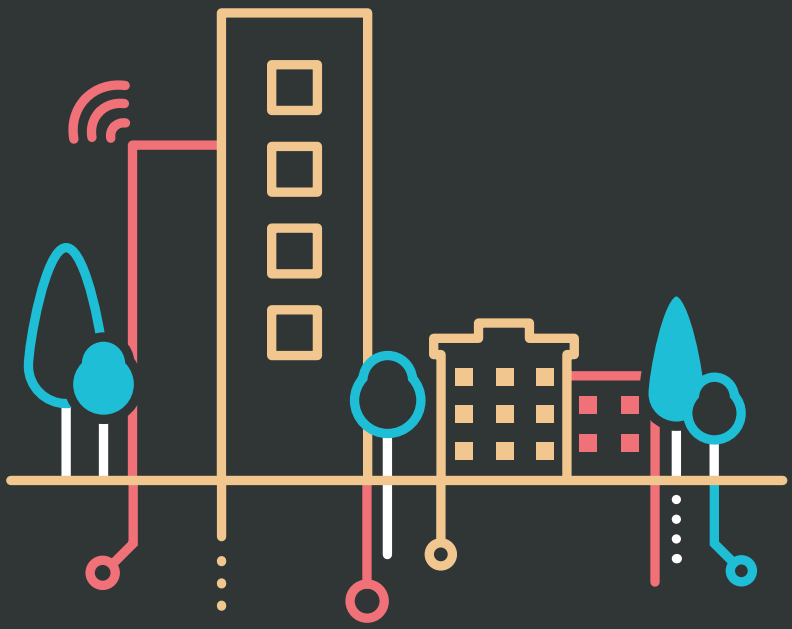


TOUS ACTEURS DES DONNÉES

Appréhender les données pour mieux les valoriser

RENAISSANCE numerique

syntec numérique

TOUS ACTEURS DES DONNÉES

Appréhender les données pour mieux les valoriser

RENAISSANCE numerique

syntec numérique

DÉPASSER LE MYTHE DES DONNÉES

Les données suscitent de nombreuses interrogations, et parfois même des inquiétudes. Elles sont une richesse singulière qui rebat les cartes de notre économie et de notre société, et redistribue les jeux de pouvoirs au sein de nos organisations et entre pays. Il est urgent que chacun puisse s'en saisir pleinement, sache en extraire la valeur, que celle-ci soit économique ou démocratique, puisse en assurer la protection. Pas seulement les *data scientists*, pas seulement les acteurs privés, mais aussi les décideurs publics, les citoyens, la société civile dans son ensemble.

Les données sont souvent présentées comme le nouvel « or noir » du XXIe siècle. Si cette expression illustre bien la profusion de cette nouvelle source de richesse, elle est cependant trompeuse. Les données ont des caractéristiques bien différentes des ressources naturelles physiques : ce sont des biens inépuisables (nous en produisons chaque jour davantage) et leur utilisation ne conduit pas à leur disparition (ce sont des biens non rivaux). Cette interprétation ne permet pas d'appréhender ce qui fait leur valeur : leur circulation, leur agrégation, leur traitement. Les données font ainsi l'objet de nombreuses confusions. Or, si nous souhaitons nous en saisir, il faut en avoir une analyse fine. Il convient mieux d'utiliser le terme dans son acception plurielle : les données sont multiples, et parfois très différentes selon les secteurs et les acteurs. La résurgence de vieux débats participe également de ces confusions, à l'instar du débat sur la patrimonialisation des données personnelles, pour lequel notre droit a arbitré depuis déjà plusieurs décennies.

Cependant, des débats sérieux restent à trancher, si nous ne souhaitons pas être dessaisis des opportunités induites par cette nouvelle économie des données. En Europe, nous avons renforcé la protection des données personnelles avec l'entrée en vigueur du règlement général sur la protection des données (RGPD). En harmonisant la vision entre États membres de l'Union européenne (UE) sur la protection des données, cette nouvelle réglementation a non seulement consolidé les droits des individus, mais elle a également permis de sensibiliser un certain nombre d'acteurs jusqu'alors éloignés de la culture des données. Mais avons-nous pris en compte toutes les dimensions économiques de cette réglementation ? Nous offre-t-elle pleinement la capacité à produire de la valeur, à favoriser l'innovation ?

Au Royaume-Uni, l'ICO (*Information Commissioner's Office*), l'équivalent de la Commission nationale de l'informatique et des libertés (CNIL) en charge du suivi de cette réglementation en France, a mis en place un « bac à sable réglementaire ». Ce dernier permet aux entreprises innovantes de bénéficier d'un allègement des obligations relatives à la protection des données personnelles, afin de tester leurs produits ou services. Les entreprises peuvent ainsi se faire accompagner pour être conformes aux règles de protection des données tout en innovant, et sans risquer de sanctions. Un tel dispositif est positif tant pour les entreprises que pour les citoyens, car l'innovation n'irait pas à l'encontre de leur protection. Encore faut-il que nous donnions les moyens (humains et financiers) à la CNIL pour développer de telles initiatives.

La circulation des données est un enjeu majeur des débats actuels autour des nouvelles régulations du numérique. Ces débats sont indispensables afin de mieux encadrer et d'encourager ce partage. Nous devons toutefois être vigilants à ce que ces nouvelles régulations ne contreviennent à d'autres réglementations, à peine mises en œuvre. En cela, le débat sur le futur règlement européen concernant le respect de la vie privée et la protection des données à caractère personnel dans les communications électroniques, dit règlement *ePrivacy*, qui devrait être adopté dans la prochaine mandature européenne, peut donner le tournis. Malgré la volonté affichée de la Commission européenne d'assurer une bonne articulation entre le projet de règlement *ePrivacy* et le RGPD, un certain nombre de ses dispositions sont susceptibles de créer des contradictions, sources d'incertitudes juridiques.

La valeur des données est souvent abordée dans sa seule dimension économique. Or, elle revêt également une dimension sociale et politique forte. La mise en données de nos activités et de nos institutions nous invite à repenser leur gouvernance, à tous les étages de notre société. Organisations ou individus, des acteurs participent aujourd'hui activement à cette transformation, par le partage volontaire de leurs données. Cela pose la question de l'évolution de leur rôle dans la prise de décision. Les citoyens peuvent également participer directement de la régulation de certains secteurs par la remontée d'information, à l'instar des cartes participatives mises en place par le régulateur des télécoms, l'Autorité de régulation des communications électroniques et des postes (ARCEP). De l'information du consommateur à la redevabilité des élus et de l'administration, le mouvement d'ouverture des données donne également

une nouvelle dimension au principe de transparence. Il permet de faire éclore de nouveaux services et de consolider la relation avec les clients ou usagers par une meilleure information. Des communautés d'utilisateurs se créent également de manière ad hoc afin d'améliorer cette information, comme la base de données sur les produits alimentaires *Open Food Facts*. Il s'agit désormais de ne pas aborder cette transparence comme une nouvelle contrainte.

Les données sont aussi devenues un objet de géopolitique. Elles sont au cœur des nouvelles relations entre États et régions du monde. L'Union européenne s'est engagée tardivement dans la révolution des données et peine aujourd'hui à se bâtir une doctrine commune au-delà de la protection des données personnelles. Son approche éthique, si elle est louable, mérite néanmoins d'être plus précise et mieux définie. La mandature européenne qui s'ouvre aura à avancer sur un certain nombre de ces aspects : encadrement de l'intelligence artificielle, définition des données d'intérêt général, cybersécurité, échange de flux de données avec les pays hors UE, etc.

Les données constituent un formidable levier de progrès. Mieux nous les appréhendons, plus nous serons en capacité de les valoriser.

Cet ouvrage, fruit d'un travail commun entre Renaissance Numérique et Syntec Numérique, souhaite y contribuer en fournissant à chacun la possibilité d'une compréhension des enjeux liés à la mise en données du monde.

Godefroy de Bentzmann, président de Syntec Numérique

Henri Isaac, président de Renaissance Numérique

SOMMAIRE

TOUS ACTEURS DES DONNÉES

PRÉFACE - DÉPASSER LE MYTHE DES DONNÉES 3

CHAPITRE 1

LES DONNÉES, CETTE RICHESSE MAL APPRÉHENDÉE 9

1_ Les données ne sont pas de l'information 11

2_ Les données ne sont pas le nouvel « or noir » 14

3_ Des données à la création de valeur 23

CHAPITRE 2

LA DONNÉE AU CŒUR DE LA TRANSFORMATION DES ORGANISATIONS 35

1_ Un actif stratégique pour l'organisation 37

2_ Les nouveaux écosystèmes de données 49

CHAPITRE 3

DE NOUVEAUX ENJEUX DE GOUVERNANCE FACE À LA MISE EN DONNÉES DE LA SOCIÉTÉ 75

1_ Les données, outil d'une démocratie renouvelée 76

2_ Relever les défis de cette nouvelle économie 86

BIBLIOGRAPHIE 93

REMERCIEMENTS 97

CHAPITRE 1

LES DONNÉES,
CETTE RICHESSE MAL APPRÉHENDÉE

Aujourd'hui, les innovations liées au numérique ne sont plus cantonnées au seul champ des technologies de la communication (TIC). Elles se déploient dans la robotique, l'intelligence artificielle ou encore les objets connectés, couvrant un spectre toujours plus large de l'activité humaine. Cette place prépondérante prise par les technologies numériques dans notre vie quotidienne a donné lieu à une prolifération inouïe des données numériques. Avec elles, une nouvelle économie a vu le jour, qui a fait de l'exploitation de la donnée, cette nouvelle « matière première », la source de création de valeur. Les activités de génération, de collecte et d'exploitation des données numériques figurent désormais au cœur de la production de services dans de nombreux secteurs jusqu'à la sphère publique. La presse, l'industrie culturelle, le marketing, la finance, l'automobile, l'hôtellerie, le transport de personnes, la santé, l'énergie, l'agriculture ont ainsi vu leur économie profondément modifiée, sinon bouleversée, par l'utilisation massive des données numériques. Ces transformations ont été initiées et portées par de nouveaux acteurs économiques qui ont réussi à acquérir en quelques années un pouvoir de marché considérable. Dans le même temps, les positions les mieux établies ont été remises en cause. Si de tels constats sont largement partagés, ils s'accompagnent rarement d'une interrogation sur la nature des données et sur le processus de création de leur valeur.

1_ Les données ne sont pas de l'information

Des données à la connaissance

Les termes « données », « informations » et « connaissances » sont souvent utilisés l'un pour l'autre. Pourtant, bien que imbriqués, ils recouvrent chacun des réalités différentes. Recourant à une métaphore restée célèbre, le théoricien des organisations Russel Ackoff expliquait : « Un gramme d'informations équivaut à un kilogramme de données. Un gramme de connaissances équivaut à un kilogramme d'informations » [1].

Au-delà de cette comparaison, la donnée peut se définir comme une description élémentaire d'une réalité : le résultat d'une observation (le nombre de voitures sur une route, le décompte du nombre d'oiseaux dans un champ) ou d'une mesure (la température d'une pièce, la durée d'un appel téléphonique) [2]. Le passage d'un ensemble de données à la production d'une information est le fruit d'un traitement de ces données primaires, généralement une agrégation, destiné à les rendre utilisables. Tout comme les données, les informations sont des descriptions d'une réalité, mais d'un segment de réalité plus ample et sous une mise en forme plus facilement exploitable (le nombre total de voitures qui passe sur une route chaque jour, le suivi annuel de la présence des oiseaux dans un département, les variations de température dans un bâtiment). Enfin, la connaissance résulte des réflexions menées à partir d'une accumulation d'informations et de leur comparaison. Cette agrégation de savoirs permet d'interpréter des phénomènes et elle sert de guide à la prise de décision en permettant de gagner en efficacité par rapport à un objectif déterminé (améliorer la circulation routière, lutter contre la disparition des oiseaux, faire des économies d'énergie).

Des données brutes aux données traitées

Bien qu'elle soit fréquemment utilisée, l'expression « données brutes » est trompeuse. La donnée est toujours produite dans le cadre de processus spécifiques, en vue de son traitement ultérieur. Elle est ensuite contextualisée, agrégée, croisée à d'autres données, etc.

[1] Ackoff Russell, "From Data to Wisdom", in *Ackoff's Best*, John Wiley & Sons, New York, 1989, p. 170.

[2] Isaac Henri, « La donnée numérique, bien public ou instrument de profit », *Pouvoirs*, n° 164, 2018, p. 75-86.

Le *Big Data*, une abondance de données

L'exploitation des données n'est pas un phénomène nouveau, l'apanage d'une « nouvelle économie » liée au numérique. Les techniques marketing les plus traditionnelles reposent en effet déjà sur la collecte, le traitement et l'analyse systématique de données, que ce soit pour connaître les préférences des clients, pour définir une campagne de communication ou pour mesurer les performances d'une entreprise. La révolution portée par le développement des technologies numériques tient davantage à l'apparition du « *Big Data* » ou « données massives » qu'à l'utilisation des données elles-mêmes. Le *Big Data* désigne à la fois l'augmentation exponentielle du volume de données produites et collectées chaque jour et les progrès de la capacité à les traiter [3].

L'augmentation du volume de données tient à deux facteurs : la multiplication des modes de collecte et la diversification des types de données collectées. Par ailleurs, les données ne se limitent plus seulement à des caractéristiques administratives telles que la date de naissance, l'adresse, la catégorie socio-professionnelle, l'âge ou le sexe. « Aujourd'hui, nos données, c'est-à-dire les traces que nous laissons dans l'environnement numérique, disent presque tout de notre intimité », selon Laurent Cytermann, maître des requêtes au Conseil d'État et co-rédacteur de l'étude annuelle du Conseil d'État sur le numérique et les droits fondamentaux [4]. La diversité des données ainsi agglomérées renseigne sur nos goûts, nos centres d'intérêt, nos déplacements, nos relations, nos paramètres biologiques et même notre humeur.

La capacité de traitement des données a augmenté de manière tout aussi spectaculaire que le volume de données disponibles. Cette évolution a été rendue possible grâce à la conjonction de trois facteurs : l'accroissement de la puissance informatique, l'augmentation des capacités de stockage de données et la rapidité des canaux de communication numérique. Conformément à la prédiction formulée en 1965 par Gordon Moore, et connue sous le nom de « loi de Moore », selon laquelle le nombre de transistors sur une puce de silicium de prix constant double tous les deux ans, les ordinateurs ont connu une croissance exponentielle de leur puissance de calcul, tout en devenant toujours moins coûteux. Dans le même temps, on estime que le prix du gigaoctet de stockage est passé de

[3] Cytermann Laurent, « Promesses et risques du big data et de l'open data », *Informations sociales*, 2015/5, n° 191, p. 82.

[4] Cytermann Laurent, « Promesses et risques du big data et de l'open data », *op. cit.*, p. 85.

40 000 dollars en 1987 à 0,05 dollar en 2015 ! S'ajoutent à cela les importants progrès réalisés en matière de connectivité. Avec le déploiement de la 4G et prochainement de la 5G, les données qui étaient autrefois uniquement accessibles via des connexions physiques le sont désormais à distance. De l'Internet des objets (IoT) aux véhicules autonomes, la connectivité sans fil offre de nouvelles possibilités d'exploitation des données.

L'augmentation du volume de données, associée aux nouvelles capacités de calcul et de stockage, a donné lieu au développement de nouvelles technologies, à l'instar des technologies d'intelligence artificielle (IA)[5] avec le *machine learning* et le *deep learning*. L'intelligence artificielle, nourrie par le jeu massif de données, permet aujourd'hui de régler en quelques minutes des problèmes qui demandaient autrefois des années, qu'il s'agisse de reconnaître un visage ou de créer des tableaux et des mélodies originales.

AlphaGo, ou l'entrée de l'intelligence artificielle dans le classement des meilleurs joueurs au monde

En 2016, AlphaGo, le programme de Deepmind, une filiale de Google, l'emportait au jeu de go contre le Coréen Lee Sedol, considéré comme l'un des meilleurs joueurs du monde. En 2017, un nouvel algorithme de Deepmind, AlphaGo Zero, battait le précédent programme, mais en ayant cette fois-ci appris à jouer tout seul, sans apport de données humaines, affinant sa stratégie à partir de millions de parties jouées contre lui-même. En janvier 2019, Deepmind a annoncé que son programme AlphaStar avait réussi à battre des joueurs professionnels au jeu Starcraft II, un jeu vidéo de stratégie qui impose aux joueurs de gérer plusieurs opérations en temps réel : extraction de ressources, construction de bâtiments, déplacements de dizaines, voire de centaines d'unités, exploration de la zone de jeu[6].

[5] L'intelligence artificielle est un terme générique qui regroupe l'ensemble des technologies informatiques qui permettent d'imiter des comportements intelligents, que ce soit en termes de perception, en raisonnement ou en génération d'informations. Ces technologies s'appliquent dans un grand nombre de domaines - santé, transports, énergie, e-commerce, banque, industrie, etc. -, au travers d'applications telles que le calcul d'itinéraire GPS, la reconnaissance d'image, l'interaction avec l'utilisateur dans les assistants personnels, l'imagerie médicale, la robotique, ou encore la recommandation de contenus basée sur les préférences de l'utilisateur (écoute de musique, films, achats, etc.).

[6] https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/

2_ Les données ne sont pas le nouvel « or noir »

Fort des possibilités offertes par le *Big Data*, les données sont souvent présentées comme un nouvel « or noir », le carburant de l'économie numérique. La donnée serait devenue le pétrole du XXIe siècle : à la fois une matière première prisée sur les marchés, une source d'enrichissement abondante pour celui qui la détient et le vecteur d'une révolution industrielle dont les *data centers* (centres de données) constitueraient, après les raffineries au XXe siècle, les nouvelles cathédrales. Le succès incontestable de certaines entreprises américaines et chinoises qui ont fait du traitement et de la valorisation des données leur cœur de métier serait la preuve irréfutable de l'existence de ce nouvel eldorado pétrolier. Ainsi, les bénéfices de Google ont pratiquement doublé entre 2012 et 2016, passant de 10,6 milliards à 19,5 milliards de dollars. Sur une période encore plus brève, le chiffre d'affaires de Facebook a été multiplié par près de deux cents entre 2012 et 2016. Et au classement 2018 des entreprises les plus valorisées en bourse, sept des dix premières appartiennent au secteur numérique. À l'exception d'Apple et de Microsoft, dont les principaux revenus proviennent respectivement de la vente de terminaux et de logiciels, toutes les autres - Google, Facebook, Amazon, Alibaba, Tencent - ont un modèle d'affaires fondé sur l'exploitation des données. Pourtant, les mécanismes de production et de valorisation des données n'ont pas grand-chose à voir avec ceux de l'industrie pétrolière. Lorsqu'on examine les caractéristiques de la donnée, on constate que sa nature économique particulière et son mode de production sont loin d'en faire une marchandise comme une autre [7].

Une nature économique singulière

D'un point de vue économique, les données sont un bien non rival. Cela signifie que leur utilisation par un agent ne diminue pas leur disponibilité pour les autres agents économiques. Contrairement au pétrole, avec la donnée il n'y a donc aucun risque d'épuisement de la ressource. Si, par exemple, une entreprise collecte et exploite les coordonnées personnelles, le sexe, la date de naissance, les centres d'intérêt ou encore les données de géolocalisation d'une personne, il est possible pour ses concurrents

[7] Isaac Henri, « La donnée, une marchandise comme les autres ? » *Enjeux numériques*, n°2, juin 2018, p. 20.

de disposer et de faire un usage simultané des mêmes informations. Sous réserve bien entendu de l'accord de la personne concernée, qui a le pouvoir de rendre cette donnée disponible ou non et d'autoriser ou non son usage par des tiers.

Les données constituent par ailleurs un bien immatériel, facile à répliquer et au coût marginal de production très faible. La production de données impose souvent des investissements de départ très importants : la mise en œuvre de dispositifs techniques aptes à produire les données (par exemple l'installation de capteurs), l'existence des capacités de stockage indispensables à leur conservation et le déploiement de réseaux de communication qui permettent leur circulation entre leur point de création et leur point de stockage. Cependant, une fois ces dispositifs mis en place, la création d'une donnée supplémentaire a un coût marginal presque nul. Les capteurs installés sur les autoroutes ou les réseaux électriques et les objets connectés peuvent par exemple collecter des données en continu. Il faut ajouter à cela qu'un nombre important de données sont produites directement et gratuitement par les utilisateurs eux-mêmes en échange de l'accès à un service numérique.

Les données existent en abondance. De nouvelles données sont produites chaque jour, toujours plus nombreuses. Les proportions sont vertigineuses. Dans une étude publiée en 2013, IBM estimait que 90 % des données créées dans le monde l'avaient été au cours des deux dernières années [8]. D'après une autre étude, le nombre total de données disponibles devrait avoir été multiplié par cinq entre 2018 et 2025 avec le développement des objets connectés [9]. Par ailleurs, comme leur utilisation ne diminue pas le stock de données disponibles, les données ne se situent pas au sein d'une économie de la rareté.

Les données s'apparentent plutôt à un bien de club [10], c'est-à-dire que le détenteur de données peut décider d'en permettre l'accès à des tiers ou non. « Si chaque entreprise peut donc en théorie acquérir des données de tierce partie, et ainsi rattraper l'avantage des entreprises établies en matière d'accès aux données, en pratique cela pourrait s'avérer impossible au regard de la quantité et de la qualité des données déjà détenues par l'entreprise en place », écrit l'Autorité de la concurrence française dans

[8] www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[9] International Data Corporation, *Data Age 2025*, rapport parrainé par Seagate, 2017.

[10] Isaac Henri, « La donnée numérique, bien public ou instrument de profit », *Pouvoirs*, n° 164, 2018, p. 77.

un rapport produit conjointement avec son homologue allemand [11]. « Dans certains secteurs, poursuit-elle, les principales entreprises peuvent disposer d'une base de clientèle et d'informations tellement large que la question de savoir si une entreprise tierce est en mesure de reproduire le même volume et la même variété de données se pose ».

Pour autant, les données ne doivent pas être considérées comme une facilité essentielle. Pour la théorie économique, une facilité essentielle est une ressource à la fois non reproductible dans des conditions économiques raisonnables et indispensable à l'exercice d'une activité. Le fait pour une entreprise de refuser l'accès à une facilité essentielle peut être considéré comme un abus de position dominante en droit de la concurrence. Or, aujourd'hui, une entreprise a généralement la faculté d'accéder à des données similaires à celles qui sont détenues par une autre entreprise ou d'accéder à des données de substitution. Plusieurs chemins permettent d'avoir accès aux mêmes données et différentes données permettent d'obtenir les mêmes informations. « Si un réseau publicitaire a déterminé que l'utilisateur d'un appareil mobile donné vit dans le Connecticut, aime voyager et possède un chien, rien n'empêche un autre réseau publicitaire d'acquérir les mêmes informations – en effet, concernant un utilisateur fréquent d'Internet, il est probable que des dizaines d'entreprises créent un profil similaire » [12]. Une entreprise qui ne peut pas avoir accès aux données relatives aux centres d'intérêt d'un de ses utilisateurs potentiels par le biais des réseaux sociaux peut par exemple s'appuyer sur l'activité de cette personne sur les moteurs de recherche ou sur les produits qu'elle a consultés sur des sites marchands.

Il n'en demeure pas moins que les données détenues par certains opérateurs peuvent ponctuellement être assimilées à des facilités essentielles. En 2014, par exemple, l'Autorité de la concurrence a pu imposer à GDF-Suez d'accorder à ses concurrents l'accès à certaines données collectées lorsque l'entreprise était bénéficiaire d'un monopole sur un marché réglementé [13]. GDF-Suez se servait des bases de données constituées à l'époque du monopole public (coordonnées des utilisateurs, données de consommation) pour proposer des offres personnalisées à ses clients. « L'objectif de cette mesure conservatoire était de permettre

[11] Autorité de la concurrence et Bunderskartellamt, *Droit de la concurrence et données*, 10 mai 2016, p. 47.

[12] Isaac Henri, « La donnée numérique, bien public ou instrument de profit », *Pouvoirs*, n° 164, 2018, p. 77.

[13] Autorité de la concurrence, Décision 14-MC-02 du 09.09.2014.

à tous les fournisseurs de bénéficier du même niveau d'informations pertinentes pour formuler leurs offres aux consommateurs (aucune information publique ou privée n'existant sur les ménages souscrivant des contrats de fourniture de gaz) », selon l'Autorité de la concurrence [14].

De la diversité des données

Il existe une grande variété de données en termes de nature et de format [15]. Toutes ne sont pas soumises au même régime juridique et toutes ne présentent pas le même intérêt économique et sociétal. De la géolocalisation à l'identification d'une personne, de ses centres d'intérêt à ses achats en ligne, de la qualité de l'air d'une pièce aux cycles de sommeil d'un dormeur en passant par les données publiques, les données ne sont pas toutes également disponibles ou exploitables, et toutes n'ont pas un même attrait.

