

« Données d'innovation »

Quelle intervention de l'Etat sur le partage de données ?

Synthèse

*Antoine Berthier,
Ingénieur du corps des
Mines, diplômé de
l'École polytechnique.
A travaillé pour un
grand groupe
autoroutier, dans un
fonds d'investissement
et dans la
cybersécurité.*

*Tristan Fulchiron,
Ingénieur du corps des
Mines, diplômé de
l'École polytechnique.
A travaillé dans un
fonds d'investissement,
une grande banque
française, et une start-
up d'objets connectés
en santé.*

10 octobre 2019

Ces dernières années, les entreprises ont en majorité pris conscience de la valeur que pouvaient représenter leurs données dans un contexte de transformation numérique de leurs métiers. Elles ont cependant assez peu tenté de valoriser leurs données auprès d'autres entreprises, si ce n'est pour les marchés du ciblage marketing et de la publicité en ligne.

Les modèles économiques de partage de données entre entreprises constituent une tendance qui s'affirmera dans les prochaines années. La Commission européenne a d'ailleurs communiqué sur le sujet en soulignant les opportunités pour l'économie numérique, et en en faisant une priorité pour la prochaine mandature¹. Le rapport Villani sur la stratégie nationale en intelligence artificielle (IA), rendu au gouvernement en mars 2018, présente quant à lui la mutualisation de données pour l'IA comme un enjeu de souveraineté.

Que ce soit pour croiser des informations de sources différentes, pour atteindre une masse critique de données nécessaire aux développements technologiques, ou pour créer des externalités non attendues avec du partage trans-sectoriel, le partage de données entre entreprises doit être anticipé dès aujourd'hui.

¹ « Commission's digital department pitches tech priorities to Vestager, Goulard », www.politico.eu, 23 septembre 2019.

Cependant, beaucoup d'entreprises sont encore réticentes, et, selon une étude rendue à la Commission européenne², seules 6,3 % d'entre elles déclarent s'engager activement dans du partage de données. Nous constatons que les entreprises préfèrent développer elles-mêmes les services à valeur ajoutée s'appuyant sur ces données, ou, dans le pire des cas, ne pas les valoriser à l'extérieur.

Nous pointons du doigt plusieurs raisons fondamentales qui, au-delà de l'immaturation des organisations sur ce sujet encore nouveau, expliquent ces difficultés. Devant des usages émergents, multiples et dont la rentabilité est difficile à estimer *a priori*, mesurer la valeur des données relève de l'impossible. En outre, les risques de se faire concurrencer sur son marché ou les coûts de développement d'interopérabilité associés viennent souvent finir de dissuader les entreprises de réutiliser leurs données.

Si ces barrières expliquent le manque de circulation des données, il faut alors s'interroger sur la façon d'y remédier afin de promouvoir la croissance de l'économie numérique en Europe. En science économique, on s'accorde à dire que la puissance publique doit pallier les déficiences de marché pour maintenir des incitations à créer de la connaissance, qui sera à terme utile au plus grand nombre. Cette intervention s'opère par exemple *via* la recherche publique ou des subventions aux entreprises. Un parallèle peut être dressé avec le partage de données, sujet que nous pensons être un terrain propice à l'intervention de la puissance publique.

Une politique publique adaptée aux enjeux de réutilisation consisterait à promouvoir des ouvertures de données privées à fort potentiel. Ces ouvertures de données s'appliqueraient à la classe des « traces numériques », ces données générées automatiquement et en grande quantité par l'utilisation des machines, et pour lesquelles les régimes de propriété intellectuelle ne peuvent pas s'appliquer. Par exemple, les données générées par une voiture connectée ne sont *a priori* ni la possession du constructeur ni celle du conducteur³ – cette question se réglant au cas par cas dans les contrats.

² DG Connect, « Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability », European Commission, 2018, p. 31.

³ « Who Owns the Data Generated by your Smart Car? », *Harvard Journal of Law & Technology*, 2018.

L'État se positionnerait alors comme un facilitateur qui interviendrait en faveur de droits d'accès plus larges de ces données possédées par les entreprises. Ces ouvertures permettront à des entreprises innovantes d'accéder à de nouvelles données et de proposer de nouveaux biens et services. Il s'agira de reproduire l'exemple probant de la directive européenne DSP2. Entrée en vigueur en septembre 2019, elle prévoit notamment que les clients particuliers de banques européennes puissent réutiliser leurs données de consommation bancaire. Ces données sont alors valorisées par des start-ups dites « fintech », qui créent des services nouveaux tels que l'analyse automatisée de dépenses, des cartes de fidélité dématérialisées, etc.

L'État ne s'étant par le passé que peu aventuré sur le terrain de la législation sur la propriété des données, il dispose de peu de leviers lui permettant d'atteindre cet objectif. La portabilité introduite par le RGPD (Règlement général sur la protection des données), qui redonne le contrôle de ses données au consommateur, est un moyen détourné mais efficace d'introduire du partage. Il ne peut cependant s'appliquer qu'aux données personnelles. Le droit de la concurrence a également permis par le passé des ouvertures ciblées de bases de données, mais n'est pas l'outil adapté pour un passage à l'échelle.

Nous proposons d'introduire un nouvel instrument réglementaire : les données d'innovation. Les données d'innovation sont des classes de données d'entreprises qui représentent un fort potentiel d'innovation et pour lesquelles l'État décide d'intervenir légalement en faveur d'un élargissement de leur accès à d'autres groupes d'acteurs. Ce nouveau concept permettra de légitimer l'action de l'État dans sa position de facilitateur du marché du partage de données. Les autorités sectorielles auront la responsabilité de définir les données d'innovation en statuant au cas par cas sur des ouvertures ciblées de données. Elles auront pour mission de préciser les points suivants : quel périmètre d'entreprises est concerné et quelles classes de données devront être échangées ? Quels standards de moyens d'échange devront être employés ? Sous quelles conditions les entreprises forcées d'ouvrir leurs données pourront-elles être rétribuées ?

INTRODUCTION

Une fois n'est pas coutume, la France peut se targuer d'être parmi les leaders mondiaux sur un sujet lié au numérique. En effet, selon une étude de l'OCDE⁴, la France se classe en deuxième position des pays les plus performants en termes d'accès aux données publiques, derrière la Corée du Sud. L'« open data » public à la française est donc parmi ce qui se fait de mieux au monde. La meilleure illustration est sans doute l'ouverture récente par la Direction générale des finances publiques des données de « demande de valeurs foncières », c'est-à-dire des historiques de transactions immobilières qui doivent faciliter, pour les citoyens, l'exercice de détermination de la valeur d'un bien immobilier.

Un rapport fondateur sur les « données d'intérêt général », remis par Laurent Cytermann au ministre de l'Économie Emmanuel Macron en 2015, a permis d'accélérer ce processus d'ouverture⁵.

Ce rapport recommandait d'inscrire dans la loi l'ouverture de principe des « données essentielles » des concessions ou des délégations de service public. Il préconisait également pour la première fois que l'État accède aux données possédées de manière privée « à condition que cette ouverture soit justifiée par des motifs d'intérêt général ». Un exemple emblématique de partage de données du privé avec la puissance publique est le programme Connected Citizens⁶ de Waze (l'entreprise en charge de l'application de navigation GPS du même nom). Comptant une trentaine de villes partenaires en France, il permet aux municipalités d'accéder aux données de trafic produites par les utilisateurs de Waze, par exemple pour tester des modifications de voirie ou pour transmettre les informations d'accident plus rapidement.

Nous pensons que la France peut aller plus loin et devenir leader mondial sur le sujet d'accès aux données possédées par les entreprises.

Alors qu'un moment politique privilégié se dessine en Europe et aux États-Unis pour la régulation des GAFA (taxe « GAFA », responsabilité des plateformes dans la lutte contre les *fake news*, enquêtes de la Commission européenne et de procureurs généraux

⁴ OECD OURdata, « Index on Open Government Data 2017 », cité par l'étude « Maturité digitale : la France est à l'avant du peloton dans la compétition mondiale mais doit accélérer pour rejoindre le groupe de tête » du Boston Consulting Group et du Medef, juin 2019.

⁵ « Rapport relatif aux données d'intérêt général », 2015.

⁶ <https://www.waze.com/fr/ccp>

américains contre les comportements anti-concurrentiels), ce rapport invite à ne pas envisager le rôle de la puissance publique uniquement au prisme des géants du numérique.

En effet, la transformation numérique des entreprises touche tous les secteurs de manière transverse. Il est donc crucial de fournir, en réponse à cette transformation massive, des politiques publiques adaptées qui permettront à chacun des acteurs de se saisir de nouvelles opportunités. Les exemples cités précédemment d'ouverture de données administratives ou de partenariats des villes avec des entreprises privées pour du partage de données montrent d'ailleurs l'étendue de la palette d'action de l'État sur le sujet.

Le concept de « données d'innovation », que nous introduisons dans cette note doit répondre à ce besoin en permettant à l'État de produire dans les prochaines années, de manière progressive, des politiques publiques sectorielles différenciées facilitant la circulation de données identifiées comme porteuses d'innovation entre les entreprises.

1. LE PARTAGE DE DONNÉES ENTRE ENTREPRISES

1.1. QU'EST-CE QUE « LES DONNÉES » ?

S'il est désormais communément admis que les données représentent pour les entreprises une source de valeur importante, il n'est jamais aisé d'en donner une définition claire et concise. En effet, sous le terme « données » se regroupent des réalités aussi diverses que des images satellites, un registre de transaction de cartes bancaires, la base des remboursements médicaux de la Cnam ou encore l'enregistrement du fonctionnement d'un moteur d'avion.

On le voit, la donnée est présente dans tous les secteurs et trouve une multitude d'utilisations différentes. Il ne faudrait d'ailleurs à ce titre pas parler de « la donnée » mais bien « des données ». Cette diversité incite à ne pas considérer les données comme un bien classique. Au contraire, le spectre des données est sans doute aussi large que celui des biens. Si la comparaison avec le pétrole revient souvent (en ce qu'elles irriguent tous les secteurs de l'économie), celle-ci trouve rapidement ses limites. Mais alors, si parler de données englobe des réalités tellement distinctes, pourquoi reste-t-il pertinent d'en étudier la valorisation au travers d'un seul prisme ? En réalité, les données partagent certaines propriétés économiques importantes.

