des années 2000 en Corée du Sud avec le projet de la ville ultra connectée Songdo, et est formulé pour la première fois par Bill Clinton en 2005 qui parle de « *smart city* » pour commander un plan de décongestion de plusieurs villes (San Francisco, Séoul, Amsterdam), visant à réduire les émissions de CO2 et à réaliser des économies.

Dans l'air du temps, la *smart city* est celle où les technologies numériques connectées à internet occupent une place importante et permettent d'améliorer les infrastructures et les services. A l'aune de l'ère numérique, l'objectif est de repenser les politiques publiques et d'optimiser le fonctionnement des villes. En France, c'est Nice qui est en passe de devenir un cas d'école, en se rangeant à la quatrième place des villes les plus connectées au monde derrière Barcelone, New York et Londres. On y trouve par exemple installés plusieurs centaines de capteurs : en haut des lampadaires, dans la chaussée ou encore sur des containers à ordures. Grâce à la collecte de données en temps réel, les lampadaires détectent les passants et adaptent leur luminosité en conséquence. Dans la même logique, les bennes à ordures ne sont vidées que si elles sont pleines grâce à un détecteur de remplissage. Ces nouvelles pratiques permettent de réaliser un gain de productivité et d'argent : une aubaine pour les villes qui cherchent s'approprier une identité verte tout en faisant des économies.

Toutefois le concept de smart city n'est pas l'apanage des métropoles « occidentales » : les pays en voie de développement se l'approprient aussi. Sur le continent africain, l'urbanisation exponentielle couplée aux prévisions démographiques ouvre la voie à la mise au point de solutions pour désengorger les villes paralysées par les embouteillages. Un des moyens est le réaménagement. A Abidjan, la baie de Cocody aujourd'hui polluée et dégradée fait l'objet d'un vaste projet : installation d'une zone pédestre autour de la berge lagunaire et développement d'une « *smart city* » au cœur de la capitale ivoirienne. A terme, elle deviendra un lieu de détente et de tourisme écologique.

En effet, la ville connectée est aussi une ville durable. Les grands groupes s'y engouffrent. La R&D d'EDF travaille depuis 2013 en partenariat avec Singapour où elle a créé en 2014 l'« EDF Lab Singapour » : un centre dédié à la recherche énergétique et à la planification urbaine, en coopération avec des universités locales. Un travail important d'aide à la décision pour la planification urbaine a été réalisé à partir de recueils des data – ouvertes et non ouvertes - permettant de modéliser des quartiers (bâtiments, air conditionné, gestion des transports, des déchets et de l'eau). Il permet de sélectionner les technologies les plus efficaces et de créer des quartiers durables et intelligents. A terme, ce type de centre pourrait s'étendre en Asie. Toujours au sein du Groupe EDF, Citelum, filiale spécialisée dans l'éclairage public et très en pointe dans le domaine de la numérisation du mobilier urbain, travaille à l'élaboration de solutions permettant une meilleure maîtrise des consommations d'énergie.

La ville connectée n'a cependant pas que des adeptes : si elle témoigne de formidables progrès dans l'efficacité des fonctions urbaines, elle soulève aussi des problèmes liés à la protection des données personnelles. « *On ne sait pas qui utilise les données en open data,*

reconnait Benoît Chabrier, délégué régional au numérique à la préfecture d'Occitanie. Par principe, les données étant ouvertes, on peut juste connaître le nombre de téléchargements de chaque jeu de données. Par exemple, les données budgétaires sont très largement consultées et téléchargées mais on ne sait pas pour quel usage, poursuit Bernadette Kessler, responsable du service innovation numérique de la métropole de Rennes. Ces données sont très peu exploitées, au sens propre, pour l'instant. Les associations et la population manquent encore cruellement de culture du numérique en général et de la donnée en particulier. »

Pour les détracteurs du « tout numérique », l'ubiquité de l'information et de la connexion présente des risques d'intrusion dans la vie privée et de surveillance accrue contre la volonté des citoyens.

Le risque est également de tomber dans un modèle de désurbanisation, où la ville serait phagocytée par les technologies numériques : c'est une des critiques qui a été formulée à l'égard de Songdo : là-bas la technologie semble avoir pris le pas sur la ville, et nuit aux qualités de dynamisme, spontanéité et d'adaptation propres au modèle depuis des siècles. Au demeurant, la ville connectée ouvre la perspective d'un retour aux origines de l'idéal démocratique au sens étymologique du terme - le pouvoir du peuple - car elle se veut davantage participative et transparente.