Les données sont souvent fournies volontairement par les utilisateurs de services. Par exemple, dans le cadre d'une transaction commerciale, le consommateur renseigne son nom, son adresse et ses coordonnées de paiement. Sur les réseaux sociaux, les utilisateurs peuvent communiquer, outre leur nom et leur adresse, leur activité, leur localisation, leurs goûts, des photos, des vidéos ou encore des commentaires sur l'actualité. Dans un moteur de recherche, les mots clés saisis et les sites Internet visités parmi les résultats proposés renseignent sur les centres d'intérêt et les préférences des utilisateurs. Les internautes peuvent également fournir des données dans une logique d'entraide, comme sur Wikipedia ou dans le cadre du projet *Open Food Facts* [16].

Les données peuvent également être obtenues par observation des traces que nous laissons au cours de nos interactions sur Internet et avec les appareils connectés. Ces données d'usage ou données ancillaires peuvent

[14] Autorité de la concurrence et Bunderskartellamt, *Droit de la concurrence et données, op. cit.*, p. 23.

[15] OCDE, *Data-driven Innovation: Big Data for Growth and Well-Being*, Éditions OCDE, 2015.

[16] L'application *Open Food Facts* attribue un score nutritionnel basé sur la lecture des étiquettes à plusieurs dizaines de milliers de produits. Ce projet, qui voit collaborer des milliers de contributeurs, renseigne le consommateur sur l'origine des produits et les additifs qu'ils contiennent. Les producteurs eux-mêmes sont invités à compléter les informations disponibles sur l'application avec des données certifiées mises à jour. Renaissance Numérique, *Les défis de l'agriculture connectée dans une société numérique*, novembre 2015.

être constituées par le suivi d'un internaute dans sa navigation (adresse IP, sites visités, achats antérieurs, le temps passé sur une page, le passage d'une page à l'autre, les clics). Il peut également s'agir des données de géolocalisation fournies par les téléphones mobiles ou de toutes celles qui sont recueillies par les capteurs et autres objets connectés.

Les données sont également susceptibles d'être inférées à partir de l'analyse d'autres données, en particulier à partir du recoupement de plusieurs données. Le recoupement des données personnelles d'utilisateurs peut par exemple être utilisé à des fins de comparaison ou d'application d'algorithmes de filtrage pour proposer des recommandations à d'autres utilisateurs. Les données d'utilisateurs obtenues par inférence sont notamment : les revenus, l'état de santé, la réactivité à la publicité, l'orientation politique, la fidélité du consommateur, les loisirs, les biais comportementaux.

Une donnée peut avoir plusieurs vies juridiques

À qui appartiennent les données ? Dans quelles conditions peut-on les utiliser ? Autant de questions à se poser pour définir sa stratégie en matière de données. L'exploitation des données est conditionnée par le cadre légal qui détermine les règles applicables à leur collecte, à leur traitement et à leur circulation. Selon la nature des données et le contexte, leurs conditions d'utilisation et le régime juridique varient : données personnelles, données publiques, données soumises aux secrets légaux, données non personnelles.

Il existe des porosités entre les différents statuts des données. Non seulement ces différents régimes ne sont pas nécessairement exclusifs les uns des autres : une donnée non personnelle peut être soumise aux secrets légaux ; une donnée personnelle peut être une donnée publique (par exemple, l'annuaire des administrations ou le répertoire national des élus). Mais, au cours de son existence, une donnée peut encore changer de statut : les données personnelles peuvent être anonymisées, les secrets légaux peuvent être levés, etc. Ainsi, les données qui identifient directement une personne sont « naturellement » personnelles, alors que celles qui peuvent identifier indirectement une personne ne le sont pas, tout dépend du contexte.

Données personnelles et données non personnelles

C'est le Règlement général sur la protection des données (RGPD) adopté par l'Union européenne le 27 avril 2016 qui fixe désormais le cadre légal général de la protection des données personnelles en France et en Europe.

Le RGPD renforce le contrôle des citoyens sur l'utilisation qui peut être faite de leurs données. Des règles spécifiques encadrent et limitent la collecte et l'utilisation des données à caractère personnel dans le but de protéger la vie privée des personnes. À l'instar des autres droits, la protection des données personnelles n'a cependant pas un caractère absolu. Elle peut connaître des limitations au nom de la conciliation avec d'autres droits fondamentaux et de la poursuite d'autres objectifs tels que la lutte contre la criminalité et la préservation de la capacité des entreprises à innover. Surtout, elle n'exclut pas l'exploitation des données personnelles à des fins économiques, ni leur circulation, éventuellement hors d'Europe.

Aux origines du droit de la protection des données personnelles

L'idée d'un droit à la protection des données personnelles est née de la prise de conscience que l'exploitation de certaines données était susceptible de se faire au détriment du respect de la vie privée des personnes. Cette prise de conscience est bien antérieure à la démocratisation des outils informatiques. Elle remonte aux années 1970 avec l'apparition des premiers moyens de traitement informatisé des données.

En France, la problématique fait son entrée dans le débat public avec la publication dans le journal *Le Monde* daté du 21 mars 1974 d'un article de Philippe Boucher « SAFARI ou la chasse aux Français ». Il révèle le projet du gouvernement de réaliser l'interconnexion des fichiers nominatifs de l'administration française [17], afin de faciliter la circulation des informations entre administrations à l'aide d'un identifiant unique et obligatoire. L'émoi suscité dans l'opinion publique par ces révélations fut à l'origine de la création de la Commission nationale de l'informatique et des libertés en 1978, la CNIL, et du premier régime de protection des données personnelles en France par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Les grands principes fixés par le législateur à l'époque ont depuis connu une remarquable stabilité. Ils sont désormais repris et complétés au niveau européen par le règlement général sur la protection des données personnelles (RGPD) entré en vigueur au 25 mai 2018.

Une donnée à caractère personnel désigne toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement. Par exemple, sont considérées comme des données personnelles : un nom, un numéro de téléphone, une adresse mail, un numéro de sécurité sociale, une adresse IP, un identifiant de connexion informatique, une photographie, un enregistrement vocal, un code ADN, etc. Une donnée personnelle peut faire l'objet d'une anonymisation de façon à rendre impossible toute identification de la personne concernée (agrégation des informations personnelles dans une base de données, visage flouté). Le processus doit être irréversible, c'est-à-dire que toutes les informations directement ou indirectement identifiantes sont supprimées ou modifiées. Lorsque l'opération est réversible, on utilise le terme de pseudonymisation. Enfin, si la personne reste malgré tout identifiable par recoupement d'informations (âge, sexe, ville, emploi, etc.), les données sont toujours considérées comme personnelles.

Le RGPD renforce les droits existants et introduit de nouveaux droits, afin de permettre aux personnes de garder la maîtrise des informations qui les concernent lorsque celles-ci sont traitées par un organisme. Ce texte adopte une nouvelle architecture régulatrice, plus équilibrée entre les schémas de responsabilité. L'effectivité de la régulation est ainsi moins dépendante de l'action individuelle (les droits du citoyen) et fait davantage supporter cette charge sur les organisations (nouvelles obligations) et les dispositifs technologiques (*Privacy by Design*) [18]. Peu importe que cet organisme soit public ou privé. Préalablement à toute collecte d'informations, le responsable de traitement doit impérativement informer les personnes concernées sur la finalité du fichier. Les personnes bénéficient par ailleurs d'un ensemble de droits sur leurs données : droit d'opposition, droit d'accès et de rectification des informations les concernant, droit à l'effacement (ou droit à l'oubli), droit à la limitation du traitement, droit à la portabilité des données. En cas de mauvaise application du règlement, le responsable de traitement risque d'importantes sanctions administratives. Les amendes administratives prévues par le règlement peuvent s'élever à 10 ou 20 millions d'euros ou 2 à 4 % du chiffre d'affaires mondial d'une entreprise.

[17] Chignard Simon, « Safari ou la chasse aux Français, 40 ans après », publié sur le blog *données ouvertes*, 26 janvier 2018 https://donneesouvertes.info/2018/01/26/safari-la-chasse-aux-francais-40-ans-apres/

[18] Renaissance Numérique, « Données personnelles : rendre ces droits effectifs », mai 2018.

De leur côté, les données non personnelles rassemblent toutes les données qui ne sont pas des données personnelles au sens du RGPD - autrement dit, les données qui ne sont pas rattachables à une personne physique identifiée ou identifiable. Il peut notamment s'agir des traces laissées par des utilisateurs non identifiés, des données agrégées et anonymisées ou encore des données produites par les objets connectés, comme les données des machines industrielles ou des machines agricoles. Même si dans ce cas, elles sont susceptibles de devenir personnelles : si nous rattachons une donnée à une machine et que nous connaissons le propriétaire qui est une personne physique.

Le 14 novembre 2018, l'Union européenne a adopté un règlement sur la libre circulation des données non personnelles [19]. Le texte vise à lever certains obstacles à la circulation de ce type de données, et notamment certaines exigences nationales qui imposaient de localiser les données sur un territoire précis, incompatibles avec les dispositions européennes.

L'utilisation des données non personnelles s'inscrit dans un cadre contractuel entre fournisseurs et réutilisateurs de données. Le contrat précise notamment les modalités d'accès aux données, les conditions de réutilisation et de commercialisation.

Les bases de données, quant à elles, relèvent de l'article 7 de la directive 96/CE, transposé en droit français et figurant désormais à l'article L. 341-1 du Code de la propriété intellectuelle. En vertu de ce droit, « le producteur d'une base de données, entendu comme la personne qui prend l'initiative et le risque des investissements correspondants, bénéficie d'une protection du contenu de la base lorsque la constitution, la vérification ou la présentation de celui-ci atteste d'un investissement financier, matériel ou humain substantiel ». Il s'agit de permettre à l'entreprise de protéger les investissements qui ont été nécessaires à la constitution d'une base de données (depuis la collecte des données jusqu'à la vérification de leur exactitude), en privant les tiers de la faculté de réutiliser une partie du contenu de celle-ci. Par ailleurs, la directive 96/9/CE reconnaît l'existence d'un droit d'auteur sur les bases de données, qui figure à l'article L. 112-3 du Code de la propriété intellectuelle. Mais un tel droit s'applique uniquement à la structure d'une base de données et ne concerne pas son contenu.

[19] Règlement (UE) 2018/1807 du Parlement européen et du Conseil du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne.

Les données publiques

La politique d'ouverture des données publiques a été initiée en France dès le début des années 2000. Ce mouvement s'est poursuivi avec la loi pour une République numérique du 7 octobre 2016, qui a précisé les règles applicables aux données publiques en France, au travers de trois grands principes :

- l'ouverture des données publiques par défaut ;
- la gratuité de leur mise à disposition ;
- la liberté de leur réutilisation par les citoyens et les organisations.

L'ouverture concerne les données détenues par les administrations de l'État, les collectivités territoriales de plus de 3 500 habitants, les établissements publics, mais aussi par les personnes privées titulaires d'une délégation de service public ou dont les activités sont subventionnées par la puissance publique. Elle s'applique à leurs bases de données et aux données qui présentent un intérêt économique, social, sanitaire ou environnemental. Elle concerne également les algorithmes publics, qui sont de plus en plus fréquents dans les prises de décision de l'administration (par exemple le système Parcoursup qui détermine l'affectation des élèves dans les établissements d'enseignement supérieur). L'ouverture ne concerne toutefois ni les informations personnelles, ni celles touchant à la sécurité nationale, ni celles couvertes par les différents secrets légaux.

La réutilisation des données publiques peut être soumise à un système de licence qui régit les conditions dans lesquelles les données peuvent être utilisées. En France, deux licences sont recommandées par l'État pour l'ouverture des données publiques :

- La licence ouverte de la mission Etalab : elle permet de reproduire, redistribuer, modifier et exploiter à titre commercial les données. Le réutilisateur est simplement tenu de mentionner la source des données qu'il utilise et d'indiquer la date de leur dernière mise à jour.
- L'Open database Licence (ODbL) : elle permet de copier, modifier et de faire un usage commercial d'une base de données à condition que la réutilisation qui en est faite soit diffusée dans des conditions de partage identiques ("*share alike*"). Le réutilisateur doit donc maintenir ouvertes les bases de données qu'il utilise et les modifications qu'il y apporte.

3_ Des données à la création de valeur

Les trois facettes de la valeur des données

Compte tenu de la grande variété des données, leur valeur est difficile à mesurer. Aucun outil de mesure statistique ne fait aujourd'hui consensus, qu'il s'agisse d'apprécier la valeur globale des données dans l'économie ou de mesurer leur valeur individuelle [20]. Ainsi, en 2017, l'Union européenne estimait qu'en 2015, l'économie fondée sur les données représentait 272 milliards d'euros, soit 1,87 % du PIB de l'Union, et qu'elle devrait représenter 643 milliards d'euros d'ici à 2020, soit 3,17 % du PIB global de l'Union [21]. Et dans une étude de 2013 consacrée à l'évaluation des données personnelles, l'OCDE notait que les courtiers en données vendent une adresse postale pour 0,50 dollar, une date de naissance pour 2 dollars et un numéro de sécurité sociale pour 8 dollars. Sur les marchés illégaux, le prix d'un numéro de carte de bancaire dérobé peut atteindre 30 dollars [22].

Contrairement à ce que l'on observe pour l'or ou le pétrole, il n'existe, pour les données, ni unité de mesure de référence ni cours qui y soient associés. Leur valeur d'échange est extrêmement volatile dans le temps et en fonction du type de données en cause. Certaines données n'ont de valeur que parce qu'elles sont disponibles en temps réel. De plus, les données ne ressemblent pas aux ressources traditionnelles : leur extraction, leur valorisation, leur échange empruntent plusieurs voies, parfois très différentes les unes des autres. Si un consensus émerge, c'est pour considérer que les données n'ont la plupart du temps pas de valeur intrinsèque. La valeur des données réside principalement dans l'usage qui peut en être fait, dans leur croisement avec d'autres données et leur circulation. Ce qui a une valeur d'échange, en revanche, c'est le service qui est commercialisé à partir de l'exploitation des données.

[20] Mandel Michael, "Beyond Goods and Services. The (Unmeasured) Rise of the Data-Drive Economy", *Progressive Policy Institute*, Policy Memo, octobre 2012 ; Brynjolfsson Erik, "Why it Matters that the GDP Ignores Free Goods", Techonomy 2012, Conference Report, 7 décembre 2012.

[21] Commission Européenne, "Créer une économie européenne fondée sur les données", Communication, COM (2017) 9 final, 10 janvier 2017.

[22] OCDE, "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value", *Documents de travail de l'OCDE sur l'Économie numérique*, n° 220, Éditions OCDE, 2013.

Il est toutefois possible de distinguer trois aspects de la valeur des données [23].

Les données constituent d'abord une matière première, une marchandise qui peut se vendre et s'acheter sur le marché. Certaines données acquises par une organisation sont directement monétisables auprès de tiers, qui peuvent, par exemple, chercher à acquérir les coordonnées de nouveaux clients potentiels ou se servir des données détenues par une entreprise pour lancer des campagnes publicitaires ciblées ou proposer de nouvelles offres de services.

Le marché de la donnée est également investi par des « data brokers » ou « courtiers en données », qui sont des entreprises spécialisées dans la commercialisation à des tiers de jeux de données sous forme de fichiers ou de flux de données qu'elles acquièrent auprès de différentes entreprises ou organisations. Leur modèle économique s'appuie sur les nouvelles possibilités offertes par la multiplication des données produites et les technologies permettant de les exploiter. Les données concernées peuvent être plus ou moins précises, par exemple le type de logement occupé par un individu ou le modèle de voiture qu'il conduit.

Aux États-Unis, plusieurs startups proposent par ailleurs d'associer les particuliers à la marchandisation de leurs données. Le principe est simple : en échange de leurs données, les utilisateurs reçoivent une rémunération. Par exemple, l'entreprise Datacoup propose de rémunérer ses utilisateurs 8 dollars par mois en échange de la collecte de leurs données sur les réseaux sociaux et de leurs relevés de transactions bancaires, qu'elle vend ensuite aux entreprises.

La commercialisation de données contre une somme d'argent ne représente toutefois pas l'opération la plus courante si l'on rapporte celle-ci à la quantité de données qui sont produites chaque jour. En dehors des données personnelles, le marché des données reste relativement embryonnaire.

Les données sont également un levier économique et un facteur d'innovation. L'agrégation de données permet de produire de la connaissance qui peut ensuite être utilisée pour mieux décider, mieux produire ou mieux agir. Les données peuvent bien sûr être mises à profit pour améliorer les performances d'une organisation, qu'il s'agisse d'optimiser les processus de production ou de développer de nouvelles

[23] Chignard Simon et Benyayer Louis-David, *Datanomics, les nouveaux business models des données*, éditions FYP, 2015.

activités. « Uber peut, par exemple, décider d'ouvrir ou non son service dans une ville à partir du nombre de personnes qui tentent chaque jour d'utiliser son application pour se déplacer », explique Simon Chignard. Les données peuvent par ailleurs servir au profilage des consommateurs. À partir des données dont on dispose sur un individu et sur la base de corrélations statistiques, on infère certaines de ses caractéristiques. Le profilage peut servir à des campagnes de publicité ciblées ou pour améliorer et personnaliser le service. Il peut également être utilisé par les banques et par les assurances pour évaluer les risques.

L'exploitation des données peut également contribuer à l'amélioration des services publics : renforcement des mobilités urbaines, amélioration de la pharmacovigilance, lutte contre la fraude sociale ou fiscale, ou encore prévention des atteintes à la sécurité publique dans le cadre de ce qu'on appelle la police prédictive. À cette fin, des logiciels d'aide à la décision peuvent être utilisés pour identifier les zones à risque où la police doit patrouiller à partir des statistiques des crimes passés [24]. De nombreuses collectivités ont également mis en place des applications mobiles qui permettent aux habitants de remonter, de façon volontaire, les dysfonctionnements constatés sur la voie publique (feux de circulation cassés, poubelles renversées, etc.).

Les données peuvent enfin représenter un actif stratégique, en ce sens que celui qui les contrôle est à même d'exercer un pouvoir de marché. Les données permettent généralement à une entreprise d'obtenir un avantage compétitif par rapport à ses rivales, et cela se vérifie tout particulièrement dans les secteurs économiques qui reposent sur une exploitation des données. On parle d'un « effet boule de neige » [25] : l'entreprise qui a accès au plus grand volume de données est en mesure d'offrir à ses clients les meilleurs services, ce qui lui permet d'attirer de nouveaux clients et donc de collecter encore davantage de données. Au final, cette entreprise est en mesure de définir de nouvelles fonctionnalités et de nouveaux produits, d'améliorer ses algorithmes et d'entrer sur des marchés adjacents. La concentration des données entre les mains de certaines entreprises peut contraindre l'arrivée sur le marché d'autres acteurs, qu'ils soient plus petits ou pas encore numérisés.

[24] Legros Claire, « À Marseille, le big data au service de la sécurité dans la ville », *Le Monde* du 8 décembre 2017 ; Paquette Emmanuel, « Police : le logiciel d'anticipation des crimes face à ses détracteurs », *L'Express* du 8 février 2018.

[25] Autorité de la concurrence et Bunderskartellamt, *Droit de la concurrence et données*, 10 mai 2016, p. 15.

Valeur unitaire et valeur agrégée des données

Une donnée prise isolément n'a que peu de valeur. Les données primaires doivent être agrégées et traitées pour qu'il puisse en être dégagé une valeur suffisamment conséquente. Le plus souvent, une donnée n'est d'ailleurs exploitable que lorsqu'elle est contextualisée, utilisée en combinaison avec les métadonnées qui l'accompagnent. Les métadonnées désignent l'ensemble des données qui servent à décrire une autre donnée, qu'il s'agisse du nom, de la taille et de la date de création d'un fichier, du titre d'une chanson ou d'un album associés à un fichier audio ou encore du descriptif d'une photographie. « De très nombreux services digitaux reposent de façon quasi exclusive sur la gestion et l'exploitation des métadonnées plutôt que des données elles-mêmes », selon Henri Isaac. « Il en va ainsi des systèmes de messagerie électronique, des moteurs de recherche, des services musicaux en ligne (les playlists proposées ne sont qu'une gestion des métadonnées), les services de catalogage des photographies, les catalogues de commerce électronique, les services de réseaux sociaux, etc. »[26]. Par exemple, sans ses métadonnées, un fichier musical n'est pas exploitable : il ne peut pas être classé ni inclus dans une liste de lecture personnalisée. En revanche, dès lors qu'il est associé à d'autres données, comme le nombre de fois où il a été écouté par les utilisateurs de services de musique en ligne, les lieux et les moments d'écoute, les fichiers musicaux qui lui sont associés dans les listes de lecture créées par d'autres usagers, il peut être utilisé pour élaborer des services complémentaires, comme la création de playlists personnalisées.

Plus encore, la constitution de larges bases de données représente un enjeu majeur pour le développement de l'intelligence artificielle contemporaine fondée sur les techniques de l'apprentissage automatique (*machine learning*)[27]. En ce sens, l'accès à la donnée procure un avantage compétitif essentiel dans la compétition mondiale pour l'IA. Or sur ce terrain, l'Europe paraît aujourd'hui distancée par les États-Unis et la Chine.

Le traitement, un levier essentiel de la valeur des données

Le simple fait de collecter un nombre important de données ne garantit pas la création de valeur. La valorisation des données est le résultat de

[26] Isaac Henri, « Données, valeur et business model », *Les Cahiers Scientifiques*, n°21, 2016, p. 18.

[27] Villani Cédric, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, rapport issu de la mission confiée par le Premier ministre, 2018, p. 27.

mécanismes complexes de collecte, de stockage et de traitement des données destinés à la production de biens ou services à valeur ajoutée pour les consommateurs ou usagers.

Un préalable indispensable à toute extraction de valeur à partir des données est la production de données de qualité. De nombreuses données sont en effet rendues inaccessibles ou inutilisables du fait de la mauvaise qualité de leur production. Cette dernière est généralement la première difficulté à laquelle se trouvent confrontées les organisations qui entendent recentrer une partie de leur activité autour de l'exploitation des données en leur possession. D'après une étude menée par le cabinet Gartner, plus de la moitié des programmes de gestion de la relation client et des projets d'entrepôts de données échouent en raison de la piètre qualité des données collectées [28]. Une autre étude a pu recenser 122 organisations en Europe et aux États-Unis qui avaient perdu plus de 1 200 milliards de dollars pour la même raison [29].

La plupart des données nécessitent donc d'être retraitées avant de pouvoir être utilisées. L'objet de ce traitement est de rendre les données fiables, cohérentes, accessibles pour les utilisateurs et conformes aux normes de sécurité et de confidentialité. Il peut par exemple s'agir d'enrichir et d'annoter les données en y adjoignant les métadonnées qui en qualifient le contenu. De telles opérations sont toutefois longues et pénibles et nécessitent la mobilisation d'importantes ressources humaines et financières. C'est pourquoi certains acteurs font appel pour ce faire à des sources de financements spécifiques ou à du *crowdsourcing*.

Ainsi, certaines plateformes de micro-travail proposent par exemple aux entreprises d'externaliser des tâches comme la collecte et la vérification de données auprès des internautes contre rémunération. Autre exemple : l'internaute qui, pour accéder à un site Internet, retranscrit les caractères ou identifie les images qui lui sont présentées, participe à la constitution de gigantesques bases de données qui serviront ensuite à alimenter les programmes d'IA.

Une fois garantie la qualité des bases de données, leur traitement par des méthodes d'analyse est une étape indispensable de la création de valeur. Le traitement des données permet aux organisations d'améliorer

[28] Gregory Adrian "Data Governance – Protecting and Unleashing the Value of your Customer Data Assets", *Journal of Direct, Data and Digital Marketing practice*, vol. 12, n° 3, 2011, p. 230-248.

[29] English Larry, *Information Quality Applied: Best Practices for Improving Business Information Processes*, Wiley Publishing Inc, 2009.

leur processus de gestion, qu'il s'agisse de leur méthode de production ou de leur relation avec leurs publics. Si l'utilisation de l'analyse de données est ancienne, les progrès récents en algorithmes, rendus possibles par la puissance de calcul des ordinateurs et la volumétrie croissante de données, ont introduit de nouvelles possibilités dans l'exploitation des données [30]. En ce sens, explique le sociologue Dominique Cardon, « les algorithmes constituent d'indispensables instruments pour classer les informations (moteur de recherche), personnaliser les affichages publicitaires (ciblage comportemental), recommander (prédictions culturelles) ou guider vers la meilleure route possible (GPS) » [31].