Tout d'abord, elles sont répliquables, c'est-à-dire qu'elles peuvent être dupliquées pour un coût très faible, et (en général) non-rivales : leur utilisation par un acteur n'empêche pas l'utilisation par un autre. Là où un baril de pétrole ne peut être extrait, vendu et utilisé qu'une fois, une donnée peut potentiellement être utilisée une infinité de fois. Par exemple, les données recueillies par Waze permettent à la fois de prédire les embouteillages, de calculer des itinéraires, voire d'optimiser l'urbanisme des villes. Alors que la propriété de non-rivalité devrait encourager sa circulation, la quasi-gratuité de la réplication peut paradoxalement constituer un obstacle à la valorisation des données. En effet, en théorie économique, dans une situation de concurrence pure et parfaite, un bien est vendu à son coût marginal de production. Or, celui-ci est ici nul. Concrètement, une entreprise qui possède de la donnée devrait donc être en mesure de la valoriser auprès de nombreuses autres entreprises, mais des difficultés pourraient résider dans l'estimation du prix de vente. Ces considérations, de prime abord plutôt abstraites, se révèlent structurer la réalité vécue par les entreprises souhaitant valoriser leurs données.

1.2. COMMENT ESTIMER LA VALEUR DES DONNÉES ?

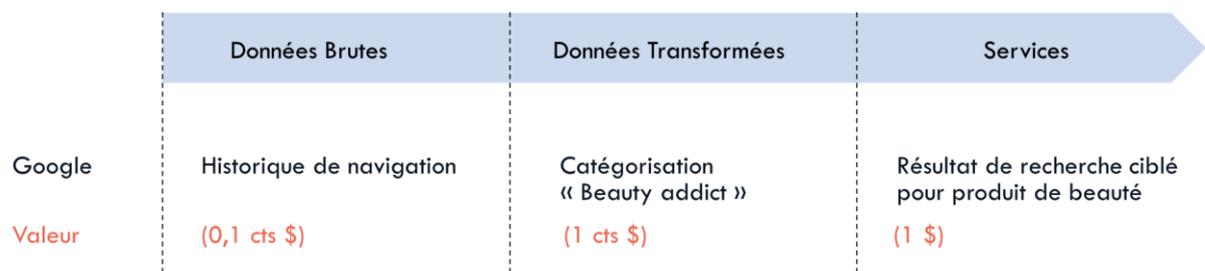
On l'a dit, les données sont unanimement reconnues comme représentant une importante source de valeur. La Commission européenne n'hésite d'ailleurs pas à affirmer que celles-ci contribuent à « la croissance économique, la compétitivité des entreprises et l'innovation, à la création d'emploi et au progrès sociétal⁷ ». Il est instructif de s'arrêter quelques instants sur les mécanismes sur lesquels repose cette création de valeur.

La diversité précédemment évoquée de la nature des données se traduit également par une diversité des caractéristiques permettant la création de valeur. Par exemple, la « fraîcheur » d'une donnée pourra ou non être primordiale à un cas d'usage. L'intérêt d'un carnet de santé numérique réside souvent dans l'historique des pathologies, là où, dans un tout autre domaine, le ciblage publicitaire nécessite des données extrêmement récentes. D'autres caractéristiques comme le volume ou la véracité constituent également des facteurs importants permettant ou non la création de valeur des données. On parle

⁷ Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, « Créer une économie européenne fondée sur les données », janvier 2017.

d'ailleurs souvent des « cinq V de la donnée⁸ » : véracité, volume, vélocité, variété, visibilité.

À l'instar du pétrole, la donnée connaît aussi plusieurs étapes de traitement avant de révéler sa vraie valeur. La chaîne de valeur des données est un outil souvent utilisé pour représenter la façon dont la donnée irrigue toutes sortes de services ou produits à valeur ajoutée.



Chaîne de valeur de la donnée : l'exemple de la publicité en ligne

Cette visualisation permet une séparation en 3 étapes, décrite ci-après avec le cas le plus emblématique et le plus important de valorisation des données : celui de la publicité en ligne.

Données Brutes : La donnée brute peut être récoltée par les entreprises sous de nombreuses formes. À ce stade, elle est totalement non-structurée, et, comme nous le verrons par la suite, difficilement valorisable directement.

Dans le cas de la publicité en ligne, il peut s'agir par exemple d'un historique de navigation. L'entreprise Google enregistre et stocke ainsi la trace de chacune des recherches web effectuées par chacun de ses utilisateurs.

Données Transformées : Afin de créer de la valeur à partir des données, il est nécessaire de la traiter pour en extraire les informations pertinentes. Ce passage obligé est généralement réalisé en entreprise par les « *data scientists* » et consiste à structurer, filtrer, nettoyer ou croiser différentes bases de données.

Pour Google, cette donnée transformée peut, par exemple, correspondre à une classification de l'utilisateur dans différents profils-types de consommateur, par exemple

⁸ Sur les (nombreux) V de la donnée, voir par exemple : <https://www.cairn.info/revue-regards-croises-sur-l-economie-2018-2-page-27.html>

« *Beauty-addict* ». L'ordre de grandeur de valeur d'une telle donnée peut être 10 fois plus élevé qu'une simple donnée brute.

Service : Le dernier stade de valorisation d'une donnée est celui d'un service ou d'un bien à valeur ajoutée reposant sur ces données. Il n'y a pas de règle générale, mais l'information apportée par les données doit permettre de rendre ce bien ou ce service plus pertinent.

Les moteurs de recherche en ligne sont un vecteur très puissant de valorisation. En effet, des entreprises vendeuses de produits de beauté accordent par exemple beaucoup de valeur à apparaître en haut des résultats de recherche d'un utilisateur ayant été classé comme un « *beauty addict* ». Une entreprise comme Google peut valoriser sa donnée *via* son service à un ordre de grandeur encore très supérieur, parfois plusieurs dollars pour apparaître dans les résultats de recherche d'un seul utilisateur.

1.3. QUELLE EST LA VALEUR DU PARTAGE DE DONNÉES ?

Si les mécanismes de création de valeur à partir des données sont de mieux en mieux compris des entreprises, ces dernières n'ont encore que trop peu exploré leur valorisation auprès d'acteurs tiers. Certains usages nés du partage et du croisement de données se sont révélés très bénéfiques pour les entreprises engagées (dans l'aéronautique notamment), d'autres plus confidentiels sont très prometteurs, mais le phénomène reste encore très peu généralisé.

Nous pensons pourtant que le partage de données entre entreprises représente un potentiel de valeur important restant à réaliser. Dans une étude rendue à la Commission européenne, le cabinet Everis note que ce phénomène est amené à prendre de l'importance dans les prochaines années et recommande à la Commission « d'organiser des campagnes de sensibilisation afin d'impliquer davantage d'entreprises dans du partage de données⁹ ».

Il convient d'étudier les raisons qui expliquent pourquoi la circulation des données représente un tel levier d'opportunité.

Nous distinguons trois sources principales de valeur relatives au partage de données.

⁹ « Study on data sharing between companies in Europe », A study prepared for the European Commission DG Communications Networks, Content & Technology by Everis, 2018.

1.3.1. Les externalités positives et la force de la multitude

Le caractère non-rival de la donnée la positionne comme un bien de choix dans la génération d'externalités positives. En effet, puisque l'utilisation de données par un acteur n'en détruit pas la valeur potentielle pour un autre, et qu'à l'inverse leur mise en commun peut aboutir à la réalisation d'innovations non-concurrentes que n'autorisaient pas les données de chaque entreprise individuellement, le partage de données d'une entreprise vers d'autres permet de réaliser *in fine* une valeur inatteignable initialement. On retrouve ce raisonnement dans les théories des biens communs en économie de l'information et de l'innovation : la connaissance partagée par une communauté, par exemple un savoir technique ou scientifique, est un bien collectif non-rival également, dont le partage génère des externalités positives. Cependant, sa nature s'accommode mal des règles de l'économie de marché : le rendement du savoir créé par un acteur privé (par exemple, la R&D d'une entreprise) est *a priori* inférieur à son rendement social puisqu'il engendrerait des innovations additionnelles s'il était partagé. Toutefois, si l'on devait imposer la diffusion des connaissances entre les entreprises, les incitations à produire de la R&D risqueraient de diminuer fortement, les entreprises ne désirant pas investir dans la production d'un savoir dont elles porteraient le coût principal (ainsi que l'incertitude), et dont les autres acteurs bénéficieraient à un coût marginal. Le caractère non-rival et collectif de la connaissance est donc à l'origine d'une déficience de marché. Pour permettre de réconcilier optimum social et optimum privé, les pouvoirs publics ont alors recours à différentes approches, qui consistent soit à contrôler le partage de la connaissance (régimes de propriété intellectuelle) soit à prendre en charge le coût de sa production afin de la rendre collective (recherche publique).

Comme nous le verrons, la propriété intellectuelle n'est malheureusement souvent pas extensible aux données des entreprises, et le parallèle entre R&D et données privées trouve ici une de ses limites. Cela permet néanmoins d'illustrer un certain nombre d'enjeux liés à la réconciliation entre les incitations des acteurs privés et la réalisation d'un optimum social.

On soulignera en outre que le caractère souvent numérique des données, leur accessibilité et les faibles barrières à l'entrée associées (si des données sont mises en ligne, tout le monde peut les utiliser à faible coût) renforcent cet effet potentiel d'externalités positives¹⁰.

Un exemple emblématique de retombée positive pour les citoyens est celui de l'ouverture des données de transport. Des applications comme Waze ou Citymapper permettent ainsi, dans les métropoles où les données en temps réel des transports en commun sont ouvertes, d'offrir des calculs de trajets astucieux et précis. Une étude du cabinet Deloitte a ainsi chiffré, pour la seule ville de Londres, le gain de temps (et donc l'externalité positive associée) à près de 70 à 90 millions de livres par an pour les citoyens¹¹.