Les Etats-Unis sont pionniers dans l'open data à l'échelle locale.

Depuis quelques années, la Silicon Valley propose un nouveau paradigme : le transhumanisme. Ce mouvement promeut l'utilisation des découvertes scientifiques et techniques pour l'amélioration des performances humaines, à l'échelle d'une ville, d'un territoire ou d'un pays. Il explique que le rôle des responsables politiques doit être réduit à sa plus simple expression et que les algorithmes pourront faire bien mieux qu'eux, non seulement pour gérer techniquement les villes, mais pour mieux comprendre les talents des habitants. Dans son livre, « *Comment les géants du numérique veulent gouverner nos villes* », Jean Haëntjens écrit « *Les politologues observent un désir de participer qui est composé de différentes aspirations : refus des hiérarchies, envie de s'impliquer, prise de conscience des enjeux écologiques et sociaux ou simple désir de d'empowerment. En somme tout citoyen peut être à la fois consommateur de services et citoyens. Et l'intérêt, bien compris, de la cité politique sera de développer chez lui la deuxième dimension, en lui proposant une offre politique motivante. Cette offre repose sur plusieurs volets qui sont la transparence de l'information, les consultations électorales et la participation aux décisions et la possibilité de s'impliquer activement dans la vie de la cité. La transparence de la gestion publique a déjà été fortement améliorée depuis vingt ans avec, par exemple, la retransmission en ligne des conseils municipaux, l'accès aux marchés publics et aux budgets. L'étape suivante consistera à rendre les nombreuses données dont disposent les collectivités accessibles aux citoyens, aux associations ou aux start-ups proposant des services urbains. »*

De nombreux exemples peuvent être cités dans le domaine des transports : du partage des données de transports publics aux emplacements des bouchons en temps réel, les possibilités sont infinies. **La ville de Chicago au Etats-Unis a mis en place un service à destination des automobilistes.** Le nettoyage de la ville se fait à grande eau, contrairement

aux pratiques des services municipaux européens utilisant de petits véhicules. Le moindre véhicule garé dans ces rues à nettoyer empêche les équipes de maintenance d'intervenir. Ainsi, pour éviter aux automobilistes des amendes, la ville met à leur service une application qui les prévient de l'intervention des services municipaux. Cela permet d'accélérer le temps de traitement et de diminuer les bouchons résultants des mauvaises pratiques des automobilistes.

New York prône la démocratie participative en incitant à la création de services à forte valeur ajoutée. L'application Healthy Out qui permet d'identifier les restaurants le plus respectueux des normes d'hygiène. En alliant l'open data et le « crowd sourcing » (production participative), un ingénieur a lancé l'initiative Brand Your Subway, en sollicitant les « maîtres d'usage » afin qu'ils proposent une nouvelle planification du métro de NY via une carte interactive.

San Francisco, territoire des géants de l'Internet, a mis en place avec une équipe pluridisciplinaire une solution qui vise à analyser des données liées aux crimes et délits pour élaborer des modèles prédictifs grâce à l'intelligence artificielle par l'usage des algorithmes de Machine Learning. Ces prédictions permettent d'établir une carte des zones à fort taux de criminalité en précisant la géolocalisation et les types de crimes commis durant une période donnée en vue d'assurer la sécurité urbaine par l'adaptation des dispositifs aux nouvelles formes de menaces.

Au Canada, à Montréal, Toronto ou encore Ottawa, les projets développés en matière d'open data ont tous été conçus à partir d'un même constat : un problème concret à résoudre, les actions menées pour y répondre, leurs résultats, et les impacts concrets, parfois inattendus. Non pas parce que les données Canadiennes seraient intrinsèquement "meilleures", mais parce que les méthodes adoptées (par Synapse C ou le Laboratoire d'Innovation Urbaine de Montréal notamment) reposent sur l'usage. C'est tout le sens de la stratégie "publishing with a purpose" de l'Open Data Charter.

