

RAPPORT DE SYNTHÈSE

Prix et valeur des données dans la plateformes numérique - Repères pour les relations interentreprises

anRT
ASSOCIATION NATIONALE
RECHERCHE TECHNOLOGIE

 **FUTURIS**

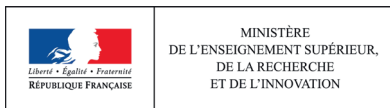
**LE POUVOIR DE
L'INTELLIGENCE
COLLECTIVE**

OCTOBRE / 2019
LES CAHIERS FUTURIS

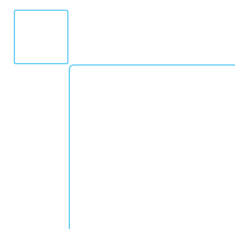
Groupe Pour une politique industrielle du numérique
sous la présidence de Gérard Roucairol, Académie des technologies
Pierre Bitard, ANRT

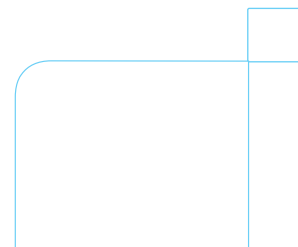
Ces travaux sont soutenus financièrement par les souscripteurs FutuRIS :

AICARNOT, AIR LIQUIDE, ANR, BERGER-LEVRAULT, BOUYGUES, BRGM, CEA, CNRS, CPU, CURIF, EDF, FRAMATOME, GE HEALTHCARE, INRIA, INSERM, INSTITUT MINES TELECOM, INSTITUT PASTEUR, IRIS SERVIER, MESRI, NOKIA, ORANGE, RENAULT, SAFRAN, SNCF, THALES, TOTAL, UBER



Le contenu n'engage que la responsabilité de l'ANRT en tant qu'auteur et non celle des institutions qui lui apportent leur soutien.

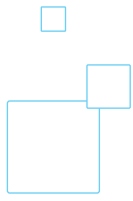




Liste des experts de haut-niveau ayant participé au groupe de travail

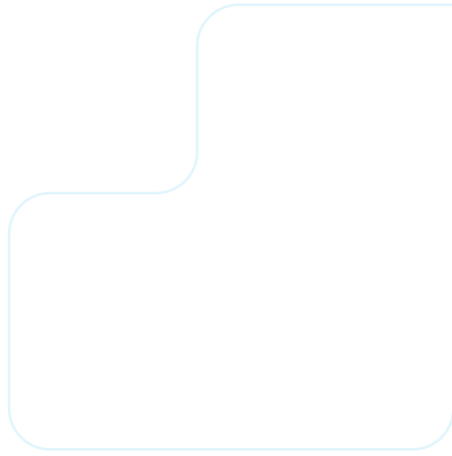
François ANDRY	PHILIPS
Dominique BOLIGNANO	PROVE&RUN
Régis CHATELLIER	CNIL
Jean-Michel DALLE	AGORANOV
Jacques-Etienne GRANDJEAN	WE::NXT
Doh-Shin JEON	TOULOUSE SCHOOL OF ECONOMICS
Olivier LE GALL	INRA et OFI
Claude LE PAPE	SCHNEIDER ELECTRIC
Luca MOLLO	PFIZER
Jean-Manuel QUIROGA	MATHWORKS
Gérard ROUCAIROL	ACADÉMIE DES TECHNOLOGIES, Président du groupe de travail
Hubert TARDIEU	ATOS
Ascension VIZINHO-COUTRY	MATHWORKS
Patrick WAELBROECK	TELECOM PARISTECH
Jonathan WEILL	EDF puis EDVANCE

Ce rapport prolonge et complète le livre blanc *Pour une politique industrielle du numérique* de mars 2018. Il résulte d'un travail d'intelligence collective et propose un état de l'art sur la question du prix et des valeurs dans la plateforme numérique B2B. La substance de ce document provient essentiellement des comptes rendus du groupe de travail des experts de haut-niveau ayant participé. Qu'ils en soient remerciés.



SOMMAIRE

RÉSUMÉ	7
EXECUTIVE SUMMARY	9
INTRODUCTION GÉNÉRALE	13
Les effets paradoxaux de la plateformesisation numérique pour les industriels	14
- Pour commencer, l'intégrateur-systémier	14
- En deuxième lieu, on doit s'intéresser aux entreprises du numérique elles-mêmes	14
Les plateformes sous un angle économique	15
Les multiples dimensions des données : conséquences industrielles	16
Une circulation organisée des données, source de création de valeur	16
<hr/>	
1 / SANTÉ	19
1.1 « La quête des données de vraie vie » ?	19
1.2 Valeur et prix dans les plateformes numériques en santé	21
1.2.1. Connaissance patient et consensus sur la qualité	21
1.2.2. Le dossier médical partagé et au-delà	22
<hr/>	
2 / AUTOMOBILE CONNECTÉE ET AUTONOME	25
2.1 L'unique « open location platform »	25
2.2 De la privacy au sein des plateformes comme bien public	27
2.3 Usages à des fins industrielles des données et algorithmes	28
<hr/>	
3 / ÉNERGIE	31
3.1 L'ubérisation vue comme « ligne rouge stratégique »	31
3.2 Vers une gestion électrique optimisée	32
3.3 Chaînes de valeur des données, trois enjeux industriels clés	33
3.3.1. Comprendre l'alerte d'un système en fonctionnement	33
3.3.2. Passer d'un « modèle appris » à des « données de vraie vie »	34
3.3.2. Améliorer les systèmes en fonctionnement, les faire évoluer grâce aux données	35
<hr/>	
4 / RETOUR SUR LA VALEUR ET LES PRIX DES DONNÉES : ENTRE STRATÉGIE D'ENTREPRISE ET POLITIQUE INDUSTRIELLE	37
4.1 Pour une compréhension commune de la chaîne de valeur des données, métier par métier	37
4.1.1. A environnement cyber-riské, nouvelle prise de conscience	37
4.1.2. La sécurité by design	38
4.2 Associer valeur des données et valeur de la protection. Vers l'industrie de la confiance	38
4.3 Pour une <i>platformonomics</i> : la question des traitements différenciés	39



RÉSUMÉ

La transformation économique et sociale à laquelle contribuent les plateformes numériques ne se limite pas à la prépondérance croissante des GAFAM (Google-Amazon-Facebook-Apple-Microsoft) dans nos vies. L'histoire des plateformes numériques, et de la réussite économique qu'il est possible d'en attendre, n'est pas finie. Ni pour la France, ni pour l'Europe. Il existe une autre plateformes numérique, moins visible bien que plus profonde, qui se manifeste au sein des relations entre entreprises (B2B). Là se situe une chance à saisir pour notre industrie.

Ce rapport constitue donc un plaidoyer qui encourage les démarches publiques et privées concourant à l'organisation des chaînes de valeur des données au plus près des métiers et secteurs concernés. C'est dans cet esprit qu'ont été émises les recommandations du livre blanc *Pour une politique industrielle du numérique* (2018). Prolongeant ce travail, nous approfondissons les conditions dans lesquelles les écosystèmes métier sont capables de se coordonner pour créer de la valeur. Et ce, à travers l'examen de trois des secteurs les plus avancés en ce domaine : **la santé, l'automobile, l'énergie**. Notre grille de lecture fait de l'écosystème du domaine technologique le juste lieu de définition des comportements des entreprises. C'est le cas en matière de prix et de valeur des données. La description des enjeux propres à un secteur est donc l'occasion d'éclairer l'action par des repères spécifiques. Ils sont au nombre de trois :

1. Ce qui fait la valeur des données : les traitements croisés dont elles peuvent faire l'objet

Avant de créer de la valeur à partir de données, encore faut-il qu'une entreprise se dote de capacités à combiner puis à traiter ces données issues de sources variées, internes comme externes à l'entreprise. Ces traitements nécessitent des investissements matériels et logiciels importants et résultent de contributions d'experts, rares. Ils concernent des maillons spécifiques de la chaîne de valeur des données. Echanges, partages, achats croisés ou trocs s'expriment au regard de quantités et de qualités des données (penser la donnée comme un vecteur, au sens mathématique, pas comme un point du plan). Nous plaidons donc pour des démarches délibérées d'établissement d'une compréhension commune de la chaîne de valeur des données, métier par métier. Ces efforts, à l'initiative des intégrateurs-systèmeurs impliqués et en coopération avec les entreprises du numérique, forment un socle préalable. La politique industrielle peut contribuer à établir les conditions de l'élaboration d'une lecture commune de la chaîne de valeur des données. C'est la clé de voûte de l'intervention publique en faveur de saines, fructueuses et structurantes relations entre les entreprises en cours de plateformes numériques.

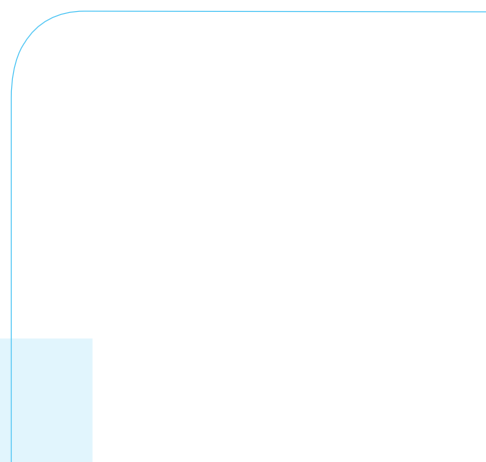
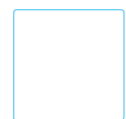
2. Pour l'avènement de l'industrie de la confiance, une sécurité by design

La valeur des données est corrélée à celle de la protection que leur attribuent leurs producteurs, usagers ou propriétaires. La qualité de l'infrastructure informatique, composée de ressources technologiques

interopérables, incorpore en outre un ensemble de règles au caractère immuable et auditable. Cette qualité constitue la condition vitale du développement de la plateforme numérique des entreprises opérant sur les sites français et européens. Une infrastructure de confiance est à portée d'action en France et en Europe. Elle procède de choix technologiques raisonnés et explicables, caractéristiques de la souveraineté. Les investissements publics et privés nécessaires à l'avènement de cette économie de la confiance sont susceptibles de rendements sociaux élevés. A encourager donc.

3. Vers une *platformonomics*, contenir les tentations de traitements différenciés

Les intégrateurs-systémiers et leurs très nombreux fournisseurs et sous-traitants œuvrent aujourd'hui à leur transformation radicale en tant que plateformes numériques, pour des métiers « cœur de leur *business* ». Tous les intervenants de la chaîne de valeur entendent obtenir des rendements économiques des données d'usages supérieurs à ceux des autres. Aucune position n'est donc privilégiée. Le modeste client d'aujourd'hui pourrait devenir le fournisseur de composants/prestataire de service incontournable de demain. Au sein de cette effervescence, les tentations pour les entreprises à adopter des comportements de traitements différenciés (par opposition à la neutralité) sont inhérentes. Les écosystèmes métier ont tout intérêt à se doter de règles explicites qui favorisent l'auditabilité des données, techniques et architectures de données sous-jacentes à la plateforme numérique. Telle pourrait être la recommandation d'une économie industrielle progressiste dédiée à l'analyse des comportements des entreprises au sein des plateformes numériques. Cette *platformonomics* reste encore largement à élaborer.



EXECUTIVE SUMMARY

The economic and social shift partly spurred by digital platforms is not just about the growing dominance of GAFAM (Google-Amazon-Facebook-Apple-Microsoft) in our lives. The story of digital platforms and their potential economic success is not over yet. Neither in France nor in Europe. Another less visible though deeper type of digital platformization is taking place within business-to-business (B2B) relations, bringing significant opportunities for our industry.

This report therefore makes the case for public and private measures aimed at organizing data value chains as close as possible to the professions and sectors concerned. This was the spirit behind the recommendations featuring in the White Paper *Pour une politique industrielle du numérique (For an industrial digital policy, 2018)*. To take this work further, we examine the conditions in which business ecosystems can work together to create value, taking three of the most advanced sectors in the domain, i.e. health, the automotive industry, and energy. Our model views the technological domain's ecosystem as the perfect place to define business behaviour, such as the making of both data price and value. In describing the challenges facing a sector, we thus elucidate action through the following three specific key pointers:

1. Data value stems from combining datasets of various origins

Before value can be created from data, a company must have the capacity to process then combine data, which come from varied sources inside and outside the company. This processing requires significant material

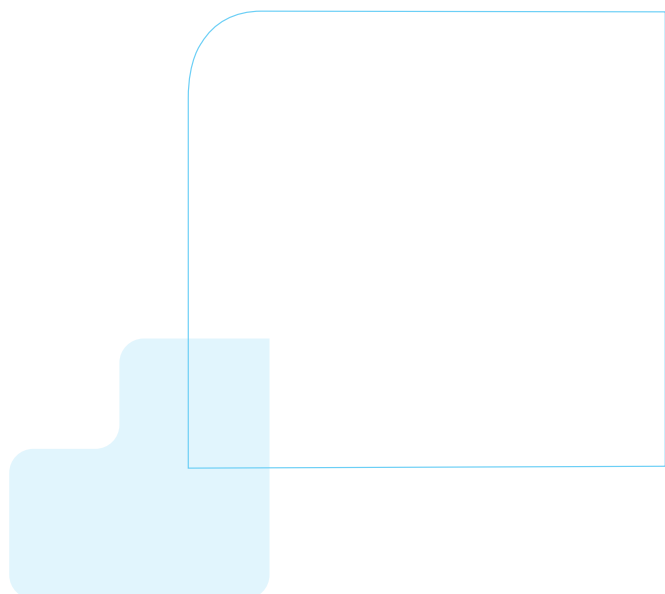
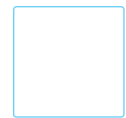
and software investments and results from high-level expert contributions -. It concerns specific links in the value chain of the data. Exchanges, sharing, cross-purchases and bartering are expressed in terms of the quantity and quality of data (thinking of data as a vector, in the mathematical sense, rather than a point on a plane). We therefore advocate taking deliberate steps to establish a common understanding of the data value chain, business by business. These efforts, initiated by the system integrators involved, and working with digital companies, would constitute a base. Industrial policy can help establish the conditions for creating a common understanding of the data value chain. It is the cornerstone of public intervention to foster sound, fruitful and structuring relations between companies in the process of digital platformization.