1.3.2. Croiser des informations de sources différentes

Le croisement de bases de données issues d'entreprises différentes permet également une importante création de valeur.

L'exemple de la publicité en ligne est sans doute parmi les plus parlants. Les utilisateurs d'internet génèrent *via* leur navigation des données personnelles qui révèlent leurs personnalités, goûts, préférences, etc. Chacune de ces données collectées par les sites Internet prises indépendamment serait probablement difficilement valorisable. À l'inverse, pour de gros acteurs (les GAFAs), plus les données qu'ils croisent sont nombreuses et issues de sources différentes, plus ils sont en mesure de fournir à leurs entreprises clientes des services de campagnes marketing ciblant des internautes pertinents. C'est ainsi que se sont construites les valeurs boursières d'entreprises comme Google ou Facebook. Le marché de la publicité en ligne est aujourd'hui estimé à 194 milliards de dollars¹². Une grande partie de cette valeur ne serait pas possible sans ces croisements de données de différentes sources.

1.3.3. Atteindre une masse critique de données *via* la mutualisation

Enfin, le besoin de données en forte quantité constitue également un argument en faveur du partage. En effet apparaissent aujourd'hui de nouveaux usages de la donnée liés à l'intelligence artificielle ou à la maintenance prédictive, qui demandent toujours plus de

¹⁰ À propos de l'importance de la « multitude » à l'ère numérique, voir Nicolas Colin et Henri Verdier, *L'Âge de la multitude. Entreprendre et gouverner après la révolution numérique*, Paris, Armand Colin, 2012.

¹¹ « Assessing the value of TfL's open data and digital partnerships », Deloitte, 2017, <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>

¹² Statista, « Valeur des dépenses mondiales en matière de publicité sur internet de 2013 à 2018 ».

données pour alimenter les algorithmes et améliorer leur performance. Pour développer ces usages, il n'est donc pas rare qu'une entreprise seule ne soit pas en mesure de regrouper assez de données. Apparaissent alors des besoins de mutualisation de bases de données, comme cela a été décrit dans le rapport Villani en 2018¹³. Ce dernier pointait en effet du doigt le besoin de création de jeux de données stratégiques, à mettre notamment à disposition des chercheurs, pour que la France puisse rattraper son retard sur la Chine et les États-Unis dans le domaine de l'intelligence artificielle. Par la suite, la DGE (Direction générale des entreprises) a lancé un appel à manifestation d'intérêt pour soutenir des projets de mutualisation de données entre acteurs privés¹⁴. Celui-ci a reçu un vif intérêt, avec 75 réponses, et sera, semble-t-il, suivi d'un appel à projets doté de financements¹⁵.

Ce besoin de mutualisation est, par exemple, apparu dans l'industrie aéronautique. Airbus a ainsi fait développer par l'entreprise américaine Palantir une plateforme logicielle nommée Skywise. Avec cette plateforme, les compagnies aériennes mutualisent les données produites par leurs avions en vol (monitoring, pannes, etc.). Cela permet à Airbus de leur fournir en retour un service de maintenance prédictive pour minimiser le temps d'immobilisation des avions au sol entre chaque trajet, ce qui *in fine* induit un gain économique pour les compagnies.

Par ailleurs, elles seules n'auraient pas la masse critique de données pour parvenir à un niveau satisfaisant de maintenance prédictive. Il était donc nécessaire de réunir des données de différentes compagnies aériennes pour créer ce nouveau service de maintenance prédictive. Pour ne pas donner un avantage concurrentiel à leurs rivales, il aurait été très peu probable que les compagnies acceptent de partager leurs données avec leurs pairs. Airbus s'est donc retrouvé en position de force pour collecter les données de toutes les compagnies et internaliser ce service. Il est intéressant de noter que, dans ce cas, c'est la structure du marché, avec Airbus en tant qu'acteur central, qui a permis l'apparition spontanée sur le marché d'un modèle de mutualisation.

¹³ « Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne », https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf, page 29.

¹⁴ <https://www.entreprises.gouv.fr/numerique/mutualisation-de-donnees-pour-intelligence-artificielle>

¹⁵ <https://www.entreprises.gouv.fr/numerique/intelligence-artificielle-au-service-des-entreprises>

2 - POURQUOI LES ENTREPRISES ONT-ELLES ENCORE DU MAL À PARTAGER LEURS DONNÉES ?

Si le partage de données représente un gros potentiel de valeur économique, pourquoi les entreprises sont-elles encore réticentes à s'engager ? Comme nous allons le voir, certaines incitations au niveau micro des entreprises ainsi que des zones de flou juridique qui créent des incertitudes empêchent l'économie globale de réaliser la valeur promise par le partage de données.

Nous ne nous intéressons dans le cadre de cette note qu'à la circulation de données entre entreprises. Il va sans dire que la mutualisation de données présente également des enjeux colossaux pour le secteur public ou la recherche.

2.1. DES ENTREPRISES ENCORE PEU MATURES

Au cours de la cinquantaine d'entretiens que nous avons menés auprès de décideurs d'acteurs privés, nous avons constaté une assez faible maturité au sujet des modèles économiques de partage de données. Les entreprises sont maintenant bien conscientes de la valeur de leurs données en interne, pour l'optimisation de leurs ventes ou de leur logistique (sans pour autant toujours réussir à les exploiter). En revanche, elles ont encore peu conscience de la valeur potentielle de leurs données en externe. Plusieurs facteurs expliquent cela.

- Le sujet est encore nouveau, et son émergence est concomitante avec le développement du big data et de l'IA.
- Les difficultés techniques sont nombreuses, parfois cumulées avec un retard de consolidation des bases de données dans les grandes entreprises. Il nous a ainsi été rapporté qu'une grande entreprise française ne disposait pas de la vue consolidée de ses ventes et que les bases de données étaient extrêmement fragmentées. Ces difficultés techniques se révèlent également être de véritables freins à l'expérimentation.
- On n'a certainement pas encore découvert les cas d'usage les plus intéressants. Il est en effet difficile pour une entreprise spécialisée dans un secteur économique d'imaginer des usages potentiels pour ses données dans un autre secteur. La

collaboration entre deux entreprises de secteurs différents n'est d'ailleurs pas forcément naturelle.

- Pour les entreprises dont le numérique n'est pas le cœur de métier, la maturité vient certainement moins vite.

Malgré ces freins, les entreprises françaises commencent à s'emparer du sujet, avec l'émergence de collaborations et de modèles économiques dans tous les secteurs. Les entreprises font d'ailleurs parfois de la valorisation externe de leurs données sans s'en apercevoir ! Avec le temps, il est certain que les entreprises se saisiront plus largement de ce phénomène émergent.

2.2. LA DIFFICULTÉ DE MONÉTISATION EXPLIQUE UNE TENDANCE À L'INTERNALISATION DES SERVICES

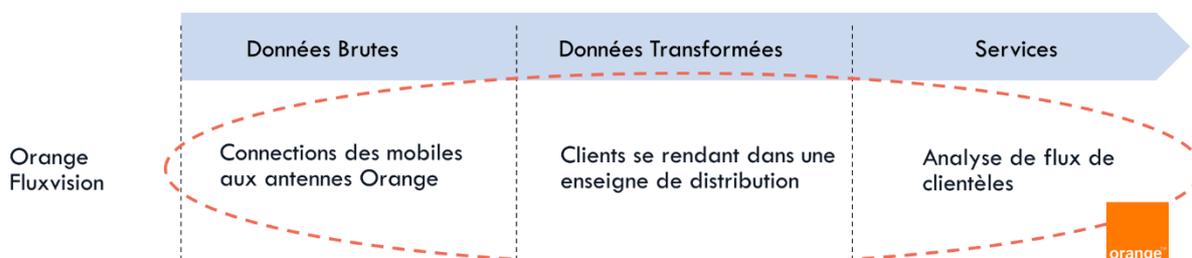
Au-delà de ces inévitables balbutiements, une raison plus fondamentale explique la difficulté des modèles économiques de partage de données. Au cours des différents entretiens que nous avons menés et des cas que nous avons étudiés, nous avons également remarqué ce que nous appelons une tendance à l'internalisation des services développés à partir des données.

Si l'on reprend la chaîne de valorisation des données présentée précédemment, cela se traduit par des difficultés ou des réticences à partager et valoriser directement des données brutes. Les entreprises qui souhaitent valoriser leurs données à l'extérieur préfèrent en effet traiter toute la chaîne de valeur de la donnée jusqu'à développer les services elles-mêmes.

Ce phénomène s'illustre avec le cas d'Orange Fluxvision¹⁶, un service développé par l'opérateur télécom Orange à destination par exemple des enseignes de grande distribution. Le fonctionnement est le suivant : sur la base des données de position générées par le téléphone de ses clients, Orange vend à des hypermarchés des outils d'analyse de leur clientèle : « Quel âge ont vos clients ? », « D'où viennent-ils ? », « Combien de temps restent-ils en magasin ? », etc. Ces informations ont bien sûr une grande valeur pour les hypermarchés.

¹⁶ <https://www.orange-business.com/fr/produits/flux-vision>

Nous pourrions imaginer que si Orange était en mesure d'ouvrir ces données brutes en les monétisant, outre les problèmes d'anonymisation¹⁷ – incontournables – que cela poserait, de nombreux services pourraient se créer et, par exemple, n'importe quelle enseigne serait en mesure de récupérer ses statistiques de fréquentation.



Chaîne de valeur de la donnée : l'exemple d'Orange Fluxvision
Orange internalise un service d'analyse de flux de clientèles à partir de ses données

Malheureusement, et au-delà du cas d'Orange, les modèles économiques de partage de données brutes restent très rares (voir l'exemple du marché des Data Brokers ci-dessous). Cela restreint la pleine réalisation de la valeur du partage de données, en diminuant les externalités, le croisement de bases différentes ou la mutualisation.