A Toronto, la smart city n'échappe pas à la polémique. Piloté par une filiale de Google, le projet Quayside se veut à la pointe de l'innovation urbaine, relève Les Echos. Malgré les nombreuses idées proposées dans ce cadre, les projets ont dû être révisés à la baisse pour calmer les inquiétudes des habitants, notamment sur la gestion des données. La ville de Toronto envisage par exemple la mise en place un service "d'open data à la demande", qui fournirait aux résidents la liste des données existantes afin qu'ils expriment eux-mêmes leurs besoins. En parallèle, Toronto réfléchit à comment prioriser l'ouverture de données.

L'open data reste en cela une composante essentielle de la ville intelligente, tout simplement car il s'agit du seul moyen d'en atteindre les objectifs. Une ville entièrement connectée, intelligible et gouvernable par la donnée ne peut techniquement exister sans que ses systèmes d'information soient rendus interopérables. Multiplier les licences, les plateformes et les modalités de partage favorise plutôt la création de "bulles" d'échange, soit des zones de circulation fluides en internes mais silotées les unes des autres. Les données sont partagées, mais restent inaccessibles pour un large panel d'acteurs, notamment ceux de la société civile. C'est cette architecture qui différencie une ville "partagée" d'une ville ouverte.

4.3. L'IA

L'intelligence artificielle (IA) se déploie à grande vitesse dans les entreprises et les administrations. Les données sont la matière première de l'intelligence artificielle. Ces systèmes sont étroitement associés aux usages de l'open data. Les initiatives se multiplient et ce n'est pas un hasard si les géants du Net américains (Google, Apple, Facebook, Amazon, Microsoft, IBM...) ainsi que leurs concurrents chinois (notamment les "BATX" Baidu, Alibaba, Tencent, Xiaomi) dominent de la tête et des épaules, y compris en matière de financement, la recherche mondiale dans l'intelligence artificielle. Comme le montre le rapport Villani, les pouvoirs publics ont identifié l'IA comme un relais de croissance potentiel pour notre économie. Les GAFAM se sont activement lancés dans une course aux données pour développer l'IA la plus avancée. Les données de référence pour le développement de l'IA sont à ce stade essentiellement voire uniquement américaines.

Amazon propose aux différents acteurs de l'open data ses infrastructures d'IA pour faciliter l'exploitation des données. Un développement de son business de plus en plus profitable et qui ne cesse de grossir dans la répartition de son CA.

Mark Zuckerberg/Facebook appelait récemment à une régulation d'État sur ces sujets. On comprend bien qu'il ne pourra bientôt plus fonctionner dans cette zone hors du droit. Dans toutes les grandes ruptures technologiques, le droit de propriété s'est peu à peu adapté : avec l'imprimerie est apparu le droit d'auteur, la révolution industrielle a créé les brevets, etc. Dans cette continuité, le droit de propriété pourrait s'étendre à la data. En témoigne l'intérêt toujours croissant des acteurs américains pour le secteur des legal techs, nouvel axe de croissance où se déploient des concentrés de technologie : IA, Big Data ou encore algorithmes poussés de recherche. Cette compétition entre titans de la Silicon Valley est bien partie pour se dérouler à l'insu des consommateurs et à leur détriment. Les systèmes d'intelligence artificielle évoluent en continu et peuvent changer leur comportement en fonction des données qu'ils manipulent. Une validation à un instant « t » d'un fonctionnement pourra ne plus être valide quelques jours ou mois après.

Mais derrière cette frénésie, beaucoup s'inquiètent et tirent la sonnette d'alarme comme le rapport sur l'intelligence artificielle (IA) rédigé par le mathématicien et député Cédric Villani rendu public le 28 mars 2018. Il prône une stratégie pour la France dans l'intelligence artificielle. Comme à l'époque de l'émergence d'Internet, la France dispose de pépites en la matière (doctrine.fr...). Mais cette prouesse technologique pourrait ne pas suffire : à l'image du numérique, dans les années 90, la France est en passe de manquer le virage de l'IA – et les entreprises les plus prometteuses, d'être rachetées par des acteurs étrangers.

4.4. Self data

Les données personnelles des individus, qu'ils soient abonnés, clients, consommateurs ou internautes, sont une denrée particulièrement prisée. Traitées, analysées, vendues, leurs utilisations par ceux qui les collectent sont souvent opaques et échappent au contrôle des citoyens qui s'inquiètent pour leur vie privée. Le chemin vers le Self data passe par les services et les usages.