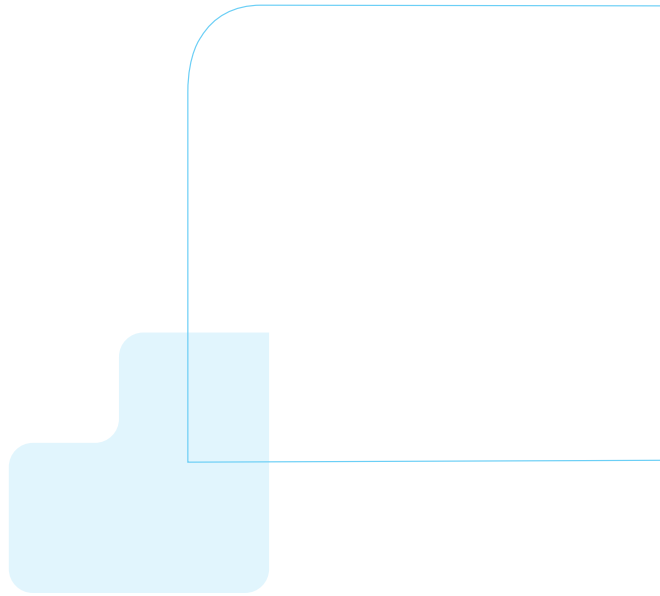
2. Security by design to foster an industry of trust

The value of data is directly related to the extent to which they are protected by their producers, users and owners. The quality of IT infrastructure, comprising interoperable technological resources, incorporates a set of rules that are immutable and auditable. The latter qualities constitute crucial components for developing the digital platformization of businesses operating on French and European sites. An infrastructure of trust is within reach in France and in Europe. It requires making thought-out, explainable technological choices, which are characteristic of sovereignty. The public and private investments required to achieve this economy of trust are likely to bring high social returns - making them well worth encouraging.

3. The move towards *platformonomics* as an approach to resist the temptation of differentiated treatments

System integrators and their numerous suppliers and sub-contractors are currently working on making a radical shift towards digital platforms, for their core businesses. Everyone involved in the value chain intends to obtain higher economic results from data usages than everyone else. No position can be taken for granted. Today's ordinary client could become tomorrow's crucial parts supplier or service provider. Caught up in the buzz, companies inevitably face the temptation of adopting differentiated treatments (as opposed to neutrality). Business ecosystems would do well to establish explicit rules that make it easier to audit the data, techniques and data architectures underlying digital platformization. This could be the recommendation of a progressive industrial policy dedicated to analysing business behaviour within digital platforms. This *platformonomics* still requires significant development.







INTRODUCTION GÉNÉRALE

Les plateformes numériques occupent les premières pages des journaux, mais le propos reste le plus souvent confiné aux seules GAFAM (Google-Amazon-Facebook-Apple-Microsoft). Soit pour les louer, dans la mesure où elles nous simplifient la vie de tous les jours, soit pour critiquer leur hégémonie, à la suite des autorités américaines et européennes de la concurrence. L'intelligence artificielle focalise l'attention comme solution exclusive à de nombreux problèmes. Enfin, « la data » est évoquée à tout propos comme « nouvel or noir », comme si la seule limite à l'enrichissement via les données était celle de l'accès à leurs flux. En outre, les trois sujets sont généralement pris ensemble, sans que cela n'éclaire le débat. La domination des GAFAM laisse peu de place à des succès mondiaux d'ampleur similaire, français ou européens sur ce créneau du B2C (relations entre les entreprises et les consommateurs), en tout cas dans un avenir proche.

Et si une lecture différente de la plateformisation numérique, de la place qu'y prend l'intelligence artificielle et du rôle des données était porteuse d'un espoir raisonné ? Et si une lame de fond montait qui est en train de bouleverser, possiblement en notre faveur, l'industrie française et européenne ?

La plateformisation numérique lorsqu'elle concerne les relations entre les entreprises (B2B) constitue une opportunité économique cruciale, qu'il convient de saisir en France et en Europe. Ce processus en cours va avoir des conséquences profondes sur les entreprises et notre économie. Il convient donc d'encourager les démarches publiques et privées qui concourent à l'organisation des chaînes de valeur des données au plus près des métiers et secteurs concernés. C'est dans cet esprit qu'ont été émises les recommandations du livre blanc *Pour*

une politique industrielle du numérique (2018). Prolongeant ce travail, nous explorons à travers ce texte d'intelligence collective les conditions dans lesquelles les écosystèmes métier sont capables de se coordonner pour créer de la valeur. Et ce, à travers l'examen de trois des secteurs les plus avancés en matière de plateformisation numérique : la santé, l'automobile, l'énergie.

La grille de lecture fait de **l'écosystème du domaine technologique** le juste lieu de définition des comportements des entreprises. Cette approche s'inscrit dans l'analyse dynamique des infrastructures numériques à laquelle s'est livré le livre blanc. Décrire le fonctionnement des infrastructures contemporaines en France - des microprocesseurs aux capteurs, à l'ordinateur, en passant par les réseaux, dont l'internet et sa programmabilité - permet d'en déduire un scénario de transformation. Dans celui-ci, la valeur créée à partir des données procède de leur cycle de mutation en connaissances. Lequel va du monde réel vers l'enrichissement des contenus proposés par les nouveaux intermédiaires de spécialité, grâce à la connectivité étendue.

Comment, dans le mouvement de plateformisation numérique B2B, les données acquièrent-elles une valeur ? Comment établir une compréhension commune de leurs chaînes de valeur qui favorise le partage de la valeur au sein des écosystèmes, grâce à de justes prix ? A quelles conditions d'infrastructures et de régulations ? La description des enjeux propres à un secteur sera l'occasion d'éclairer l'action par des repères issus de la pratique. Aux décideurs dans nos entreprises, au sein des organisations publiques et des politiques publiques de saisir la chance offerte par la plateformisation numérique B2B. L'examen qui suit, au plus près des enjeux, a pour objet de partager, de manière aussi transparente et approfondie que possible, l'état

de l'art sur la question du prix et des valeurs dans la plateformes numérique interentreprises (B2B).

LES EFFETS PARADOXAUX DE LA PLATEFORMISATION NUMÉRIQUE POUR LES INDUSTRIELS

Il est utile d'envisager la plateformes numérique selon deux angles de vue complémentaires, qui correspondent aux deux grandes catégories d'entreprises concernées.

Pour commencer, l'intégrateur-systémier

Chaque constructeur d'automobiles, d'avions, chaque fabricant de tracteurs ou de machines agricoles, de systèmes électriques, chaque entreprise pharmaceutique productrice de médicaments ou de vaccins, chaque fabricant d'équipements médicaux, mais aussi chaque producteur et chaque distributeur d'énergie mobilise ses ressources pour saisir l'opportunité numérique afin de conforter sa position concurrentielle. Par la plateformes numérique, chacun d'entre eux entend bien, sur son marché principal et sur ses marchés connexes en amont ou aval de celui-ci, passer du stade de compétiteur à celui de monopoliste. Cette position offre à l'entreprise plus de latitude à l'égard de sa clientèle en matière de prix, de quantité, de délais.

Le constructeur automobile souhaite accroître sa part de la valeur ajoutée issue de l'usage du véhicule en fournissant des services au-delà de la vente. L'entreprise de médicaments veut fournir des dispositifs de suivi de l'observance ou aider le futur patient à participer à un meilleur diagnostic de son état. Le distributeur d'électricité pourrait souhaiter encourager certains usages de l'électricité qu'il achemine, au regard de ses propres contraintes techno-économiques. Tout se passe comme s'ils devenaient les propriétaires des données engendrées lors de l'usage du produit. Cette appropriation s'effectue au prix d'investissements dans une infrastructure informatique et dans des compétences nouvelles. Elle favorise l'extension de leur pouvoir de marché.

Mais dans quelle mesure l'installation de capteurs sur le tracteur fait-il du tractoriste le propriétaire des informations issues de la circulation du tracteur dans tel champ

pour tel usage ? Le tracteur récolte-t-il des informations géolocalisées de qualité des sols ou de consommation de carburant - qui peuvent avoir de multiples usages - ou permet-il au cultivateur de mieux répandre son engrais ? Le motoriste aéronautique fournit-il à l'avionneur des performances d'heures de vol, de puissance, de poussée et de consommation ou des informations sur les paramètres issues du vol des avions équipés ? Le producteur de lait infantile fournit-il des nutriments adaptés aux besoins des bébés, selon leur âge et seulement cela... sans rapport avec la mise à disposition gratuite d'une application de suivi de la croissance de bébé ? A ces questions, une réponse évidente : les données semblent ici intimement liées au produit à travers la relation à la clientèle et les en distinguer s'avère difficile. Aussi, le client s'est vu proposé une amélioration des services rendus par le produit grâce aux informations d'usage collectées. Le tracteur suggère au cultivateur un trajet optimisé de récolte, un emploi optimisé des intrants dispensés par le tracteur, etc. La performance d'usage d'un tracteur est une fonction croissante du nombre d'utilisateurs desdits tracteurs. En effet, le tractoriste bénéficie alors d'une base d'analyse riche de la totalité des usages et données d'usage qu'il peut mobiliser et traiter pour améliorer ses modèles et prédictions. Et les rediffuser sous formes de services de façon ciblée à ses clients.

Ces intégrateurs-systémiers voient leur position sur le marché nettement confortée par la plateformes numérique : prix supérieur à celui que permettrait la concurrence, capacité de lock-in accrue, coûts du changement de produit plus élevé, etc. A l'aune classique des autorités de la concurrence, la question qui se pose est la suivante : à quel point le client y trouve-t-il son compte économique ? L'avantage retiré de la plateformes numérique ne distord-il pas trop la relation à l'avantage du vendeur ?

En deuxième lieu, on doit s'intéresser aux entreprises du numérique elles-mêmes

Se focaliser sur les effets de la plateformes numérique sur les relations interentreprises conduit à envisager les évolutions auxquelles elles ont à faire face. Nul doute que la plateformes numérique remodèle les interactions entre l'intégrateur-systémier et les fournisseurs et sous-traitants grâce auxquels il fabrique ou

offre ses services. Toutefois, il n'est plus un produit-service complexe dont la totalité des composants et sous-systèmes soit issu des chaînes de production de l'intégrateur-systémier. La plateformes numérique renforce ceci. Ainsi, si le fabricant d'équipement médical ou le motoriste aéronautique intègrent capteurs et actionneurs dans leurs produits, il est douteux qu'ils en soient eux-mêmes les concepteurs et les fabricants. De même, une partie des compétences de traitement et d'analyse des données ainsi recueillies se trouveraient, temporairement ou de manière plus durable, à l'extérieur de l'entreprise intégratrice. Une automobile récente comporte entre 100 et 150 capteurs, dispose d'ordinateurs de bord, de nombreux logiciels embarqués et d'un accès réseau. Ce dernier permet de transférer, de sauvegarder et de traiter les données sur des serveurs dédiés. Seuls certains de ces composants sont conçus et produits par le constructeur lui-même. Réciproquement, l'essentiel de la valeur nouvelle que peuvent permettre d'apporter au véhicule ces composants, à l'issue de traitements de données, implique une myriade d'intervenants.

Pour les intégrateurs-systémiers, la plateformes numérique a donc des effets paradoxaux vis-à-vis des entreprises du numérique, quel qu'en soit le secteur (pharma, auto, aéronautique, défense, éducation, électricité, etc.) :

- leur dépendance à l'égard des composants et des compétences s'accroît ;
- leur pouvoir s'en trouve aussi renforcé, les entreprises du numérique ayant besoin des données de l'usage des IoT qu'elles ont conçues, ne serait-ce que pour maintenir la qualité du service rendu par leur composant ou ensemble de composants.

Du fait des technologies du numérique, la tension concurrentielle s'est donc étendue : aux effets classiques qui touchent les consommateurs se surajoutent ceux affectant la structure des marchés. L'insertion dans les chaînes de valeur de la donnée modifie autant la capacité des entreprises à orienter leur système de prix en aval que leur aptitude à influencer sur leurs contraintes technologiques en amont.

LES PLATEFORMES SOUS UN ANGLE ÉCONOMIQUE

Les plateformes numériques maintenant bien connues du grand public comme celles de Google, Amazon, TripAdvisor, Airbnb ou BlablaCar fonctionnent très fréquemment comme des « marchés bifaces ». Le plus souvent, une « face », soit l'un des marchés auxquels la plateforme vend, répond à un besoin d'autres entreprises (B2B) ; l'autre face, l'autre marché de la plateforme, répond à un besoin de consommateurs finaux (B2C). Les utilisateurs de même nature bénéficient « sur leur face » de services spécifiques, propres à leur appartenance à ce réseau-ci et dont l'intérêt est fonction du nombre de ses membres. Simultanément, bien que de manière indirecte, les utilisateurs d'une face enrichissent la performance du réseau de l'autre face, y attirant ainsi plus d'utilisateurs. Ces plateformes permettent une intermédiation, source de création de valeur. Certaines ont d'ailleurs pour vocation directe d'établir le prix des biens qui y sont échangés, à l'instar d'eBay ou de Datalogix parmi les *information (ou data) brokers*.