IA, *machine learning* et *deep learning*

Parmi les différentes méthodes algorithmiques existantes, les techniques d'apprentissage par données issues des recherches en intelligence artificielle marquent une rupture. Comme le souligne le rapport Villani : « elles marquent le passage progressif d'une logique de programmation à une logique d'apprentissage » [32]. L'apprentissage automatique (*machine learning*) et l'apprentissage profond (*deep learning*) sont des techniques d'apprentissage qui permettent de produire des modèles directement à partir des données. Autrement dit, c'est le programme informatique qui établit les variables pertinentes à partir de l'objectif qui lui a été assigné et qui ajuste ensuite le modèle sur la base des nouvelles données qu'il continue de recevoir. Les données combinées à ces algorithmes permettent par conséquent la création de services à forte valeur ajoutée qui apprennent continuellement des données d'usage ou d'un quelconque processus opérationnel pour améliorer leur offre.

Le *machine learning* permet d'entraîner l'algorithme à partir d'exemples pour créer un modèle prédictif.

[30] « Les algorithmes sont des procédures informatiques réglées qui permettent d'opérer des calculs à partir de données », Cardon Dominique, « Le pouvoir des algorithmes », *Pouvoirs*, n°164, p. 64.

[31] Cardon Dominique, « Le pouvoir des algorithmes », *op. cit.*

[32] Villani Cédric, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, rapport issu de la mission confiée par le Premier ministre, 2018, p. 26.

Henri Isaac, « Données, valeur et business model », Cahiers scientifiques, p. 23-24 (extrait)

Le *machine learning* chez Uber

Tout comme Airbnb, Uber utilise des algorithmes de *machine learning* pour déployer son modèle d'affaires de mise en relation entre une demande de déplacement urbain et d'offre de transports des chauffeurs privés (VTC). Uber utilise par exemple un modèle Bayésien qui lui permet de prédire la destination du client et ainsi de pré-positionner ses chauffeurs dans les localisations les plus susceptibles de voir une demande exister. Mais la proposition de valeur d'Uber réside dans la mise à disposition d'un chauffeur dans un temps minimum. Uber doit minimiser le temps d'attente du passager et maximiser le nombre de passagers pour un chauffeur. Il faut donc un modèle qui permette à l'offre et à la demande de répondre à ce double programme de maximisation. Ainsi dans la ville de San Francisco seulement 16 % des taxis arrivent en moins de dix minutes contre 97 % des chauffeurs VTC. Cette disponibilité des véhicules Uber est au cœur de l'avantage concurrentiel et d'autres études démontrent que plus Uber est présent depuis longtemps, moins les clients ont de patience pour attendre un véhicule VTC.

Ce programme s'avère complexe, car il peut différer d'une ville à l'autre, mais aussi dans la journée (heure creuse versus heure pleine). En effet, pendant les heures creuses, les chauffeurs sont prêts à parcourir de plus longues distances pour obtenir une course, ce qui n'est pas le cas aux heures d'affluence où une demande peut émerger très rapidement à proximité du chauffeur et où le chauffeur a intérêt à attendre quelques minutes plutôt que parcourir une plus longue distance pour obtenir un nouveau client. En utilisant un système multi-agents et en intégrant différents comportements du côté des passagers et des chauffeurs, Uber a pu déterminer que les conducteurs utilisant son système de dispatching central gagnent entre 25-50 % de plus que les chauffeurs qui utilisent leurs seules connaissances de la ville pour la recherche d'un passager.

Mais là où l'utilisation des algorithmes est encore plus saillante dans le modèle d'Uber est son modèle de tarification dynamique qui équilibre l'offre et la demande. Le système de tarification dynamique « *surge price* », mis en œuvre aux heures de pointe et à certaines périodes (soirées du vendredi et samedi, nuit du réveillon, etc.), qui multiplie le prix par un facteur 2 ou plus, incite les chauffeurs à travailler à ces heures, car ils augmentent leur revenu. Ainsi, le service Uber continue d'offrir des chauffeurs à des périodes où il est traditionnellement difficile de trouver un taxi. Les données montrent que l'offre augmente de 70 à 80 % pendant ces périodes et que deux tiers des demandes

sont absorbées. Cet algorithme, introduit en 2011, a depuis évolué grâce aux données collectées sur toutes les courses et le modèle a pu se raffiner. Il concerne environ 10 % de l'ensemble des courses sur la plateforme Uber. Désormais, les clients sont avertis, lorsque la période est en « *surge price* », tout comme lorsque cette période se termine.

La collecte des évaluations des clients et des chauffeurs constitue également une source importante de données utilisées pour réguler le système et maintenir sa qualité et garantir la sécurité des passagers.

L'apprentissage profond, ou *deep learning*, est une technique d'apprentissage dérivée de l'apprentissage automatique qui fonctionne grâce à un réseau de neurones artificiels. Une fois qu'il est suffisamment entraîné, le réseau de neurones est capable d'identifier les caractéristiques essentielles du traitement sans nécessiter aucune intervention humaine. Le *deep learning* a notamment permis de faire des progrès spectaculaires en matière de reconnaissance d'images. Par exemple, une machine est capable d'identifier une voiture sur une photographie en utilisant les caractéristiques qu'elle a elle-même identifiées à partir de millions de photographies contenant des voitures. L'apprentissage profond a ceci de particulier que le modèle est directement « appris » des données en recourant à des caractéristiques et des relations entre données qui sont parfois impossibles à identifier pour un être humain. Les applications de l'apprentissage profond sont innombrables. Ce dernier peut notamment être utilisé pour identifier un individu sur une photo ou une vidéo, analyser des émotions sur un visage, lire sur les lèvres, établir un système de recommandation musicale, détecter des fraudes ou encore proposer un diagnostic à partir d'une imagerie médicale.

La clé du développement du *deep learning* tient toutefois autant à la qualité des algorithmes qu'au volume de données disponibles. Pour être correctement entraînés, les algorithmes ont en effet besoin d'un nombre très élevé d'exemples. À en croire Hal Varian, économiste en chef chez Google, les données auraient cependant « des rendements d'échelle décroissants » [33]. Autrement dit, l'utilité marginale d'une donnée supplémentaire diminuerait avec la quantité de données déjà collectées. Tant et si bien qu'à partir d'un certain seuil, il deviendrait inutile de

[33] Cité par *The Economist*, "Fuel of the future", 6 mai 2017.

recueillir davantage de données. Selon lui, le succès d'entreprises comme Google tiendrait donc plus aux recettes qu'aux ingrédients employés, à la qualité de ses algorithmes plutôt qu'au volume de ses données. Mais de telles analyses sont aujourd'hui remises en cause au regard des avancées réalisées en matière d'IA : la quantité de données disponibles et leur fraîcheur deviennent une ressource stratégique dans l'amélioration des algorithmes auto-apprenants. En ce sens, les données constituent la matière première de l'IA et l'accès aux données tend, au-delà des enjeux économiques, à devenir un enjeu de souveraineté et d'autonomie pour les États [34].

La valeur des données n'est donc pas seulement liée à leur abondance et à leur qualité. Leur traitement apparaît tout aussi indispensable à la création de biens et services performants. Tout au bout de la chaîne, la dernière étape du processus de création de valeur à partir des données tient à leur restitution sous forme de services à destination de l'utilisateur.

L'étape de la restitution, ou la valeur pour les utilisateurs

L'enjeu pour les organisations est d'être en mesure de proposer un service à plus forte valeur ajoutée à l'issue de la phase de traitement des données.

L'un des objectifs pour elles est de réussir à impliquer leurs utilisateurs dès la phase de conception du service numérique. Elles peuvent pour cela recourir au *design thinking* développé à l'université de Stanford dans les années 1980 par Rolf Faste sur la base des travaux de Robert McKim. Cette méthode prend notamment pour point de départ les besoins des utilisateurs finaux et leur ressenti et intègre ensuite leurs retours pour améliorer le produit ou l'offre de service.

Prenons l'exemple des compteurs de vélo connectés. La fonction première de ce type d'objets est de fournir à leur utilisateur des indications sur sa vitesse et sur les distances qu'il a parcourues. Mais le service proposé aura d'autant plus de valeur aux yeux de l'utilisateur qu'il sera en mesure de l'inciter à exercer une activité physique régulière et à améliorer ses performances. Le compteur pourra ainsi par exemple accompagner l'affichage de données brutes de messages incitatifs liés à l'évolution des performances du cycliste et aux caractéristiques de son parcours. Il pourra également informer l'utilisateur sur le nombre de calories qu'il

[34] Morin-Dessailly Catherine, *L'Union européenne, colonie du monde numérique ?* Rapport d'information du Sénat, 2013.

a dépensées durant son trajet ou encore calculer les économies qu'il a réalisées en choisissant d'effectuer ses déplacements à vélo plutôt qu'en voiture. Le compteur pourra encore être relié à une application mobile qui restituera à l'utilisateur sous une forme visuelle et interactive certaines caractéristiques de ses courses : carte des trajets parcourus, graphique représentant les temps de trajet, les dénivelés réalisés ou la vitesse moyenne, etc.

L'amélioration du service proposé aux utilisateurs passe également par le recours aux outils de datavisualisation [35]. La datavisualisation est l'art de présenter les données sous forme visuelle, qu'il s'agisse de graphiques, de camemberts, de chronologies, de cartographies, d'infographies ou de toute autre forme de création graphique. Au-delà de l'aspect esthétique, l'enjeu est d'aider les utilisateurs ou tout autre destinataire à mieux comprendre des informations en apparence complexes ou qui contiennent de nombreux paramètres.

Les enjeux ne sont pas tellement différents si l'on se place cette fois du point de vue de l'administration. L'ouverture des données détenues par l'administration n'a de sens, en tant que telle, que si ces données sont suffisamment lisibles et compréhensibles pour les citoyens et si elles leur procurent un surcroît d'informations. Il faut encore que le format dans lequel elles sont restituées permette leur réutilisation par d'autres acteurs. La publication d'immenses tableaux en format pdf qui contiendraient pêle-mêle l'intégralité des données d'une administration ne présente à ce titre qu'un faible intérêt. Pour être efficace, toute politique d'ouverture des données publiques (*Open Data*) doit en principe s'accompagner d'une réflexion de l'administration portant à la fois sur la pertinence des informations qui sont mises à la disposition du public et sur la forme donnée à cette restitution.

Par exemple, dans le cadre de ses missions de régulation, l'ARCEP (Autorité de régulation des communications électroniques et des postes) a revu depuis quelques années de sa façon de s'adresser aux utilisateurs des télécoms et du secteur postal. En lieu et place de la publication de rapports longs et difficilement compréhensibles pour les non-experts, l'autorité a développé des outils qui prennent pour point de départ les préoccupations des utilisateurs. Le site Internet MonReseauMobile.fr propose ainsi aux visiteurs une carte interactive de la couverture mobile en France. Les utilisateurs peuvent désormais connaître et comparer

[35] Isaac Henri, « Données, valeur et *business model* », *op. cit.*

les performances des opérateurs français sur les différents points du territoire pour les appels, les sms et l'Internet mobile. Ils peuvent ensuite arbitrer leurs choix en fonction de ces informations et décider de changer d'opérateur.

Piloter la donnée au sein de l'organisation

Notons enfin que l'extraction de valeur à partir des données est loin de pouvoir se résumer à un ensemble d'opérations techniques. De nombreux travaux ont souligné que la valorisation des données au sein d'une organisation ne tient pas seulement à la collecte massive de données, à leur traitement par des algorithmes et à leur restitution sous forme de produits ou services proposés aux utilisateurs, mais qu'elle relève aussi de processus organisationnels spécifiques [36].

L'un des enjeux est de réussir à développer une véritable culture de la donnée et de son partage au sein de l'organisation. L'exploitation des données requiert en effet une collaboration cohérente entre les différents services d'une même organisation pour casser les silos dans le traitement des informations. Alors que certains métiers ont une culture de la donnée très développée, d'autres manquent de maturité sur le sujet, sans qu'ils soient pour autant moins concernés.

Compte tenu de la volumétrie croissante des données, il est indispensable de penser ensemble les mécanismes de création de valeur, le modèle d'affaires et les capacités techniques de l'organisation. Par conséquent, à côté des investissements en infrastructures informatiques, la création de valeur à partir des données nécessite des investissements en termes d'organisation et de compétences afin de mettre en œuvre une véritable gouvernance des données au sein de l'organisation.

Cette gouvernance se répartit en trois niveaux [37]. Tout d'abord, la gouvernance des données ne doit pas être considérée comme une sous-catégorie de la gouvernance des systèmes d'information. Pour qu'une démarche de valorisation des données se mette en place dans une organisation, elle doit être rattachée à la gouvernance globale de l'entreprise au plus haut niveau. C'est à ce premier niveau que les données doivent être considérées comme un actif stratégique et qu'il convient de définir la politique de gestion des données et les principes de leur intégration dans la chaîne de valeur de l'organisation. Il appartient ensuite

[36] Isaac Henri, « Données, valeur et *business model* », *op. cit.*
[37] Isaac Henri, « Données, valeur et *business model* », *op. cit.*

au niveau managérial de définir le processus de gestion des données (la propriété des données, leurs mécanismes de gestion et de contrôle). Enfin, le dernier niveau est en charge des aspects techniques de la gestion des données.

La coordination de ces trois niveaux peut éventuellement être confiée à un nouvel acteur, le *chief data officer*, qui aura non seulement la responsabilité de la gouvernance et de la stratégie des données, mais qui sera également chargé de diffuser la culture de la donnée à tous les niveaux de l'organisation. Cependant, le *chief data officer* ne peut rien faire tout seul. Pour que son action soit efficace, il doit pouvoir s'appuyer sur des interlocuteurs métiers au sein de l'organisation, voire, en fonction de la taille de l'organisation, sur un département dédié, avec des *data scientists* qui pourront faciliter la diffusion des pratiques centrées sur la donnée dans l'organisation [38].

[38] CIGREF, *Valorisation des données dans les grandes entreprises. Maturité, pratiques et modèles*, rapport, novembre 2016, p. 16 et s.

CHAPITRE 2

LA DONNÉE AU CŒUR DE LA TRANSFORMATION DES ORGANISATIONS

En réorganisant son activité autour de l'utilisation des données, une organisation peut espérer améliorer sa productivité et ses performances économiques. L'exploitation des données permet également aux entreprises de déployer leurs activités dans de nouveaux secteurs. Mais la révolution industrielle portée par la donnée ne se limite pas à la transformation des organisations et de leur modèle d'affaires. Elle donne lieu à l'apparition d'un nouvel écosystème économique, qui se caractérise par des interactions et des coopérations entre des acteurs économiques situés à différents niveaux de la chaîne de valeur des données. Au sein de cet écosystème, la question de la circulation des données devient un enjeu économique et politique crucial.

1_ Un actif stratégique pour l'organisation

La donnée est un actif stratégique autour duquel une organisation peut réorganiser son activité et déployer de nouveaux modèles d'affaires. Les modalités de collecte et de traitement peuvent différer en fonction de l'activité en cause, mais plusieurs tendances sont observables. L'exploitation des données permet tout d'abord aux organisations d'améliorer leur offre de produits et services, en proposant à leurs publics des services enrichis et personnalisés. Les capacités d'analyse offertes par la collecte massive de données sont aussi la source de gains de productivité : elles permettent une aide à la décision de plus en plus intégrée en continu à la production et une optimisation des produits et des services. Enfin, au-delà de l'utilisation des données pour consolider une activité existante, certains modèles d'affaires reposent spécifiquement sur l'exploitation des données (*data-driven business models*).

Vers l'amélioration de la relation client et usager

Les données générées par un utilisateur peuvent servir à personnaliser le service qui lui est offert. La prise en compte des pratiques de l'utilisateur permet de lui faire des recommandations adaptées à ses centres d'intérêt. La personnalisation du service est un facteur de fidélisation des utilisateurs : plus un service collecte de données sur ses utilisateurs à partir de leurs usages, plus il est en mesure de leur proposer un service personnalisé et adapté à leurs attentes.

La personnalisation du service

Les exemples de fourniture de services personnalisés sont légion dans l'économie numérique. L'algorithme PageRank de Google propose par exemple à ses utilisateurs des résultats personnalisés en tenant compte de l'historique de leurs recherches. En exploitant la liste des morceaux écoutés par ses utilisateurs, le service de streaming musical Spotify est en mesure de créer des playlists personnalisées. Facebook sélectionne les publications qui apparaissent sur le fil d'actualité d'un utilisateur en fonction de ses activités, de la liste de ses contacts, du nombre de mentions « j'aime » et de commentaires suscités par une publication. L'assistant vocal Siri d'Apple intègre des algorithmes d'apprentissage qui permettent son adaptation progressive à la diction des utilisateurs.

De même, les programmes de VOD ont fait le choix d'une offre de services personnalisés à destination d'audiences spécifiques.

Les algorithmes peuvent pousser la personnalisation du service jusqu'à devancer les besoins des utilisateurs. L'entreprise de grande distribution Target estime ainsi les probabilités de grossesse de ses clientes et fait parvenir des coupons de produits pour bébés aux consommatrices qui ont les probabilités les plus élevées de débuter une grossesse.

La singularisation des prix

La différenciation des prix en fonction des prévisions de demande et du profil des consommateurs est née dans les années 1980 aux États-Unis dans le secteur du transport aérien et de l'hôtellerie (*yield management*). Le développement du commerce sur Internet, conjugué aux évolutions technologiques (traçage des adresses, cookies, augmentation du volume de données disponibles), permet désormais d'obtenir en temps réel des informations sur les caractéristiques et le comportement des clients [39]. Autant d'informations à partir desquelles il est possible d'estimer le « consentement à payer » des consommateurs. Dès lors, une entreprise est à même de proposer à ses clients des tarifs adaptés à leur portefeuille.

La personnalisation des prix est une pratique relativement peu documentée, mais, d'après une étude de 2018 du cabinet Deloitte portant sur 500 entreprises, 40 % des détaillants ayant adopté des outils d'intelligence artificielle pour individualiser l'expérience du consommateur personnalisent leurs prix et leurs offres promotionnelles en temps réel [40]. Parmi les exemples observés, la compagnie aérienne AirAsia Berhad recourt depuis 2017 à l'apprentissage automatique pour déterminer ce que ses clients sont prêts à payer pour un bagage.

Aux États-Unis, les compagnies d'assurance ont commencé à proposer à leurs assurés des tarifs modulables en fonction de leurs comportements. Les contrats dits "*pay as you drive*" adaptent le montant de la prime d'assurance au nombre de kilomètres effectivement parcourus et aux conditions de conduite (le week-end, la nuit, à la ville ou à la campagne). Pour ce faire, l'assuré doit installer dans son véhicule un boîtier de géolocalisation qui collecte pour le compte de la compagnie d'assurance

[39] OCDE, *La personnalisation des prix à l'ère numérique*, note de référence, 28 novembre 2018.

[40] Deloitte Digital, *Consumer Experience in the Retail Renaissance: How Leading Brands Build a Bedrock with Data*, 2018.

une masse d'informations. Un cran plus loin dans la personnalisation des contrats d'assurance, on trouve les contrats qualifiés de "*pay how you drive*". Outre le kilométrage et les trajets, l'assureur enregistre la vitesse de conduite, l'allure dans les virages, le freinage ou encore l'accélération pour déterminer le montant de la prime d'assurance. Les assurances proposent une réduction de 3 % aux conducteurs qui acceptent d'utiliser une application téléphonique qui détecte leurs comportements : par exemple si l'automobiliste tient son téléphone à la main ou utilise une connexion Bluetooth pour téléphoner [41].

Toujours aux États-Unis, les assureurs se sont aussi lancés dans la collecte de données sur leurs assurés à travers des programmes de coaching en ligne. Il est alors possible d'imaginer qu'à l'avenir les compagnies d'assurance s'emparent des données collectées par les objets connectés pour proposer des tarifs adaptés au comportement de l'assuré, sur la base d'une anticipation individualisée des risques. Les assurés, dont les comportements vertueux seront attestés par toute une panoplie d'objets connectés (podomètre, balance, appareil d'analyse du sommeil), pourront alors bénéficier de primes d'assurance avantageuses ou de réductions (et inversement !).

L'optimisation du métier et du produit

Les données constituent un précieux outil d'aide à la décision. En permettant aux organisations d'obtenir un retour sur leur activité quasiment en temps réel et d'anticiper des évènements, l'analyse des données permet l'adoption de décisions mieux informées. Cela peut se traduire par l'amélioration de la qualité des services et des produits qu'elles proposent ou encore par l'optimisation de leurs processus de production. Vue sous cet angle, l'économie fondée sur la donnée participe moins au développement de nouveaux services ou de nouvelles activités qu'elle n'offre de nouveaux outils pour améliorer des activités traditionnelles et permettre aux organisations de réaliser d'importants gains de productivité.

Les organisations peuvent utiliser les données massives pour améliorer l'efficacité de leur système d'information dans le cadre de la mise en place de *customer relationship management* (gestion de relation client) et d'*entreprise ressource planning* (progiciel de gestion intégrée).

[41] Quain John, "Cars Suck Up Data About You. Where Does It All Go?", *New York Times*, 27 juillet 2017.

Le concept de *customer relationship management* remonte aux années 1970. Le numérique permet désormais l'intégration et l'analyse en continu des données des clients ou usagers issues de différentes sources (historique d'achat, profilage marketing, service client, réseaux sociaux, questionnaires en ligne, etc.) pour aider à la prise de décision dans l'organisation. À partir de ces données, une organisation est en mesure d'établir le profilage de sa clientèle ou de ses usagers et de lancer des opérations marketing adaptées (par exemple, identifier les clients potentiels, proposer des offres ciblées, trouver le bon moment pour relancer un client par email) ou encore développer de nouveaux services ou de nouveaux produits adaptés aux attentes des consommateurs ou usagers.

En suivant l'activité des utilisateurs, une organisation peut mesurer avec précision ses performances et décider d'apporter des améliorations et des corrections ciblées à ses services. La communication en ligne permet ainsi de pratiquer le « test A/B » à grande échelle. La technique consiste à proposer plusieurs variantes d'une page web, d'une application mobile, d'une bannière publicitaire ou d'un e-mail à différents groupes d'utilisateurs afin de déterminer celle qui rencontre le plus de succès. À partir des résultats obtenus, l'amélioration du service peut se faire en continu. Ainsi, le test A/B a été mis en œuvre sur les pages d'appel aux dons de la campagne américaine de 2012 par l'équipe de Barack Obama. Il a permis d'augmenter sensiblement la conversion du nombre de visiteurs des sites Internet du candidat en souscripteurs.

Dans une optique de production participative (*crowdsourcing*), les données issues d'utilisateurs peuvent par ailleurs servir à améliorer la qualité du service proposé à d'autres utilisateurs. Les notes et les commentaires laissés par les utilisateurs sur les plateformes numériques permettent directement de guider le choix des futurs utilisateurs, que ce soit vis-à-vis d'un produit, d'une application, d'un vendeur, d'un chauffeur ou d'un logement. Une entreprise en ligne peut elle aussi s'appuyer sur les informations générées par les utilisateurs pour établir ses choix éditoriaux et mettre en avant les offres les plus susceptibles de répondre aux attentes de ses clients. Parfois, ce choix éditorial est personnalisé : les données de comportement d'un utilisateur sont alors utilisées pour formuler des recommandations aux utilisateurs qui présentent des caractéristiques similaires. Une fois indiqué que nous voyageons en couple, un site de réservation d'hébergement en ligne nous proposera par exemple les chambres qui ont obtenu le plus haut taux de satisfaction auprès des couples. Après un premier achat, la plateforme de vente affichera les autres objets qui ont également été achetés par les acheteurs qui ont fait le même achat.

Le secteur industriel n'est pas en reste. D'après un rapport de l'OCDE, « les données occuperont une place centrale dans la production industrielle du XXIe siècle »[42]. La transformation numérique de la production, désignée par les expressions « usine du futur » ou « industrie 4.0 », devrait à l'horizon 2030 très largement prendre appui sur une exploitation intensive des données.

L'industrie 4.0 et la prochaine révolution de la production

Extrait du rapport de l'OCDE : La prochaine révolution de la production : Conséquences pour les pouvoirs publics et les entreprises, éditions OCDE, 2018, p. 23

L'expression « quatrième révolution industrielle », également appelée « industrie 4.0 », renvoie à l'utilisation, dans la production industrielle, des technologies numériques récentes et souvent interdépendantes pour concevoir des procédés ou améliorer ceux existants, conduisant parfois à l'apparition de nouveaux biens et services. Les technologies concernées sont nombreuses et diverses : programmes d'apprentissage automatique ou applications de la science des données débouchant sur des systèmes de plus en plus autonomes et intelligents, capteurs à faibles coûts utilisés dans l'Internet des Objets, ou encore dispositifs de contrôle novateurs rendant possible l'arrivée des robots industriels de deuxième génération.