¹⁷ Pour approfondir la question de l'anonymisation des données personnelles et des risques inévitables de ré-identification, voir par exemple Boris Lubarsky, « Re-Identification of "Anonymized Data" », *Georgetown Law Technology Review*, 2017.

Zoom - Data Brokers

Le secteur de la publicité en ligne constitue un exemple intéressant, où les acteurs se sont naturellement répartis le long de la chaîne de valeur. Ainsi, des entreprises appelées *data brokers* achètent des données brutes de navigation à des sites web, pour les croiser et revendre de véritables profils de consommateurs. Ces profils sont vendus à des entreprises souhaitant mener des campagnes marketing ciblées. C'est d'ailleurs grâce à ce secteur que nous disposons d'ordres de grandeur de la valeur effective de ces données individuelles¹⁸.

Le rôle de ces *data brokers* est évidemment très controversé, notamment suite au scandale Facebook/Cambridge Analytica¹⁹, et ils ont été épinglés dès 2013 dans un rapport sénatorial américain²⁰. Les dérives pointées par ce rapport sont édifiantes : vente de listes de profils catégorisés comme « vieillards souffrants » ou « parent célibataire pauvre²¹ ». Leur activité est très réduite en Europe, notamment depuis l'entrée en vigueur du RGPD, qui interdit la revente de données personnelles sans consentement explicite et *a priori* des individus.

Différents éléments contribuent à expliquer cette exception sectorielle. Historiquement, les *data brokers* (Acxiom, Experian, Equifax, etc.) travaillaient déjà sur des données personnelles en proposant des services de notation (*credit rating*) aux États-Unis. Il s'agit de recueillir et croiser un maximum de données sur le profil d'un individu pour aider les banques à estimer leur risque de non-remboursement d'un prêt. Les deux activités étant très similaires, la diversification était donc naturelle. Par ailleurs, tous les acteurs de ces chaînes de valeur de la donnée financière individuelle avaient déjà conscience de la valeur des données. Enfin, il s'agit d'acteurs dont le cœur de métier est le traitement de la donnée, au contraire des industries plus classiques.

Deux éléments viennent selon nous expliquer cette tendance à l'internalisation et ces difficultés de partage.

2.2.1. De l'impossibilité de valoriser sa donnée sans en connaître *a priori* l'usage

Partant du postulat important que la donnée n'a pas de valeur intrinsèque – mais uniquement une valeur tirée de l'usage qui est construit à partir des données – il est important pour une entreprise de connaître l'usage qui sera fait de ses données partagées. Cela a une importance primordiale pour estimer à quel prix il est possible de les vendre. Or, on l'a dit, l'une des sources de valeur importantes du partage de données est la création d'externalités et de nouveaux modèles économiques, qui, par définition, ne sont pas attendus au moment du partage. Les entreprises n'ont donc pas les incitations nécessaires

¹⁸ Voir par exemple OECD (2013), « Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value ».

¹⁹ https://fr.wikipedia.org/wiki/Scandale_Facebook-Cambridge_Analytica

²⁰ Staff Report for Chairman Rockefeller, « A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes », US Senate, 2013.

²¹ Pages 24-26 du rapport précédemment cité, voir également *The New York Times*, « Bilking the Elderly, With a Corporate Assist », <https://www.nytimes.com/2007/05/20/business/20tele.html>

pour créer des externalités avec leurs données, puisqu'elles n'arrivent pas à les monétiser correctement sans connaître l'usage par avance.

Un contre-exemple marquant reste celui de l'entreprise américaine et réseau social Twitter. En effet cette dernière monétise directement – bien que cela représente en 2015 moins de 10 % de ses revenus bruts²² – sa donnée brute générée sur sa plateforme par des API (interface de programmation applicative) payantes. De nombreux usages différents peuvent être faits à partir de cette donnée, l'entreprise Twitter a tout de même fait le choix de les ouvrir à un prix fixé, la seule différenciation sur le prix se situant principalement dans la quantité de données qu'il est possible de récupérer par intervalle de temps.

2.2.2. Pour un usage connu, quelle part de la valeur ajoutée provient de la donnée ?

Mais même lorsque l'usage est connu à l'avance, monétiser ses données au juste prix reste une question délicate. En effet, mesurer la part de valeur ajoutée provenant des données est souvent une mission perdue d'avance.

Prenons l'exemple théorique du développement d'un service de reconnaissance d'image, basé sur des algorithmes d'intelligence artificielle. Beaucoup d'usages permettent probablement de monétiser ce service. Supposons qu'il existe un usage connu pour lequel l'entreprise ayant développé l'algorithme parvienne à en tirer un revenu, disons 100. Mettons-nous maintenant à la place de l'entreprise qui a fourni tout ou une partie des données permettant l'entraînement de l'algorithme. Quelle part de cette valeur suis-je en droit de réclamer ? 5, 10, 50 ? Pire, une fois que l'algorithme a été développé et entraîné, il n'a plus besoin de mes données et les leviers restant à ma disposition pour négocier un bon prix sont malheureusement assez restreints...

En résumé, monétiser de la donnée brute s'avère ardu pour les entreprises. C'est ainsi que face à la difficulté de fixer un prix, les entreprises qui possèdent de la donnée préfèrent tout simplement ne pas la partager. Beaucoup d'acteurs peuvent se dire : « Aucune envie que quelqu'un fasse de l'argent avec mes données. »

En outre, certains risques ou coûts associés achèvent de dissuader les entreprises envisageant de monétiser leurs données. Simon Chignard, conseiller stratégique chez Etalab, en dresse un portrait convaincant dans son article pour *Management & Data*

²² Surobh Das, « How does Twitter Make Money? Twitter Business Model », www.feedough.com, février 2019.

*Science*²³. On peut citer notamment le risque que la donnée partagée ne mette l'entreprise en difficulté d'un point de vue concurrentiel en favorisant l'émergence de nouveaux acteurs. Les coûts de développement technique liés à l'interopérabilité des systèmes sont également un obstacle. Enfin, dans les initiatives de mutualisation, il y a souvent un intérêt à ne pas être le premier, voire à contribuer moins que ses pairs, problème bien connu en économie : le passager clandestin.

Du point de vue d'une entreprise (« micro »), la balance entre les gains et les risques et coûts potentiels est donc souvent défavorable, freinant ainsi le développement de modèles économiques liés au partage de données. Aux niveaux national et européen (« macro »), une meilleure circulation des données privées débloquerait davantage d'innovation. Ce décalage entre incitations micro et optimum macro représente un terrain propice à une éventuelle intervention de la puissance publique.

3. OUVRIR LES DONNÉES D'ENTREPRISE ?

Pour faciliter la circulation des données, plusieurs modes d'action de la puissance publique sont envisageables. Des initiatives publiques « soft » existent déjà, notamment *via* les travaux, rapports et communications de la Commission européenne. On pourrait aussi envisager des incitations financières pour les entreprises prenant activement part à des initiatives de partage de données. À notre sens, une intervention plus contraignante permettrait de répondre plus adroitement aux barrières identifiées précédemment. Nous étudions, dans cette note, la forme d'intervention nous paraissant la plus naturelle, à savoir l'ouverture de bases de données possédées par les entreprises. Ces ouvertures de données privées sont d'ailleurs de plus en plus envisagées, par exemple dans le très récent rapport de la commission d'enquête du Sénat sur la souveraineté numérique²⁴.

Il faut, avant d'intervenir, comprendre qui est réellement propriétaire de ces données à forts enjeux. Comme nous allons le voir, il existe un flou (voire un vide) juridique sur la notion de propriété des données. Par ailleurs, intervenir sur la « propriété » privée et l'ouverture de données est un terrain sur lequel l'État s'est encore peu aventuré. Enfin, des ouvertures de données mal maîtrisées pourraient induire des dommages notables pour l'économie. Les

²³ Simon Chignard, « Partager les données du secteur privé : mission impossible ? », *Management & Data Science*, 2018.

²⁴ http://www.senat.fr/commission/enquete/souverainete_numerique.html

interventions de l'État devront donc être très ciblées, justifiées, et le périmètre de l'ouverture des données devra être déterminé précautionneusement.

3.1. À QUI APPARTIENNENT LES DONNÉES DE TRACE ?

Avant de décider ou non d'interventions et d'ouvertures de données, il s'agit de savoir à qui elles appartiennent vraiment ! Si l'interrogation semble de prime abord être une considération théorique, la question de la propriété des données structure pourtant les leviers d'intervention de la puissance publique. Les spécificités économiques des données évoquées précédemment rendent particulièrement délicat le fait de parler de propriété des données.

La bonne notion pour s'intéresser aux enjeux juridiques du partage de données est la notion de données de trace. Nous appelons données de trace les données générées par l'utilisation d'un système numérique, en grande quantité et collectées passivement. Quelques exemples de traces : un historique de recherche Google, les données collectées par une machine à traire les vaches (plus de 120 variables collectées par vache et par jour²⁵ !) ou l'historique de validations d'un pass Navigo. Cette notion s'oppose principalement aux créations ou œuvres numériques, produites de manière unique et ponctuelle par un humain. Par exemple, un morceau de musique au format mp3, un modèle de conception 3D de véhicule ou encore un e-book ne rentrent pas dans le cadre de cette définition. Une autre définition serait d'envisager les données de trace comme celles n'étant pas naturellement protégées par la propriété intellectuelle. En effet, les données de trace décrivent des faits – et non des créations ou inventions – et ne tombent donc pas sous les régimes de propriété intellectuelle. Par exemple, il ne ferait pas sens de protéger les données sur les embouteillages possédées par Waze comme on protège un morceau de musique : une autre entreprise pourrait tout aussi bien mesurer les bouchons (c'est d'ailleurs le cas) sans être dans une situation répréhensible.