Ces données "personnelles" sont aussi, très souvent, familiales et sociales. Le RGPD introduit en particulier le droit à la portabilité des données personnelles. Il prévoit que les personnes peuvent demander à récupérer leurs données, non pas seulement pour les vérifier ou les faire effacer, mais pour les réutiliser dans d'autres contextes, à leur initiative. Daniel Kaplan, l'un des fondateurs de la Fondation Internet Nouvelle Génération, appelle le Self Data "*qui désigne la production, l'exploitation et le partage de données personnelles par les individus, sous leur contrôle et à leurs propres fins : pour mieux se connaître, prendre de meilleures décisions, évaluer leurs décisions passées, se faciliter la vie, etc. En plus bref encore, le Self data désigne l'empowerment des gens à partir de leurs propres données. En France, le projet que nous avons mené avec la Fondation Internet Nouvelle Génération (FING) autour du concept de Self data s'appelle MesInfos* ».

Les particuliers prennent petit à petit conscience de l'importance de ces informations qu'ils livrent sans frais aux multinationales. Pour preuve, on estime que 30% des surfeurs français sur le web utilisent un bloqueur de publicité, pour ne pas être dérangés par les pop-up ouvrant à l'improviste, et le taux de clics moyen d'une campagne d'e-mailing atteint 4,2% selon les statistiques d'IBM.

Il est temps que cela change. Pour cela le RGPD est une première marche dans l'escalier de la propriété, mais une start-up Varanida va plus loin. « *Nous offrons à l'utilisateur de reprendre la main sur ses informations jusqu'à la fin de sa vie en lui offrant la possibilité de la monétiser suivant ses propres envies,* » explique le responsable de la start-up. Les intérêts ne sont pas seulement pour les consommateurs mais aussi pour les publicitaires : « *20% de leur budget part en fraude, cela représente 46 milliards d'euros annuellement* » rapporte le responsable de Varanida. La start-up ambitionne de devenir un «ad contrôleur», pour résumer les publicités sont bloquées, mais l'internaute peut décider de délivrer différents niveaux de données, tout en étant rémunéré pour ce geste. « *Prenez l'exemple d'une personne qui prépare ses futures vacances, elle peut décider de délivrer une partie de ses données à certains sites, afin de recevoir des publicités prédictives.* ». La communauté d'adblockers (bloqueur de publicité, ndlr) représente 100 millions de personnes dans le monde.



D'autres initiatives voient le jour. Tim Berners-Lee dévoilé son nouveau projet, baptisé Solid, «*open source visant à restaurer le pouvoir et l'influence des individus sur le Web.*» Ainsi l'internaute bénéficiera d'un coffre-fort pour stocker ses données personnelles des utilisateurs. «*Nous faisons évoluer le Web de façon à restaurer un équilibre, en donnant à chacun de nous le contrôle complet sur ses données, personnelles ou non, d'une manière révolutionnaire,*» déclarait Tim Berners-Lee. Les internautes ne veulent pas d'un monde sans publicité, mais d'un monde avec moins de publicité, moins d'intrusion, et un meilleur ciblage.

Au **Royaume-Uni** avec le programme gouvernemental MiData. Aux États-Unis, l'administration Obama étendait ses ambitieux programmes d'open data vers un « feedback » aux gens de leurs propres données. Cela a pris la forme de programmes sectoriels : "Blue Button" concerne les données de santé, "Green Button" celles de l'énergie. Il y avait d'autres projets dans l'éducation, par exemple, mais apparemment, ils sont arrêtés. Toujours aux États-Unis, le "Projet VRM" (pour Vendor Relationship Management, le symétrique du CRM ou Customer Relationship Management) explore des sujets proches depuis plusieurs années.

Les projets locaux se multiplient comme à Lyon ou encore Helsinki autour de thématiques variées (données de mobilité en Europe du Nord, santé) et portés par des entreprises installées ou des startups : Cozy, Digi.me, Fair & Smart... La FING, think Tank de référence pour anticiper les transformations numériques, contribue depuis 2017 à la mise en place d'une organisation internationale correspondant à ces logiques de Self Data : Mydata.org organise une grande conférence internationale chaque année et dispose d'une vingtaine de "hubs" locaux de tailles très diverses.