Les modèles économiques des plateformes citées ci-dessus, le plus souvent classées parmi les plateformes B2C, peinent à s'imposer chez les *pure players* du B2B. Dans le cadre de relations interentreprises en effet, la valorisation des données procède des principes techniques propres au problème à traiter. Ils varient donc selon les secteurs socioéconomiques, et de manière plus précise, selon les spécialités définies : dans la santé, l'anatomo-pathologie par exemple ; dans le transport, l'automobile ; dans l'énergie, l'électricité, voire une source d'énergie intermittente. **A l'échelle industrielle, la maîtrise de la physique du domaine est indispensable, elle s'incarne dans des savoir-faire métier spécifiques.** La valeur du service procuré résulte de la capacité du système-solution à prévoir avec justesse un fonctionnement ou dysfonctionnement : l'occurrence d'une panne, d'un risque pathologique, d'un pic de demande, etc. La donnée – une mesure produite par un capteur par exemple – constitue la source de la valeur créée et la valeur elle-même à l'issue des traitements et analyses proposés. Entre les deux, de nombreux traitements ont permis de transformer les données brutes en informations directement utiles à la précision du traitement.

A un certain stade du cycle de vie d'une brique d'informations correspond un maillon d'une chaîne de valeur globale. A chaque stade du cycle, de la valeur est créée.

LES MULTIPLES DIMENSIONS DES DONNÉES : CONSÉQUENCES INDUSTRIELLES

Confrontés à la domination des GAFAM et des BATX¹, les intégrateurs-systémiers européens et français ont tendance, lorsqu'ils ont besoin d'entreprises du numérique, à s'allier avec ces mêmes GAFAM. Plusieurs de ces alliances ont récemment été conclues par exemple par Renault, Atos ou Sanofi. Pourtant, il n'est probablement pas encore trop tard pour « jouer collectif » et développer des plateformes numériques interindustrielles en santé, dans le bâtiment ou dans l'énergie. L'âpreté d'une concurrence portant simultanément sur de nombreux maillons de la chaîne de valeur des données semble inciter celles qui ne sont pas encore entreprises-plateformes à acheter les services des grandes plateformes numériques installées.

Pourtant, des coalitions d'entreprises complémentaires sur la chaîne de valeur des données auraient du sens et véhiculeraient un avantage économique réel. Aucun intervenant ne peut à lui seul détenir la totalité des informations susceptibles de permettre une valorisation ponctuelle. La valeur des données circonscrite à un maillon de la chaîne résulte toujours d'une agrégation fondée sur l'échange. Echanges, partages, achats croisés ou trocs s'expriment dans les deux dimensions des données. On a tendance à penser d'abord au volume, à la quantité des données, prisme naturel que traduit l'expression *big data*. Pourtant, les données se caractérisent aussi relativement à leur nombre d'attributs (et au degré d'impact de ces attributs sur le phénomène analysé). Chaque individu ou point de donnée se caractérise par un ensemble de qualités. Si l'objectif est de parvenir à une meilleure compréhension d'une certaine pathologie, la recension de la totalité des personnes atteintes à un certain moment pourrait être utilement enrichie. Ainsi, quid des éléments tels leur localisation géographique, la dynamique de celle-ci au cours de l'année qui précède, la connaissance des pratiques

sportives ou à l'inverse de l'absence d'activité physique, la fréquence de leur consommation d'alcool, l'existence d'autres pathologies et des traitements reçus, etc. ? La valeur qui résulterait d'une analyse statistique des prédictors d'une maladie courante pour les patients affectés dépend de la liste des attributs desdits patients qu'il est possible d'analyser. **Ces caractéristiques s'enrichissent les unes les autres et gagnent donc à bénéficier d'une circulation aussi fréquente et étendue que possible.** Au lieu de l'achat d'une compétence technique et de jeux de données génériques déjà établis auprès des GAFAM, les intégrateurs-systémiers européens et français pourraient bâtir des associations nouvelles, au plus près des besoins.

UNE CIRCULATION ORGANISÉE DES DONNÉES, SOURCE DE CRÉATION DE VALEUR

C'est à l'échelle des domaines d'usage que la circulation des données nécessite d'être organisée. Aussi, des standards et normes de qualité sont à développer et à adopter au plus près des usages. La sécurité apportée par les protections normalisées véhiculées par les données est partie intrinsèque de leur valeur. En l'absence d'initiatives concertées et urgentes, les données sous-jacentes à la création de valeur des secteurs industriels cités courent un risque élevé de trouver place au sein des *business models* des GAFAM. La politique industrielle retrouve ici ses lettres de noblesse, en accompagnant les stratégies d'entreprises qui concurrencent les GAFAM : depuis la collecte jusqu'aux données des usagers finaux desdites entreprises industrielles. Une exploitation satisfaisante pour les parties prenantes procèdera toutefois d'un accord structurant, d'une lecture commune, quant aux divers maillons de la chaîne de valeur. L'élaboration d'une lecture commune de la chaîne de valeur des données constitue la clé de voûte de l'intervention publique et des relations interindustrielles structurantes des écosystèmes.

Ainsi, à défaut d'un accord structurant des parties prenantes, validé en droit, les données issues de la circulation d'un véhicule automobile peuvent appartenir au propriétaire du véhicule, au constructeur, au loueur, au conducteur, au fabricant d'un des dispositifs sous-jacents aux

¹- Les initiales de Google, Amazon, Facebook, Apple et Microsoft forment l'acronyme GAFAM, celles de leurs équivalents chinois Baidu, Alibaba, Tencent et Xiaomi, BATX.

services numériques embarqués ou sollicités ponctuellement (bornes de recharge, passage au péage, en caisse dans les stations-services), et utilisés lors du déplacement, etc. Dans tous les cas, en Europe, dès lors que ces données sont à caractère personnel, reliées de manière directe ou indirecte à un individu, celui qui en a le contrôle doit s'inscrire dans le cadre du règlement général sur la protection des données (RGPD), comme le rappelle la Commission nationale de l'informatique et des libertés (CNIL). Et ainsi garantir les droits des individus en matière de protection et de respect des principes de transparence et d'information, le cas échéant via leur consentement. Si celui qui a le contrôle des données et effectue les traitements les rend « anonymes » par agrégation suffisante et/ou d'autres techniques, il s'abstrait du cadre RGPD.

Ce niveau de complexité est commun aux domaines aéronautique, maritime, ou ferroviaire si l'on s'en tient au transport. Des efforts conjoints méritent d'être faits pour le développement de normes communes de communication de l'internet des objets ou IoT², voire pour des possibilités de traitements standardisés. Dans quelle mesure les constructeurs automobiles français et leurs sous-traitants de premier rang mutualisent-ils leurs efforts de collecte ? Dans quelles conditions les informations obtenues grâce aux algorithmes de reconnaissance vocale peuvent-elles être combinées avec d'autres informations d'usage du véhicule ?

A l'instar de celles de l'automobile, les données de santé peuvent, selon les conditions spécifiques d'usage, être de nature industrielle, publique, privée, successivement voire simultanément. Leur interopérabilité s'avère par conséquent un attribut crucial qui potentialise les données de santé. Ainsi, les données du Système national d'information inter-régimes de l'assurance maladie (Sniiram) et celles du Programme de médicalisation des systèmes d'information (PMSI) qui couvrent plus de 85 % de la population française seraient utilement croisées avec d'autres. On peut penser aux données de l'Assistance publique-Hôpitaux de Paris (APHP) ou celles issues des systèmes de santé d'autres pays par exemple. Les industriels de la pharmacie sont le plus souvent présents dans plusieurs

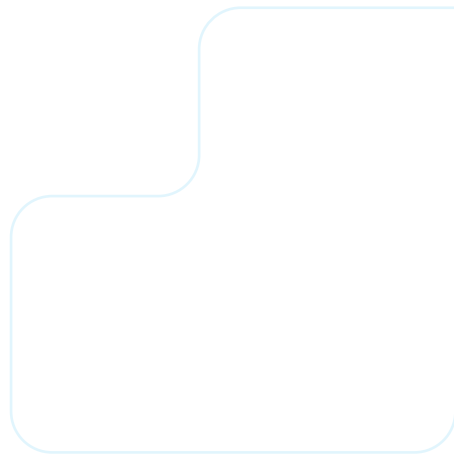
pays. L'efficacité du traitement proposé sur une pathologie dans un pays donné bénéficie de la connaissance de cette même pathologie, et de l'usage du traitement, dans de nombreux pays. L'ouverture de ces données administratives et l'adoption, pour la partie « anonymisable », de normes d'échange reconnues internationalement auront des conséquences vertueuses pour les patients. Cela implique toutefois, à l'échelle des entreprises et des institutions (par ex. des hôpitaux, des centres de recherche), le partage d'une vision intégrée de la chaîne de valeur des données dont chaque maillon est clairement distingué et caractérisé.

L'usage des données, à chaque stade de leur cycle de vie, procède de règles dont les solutions les plus solides sont technologiques. Lesquelles s'accompagnent de la mise en œuvre de méthodologies de conception de contrôles qualité, de processus opérationnels et d'audits adéquats. Les inquiétudes de nature éthique sur lesquelles se focalise le débat public peuvent être levées par la mise en œuvre de choix techniques informés et explicables au public concerné. Ces appréhensions s'expliquent, pour partie, par une insuffisante connaissance des sciences informatiques dans la société. Par conséquent, l'éducation et la formation à l'informatique sont des enjeux socioéconomiques majeurs. Tirer le meilleur parti économique des plateformes numériques passe par une accélération de la diffusion des connaissances en informatique dans l'ensemble du système éducatif français³.

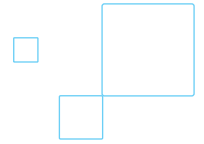
Après cette introduction, nous analyserons les conditions spécifiques de plateformes numériques, et les relations entre entreprises associées, des trois secteurs parmi les plus avancés en ce domaine (la santé, l'automobile, l'énergie). Fidèles à notre grille de lecture (cf. *Pour une politique industrielle du numérique*) qui fait de l'écosystème du domaine le juste lieu de définition des comportements des entreprises, en particulier en matière de prix et de valeur des données, chaque secteur sera l'occasion d'extraire des points de repères dans ce qu'ils ont de plus spécifiques. La conclusion synthétisera les trois enseignements principaux de ce travail d'intelligence collective.

2- On utilise généralement les trois lettres du sigle anglais IoT pour 'Internet of Things', comme dans la suite de ce rapport.

3- Ce sujet majuscule mériterait bien plus qu'une mention. Rappelons seulement ici, à la suite d'A. Vizinho-Coutry (MathWorks) quelques fondamentaux du sujet. L'éducation des jeunes générations est essentielle pour leur permettre de comprendre ce qu'ils peuvent apporter à la société dès le lycée (car c'est là qu'ils choisissent leurs parcours supérieurs) et qu'ils soient plus confiants dans leur quête de métiers. Mais évoquer l'importance de l'apprentissage du codage et de l'algorithmie ne peut suffire. Un point clé concerne l'expertise. Le lieu de la construction de l'expertise, en particulier système, se trouve dans les filières telles que les sciences de l'ingénieur ou la filière sciences et technologies de l'industrie et du développement durable (STI2D). Lesquelles intègrent déjà dans leurs enseignements les systèmes IoT, la gestion de l'énergie, la mise en place de systèmes d'aide à la personne, etc. dans les salles de travaux pratiques.



1 SANTÉ



1.1

« LA QUÊTE DES DONNÉES DE VRAIE VIE » ?

Les développements envisagés vers la médecine prédictive mettent en lumière plusieurs enjeux fondamentaux de la plateformes numériques. La médecine prédictive repose sur une prise en compte holistique de la santé de l'individu. Pour que la médecine préventive puisse apporter tous ses bénéfices aux patients, il ne faudrait pas se contenter de collecter des informations sur la personne une fois malade – i.e. une fois atteinte la condition de patient, dans le système de soin. Un suivi personnalisé fin tout au long de la vie, y compris en tant que personne en bonne santé, permettra, seul, s'il est mis en place, de tirer pleinement parti de la médecine personnalisée, préventive, prédictive et participative, dite des 4P⁴. Une bienveillante vigilance à l'égard des données caractéristiques de la santé en continu, tout au long de la vie, en favoriserait ainsi l'avènement. Les risques d'occurrence de certaines pathologies pourraient ainsi être appréciés. Des alertes précoces permises par l'informatique prédictive réduiraient au maximum l'expression de tel ou tel risque pathologique. Le *monitoring* individualisé, à l'échelle de cohortes aussi vastes que possible, de la médecine prédictive préviendrait l'errance médicale, très coûteuse. Une telle médecine bouleverserait le fonctionnement actuel du système de soin. Elle nécessite d'importants investissements dans des infrastructures ajustées, avec un haut niveau d'interopérabilité. En Europe, et en France en particulier, ce sont généralement des opérateurs publics qui incarnent les garants

d'une opérationnalité désintéressée et protectrice. Leur mobilisation constitue donc un préalable à l'avènement de cette nouvelle économie de la santé fondée sur la médecine prédictive. Les plateformes numériques spécialisées, offrant chacune des services de médecine préventive, requièrent cette couche infrastructurelle de confiance.

La valeur des données découle de l'établissement de standards, lesquels ne se développeront que sous des conditions restrictives d'usage. Il est admis que l'on adhère à une plateforme numérique de services seulement si notre contribution d'aujourd'hui fait notre bénéfice personnel demain.