Pourquoi ces données de trace sont-elles particulièrement pertinentes ? Il s'avère que les données de trace impliquent souvent plusieurs acteurs : le fabricant du système, son utilisateur final, potentiellement les autres acteurs d'une chaîne de valeur. Par exemple, dans un système aussi complexe qu'un véhicule connecté, chaque pièce génère

²⁵ <https://www.progressivedairy.com/topics/management/management-decisions-enhanced-with-robotic-milking-data>

énormément de données de trace^{26,27}. Si le constructeur y a nativement accès, les sous-traitants peuvent aussi revendiquer leur accès. Naturellement, le consommateur final qui conduit le véhicule pourrait aussi légitimement prétendre avoir accès à ces traces. De plus, les données de trace sont par définition générées comme coproduit d'une activité, et ne sont donc que rarement valorisées directement. Il est donc crucial d'apprendre à mieux les exploiter. À propos des véhicules connectés, il est intéressant de noter que les juristes américains ne savent pas répondre avec certitude à la question « qui possède les données produites par votre voiture intelligente ? » (« *who owns the data generated by your smart car ?* »)²⁸.

Concernant le droit de propriété des données en général, la loi reste assez floue. On l'a dit, les données de traces ne sont pas naturellement couvertes par la propriété intellectuelle. Les données décrivant des faits ne sont pas protégées par le copyright aux États-Unis²⁹. En Europe, une notion faible de propriété intellectuelle sur les bases de données existe, *via* une directive de 1996³⁰. Il s'agit d'un droit *sui generis* qui protège l'acteur ayant réalisé un « investissement financier, matériel ou humain conséquent » pour constituer la base de données. Des arrêts de la Cour de Justice européenne de 2004 sont cependant venus préciser et limiter le périmètre de cette mesure : l'investissement doit avoir porté sur l'obtention des données. De même, au travers de plusieurs décisions, le Conseil d'État a jugé qu'il n'y avait pas de droit de propriété sur les données brutes. Cela exclut de fait les données de trace telles que nous les avons définies : leur obtention est quasiment passive, en tant que « produit secondaire » du fonctionnement d'un système. Il n'existe donc pas de droit de propriété prévu par la loi pour les données de trace. En somme, personne ne sait dire à qui *a priori* devraient appartenir les traces d'utilisation d'un tracteur agricole : à l'agriculteur qui opère l'équipement ou au fabricant ? Nous reviendrons ultérieurement et plus avant sur le secteur de l'agriculture.

²⁶ Voir par exemple « Monetizing data in the age of connected vehicles », Deloitte, 2019.

²⁷ Une Tesla renvoie ainsi au fabricant environ 100Mb/jour.

²⁸ « *Although the consumer owns the smart car itself, data collected by the vehicle cannot be directly "owned" like traditional intellectual property. Instead, the rights to access, limit access to, use, and destroy data are likely the closest proxies for "ownership"* », « Who Owns the Data Generated by your Smart Car? », *Harvard Journal of Law & Technology*, 2018.

²⁹ « Introduction to intellectual property rights in data management », Cornell University, <https://data.research.cornell.edu/content/intellectual-property>

³⁰ Directive 96/9/CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données, voir aussi <https://www.legavox.fr/blog/murielle-cahen/protection-bases-donnees-droit-generis-984.htm>

L'ancien commissaire européen au Numérique Günther Oettinger a d'ailleurs proposé³¹ de trancher de manière définitive et transverse cette question du droit de propriété des traces en donnant systématiquement un droit de propriété intellectuelle à l'utilisateur de la machine.

Dans la pratique, cette incertitude sur l'aspect juridique de la propriété des données est réglée au cas par cas, de manière très pragmatique, par le droit des contrats. Des contrats de droit privé régulent qui a accès aux données, selon les rapports de force du secteur. Par exemple, l'agriculteur acquérant un tracteur John Deere signe un contrat de licence l'obligeant à partager les données générées par l'engin avec son fabricant. L'équilibre effectif de la propriété des données dépend donc à la fois de l'identité de l'acteur qui les collecte nativement et de ses rapports avec les autres acteurs du secteur en question. Cette allocation des données n'est donc pas forcément toujours optimale pour favoriser l'innovation. De mauvaises surprises peuvent parfois apparaître concernant la propriété de ces traces : l'armée norvégienne a ainsi découvert que ses avions de chasse F-35 renvoyaient des données sensibles à son constructeur américain³².

Enfin, en matière de données personnelles, la question du droit de propriété est claire : il n'existe pas. En effet, philosophiquement, et dans une conception européenne, les données personnelles constituent une extension de l'individu et ne peuvent donc pas être vendues, au même titre qu'on ne peut vendre un organe. On peut cependant envisager le droit de contrôle (consentement, contrôle des finalités de collecte, etc.) accordé à l'individu sur ses données personnelles par le RGPD comme une forme très atténuée de droit de propriété.

Quelles conclusions faut-il tirer de ces considérations théoriques ?

- Les données de trace concentrent une bonne partie des enjeux liés au partage de données. Ces dernières ne rentrant pas dans le cadre de la propriété intellectuelle, les entreprises sont incitées à les accumuler, sans les partager, de peur d'en perdre le contrôle. La typologie des acteurs (sous-traitant, fabricant, utilisateur) ayant accès aux données dépend alors des rapports de force du secteur. Les

³¹ European Commission, « Building a European data economy », 22 mai 2019, p. 13, <https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy>

³² « "Spy" F-35s send sensitive Norwegian military data back to Lockheed Martin in the United States », <https://www.news.com.au/technology/online/security/spy-f35s-send-sensitive-norwegian-military-data-back-to-lockheed-martin-in-the-united-states/news-story/12b4fafce6b579448cc8416518063d1f>

incertitudes juridiques sont donc une raison supplémentaire pour laquelle les entreprises s'engagent encore trop peu dans du partage.

- En dehors de cas très spécifiques, il n'existe pas de droit de propriété fort sur les données des entreprises. Cela permet donc à une éventuelle ouverture forcée de ces données privées de ne pas tomber sous la qualification d'expropriation³³.

3.2. POURQUOI UNE OUVERTURE POURRAIT RÉPONDRE AUX BESOINS D'INNOVATION – L'EXEMPLE DE DSP2

Nous examinons maintenant en quoi une ouverture ciblée de données permettrait de répondre à des enjeux d'innovations.

L'intervention publique la plus probante en faveur d'une ouverture des données pour des besoins d'innovation provient du secteur bancaire et de la directive DSP2 (Directive sur les systèmes de paiement, octobre 2017). Cette directive sectorielle utilise le principe de portabilité des données personnelles introduit par le RGPD pour forcer les banques à ouvrir *via* des API standardisées les données de consommation bancaire de leurs clients. Ces derniers peuvent désormais connecter automatiquement leurs comptes bancaires à d'autres applications pour recevoir des services tels que de l'analyse automatique de comptes, des levées d'alerte en cas de grosses dépenses, des cartes de fidélité dématérialisées, etc. Cela a permis le foisonnement de start-ups (les « fintech ») innovantes proposant ces services à forte valeur ajoutée pour les consommateurs. Parmi elles, la start-up espagnole Fintonic, créée en 2012 et qui revendique plus de 700 000 utilisateurs actifs³⁴, se sert des données historiques des comptes bancaires pour réaliser des offres de crédit. Puisque les banques avaient à leur disposition des données qu'elles ne valorisaient pas, le régulateur a donc décidé d'en forcer l'ouverture vers des acteurs tiers, en l'occurrence des fintechs (en s'assurant bien sûr du consentement de l'utilisateur final).

Au vu de la jeunesse de la directive, il est difficile d'établir un chiffrage du gain économique associé à cette législation. Cependant, le foisonnement de start-ups issues de cette

³³ Pour des considérations plus précises sur le risque juridique de l'ouverture forcée de données privées, et notamment le risque que ces ouvertures soient cassées par le Conseil constitutionnel, voir le « Rapport sur les données d'intérêt général », paragraphe 4.2.

³⁴ <https://www.fintonic.com/blog/fintonic-secures-19-million-euros-in-a-new-funding-round-that-sets-the-valuation-of-the-company-at-160-million-euros/>

opportunité réglementaire ainsi que le succès de certaines d'entre elles comme Fintonic laissent à penser que cet impact est tout à fait notable.

Exemple novateur et fondamental de législation sectorielle sur des données privées, la directive DSP2 concentre selon nous tous les aspects positifs d'une régulation sectorielle sur le sujet du partage de données. Nous pensons, et c'est le sens de cette note, que l'État pourrait reproduire cet exemple dans d'autres secteurs.

3.3. LES RISQUES CONCOMITANTS À UNE OUVERTURE

Si l'intervention de la puissance publique peut se révéler pertinente et bénéfique, comme dans le cas de DSP2, elle n'est toutefois pas sans risque.

3.3.1. Risques sur la concurrence

D'abord, différents risques liés à la concurrence sont évoqués par des rapports de la Commission Européenne³⁵.

- Risques de collusion : certains partages d'informations entre concurrents sont interdits, notamment pour éviter des ententes et collusions entre concurrents³⁶. Citons par exemple l'amende de 192,7 millions d'euros contre le « cartel des yaourts » en 2015³⁷.
- Désincitations à investir dans la collecte de données : si une entreprise profite d'un accès à un jeu de données mutualisées, elle pourrait choisir d'investir moins dans ses propres moyens de collecte et de traitement des données.
- Risques de désavantages concurrentiels : les entreprises n'ayant pas accès à un jeu de données risquent de devenir moins efficaces. Ce risque dépend du périmètre d'ouverture et de partage des données, point que nous discutons par la suite.

Il s'agira donc de veiller à ce que d'éventuelles ouvertures de données ne viennent pas nuire à l'intensité concurrentielle d'un secteur.

³⁵ Voir notamment Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitzer, « Competition Policy for the digital era », European Commission, avril 2019, p. 9.

³⁶ Pour approfondir la notion d'échange d'informations, voir : « Table ronde sur les échanges d'informations entre concurrents en droit de la concurrence », 2010 http://www.autoritedelaconcurrence.fr/doc/ocde_cp_10_2010.pdf

³⁷ https://fr.wikipedia.org/wiki/Cartel_du_yaourt

3.3.2. Risques sur le secret industriel

Par ailleurs, bien que non protégées *a priori* par la propriété intellectuelle, il peut arriver que les données de trace possédées par des entreprises révèlent des informations relevant du secret des affaires. Dans cette situation, une ouverture forcée pourrait porter un préjudice sévère à l'entreprise détentrice, écueil qu'il s'agira d'éviter.