Le mouvement vers le "Self Data" est désormais engagé et la monétisation des données paraît inéluctable.

5. Conclusion

Aujourd'hui, en France, force est de constater que l'engouement a surtout été médiatique mais que faute de volonté (crainte pour les entreprises de partager de la donnée qui pourrait être utilisée par des concurrents), faute de moyens (cf. supra) notamment du côté des collectivités qui sont en retard dans l'appréhension et le déploiement de ces technologies, faute d'une volonté politique claire, l'open data n'a pas (encore) permis la révolution politique escomptée. Les collectivités territoriales devraient être accompagnées techniquement (modernisation des services) et « socialement » (conduite de changement), dans la conduite de tels projets. L'accent doit être mis sur la formation de l'ensemble des acteurs.

Au-delà de ces freins identifiés, une volonté politique nationale qui permettrait d'intégrer la démarche d'open data, implique des investissements massifs dans la recherche et développement. Chacune des initiatives locales, ne permet pas la mutualisation, et l'exploitation performante de ces données.

Un des enjeux premiers seraient de mettre le monde de l'entreprise devant ses responsabilités sur l'open data et plus largement sur l'économie numérique et d'opérer une nécessaire convergence entre les intérêts des politiques, des représentants de

l'administration et les acteurs privés sur la manière de concevoir une politique d'accroissement de puissance dans l'économie numérique, le débat se figera sur les plus petits dénominateurs communs d'une politique de souveraineté.

Certaines entreprises européennes ont compris le potentiel de l'open data, à condition de récupérer et traiter un grand volume de données. La riposte européenne aux GAFAM et BATX est peut-être en train de s'organiser.

6. Bibliographie

Entretiens et rencontres

- Jean-Marc Lazare, CEO OpenDataSoft
- Guillaume Sylvestre, lobbyiste
- Mathilde Biquet, directrice générale FEDMA
- Nicolas Ravailhe, lobbyiste
- Georges Hebrail, directeur data science et interaction EDF R&D

Livres et ouvrages

- « *Comment les géants du numérique veulent gouverner nos villes. La cité face aux algorithmes* », Jean Haëntjens
- « *La géopolitique pour comprendre le cyber espace* », Frédéric Douzet
- « *Open data et business models* », Etienne Roché
- « *Géopolitique des données numériques* », Amaël Cattaruzza
- « *Enjeux et méthodes de la gestion des données (Management des systèmes d'information)* » Franck Régnier-Pécastaing , Michel Gabassi
- « *Big Data, Smart Data, Stupid Data... : Comment (vraiment) valoriser vos données* », Antoine Denoix
- « *La boîte à outils de la Stratégie big data* », Romain Rissoan et Romain Jouin
- « *Big Data et Machine Learning - 3e éd. - Les concepts et les outils de la data science* » de Pirmin Lemberger, Marc Batty

E-learning

- <https://www.europeandataportal.eu/elearning/fr/#/id/co-01>
- <https://www.europeandataportal.eu/elearning/fr/module1/#/id/co-01>

Articles

- <https://www.etalab.gouv.fr/lancement-dun-nouvel-appel-a-projets-opendata-dans-le-cadre-du-programme-dinvestissements-davenir-pia/>
- https://empruntnational.achatpublic.com/sdm/ent/gen/ent_detail.do?selected=0&PCSLID=CSL_2017_EVlcl5PqHD/
- <https://www.village-justice.com/articles/cour-des-comptes-interesse-open-data-milite-pour-une-nouvelle-phase-maturite,30994.html>
- <https://www.lagazettedescommunes.com/184790/open-data-quels-couts-pour-les-collectivites-territoriales/>
- <https://www.lebigdata.fr/open-data-definition>
- <https://www.forbes.com/sites/forbestechcouncil/2019/08/22/what-is-self-service-data-science/#3dc1c5df2d63>
- <https://www.ladn.eu/tech-a-suivre/fabrique-des-futurs/protéger-donnees-payer-avec/>