Là encore, la technique n'est pas loin. La capacité de traitements anonymisés de données collectées à l'échelle individuelle pour des résultats d'analyse personnels et confidentiels, qui mettent en jeu des croisements d'immenses bases de données n'est pas à la portée de tous. En santé en particulier, l'anonymisation de certains jeux de données continue de poser de redoutables problèmes. La pseudonymisation est alors à envisager, y compris en respect du cadre RGPD. Les individus auraient à être informés (consentement éclairé, droit possible de retrait), étape par étape, tout au long de la chaîne de traitements de leurs données. Ce processus peut sembler complexe de prime abord, mais il est à mettre en perspective durable de construction progressive de la confiance avec les utilisateurs⁵.

4- Cf. <https://www.carnetsdesante.fr/+La-medecine-des-4P-prediction+>

5- Cf. Cahier IP n° 5, CNIL, septembre 2017 : https://linc.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip5.pdf

Au-delà, cette maîtrise prouvée ne peut suffire à susciter l'adhésion, laquelle nécessite d'être entretenue par une communication claire qui valorise le « bénéficiaire utilisateur ». Ici, il est possible de s'interroger sur le caractère favorable du coup de projecteur actuel sur les techniques d'apprentissage artificiel. Puisque deux niveaux d'argumentaires de confiance sont à développer :

- celui du bénéficiaire utilisateur du partage tout au long de la vie de ses données de santé ;
- celui de la solidité des algorithmes d'apprentissage profond (dont même les partisans qualifient le fonctionnement de « boîte noire »).

La dimension méthodologique mérite d'être mise en avant, en distinguant les différents temps des données de vraie vie : leur production et les situations de production, dont le recours à de l'instrumentation ; la collecte et ses conditions, dont le contrôle qualité⁶; l'agrégation, le classement, l'indexation, la hiérarchisation, l'organisation des données ; les traitements statistiques combinés à l'emploi de modèles exacts ; la restitution de l'analyse du diagnostic vers le patient, par le professionnel de santé. **A aucun moment de la chaîne de valeur de la donnée, « la machine n'agit seule ».** Ainsi les outils d'analyse d'images radiologiques ou d'anatomopathologie de Philips facilitent certains tris préliminaires. Mais le radiologue ou l'oncologue finissent toujours par prendre la décision. L'accréditation FDA (U.S. Food and Drug Administration) porte sur le système d'anatomopathologie Digital Pathology - Intellispace Pathology. Le dispositif numérise des lames cytologiques et histologiques (e.g. biopsies). Ce n'est donc pas spécifiquement de l'analyse radiologique grâce à l'intelligence artificielle. Des principes similaires régissent le fonctionnement des outils d'annotations et segmentations automatiques sur ces deux types d'images (radios ou lames numérisées). Avec un même résultat visé, le repérage de cas suspects.

A l'échelle du système de soin, dotées d'une approche méthodologique clarifiée et explicable, les autorités de santé pourraient disposer d'une capacité améliorée de pharmacovigilance. Sur un maillon de la chaîne de valeur des données de santé, l'emploi de techniques d'apprentissage artificiel à partir de flux d'informations des réseaux sociaux, offrirait une capacité nouvelle et complémentaire de détection de signaux d'alerte. L'imagerie médicale assistée par ordinateur bénéficie déjà des apports des techniques d'apprentissage artificiel. En cas de diffusion étendue, les conséquences sur la fluidité du parcours de soin, de la prise de rendez-vous à la consultation, en passant par le diagnostic et la délivrance du soin seraient importantes. Et permettraient très certainement de réaliser des économies.

Dans la chaîne des données médicales, la traçabilité sur l'origine constitue le point clé. Fonder son approche sur les « données de vraie vie » devrait s'accompagner de la description précise des protocoles de collecte, y compris l'homologation des matériels et la formation du personnel. En France, l'hôpital conserve les données personnelles de ses patients et est responsable des traitements réalisés sur les données. Les patients sont informés et signent un accord avec l'établissement hospitalier pour les finalités définies dans cet accord. Les établissements hospitaliers sont des lieux de respect absolu de normes, et considérés comme tels par la population. La mise en œuvre des tâches d'indexation et d'anonymisation représente ensuite un coût élevé. L'hôpital paye pour le matériel, le logiciel et les services du dispositif médical. Les volumes de données à gérer et stocker sont gigantesques, on parle d'une dizaine de gigaoctets par patient et par an pour un hôpital⁷. Un tel volume peut justifier une colocation des données sur leur lieu d'usage⁸.

6- C'est un point central en santé, cf. pour la cohorte Constances : « Le contrôle qualité est un élément essentiel de la réussite de Constances car il est crucial que les données qui feront l'objet d'analyses pour la recherche soient rigoureusement validées, particulièrement dans le contexte d'un projet « multicentrique » où les personnes volontaires peuvent effectuer un examen dans une vingtaine de centres différents : il faut faire en sorte que le résultat des examens pratiqués soit strictement comparable, quel que soit le centre où ils ont eu lieu. Idéalement, la même personne qui passerait le même jour un examen de santé à Lille, Marseille ou Bordeaux, devrait avoir exactement les mêmes résultats ». <http://www.constances.fr/coulisses-constances/contrôle-qualité.php>

7-En janvier 2017, la Commission nationale de l'informatique et des libertés (Cnil) a autorisé l'Assistance publique-hôpitaux de Paris (AP-HP), avec ses 39 établissements, à établir son propre entrepôt de données (EDS). Cet entrepôt héberge les informations administratives, médicales et sociales de 8 millions de patients, 163 millions de résultats d'examens biologiques et 5 millions de comptes rendus médicaux. L'EDS entend favoriser « l'accélération » de la recherche scientifique grâce aux études qu'il autorise. Cf. <https://recherche.aphp.fr/eds/>

8- D'autant que la dizaine de Go évoquée représente une moyenne, à un certain niveau de numérisation, en France. Il faut noter que lorsqu'il s'agit de données numériques radiologiques, génomiques ou de pathologies (« anapath ») alors le volume peut s'élever à plusieurs terabytes par patient.

1.2

VALEUR ET PRIX DANS LES PLATEFORMES NUMÉRIQUES EN SANTÉ

Ces quelques illustrations incitent à penser que les entreprises qui entendent installer leur plateforme numérique sont nombreuses et variées au sein du système de santé. Pour des raisons de logique industrielle, l'initiative se trouve souvent du côté des fabricants d'équipements et dispositifs médicaux. Lesquels développent et proposent des services de plus en plus enrichis grâce à la numérisation.

L'usage de ces plateformes numériques suppose des mécanismes de tarification qui engagent un entrelacs complexe de données. La valeur créée se matérialisera par des prix qui lient les industriels de la santé aux organisations et professionnels du système de santé. Et ce, sur tous les maillons d'une chaîne de relations qui va jusqu'au patient.

Ces prix peuvent être temporaires ou forfaitaires ; le marché peut aussi prendre la forme d'un troc. Les prix dépendent directement des données qui sont produites, mobilisées ou traitées par la plateforme. A chaque stade de leur cycle de vie - brutes, indexées, traitées, modélisées, etc. -, au sein d'une chaîne de valeur industrielle spécifique, peuvent correspondre des valeurs et des prix distincts. Mais une partie de ces données peut aussi être échangée, voire valorisée directement sur un marché. Trocs, échanges, ou marchandisation concernent le plus souvent des stades de transformation différents. L'identification de la signature génomique d'une maladie, le traitement statistique totalement anonymisé d'un ensemble de patients, l'élaboration et la maintenance d'une base de connaissances faisant apparaître des causalités, autant de stades de transformation, autant de valeur créée.

En santé, le bout de la chaîne de valeur des données se matérialise en une base de connaissances faisant apparaître des causalités. Cette base de connaissances résulte d'un processus de modélisation rigoureux, sous-tendu par un contrôle qualité sévère. Cet outil constitue le « produit final » de l'informatique

prédictive. Plusieurs modèles de prix sont alors possibles, qui combinent achats et locations, et s'appuient sur la propriété intellectuelle créée. Sont concernés aussi bien les algorithmes, le modèle sous-jacent, les process et les structures. Les données annotées peuvent faire l'objet d'un emploi pour un usage propre, et sous certaines conditions, être vendues ou échangées.

Il peut être utile de distinguer entre les données sources et leurs usages à des fins de modélisations. En effet, entre les établissements hospitaliers et les sociétés d'équipements et dispositifs médicaux, c'est sur cette distinction que repose le troc selon lequel s'organise la relation entre les parties. L'accès aux données produites par les équipements est échangé contre l'usage des modèles et de leurs résultats. De tels mécanismes sont adoptés dans d'autres secteurs. Ils sont bien adaptés au monde de la recherche, lorsque l'atteinte des résultats demeure aussi incertaine que les gains économiques potentiels associés.

1.2.1 Connaissance patient et consensus sur la qualité

Les phases d'études cliniques sont longues et très coûteuses. Les grandes compagnies pharmaceutiques se transforment pour opérationnaliser la médecine des 4P : personnalisée, préventive, prédictive et participative. Ces dimensions reposent essentiellement sur une gestion intelligente des données de santé. La « connaissance patient » préalable aux essais cliniques de phase 3 procède de la sélection d'une population la mieux adaptée aux tests qui nécessitent d'être effectués. Franchir un cap en matière de « connaissance patient » ferait gagner un temps précieux et ferait économiser beaucoup d'argent aux compagnies pharmaceutiques. Des bases de fichiers de jumeaux numériques de patients choisis seraient pertinentes. Les méthodes d'exploitation de ces données « vie réelle » s'améliorent par un usage récurrent. Un marché pour les plateformes numériques interindustrielles de la santé s'est développé : le *population health management*. Marché sur lequel Philips s'est positionné⁹. Certaines barrières légales demeurent, dans de nombreux pays, à cet usage des données de vie réelle pour révéler des certitudes précoces. Nul doute que l'établissement de consensus sur la qualité

9-Rachat de deux compagnies spécialisées : Wellcentive en 2016 et VitalHealth en 2017.

des données et modèles associés favorisera l'émergence des marchés. Ces consensus doivent passer par des collaborations entre organisations du système de santé. Le projet HU-PRECIMED de l'Association française des sociétés de services et d'innovation pour les sciences de la vie (AFSSI)¹⁰, qui vise à structurer une filière de médecine de précision en France peut être cité comme exemple.

Des collaborations entre acteurs concurrents, i.e. présents sur le même maillon de la chaîne de valeur de la donnée de santé, sont possibles lorsque l'objet du consensus concerne les standards de représentation. Les standards et normes de qualité des données et modèles constituent par conséquent un enjeu vital pour le développement de nouveaux marchés de la santé. L'entreprise technologique conjointe Innovative Medicines Initiative (IMI)¹¹, deuxième itération (2014-2020) de l'UE, dispose d'un budget public-privé de près de 3,3 milliards d'euros et s'attache à développer de tels modèles.

1.2.2 Le dossier médical partagé et au-delà

En France, depuis novembre 2018, la mise en œuvre du dossier médical partagé (DMP, sous l'égide de l'Assurance maladie) pourrait ouvrir des perspectives économiques à l'ensemble des acteurs du système de santé.

Le DMP est considéré comme un carnet de santé numérique, avec comme propriétaire des données, le patient. Le dossier conserve la mémoire de la santé des patients : historique de soins des 24 derniers mois, traitements suivis, résultats d'examens, antécédents médicaux, comptes rendus d'hospitalisation, etc.

Pour l'Assurance maladie, le DMP favorise la coordination et la qualité des soins entre tous les professionnels de santé, y compris à l'hôpital¹². Les informations de santé du patient peuvent être déposées dans le DMP par les professionnels de santé qui le prennent en charge ou par le patient lui-même. Elles sont accessibles à tout moment sur le site dmp.fr ou via l'application mobile DMP. Elles peuvent être consultées, avec l'accord du patient, par son médecin traitant ou par tout professionnel de santé qui est amené à le prendre en charge. À chaque instant, le patient

garde le contrôle de l'accès aux données par un service sécurisé. Il peut ajouter une information (personne à prévenir en cas d'urgence par exemple) ou décider de masquer certains documents. Enfin, il est averti par courriel si un nouveau document est ajouté.

En juin 2019, plus de 6 millions de personnes disposent de leur DMP. Dans son fonctionnement actuel, il est déjà possible d'anticiper des économies, voire des améliorations de la prise en charge des patients. L'enjeu pressant des entreprises de santé consiste à obtenir l'accès à cet entrepôt unique pour effectuer des recherches nouvelles, toutes conditions d'anonymat et de *privacy* prises. Le Centre d'accès sécurisé aux données (CASD)¹³ est l'un des spécialistes français qui proposent des solutions à ces problématiques d'accès sûr. L'autorisation d'accès aux données de tels entrepôts permettra la floraison de nouveaux services créateurs de valeur. Ces services supplémentaires seront offerts aux patients grâce à des croisements judicieux avec d'autres sources de données. Au-delà des aspects réglementaires, l'émergence de tels services suppose que les personnes soient convaincues de la pertinence des services qui pourront leur être apportés en échange. Et ce, en toute confiance et sécurité.

Les traitements statistiques massifs des données de santé doivent pouvoir être effectués en préservant l'anonymat du patient. La preuve de cet anonymat devrait être démontrable et démontrée par l'acteur qui effectue les traitements. Les techniques de la chaîne de blocs pourraient être mobilisées à cette fin, en tant qu'« outil de traçabilité », selon certains. L'informatique industrielle classique met en œuvre des systèmes de jetons (token), des techniques de traitement rapides, précises et fiables de preuves d'anonymat par authentification.