L'expression « Industrie 4.0 » établit un contraste avec les trois précédentes révolutions industrielles, qu'il est difficile de situer avec précision dans le temps : i) la mécanisation de la production, avec l'apparition de la machine à vapeur (autour de 1780), ii) le développement de la production de masse grâce à l'énergie électrique (années 1870), et iii) l'automatisation de la production grâce à l'électronique (années 1960). L'électronique de l'industrie 4.0 diffère à maints égards, notamment en termes de coût, de taille, de puissance de calcul, d'intelligence et d'interconnectivité ainsi que par l'intégration d'objets physiques.

L'utilisation de capteurs et de machines connectées et le recours au *cloud computing* rendent la collecte, le stockage et le traitement des données industrielles de plus en plus simples. Par ailleurs, couplée à l'analyse des données et l'apprentissage automatique, l'interconnexion des machines et des systèmes d'informations sur les sites de production, mais aussi à

[42] OCDE, *La prochaine révolution de la production : Conséquences pour les pouvoirs publics et les entreprises*, Éditions OCDE, 2018.

l'extérieur des sites de production (par exemple pour intégrer les données de clientèle, les données des partenaires, les données de suivi de produits en sortie d'usine), devrait permettre d'obtenir des machines et des systèmes de plus en plus intelligents et autonomes. De multiples gains de productivité sont attendus à tous les stades de la chaîne d'activité, sans exhaustivité :

- L'élimination des erreurs dans certains processus de production : les capteurs rendant possible la surveillance de chaque élément, il devient inutile de réaliser des tests sur des échantillons de chaque lot.
- La maintenance prédictive : les machines connectées communiquent en temps réel des données relatives à leur fonctionnement sur un cloud pour qu'elles puissent être analysées. À partir de ces données et en exploitant l'historique des défaillances, des algorithmes de *machine learning* apprennent à reconnaître les signes avant-coureurs de pannes. Il est alors possible d'anticiper les moments où une panne pourrait intervenir et de procéder aux réparations en temps voulu, de manière à réduire les coûts et la durée d'immobilisation des machines.
- L'organisation de la production pour répondre à la demande réelle : les logiciels de commande ou de gestion de stocks interagissent directement avec les chaînes de production, rendant possible une gestion des stocks en temps réel et l'automatisation du processus de commande auprès des fournisseurs.
- L'optimisation des produits : avec les logiciels dits « génératifs » qui utilisent l'IA, il devient possible de créer des modèles industriels optimaux en matière de poids et de résistance. Le logiciel Dreamcatcher est par exemple à l'origine de la carrosserie de la moto électrique la plus rapide du monde, Lightning, et de la création d'une cloison d'avion 50 % plus légère que les modèles antérieurs.
- L'automation de la logistique : les robots de la société Kiva permettent de traiter quatre fois plus de commandes dans les entrepôts qui en sont équipés [43]. Alors que dans les entrepôts qui n'ont pas été automatisés les travailleurs peuvent passer jusqu'à 70 % de leur temps de travail à marcher pour récupérer les produits commandés, les petits robots orange roulants soulèvent les étagères pour les apporter aux préparateurs de commandes.

[43] Rotman David, "How technology is destroying jobs", MIT Technology Review, juin 2013, www.technologyreview.com/insider/pricing

- La réalisation d'économies d'énergie : en 2016, Google annonçait une réduction de 40 % de la consommation d'énergie nécessaire au refroidissement de ses centres de données grâce au recours au *machine learning* [44]. D'après une étude de Citigroup - Oxford Martin School, les camions autonomes pourraient accroître la production dans l'industrie minière de 15 % à 20 % et réduire la consommation de carburant de 10 % à 15 % [45].

De nouveaux modèles d'affaires et de services

Les données se trouvent au centre de nouveaux modèles d'affaires spécifiquement fondés sur leur exploitation. Il ne s'agit alors plus seulement d'utiliser les données pour améliorer un service ou un produit préexistant ou gagner en productivité. Dans ces modèles d'affaires, la donnée représente une ressource clé, consubstantielle à l'activité de l'entreprise [46].

De nouvelles activités se développent, qui sont entièrement fondées sur la donnée. Ainsi, dans le secteur de la publicité en ligne, les *third party data providers* sont des entreprises spécialisées dans la collecte, le traitement et la qualification de données comportementales ou déclaratives, qui sont ensuite fournies aux annonceurs dans le cadre du ciblage publicitaire.

[44] https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/

[45] Citigroup - Oxford Martin School, *Technology at Work: The Future of Innovation and Employment*, University of Oxford, février 2015.

[46] Hartmann Philipp Max et al., "Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms", working paper, University of Cambridge.

L'usine du futur,
connectée et pilotée par les données

Achats

L'entreprise échange avec ses partenaires par le biais de plateformes collaboratives et d'intermédiation.



Logistique



Les aléas de la logistique sont absorbés car communiqués en temps réel à la chaîne qui se réorganise automatiquement.

Conception

L'intégration numérique de la conception et de la fabrication permet la continuité de l'information entre les deux domaines.

① Les arrêts non planifiés sont pris en compte instantanément et la planification se réajuste automatiquement.

② Les opérateurs, connectés et mobiles, reçoivent une information personnalisée.

③ Le traçage de toutes les pièces permet un contrôle qualité en continu sur toute la chaîne.



Finance

La gestion de la production s'interface avec l'ERP.

Services commerciaux relation client



La chaîne industrielle est pilotée par la demande.

Distribution



La dématérialisation de la chaîne industrielle permet de produire localement, à proximité du client.

④ Les flux physiques sont gérés de manière automatique et optimisée.

⑤ La production est optimisée grâce à l'analyse en temps réel de l'ensemble des données. collectées.

⑥ Les opérateurs et les robots collaborent.

Source : Syntec Numérique, Livre Blanc 2016, *Transformer l'industrie par le numérique*

Les données figurent également au cœur des services d'informatique en nuage (*cloud computing*). Le cloud computing désigne un modèle de services informatiques qui repose sur le partage de ressources informatiques accessibles sur demande, généralement par Internet, et qui est associé à une grande souplesse d'utilisation et à de faibles coûts de maintenance. Au lieu d'obtenir de la puissance de calcul par l'achat de matériel et de logiciel, le client paye un fournisseur de services de cloud pour accéder à sa puissance de calcul depuis n'importe quel lieu. La facturation se fait à l'utilisation : le fournisseur de services garde à sa charge les investissements financiers matériels et logiciels et les frais de maintenance. Le *cloud computing* comprend trois catégories de services : l'infrastructure en tant que service (Iaas), la plateforme en tant que service (PaaS) et le logiciel en tant que service (SaaS). Les entreprises de *cloud computing* peuvent proposer une grande variété de services à leurs utilisateurs, tels que le stockage et l'analyse de données, des solutions logicielles ou des services de *machine learning*. Le modèle de facturation à l'usage et le fait que l'utilisation des services ne nécessite pas d'importants investissements de départs (matériel et logiciel) contribuent à démocratiser l'accès à des services de traitement de données auprès des PME et des administrations.

Enfin, des modèles d'affaires fondés sur la donnée peuvent venir compléter l'activité principale de l'entreprise. Une entreprise qui exerce son activité dans un secteur économique traditionnel peut exploiter les données qu'elle y collecte pour pénétrer sur de nouveaux marchés, lesquels sont parfois très éloignés de son cœur de métier. Ainsi, l'entreprise Elis, leader européen du nettoyage et de la location d'articles textiles, dispose de données issues de ses centres de nettoyage qui lui permettent de connaître avec précision le nombre de nuitées dans l'hôtellerie. La société peut commercialiser les données qu'elle détient sur l'occupation des hôtels par ville et par quartier auprès de l'industrie hôtelière, complétant son cœur de métier par une prestation de service numérique.

De son côté, l'Internet des Objets conduit à un rapprochement notable des secteurs de l'industrie et des services, parfois qualifié de « servicialisation ». Avec l'IoT, ce ne sont désormais plus seulement des objets qui sont proposés par l'industrie manufacturière. Les données générées par les objets connectés et leur transmission en temps réel sont à la base de nouveaux services facturés à l'utilisateur par le fabricant : un réfrigérateur qui propose des recettes correspondant aux produits qui s'y trouvent, des machines agricoles qui recueillent des données relatives à l'exploitation et les restituent à l'agriculteur sous forme de services

de gestion, mais également des services de maintenance prédictive qui permettent d'anticiper les pannes de l'appareil. Cette transformation vient profondément redessiner les contours de l'industrie.

Après le téléphone portable, la voiture connectée illustre la manière dont un produit traditionnel peut devenir pour son fabricant le support de nouvelles sources de valeur. Selon une étude du cabinet Frost & Sullivan, le marché de la donnée automobile pourrait s'élever à 33 milliards de dollars d'ici 2025. Peut-être demain les constructeurs automobiles tireront-ils davantage de revenus de la commercialisation des données de leurs véhicules que de la vente des véhicules elle-même.

Une première évolution du secteur liée à la connectivité croissante des véhicules consiste dans l'intégration d'une interface servicielle aux voitures, comprenant par exemple des fonctionnalités de commerce électronique. Fiat Chrysler Automobile propose ainsi à ses automobilistes un service « Market ». En cliquant sur l'écran de son tableau de bord, le conducteur a alors la possibilité de commander un café ou une pizza à un commerçant préalablement inscrit. Le système propose ensuite un guidage GPS jusqu'à la boutique et permet de payer directement depuis le véhicule.

Les constructeurs cherchent par ailleurs à monétiser les quantités de données de plus en plus importantes produites par les voitures connectées. Par exemple, BMW a lancé sa place de marché de données appelée BMW CarData, qui permet à des prestataires de services extérieurs à l'entreprise (compagnies d'assurance, garagistes, gestionnaire de parcs automobiles) d'accéder aux données provenant des véhicules connectés contre rémunération. Le constructeur allemand Daimler s'est quant à lui associé avec l'entreprise israélienne Otonomo pour proposer une plateforme de partage des données des véhicules. Otonomo propose au propriétaire du véhicule de partager ses informations avec d'autres entreprises, fournisseurs de services ou développeurs d'applications, en échange de rabais ou de récompenses. Par l'intermédiaire de la plateforme, le conducteur peut par exemple souscrire à un système d'assurance individualisé correspondant au nombre de kilomètres parcourus.

Ainsi, les frontières entre le développement de logiciels, la fabrication d'appareils et la prestation de services se brouillent et ces activités se rejoignent progressivement.

Les modes de rémunération dans l'écosystème numérique

Source : OCDE, Data-Driven Innovation: Big Data for Growth and Well-Being, 2015

Lorsqu'elle se développe dans l'écosystème numérique, une entreprise a différents moyens de se rémunérer.

Le freemium : Il s'agit d'un des modèles de rémunération les plus développés chez les startups. Une entreprise propose gratuitement à ses clients certains services de base, mais elle facture leur amélioration ou l'ajout de nouveaux services (premium). Souvent, ce mode de rémunération est combiné avec une rémunération tirée de la publicité.

L'abonnement : L'abonnement est particulièrement fréquent dans les relations *Business to Business* (B2B). L'utilisateur paie régulièrement l'entreprise, qui lui offre en contrepartie accès à ses services.

La publicité : Elle est très utilisée, notamment dans le cadre des relations *Business to Consumer* (B2C). Une entreprise propose gratuitement ses services aux utilisateurs, en échange de quoi elle intègre des annonces publicitaires qui lui permettent de se rémunérer.

La redevance : La redevance se retrouve dans beaucoup d'offres de *cloud computing*. L'utilisateur paie l'entreprise à hauteur de l'utilisation qu'il fait de ses services.

La vente à l'unité : Ce mode de rémunération traditionnel ne disparaît pas dans l'économie de la donnée. Un client peut payer pour accéder au téléchargement de contenus, qu'il s'agisse de livres numériques, de morceaux de musique, d'applications, de jeux vidéos ou de films.

La vente de services : La vente de services se retrouve dans de nombreuses relations B2B (maintenance, consultation, développement informatique, hébergement de site Internet).

Les frais de commission : Les frais de commission se retrouvent fréquemment dans les relations B2C, B2B et B2B2C. Les plateformes qui font office de places de marché prélèvent un pourcentage sur les transactions conclues par leur intermédiaire entre vendeurs et acheteurs.

2_ Les nouveaux écosystèmes des données

La circulation des données peut être porteuse de gains économiques, politiques, et constituer une source d'innovation bénéfique à la société. Cette circulation est d'autant plus nécessaire que les acteurs en capacité d'extraire de la valeur à partir des données ne sont pas nécessairement ceux qui les détiennent. L'opération de valorisation des données est le plus souvent complexe et nécessite l'intervention de différentes catégories d'acteurs répartis le long de la chaîne de valeur : les fournisseurs de données, les services d'analyse de données, les utilisateurs qui se servent des résultats de l'analyse des données dans leur activité et les consommateurs ou usagers finaux. L'économie fondée sur les données donne lieu à l'émergence d'un écosystème économique à l'intérieur duquel apparaissent de nouveaux modèles d'affaires et des formes originales de collaboration intra et inter-secteurs, privés ou publics.

Les enjeux de la circulation des données

La circulation des données devient un enjeu économique et politique, jusqu'au consommateur et au citoyen.

D'abord, en raison du caractère non rival des données, l'utilisation d'une donnée par un agent ne réduit aucunement sa disponibilité pour les autres agents. Plusieurs agents peuvent donc faire un usage simultané des mêmes données. S'ajoutent à cela les faibles coûts liés au stockage et à la circulation des données numériques. À l'époque des technologies analogiques, copier et réutiliser des données étaient encore des opérations coûteuses. Le passage à l'ère numérique change complètement la situation, en réduisant drastiquement les coûts et en simplifiant le transfert de données.

Deuxièmement, les données voient leur valeur et leurs possibilités d'usage démultipliées lorsqu'elles sont partagées et croisées avec d'autres jeux de données [47]. Pour reprendre l'image du sociologue Bruno Marzloff, « la valeur d'une donnée est proportionnelle au carré du nombre de données auxquelles elle est associée » [48]. C'est donc l'association de différents

[47] Duchesne Claudine, Cytermann Laurent, Vachey et al. *Rapport relatif aux données d'intérêt général*, rapport du Conseil général de l'économie, de l'industrie, de l'énergie et des technologies, du Conseil d'État et de l'Inspection général des finances, septembre 2015 ; OCDE, Data-Driven Innovation. *Big Data for Growth and Well-Being, op. cit.*

[48] Marzloff Bruno, *Sans bureau fixe*, FYP éditions, 2013.

jeux de données, des données internes et externes à une organisation, qui permet d'en révéler tout le potentiel exploitable.

De plus, l'ouverture des données rend possible leur réutilisation par de nouveaux acteurs et pour des finalités qui n'ont pas nécessairement été prévues au moment où les données ont été collectées. Par exemple, l'analyse massive de données de santé recueillies dans le cadre de soins, avec l'accord des patients concernés, permet de développer des logiciels d'aide à la décision médicale. Les chances de réutilisation augmentent donc en proportion du périmètre d'ouverture des données. C'est la raison pour laquelle la circulation des données entre différentes organisations et entre différents secteurs d'activité devrait généralement être favorisée.

À l'inverse, une sous-utilisation des données n'engendre pas de gains économiques et sociaux. Si les données sont retenues dans des silos, une part importante de leur valeur risque de ne jamais en être extraite. À ce titre, l'interopérabilité des données joue un rôle clé dans la circulation des données.

Cependant, il n'est pas facile de mesurer précisément les retombées positives des données[49]. Non seulement la valeur économique des données reste difficile à appréhender, mais quantifier leur influence positive sur les échanges scientifiques et culturels, sur les changements sociaux ou sur la confiance est une opération plus délicate encore. C'est pourquoi l'OCDE craint que les difficultés à saisir l'ensemble des externalités positives liées aux données ne se traduisent en termes de sous-investissements en matière de données et d'analyse de données et ne conduisent à une situation dans laquelle l'accès aux données est socialement sous-optimal.

Pour autant, l'ouverture des données peut connaître des limites et leur accès ne doit pas nécessairement être libre et gratuit. Tous les travaux portant sur le sujet soulignent l'importance de trouver un juste équilibre entre les bénéfices économiques et sociaux de l'ouverture des données et la protection de l'intérêt légitime de leurs détenteurs. Ces derniers peuvent avoir de bonnes raisons de vouloir conserver pour eux certaines données, dont certaines sont d'ores et déjà consacrées par le droit : sécurité, protection de la vie privée, du secret des affaires ou des droits de propriété intellectuelle. En matière d'ouverture, l'alternative ne se réduit pas à une opposition binaire entre « données ouvertes » et « données fermées ». Entre les deux termes, il existe différents degrés d'ouverture possibles. Toute la question est de savoir où placer le curseur. Une organisation peut

[49] OCDE, *Data-Driven Innovation. Big Data for Growth and Well-Being, op. cit.*

choisir de rendre ses données accessibles gratuitement, d'autres peuvent décider de les vendre. L'accès aux données peut être limité à certaines personnes, il peut au contraire ne connaître aucune restriction, etc.

Le niveau d'ouverture adapté dépend chaque fois du secteur d'activité concerné et du degré de maturité de son écosystème. Préalablement à toute ouverture de ses données, une organisation doit se poser certaines questions : quelles sont les données concernées par l'ouverture ? Qui peut avoir accès à mes données ? Quelles sont les conditions d'accès et de réutilisation des données ?

Quelles sont les données concernées ?

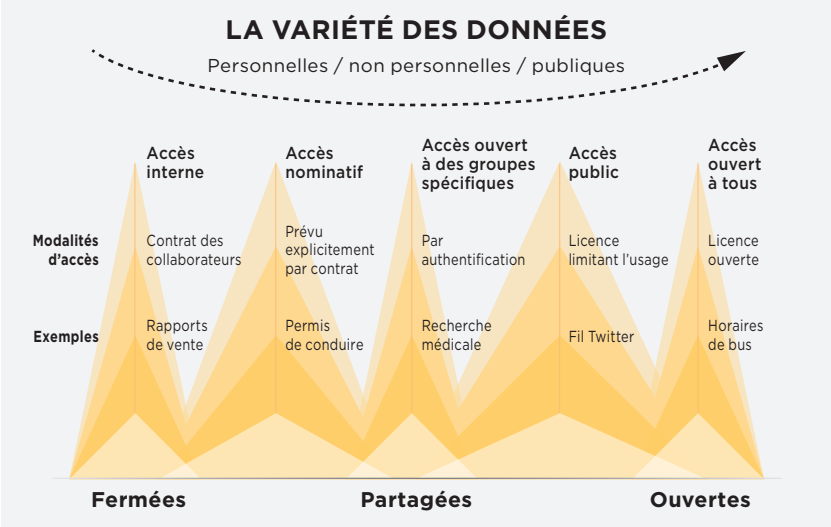
L'ouverture peut être limitée à certains jeux de données. Toutes les données n'ont pas la même valeur stratégique au sein d'une organisation. Certaines peuvent être considérées comme accessoires par rapport à ce que l'organisation considère comme le centre de son activité ; d'autres, au contraire, touchent au cœur de la valeur créée par l'organisation. De plus, une organisation peut n'exploiter qu'une partie des données en sa possession, laissant en jachère une partie d'entre elles. Il appartient à l'organisation de dresser l'inventaire des données qu'elle détient et d'engager une réflexion au cas par cas sur les bénéfices qu'elle peut attendre de leur ouverture, au-delà des données publiques dont l'ouverture est prévue par la loi.

Qui a accès aux données ?

Ouvrir ses données ne signifie pas nécessairement y donner un accès libre à tous, excepté pour certains acteurs concernés par des obligations d'ouverture. Une organisation est libre de déterminer le degré d'ouverture qu'elle souhaite conférer à ses données : ne pas ouvrir ses données, les partager avec des organisations partenaires, mutualiser ses données avec ses concurrents sur une même plateforme, les partager avec les autorités publiques ou encore les rendre accessibles au public. Tout dépend des objectifs qui sont recherchés par l'organisation. La mutualisation des données de filière permet par exemple à des entreprises concurrentes d'enrichir la connaissance respective qu'elles ont de leur marché tout en partageant les coûts d'analyse. C'est dans un objectif de transparence que Nike a rendu publiques des informations sur sa chaîne d'approvisionnement, à la suite du scandale lié au travail des enfants dans les années 1990. De son côté, Airbnb a créé le portail Dataville qui permet « aux municipalités de mieux suivre le développement de l'activité touristique via Airbnb

sur leur commune, son impact positif sur l'attractivité de leur commune comme sur le pouvoir d'achat de leurs administrés », comme l'explique l'entreprise dans le communiqué de lancement du portail. La loi peut par ailleurs imposer la communication de certains jeux de données détenus par des acteurs privés dans le cadre de certaines délégations de services publics.

Un certain nombre d'entreprises de *cloud computing* proposent des services de partage de données qui permettent à leurs clients de rendre leurs données publiques, ou de les partager avec des partenaires spécifiques. Ces solutions permettent également de partager des données au sein d'une même organisation.



Source : Open Data Institute, « The Data Spectrum »

Quelles sont les conditions d'accès et de réutilisation des données ?

Une fois qu'elle a décidé d'ouvrir une partie de ses données, une organisation peut encore fixer les conditions auxquelles les tiers pourront y accéder. L'accès aux données peut être gratuit ou payant.

Le marché des données reste embryonnaire. S'il existe quelques places de marchés comme Infochimp, Datamarket, Factual ou Microsoft Azure, ou des plateformes spécialisées dans des secteurs intégrés comme la

mobilité, l'agriculture ou la logistique, la plupart des données ne sont pas directement commercialisées sur un marché qui organise la rencontre de l'offre et de la demande. Elles sont souvent exploitées en interne ou dans le cadre d'accords de partenariat. Elles fournissent un modèle économique qui rend accessibles des données qu'il serait autrement difficile d'échanger. De plus, elles contribuent à rendre les données directement utilisables par des tiers, en prenant en charge les opérations de nettoyage et de formatage, et en les complétant par des métadonnées (par exemple, des indicateurs de qualité).

Certains considèrent que l'échange de données contre de l'argent est un modèle en perte de vitesse. Comme le soulignent Simon Chignard et Louis-David Benyayer, « Seuls les producteurs les plus puissants, ou ceux qui sont capables de garantir l'exclusivité de leur service sont en mesure de maintenir durablement leur position sur le marché des données » [50]. Thésauriser les données dans l'attente d'en tirer un prix élevé est une stratégie improductive et inadaptée aux spécificités de l'économie numérique.

Des modèles rivaux à l'échange monétaire existent :

- Les données peuvent tout d'abord être échangées contre d'autres données : l'accès à une base de données a alors pour contrepartie l'enrichissement de cette même base par son utilisateur.
- Les données peuvent être échangées contre des services gratuits. Ce modèle est à la base des modèles économiques bifaces adoptés par des entreprises comme Google ou Facebook. Elles fournissent des services gratuits à leurs utilisateurs, en échange de la collecte de leurs données personnelles.
- Un autre modèle d'échange consiste à ne plus attacher de valeur pécuniaire aux données brutes mais aux services issus des données. Par exemple, la société Avuxi rassemble l'ensemble des avis laissés par les touristes dans une ville pour identifier les hôtels et les restaurants les plus populaires. Ces informations sont ensuite revendues aux voyagistes et aux portails de réservation.

Après avoir fixé les conditions d'accès à ses données, une organisation peut encore en déterminer les conditions d'utilisation. Elle peut par exemple interdire certains usages de ses données ou empêcher leur diffusion auprès de tiers.

[50] Chignard Simon et Benyayer Louis-David, *Datanomics, les nouveaux business models des données, op. cit.*, p. 59.