- <https://www.data.gouv.fr/fr/datasets/catalogue-des-jeux-de-donnees-publies-en-opendata/>
- <https://www.data.gouv.fr/fr/datasets/liste-des-producteurs-des-donnees-geographiques-1/>
- <https://www.etalab.gouv.fr/demain-tous-producteurs-de-donnees-les-sciences-participatives-nous-montrent-la-voie-data-literacy-conference-2018>
- <https://www.forbes.fr/technologie/opendata-nouvelle-arme-puissante-au-profit-de-villes-performantes/?cn-reloaded=1>
- <https://www.nextinpact.com/news/107708-le-premier-ministre-veut-bilan-lopen-data-et-fin-redevances-a-horizon-2022.html>
- <https://andreadata.frama.site/data/dans-les-coulisses-du-portail-data.gouv.fr>
- <https://www.culture.gouv.fr/Sites-thematiques/Innovation-numerique/Donnees-publiques/Guide-pratique-Open-data>
- <https://medium.com/dataactivist/des-analytics-%C3%A0-l'impact-pourquoi-%C3%A9valuer-une-politique-open-data>
- <https://www.oreilly.com/ideas/open-data-business-models-deloitte-insight>
- <https://infoguerre.fr/2019/05/19602/>
- <https://infoguerre.fr/2019/04/guerre-economique-question-essentielle-encerclements-cognitifs/>
- <https://infoguerre.fr/2019/04/axes-de-recherche-actuels-de-lecole-de-guerre-economique/>
- <https://infoguerre.fr/2019/03/notion-de-marche-prive-de-linformation/>
- <https://infoguerre.fr/2018/09/fil-conducteur-de-letude-de-problematique-de-puissance-francaise/>
- <https://infoguerre.fr/2018/08/recherche-de-liberte-strategique/>
- <https://infoguerre.fr/2018/07/limperieuse-necessite-occidentale-delaborer-de-nouvelles-grilles-de-lecture/>
- <https://infoguerre.fr/2018/04/contradictions-francaises/>
- https://www.idcite.com/Open-data-la-CNIL-et-la-CADA-publient-un-guide-pratique-de-la-publication-en-ligne-et-de-la-reutilisation-des-donnees_a44527.html

7. Glossaire

Data : une donnée est la représentation d'une information dans un programme : soit dans le texte du programme (code source), soit en mémoire durant l'exécution. Les données, souvent codées, décrivent les éléments du logiciel tels qu'une entité (chose), une interaction, une transaction, un évènement, un sous-système, etc.

Big data : « grosses données » en anglais, les mégadonnées ou les données massives, désigne des ensembles de données devenus si volumineux qu'ils dépassent l'intuition et les capacités humaines d'analyse et même celles des outils informatiques classiques de gestion de base de données ou de l'information.

Code source : texte qui représente les instructions d'un programme telles qu'elles ont été écrites dans un langage de programmation sous une forme humainement lisible par un programmeur. Le code source se matérialise souvent sous la forme d'un ensemble de fichiers textes.

Backbone : Il constitue le centre névralgique d'un réseau à haut débit. Littéralement épine dorsale. Dans le contexte des réseaux de télécommunications désigne la partie qui supporte le gros du trafic, en utilisant les technologies les plus rapides et une grande bande passante sur des distances importantes. Les petits réseaux (internes à une entreprise ou à une région) se rattachent à ce réseau fédérateur, comme les rivières viennent grossir le cours d'un fleuve.

Data broker : courtier de données, qui désigne une entreprise dont l'activité est basée sur la revente de données à des annonceurs ou à des prestataires marketing.

Hackathon : événement au cours duquel des spécialistes se réunissent durant plusieurs jours autour d'un projet collaboratif de programmation informatique ou de création numérique.

Intelligence artificielle (IA) : l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence. Elle correspond donc à un ensemble de concepts et de technologies plus qu'à une discipline autonome constituée

Interopérabilité- Capacité de mélanger différents ensembles de données.

Open data : Définition normalisée. Le terme open data désigne des données auxquelles n'importe qui peut accéder, que tout le monde peut utiliser ou partager. Les critères essentiels de l'open data sont la disponibilité, la réutilisation et la distribution, et la participation universelle. Il s'agit là de la définition donnée par l'Open Knowledge Foundation en 2005.

Open Innovation : Innovation ouverte ou parfois aussi Innovation distribuée désignent dans les domaines de la recherche et du développement des modes d'innovation fondés sur le partage, la collaboration (entre parties prenantes).