Pour répondre aux besoins qu'elles ont identifiés, les plateformes numériques de la santé ont à pratiquer une dialectique complexe d'accès et d'usages de données nominatives et de données anonymisées. L'exploitation de données agrégées, correspondant à des profils types par exemple, ouvrirait de nombreuses possibilités d'améliorations thérapeutiques. La méga base

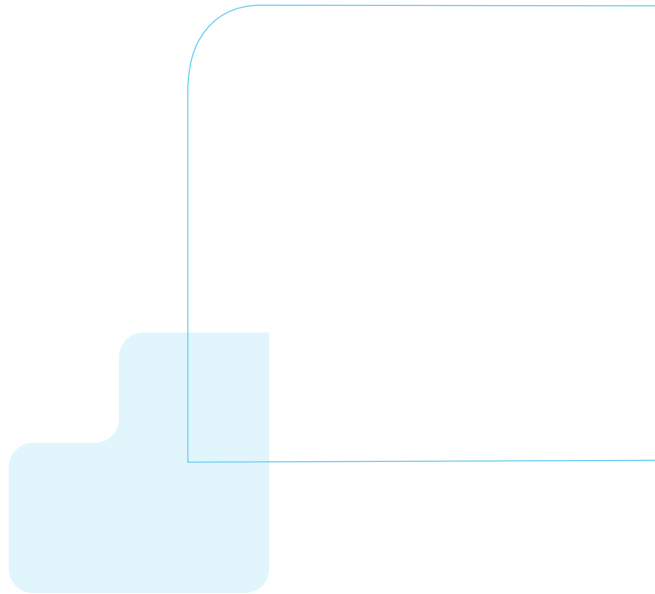
10- <https://www.afssi.fr/blog/hu-precimed/>

11- <https://www.imi.europa.eu/>

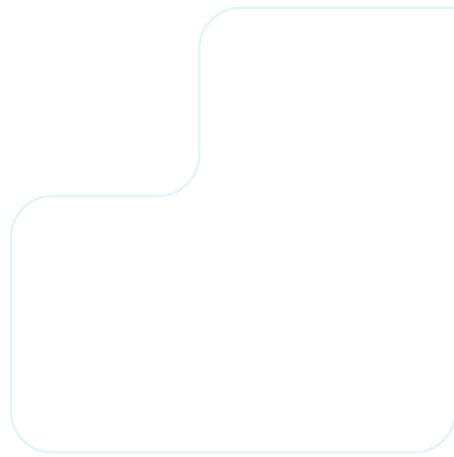
12- <https://www.ameli.fr/>

13- <https://www.casd.eu>. Ce groupement d'intérêt public rassemble l'État représenté par INSEE, le GENES, le CNRS, l'École polytechnique et HEC Paris ; il bénéficie d'un financement « Equipex » du Plan Investissement d'Avenir.

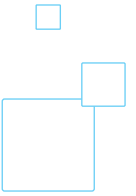
de données qu'est le Sniiram est conçue pour le remboursement des frais de santé. Elle procède donc du traitement d'informations nominatives. Etant donné les enjeux et la complexité des informations manipulées, la maintenance, et le maintien du haut niveau technologique nécessaire sont coûteuses. Se pose ensuite la question de la nature des traitements statistiques et algorithmiques réalisés avec des données. En particulier, dans le contexte de la diffusion des techniques d'apprentissage artificiel. Focalisés sur les découvertes de corrélations permises par ces analyses de données, certains praticiens de l'apprentissage artificiel peuvent négliger les précautions élémentaires d'usage de ces outils statistiques. Des biais bien connus des statisticiens peuvent être rencontrés : biais de sélection et d'endogénéité, biais de présentation, voire biais cognitifs du programmeur, mais aussi des biais économiques, parfois mobilisés à dessein¹⁴.



14- <https://www.telecom-paristech.fr/recherche/publications/algorithmes-biais-discrimination-equite.html>



2 AUTOMOBILE CONNECTÉE ET AUTONOME



2.1

L'UNIQUE OPEN LOCATION PLATFORM

Le secteur automobile apparaît comme l'un de ceux où les attentes sont les plus grandes à l'égard de l'informatique prédictive et où la concurrence est rude pour y imposer sa plateforme. Constructeurs et équipementiers entendent bien imposer la leur. En outre, des plateformes métier ou de spécialité s'attachent à s'imposer, à l'échelle mondiale, sur le plus grand nombre possible de véhicules. En matière de plateformes interindustrielles cependant, il y a pour le moment bien peu d'exemples de domination du marché. Peu de *use cases* dont les gestionnaires, stratèges et économistes auraient pu tirer les enseignements. Ce qui est vrai pour l'automobile reste vrai pour de nombreux secteurs aujourd'hui. La tension entre intégrateurs-systémiers et entreprises du numérique pour s'imposer partout, indépendamment d'une compréhension commune de la chaîne de valeur des données, freine l'émergence d'un *dominant design*. Si l'intérêt pour la brique technologique «services cartographiques de précision» ne fait pas tout à fait l'unanimité dans le secteur¹⁵, ce créneau fournit pour le moment l'exemple le plus abouti de plateforme interindustrielle numérique qui réunit le plus d'atouts pour faire école.

L'entreprise HERE¹⁶ Technologies fournit des services de cartographie et de navigation avancées pour les automobiles d'une part, et des services d'informations et d'alertes avancés pour véhicules connectés d'autre part. Elle fournit enfin des services cartographiques avancés pour les applications industrielles ou du secteur public.

L'entreprise collecte et achète des données cartographiques des réseaux routiers, des bâtiments, des parkings ou des données de localisation du trafic routier, et celles concernant les conditions météorologiques. Avec ce matériau, elle établit des cartes dynamiques et à haut degré de précision pour près de 200 pays. Elle vend et licencie ses services de cartographie et de navigation à la plupart des constructeurs automobiles dans le monde, avec une part de marché mondiale d'environ 80 % (hors Chine, Corée et Japon). Les actionnaires de la société, Daimler/Mercedes, BMW, Audi, Bosch, Nokia, Intel, Continental et Pioneer, représentent les principaux clients. HERE Technologies vend aussi ses services à d'autres entreprises telles Amazon ou Garmin, et a développé plusieurs applications grand public pour simplifier la mobilité.

HERE Technologies a fait preuve d'opportunisme et s'est attachée, au-delà du secteur automobile, à développer des marchés adjacents auxquels les services de cartographie avancée pour les automobiles peuvent donner accès. Sa flotte de 400 véhicules circulant en permanence dans le monde collecte les informations contextuelles associées. HERE se fait alors *information broker* et revend ces informations. Elle fournit également des services cartographiques en ligne. De fait, la plateforme HERE sert de multiples marchés, essentiellement B2B, des secteurs automobile, transport-logistique, infrastructures et secteur public, média, assurance, électronique, Telco et utilities, grande distribution, santé ou immobilier.

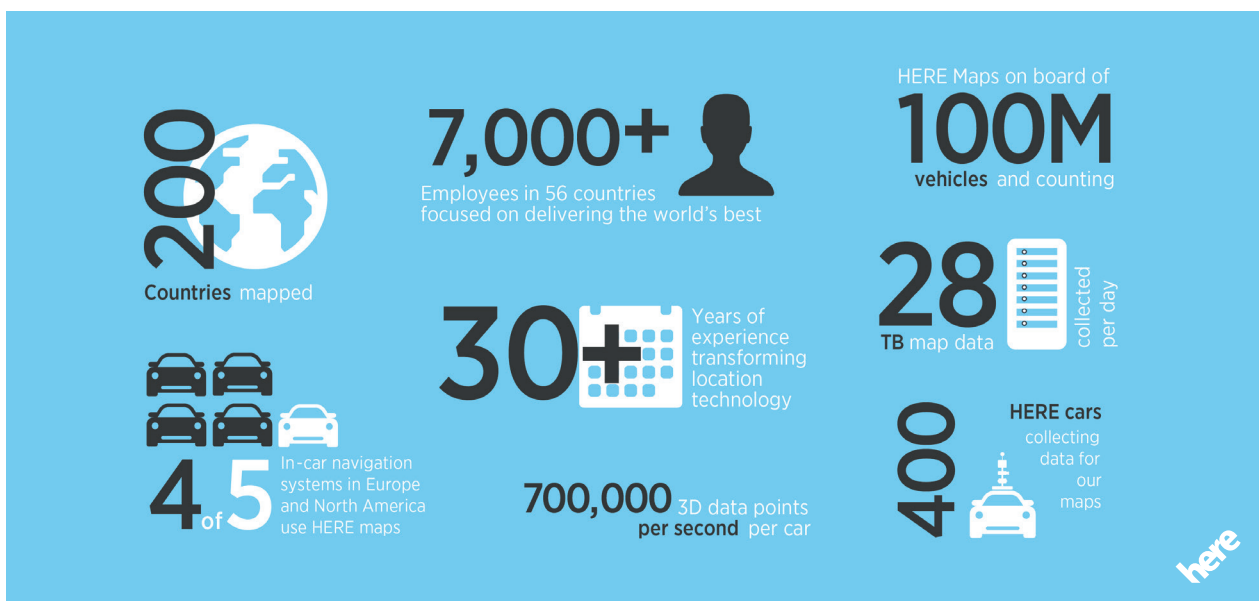
15- « High-precision GPS maps for self-driving cars are a "really bad idea," resulting in a "system [that] becomes extremely brittle" by being too dependent and not being able to adapt», Elon Musk, Autonomy Day Event, The Verge, 24 avril 2019 cf. <https://www.theverge.com/2019/4/24/18512580/elon-musk-tes-la-driverless-cars-lidar-simulation-waymo>

16- L'histoire de cette société fondée en 1985 par Karlin & Collins, Inc., sous le nom de Navigation Technologies Corporation mériterait des développements plus amples que ne nous le permet ce rapport. Le lecteur intéressé pourra consulter [https://en.wikipedia.org/wiki/Here_\(company\)](https://en.wikipedia.org/wiki/Here_(company))

HERE Maps et HERE AUTO fournissent des services aux constructeurs automobiles, leur première clientèle, alors que HERE Location Services, HERE Mobile SDK et HERE Data Lens ciblent les entreprises des autres secteurs industriels. Enfin, HERE WeGo et HERE SoMo sont des services destinés à une troisième catégorie de clients, les consommateurs utilisateurs de moyens de transport multimodaux. Selon D&B Hoovers, l'entreprise serait en déficit de 255 millions d'euros en 2017/2018.

et traiter en temps réel les données issues des capteurs des automobiles connectées, afin de fournir en temps réel des services de cartographie haute définition (HERE HD Live Maps) et des services d'information et d'alerte avancés pour véhicules connectés (HERE Connected Vehicle Services). Plus d'un million de véhicules Audi, BMW et Mercedes connectés alimentent aujourd'hui cette plateforme en temps réel grâce à des données anonymisées. Cependant, la plateforme est « ouverte » à tous, en entrée comme en sortie : le portefeuille de

HERE, chiffres clés



Source : Présentation de J.-E. Grandjean pour le groupe de travail de FutuRIS, 17.01.19

L'avènement du véhicule autonome constitue l'horizon stratégique dans lequel se projette HERE. La navigation temps réel d'un véhicule autonome de niveau 4 ou 5 (sans intervention du conducteur) requiert une précision de géolocalisation à 10 cm près. La position exacte du véhicule, au regard de l'ensemble de ses contraintes d'environnement, y compris les voies et les autres véhicules, représente la matière première complexe de tous les modèles de comportement en interaction. De manière générale, les immenses quantités de données issues des capteurs de l'automobile sont loin d'être toutes exploitées.

A partir de ce constat, HERE Technologies a développé son offre *Open Location Platform* (OLP). Cette plateforme permet de collecter

services à valeur ajoutée générés par HERE sur OLP est disponible pour tout client automobile ou industriel, indépendamment de la contribution ou non des données sous-jacentes. De plus, tout acteur peut utiliser OLP pour développer ses propres services à valeur ajoutée, pour son compte ou en vue de commercialisation.

Du fait de l'intégration, la plateforme HERE contractualise directement avec les constructeurs et bénéficie ainsi d'un accès étendu et profond. Sur la forme, il s'agit d'une situation comparable à celle des entreprises de l'instrumentation médicale qui contractualisent avec les hôpitaux. La valeur potentielle engendrée par l'utilisation de tels systèmes de navigation avancée naît du croisement et du partage d'informations émanant de capteurs de natures et de marques

variées. Un degré de concentration plus ou moins élevé des informations, en un lieu unique, ou en tout cas chez un unique acteur propriétaire des données, est donc évoqué afin d'en simplifier l'accès. En Europe, et à l'échelle internationale, le concept de « véhicule étendu », développé sous forme de norme (ISO 20077-1), fait donc l'objet de travaux. Les constructeurs automobiles sont partisans de cette solution, seule à même selon eux d'apporter l'efficacité et le bon niveau de protection de la *privacy*.

Parallèlement, afin de garantir l'interopérabilité nécessaire à l'exploitation des données produites, HERE a rendu public un format homogène d'ingestion de données de capteurs : SENSORIS, pour Sensor Ingestion Interface Specification. Ce format décrit les prérequis techniques pour la production et l'échange des données issues des capteurs des véhicules roulants. Grâce à lui, les données recueillies en roulant par les capteurs sont envoyées dans le nuage et les cartes des véhicules sont actualisées à la volée. Les véhicules abonnés aux services à valeur ajoutée de la plateforme OLP sont alors aussi alertés en temps quasi-réel des obstacles ou des accidents de la route à proximité.

Au cœur de la proposition de valeur d'HERE Connected Vehicle Services se trouvent l'info-traffic temps réel, la mise à jour des permanents changements de signalisations routières, les signalements de risques divers ou encore la détection de places de parking libres. Autant de services qui pourraient être considérés comme faisant appel à des données d'intérêt général. Conformément à la directive européenne du 20 juin 2019¹⁷, le secteur public devrait alors les mettre à disposition à titre gratuit. Cela ouvrirait la porte à des possibilités de développements économiques. La création de valeur procède des traitements appliqués aux données : nettoyages et vérifications de qualité, éventuelles indexations, mises en forme pour des traitements statistiques de masse, éventuels dressages des algorithmes, etc. Autant de tâches onéreuses dont les résultats doivent être rendus désirables aux yeux de clientèles bien identifiées.