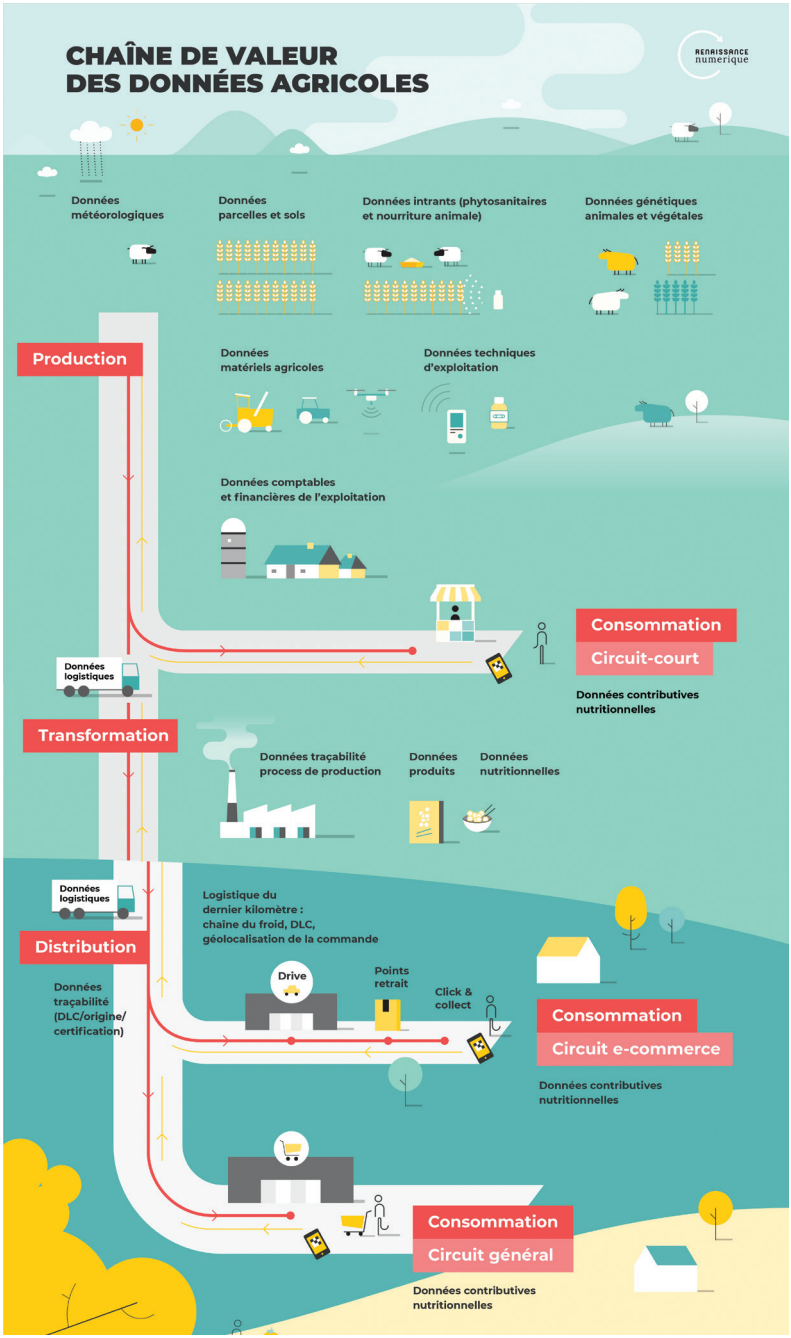
L'émergence de nouvelles modalités d'interaction entre acteurs : vers la « coopétition »

L'écosystème des données donne lieu à des interactions nombreuses entre des acteurs qui se positionnent à différents niveaux de la chaîne de valeur : fabricants de logiciels et de matériel, producteurs de données, prestataires de services de traitement des données, fournisseurs de produits et de services qui intègrent les données dans leur activité, utilisateurs finaux. Les acteurs occupent parfois plusieurs niveaux à la fois de la chaîne de valeur.

L'activité des organisations génère, par exemple, des données qui peuvent servir à améliorer la qualité du service proposé et au développement de compétences en matière d'analyse de données. Mais ces données peuvent aussi être revendues sous forme de services à d'autres organisations. Ainsi, les machines agricoles du fabricant John Deere collectent des données. Ces dernières sont exploitées par la société, qui les restitue à ses clients sous forme de services de gestion agricole. Dans le même temps, ces données intéressent aussi les entreprises de biotechnologie, les compagnies d'assurance et les marchés des matières premières, qui sont prêts à payer pour se les procurer.

Au sein de cet écosystème, l'exploitation des données massives prend la forme d'un mélange de compétition et de coopération entre les différents acteurs, phénomène qui est parfois qualifié de « coopétition ». Parce qu'une entreprise n'a pas forcément l'expérience nécessaire ou les compétences techniques requises pour traiter et analyser les données qu'elle détient, elle va chercher à s'attacher les services d'une autre entreprise qui dispose de ces compétences, mais qui, elle, ne possède pas nécessairement les bonnes données, ou qui ne serait pas en mesure de restituer la valeur de ces données sous forme de produits ou de services. Cette association peut prendre la forme d'un accord de partenariat, de l'achat d'une prestation de services ou encore de l'acquisition d'une entreprise par une autre.

L'écosystème numérique évolue rapidement sous l'effet conjugué des innovations technologiques et de l'arrivée de nouveaux acteurs. Comme dans d'autres secteurs, les entreprises les mieux établies ne cherchent pas seulement à consolider leur position en établissant des partenariats ou en proposant des produits innovants, mais également en rachetant d'autres entreprises et les startups les plus prometteuses. Pour elles, c'est une façon de proposer de nouveaux services, d'obtenir de nouveaux brevets, d'éliminer un concurrent potentiel ou encore d'empêcher qu'un de leur concurrent ne s'en empare avant eux. Les collaborations sont un



Source : Renaissance Numérique, *La valeur des données agricultures*, février 2018

autre moyen pour les entreprises du secteur de renforcer leur position au sein de l'écosystème. Ces partenariats laissent par ailleurs apparaître une convergence entre l'économie numérique et d'autres secteurs d'activité.

Les *Application programming interfaces* (API)

En permettant à des développeurs extérieurs d'accéder à un ensemble de données ou de ressources logicielles pour les intégrer à des applications externes, les API (*application programming interfaces*) sont devenues un vecteur essentiel de la circulation des données. Elles sont à l'origine du développement des applications et de la croissance des plateformes numériques. Les API permettent à des tiers de développer de nouveaux services autour des données détenues par une organisation, tout en garantissant au détenteur des ressources la faculté de contrôler leur utilisation. Comme le résume Mehdi Medjaoui, fondateur des API Days, les API offrent un « accès et non pas un actif »[51]. Le détenteur des ressources peut continuer à contrôler qui a accès à quoi, où, comment et combien de fois. Il peut par exemple fixer les conditions d'utilisation, refuser l'accès à son API à certains développeurs ou décider d'arrêter un contrat à tout moment s'il l'estime nécessaire. C'est un modèle de partage qui permet donc d'être à la fois ouvert et propriétaire, un point d'ouverture, mais aussi un point de contrôle.

Les avantages d'une API pour le détenteur des ressources sont nombreux. L'ouverture d'API à des développeurs extérieurs permet l'émergence d'un écosystème d'applications couvrant un « périmètre beaucoup plus large que celui qu'aurait pu mettre en chantier une seule entreprise »[52]. De plus, lorsqu'il est payant, l'accès à l'API constitue une source de rémunération pour son propriétaire. Celui-ci peut par exemple se rémunérer sur le chiffre d'affaires réalisé par les développeurs. Enfin, les API permettent d'attirer au sein d'un écosystème des entreprises qui auraient autrement pu développer des services concurrents. Pour les développeurs, les API se présentent sous la forme de briques technologiques facilement paramétrables et prêtes à l'emploi, qui ne nécessitent pas de connaître la logique interne du logiciel tiers. Un autre avantage de l'API est d'offrir un accès à une donnée « ubiquitaire et instantanée » : lorsqu'une donnée change dans un système, elle est automatiquement mise à jour dans les autres systèmes[53]. Inversement, lorsque les données sont partagées sous forme de fichier, les réutilisateurs doivent recharger l'ensemble du fichier à chacune de ses mises à jour. C'est ainsi l'obtention des horaires de la RATP en temps réel via une API, et non uniquement des horaires théoriques sous forme de fichier, qui a permis à Citymapper de développer un véritable service[54].

L'ouverture des données peut par ailleurs s'inscrire dans des démarches collaboratives d'innovation ouverte ou de *crowdsourcing*. Une organisation peut procéder à une ouverture ponctuelle de ses jeux de données dans le cadre d'une compétition ouverte (*hackathon*) pour obtenir les meilleures analyses. Le travail collaboratif entre développeurs et analystes de données peut permettre d'atteindre des résultats plus rapidement et à moindre coût que des travaux menés individuellement ou en petit groupe pour le compte de l'organisation. La plateforme InnoCentive organise par exemple des concours dans lesquels elle propose à des « scientifiques » externes (la plateforme compte 380 000 inscrits) de résoudre des problèmes pour le compte d'autres organisations. Les sujets proposés au concours sont généralement spécifiques et concernent des questions complexes qui n'ont pas trouvé de solutions en interne ou des questions secondaires sur lesquelles l'organisation ne souhaite pas engager de ressources. La personne qui remporte le concours se voit offrir des récompenses pouvant atteindre plusieurs dizaines de milliers de dollars. Dans un autre registre, le portail Internet Zooniverse propose aux internautes de participer à des projets de recherche scientifique en accomplissant certaines tâches qui ne peuvent pas être accomplies par des ordinateurs. Par exemple, dans le cadre du projet Old Weather, 16 400 volontaires ont retranscrit les données météorologiques contenues dans plus d'un million de pages de carnets de la Royal Navy.

Faciliter la circulation et la réutilisation des données : portabilité et interopérabilité

La portabilité et l'interopérabilité sont des éléments essentiels à la circulation des données et au partage de la valeur issue de leur utilisation. Le règlement européen sur les données personnelles a introduit un droit à la portabilité qui permet aux individus dans des cas expressément définis

[51] Medjaoui Mehdi, « Les APIs sont les nouveaux contrats du monde numérique », entretien réalisé par Ouishare pour son magazine, 31 juillet 2018. https://www.ouishare.net/article/les-apis-sont-les-nouveaux-contrats-du-monde-numerique?locale=en_us.

[52] Collin Pierre et Colin Nicolas, *Mission d'expertise sur la fiscalité de l'économie numérique*, rapport pour le ministère de l'Economie et des Finances et le ministère du Redressement productif, 2015, p. 37.

[53] Medjaoui Mehdi, « Les APIs sont les nouveaux contrats du monde numérique », art. cit.

[54] Jacque Philippe, « La RATP ouvre (enfin) ses données temps réel », *Le Monde*, 5 janvier 2017. https://www.lemonde.fr/economie/article/2017/01/05/la-ratp-ouvre-enfin-ses-donnees-temps-reel_5057926_3234.html

de récupérer les données qu'ils ont générées et qui sont stockées et traitées par un prestataire, en demandant qu'elles leur soient communiquées directement ou transmises à une autre entité, là aussi dans des cas définis. La circulation des données passe aussi par la portabilité des données à caractère non personnel, qui n'est pas, à l'heure actuelle, consacrée par la loi, malgré de timides avancées au niveau européen [55].

En facilitant la migration des particuliers et des professionnels entre deux prestataires, la portabilité des données favorise la concurrence entre les services. Elle permet également aux fournisseurs de données de conserver la maîtrise de leurs données : une organisation pourra, par exemple, récupérer ses données pour les réutiliser en interne ou avec d'autres acteurs, en vue de développer de nouveaux services. Il s'agit d'éviter les effets de « verrouillage » qui rendent excessivement complexe, sinon impossible, la récupération des données auprès d'un prestataire et qui aboutissent à ce que ce dernier capte l'intégralité de la valeur créée à partir de celles-ci [56].

Or, la portabilité est en grande partie conditionnée par les modalités de sa mise en œuvre, qui a un coût non négligeable pour les organisations, et dont les effets ne sont pas encore évalués. Le règlement européen sur les données personnelles précise que les données doivent être fournies dans un format « structuré, couramment utilisé et lisible par une machine ». Le règlement européen sur le libre flux des données à caractère non personnel va dans le même sens. L'utilisation de normes ou de standards informatiques fermés - quand le référentiel n'est pas diffusé, ou quand il est soumis à des restrictions d'accès - est un frein à l'échange et à la réutilisation des données [57]. Même quand une donnée est rendue accessible, elle ne peut pas nécessairement être réutilisée si son format n'est pas pris en charge par un autre système, ou si elle est diffusée par une API propriétaire qui en limite l'usage. L'interopérabilité des données est facilitée par l'adoption de standards internationaux, tant sur le plan

[55] Règlement (UE) 2018/1807 du 14 novembre 2018, établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne.

[56] CNNum, *Avis du Conseil national du numérique sur la libre circulation des données en Europe : Fiche sur la consécration d'un droit à la portabilité des données non personnelles*, avril 2017
https://cnnumerique.fr/nos-travaux/libre-circulation-des-donnees-en-europe

[57] D'après la définition donnée par Wikipédia : « Un standard est fermé quand le référentiel n'est pas diffusé, ou quand il est soumis à des restrictions d'accès, par exemple si sa mise en œuvre nécessite le paiement de royalties à cause de brevets, ou si l'octroi de licence sur les brevets est soumis à une acceptation préalable par son éditeur ou si son éditeur impose une clause contractuelle de confidentialité ».

sémantique que technique (XML, JSON, CSV, etc.) et dont l'utilisation n'implique pas l'acquisition d'un logiciel ou d'une licence payante. Le Conseil national du numérique (CNNum) recommande à ce sujet la définition de standards techniques d'interopérabilité de concert avec les industriels et les associations professionnelles et l'utilisation d'interfaces de programmation (API), permettant de transférer directement les données d'un opérateur à un autre et d'interroger automatiquement de vastes bases de données [58].

Vers la plateformalisation pour organiser ces nouveaux écosystèmes

Les plateformes, pivots des écosystèmes des données

La notion de plateforme rassemble une grande variété d'activités, puisqu'on y retrouve aussi bien les places de marché en ligne, les moteurs de recherche, les magasins d'applications, les comparateurs de prix, les réseaux sociaux, les systèmes de paiement, les plateformes d'économie collaborative, etc. [59] En dépit de cette diversité, les plateformes en ligne partagent certaines caractéristiques clés, telles que « l'utilisation des technologies de l'information et de la communication pour faciliter les interactions (y compris les transactions commerciales) entre utilisateurs, la collecte et l'utilisation de données sur ces interactions et les effets de réseau qui rendent l'utilisation des plateformes avec le plus grand nombre d'utilisateurs plus utile » [60]. Ces entreprises jouent un rôle pivot dans les écosystèmes des données. Intervenant aux différents stades de la chaîne de valeur des données, les plus grandes d'entre elles disposent d'un pouvoir de marché considérable qui leur permet d'animer et d'organiser autour d'elles tout les écosystèmes d'acteurs [61].

L'intermédiation est au cœur du modèle économique de la plateforme : les plateformes mettent en relation des vendeurs et des acheteurs, des prestataires de service et des utilisateurs, etc. La plateforme est à la fois

[58] CNNum, *Avis du Conseil national du numérique sur la libre circulation des données en Europe*, avril 2017, *op. cit.*

[59] eBay, Amazon Marketplace, ManoMano, Leboncoin, Facebook et YouTube, Google Play et App Store, PayPal, Waze, Airbnb, Blablacar et Uber sont des exemples de plateformes.

[60] https://ec.europa.eu/digital-single-market/en/glossary#Online_Platforms

[61] Isaac Henri , *Plateforme et dynamiques concurrentielles*, Renaissance Numérique, octobre 2015 http://www.renaissancenumerique.org/publications/plateformes-et-dynamiques-concurrentielles.

un point d'accès et un point de contrôle autour duquel s'organise un écosystème. En tant que point d'accès, la plateforme est incontournable : elle possède une communauté d'utilisateurs à laquelle ses partenaires ne peuvent avoir accès que par son intermédiaire. Cette position centrale lui confère par ailleurs un très important pouvoir de contrôle. C'est elle qui définit les modalités techniques, juridiques et financières de ces partenariats.

Les plateformes sont par ailleurs les acteurs économiques qui extraient le plus de valeur à partir des données. Leur modèle économique a d'ailleurs largement contribué à la prise de conscience de cette valeur. Certaines entreprises ont bouleversé les règles dans l'industrie du transport et de l'hôtellerie, en ne détenant pas d'actifs physiques, tels que des biens immobiliers ou des véhicules, et en n'investissant pas dans leur gestion [62]. Les services que ces plateformes proposent reposent intégralement sur le traitement et l'organisation des informations fournies par les utilisateurs. À partir de ces informations, l'entreprise est en mesure d'enrichir son service d'intermédiation, notamment en faisant des recommandations sur les prix ou au travers d'un système de notation par les pairs pour guider les utilisateurs dans leurs futurs choix et ainsi renforcer la confiance des utilisateurs dans la plateforme.

De plus, les plateformes exploitent systématiquement les effets de réseau [63]. C'est-à-dire que l'utilité de la plateforme est proportionnelle au nombre de ses utilisateurs. La plateforme peut s'appuyer sur des effets de réseau indirects : elle ouvre ses ressources, notamment ses jeux de données, à des développeurs extérieurs, afin qu'ils développent une offre complémentaire de services qui renforcera en retour l'attractivité de la plateforme. C'est ainsi qu'Apple bénéficie pour ses téléphones d'une riche offre d'applications mobiles développées par des entreprises tierces. La plateforme peut aussi profiter d'effets de réseau croisés. En exerçant son activité sur un marché, une entreprise constitue une audience qu'elle va ensuite pouvoir valoriser sur d'autres marchés. L'exemple le plus caractéristique est sans doute celui de la publicité en ligne. Au départ, les plateformes proposent le plus souvent un service gratuit afin de se constituer une vaste audience auprès des utilisateurs. Une fois cette audience constituée, la plateforme commercialise les données de ses

[62] Chignard Simon et Benyayer Louis-David, *Datanomics, les nouveaux business models des données, op. cit.*, p. 94.

[63] Isaac Henri, *Plateforme et dynamiques concurrentielles*, Renaissance Numérique, *op. cit.*

utilisateurs sur le marché de la publicité en ligne. Mais cette valorisation de l'audience peut également prendre la forme de la vente de mots-clés, de l'intéressement, de commissions de rétrocession, etc. Comme le remarque le CNNum, « parfois, la plateforme va plus loin et passe des accords de quasi-exclusivité technique avec certains de ses fournisseurs qui deviennent des partenaires, voire elle devient elle-même un opérateur venant concurrencer directement les entreprises qui étaient auparavant ses clientes » [64].

Adopter une stratégie de plateformisation

Combinés à des rendements d'échelle croissants, les effets de réseau aboutissent rapidement à des situations oligopolistiques, dans lesquelles le vainqueur remporte l'essentiel du marché ("*the winner takes all*" ou "*the winner takes the most*") et d'importantes marges de profits. Pour les acteurs établis, la plateformisation de l'économie prend d'abord la forme d'une menace de désintermédiation. De nouveaux acteurs économiques arrivent et bouleversent les règles du jeu sur leur marché, voire en modifient les frontières. Le phénomène est particulièrement prégnant dans certains secteurs, si l'on veut bien songer à l'industrie culturelle et aux services de *streaming*, à l'hôtellerie, aux taxis ou encore au transport ferroviaire.

En réaction, de plus en plus d'entreprises issues des secteurs d'activité traditionnels entendent à leur tour se positionner comme des plateformes dans leur secteur d'activité. Le président de Nexity, Alain Dinin, entend ainsi faire de son organisation une plateforme de services de l'immobilier, qui accompagne le client tout au long de sa vie, et sur laquelle des startups viennent se greffer [65]. Guillaume Pepy, le président de la SNCF, va dans le même sens lorsqu'il présente son entreprise comme une plateforme de solutions de mobilité, capable d'agréger différents types de services, du train au VTC, en passant par de la location de voiture, du car ou du covoiturage [66]. Pour une organisation, se positionner comme

[64] CNNum, *Neutralité des plateformes : Réunir les conditions d'un environnement numérique ouvert et soutenable*, mai 2014, p. 83.

[65] Sabbah Catherine, « Nexity va muter en plateforme de services à l'immobilier », *Les Échos*, 23 janvier 2017, https://www.lesechos.fr/23/01/2017/lesechos.fr/0211721902733_nexity-va-muter-en-plate-forme-de-services-a-l-immobilier.htm

[66] Steinmann Lionel, « Après la réforme, priorité à la plateforme », *Les Échos*, 8 janvier 2019. https://www.lesechos.fr/idees-debats/editos-analyses/0600468947163-sncf-apres-la-reforme-priorite-a-la-plate-forme-2234521.php

une plateforme, c'est se positionner comme le réceptacle de données, y compris des données issues de tierces parties, auxquelles d'autres acteurs pourront avoir accès pour créer des services complémentaires. La plateforme qui réussit à s'imposer devient alors le point de passage obligé, un acteur incontournable qui ne peut pas être écarté. Mais comme le rappellent Simon Chignard et Louis-David Benyayer, « il n'y a pas de place pour plusieurs plateformes dans une même arène concurrentielle » [67].

Cependant, contre la tendance oligopolistique, un autre modèle, encore balbutiant, tend à émerger, celui des plateformes de partage, dans lequel plusieurs acteurs économiques coopèrent pour organiser la collecte, la mise en partage et le traitement d'une partie de leurs données. L'objectif de ces plateformes de partage est double. Il s'agit, tout d'abord, d'optimiser la valorisation des données, non seulement en accroissant la quantité et la variété des données disponibles, mais aussi en mutualisant les ressources humaines, techniques et financières nécessaires à leur traitement. Ces plateformes de partage ont aussi une vocation défensive, puisqu'il s'agit en partie d'éviter que la valorisation intervienne uniquement au profit de quelques grands acteurs.

Le secteur agricole est emblématique de cette tension. Des sociétés comme le géant agrochimique Monsanto ou le fabricant de machines agricoles John Deere collectent d'importantes quantités de données sur les exploitations agricoles de leurs clients. Croisées avec d'autres informations, par exemple les conditions météo, ces données peuvent être utilisées pour prodiguer des conseils aux exploitants. Mais elles peuvent également servir à leur vendre des services agricoles complémentaires, voire être directement commercialisées auprès de tiers dont les intérêts ne convergent pas nécessairement avec ceux des agriculteurs : courtiers sur les marchés agricoles, assureurs, fabricants de produits phytosanitaires, etc. Les agriculteurs encourent donc un double risque : celui de se voir dépossédés de la valeur des données qu'ils produisent et celui de voir ces données utilisées contre eux. La révision des modalités contractuelles, ainsi que la mise en place de plateformes de partage des données agricoles accessibles aux contributeurs et aux partenaires que ceux-ci auront choisi, font partie des solutions choisies dans le secteur [68].

[67] Chignard Simon et Benyayer Louis-David, *Datanomics, les nouveaux business models des données, op. cit.*, p. 87.

[68] Renaissance numérique, *La valeur des données en agriculture*, février 2018.

Les collaborations peuvent prendre place entre des acteurs qui sont en concurrence sur les marchés, comme l'illustre le rachat du service de navigation HERE par les constructeurs automobiles allemands Volkswagen, BMW et Mercedes. Avec ce rachat, les industriels allemands créent une plateforme commune de connectivité et de navigation qui entend être une alternative crédible à Google Maps. L'enjeu pour les constructeurs est de garder la main sur les données de leurs véhicules et les données de leurs clients, qui constituent des ressources vitales dans la course au véhicule autonome. Les collaborations peuvent également se faire au sein d'une même filière [69], avec des acteurs situés en différents points de la chaîne de valeur.

Ces initiatives témoignent d'une prise de conscience que la valeur des données passe par leur mise en partage et leur circulation. Mais mutualiser des données, en les regroupant dans un pot commun, n'est pas, en soi, une promesse de création de valeur. Là où les plateformes extraient de la valeur, c'est dans l'agrégation, l'enrichissement, le croisement, le traitement des données et dans leur restitution sous forme de services à valeur ajoutée pour les tiers. Que ce soit pour l'industrie automobile ou pour la filière agricole, la mise en partage de données doit s'accompagner d'une véritable stratégie de valorisation des données, laquelle passe par la définition d'objectifs précis, le développement de partenariats avec des acteurs tiers et la conception d'offres de services innovants. C'est seulement de cette manière que la plateforme peut s'imposer comme un lieu de création de valeur et attirer de nouvelles données et de nouveaux partenaires.

Les plateformes de partage sont également confrontées au problème de la répartition de la valeur entre les fournisseurs de données, les utilisateurs et la plateforme elle-même. Le sujet est loin d'être simple, d'autant que la mise en place d'une plateforme et sa gestion exigent d'importants investissements [70]. C'est pourquoi dans certains secteurs l'intervention de l'État peut accompagner le mouvement de plateformisation. À ce titre, la stratégie nationale d'intelligence artificielle a identifié quatre secteurs prioritaires pour la mise en commun de données : la santé, la mobilité, l'environnement et la sécurité.

[69] En France, voir l'initiative Valdriv** PLM qui réunit les constructeurs et les équipementiers du secteur.

[70] Choukroun Marc-David, « Non, la coopérative n'est pas l'avenir de l'économie collaborative », Les Échos, 29 décembre 2015. https://www.lesechos.fr/29/12/2015/LesEchos/22096-028-ECH_non--la-cooperative-n-est-pas-l-avenir-de-l-economie-collaborative.htm

Quel rôle pour les acteurs publics dans ces écosystèmes ?