3.3.3. Risques sur les données personnelles

Le partage de données personnelles est strictement encadré dans l'Union européenne depuis le RGPD (il faut *a minima* recueillir le consentement et préciser une finalité). L'anonymisation et l'agrégation soigneuses de telles données permettent souvent de les rendre non-personnelles. Par exemple, Waze agrège les positions individuelles de ses utilisateurs (données hautement personnelles) pour constituer des estimations du trafic sans lien avec des individus précis (données non personnelles). Cependant, l'autorité compétente devra estimer et peser les risques de désanonymisation liés à une éventuelle ouverture de données.

3.3.4. Risque politique

Enfin, si les ouvertures de données d'entreprises doivent permettre la création de valeur économique, elles sont inséparables de considérations politiques. Dans cette note, nous étudions les mécanismes de création de valeur *via* des ouvertures de données tout en restant agnostique à la nationalité des entreprises qui en bénéficieront. Par exemple, faudra-t-il promouvoir des ouvertures de données de nos industriels français sur des segments où l'on sait par avance que les entreprises qui en bénéficieront le plus seront chinoises ? Le politique sera chargé de répondre au cas par cas à ces questions.

Ainsi, bien que l'État puisse agir favorablement pour la circulation des données, chaque ouverture possède ses propres enjeux et ses propres risques associés. Nous pensons donc que toute ouverture de données privées doit être considérée de manière sectorielle, après cartographie des données pertinentes, et en pesant précautionneusement les aspects positifs sur l'innovation mais également les potentiels aspects négatifs. En particulier, aucune législation sur les données privées ne saurait être pensée de manière trans-sectorielle.

4 - LES LEVIERS À DISPOSITION DE L'ÉTAT

Si l'État souhaite avancer sur le sujet de l'ouverture de données entre entreprises, il convient de s'interroger sur les leviers à sa disposition. Nous avons identifié trois leviers principaux, qui ont permis par le passé et permettent encore à l'État de procéder à des ouvertures de données. Nous passons ici ces trois leviers en revue.

4.1. DROIT DE LA CONCURRENCE

Le droit de la concurrence est aujourd'hui régulièrement mis en avant comme étant un levier primordial dans la régulation du secteur numérique. Les amendes infligées à Google par la Commission européenne ces deux dernières années (8,25 milliards d'euros d'amende au total³⁸) font figure d'exemple et sont sans cesse reprises par les partisans d'une régulation plus ferme envers les grandes plateformes.

Par ailleurs, la question d'adapter le droit de la concurrence à l'ère digitale pour une meilleure prise en compte des enjeux numériques reçoit aujourd'hui un intérêt au plus haut niveau. Les rapports Furman pour le gouvernement anglais³⁹ ou Crémer-Montjoye-Schweitzer⁴⁰ pour la Commission européenne font figure d'exemples. Plusieurs idées majeures y figurent, telles qu'une régulation *ex ante* des grandes plateformes, *via* des codes de conduite.

En outre, les autorités de la concurrence se sont, par le passé, plusieurs fois intéressées aux données possédées par des entreprises. Lors de la fin des tarifs réglementés du gaz en 2014, l'Autorité de la concurrence française a, par exemple, prononcé une mesure conservatoire contre GDF Suez. Le groupe était en effet soupçonné de profiter de son fichier client provenant de son ancien monopole sur le gaz pour faire une concurrence déloyale aux nouveaux entrants. GDF a donc été contraint de partager cette base de données client avec ses nouveaux concurrents⁴¹.

La Commission européenne, dès 2004, s'est intéressée aux enjeux d'interopérabilité, avec de lourdes sanctions financières contre Microsoft. Microsoft a, par exemple, été condamné

³⁸ http://europa.eu/rapid/press-release_IP-19-1770_fr.htm
http://europa.eu/rapid/press-release_IP-17-1784_fr.htm
http://europa.eu/rapid/press-release_IP-18-4581_fr.htm

³⁹ Jason Furman, « Unlocking Digital Competition », mars 2019.

⁴⁰ Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitzer, « Competition Policy for the digital era », European Commission, avril 2019.

⁴¹ <https://www.lesechos.fr/2014/09/gaz-gdf-suez-devra-ouvrir-son-fichier-client-a-ses-concurrents-309240>

pour ne pas avoir offert à des développeurs tiers la documentation permettant de communiquer avec des serveurs utilisant un système d'exploitation Microsoft⁴². Plus récemment, la Commission européenne a également annoncé *via* un communiqué l'ouverture d'une enquête approfondie sur l'utilisation faite par le géant Amazon des données générées par les vendeurs tiers sur sa plateforme. En effet, le géant américain de l'e-commerce exerce sur son site web le double rôle de détaillant mais également de commerçant *via* les produits qu'il vend sous sa propre marque. Il s'agira pour l'autorité européenne de déterminer si Amazon n'a pas utilisé les données de ses vendeurs tiers à son avantage, par exemple en favorisant certains produits d'une façon non conforme aux règles de la concurrence. L'entreprise Google, *via* son produit Google Shopping, a d'ailleurs été condamnée à une amende colossale en 2017 pour des faits similaires⁴³.

Cependant, sur le sujet précis du partage de données, plusieurs éléments démontrent que le droit de la concurrence n'est pas à même de répondre au besoin identifié de produire des politiques publiques sectorielles pour du partage de données entre entreprises :

- Les articles 101 et 102 du Traité de fonctionnement de l'Union européenne (TFUE)⁴⁴ ne permettent d'appliquer des mesures qu'aux entreprises en situation de position dominante sur leur marché. Ce n'est donc pas un outil permettant des politiques d'ouvertures de données sectorielles, s'appliquant à plusieurs acteurs, comme c'est le cas de la directive DSP2.
- Le droit de la concurrence ne peut que s'appliquer *a posteriori* si un dommage a été commis de la part d'une entreprise.
- Le temps juridique nécessaire à l'application des sanctions et à leur éventuel passage devant la Cour de Justice européenne ne semble pas adapté à l'objectif d'encourager l'innovation par l'intermédiaire du partage de données.

⁴² Voir par exemple : <https://www.usinenouvelle.com/article/l-affaire-microsoft-illustre-l-enjeu-de-l-interoperabilite.N58100>

⁴³ Voir par exemple : <https://www.nextinpact.com/news/104669-google-shopping-commission-europeenne-inflige-amende-242-milliards-deuros.htm>

⁴⁴ <https://concurrence.public.lu/dam-assets/fr/legislation/Article-101-et-102-du-TFUE.pdf>

4.2. PORTABILITÉ DES DONNÉES PERSONNELLES

Si le droit de la concurrence *stricto sensu* ne semble pas répondre au besoin d'encourager la circulation des données sur différents secteurs, la portabilité des données personnelles est un moyen détourné mais efficace de répondre à ce besoin.

C'est sur cette base qu'a été rédigée la directive DSP2, qui fait figure d'exemple sur le partage de données d'entreprises et que cette note propose d'étendre à d'autres secteurs.

Le rapport Crémer-Montjoye-Schwiezter la cite d'ailleurs comme exemple et invoque la « création de nouveaux services à partir de l'accès aux données⁴⁵ » comme une raison devant permettre d'autres régulations sectorielles pro-ouverture des données.

Malheureusement, le principe de portabilité ne s'applique qu'aux données personnelles, qui ne correspondent pas à toutes les données représentant un potentiel fort d'innovation et possédées par les entreprises. Nous pensons donc que ce levier, bien qu'efficace, n'est pas suffisant.

4.3. DONNÉES D'INTÉRÊT GÉNÉRAL

Introduites en 2015 dans un rapport remis au ministre de l'Économie Emmanuel Macron^{46,47}, les « données d'intérêt général » sont des classes de données pouvant être utiles à l'intérêt général et pour lesquelles l'État a donc une légitimité à y avoir accès.

Ce concept s'est matérialisé dans la Loi pour une République numérique de trois façons :

- la récupération par défaut par l'État des données « essentielles » générées par les entreprises en position de délégation de service public ;
- la récupération par défaut par l'État des données essentielles des conventions de subvention ;
- la transmission obligatoire de certaines données des entreprises pour des enquêtes de l'Insee.

⁴⁵ Jacques Crémer, Yves-Alexandre de Montjoye, Heike Schweitzer, « Competition Policy for the digital era », European Commission, avril 2019.

⁴⁶ Voir aussi Bertrand Pailhès, « Comment définir et réguler les "données d'intérêt général" ? », *Annales des Mines*, 2018.

⁴⁷ « Rapport relatif aux données d'intérêt général », 2015.

En outre, le rapport incluait initialement le développement économique comme un « intérêt général » et discutait des modalités d'une récupération potentielle par l'État de données d'entreprises privées. La potentielle inconstitutionnalité de lois généralisées sur les ouvertures de données privées (pour des raisons d'expropriation) et le besoin de produire des législations sectorielles y étaient d'ailleurs déjà identifiés. Cependant, la puissance publique n'a, à ce jour, pas fait part de son intention de matérialiser cette partie du rapport *via* des ouvertures de données privées, et l'implémentation des données d'intérêt général dans la loi reste à ce stade limitée aux trois points évoqués ci-dessus.

En tout état de cause, les données d'intérêt général n'ont vocation à permettre qu'une récupération des données par l'État, qui pourrait alors les ouvrir en mode « open data » pour une réutilisation par d'autres entreprises. Nous pensons que pour améliorer le partage de données entre entreprises, le rôle de la puissance publique ne doit pas se limiter à des ouvertures « open data ». Au contraire, l'État devra légiférer sur la propriété des données dans un périmètre plus restreint, entre des acteurs bien identifiés. Par exemple, dans le cas de la directive DSP2, les données de consommation bancaires ne sont pas ouvertes à tout le monde, mais le consommateur acquiert le droit de les réutiliser auprès d'autres entreprises.