La « compétence cœur » d'HERE consiste à mettre en œuvre, pour chaque strate de sa plateforme OLP, le *business model* qui sert le mieux sa stratégie

de recherche de rentabilité. A chaque strate de la plateforme, soit « *data market place* », « *data* », « *developer environment and platform foundation* », « *services and solutions* », des conditions techno-économiques spécifiques s'appliquent dont il s'agit de savoir exploiter le potentiel pour un prix profitable.

Les efforts managériaux et de gouvernance sous-jacents visent à faire cohabiter :

- la nécessité d'un développement international aux conditions locales contraignantes avec
- un très conséquent besoin de *cash* pour les investissements dans les infrastructures technologiques et les compétences rares associées.

Les caractéristiques spécifiques de cette plateforme, son caractère multicouche pour des conditions locales de rentabilité variables, constituent une source de dynamisme pour son *business model*. Que des informations de même nature puissent être vendues à des conditions de marge différentes selon les clientèles visées pourrait être source d'inconfort pour les principaux clients-actionnaires. Ce qui compte est l'usage ultime des données.

2.2 DE LA PRIVACY AU SEIN DES PLATEFORMES COMME BIEN PUBLIC

Ce type de plateformes offre des services à destination principale d'autres entreprises. Elles se nourrissent toutefois de données produites par des capteurs qui dressent en creux un portrait du comportement d'un usager final (mobilité, santé). En l'occurrence, les données reçues par HERE Technologies sont pseudonymisées par les constructeurs automobiles. HERE héberge donc des informations anonymisées mais détaillées concernant des véhicules. Bien que HERE se conforme strictement aux règles européennes en matière de données personnelles (RGPD), la distinction pratique et légale entre données industrielles (anonymisées) et données personnelles s'avère parfois difficile à tenir, notamment dans des zones rurales ou à très faible densité.

17- Cf. la directive 2019/1024 du Parlement européen et du Conseil concernant les données ouvertes et la réutilisation des informations du secteur public (refonte), <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32019L1024&from=EN>

Des travaux de recherche en économie ont entrepris de modéliser les problématiques d'*e-privacy*^{18 19}. Le nouveau contexte de plateformes numériques se caractérise par un usage intense de réseaux sociaux et autres plateformes d'échanges sur internet. Dans un tel contexte, la confidentialité peut être considérée comme un « bien public ». Il est démontré qu'il suffit qu'un très faible pourcentage des participants à un réseau renonce à leur *privacy* (en échange de l'usage des services du réseau par exemple) pour qu'il soit possible de reconstituer la totalité des informations de tous les participants. Y compris de ceux qui n'ont pas consenti, voire les informations d'autres personnes qui ne sont même pas sur ce réseau-ci. Le consentement individuel ne suffisant pas, seule une solution politique, réglementaire, forte est susceptible d'apporter des éléments de réponse. L'approche réglementaire puissante et généralisée s'appuie sur les moyens et méthodes techniques qui permettent le respect de l'anonymat. Le règlement général (européen) de protection des données personnelles (RGPD) est perçu comme un premier pas encourageant en matière d'e-libertés.

2.3

USAGES À DES FINS INDUSTRIELLES DES DONNÉES ET ALGORITHMES

Une insuffisante maîtrise des concepts statistiques sous-jacents à certaines applications des plateformes interindustrielles est susceptible d'avoir des effets dommageables. Les velléités de modélisation de comportements de systèmes complexes via des algorithmes d'apprentissage peuvent aussi se heurter à une insuffisante prise en compte a priori des principes physiques fondamentaux régissant ces systèmes. Pour obtenir des prévisions d'une qualité industrielle irréprochable, il est vital de qualifier les relations, sens et puissance, et d'être capable de pondérer l'importance des différents attributs. Ces tâches de préparation des données, qui nécessitent une connaissance fine des fonctionnements sous-jacents, sont donc cruciales et coûteuses. Mais d'elles dépendent la fiabilité et la robustesse des résultats du modèle. L'échec de d'IBM Watson Health procède principalement d'une mauvaise annotation des données. Laquelle requiert une expertise équivalente à celle

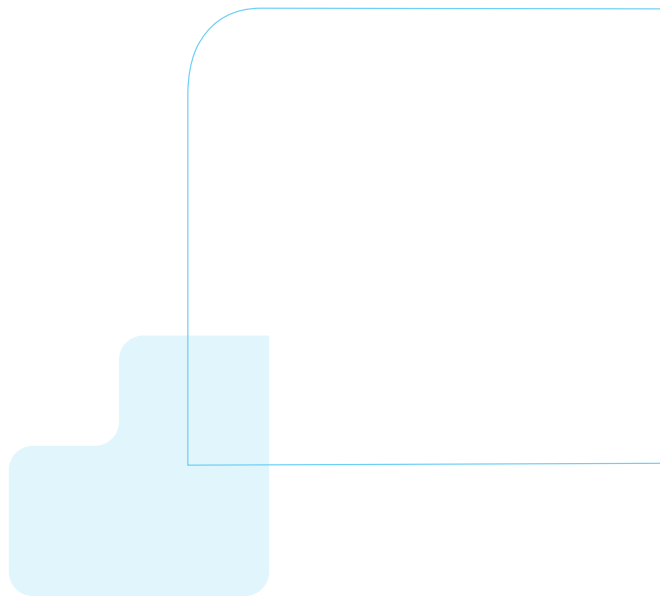
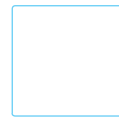
des experts capables de se passer de l'outil d'aide à la prédiction, comme des cardiologues ou des neurologues expérimentés, par exemple.

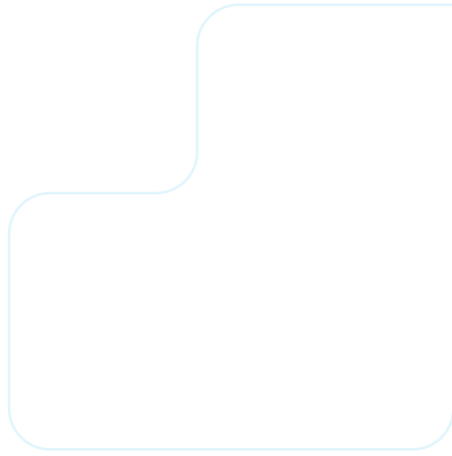
Pour que les données et donc les résultats des traitements statistiques à partir de ces données, c.-à-d. les prévisions, atteignent une qualité industrielle, le recours à des compétences rares d'un haut niveau d'expertise s'impose.

En contraste avec l'image d'automatisme et d'immédiateté de services que l'on prête à ces algorithmes pour des usages à la Facebook ou Amazon. Comme l'illustre bien HERE, le prix des services fournis par les plateformes résulte d'investissements dans des technologies et dans des compétences scientifiques et techniques de haut niveau. Certes, les économies d'échelle font partie, avec les externalités de réseau, des caractéristiques des *business models* de plateformes numériques. Mais les coûts infrastructurels à engager peuvent demeurer élevés et parfois irrécouvrables.

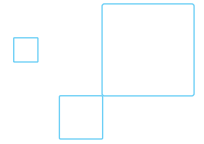
18- Fairfield, J. A. T., Engel, C., 2015, Privacy as a Public Good, Duke Law Journal, 65, 385-457.

19- Choi, J.P., Jeon, D.-S., Kim, B.-C., 2016, Privacy and Personal Data Collection with Information Externalities, Toulouse School of Economics, Preliminary version, October 7.





3 ÉNERGIE



3.1

L'UBÉRISATION VUE COMME « LIGNE ROUGE STRATÉGIQUE »

En tant qu'énergéticien²⁰, il est possible de considérer qu'il y aurait deux types de plateformes numériques : B2B, elles seraient de nature technologique ; B2C, de nature économique. S'il est éventuellement possible de faire évoluer certaines fonctionnalités d'une plateforme B2C vers du B2B, l'inverse n'est pas envisageable. Adopter un tel point de vue et le faire savoir offre la possibilité de relations économiques basées sur la confiance au sein des plateformes numériques interindustrielles.

Les risques encourus par cette évolution – notamment celui d'une perte radicale de confiance – expliqueraient l'échec relatif de tentatives de ce type. Imaginons qu'EDF considère telle plateforme numérique P comme indispensable au monitoring de certains paramètres de ses centrales nucléaires. La relation entre EDF et la plateforme P s'établit sur des bases technologiques. Aucun développement commercial n'a à être recherché à partir des données ou résultats des analyses proposées. Les entreprises de l'écosystème d'EDF demeurent donc préservées, de même que la qualité des interactions, sur la durée. Aucune initiative ne sera prise qui viserait à valoriser au profit d'EDF les données issues du fonctionnement des systèmes tiers à partir desquelles elle opère. Dans le cadre d'une relation de plateforme avec ses clients finaux, des services à valeur ajoutée sont susceptibles de résulter de la valorisation des données associées. Par exemple, dans le cadre d'un réseau de bornes de recharge pour véhicules électriques, EDF pourrait s'appuyer sur les données échangées entre le véhicule et la

borne pour optimiser le coût de la recharge, voire « rémunérer » le client pour fournir des services au réseau.

La propriété des informations présentes et utilisées peut et doit être garantie par des solutions techniques. Encore faut-il atteindre et parvenir à maîtriser cette couche sous-jacente à l'analyse des données, puis être capable de communiquer sur les méthodologies concernées. Des cas d'usage existent qui pourraient mériter d'être plus largement diffusés. En Europe, le règlement RGPD et la directive DSP2 (Deuxième directive sur les services de paiement bancaire), stipulent la propriété des informations au consommateur usager et l'obligation de demande du consentement éclairé. Elles favorisent des initiatives qui sont autant de cas d'usage intéressants. Du point de vue du droit de la concurrence, la compétition fonctionne grâce à la transparence des informations. Plus claire est la distribution initiale des droits de propriété, moins les frictions susceptibles de perturber le fonctionnement du marché se manifestent. **La mise à disposition d'API - interfaces de programmation applicative - joue un rôle clé dans la fluidité des échanges.** Les API fournissent, au sein des plateformes, les informations détaillées pour une certaine fonctionnalité. Ces « façades contractuelles » permettent à un logiciel tiers de mobiliser une fonctionnalité de la plateforme. Leur existence apporte de la clarté aux partages de propriété des données entre les différentes entreprises interagissant par le truchement de la plateforme.

Les énergéticiens du secteur du bâtiment, les 170 opérateurs de réseaux d'énergie de France, se sont regroupés au sein de l'Agence ORE (opérateurs Réseaux d'énergie). Cette dernière pousse à la numérisation des réseaux énergétiques,

20- Selon les travaux des prospectivistes d'EDF, cf. leur participation à notre groupe de travail le 15 novembre 2018.

à la gestion et à la valorisation des données selon une doctrine *open data*. Cette agence promeut et facilite la transition énergétique grâce au numérique. Elle développe trois axes de services : la visualisation des données de l'énergie, des bilans énergétiques et des outils pour les acteurs de marchés (comme la demande de certification de capacité aiguilleur, ou des mécanismes des compléments de rémunération). Les API proposées permettent la prise de contrôle par le propriétaire des données.

3.2

VERS UNE GESTION ÉLECTRIQUE OPTIMISÉE

EcoStruxure sous-tend l'architecture générale des plateformes IoT de Schneider Electric. Chacune opère un groupe de solutions techniques appliquées à des problèmes d'optimisation énergétique spécifiques dans les maisons, les bâtiments, les centres de données, les infrastructures et les entreprises industrielles. Ces solutions contiennent des dimensions de sécurité, fiabilité, efficacité opérationnelle, durabilité ou connectivité. EcoStruxure se décline en plusieurs versions au plus près des domaines d'expertise métier *alimentation électrique, informatique, bâtiment, équipement/machine, installation usine, et réseau*.

Ce que propose un département d'intelligence artificielle dans l'industrie

L'équipe *Analytics et Intelligence Artificielle* de **Schneider Electric** utilise toutes les techniques d'analyse de données et d'optimisation pour réaliser des applications utiles à ses clients. Sept familles de fonctionnalités sont distinguées :

- Evaluation de la performance et parangonnage
- Désagrégation de données et découverte d'informations
- Corrélation de données, prédiction, révélation de structures et modèles sous-jacents
- Monitoring et diagnostic d'équipements et de systèmes, maintenance prédictive
- Planification et ordonnancement des activités et de ressources (notamment énergétiques)
- Réglage de contrôles en fonction du contexte
- Aide à la décision par la simulation

Source : Présentation de C. Le Pape pour le groupe de travail de FutuRIS, 21.02.19

Une classe spécifique de services rendus par la plateforme EcoStruxure, « les conseillers » ou *advisors*, permet d'illustrer plusieurs des traits cruciaux des plateformes numériques pour l'industrie au regard des données. *Ecostruxure Asset Advisor* et *Ecostruxure Microgrid Advisor* démontrent l'intérêt d'exploiter la complémentarité de données multiples, de différentes sources, et d'approches conceptuelles variées en matière d'informatique prédictive. Des données météorologiques et tarifaires sont ainsi croisées avec celles directement issues des systèmes électriques et de leurs usages. Des modèles multiéchelles et multiphysiques sont combinés avec des techniques d'analyse de données et d'apprentissage statistique, lors des différents moments de résolution de problèmes : phase d'identification du modèle sous-jacent, élaboration du modèle de référence au plus proche des données collectées, tests de validation du modèle, puis afin d'apprécier les éventuels écarts avec les normes de comportement attendus du système en fonctionnement, et enfin déclencher l'alerte ou la commande correctrice adaptée.