Open source : « code source ouvert », s'applique aux logiciels (et s'étend maintenant aux œuvres de l'esprit) dont la licence respecte des critères précisément établis par l'Open Source Initiative, c'est-à-dire les possibilités de libre redistribution, d'accès au code source et de création de travaux dérivés. Mis à la disposition du grand public, ce code source est généralement le résultat d'une collaboration entre programmeurs.

RGPD : Le règlement no 2016/679, dit règlement général sur la protection des données (RGPD, ou encore GDPR, de l'anglais General Data Protection Regulation), est un règlement de l'Union européenne qui constitue le texte de référence en matière de protection des données à caractère personnel¹. Il renforce et unifie la protection des données pour les individus au sein de l'Union européenne.

Self Data : la production, l'exploitation et le partage de données personnelles par les individus, sous leur contrôle et à leurs propres fins.

Smart City - « ville intelligente », désigne une ville utilisant les technologies de l'information et de la communication (TIC) pour « améliorer » la qualité des services urbains ou encore réduire ses coûts. Une ville intelligente est une zone urbaine qui utilise différents capteurs de collecte de données électroniques pour fournir des informations permettant de gérer efficacement les ressources et les actifs. Cela comprend les données collectées auprès des citoyens, des dispositifs mécaniques, des actifs, traitées et analysées pour surveiller et gérer les systèmes de circulation et de transport, les centrales électriques, les réseaux d'approvisionnement en eau, la gestion des déchets, les systèmes d'information, les écoles, les bibliothèques et les hôpitaux.

Le gouvernement ouvert ou "Open Gov" : La transparence, l'intégrité, le devoir pour l'Etat de rendre des comptes et la participation des citoyens aux décisions qui les concernent sont des piliers de la démocratie. La Déclaration des Droits de l'Homme et du Citoyen de 1789 proclamait déjà que les citoyens doivent pouvoir concourir à l'élaboration de la loi, constater par eux-mêmes l'efficacité de la dépense publique, demander compte à tout agent public de son administration. Depuis, la France a progressivement inscrit ces principes dans sa loi et dans ses institutions : Service public des archives, Service de la statistique publique, loi de 1978 sur l'accès aux documents administratifs, Service public de la diffusion du droit, Commission nationale du débat public, Conseil économique social et environnemental, loi sur la transparence de la vie publique en 2013... La diffusion massive d'internet permet aujourd'hui à des milliards de citoyens du monde entier de s'informer et de s'engager. Elle ouvre sur de nouvelles pratiques sociales, de nouvelles aspirations et de nouveaux pouvoirs d'agir pour la société civile. Le numérique est ainsi une opportunité pour renouveler, enrichir, étendre la participation et la co-construction entre l'Etat et les citoyens. Comme l'a réaffirmé le président de la République lors de son discours à New York, la France poursuit cette ambition d'une action publique plus ouverte, plus transparente et plus efficace.

Données personnelles : elles correspondent à toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres (article 2 de

la loi informatique et liberté). Elles sont protégées par divers instruments juridiques concernant le droit à la vie privée, dont notamment la loi Informatique, fichiers et libertés de 1978, la directive 95/46/CE au niveau communautaire ainsi que la Convention n°108 pour la protection des données personnelles du Conseil de l'Europe. À l'instar de la Cnil française, beaucoup de pays disposent aujourd'hui d'autorités chargées de la protection des données personnelles, qui sont souvent des autorités administratives indépendantes.

Données publiques : cette notion couvre l'ensemble des données qui sont ou devraient être (légalement ou volontairement) publiées ou tenues à disposition du public, et qui sont produites ou collectées par un État, une collectivité territoriale, un organe parapublic, dans le cadre de leurs activités de service public. L'accessibilité de la donnée publique (qui implique aussi la liberté d'accès aux documents administratifs) est un des éléments de la transparence d'une gouvernance, considérée par Guy Braibant comme faisant partie de la "troisième génération des droits de l'Homme.

Data puking : il se définit comme une profusion de données. C'est le fait de disposer d'une multitude de data, de ne pas savoir les hiérarchiser, de ne pas en extraire la valeur et de vouloir absolument toutes les utiliser. Cela aboutit finalement à perdre du temps, perdre de vue les objectifs et ne donner aucun résultat concret.