Si les trois leviers évoqués précédemment ont chacun un intérêt, ils ne semblent pas adaptés pour répondre au besoin de générer plus d'innovation avec des données possédées par des entreprises. Nous avons choisi d'illustrer ce manque d'outils réglementaires et les freins au partage des données avec le secteur de l'agriculture, qui se révèle être un terreau fertile d'innovation et d'intervention de la puissance publique.

5. L'EXEMPLE DE L'AGRICULTURE

Contrairement aux idées reçues, l'agriculture est un secteur économique où la révolution numérique s'est opérée en profondeur. Les exploitants agricoles sont dorénavant dotés de tracteurs connectés, de robots de traite, et la révolution de l'IoT (Internet of Things), la démocratisation des drones et le déploiement de la 5G promettent de nouvelles applications. Quel que soit le type d'exploitation (céréales, élevage, etc.), cette évolution permet de meilleurs rendements, *via* des optimisations, des économies (d'engrais par

exemple) et une automatisation croissante. On parle d'ailleurs de « *precision farming*⁴⁸ », et le marché de la robotique agricole, qui représentait 3 milliards d'euros en 2015, est promis à une croissance fulgurante (+ de 70 milliards en 2024⁴⁹).

Citons quelques exemples.

- Les (incontournables) tracteurs agricoles connectés. Ceux-ci recueillent par exemple des données de position, de nature des sols (avec une précision de l'ordre du centimètre), ou encore d'irrigation. Ils deviennent également de plus en plus autonomes.
- Les robots de traite permettent aux exploitations laitières d'améliorer leurs rendements et le bien-être des animaux. Des centaines de variables sont mesurées à chaque traite, pour chaque animal⁵⁰.

5.1. DES DONNÉES VERROUILLÉES...

Ces machines connectées sont donc autant de générateurs de données de trace dont l'exploitation pourrait encore stimuler l'innovation. Il s'avère cependant que ces données sont sous-utilisées, car elles circulent mal entre des acteurs souhaitant les contrôler et les garder.

Comme nous l'avons vu, face au flou du droit de propriété des données *a priori*, les acteurs utilisent entre eux le droit des contrats. Ainsi, les fabricants de matériel tentent, souvent avec succès, de récupérer les données de trace des équipements. Il est souvent inclus dans les conditions d'achat que les données générées par les équipements (tracteurs, robots, etc.) reviennent au fabricant. Celles-ci leur sont utiles pour améliorer les performances de leurs produits, ou mieux comprendre leurs usages. Ces données ont de la valeur pour tout l'écosystème : certains fournisseurs de logiciels de gestion de ferme offrent ainsi des promotions lorsque les agriculteurs consentent à leur partager les données de fonctionnement de leur ferme entière⁵¹.

⁴⁸ <https://agfundernews.com/what-is-precision-agriculture.html>

⁴⁹ <http://www.lafranceagricole.fr/actualites/agriculture-connectee-la-demande-pour-les-robots-des-champs-va-exploser-1,0,414055413.html>

⁵⁰ <https://www.progressivedairy.com/topics/management/management-decisions-enhanced-with-robotic-milking-data>

⁵¹ DG Connect, « Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability », European Commission, 2018, p. 220.

De plus, les équipementiers semblent vouloir garder un écosystème fermé autour de leurs produits, afin de capter toute l'innovation possible. Les enjeux d'interopérabilité des équipements agricoles⁵², bien que connexes aux enjeux de partage de données, sont tout à fait notables.

Enfin, ces tensions autour des données cristallisent des craintes chez les agriculteurs. Par exemple, si des données relatives aux rendements de l'exploitation sont transmises à leurs clients, la négociation n'en sera que plus difficile. Une autre crainte, certainement moins fondée, revient régulièrement : les données pourraient être utilisées pour spéculer sur les cours des matières premières.

Cette lutte pour le contrôle des données de trace agricoles, *via* des contrats, génère donc des tensions et une perte de confiance généralisée entre les acteurs.

5.2. ... DONC SOUS-EXPLOITÉES

Comme nous l'avons vu au début de cette note, croiser des données de sources différentes permet d'améliorer l'existant et de créer de nouveaux biens ou services. Faute de standards communs et de circulation suffisante des données de trace, le secteur agricole souffre d'un déficit d'innovation. On pourrait, par exemple, imaginer qu'un logiciel croise les données de drones survolant les parcelles pour mesurer la santé des céréales et de capteurs passifs d'hygrométrie afin d'optimiser la dispersion d'engrais par des tracteurs intelligents. Cela nécessiterait que les entreprises impliquées acceptent de partager avec des tiers les données qu'elles contrôlent.

Quelques start-ups, telles qu'Augmenta⁵³, s'intéressent à ce type d'applications, mais doivent ajouter des dispositifs physiques supplémentaires aux équipements car elles n'ont pas accès aux données déjà mesurées par les machines agricoles. Un parallèle peut être dressé avec la situation des fintechs avant DSP2. Avant les standards de portabilité de DSP2, les fintechs souhaitant valoriser les données bancaires de leurs clients les récupéraient de manière artisanale et non sécurisée.

⁵² Voir par exemple <https://www.iof2020.eu/trials/arable/farm-machine-interoperability>

⁵³ <https://www.augmenta.ag/>

5.3. VERS UNE AUTORÉGULATION ?

Devant les tensions associées aux enjeux des données agricoles, différentes initiatives d'autorégulation ont émergé des acteurs du secteur. Celles-ci tentent souvent d'encadrer l'utilisation des données de trace, recueillies lors du fonctionnement des machines agricoles. Citons par exemple la charte Data-Agri sur l'utilisation et la propriété des données agricoles, initiative française lancée par la FNSEA en 2018⁵⁴. La philosophie de cette charte de bonne conduite est proche d'un « RGPD agricole » : les agriculteurs doivent donner leur consentement pour le traitement des données produites par leurs machines. Ces traitements ne doivent pas non plus aller contre l'intérêt de l'exploitant. Aux États-Unis, le Farm Bureau syndicat agricole comparable à la FNSEA, a lancé dès 2015 une charte avec des concepts proches sur l'utilisation des données agricoles⁵⁵.

Les machinistes agricoles, conscients des craintes des exploitants, proposent également des chartes relatives aux données. Ainsi, le constructeur de robots de traite et de logiciels Lely s'engage par exemple, dans sa charte, à permettre à l'utilisateur de décider si ses données peuvent être partagées à un tiers⁵⁶.

De telles initiatives permettent donc de lever une partie des tensions liées au contrôle des données, et se retrouvent d'ailleurs dans d'autres secteurs tels que l'automobile. Rien n'assure cependant que ces chartes de bonne conduite seront largement adoptées ni appliquées, si les industriels n'y voient pas leur intérêt. Par ailleurs, ces initiatives non-contraignantes n'abordent pas réellement le déficit d'innovation évoqué auparavant. Une intervention de la puissance publique, que ce soit au niveau français ou européen, pourrait s'inspirer de la DSP2 dans le secteur bancaire, en arbitrant le contrôle des données de trace générées par les machines agricoles entre les industriels et les exploitants.

5.4. QUELS LEVIERS D'INTERVENTION ?

Dans le cas du secteur agricole, les leviers existants d'intervention de la puissance publique évoqués précédemment, se révèlent insuffisants ou inadaptés.

⁵⁴ <https://www.data-agri.fr/>

⁵⁵ À ce sujet, voir cette présentation par l'ACTA (Association de coordination technique agricole) : http://www.acta.asso.fr/fileadmin/ressources/R_D/themes/numerique/sem8octobre2015/08_Charte_LaugaCimino_Numerique8octobre2015.pdf

⁵⁶ Voir la Charte relative aux données Lely – Lely T4C, https://www.lely.com/media/filer_public/cc/f4/ccf4f62e-6482-4f1d-bbcf-595c6908d29c/lely_data_charter_2017_fr.pdf

5.4.1. Droit de la concurrence

Le droit de la concurrence a déjà été utilisé dans le secteur agricole, sur le sujet connexe de la maintenance des équipements, ou « *right-to-repair* ». S'abritant derrière les lois sur la propriété intellectuelle, des fabricants d'équipements agricoles tels que John Deere⁵⁷ ont verrouillé l'accès aux standards et aux logiciels nécessaires à la maintenance des tracteurs. La justice s'est d'ailleurs prononcée plusieurs fois, dans différents pays, mais sans cohérence globale. Un parallèle se dresse avec l'enjeu de l'accès aux données de maintenance des véhicules légers. Une directive européenne de 2007⁵⁸ a imposé aux constructeurs d'accorder aux réparateurs l'accès aux données nécessaires à la réparation des voitures.

Cependant, le droit de la concurrence sous sa forme actuelle se révèle insuffisant dans la plupart des cas, où il n'y a pas de position dominante avérée.

5.4.2. Portabilité des données personnelles

L'utilisation du levier de portabilité des données permettrait de résoudre certaines tensions et de stimuler l'innovation dans le secteur. Par exemple, un exploitant aurait le droit de partager les données de trace de son tracteur à un tiers, qui pourrait en extraire de nouveaux services. Cependant, la plupart des données à fort enjeu ne sont pas des données personnelles⁵⁹.

5.4.3. Données d'intérêt général

Une application de l'outil « données d'intérêt général » paraît ici hors de propos. Il s'agit en effet d'arbitrer le contrôle de données de traces entre un fabricant de machine et un exploitant, non de décider d'une ouverture type « open data ».

L'agriculture est donc un exemple illustrant les insuffisances du cadre réglementaire actuel concernant les données privées. Au-delà de cet exemple particulièrement parlant, les tensions illustrées précédemment se retrouvent déjà ou seront amenées à se retrouver

⁵⁷ Voir par exemple <https://www.wired.com/story/john-deere-farmers-right-to-repair/>

⁵⁸ « Access to vehicle repair and maintenance information » : https://ec.europa.eu/growth/sectors/automotive/technical-harmonisation/vehicle-repair-maintenance_en

⁵⁹ On peut cependant noter qu'un flou existe sur la notion de données personnelles dans le secteur agricole. Ainsi, une donnée d'utilisation (position, horaires, etc.) de tracteur sera personnelle si le tracteur appartient à une exploitation unipersonnelle, mais probablement pas s'il est partagé entre des employés.