Les microréseaux électriques intelligents ou *microgrids*²¹ sont des réseaux électriques de petite taille, conçus pour fournir un approvisionnement électrique fiable et de meilleure qualité à un petit nombre de consommateurs. L'étude de leur exemple favorise la mise en lumière de la diversité des données en jeu ainsi que leurs multiples usages, dont la valeur dépend des contextes d'utilisation. L'accompagnement offert par un petit nombre de sociétés, comme Schneider Electric, démontre comment ces données peuvent être combinées pour améliorer la prise de décision. Et ce, aussi bien en amont de l'investissement initial que lors de l'installation, de la maintenance et du contrôle au jour le jour.

Une première étape consiste habituellement en l'analyse technicoéconomique préalable à la mise en place de la *microgrid*. Il peut par exemple s'agir d'un réseau de panneaux solaires et de moyens de stockage d'énergie sur un site. *EcoStruxure Microgrid Advisor* prévoit les évolutions de la production et de la charge du site. Ces évolutions dépendent des conditions météorologiques et de la bonne utilisation des batteries. Maîtriser ces évolutions permet d'établir des propositions afin d'alléger les futures factures d'électricité. La durée de vie des batteries dépend du nombre de cycles de charge

et de décharge. Les contrats d'achat ou de vente d'électricité varient aussi grandement. Le prix de l'électricité et les revenus tirés de la revente ou de la contribution à divers services rendus au réseau (régulation de fréquences, par exemple) affectent la production et la consommation électriques. Les données caractéristiques du fonctionnement du site en termes de production et de consommation ainsi que les facteurs qui influencent ce fonctionnement sont donc mobilisés. Renseignant sur l'activité du site, ces informations sont généralement considérées comme extrêmement confidentielles. Ces mêmes données ont, par ailleurs, d'autres usages comme l'évaluation de la performance énergétique du site, l'amélioration du contrôle des systèmes de chauffage, de ventilation et de conditionnement d'air ou la détection d'anomalies.

La tarification électrique, dont les éventuels tarifs de rachat d'énergie d'origine éolienne ou solaire, varie structurellement d'une grande région du monde à l'autre. Dans les cas simples, l'optimisation de la facture électrique dépend souvent de deux composantes aux États-Unis et d'une seule en France (la première citée ci-après) : (i) le coût total de la consommation sur des périodes de temps données (heures pleines, heures creuses), soit le prix unitaire de la période multiplié par l'énergie utilisée au cours de la période (en kWh) et (ii) le prix de la plus haute demande ou « demande d'appel » (*demand charge*) sur une période donnée, par exemple le mois. Pour l'opérateur, la réduction de la *demand charge* représente une contrainte majeure, compliquée et onéreuse à garantir. La demande d'appel constitue un paramètre clé dans l'optimisation de la conception puis du pilotage du microréseau électrique. La valeur apportée par *Microgrid Advisor* à ses clients résulte donc de la meilleure estimation du montant des frais de la demande d'appel, compte tenu du bilan économique coûts-avantages des panneaux solaires et des batteries.

Enfin, on voit que les conditions de développement économique de l'informatique prédictive sont dépendantes de choix de politiques publiques en matière de normes et de règlements. Ces derniers ayant des implications majeures en termes technologiques et industriels. Une

plateforme pourra offrir un service ou pas selon les modalités de tarification de l'électricité de la zone économique concernée. En l'occurrence, la présence de la composante *demand charge* dans la tarification électrique aux États-Unis complique la réalisation mais accélère la rentabilisation de la microgrid. Et facilite donc la vente des services rendus par *Microgrid Advisor*.

Le thermostat connecté *Wiser Air*, proposé seulement au Canada et aux États-Unis, présente un autre versant des méthodes d'optimisation. *Wiser Air* apprend grâce aux interventions des usagers du logement qui, en indiquant s'ils ont trop chaud ou trop froid, permettent au système de calibrer par apprentissage la température désirée selon le contexte. La bonne température, celle qui procure le plus grand confort aux habitants, résulte de leurs interactions avec le système. Une fois le paramétrage dynamique établi grâce aux réglages humains, *Wiser Air* effectue des tests d'ajustement plus fins afin de n'utiliser que la bonne quantité d'énergie nécessaire au confort désiré. A ces tests répondent d'éventuelles corrections des humains, pour que s'établisse progressivement la convergence vers un optimum confort/prix de la consommation énergétique.

3.3

CHAÎNES DE VALEUR DES DONNÉES, TROIS ENJEUX INDUSTRIELS CLÉS

Trois enjeux industriels majeurs peuvent être mis en exergue à partir de la pratique d'analyse de données de systèmes complexes. Tout commence lorsqu'un système électrique renvoie un signal d'alerte, y compris une prédiction de panne. Des questions d'interprétation, de compréhension, voire d'intervention-réparation se posent alors. Et les enjeux soulevés peuvent avoir des répercussions organisationnelles et économiques importantes.

3.3.1 Comprendre l'alerte d'un système en fonctionnement

Pour commencer, il s'agit d'adopter le meilleur modèle explicatif du problème rencontré. Ce modèle reste le plus souvent multiphysique (combinaison par exemple de thermique, de

chimique et d'électromagnétique). Il faut donc disposer d'une bibliothèque de modèles à jour pour tous systèmes, sous-systèmes et composants opérés. Ensuite, les analyses et tests doivent s'effectuer à partir d'une source de données de haute qualité garantie. Issues du fonctionnement des systèmes en opération, les données nécessitent d'être captées selon un procédé contrôlé et d'être sélectionnées en quantité suffisante pour assurer leur représentativité. S'ensuit alors un exercice systématique de parangonnage des algorithmes, afin de choisir celui qui fournit l'estimation la plus convaincante. D'autant que l'information sur la mesure de la fiabilité des estimations proposées constitue une caractéristique des algorithmes testés. Les tests de robustesse des algorithmes se développent. Toutefois, à ce jour, le parangonnage temps réel d'algorithmes à partir des jumeaux numériques n'est que rarement implémenté sur des systèmes complexes.

Les modèles théoriques à jour sont alimentés et complétés par les standards et lignes directrices du domaine. Pour notre exemple, aux États-Unis, ces informations indispensables sont disponibles auprès de l'American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE). Et ce, autant pour le choix des indicateurs et les valeurs des performances attendues et légales que pour l'interprétation. Ces étapes de traitement et d'analyse des données obligent à des manipulations préparatoires compliquées et coûteuses.

3.3.2 Passer d'un modèle appris à des données « de vraie vie » (hors de l'espace d'apprentissage d'origine)

La performance des modélisations fondées sur des algorithmes d'apprentissage artificiel procède du jeu de données à partir duquel un modèle a été « appris ». L'utilisation sur de nouvelles données, celles de la vraie vie du système, ne va pas de soi. En effet, l'emploi du modèle appris présuppose une très forte ressemblance qualitative et structurelle entre l'ensemble des données d'apprentissage et celui des données d'application.

Cette très grande proximité demeure un construit. Elle nécessite des tâches de nettoyage,

de recalibrage, de marquage et d'indexation de même qualité pour les deux jeux de données. Il faut aussi que le jeu de données d'apprentissage soit représentatif des situations rencontrées dans le monde réel. Ceci doit être établi et vérifié avec l'homme de l'art. Dans le cas d'une tâche de classification, par exemple des causes des anomalies détectées sur un appareil ou un système, les données d'apprentissage doivent être labellisées avec les causes des anomalies constatées dans le passé et les labels vérifiés avec des experts du métier. Ce recours aux experts métier constitue un goulet d'étranglement au développement de modèles spécialisés par apprentissage artificiel. Les compétences nécessaires à la validation et à la labellisation des données sont précieuses. Les personnes qui les détiennent sont des atouts pour les organisations qui les emploient, de la conception des nouveaux produits au support de clients stratégiques et à la compréhension des besoins émergents de ces clients. Mobiliser ces experts de manière significative et répétitive pour préparer le travail de la machine, voire pour le vérifier, peut sembler peu flatteur et moins directement rentable que les autres contributions de ces mêmes experts.

Deux techniques sont souvent mises en avant pour permettre une adaptation ou une généralisation d'un modèle appris à un nouveau contexte. Dans la première, la simulation d'un jumeau numérique (*digital twin*) permet de générer de nouvelles données, lesquelles peuvent être utilisées pour apprendre un nouveau modèle. Dans la seconde, l'apprentissage par transfert (*transfer learning*) permet de transformer un modèle précédemment appris dans un contexte donné en un modèle applicable à un nouveau contexte et donc à des données situées hors de l'espace d'apprentissage initial. L'utilisation de telles techniques ne fait cependant pas disparaître le besoin de validation. Dans le cas de la simulation, le besoin est en réalité déplacé, puisque le modèle utilisé pour simuler le système réel doit lui-même être validé.

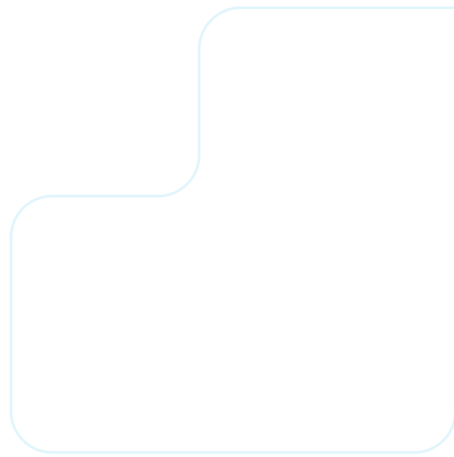
Les délimitations de l'espace d'apprentissage et de l'espace d'application sont donc des questions autant scientifiques et techniques qu'économiques. Les données raffinées – convenablement nettoyées, organisées,

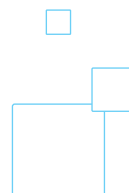
labellisées et étendues au fil du temps pour tenir compte de changements de contexte – ont acquis une valeur que les données brutes initiales n'avaient pas.

3.3.3 Améliorer les systèmes en fonctionnement, les faire évoluer grâce aux données

Les entreprises qui fournissent des services par l'intermédiaire d'une plateforme numérique opérant plusieurs produits (un transformateur, un chauffe-eau, un moteur, un automatisme) cherchent à les améliorer grâce à l'accès aux données de fonctionnement. Il existe un lien à entretenir entre la conception et les opérations. Ces boucles de rétroaction sont les canaux par lesquels les principaux facteurs d'influence sur la performance du système considéré s'expriment. La véracité de l'alerte renvoyée par le système doit être absolument validée. Il faut s'assurer d'une compréhension réaliste des mécanismes et ne pas se contenter du constat d'un certain nombre de corrélations convergentes.

L'affinage des modèles s'avère nécessaire lorsque le système suivi évolue. Les méthodes utilisées peuvent consister à introduire de manière systématique et graduée des perturbations dans le modèle appris, à faire usage de simulations, voire de jumeaux numériques si le niveau de complexité du système l'autorise. Si les tests indiquent que la qualité des prévisions a tendance à se dégrader au fil du temps, il se peut alors que les bons facteurs explicatifs n'aient pas été pris en compte... ou que les conditions aient effectivement évolué. Des tests statistiques peuvent alors être effectués pour vérifier le niveau d'anomalie. Classiquement, dans le cas où la prédiction se manifeste par une valeur numérique, on peut suivre la façon dont évolue la somme cumulative des valeurs d'erreurs successives ou vérifier que les valeurs de la série restent bien dans l'intervalle centré autour de la moyenne entre 2 et 3 écarts-types. Dans le cas contraire, il convient d'expliquer l'anomalie : insuffisance d'information, évolutions, dysfonctionnement des capteurs, mauvaise prédictions intermédiaires, évènement exceptionnel ? Avant d'améliorer, voire de corriger un modèle au regard d'alertes, un certain nombre de précautions doivent être prises.





4 RETOUR SUR LA VALEUR ET LES PRIX DES DONNÉES : — ENTRE STRATÉGIE D'ENTREPRISE ET POLITIQUE INDUSTRIELLE

N'ayant pas vocation à clore une discussion dont ce document ne vise qu'à rendre compte, cette dernière partie insiste sur les trois aspects centraux de l'analyse des plateformes numériques interentreprises : l'élaboration d'une compréhension commune de la chaîne de valeur des données ; la condition associée du développement et de la maîtrise de l'économie de la confiance ; l'exploration d'une *platformonomics* visant à contenir les comportements de traitements différenciés.

4.1

POUR UNE COMPRÉHENSION COMMUNE DE LA CHAÎNE DE VALEUR DES DONNÉES, MÉTIER PAR MÉTIER

Au service du monde industriel, l'informatique prédictive procède de la combinaison de méthodes de modélisation variées, lesquelles pour être performantes ne peuvent s'abstraire des expertises métier. Ce sont les efforts de validation, certification, vérification, homologation des données et des modèles explicatifs dans lesquels investissent les industriels du numérique qui font la qualité des prédictions proposées. La valeur des données industrielles, tout au long de leur chaîne, résulte de ces traitements experts, le plus souvent de très haut niveau. De tels investissements sont intégrés au prix d'usage des plateformes numériques. Les standards ouverts qui favorisent l'interopérabilité sont à l'origine des gains économiques extraits des appropriations privées permises par les droits de propriété intellectuelle afférents à un autre maillon de la chaîne de valeur des données. Les efforts de compréhension commune, maillon

métier par maillon métier, principalement à l'initiative des intégrateurs-systémiers impliqués et en coopération avec les entreprises du numérique constituent un préalable.