L'État, et plus globalement l'administration, a un rôle clé à jouer dans l'organisation et l'animation des écosystèmes des données. Cette dernière est à la fois productrice et utilisatrice de données, et il lui appartient de fixer le cadre réglementaire applicable aux données, à leur circulation et à leur utilisation. Elle pourrait par exemple mettre en place des mesures incitatives ou se positionner comme « État tiers de confiance ».

L'*Open Data*, vecteur d'enrichissement des écosystèmes

Au sens large, l'*Open Data* désigne une démarche d'ouverture des données qui n'est pas limitée aux organismes publics [71]. En France, l'expression est généralement utilisée dans un sens plus restreint pour désigner le mouvement d'ouverture des données publiques.

Les administrations sont traditionnellement de très importantes productrices de données. La statistique publique, les enquêtes publiques, le cadastre, les bases topographiques, les circulaires, les registres d'état civil et des sociétés sont des instruments sur lesquels s'appuient classiquement les autorités publiques pour mener à bien leurs missions. Cependant, pour des administrations qui ont longtemps fonctionné dans la culture du secret, la diffusion de ces informations au public, qui plus est sous un format exploitable pour un traitement informatique, ne va pas de soi. Selon les mots de l'ancien administrateur général des données, Henri Verdier, « jusqu'à une date récente, la situation était relativement paradoxale. D'un côté, les grands producteurs de données, opérateurs ou services d'administration centrale, monétisaient leurs données (via des redevances) auprès d'une poignée d'acteurs économiques. De l'autre, l'exploitation de la très grande majorité des données était limitée à l'administration qui les produisait ou les collectait, entraînant ainsi une perte d'opportunité » [72].

Depuis la fin des années 2000, s'appuyant sur les possibilités offertes par les nouvelles technologies, de nombreux gouvernements et des collectivités locales se sont engagés dans une démarche d'ouverture de données publiques. Les objectifs poursuivis par les politiques d'*Open Data* sont de deux sortes.

[71] À ce sujet, voir les travaux de l'Open Data Institute disponibles sur le site Internet de l'organisation : https://theodi.org/

[72] Administrateur général des données, *La donnée comme infrastructure essentielle*, rapport annuel 2016-2017, p. 15.

En premier lieu, l'ouverture des données publiques revêt une dimension éminemment démocratique : la transparence de l'action publique qui permet aux citoyens de juger de l'action de leurs gouvernants. Le principe de redevabilité est inhérent au fonctionnement de tout système démocratique et on peut le trouver exprimé à l'article 15 de la Déclaration des droits de l'homme et du citoyen de 1789 : « la société a droit de demander compte à tout agent public de son administration ». Afin de pouvoir exercer un contrôle sur l'action des gouvernants et sur l'utilisation de l'argent public, les citoyens et leurs représentants doivent pouvoir accéder à une information transparente sur l'activité des autorités publiques et des élus.

En second lieu, l'ouverture des données publiques est un instrument au service de l'innovation et du développement économique. En mettant ses données à disposition du public, l'administration entend encourager leur réutilisation par des tiers en dehors de leur vocation d'origine. L'ouverture permet d'explorer et de révéler la valeur des données à travers la création de nouveaux services : des services marchands ou non marchands et des services venant enrichir les services publics existants. Cette ouverture constitue désormais un pan de la stratégie française pour l'intelligence artificielle : rendre disponibles d'importantes quantités de données (publiques ou privées) pour entraîner les algorithmes du *machine learning* et du *deep learning*.

Aujourd'hui, le cadre juridique français est très complet. Il s'articule autour de trois grands principes : l'accès et la diffusion des documents administratifs, le principe de réutilisation et le principe de gratuité.

Un acte important de l'ouverture des données publiques en France est la loi dite « loi CADA » de 1978, du nom de la commission qu'elle a instituée (Commission d'accès aux documents administratifs)[73]. Elle instaure un droit d'accès aux informations produites par les administrations dans le cadre d'une mission de service public. Le droit d'accès a par la suite été enrichi par de nombreux textes qui sont progressivement venus étendre le périmètre de l'ouverture et qui ont complété le droit d'accès par un droit à la réutilisation des données publiques. Avec l'arrivée d'Internet, la question de l'accès aux documents administratifs change de nature.

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

Désormais, l'enjeu n'est plus seulement de garantir l'accessibilité dans le cadre de requêtes individuelles, mais de mettre à disposition du public les données de l'administration de manière permanente.

La loi pour une République numérique de 2016 est venue consacrer le principe de l'accès par défaut aux informations administratives. Elle crée l'obligation pour les organisations publiques et pour les collectivités territoriales de plus de 3500 habitants de publier sur Internet leurs bases de données dans un standard ouvert et aisément réutilisable. Par ailleurs, elle introduit dans le droit la notion de données d'intérêt général, qui impose l'ouverture des données détenues par les acteurs privés délégataires d'une mission de service public ou qui bénéficient de subventions publiques. L'ouverture de ces données permet aux autorités publiques de contrôler l'activité des délégataires et de partager l'information avec eux.

La valeur économique et sociale qui peut être créée par la mise à disposition à titre gratuit des données publiques est bien supérieure à la valeur que l'État pourrait tirer de leur vente[74]. La gratuité augmente les chances de réutilisation des données publiques en réduisant le coût des usages commerciaux et en permettant le développement d'usages non commerciaux qui n'auraient pas été possibles si l'accès aux données avait été payant. Elle permet à des entreprises de taille modeste, et même à l'administration, d'utiliser des données auxquelles elles n'auraient pas eu accès si elles avaient dû payer pour les acquérir.

La disponibilité des données est un élément essentiel de l'*Open Data*, mais ce n'est pas le seul. Il faut encore que les données soient accessibles et, pour un objectif démocratique, simples à comprendre par les tiers. Le format utilisé pour la publication des données, mais aussi l'interface qui permet d'y accéder, sont des éléments fondamentaux des politiques d'ouverture des données publiques.

En 2011, la création de la mission Etalab, service du secrétariat général à la modernisation de l'État chargé de coordonner la politique d'ouverture et de partage des données publiques, marque une étape importante. Etalab administre la plateforme *data.gouv.fr* qui réunit et met à disposition librement l'ensemble des données publiques de l'État et de ses établissements publics. Si elles le souhaitent, les collectivités territoriales et les personnes de droit public ou de droit privé chargées d'une mission

[74] Trojette Mohammed Adnène, *Ouverture des données publiques. Les exceptions au principe de gratuité sont-elles toutes légitimes ?* Rapport au Premier Ministre, juillet 2013 ; OCDE, *Council Recommendation for Enhanced Access and More Effective Use of Public Sector Information*, 2008.

de service public peuvent également utiliser cette plateforme. Début 2019, le site annonçait mettre à disposition 36 034 jeux de données et 1915 réutilisations au-delà de leur utilisation première par l'administration. L'une des originalités de la plateforme est de permettre aux usagers d'enrichir et d'améliorer les données disponibles, et également d'en déposer de nouvelles, dans une logique contributive. Par exemple, Blablacar a utilisé le portail pour partager le fichier des aires de parking de covoiturage qu'il avait constitué. Cette logique de co-construction s'étend aux données produites par les citoyens, notamment les associations comme Open Street Map, Open Météo Foundation, Open Food Facts, etc.

Le service public de la donnée

L'article 14 de la loi pour une République numérique crée un service public de la donnée dont la gouvernance a été confiée à la mission Etalab. Celui-ci vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui ont le plus fort impact économique et social, et dont la qualité de mise à disposition est critique pour les utilisateurs. À ce jour, neuf jeux de données ont été identifiés comme des données de référence : le répertoire des entreprises et des établissements (base Sirene), le répertoire national des associations (RNA), le plan cadastral informatisé (Cadastre), le registre parcellaire graphique (RPG), le référentiel à grande échelle (RGE), la base adresse nationale (BAN), le répertoire opérationnel des métiers et des emplois (ROME), le référentiel de l'organisation administrative de l'État, et le code officiel géographique (COG).

Les données d'intérêt général : du constat partagé aux difficultés de mise en œuvre

Les écosystèmes des données peuvent également être enrichis par l'ouverture de jeux de données détenus par des organismes privés. Si cette ouverture peut produire de nombreuses externalités positives pour la collectivité, les détenteurs de données ne sont pas toujours enclins à partager des données qui leur procurent un avantage concurrentiel ou qu'ils entendent exploiter eux-mêmes. Leur intérêt économique peut même être, au contraire, de garder pour eux certaines données, ou de ne les partager qu'avec les partenaires qu'ils auront. Face à ce paradoxe, l'État a engagé une réflexion sur la définition d'une nouvelle catégorie juridique de données : « les données d'intérêt général », une catégorie de données privées dont l'ouverture pourrait être imposée par la loi pour des motifs d'intérêt général [75].

L'origine du sujet remonte à 2015 et à la prise de conscience qu'il manquait un outil juridique pour accéder à des données de transport qui ne sont pas produites par une autorité publique ou dans le cadre de délégations de services publics, mais qui n'en participent pas moins de l'intérêt public. Par exemple, une Ville comme Paris peut accéder aux données de mobilité de la RATP et des délégations de services publics, mais pas aux données produites par le secteur privé, qui représentent une part essentielle des données de mobilité sur son territoire : les taxis, les VTC, les trottinettes électriques, les deux-roues en libre-service, le covoiturage ou encore les applications de navigation GPS comme Waze ou Google Maps. Pourtant, l'ouverture de ces données présenterait des avantages certains. La Ville pourrait s'appuyer sur une meilleure connaissance des flux de circulation sur son territoire pour réorienter son offre de transport public et décider de nouveaux aménagements dans l'espace public. Les usagers seraient en mesure d'identifier les différentes offres de transport disponibles pour effectuer un même trajet. La disponibilité de ces données pourrait susciter la création de services innovants de transport multi-modaux ou d'information des voyageurs (*mobility as a service* ou MaaS).

En Finlande, pays pionnier dans l'ouverture des données de transport, les opérateurs doivent obligatoirement partager leurs données en temps réel, utiliser des interfaces logicielles ouvertes et donner accès à leur billetterie[76]. À Helsinki, l'application Whim permet d'organiser ses déplacements en vélo libre-service, en métro, en bus, en tramway, en taxi ou en voiture de location. Les utilisateurs peuvent réserver et payer un trajet en un clic, combiner les modes de déplacement et souscrire à des

[75] Duchesne Claudine, Cytermann Laurent, Vachey et al. *Rapport relatif aux données d'intérêt général*, op. cit. ; Vachey Laurent, Duchesne Claudine, Meyer Marc et al., *Les données d'intérêt général : Phase 2*, rapport de l'Inspection générale des finances et du Conseil général de l'économie, de l'industrie, de l'énergie et des technologies, mars 2016.
La Commission européenne a également engagé une réflexion sur l'accès aux données du secteur privé à des fins d'intérêt public dans le cadre de la révision de la directive concernant la réutilisation des informations du secteur public. La question sera vraisemblablement à l'agenda de la prochaine Commission. Voir Commission Européenne, « Vers un espace européen commun des données », communication, COM (2018) 232 final, Bruxelles, le 25 avril 2018.

[76] Allix Grégoire, « À Helsinki, une appli permet d'organiser tous ses déplacements en un clic », *Le Monde*, 13 octobre 2018.
https://www.lemonde.fr/economie/article/2018/10/13/a-helsinki-une-appli-permet-d-organiser-tous-ses-deplacements-en-un-clic_5368833_3234.html?xtmc=mobility_as_a_service&xtcr=5

abonnements qui intègrent les différents moyens de transport. Le sujet du *MaaS* a également émergé en France, notamment dans le cadre de la Loi d'orientation des mobilités, qui contient des mesures visant à favoriser la circulation des données de mobilité.

Le concept de « données d'intérêt général » laisse toutefois plusieurs questions en suspens :

- Comme pour l'*Open Data*, une politique d'ouverture des données privées à des fins d'intérêt général peut poursuivre des objectifs qui apparaissent très différents les uns des autres. Les données privées peuvent, tout d'abord, servir à améliorer l'efficacité des politiques publiques en permettant l'adoption de décisions politiques mieux informées. Elles peuvent également, dans une logique de transparence, procurer une meilleure information aux citoyens et aux consommateurs. Enfin, l'accessibilité des données privées est à même de contribuer au développement économique, et cela de diverses manières : en stimulant l'innovation et le développement de nouveaux services, en contribuant à l'essor des travaux sur l'intelligence artificielle ou encore en favorisant la concurrence sur les marchés. Selon la finalité recherchée, la diffusion des données d'intérêt général se fera auprès de différents cercles : les autorités publiques (*BtoG, Business to Government*), les citoyens (*BtoC, Business to Citizens*), les entreprises (*BtoB, Business to Business*) ou l'ensemble des acteurs.
- Le périmètre de l'ouverture est difficile à déterminer. Il ressort de la réflexion publique sur le sujet que l'adoption d'une loi transversale sur l'ouverture des données privées d'intérêt général n'est ni souhaitable ni possible juridiquement, et qu'il convient plutôt de retenir une approche sectorielle et différenciée selon les données concernées [77]. Le transport, l'emploi, la formation, la santé, l'énergie et la finance constituent ainsi des secteurs clés pour l'ouverture. Au niveau européen, la directive sur les services de paiement (DSP2) impose aux banques de fournir l'accès aux données de leurs clients (avec leur consentement) à des acteurs tiers via des API.
- Les règles d'accès et de réutilisation des données doivent être établies au cas par cas, selon les données et leur destinataire. Une question qui se pose est notamment de savoir si les entreprises auxquelles on impose de rendre leurs données accessibles à des tiers ont droit

[77] Duchesne Claudine, Cytermann Laurent, Vachey et *al. Rapport relatif aux données d'intérêt général, op. cit.*

à une compensation financière. Dans la mesure où l'ouverture des données privées peut se heurter aux droits de propriété, à la liberté d'entreprendre et au secret industriel et commercial et qu'elle implique des coûts, les détenteurs de données pourraient demander à être rémunérés pour ce service. On peut aussi imaginer que l'accès des autorités publiques à des fins de politiques publiques reste gratuit. Quant aux modalités du partage de données, plusieurs formules sont envisageables : accès aux données directement auprès de leur détenteur ou hébergement des données au sein d'un organisme tiers chargé d'assurer leur circulation, création d'un organe de règlement des différends pour régler les conflits relatifs au partage, etc.

Notons que plusieurs acteurs numériques se sont engagés ces dernières années dans des politiques d'ouverture volontaire de leurs données. Depuis 2014, le programme *Connected Citizens* de Waze permet ainsi à des collectivités locales et à certaines entreprises privées d'accéder aux données sur les accidents et les ralentissements signalés par les utilisateurs de l'application ; en échange, ces dernières communiquent à Waze des informations sur les routes fermées, les travaux et les évènements susceptibles de perturber le trafic. Lancé en 2017, *Uber Movement* permet de connaître le temps de transport entre deux points et d'obtenir des statistiques sur le trafic routier. Enfin, le portail *Dataville* de Airbnb permet aux collectivités d'accéder aux données de la plateforme concernant le nombre d'annonces sur leur territoire, le nombre de voyageurs accueillis, les pays d'origine des voyageurs et le revenu annuel médian d'un hôte, et d'enrichir ainsi leur connaissance de l'activité touristique de leur territoire.

Dans ce type de démarche, les entreprises choisissent les informations qu'elles rendent disponibles et leur niveau de granularité, les modalités d'accès aux données concernées (Waze et Uber demandent une inscription préalable) et elles en fixent les conditions d'utilisation (Uber refuse tout usage commercial ; Waze refuse que les données diffusées soient partagées avec des tiers).

L'État, moteur de la plateformisation des secteurs

L'État peut adopter des mesures réglementaires ou incitatives pour faciliter la rencontre et encourager le partage entre les organisations qui détiennent les données et les organisations qui recherchent des données pour développer leur activité. Sans imposer l'ouverture, les politiques publiques peuvent s'attaquer aux freins techniques, culturels et juridiques au partage.

Les organismes publics détenteurs de données peuvent être frileux à l'idée d'offrir un accès à leurs données, d'autant plus qu'ils ne perçoivent pas nécessairement les bénéfices qu'ils peuvent en retirer. Ils peuvent notamment craindre de se dessaisir d'un actif stratégique, ou que cette ouverture ne les place en situation de violation de la législation sur la protection des données. À ce titre, la Cour des comptes a publié, le 11 mars 2019, un référé portant sur l'ouverture et la valorisation des données dans trois établissements publics (IGN, Météo France et Cerema) concernés, à des degrés divers, par l'obligation de gratuité de l'*Open Data*. Pour la Cour des comptes, ces acteurs sont « confrontés à des difficultés de mise en œuvre des dispositions légales relatives à l'ouverture des données ; d'autre part, ils voient leur modèle économique remis en cause, particulièrement l'IGN ». Dans sa réponse à ce référé, le Premier Ministre indique avoir demandé à « l'Inspection Générale des Finances, avec l'appui, le cas échéant, de la Direction interministérielle du numérique et des systèmes d'information et de communication de l'État (DINSIC) qui pilote au niveau interministériel la politique d'ouverture des données et le service public de mise à disposition des données de référence, d'établir sous six mois un premier bilan de la mise en œuvre de l'ouverture des données et de ses impacts, des difficultés rencontrées par les ministères et leurs opérateurs et de proposer les mesures d'accompagnement adaptées. »

Par ailleurs, les données produites doivent l'être suivant des formats standards pour pouvoir être agrégées avec des données de sources tierces et donc utilement partageables. Comment garantir aux organisations qui investissent aujourd'hui dans des systèmes d'information lourds et coûteux que ces derniers sont interopérables avec ceux de leurs homologues, au sein de mêmes filières ? De plus, même lorsque deux organisations sont désireuses de partager des données, elles ne disposent pas toujours du savoir-faire nécessaire pour conclure un accord et organiser le partage en pratique.

Pour aider les acteurs concernés à investir dans des systèmes capables d'assurer l'interopérabilité requise, l'État pourrait flécher les aides en matière d'investissement, par exemple au travers de labels, sur la base d'un cahier des charges en matière d'interopérabilité, tant technique que sémantique.

Dans un rapport remis au gouvernement britannique en 2017, Dame Wendy Hall et Jérôme Pesenti identifient deux leviers à même de favoriser l'accès, la circulation et le partage des données privées. Une première manière pourrait prendre la forme de l'établissement de lignes directrices

et de contrats types. À l'heure actuelle, les coûts de transaction pour la conclusion d'accords entre deux organisations sont élevés, ce qui rend leur conclusion difficile pour de petits acteurs.

L'établissement de contrats types pourrait faciliter l'accès aux données de plus petites structures en réduisant leurs coûts de transaction. Mais il pourrait également contribuer à supprimer les réserves au partage des détenteurs de données en garantissant le caractère équitable et sécurisé de l'échange. C'est là un des points clés issus des auditions menées par les préfigurateurs du Health Data Hub : la difficulté pour les start-ups innovantes du secteur des données de santé à se développer en France, l'établissement de contrats d'accès aux données nécessaires pour mettre au point et entraîner leurs algorithmes devant se faire auprès de chaque établissement hospitalier individuellement. Une des missions du Health Data Hub sera de fournir une assistance sous la forme d'expertise de valorisation et de modèles de contrat pour faciliter les négociations et garantir un accès effectif aux données dans des délais raisonnables et annoncés. Dans le secteur de l'agriculture, ce sont les acteurs eux-mêmes, notamment par le biais des filières, qui commencent à établir des modèles de contrat encadrant le partage des données.

En second lieu, le rapport Hall-Pesenti préconise la création de « *data trusts* » qui sont peu ou prou l'équivalent des « plateformes de mutualisation sectorielle » recommandées par la mission Villani relative à la stratégie nationale de la France sur l'IA. Ces organismes joueraient le rôle de tiers de confiance entre les parties qui souhaitent s'engager dans un partage de données et les potentiels réutilisateurs. La question de leur forme juridique (nomment publique ou privée) et de leur modèle économique reste ouverte. Ces organismes seraient notamment en charge de fixer un cadre à l'échange de données et de promouvoir les bonnes pratiques. Ils pourraient aider les parties à s'accorder sur les données partagées, ainsi que sur les conditions de transfert, de stockage, d'accès et d'utilisation ou encore sur la répartition de la valeur générée par leur utilisation. Ces tiers de confiance pourraient encore aider les parties à évaluer la valeur de leurs données, fixer des lignes directrices sur l'anonymisation et la sécurisation des données ou délivrer des conseils sur le respect de la réglementation en vigueur (notamment le RGPD). Ainsi, ils pourraient porter toutes mesures à même de favoriser le partage entre organisations et de renforcer la confiance des utilisateurs.

L'intervention publique dans ces plateformes de partage de données sectorielles n'est justifiée que si ce service apporte des garanties supplémentaires aux fournisseurs et aux utilisateurs de données, ou parce qu'il intervient dans des domaines dans lesquels il existe une carence de l'initiative privée.

En France, le projet de « Health Data Hub » ouvre la voie à la création de plateformes de partage de données pilotées par les autorités publiques [78], en lien et en coordination avec l'ensemble des acteurs du secteur. À ce titre, il témoigne également de la difficulté de la mise en œuvre de ces structures et de leur gouvernance. Le Health Data Hub est destiné à faciliter l'accès et la réutilisation par des tiers des nombreux jeux de données de santé produits dans le cadre des soins, au-delà des données médico-économiques de l'Assurance Maladie. Il devrait prendre la forme d'un guichet unique permettant aux porteurs de projets présentant un intérêt public d'accéder aux données de santé via une plateforme sécurisée. Le groupement d'intérêt public (GIP) Health Data Hub aura pour mission de faire émerger et d'animer un écosystème autour des données de santé :

- en accompagnant les producteurs de données de santé dans la collecte et l'amélioration de la qualité des données ;
- en aidant les utilisateurs à comprendre quelles sont les données disponibles et leur potentielle valorisation ;
- en assurant la mise en relation des producteurs et des utilisateurs publics et privés selon un processus standardisé ;
- en mettant à disposition des utilisateurs des ressources technologiques et humaines mutualisées.

À la suite des conclusions de la mission Villani, la France a également lancé fin 2018 un Appel à Manifestation d'Intérêt sur la mutualisation des données au sein de plateformes sectorielles ou cross-sectorielles. L'État entend favoriser la constitution de grandes bases de données mutualisées susceptibles de servir au développement de l'IA. L'intervention de l'État prendra la forme d'un soutien financier à des opérations souvent consommatrices de temps et de ressources humaines et financières, telles que l'enrichissement des jeux de données, l'annotation des données, le développement d'interfaces techniques ou le développement d'outils de gestion technique et de facturation.

[78] Polton Dominique, Cuggia Marc et Wainrib Gilles, *Health Data Hub : Mission de préfiguration*, rapport à la ministre des Solidarités et de la Santé, 2018.

Des initiatives de partage de données voient également le jour au niveau local. Des projets comme DECODE (*DEcentralised Citizen-owned Data Ecosystems*) à Barcelone et à Amsterdam et Mes.Infos à Lyon, Rennes et La Rochelle cherchent à initier de nouvelles formes de partage des données entre les citoyens et leur collectivité. Les citoyens acceptent de partager leurs données personnelles avec la collectivité pour qu'elles soient utilisées dans la conduite des politiques publiques.

CHAPITRE 3

DE NOUVEAUX ENJEUX DE GOUVERNANCE FACE À LA MISE EN DONNÉES DE LA SOCIÉTÉ

1_ Les données, outil d'une démocratie renouvelée

La valeur d'usage des données va au-delà de sa seule dimension économique, que l'on songe à la dimension démocratique des politiques d'*Open Data* ou à l'utilisation qui peut être faite des données dans l'amélioration des politiques publiques. Encore faut-il que les administrations et les citoyens soient suffisamment formés pour s'emparer de ces nouveaux instruments.

Se saisir des enjeux démocratiques de l'ouverture des données

Favoriser une « vigie » citoyenne

Permettre aux citoyens d'accéder aux informations publiques est une idée ancienne, qui n'a pas attendu la fin des années 2000 et la mise en place des premières politiques d'ouverture des données publiques. L'objectif visé, celui de « favoriser l'émergence d'une contre-expertise citoyenne par la réduction des asymétries d'information », est inhérent à l'exigence de transparence dans une société démocratique [79]. La démocratisation d'Internet, en rendant possible une mise à disposition de l'ensemble des données publiques de manière immédiate et permanente, facilite l'accès aux informations et ouvre de nouvelles perspectives de participation citoyenne.