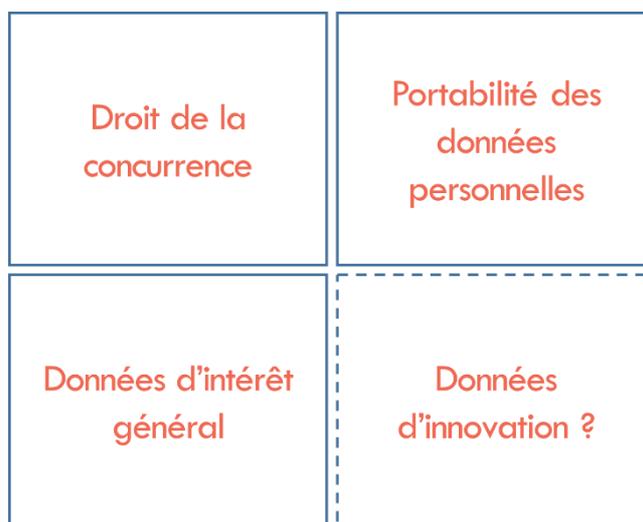
dans de nombreux autres secteurs. Ainsi, des secteurs tels que l'automobile, la santé ou les mobilités devront certainement faire l'objet d'interventions des pouvoirs publics.

Nous pensons donc qu'il est nécessaire d'introduire un autre concept généraliste pour justifier de statuer au cas par cas sur des données de trace, non personnelles et dont une meilleure circulation permettrait plus d'innovation et de concurrence. C'est le cœur de la proposition de cette note.

6. DONNÉES D'INNOVATION : NOTRE PROPOSITION

6.1. DÉFINITION

Les données d'innovation sont des données possédées par des acteurs privés pour lesquelles une circulation entre les agents économiques pourrait concourir au développement économique et à une meilleure concurrence *via* la création de nouveaux services/biens à valeur ajoutée s'appuyant sur ces données. La puissance publique peut alors légiférer sectoriellement, pour l'ouverture standardisée d'une classe circonscrite et très précise de données possédées par un groupe d'acteurs privés en faveur d'un autre groupe d'acteurs.



Les données d'innovation sont l'extension de concepts existants

Voici d'autres interprétations possibles des données d'innovation.

- **Extension du concept des données d'intérêt général à « l'intérêt d'innovation »** : les données d'innovation invoquent le besoin d'augmenter l'innovation, l'activité économique ou la concurrence entre les acteurs comme un motif légitime pour l'État d'intervention locale sur la propriété des données. De ce point de vue, il s'agit d'une extension du concept « d'intérêt général ». Cependant, les ouvertures de données décidées doivent profiter à des classes d'agents économiques, et pas uniquement à l'État.
- **Législation locale sur le droit de propriété des données de trace** : comme vu précédemment, il n'existe pas de droit de propriété prédéfini pour les données de « traces numériques ». La propriété est réglée au cas par cas par le droit des contrats. Les données d'innovation doivent permettre dans certains cas précis à la puissance publique de trancher en faveur d'une « copossession⁶⁰ » de ces données entre plusieurs classes d'acteurs (ex. les données doivent être à disposition à la fois du constructeur et du sous-traitant).
- **Portabilité pour les données non-personnelles** : dans certains cas d'applications, les données d'innovation pourraient être assimilées aux données personnelles circulant entre les acteurs pour une meilleure concurrence grâce au levier de la portabilité. Dans la directive DSP2, le principe de portabilité redonne au client bancaire le contrôle de ses données, afin que ce dernier puisse les réutiliser auprès d'autres entreprises (fintech). Les données d'innovation pourraient permettre de reproduire ce schéma, par exemple pour de la donnée industrielle. Un exploitant serait en mesure de récupérer la donnée qu'il génère *via* ses machines pour l'utiliser auprès d'un autre machiniste.

6.2. IMPLÉMENTATION

Si le concept des données d'innovation reste très large, et potentiellement applicable à des classes de données très diverses, il convient d'encadrer sa mise en application.

On l'a dit, aucune intervention sur les droits de propriété des données ne saurait être pensée de manière trans-sectorielle. Nous proposons à ce titre que, pour chaque secteur,

⁶⁰ Paul Hofheinz & David Osimo, « Making Europe a Data Economy: A New Framework for Free Movement of Data in the Digital Age », The Lisbon Council, 2017, p. 29-31.

une autorité ayant le pouvoir législatif d'implémenter des ouvertures de données d'innovation pour les entreprises du secteur soit désignée. Cette autorité pourrait être une autorité administrative indépendante ou directement une administration compétente selon les secteurs⁶¹. Nous suggérons ensuite que le principe de données d'innovation soit inscrit dans la loi seulement pour décrire la procédure que devra respecter cette autorité sectorielle lors de sa mise en application.

La procédure se déroulerait en deux étapes.

Tout d'abord, l'autorité en charge devrait coordonner une mission sectorielle, qui pourrait être menée par la DGE (Direction générale des entreprises). Ces missions auraient pour objectif d'étudier en co-construction avec les acteurs économiques la faisabilité de l'ouverture de certaines données d'innovation et la réalisation de premières études d'impact. Il conviendra de cartographier par secteur les données à forts enjeux, en mesurant l'intérêt des différentes classes d'acteurs pour des données privées et en évaluant le potentiel d'innovation relatif à chacune des classes de données.

Si nous envisageons à ce stade le dispositif à l'échelle nationale, une vision plus ambitieuse serait de le porter à l'échelle européenne *via* des études des directions sectorielles de la Commission et des règlements européens.

Nous proposons qu'ensuite, au moment de la législation en faveur de l'ouverture d'une classe de données identifiée au principe des données d'innovation, le régulateur se prononce sur les trois points suivants.

6.2.1. Périmètre de l'ouverture

L'ouverture des données d'innovation doit répondre à deux questions : De qui veut-on ouvrir les données ? À qui veut-on les rendre accessibles ?

Pour la première, la réponse doit s'adresser à une classe d'acteurs sectoriels. La mesure ne doit évidemment pas viser une seule entreprise en particulier (dans ce cas, l'ouverture de données relève des autorités de la concurrence contre l'abus de position dominante).

Pour les destinataires de l'ouverture, nous envisageons trois types de réponse. Le régulateur pourrait choisir selon les besoins identifiés d'ouvrir ces données.

⁶¹ Cette proposition de mise en application par l'intermédiaire d'autorités désignées sectoriellement est reprise du « Rapport sur les données d'intérêt général ».

- À une classe d'agents impliqués dans la génération des données. Dans ce cas, chaque exploitant n'aurait de droit que sur les données le concernant. Par exemple, un exploitant serait en mesure de récupérer les données qu'il a générées avec sa machine, mais pas les données générées par les autres exploitants avec la même machine.
- À une classe d'agents sectoriels. Tous les agents de la classe pourraient revendiquer un accès intégral aux données. Cette configuration devrait rester exceptionnelle pour ne pas donner d'avantage compétitif indu à certains acteurs. Elle pourrait être utilisée dans le cas de données sensibles auxquelles on ne voudrait donner l'accès qu'à une classe d'acteurs régulés. Ce cas serait par exemple pertinent pour rendre des données de santé accessibles aux assureurs ou des données issues de matériels d'armement aux entreprises françaises de défense.
- En open data : tout le monde serait en mesure d'accéder aux données ouvertes.

6.2.2. Standards des données et du moyen d'échange employé

Avant de procéder à une quelconque ouverture de données, le régulateur aura dû co-construire avec les acteurs un standard très précis sur les données partagées. Le type de données partagées devra ainsi être tout à fait similaire d'un acteur à l'autre. Un standard de qualité minimum de la donnée partagée pourra également être envisagé.

Des standards devront également être communiqués pour le mode d'ouverture des données. Dans le cas par exemple d'une ouverture par des API sécurisées, ces API pourraient nécessiter elles-mêmes d'être contrôlées (comme c'est le cas pour la directive DSP2).

6.2.3. Rétribution des entreprises ouvrant leurs données

Dans les cas où les acteurs concernés par une ouverture de leurs données devraient engager des coûts substantiels pour standardiser les données, ou les mettre à disposition de manière sécurisée à d'autres acteurs, le législateur pourra trancher en faveur d'une rétribution obligatoire. Cette rétribution devra uniquement permettre de couvrir les coûts de développement associés, et sera calculée par le régulateur. Elle ne doit pas s'appuyer sur

une potentielle valeur de la donnée (dont on a dit qu'elle n'était pas définissable) et un manque à gagner de l'entreprise qui partage.

Nous envisageons trois modes de rétribution.

- L'État prend intégralement en charge cette rétribution.
- La contribution est demandée aux acteurs souhaitant bénéficier des données sur la base d'un coût fixe.
- La contribution est demandée aux acteurs bénéficiant de l'accès aux données et calculée sur un modèle d'API, avec un « prix à la requête » ou au volume de données téléchargé par les bénéficiaires.

Dans la mesure du possible, le dernier mode de rétribution devra être privilégié. Notons que dans le cas de la directive DSP2, les banques n'ont pas été rétribuées pour le développement des API, conformément aux principes du RGPD.

Les données d'innovation sont un nouvel instrument réglementaire. Il permet à la puissance publique de légiférer en faveur de l'ouverture à des groupes d'acteurs de certains types précis de données d'entreprises représentant un fort potentiel d'innovation. Nous proposons de déléguer aux autorités sectorielles la responsabilité d'identifier et de cartographier les données d'innovation, en partenariat avec les entreprises. Ensuite, les autorités devront respecter la procédure suivante : définir précisément le périmètre d'application des échanges de données, les standards techniques devant être employés pour ces échanges et trancher sur une éventuelle rétribution des entreprises forcées d'ouvrir leurs données.