4.1.1 A environnement cyber-risqué, nouvelle prise de conscience

La cybersécurité a considérablement évolué au cours des cinq dernières années. De physiques et coûteuses, les attaques sont devenues logiques et économiquement profitables. Ceci du fait de la déferlante de dispositifs connectés, intégrés à la majorité des objets du quotidien. Pour les malfaiteurs, réussir en fonction d'un investissement donné à prendre le contrôle d'un objet, c'est s'assurer un rendement économique immense à l'échelle de sa diffusion. Le coût moyen des attaques a considérablement augmenté. Deux millions d'euros peuvent aujourd'hui s'avérer tout à fait rentable. Par exemple, tous les terminaux de paiement d'un certain type, toutes les voitures de tel modèle voire de toutes les marques d'une certaine génération, toutes les centrales électriques de tel type etc. L'année 2015 a été marquée par des attaques remarquables de grande ampleur, et par une prise de conscience de la vulnérabilité de nombreux systèmes. Ainsi, aux États-Unis, le Bureau d'enquêtes fédérales (FBI) en collaboration avec le ministère des transports (DoT) et l'Administration nationale pour la sécurité du trafic sur les autoroutes des États-Unis (National Highway Traffic Safety Administration, NHTSA) a publié l'alerte selon laquelle « les véhicules automobiles deviennent de plus en plus vulnérables aux exploitations à distance »²². Le problème ne fait qu'empirer puisque le nombre d'objets connectés s'accroît

vite, la pertinence d'usage desdits objets dans le quotidien et donc la valeur économique concernée aussi. La vulnérabilité et l'incitation aux attaques augmentent. L'intérêt économique des cyberattaques est plus grand et pourtant, pour le moment, elles sont surtout le fait de certains États.

Face à ce nouvel environnement cyber-risqué, les spécialistes se mobilisent. En témoigne la montée en puissance des conférences des professionnels de la cybersécurité comme Black Hat qui, à l'origine, avait une conférence annuelle aux États-Unis, et maintenant plusieurs par an, partout dans le monde. S'y partage l'état de l'art des attaques et des solutions proposées pour y faire face. Le National Institute of Standard and Technology américain (NIST) a publié en octobre 2018 le livre blanc²³ qui répertorie les problèmes de sécurité de l'internet des objets (IoT). Au fondement de la typologie des alertes proposées, le constat circonstancié que la diffusion de l'IoT impose un regard nouveau en matière de sécurité. En effet, comparant avec le fonctionnement des périphériques informatiques classiques, le NIST souligne que de nombreux périphériques IoT:

- interagissent différemment avec le monde physique ;
- ne peuvent pas être atteints, gérés ou surveillés de la même manière ;
- ont une efficacité, une efficacité et une disponibilité des fonctionnalités de cybersécurité et de confidentialité souvent différentes.

Insistons sur les conséquences de ce dernier aspect. La capacité à établir des mesures fiables de la qualité du fonctionnement des IoT devient l'un des enjeux majeurs de ces nouvelles menaces informatiques. Avec la mesure peut venir la confiance. A contrario, à défaut de mesures et d'indicateurs de performance des IoT, la confiance devient un sujet de préoccupation. Les métriques et les mesures sont les fondements de la confiance. L'IoT relève de technologies relativement jeunes. À ce jour, il existe peu de moyens de mesurer les systèmes IoT autrement

qu'en comptant ou par des tests dynamiques spécifiques. Ce manque de lisibilité objectivée constitue une gêne à la diffusion des IoT les plus fiables. Il est compliqué de faire valoir qu'un système est fiable ou même d'estimer le nombre de tests qu'il devrait recevoir.

4.1.2 La sécurité *by design*

Ces objets IoT, qu'ils soient des capteurs, des agrégateurs ou de e-utilities, sont souvent difficiles à tester et à certifier y compris par leur producteurs du fait des interactions avec leur contexte précis d'usage. Un certain degré d'incertitude demeure quant à leur perméabilité. Ils sont donc attaqués, y compris sans que l'utilisateur puisse en prendre conscience. Il n'existe qu'une seule manière pour repérer et donc protéger l'IoT de toute attaque : les méthodes formelles. Elles sont l'équivalent informatique de la démonstration logique mathématique. Les solutions commerciales proposées - des systèmes d'exploitation dédiés - fournissent la preuve automatique des théorèmes sous-jacents au système vérifié (voir celles de Prove&Run²⁴).

Le coût à l'unité de l'adjonction d'une telle sécurité prouvée sur n'importe quel IoT est très faible. Pour des raisons multiples, dont une insuffisante prise de conscience des risques, notamment réputationnels, ces solutions *by design* ne sont pas encore généralisées. Le secteur le plus propice est celui de la défense et de la sécurité nationale. Des applications spécifiques porteuses concernent actuellement la téléphonie mobile, ou encore les sites de vidéo en streaming (passage à un 4K secure), les systèmes utilisant des caméras pour des échanges sécurisés etc.

4.2

ASSOCIER VALEUR DES DONNÉES ET VALEUR DE LA PROTECTION. VERS L'INDUSTRIE DE LA CONFIANCE

La valeur des données est corrélée à celle de la protection dont elles bénéficient. Le lien va dans les deux sens. Des données de haute valeur appellent un niveau de sécurité élevé. Le fait de garantir leur sécurité en élève la valeur en

23- <https://csrc.nist.gov/CSRC/media/Publications/white-paper/2018/10/17/iot-trust-concerns/draft/documents/iot-trust-concerns-draft.pdf>

24- Cf. <https://www.provenrun.com/solutions/>. Les solutions de sécurité informatique de Prove&Run sont fondées sur cette approche. Elles atteignent le niveau d'assurance d'évaluation le plus élevé (EAL7 ou ITSEC E6). Ces systèmes d'exploitation sécurisent n'importe quel microprocesseur, basé sur toute architecture. Les revenus de Prove&Run résultent des licences sur ses OS. Il existe toutefois plusieurs façons de concevoir les microprocesseurs, et elles ne sont pas neutres quant à l'établissement de la preuve formelle. L'idéal étant des puces « sûres by design ».

retour. Les choix des opérateurs et propriétaires – qui peuvent être conditionnés par des aspects techniques – en matière de production, de stockage et/ou de traitement dans le cloud, de manière hybride en *edge-computing* impliquent des contraintes de sécurité particulières. Tous sont identiquement passibles d'une approche de sécurité *by design*. Quel que soit le domaine, la question de la localisation (géographique) des données est seconde au regard des enjeux de sécurisation.

La souveraineté nationale se matérialise ainsi d'abord par une ossature infrastructurelle simple et auditable. Ces deux qualités forment la base de confiance que représente la souveraineté. En termes industriels et commerciaux, l'enjeu majeur concerne bien celui de la réputation. Elle est la confiance accordée et renouvelée. Un accroc à l'accord et la réputation est minée. A ce titre, la revendication claire et simple de l'usage de la meilleure technologie sous-jacente aux services et produits vendus peut constituer un gage essentiel.

Du point de vue des entreprises, ce qui fait d'un industriel ce qu'il est, c'est (entre autres) sa capacité à apprécier avec justesse les risques potentiels encourus par ses process. Cette aptitude, ancrée dans l'expérience et traduite en considérations de contrôle qualité, conduit à la mesure de degrés de gravité. Lesquels sont à associés à des niveaux de sécurité à établir. La puissance des protections à mettre en œuvre, et donc des investissements humains et financiers à engager, dépend, dans les organisations, du partage de données techniques dans l'ensemble des strates décisionnaires. Il peut toutefois arriver que des éléments pertinents de décision peinent à atteindre les niveaux de décision pertinents.

Les deux facteurs cruciaux de la confiance informatique sont l'auditabilité et l'immutabilité. Cela vaut pour les systèmes et pour les données. La diffusion la plus large des méthodes formelles à même de garantir effectivement auditabilité et immutabilité est susceptible de favoriser l'émergence d'une nouvelle industrie : celle de la confiance. Cette promesse forte d'une

garantie formelle de l'exécution d'une fonction exactement comme attendu est susceptible de permettre le partage équitable de la valeur créée. Enfin, le cadre réglementaire fait partie de l'ossature infrastructurelle propice à une *accountability*. Le RGPD invite à propager la confiance que les individus peuvent accorder au système de collecte et de traitements mis en œuvre par les responsables des traitements. Charge à ces derniers de démontrer, en pratique, leur respect des règles, en toute transparence : preuves de mise en œuvre de la sécurité et du respect des droits d'accès ou de modification par exemple.

4.3

POUR UNE PLATFORMONOMICS : LA QUESTION DES TRAITEMENTS DIFFÉRENCIÉS²⁵

A l'instar de « *Competition policy for the digital era* »²⁶ de J. CREMER et al. (2019), les apports de l'économie à la compréhension des conséquences de la plateformisation numérique s'inscrivent très largement en appui à la doctrine classique du droit de la concurrence. Si la notion de marché biface s'avère un efficace plaidoyer pour un examen approfondi au cas par cas des conditions d'établissement du prix, le censeur théorique final est le consommateur. Il ne doit jamais pâtir d'une distorsion éventuelle de la concurrence du fait de la position avantageuse dont pourrait bénéficier une entreprise.

Parmi les problèmes d'organisation industrielle qu'aborde la théorie économique en présence de plateformes numériques, il y a celui des traitements différenciés. Un site de commerce en ligne n'est censé privilégier aucun de ses fournisseurs ni aucun des produits mis en vente. Il joue son rôle de place de marché en toute neutralité. Les plateformes de vente en ligne sont théoriquement poussées vers l'adoption d'un tel comportement neutre par la concurrence qui s'exerce entre elles. Mais la concurrence incite chacune à se démarquer des autres, de manière plus ou moins radicale, et ainsi échapper à une comparaison facile sur

25- Les paragraphes qui suivent ne font que s'inspirer modestement de la présentation de Doh-Shin JEON (Toulouse School of Economics) le 19 avril, sans toutefois prétendre ni à la profondeur de raisonnement proposé, ni au haut niveau de technicité du sujet. Il s'agit d'une extrapolation en relation avec le sujet de ce rapport.

26- <http://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf>

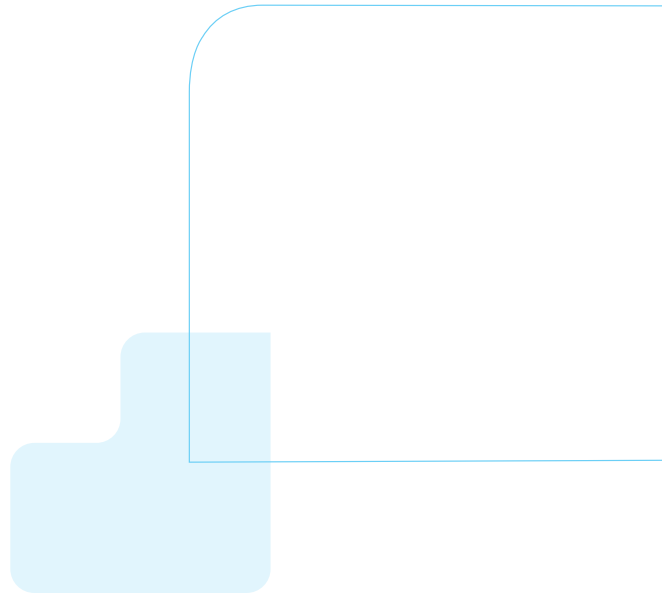
le prix des produits vendus. Elles ont tendance à éviter la confrontation directe. Par exemple, en s'arrangeant pour ne pas être compatibles entre elles aux yeux des fournisseurs et producteurs dont elles vendent les produits. Ou bien en utilisant des algorithmes d'aide à la décision des clients propres à encourager certains choix plutôt que d'autres, fondés sur des critères autres que ceux choisis par le client. L'identification d'un possible traitement différencié se révèle en outre à l'analyste d'une redoutable complexité du fait de la tendance des plateformes dominantes à devenir « plateformes de plateformes » et à développer des « marques distributeur ». Lesquelles fournissent, à partir des données de millions d'actes d'achats, des produits similaires aux produits des marques beaucoup vendues, à des prix compétitifs et à qualité similaire. Sans forcément signaler qu'il s'agit de produits de « marque distributeur ».

Du point de vue économique, de telles opérations accroîtraient de manière indue le profit des plateformes et réduiraient le bien-être des consommateurs. Les règles du jeu concurrentiel que s'attachent à garantir les autorités de la concurrence partout dans le monde sont donc mises en brèche par les comportements des plateformes. Pour garantir des pratiques justes et équitables, le législateur tend à s'attacher les services des meilleurs techniciens de l'algorithmie, de l'analyse de données en masse, de l'architecture des plateformes, etc.

Ces problématiques de traitements différenciés concernent d'abord les plateformes de vente en ligne, et donc les relations entre consommateurs et entreprises. Mais elles débouchent naturellement sur des questions techniques de comportements entre entreprises au sein de la chaîne de valeur. L'accès et le contrôle des données mixtes sont au cœur des enjeux. Dans le cas des plateformes numériques qui connectent des entreprises, les pratiques ne sont pas encore stabilisées et sont donc mal connues. Par conséquent, la théorisation économique et le droit de la concurrence restent à la traîne.

La proposition centrale de ce rapport, à l'initiative de mise en œuvre des intégrateurs-systémiers eux-mêmes lors de leur plateformes,

concerne l'élaboration d'une compréhension commune, par sous-secteur, de la chaîne de valeur des données. Notre hypothèse forte (cf. *Pour une politique industrielle du numérique*) étant que les écosystèmes sont les lieux par excellence de cette élaboration. Ces écosystèmes métier autour de l'informatique prédictive sont porteurs d'un espoir que des mesures de politique industrielle adaptées pourraient faire advenir.





anRT

ASSOCIATION NATIONALE
RECHERCHE TECHNOLOGIE

33 rue Rennequin
75017 - PARIS
Tél : +33(0)1 55 35 25 50
com@anrt.asso.fr
www.anrt.asso.fr