L'ouverture des données publiques permet d'introduire une « culture des faits » dans l'action publique. Désormais, les citoyens et la société civile disposent d'un accès aux informations nécessaires pour contrôler l'activité des élus et de l'administration, et le cas échéant demander des comptes sur les actions menées, sur l'affectation de l'argent public et plus largement sur les tenants de la décision publique. Tout cela dans une logique de redevabilité (*accountability* en anglais). Encore faut-il que l'ouverture ne se limite pas aux données dites « inoffensives » (comme les horaires d'ouvertures de bâtiments publics), mais qu'elle concerne aussi celles qui revêtent une dimension politique (qualité de l'air, détails des votes au conseil municipal, études d'impact, décisions de justice, qualité des services publics, données d'études cliniques présidant à l'autorisation sur le marché des produits de santé, etc.) [80].

[79] Renaissance Numérique, *Démocratie : mise à jour. 13 propositions pour une version améliorée de l'État, sa posture et son équilibre démocratique*, avril 2016.

[80] Voir la partie relative à l'*Open Data* dans le chapitre 2.

Les effets vertueux attendus d'une libération de la donnée publique sont également de plusieurs sortes pour l'action publique[81]. D'une part, la transparence peut avoir une incidence sur l'efficacité de l'action publique. L'administration est incitée à expliquer la répartition des dépenses publiques et à poursuivre l'amélioration de la qualité de ses services. Elle accède aussi à une meilleure information grâce au partage des autres services et informations. En imposant l'ouverture des données des entreprises qui exercent une délégation de service public, la loi pour une République numérique de 2016 permet également de contrôler au plus près l'activité des délégataires[82]. Par exemple, dans le domaine de l'eau, l'organisme chargé du service dispose de bases de données sur la consommation des ménages et des entreprises, sur les fuites et les opérations de maintenance. Les services publics de vélopartage génèrent des données sur les déplacements, les durées d'utilisation ou encore l'usure du parc. Autant d'informations qui ont de la valeur aux yeux des autorités publiques, qu'il s'agisse d'évaluer la qualité du service, de faire évoluer son organisation ou de s'assurer que le délégataire respecte effectivement le cahier des charges.

D'autre part, la transparence démocratique peut être une source de légitimité supplémentaire pour l'action publique. L'ouverture des données conduit les autorités publiques à se rapprocher des citoyens pour dialoguer et interagir avec eux. Des citoyens mieux informés et associés à la conduite des politiques publiques gagnent en autonomie et en capacité d'action. Ils sont de ce fait plus à même de contribuer au bon fonctionnement des institutions démocratiques.

L'*Open Government Partnership* : un accélérateur pour la transition numérique des États ?

L'*Open Government Partnership* est une organisation internationale lancée en 2011 et composée de 80 pays et de centaines d'organisations de la société civile qui agissent pour promouvoir la transparence de l'action publique, sa construction avec la société civile et l'innovation démocratique.

[81] Peugeot Valérie, « L'ouverture des données publiques : convergence ou malentendu ? », in (dir.) Stiegler Bernard, *Confiance, croyance, crédit dans les mondes industriels*, éditions FYP, 2011.

[82] Duchesne Claudine, Cytermann Laurent, Vachey *et al.*, *Rapport relatif aux données d'intérêt général, op. cit.*, p. 34.

Ses membres s'engagent à respecter les grands principes de la Déclaration pour un gouvernement ouvert :

- Transparence de l'action publique, notamment par l'intermédiaire de l'ouverture des données publiques
- Participation des citoyens à l'élaboration et à l'évaluation des politiques publiques
- Intégrité de l'action publique et des agents publics
- Utilisation des nouvelles technologies en faveur de l'ouverture et de la redevabilité

Tous les deux ans, les pays membres élaborent, conjointement avec la société civile, un plan d'action sur les mesures qui seront développées en faveur d'une action publique transparente et collaborative. L'organisation ambitionne notamment d'établir un cadre légal pour la publication des revenus des hauts fonctionnaires, de promouvoir la transparence budgétaire et fiscale et d'augmenter la participation des citoyens dans l'action publique. À ce titre, chaque pays membre doit produire un Plan d'action national visant à faire progresser la transparence et l'ouverture de l'action publique dans les domaines promus par le partenariat.

De l'intérêt d'une politique publique « *data-driven* »

Les services publics maintiennent traditionnellement le public dans un rôle passif, comme destinataire d'une offre de services standardisés. À l'heure de la révolution numérique, une autre approche est toutefois possible. À l'exemple des innovations en cours dans l'économie numérique, l'administration peut désormais s'appuyer sur ce que Nicolas Colin et Henri Verdier désignent comme « la multitude », ces millions de citoyens instruits et connectés, mais aussi des chercheurs, des associations et des entreprises, pour élaborer et conduire ses politiques publiques [83].

La figure de l'usager-contributeur peine encore à s'imposer, faute d'avoir réussi à trouver les formes institutionnelles appropriées pour mettre les ressources de l'intelligence collective au service de l'administration, sans ébranler l'autorité de cette dernière. Mais les plaidoyers pour une rupture stratégique de la part de l'État - sa plateformisation - se multiplient [84].

[83] Colin Nicolas et Verdier Henri, *L'âge de la multitude. Entreprendre et gouverner après la révolution numérique*, Armand Colin, 2015.

[84] O'Reilly Tim, "Government as a Platform", *Innovations: Technology, Governance, Globalization*, vol. 6, n° 1, pp. 13-40 ; Pezziardi Pierre et Verdier Henri, *Des startups d'État à l'État plateforme*, Fondation pour l'innovation politique, fondapol. org, janvier 2017.

Avec en arrière-plan le risque que si rien n'est fait, l'État et les services publics finissent à leur tour par se faire « ubériser » [85].

Les secteurs marchands comme ceux des postes et des transports ont déjà été investis par les acteurs numériques. Pour autant, les services non marchands ne sont pas immunisés. Le gouvernement français a récemment abandonné l'application nationale d'alerte attentat SAIP (Système d'alerte d'information des populations), au profit d'accords passés avec les plateformes numériques que sont Google, Facebook et Twitter. Google, via sa filiale Sidewalk Labs, ambitionne de devenir un acteur majeur dans la conception et la gestion de la ville intelligente. Les premières expérimentations de l'entreprise devraient d'ailleurs prochainement voir le jour à Toronto sur une friche portuaire de cinq hectares.

Face à cette tendance, plusieurs initiatives récentes ouvrent d'intéressantes perspectives de coproduction du service public entre le public et le privé fondées sur la donnée.

L'administration peut tout d'abord recourir aux ressources du *crowdsourcing* pour faire remonter les signaux faibles et orienter son action. L'Autorité de régulation des communications électroniques et des postes (ARCEP), le gendarme français des télécoms, s'est ainsi engagée dans un programme de « régulation par la data » qui entend « faire de chaque citoyen un micro-régulateur » [86]. La plateforme « j'alerte l'ARCEP » permet aux particuliers de signaler les dysfonctionnements qu'ils rencontrent dans leurs relations avec les opérateurs de télécommunications. Les alertes recueillies permettent à l'ARCEP d'identifier les dysfonctionnements récurrents et de gagner ainsi en efficacité dans ses actions de régulation. L'ARCEP s'est également associée aux applications mobiles de mesure des débits pour compléter les données qu'elle produit elle-même sur la couverture et la qualité des réseaux. Dans le même esprit, elle a soutenu le développement de l'application « Wehe » par la Northeastern University de Boston. L'objet de l'application est de permettre à ses utilisateurs de contribuer à la neutralité d'Internet en détectant les bridages de flux Internet et en faisant remonter les signalements positifs au régulateur. En 2017, l'État a également lancé la plateforme *cybermalveillance.gouv.fr*, qui implique les acteurs du secteur de la cybersécurité, notamment les entreprises du numérique. Destinée à aider les victimes de cybermalveillance, qu'il

[85] Bertholet Clément et Letourneau Laura, *Ubérisons l'État avant que d'autres ne s'en chargent*, Armand Colin, 2017.

[86] https://www.arcep.fr/la-regulation/grands-dossiers-thematiques-transverses/la-regulation-par-la-data.html

s'agisse d'acteurs publics, privés, ou d'individus, le dispositif devrait également à terme permettre d'enrichir les informations sur la nature de la cybercriminalité [87].

L'administration s'est par ailleurs engagée dans l'emploi des *data sciences* pour améliorer l'efficacité des politiques publiques. Plusieurs dispositifs contribuent à diffuser l'usage de la donnée au sein des ministères [88]. Le ministère des Finances a généralisé le recours au *data mining* pour lutter contre la fraude fiscale [89]. La Bonne Boîte de Pôle emploi permet aux demandeurs d'emplois d'accéder au marché caché de l'emploi. Le service utilise un algorithme de prédiction de la probabilité d'embauche des entreprises dans une région donnée. MapVHL, développé par la Direction interministérielle du numérique et du système d'information et de communication (DINSIC) et le Service des technologies et des systèmes d'information de la Sécurité intérieure (ST(SI)^2), est un outil d'aide à la décision qui permet de connaître l'historique des vols et des découvertes de véhicules.

L'État encourage également la constitution d'un écosystème autour de la production des services publics pour utiliser et favoriser l'innovation du plus grand nombre. Au-delà d'une simple mise à disposition des données publiques, il s'agit de nouer des partenariats et d'accompagner les innovateurs dans la création de services. C'est le sens de la mise à disposition des API d'État sur le site *api.gouv.fr* : les API s'adressent aux créateurs de services, qu'ils soient issus de la sphère publique ou de la sphère privée. Mais on peut également imaginer que l'administration noue des partenariats pour accéder aux données privées qu'elle juge utiles à l'activité de ses services publics. Une ville pourrait par exemple concentrer ses contrôles sanitaires sur les restaurants qui ont obtenu de mauvaises notes de leur clientèle sur des sites Internet comme TripAdvisor ou La Fourchette ou en fonction de la récurrence de certains termes sur le manque d'hygiène dans les commentaires.

[87] Untersinger Martin, « Plus de 36 00 victimes de 'cybermalveillance' assistées par la plateforme de l'Etat », *Le Monde*, 25 janvier 2019 https://www.lemonde.fr/pixels/article/2019/01/25/plus-de-36-000-victimes-de-cybermalveillance-assistees-par-la-plate-forme-de-l-etat_5414462_4408996.html

[88] Administrateur général des données, *La donnée comme infrastructure essentielle*, rapport annuel 2016-2017.

[89] Floc'h Benoît, « Bercy mise sur les nouvelles technologies pour traquer les fraudeurs au fisc », *Le Monde*, 3 janvier 2019 https://www.lemonde.fr/politique/article/2019/01/03/bercy-mise-sur-les-nouvelles-technologies-pour-traquer-les-fraudeurs-au-fisc_5404663_823448.html

Faire du citoyen un acteur à part entière de la donnée

"*You can see the computer age everywhere but in the productivity statistics*" [90]. C'est ainsi que l'économiste Robert Solow s'étonnait à la fin des années 1980 de la situation paradoxale de l'économie américaine. Malgré des investissements considérables réalisés dans les nouvelles technologies de l'information et de la communication, la croissance de la productivité était plus faible dans les années 1980 qu'elle ne l'avait été dans les années 1960 et 1970. L'une des explications régulièrement avancées pour expliquer le « paradoxe de Solow » est celle du décalage qui existe entre l'introduction d'une innovation technologique et son incidence effective sur les évolutions économiques et sociales. La diffusion des innovations technologiques, la formation d'individus aptes à s'en emparer, la réorganisation sociale autour de leur usage, tout cela prend du temps. Et de fait, dès les années 1990, la productivité américaine liée aux nouvelles technologies repartait à la hausse.

Le paradoxe de Solow peut être utile pour comprendre la situation actuelle. Si nul ne conteste que les perspectives ouvertes par l'exploitation des données soient immenses, il reste encore du chemin à parcourir avant que les entreprises, les administrations et les citoyens soient en capacité d'en tirer tous les bénéfices. D'après l'Organisation de coopération et de développement économiques (OCDE), les spécialistes des données représentent aujourd'hui moins de 1 % de l'emploi total parmi les États membres de l'organisation [91]. Faire émerger les compétences requises, c'est l'enjeu du développement de la littératie des données (de l'anglais *data literacy*) ou culture des données, qui se définit comme la capacité à produire, à comprendre et à utiliser les données numériques.

Si nous ne sommes pas tous destinés à devenir des *data scientists* aguerris, un travail pédagogique reste à faire pour permettre à chacun de lire les données, de les manipuler et de les mettre en récit. Au-delà de sa dimension économique, la littératie de la donnée est en enjeu démocratique crucial, puisqu'il en va de la faculté des citoyens à maîtriser leurs données et à peser sur les grands débats qui s'annoncent autour de la régulation des données [92].

[90] « Vous pouvez voir l'ère informatique partout, mais pas dans les statistiques de productivité. »

[91] OCDE, *Science, Technology and Innovation, Policy Note on Data-Driven Innovation for Growth and Well-being*, Édition OCDE, octobre 2015.

Dans les entreprises, la culture de la donnée demeure souvent faible. Se doter d'un responsable des données (en anglais CDO, *Chief Data Officer*) ou de *data scientists* n'est pas suffisant. La donnée ne doit pas rester une affaire de spécialistes, mais être partagée par le plus grand nombre dans l'organisation, depuis le top management jusqu'au client. Pour cela, il est nécessaire de sensibiliser les acteurs en renforçant l'information et en démystifiant la donnée, notamment à travers la création de « datalabs » et d'« infolabs » qui organisent le partage des données et informent les salariés sur les ressources disponibles, les questions d'éthique et de sécurité. La prise de conscience de la valeur des données passe également par l'expérimentation, via des *hackathons* [93], etc.

De même, les agents publics ne perçoivent pas toujours les opportunités qu'offre le numérique pour faire évoluer leurs méthodes et leurs missions. Faire figurer les opérations de collecte, d'entretien et de publication des données sur la fiche de poste des agents publics n'est pas suffisant. Les agents doivent comprendre concrètement les opportunités offertes par une exploitation des données dans le cadre de leur activité. Au-delà des offres classiques de formation, la mise en place de programmes comme celui des startups d'État ou des entrepreneurs d'intérêt général renforce la culture de la donnée au sein de l'administration, en suscitant des synergies avec les acteurs de la tech.

Vers l'acculturation de l'État plateforme

L'Administrateur général des données

En 2014, la France a créé la fonction d'administrateur général des données, chargé de coordonner l'action des administrations en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données de l'administration. Il organise la meilleure exploitation des données et leur plus large circulation, notamment aux fins d'évaluation des politiques publiques, d'amélioration et de transparence de l'action publique et de stimulation de la recherche et de l'innovation. Son rôle est aussi d'encourager l'emploi des datasciences au sein de l'administration. L'administrateur général des données peut s'appuyer sur les services de la mission Etalab et de la DINSIC, ainsi que sur le réseau des administrateurs ministériels des données nommés dans les différents ministères. L'administrateur général remet chaque

[92] Voir le chapitre sur les défis de la régulation des données.

[93] Le terme *hackathon* désigne un événement où des personnes se rassemblent pour faire de la programmation informatique collaborative, généralement sur plusieurs jours, autour d'une ou plusieurs problématiques.

année un rapport public sur son activité. Dans son rapport 2016-2017, intitulé *La donnée comme infrastructure essentielle*, l'administrateur général des données dresse l'état des lieux de la politique française des données et prône la création d'une infrastructure de la donnée.

Les startups d'État

« Vous vous reconnaissez dans le profil d'un agent public indigné par une situation irritante pour les usagers ? Vous avez une idée précise de la solution qui pourrait faire levier sur du numérique ? Vous êtes prêts à vous investir pour la mettre en œuvre ? ». C'est ainsi que le site beta.gouv.fr invite n'importe quel agent public à se transformer en intrapreneur. Lancé en 2013, le programme des startups d'État permet d'accompagner les innovateurs internes à l'administration dans la réalisation de leur projet et de mettre en place des expérimentations en dehors des canons traditionnels de l'administration.

Une startup d'État est composée d'une équipe de deux à quatre personnes (un chef de produit et des développeurs), financée par une administration pour apporter une réponse à un problème identifié. Tout cela, « pour moins de 200 000 euros et en moins de six mois », selon Henri Verdier, l'ancien directeur de la DINSIC. Par exemple, le projet Mes-aides permet à tout usager de connaître les aides sociales auxquelles il a droit à partir d'un questionnaire en ligne. La plateforme Le Taxi permet de géolocaliser les taxis. L'API Particuliers met en œuvre le principe du « Dites-le-nous une fois », en permettant aux acteurs publics de s'échanger les informations sur les usagers.

Les entrepreneurs d'intérêt général

Lancé en 2016, le programme « Entrepreneur.e d'Intérêt Général » vise à recruter des spécialistes du numérique pour répondre à des défis lancés par les ministères. Le principe est simple : une administration soumet un problème précis à des experts extérieurs à l'administration qui doivent lui apporter une réponse en dix mois. Le programme permet à l'administration de s'appuyer sur l'innovation ouverte en attirant à elle une communauté de développeurs et de *data scientists* appelée à travailler de concert avec les membres de l'administration.

Parallèlement, cette acculturation doit emprunter la voie du développement d'une expertise en traitement des données au sein même des administrations et du développement de programmes de régulation par la donnée. L'un des enjeux, déjà soulevé par Renaissance Numérique et la Fondation Jean Jaurès, tient toutefois à la difficulté pour l'administration

de recruter des experts de la donnée, compte tenu de la rareté des *data scientists* sur le marché du travail, et du fait que les salaires dans la fonction publique ne sont pas en mesure de rivaliser avec ceux qui sont proposés à ces profils dans le secteur privé [94]. L'État et les collectivités territoriales doivent débloquer les ressources financières nécessaires au recrutement des *data scientists* qui appliqueront le traitement des données au service des politiques publiques.

La littératie des données doit permettre aux individus et aux organisations de la société civile de s'impliquer dans les débats de la société de l'information [95]. Sous l'angle du contrôle démocratique, l'ouverture des données publiques n'a d'ailleurs de sens que s'il existe des citoyens dotés des compétences adéquates pour s'en saisir. Dès l'école, il faudrait mettre en place des programmes d'apprentissage numérique pour que les élèves apprennent à analyser et à manipuler les données. Plus largement, les acteurs du secteur numérique, tout autant que de nombreux acteurs de l'enseignement, appellent de longue date à faire entrer le numérique dans le socle commun de compétences de l'ensemble des Français, via la généralisation de l'apprentissage des « Humanités numériques » tout au long de la scolarité des élèves, au même titre que le français ou les mathématiques [96].

Une politique volontariste pourrait s'atteler à lutter contre la fracture numérique et à faire émerger une strate d'acteurs intermédiaires dans la société civile (journalistes, associations, etc.) capables de s'emparer des données et de construire des analyses pour les restituer aux citoyens [97].

Certains considèrent qu'un des grands chantiers qui s'annoncent est celui de la restitution des données générées par les utilisateurs. Si le règlement général sur la protection des données pourrait représenter en cela un premier pas, en clarifiant les droits des utilisateurs et en introduisant un droit à la portabilité des données personnelles, ils considèrent que l'approche juridique doit être complétée par la promotion de technologies respectueuses de la vie privée et par l'amélioration de l'information sur le traitement des données et leur portabilité. Les initiatives destinées à permettre aux utilisateurs de devenir acteurs de leurs données

[94] Renaissance Numérique et Fondation Jean Jaurès, Démocratie : le réenchantement numérique, Groupe de travail sur la citoyenneté numérique, 2017.

[95] CNNum, *Neutralité des plateformes : Réunir les conditions d'un environnement numérique ouvert et soutenable, op. cit*, p. 35.

[96] Syntec Numérique, *Cahier de Campagne « Éducation et Formation »,* janvier 2017.

[97] Isaac Henri, « La donnée numérique, bien public ou instrument de profit », *op. cit.*

personnelles se multiplient : Self Data en France, Midata en Angleterre, Smart Disclosure aux États-Unis, Cloud Personnel, Personal Data Store, Personal Information Management System (PIMS), etc. Il s'agit de rompre avec la logique dominante actuelle dans laquelle les utilisateurs produisent des données qui sont exploitées par d'autres acteurs, bien souvent en dehors de leur contrôle. Pour cela, il faut créer les outils qui permettent aux individus de récupérer la maîtrise de leurs données pour les utiliser à leurs propres fins.

Des appels au partage de données volontaires – données contributives – voient également le jour, à l'initiative de la société civile ou des pouvoirs publics, dans l'objectif de renforcer l'information de ces acteurs. Ces nouvelles formes de contribution à l'action publique participent du renouvellement de la relation des usagers-citoyens avec les institutions. Elles nous interrogent sur nos modes de gouvernance et la place que nous souhaitons confier à ces individus et acteurs, qui participent plus activement à la vie de la cité. Au-delà de la refonte du rapport démocratique, cette réflexion sur un rôle plus effectif dans la gouvernance permettrait aussi d'encourager ces nouvelles formes de contribution par une incitation plus grande.

Self Data, comment ça marche ?

Source : http://mesinfos.fing.org/selfdata-2/ *(extraits)*

D'après le Fing, le Self Data se définit comme : « la production, l'exploration et le partage de données personnelles par les individus, sous leur contrôle et à leurs propres fins ».

1. Un individu obtient ses données personnelles des organisations avec lesquelles il est en relation (aujourd'hui ces données sont dans les systèmes d'information des entreprises, difficilement obtenables par les individus) et en ajoute d'autres qu'il capte ou produit lui-même.
2. Il les stocke et les administre de manière sécurisée (par exemple à l'aide de PIMS – Personal Information Management Systems, comme les Cloud Personnels).
3. Et il les exploite à ses propres fins, à l'aide d'applications et de services tiers pour mieux se connaître, prendre de meilleures décisions, évaluer ses décisions passées, se faciliter la vie, etc.

2_ Relever les défis de cette nouvelle économie

L'ère de la donnée offre de nombreuses opportunités de développement économique et social, au bénéfice des citoyens comme des entreprises. Pour autant, accompagner son essor pose des défis aux pouvoirs publics, tant au niveau français qu'européen, qu'il s'agisse de la recomposition des chaînes de valeur, de l'adaptation des compétences à cette nouvelle donne, ou encore de la capacité à favoriser l'innovation. Et certains débats, déjà tranchés, resurgissent régulièrement, tel celui concernant une approche patrimoniale des données personnelles.

Dans ce contexte, une stratégie européenne forte est indispensable pour passer à l'échelle et permettre à tous d'appréhender pleinement les données et d'en devenir véritablement acteurs.

La patrimonialisation des données personnelles, une vieille antienne

Dans son livre *Who Owns the Future?* l'informaticien Jaron Lanier propose le postulat suivant : alors que l'économie des données génère des profits croissants, les individus qui sont à l'origine de ces données semblent laissés pour compte au moment du partage de la valeur[98]. La relation qui se noue entre certains services numériques et leurs utilisateurs repose sur un troc implicite : en échange de services gratuits, les utilisateurs abandonnent des informations à des entreprises qui peuvent les utiliser pour améliorer ces services et créer de la valeur.

Face à cette situation, certains courants estiment que la reconnaissance d'un droit de propriété sur les données personnelles pourrait être une solution pour réintégrer les utilisateurs dans la chaîne de valeur[99] et rompre avec ce qu'ils qualifient de travail gratuit[100]. La propriété permettrait également aux individus d'exercer un pouvoir de contrôle sur leurs données, en choisissant à qui ils veulent les vendre et à quelles fins.

[98] Lanier Jaron, *Who Owns the Future?*, Simon & Schuster, 2013.

[99] « Nos « données personnelles » nous appartiennent : monétisons-les ! », tribune publiée dans le journal *Le Monde*, 5 février 2018 ; Bensoussan Alain, « Pour un droit de propriété et une monétisation des données personnelles », *blog.lefigaro.fr,* 28 février 2018.

[100] Ibarra Immanol Arrieta, Goff Leonard, Hernandez Diego Jimenez, Lanier Jaron et Weyl E. Glen, "Should We Treat Data as Labor : Moving Beyond 'Free'", *American Economic Association Papers & Proceedings*, vol. 1, n°1, 2017.

Enfin, un tel régime pourrait avoir pour effet de faciliter la circulation des données et leur commercialisation dans le cadre de marchés qui peinent aujourd'hui à se développer faute d'un régime de propriété clairement défini.

Face aux atouts de la création d'un droit de propriété avancés par ses partisans, la patrimonialisation des données personnelles présente cependant un certain nombre de difficultés d'ordre juridique et économique.

Rappelons d'abord que le droit européen a déjà tranché ce débat : les données personnelles étant directement rattachées à la personne physique, elles sont un prolongement de la personne. De la même façon que nous ne pouvons être propriétaires d'une personne humaine ou vendre un de nos organes, nous ne pouvons commercer sur ces attributs et ne cédons à des tiers que des droits d'usage de nos données.