Data pool : c'est un espace de stockage de l'information, davantage structurée que le data lake. Les données sont organisées en Datamart (sous-ensembles de data warehouse). Ainsi, l'accès à tous les utilisateurs, même sans expertise data, est facilité.

Chaîne de valeur data : elle désigne les données qui contribuent le plus à apporter de la valeur ajoutée à une entreprise. Il s'agit en clair des données qui assurent un avantage concurrentiel et qui présentent les meilleurs avantages en termes de coûts, de rendement et d'image. Définir la chaîne de valeur data permet de se centrer sur ces données pour les pérenniser.

Data utile (vs stupid data) : elle désigne la donnée qui a du sens et de l'impact pour votre activité. Ce sont les données qui sont indispensables à la chaîne de valeur de votre entreprise. Ce sont sur ces data que vous devez vous concentrer pour garantir une qualité et une fiabilité maximale de votre chaîne de valeur. Au contraire, la "stupid data" désigne les données inutiles sur lesquelles certaines entreprises dépensent du temps et de l'énergie afin de les utiliser, sans que cela ne donne de résultat efficace.

Self-service data (ou data discovery) : La data discovery permet à un utilisateur métier d'extraire et d'interpréter des données pouvant provenir de différentes sources. Cela permet de trouver, de visualiser et de raconter des faits autour des données, sans qu'aucun algorithme n'ait été paramétré au préalable.

Data for & from IoT : ce sont toutes les données générées pour ou par l'Internet des Objets.

Algorithme : c'est une formule mathématique ou un traitement statistique utilisé pour effectuer une analyse de données. Les algorithmes sont utilisés dans de nombreux autres domaines, mais ont une importance capitale dans le secteur du Big Data.

Analyse prédictive : il s'agit de prévoir les probabilités de ce qui peut survenir. Les data scientists utilisent des techniques avancées comme le Machine Learning et les traitements statistiques avancés pour faire des prédictions importantes comme les changements économiques ou les prévisions météorologiques.

Analyse prescriptive : elle permet de prédire les décisions à prendre pour un impact maximal. Par exemple, toujours dans l'exemple du relevé de compte, il peut s'agir de déterminer quels sont les postes de dépense à cibler pour réduire efficacement les dépenses. Les analyses prescriptives reprennent le principe des analyses prédictives en y ajoutant la notion d'action. L'analyse des résultats permet de déterminer des cibles sur lesquelles agir. Dans le domaine du Big Data, les entreprises prennent des décisions en observant l'impact potentiel des actions qu'ils peuvent effectuer.

Cloud : omniprésent depuis quelques années, il est devenu inutile de présenter cette technologie. Concrètement, il s'agit de logiciels ou de données hébergées ou lancées sur des serveurs distants, accessibles depuis n'importe où sur internet.

Data Lake : un Data Lake est un répertoire où sont stockées de nombreuses données d'entreprises au format brut. Le Data Lake se différencie des Data Warehouses, qui servent à stocker les données structurées après qu'elles aient été nettoyées et intégrées avec d'autres sources. Un Data Lake permet de faciliter l'accès aux données.

Data Mining : il permet de trouver des patterns et d'extraire des informations pertinentes en provenance de larges ensembles de données en utilisant des techniques de reconnaissance de patterns sophistiquées. Ce terme est étroitement lié à la notion d'analyse de données. Pour extraire des patterns, les data miners utilisent des statistiques, des algorithmes de machine Learning, et l'intelligence artificielle.

Data Scientist : c'est la personne qui se charge de donner un sens au Big Data en extrayant les données brutes du Data Lake, en les traitant, et en en tirant des insights. Parmi les compétences requises pour le data scientist, on compte les compétences en analyse, en statistiques, en science informatique, de la créativité, un talent pour la narration et une compréhension du contexte de l'entreprise. Il s'agit donc d'un métier exigeant, ce qui explique les salaires très élevés des data scientists.

Machine Learning : c'est une technologie permettant aux systèmes informatiques d'apprendre, de s'ajuster et de s'améliorer grâce aux données. Ces machines embarquent des algorithmes de prédictifs et statistiques afin d'apprendre et d'adopter le comportement correct. Plus les ordinateurs reçoivent de données, plus ils s'améliorent.

8. Echiquiers