Par ailleurs, les données n'ont pas de valeur intrinsèque ou alors infinitésimale (voir chapitre 1). La valeur des données réside dans leur agrégation et leur analyse. Si les fournisseurs de données sont des acteurs indispensables dans la chaîne de valeur, la valeur des données réside en grande partie dans les autres étapes. Attribuer une juste part de la valeur au fournisseur d'une seule donnée en amont de la chaîne est dès lors relativement délicat, sauf à considérer que la part qui lui revient est tout simplement insignifiante.

Enfin, même au regard de l'objectif de circulation des données, la création d'un droit de propriété des données personnelles pourrait finalement s'avérer contre-productive [101]. En effet, en complexifiant les transactions, un régime de propriété risque de conduire à une limitation des échanges.

Les données, un bien commun ?

La logique des « communs » offre un modèle alternatif de production et de gouvernance des données [102]. Popularisés par les travaux de l'économiste Elinor Ostrom, les communs désignent alors des ressources partagées, gérées et maintenues collectivement par une communauté qui possède un intérêt direct dans leur protection. Historiquement, la notion trouve

[101] CNNum, Avis du Conseil sur la libre circulation des données dans l'Union européenne, Avril 2017.

[102] Peugeot Valérie, « Les Communs, une brèche politique à l'heure du numérique », in (dir.) Carmes et Noyer Maryse Jean-Max, *Les débats du numériques*, Presses des Mines, 2013 pp. 77-98 https://books.openedition.org/pressesmines/1663

son origine dans l'existence de pâturages, de forêts ou de rivières situés à proximité des villages, utilisables par tous les membres d'une communauté. La puissance publique, pas plus qu'un particulier ou que toute autre entité de droit moral, ne peut en revendiquer l'exclusivité. C'est aux membres de la communauté qu'il appartient de définir ensemble un certain nombre de droits d'usages et d'obligations de protection de la ressource.

À travers la multiplication des pratiques de partage, de co-construction et d'échanges de pair à pair, l'essor des technologies numériques ouvre l'espace pour un renouveau des communs. Dès les années 1980, les entreprises du logiciel libre ont ainsi organisé le partage des codes informatiques au sein de communautés d'utilisateurs qui occupent en même temps le rôle de coproducteurs. À partir des années 2000, ces « communs informationnels » prennent également la forme de contenus coproduits par les utilisateurs, dont l'encyclopédie en libre accès Wikipedia reste sans doute l'exemple le plus emblématique. Des plateformes de partage des données voient également le jour, alimentées par une communauté de contributeurs, comme la base de données cartographique *Open Street Map* ou la base de données sur les produits alimentaires *Open Food Facts*. Ces initiatives peuvent avoir une vocation scientifique, à l'instar du « projet génome humain », dans lequel une vingtaine d'institutions partagent leur travail sur le séquençage du génome. Aux Pays-Bas, le projet de recherche collaborative iSPEX a associé plus de 10 000 utilisateurs d'iPhone à la mesure des particules de poussière dans l'atmosphère.

La liberté d'accès et l'usage non exclusif des communs informationnels sont garantis par l'utilisation de licences ouvertes de type GPL (*General Public Licence*) pour les logiciels et de type Creative Commons pour les contenus. L'ouverture des données a donné lieu à la création de plusieurs licences spécifiques [103]. L'*Open Knowledge Foundation* au Royaume-Uni a été la première à se saisir du sujet, dès 2007, et à élaborer un jeu de licences, dont la plus populaire est l'OdBL (*Open database licence*), qui contient une clause de viralité imposant au réutilisateur de partager la base de données enrichie sous la même licence. Les licences *Creative Commons* ont été adaptées en 2009 sous forme de licence CC0. En France, les administrations centrales recourent en principe à la « licence ouverte » élaborée par la mission Etalab.

[103] Peugeot Valérie, « Données contributives ou ascendantes », *in* (dir.) Orsi Fabienne, Rochfeld Judith et Cornu-Volatron Marie, Dictionnaire des biens communs, PUF, 2017

La question de la création d'un cadre juridique pour les communs a fait l'objet de débats en France au moment de la discussion sur la loi pour une République numérique en 2016, entre les tenants de cette approche et ceux qui craignaient une remise en cause du droit d'auteur. Au-delà, l'enjeu premier est celui de la gouvernance de ces données : fabriquer des communs demande des infrastructures adéquates et des investissements dans la maintenance des jeux de données et interroge sur les logiques de leur réutilisation.

Vers de nouveaux outils anti abus de position dominante

En France et en Europe, certaines autorités de la concurrence s'interrogent sur le fait de savoir si la possession de données constitue en soi une barrière à l'entrée ou un abus de position dominante. Début avril 2019, la Commission européenne a proposé un cadre de réflexion pour les sujets de la concurrence à l'ère du numérique dans lequel la question des données fait l'objet d'une réflexion approfondie. Elle met notamment en évidence la difficulté qu'il y a à bâtir un cadre global et prône une analyse au cas par cas [104].

En effet, l'activité des grandes plateformes numériques défie les analyses anticoncurrentielles traditionnelles [105]. D'abord, parce que l'exploitation des données numériques est à l'origine d'importants bénéfices pour le consommateur des services que ces plateformes proposent : les prix sont en baisse, les services sont gratuits et de qualité, l'innovation est rapide. Ensuite, parce que la détention d'importantes bases de données peut difficilement être considérée à elle seule comme un facteur de distorsion de concurrence : les données sont non rivales, facilement réplicables et ont une durée de vie limitée. En particulier, la théorie des facilités essentielles est difficilement mobilisable en raison de l'existence de données de substitution dans la majorité des cas.

Or, la valeur des données tient à la capacité d'agrégation, de traitement et au final de production de services. Dans cette perspective, ce n'est pas la possession d'un grand volume de données qui constituerait en soi un abus de position dominante.

[104] Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitzer, Commission européenne, « Competition policy for the digital era », 2019. HTTP://EC.EUROPA.EU/COMPETITION/PUBLICATIONS/REPORTS/KD0419345ENN.PDF

[105] Autorité de la concurrence et Bundeskrtellamt, *Droit de la concurrence et données, op. cit.*

Plus largement, la plateformisation de l'économie amène des réflexions sur une adaptation des règles de la concurrence à cette nouvelle donne, la transformation des chaînes de valeur questionnant les approches économiques traditionnelles.

Face à la compétition mondiale, porter l'ambition européenne

L'Europe s'est engagée avec retard dans la révolution de la donnée. En 2015, les spécialistes de la donnée y représentaient moins de 1% de la part totale de l'emploi [106]. Par ailleurs, seulement 6 % des entreprises de services aux professionnels de l'information et des télécommunications faisaient une utilisation stratégique et intensive des données. Contrairement aux États européens, la Chine et la Russie ont par exemple adopté une politique volontariste pour faire émerger leurs propres champions numériques [107].

Dans le cadre de la mandature qui s'achève, la Commission européenne a fait de la réalisation du marché unique numérique une de ses priorités, avec plusieurs réussites à son actif, comme la création d'un cadre à la fois solide et ambitieux pour la protection des données personnelles, dont d'autres États à travers le monde disent vouloir s'inspirer, comme le Japon. De même, la libre circulation des données au sein de l'Union européenne est une réalité pour les entreprises. Pour autant, il reste encore beaucoup à faire au niveau européen pour ne pas rester sur des marchés éclatés et entrer pleinement dans l'ère de la donnée. En effet, si l'Union européenne a de longue date amorcé une réflexion sur les enjeux éthiques de l'exploitation des données, réflexion aujourd'hui prolongée par l'élaboration de recommandations sur un usage éthique des technologies d'intelligence artificielle, elle n'a pas encore su construire une stratégie industrielle commune en matière de données.

Les institutions européennes déploient actuellement un cadre pour accompagner leur stratégie industrielle autour de l'innovation, intégrant la circulation des données. Mais le travail est loin d'être achevé.

Le règlement *ePrivacy* offre un bon exemple de la complexité de certaines approches de régulation européennes, allant jusqu'à la confrontation entre textes européens avec le RGPD. Dans le sillage de ce dernier, le règlement *ePrivacy* vise en effet à actualiser les règles en vigueur en matière de

[106] Commission Européene, "Enter the Data Economy: EU Policies for a Thriving Data Ecosystem", *EPSC Strategic Notes*, numéro 2, 11 janvier 2017, p. 4.

[107] Villani Cédric, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, rapport issu de la mission confiée par le Premier ministre, 2018.

respect de la vie privée et de la protection des données personnelles, en étendant le champ d'application à l'ensemble des fournisseurs de services de communications électroniques, et en encadrant plus strictement les échanges d'informations qui transitent au sein de ces fournisseurs. Du fait de son acception large, le texte vise également les communications *machine-to-machine* (M2M), ce qui reviendrait à inclure les systèmes industriels embarqués (automobile, énergie, etc.) et les systèmes applicatifs (objets connectés, robotique, dispositifs médicaux, cybersécurité, etc.). Tant les données personnelles que non personnelles recueillies seraient ainsi concernées, soumettant aux nouvelles exigences en matière de confidentialité, de consentement, etc., des données telles que l'humidité de l'air captées dans le champ d'un agriculteur. Cela créerait de fait des impasses. En particulier, l'impossibilité d'identifier des utilisateurs finaux dans de nombreux cas empêcherait de recueillir leur consentement et de pouvoir procéder à un traitement des données à des fins analytiques.

L'Europe doit ainsi être à l'initiative de nouvelles modalités de gouvernance, plus agiles, afin de créer un terreau favorable à l'innovation dans le respect des citoyens et consommateurs. Elle pourrait notamment s'inspirer des « bacs à sable réglementaires » mis en place par certaines autorités régulatrices nationales, à l'instar de l'ICO au Royaume-Uni.

Mais cette stratégie européenne ne peut reposer que sur une interrogation de son cadre juridique. L'Europe, et en particulier la France, pâtit encore de la faiblesse de l'accompagnement de ses TPE-PME vers cette culture de la donnée, et plus globalement numérique. Au-delà de la quête vers les futurs champions numériques – indispensable, mais pas toujours bien définie -, elle doit renforcer ses politiques de transfert de l'innovation dans les différents secteurs de l'économie. Cela passe également par une politique ambitieuse d'accompagnement des travailleurs et futurs travailleurs à l'ère de la donnée, dès le plus jeune âge et tout au long de leur vie : démultiplication des formations spécialisées en numérique dans l'enseignement supérieur, dont l'excellence est déjà reconnue dans le monde entier, mais aussi amplification du soutien aux programmes d'adaptation des compétences (*digital skills*) des professionnels déjà en exercice. Dans la compétition mondiale, l'attractivité de la recherche est également un défi majeur pour les États membres, avec une fuite des cerveaux déjà bien enclenchée [108].

[108] Renaissance Numérique, EuropaNova, Das Progressive Zentrum, « Intelligence artificielle : bâtir la voie européenne », octobre 2018.

Enfin, cette politique des données ne doit pas se résumer pas à cette seule dimension économique et doit veiller à la préservation des valeurs fondamentales de l'Union européenne. Le cadre réglementaire, et notamment les dispositifs d'accès aux données, doit être lisible, permettre de protéger les citoyens européens, et favoriser l'innovation de tous les acteurs.

La nouvelle mandature européenne est ainsi une occasion à ne pas manquer si nous souhaitons que tous les Européens deviennent pleinement acteurs des données.

BIBLIOGRAPHIE

Ackoff Russell, "From Data to Wisdom", in *Ackoff's Best*, John Wiley & Sons, New York, 1989, p. 170.

Administrateur général des données, *La donnée comme infrastructure essentielle*, rapport annuel 2016-2017.

Arcep, *Smartphones, tablettes, assistants vocaux : Les terminaux, maillon faible de l'ouverture d'Internet, Rapport sur leurs limites et sur les actions à envisager*, février 2018.

Autorité de la concurrence et Bundeskartellamt, *Droit de la concurrence et données*, 10 mai 2016.

Bertholet Clément et Letourneau Laura, *Ubérisons l'État avant que d'autres ne s'en chargent*, Armand Colin, 2017.

Boston Consulting Group, en collaboration avec le Liberty Global, *The Value of our Digital Identity*, novembre 2012.

Brynjolfsson Erik, "Why it Matters that the GDP Ignores Free Goods", Techonomy 2012, Conference Report, 7 décembre 2012.

Capgemini Consulting, *Intelligences Artificielles dans les services publics : le moment d'accélérer*, 2017.

Cardon Dominique, « Le pouvoir des algorithmes », *Pouvoirs*, n°164, p. 63-73.

Chignard Simon et Benyayer Louis-David, *Datanomics, les nouveaux business models des données*, éditions FYP, 2015.

Chignard Simon, « J'ai testé les données 'ouvertes' d'Uber et Airbnb en France », blog « données ouvertes », 17 décembre 2017 https://donneesouvertes.info/2017/12/17/jai-teste-les-donnees-ouvertes-duber-et-airbnb-en-france/

Chignard Simon, « "Safari ou la chasse aux Français, 40 ans après" », blog « données ouvertes », 26 janvier 2018 https://donneesouvertes.info/2018/01/26/safari-la-chasse-aux-francais-40-ans-apres/

Citigroup - Oxford Martin School, *Technology at Work: The Future of Innovation and Employment*, University of Oxford, février 2015.

Cigref, *Valorisation des données dans les grandes entreprises. Maturité, pratiques et modèles*, rapport, novembre 2016.

CNNum, *Neutralité des plateformes : Réunir les conditions d'un environnement numérique ouvert et soutenable*, mai 2014.

CNNum, *Avis du Conseil national du numérique sur la libre circulation des données en Europe*, avril 2017.

Colin Nicolas et Verdier Henri, *L'âge de la multitude. Entreprendre et gouverner après la révolution numérique*, Armand Colin, 2015.

Collin Pierre et Colin Nicolas, *Mission d'expertise sur la fiscalité de l'économie numérique*, rapport pour le ministère de l'Economie et des Finances et le ministère du Redressement Productif, 2015.

Commission Européenne, « Créer une économie européenne fondée sur les données », Communication, COM(2017) 9 final, 10 janvier 2017.

Commission Européenne, "Enter the Data Economy: EU Policies for a Thriving Data Ecosystem", *EPSC Strategic Notes*, numéro 2, 11 janvier 2017.

Commission Européenne, « Vers un espace européen commun des données », communication, COM(2018) 232 final, Bruxelles, le 25 avril 2018.

Cytermann Laurent, « Promesses et risques du big data et de l'open data », *Informations sociales*, 2015/5, n° 191, p. 82.

Deloitte Digital, *Consumer Experience in the Retail Renaissance: How Leading Brands Build a Bedrock with Data*, 2018.

Duchesne Claudine, Cytermann Laurent, Vachey et al., *Rapport relatif aux données d'intérêt général*, Rapport du Conseil général de l'économie, de l'industrie, de l'énergie et des technologies, du Conseil d'État et de l'Inspection général des finances, septembre 2015.

English Larry, *Information Quality Applied: Best Practices for Improving Business Information Processes*, Wiley Publishing Inc, 2009.

Federal Trade Commission, *Data Brokers. A Call for Transparency and Accountability*, mai 2014.

Fondation Jean Jaurès et Renaissance Numérique, *Démocratie : le réenchantement numérique*, Groupe de travail sur la citoyenneté numérique, 2017.

Gregory Adrian "Data Governance – Protecting and Unleashing the Value of your Customer Data Assets", *Journal of Direct, Data and Digital Marketing practice*, vol. 12, n° 3, 2011, pp. 230-248.

Hartmann Philipp Max *et al.*, "Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms", working paper, University of Cambridge.

Ibarra Immanol Arrieta, Goff Leonard, Hernandez Diego Jimenez, Lanier Jaron et Weyl E. Glen, "Should We Treat Data as Labor: Moving Beyond 'Free'", *American Economic Association Papers & Proceedings*, vol. 1, n°1, 2017.

Instituts Techniques Agricoles et Acta, *L'accès aux données pour la Recherche et l'Innovation en Agriculture*, positions, 2016.

International Data Corporation, *Data Age 2025*, parrainé par Seagate, 2017.

Isaac Henri , « Plateforme et dynamiques concurrentielles », Renaissance Numérique, octobre 2015.

Isaac Henri, « Données, valeur et *business model* », *Les Cahiers Scientifiques*, n°21, 2016, p. 18.

Isaac Henri, « La donnée, une marchandise comme les autres ? », *Enjeux numériques*, n°2, juin 2018, p. 20.

Isaac Henri, « La donnée numérique, bien public ou instrument de profit », *Pouvoirs*, n° 164, 2018, p. 75-86.

Lanier Jaron, *Who Owns the Future?* Simon & Schuster, 2013.

Mandel Michael, "Beyond Goods and Services. The (Unmeasured) Rise of the Data-Drive Economy", *Progressive Policy Institute*, Policy Memo, octobre 2012.

Marzloff Bruno, *Sans bureau fixe*, FYP éditions, 2013.

McKinsey Global Institute, *Big data: The next frontier for innovation, competition and productivity*, mai 2011.

Missika Jean-Louis, Faivret Jean-Philippe, « Informatique et libertés », *Les Temps Modernes*, septembre-octobre 1977.

Morin-Desailly Catherine, *L'Union européenne, colonie du monde numérique ?* Rapport d'information du Sénat, 2013.

OCDE, *Council Recommendation for Enhanced Access and More Effective Use of Public Sector Information*, 2008.

OCDE, "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value", *Documents de travail de l'OCDE sur l'économie numérique*, n° 220, Éditions OCDE, 2013.

OCDE, *Data-driven Innovation: Big Data for Growth and Well-Being*, Éditions OCDE, 2015.

OCDE, « La personnalisation des prix à l'ère numérique », note de référence, 28 novembre 2018.

OCDE, *La prochaine révolution de la production : Conséquences pour les pouvoirs publics et les entreprises*, Éditions OCDE, 2018.

O'Reilly Tim, "Government as a Platform", *Innovations: Technology, Governance, Globalization*, vol. 6, n° 1, p. 13-40.

Peugeot Valérie, « L'ouverture des données publiques : convergence ou malentendu ? », in (dir.) Stiegler Bernard, *Confiance, croyance, crédit dans les mondes industriels*, éditions FYP, 2011.

Peugeot Valérie, « Les Communs, une brèche politique à l'heure du numérique », in (dir.) Carmes et Noyer Maryse Jean-Max, *Les débats du numérique*, Presses des Mines, 2013 p. 77-98.

Peugeot Valérie, « Données contributives ou ascendantes », in (dir.) Orsi Fabienne, Rochfeld Judith et Cornu-Volatron Marie, *Dictionnaire des biens communs*, PUF, 2017.

Pezziardi Pierre et Verdier Henri, *Des startups d'État à l'État plateforme*, Fondation pour l'innovation politique, janvier 2017.

Polton Dominique, Cuggia Marc et Wainrib Gilles, *Health Data Hub : Mission de préfiguration*, 2018.

Renaissance Numérique, *Démocratie : mise à jour. 13 propositions pour une version améliorée de l'État, sa posture et son équilibre démocratique*, avril 2016.

Renaissance Numérique, *La valeur des données en agriculture*, février 2018.

Renaissance Numérique, *Données personnelles : rendre ces droits effectifs*, mai 2018

Rotman David, "How technology is destroying jobs", *MIT Technology Review*, juin 2013.

Trojette Mohammed Adnène, *Ouverture des données publiques. Les exceptions au principe de gratuité sont-elles toutes légitimes ?* Rapport au Premier ministre, juillet 2013.

Tucker Darren S. et Welford Hill B, 2014 "Big Mistakes Regarding Big Data", Antitrust Source, American bar Association, décembre 2014.

Vachey Laurent, Duchesne Claudine, Meyer Marc *et al.*, *Les données d'intérêt général : Phase 2*, rapport de l'Inspection générale des finances et du Conseil général de l'économie, de l'industrie, de l'énergie et des technologies, mars 2016.

Villani Cédric, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, rapport issu de la mission confiée par le Premier ministre, 2018.

REMERCIEMENTS

Nous remercions les différents acteurs qui ont participé aux auditions :

- Guillaume Buffet, *président de U Change*
- Paul Bounaud, *responsable des filières agricoles et RHD, GS1France*
- Sabine Bruaux, *chargée de mission au bureau de l'économie de la donnée, Direction générale des Entreprises, ministère de l'Économie et des Finances*
- Simon Chignard, *data editor, mission Etalab*
- Laurent Cytermann, *rapporteur public, Conseil d'État*
- Nicolas Deffieux, *rapporteur adjoint, Autorité de la concurrence*
- Jean-Baptiste Demaison, *conseiller principal en sécurité numérique auprès du directeur de la stratégie, Agence nationale de la sécurité des systèmes d'information (ANSSI), président du conseil d'administration de l'Agence européenne de la sécurité des réseaux et de l'information (ENISA)*
- Étienne Drouard, Avocat, *Cabinet K&L Gates*
- Georges-Etienne Faure, *directeur de l'innovation et des nouveaux métiers, Nexity*
- Mathieu Guennec, *rapporteur, Autorité de la concurrence*
- Faÿçal Hafied, *adjoint au chef de bureau « activités tertiaires et concurrences », Direction générale du Trésor*
- Jean-Marc Lazard, *cofondateur et directeur général, OpenDataSoft*
- Sophie Le Pallec, *responsable Relations Institutionnelles, GS1France*
- Laura Letourneau, *cheffe de l'unité « Internet ouvert », Autorité de* régulation des communications électroniques et des postes (ARCEP)
- Laure Lucchesi, *directrice de la Mission Etalab*
- Pierre Musseau, *conseiller en charge de l'urbanisme et de la ville intelligente, Ville de Paris*
- Bertrand Pailhès, *coordonnateur national de la stratégie d'intelligence artificielle, Direction interministérielle du numérique et du système d'information (DINSIC)*
- Valérie Peugeot, *chercheuse à l'Orange Labs, membre de la Commission nationale informatique et liberté (CNIL)*
- Etienne Pfister, *chef économiste, Autorité de la concurrence*
- Marine Pouyat, *consultante, W Talents*
- Perica Sucevic, *conseiller juridique, Direction interministérielle du numérique et du système d'information (DINSIC)*
- Laurent Toustou, *chef de l'unité « Régulation par la data », Autorité de régulation des communications électroniques et des postes (ARCEP)*
- Carole Vachet, *conseillère innovation et économie, cabinet du secrétaire d'État chargé du Numérique*
- Cédric Villani, *député*

Direction de la publication
Godefroy de Bentzmann, *président de Syntec Numérique*
Henri Isaac, *président de Renaissance Numérique*

Coordination
Jennyfer Chrétien, *déléguée générale de Renaissance Numérique*
Raphaëlle Frija, *déléguée aux usages du numérique de Syntec Numérique*

Rapporteur
Damien Elkind

Renaissance Numérique est le principal think tank français indépendant dédié aux enjeux de transformation numérique de la société. Réunissant des universitaires, des associations, des grandes entreprises, des start-ups et des écoles, il vise à élaborer des propositions opérationnelles pour accompagner les acteurs publics, les citoyens et les acteurs économiques dans la promotion d'une société numérique inclusive.
www.renaissancenumerique.org

Syntec Numérique est l'organisation professionnelle des entreprises de services du numérique (ESN), des éditeurs de logiciels et des sociétés de conseil en technologies. Il regroupe plus de 2000 entreprises adhérentes qui réalisent 80 % du chiffre d'affaires total du secteur (plus de 56 Md€ de chiffre d'affaires, 474 000 employés dans le secteur). Il compte 30 grands groupes, 120 ETI, 1000 PME, 850 startups et TPE ; 11 délégations régionales. Syntec Numérique est présidé par Godefroy de Bentzmann depuis juin 2016.
www.syntec-numerique.fr

Tous droits réservés © Renaissance Numérique et Syntec Numérique.
Achevé d'imprimer en mai 2019 par Brakisto.
Dépôt légal en cours.

La mise en données du monde est une réalité qui n'est pas toujours bien maîtrisée, y compris par ceux qui doivent lui en donner un cadre. Or, face à son accélération, son appropriation par tous - organismes publics, entreprises, société civile - devient un enjeu crucial pour le devenir de notre économie et de notre société. À l'heure où l'Europe entame la deuxième mandature du Marché unique numérique (*Digital Single Market*), avec un certain nombre de débats sur les données à son agenda, le think tank Renaissance Numérique et l'organisation professionnelle Syntec Numérique ont décidé de réunir leurs expertises pour offrir des clés de compréhension dans ce nouveau contexte.

RENAISSANCE numerique

renaissancenumerique.org

syntec numérique

syntec-numerique.fr