SYNTHESIS REPORT

Data prices and value in digital platformization – Key pointers for business-to-business relations

ANRT
ASSOCIATION NATIONALE RECHERCHE TECHNOLOGIE

FUTURIS

THE POWER OF COLLECTIVE INTELLIGENCE

OCTOBER / 2019
LES CAHIERS FUTURIS

For an industrial digital policy Working group
Chaired by Gérard Roucairol, Académie des technologies
Pierre Bitard, ANRT

This work is supported financially by FutuRIS members:

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

List of high-level experts that participated in the working group

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 and OFI
Claude LE PAPE	SCHNEIDER ELECTRIC
Luca MOLLO	PFIZER
Jean-Manuel QUIROGA	MATHWORKS
Gérard ROUCAIROL	ACADÉMIE DES TECHNOLOGIES, chairman of the working group
Hubert TARDIEU	ATOS
Ascension VIZINHO-COUTRY	MATHWORKS
Patrick WAELBROECK	TELECOM PARISTECH
Jonathan WEILL	EDF then EDVANCE

This report extends and completes the white paper of March 2018, *Pour une politique industrielle du numérique* (for an industrial digital policy). It is the result of collective intelligence and proposes a state of the art on the question of price and value in B2B digital platformization. The substance of this document comes mostly from the minutes of the working group featuring high-level experts, whom we warmly thank.

SOMMAIRE

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 plaformization 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.

GENERAL INTRODUCTION

Digital platforms frequently make newspaper headlines, mostly centred on GAFAM (Google-Amazon-Facebook-Apple-Microsoft). These articles either praise the platforms for simplifying our everyday lives, or criticize their hegemony, following in the steps of US and European competition authorities. Artificial intelligence tends to be hailed as an exclusive solution to a whole range of problems. Lastly, data is described everywhere as "the new black gold", as if the only limit to getting rich from data is accessing their flows. Moreover, these three subjects are frequently treated together, without shedding much light on the debate. The domination of GAFAM leaves little room for French or European global successes on a similar scale in this B2C (business-to-consumer) niche – or in any case, not in the near future.

Yet could taking a different view of digital platformization, the place occupied by artificial intelligence, and the role of data offer well-grounded hope? And what if a growing groundswell were in the process of shaking up French and European industry, possibly to our advantage?

Digital platformization concerning business-to-business (B2B) relations constitutes a crucial economic opportunity that needs to be grasped in France and Europe. This ongoing process is about to significantly impact companies and our economy. Public and private approaches that contribute to organizing data value chains closer to the relevant businesses and sectors should therefore be encouraged. This was the spirit behind the recommendations of the 2018 white paper Pour une politique industrielle du numérique (for an industrial digital policy). Extending this analysis, in the present collective intelligence work we explore the conditions in which business ecosystems are capable of working together to create value. To do this, we examine three of the most advanced digital platformization sectors, i.e. health, the automotive industry, and energy.

The technology ecosystem is a good starting point for defining business behaviours. This approach fits in with the dynamic analysis of digital infrastructures that featured in the white paper. By describing how contemporary infrastructures operate in France – ranging from microprocessors to sensors, computers and networks, including the internet and its programmability – we can deduce a transformation scenario. In this scenario, the value created from data comes from their changing knowledge cycle, which moves from the real world to the enriched content proposed by new specialist intermediaries, thanks to widespread connectivity.

How, in this B2B digital plaformization movement, do data acquire value? How can we establish a common understanding of their value chains that fosters value-sharing within ecosystems, thanks to fair prices? On what infrastructure and regulation conditions? In describing sector-specific challenges, we are able to elucidate action through pointers established by practice. It is up to decision-makers in our businesses and within public organisations, and public policies, to grasp the opportunity offered by B2B digital platformization. The following analysis looks closely at these challenges with the object of setting out as transparently and comprehensively as possible the state of the art of price and value in B2B digital platformization.

.

THE PARADOXICAL IMPACTS OF DIGITAL PLATFORMIZATION FOR INDUSTRIALS

It is worth approaching digital platformization from two complementary angles that correspond to the two major types of business concerned.

Firstly, system integrators

All car and plane manufacturers, makers of tractors and agricultural machines, pharmaceutical companies producing drugs or vaccines, medical equipment manufacturers, and energy producers and suppliers employ resources to grasp the digital opportunity and strengthen their competitive position. With digital platformization, they all intend moving from a position of competitor to that of monopolist, on their main and connected markets, upstream and downstream. This position gives companies greater latitude regarding their clientele in terms of prices, quantity and timeframes.

Car manufacturers want to increase their share of the added value once a vehicle has been sold by providing usage services. Drugs companies look to provide compliance-monitoring devices or help future patients participate in making a better diagnosis of their condition. Electricity suppliers might want to encourage certain uses of the electricity they deliver, in line with their own techno-economic constraints. This implies that they become the owners of the data generated from the use of their products. This ownership requires investing in a computer infrastructure and new skills. It promotes the expansion of their market power.

But to what extent does installing sensors on a tractor make the tractor manufacturer the owner of the information gathered by the tractor's movement round a particular field for a particular usage? Does the tractor collect geolocalized information on soil quality or fuel consumption – which have multiple potential uses – or does it simply allow the farmer to fertilize his crops? Do aeroplane engine manufacturers provide aircraft constructors with flight time, power, thrust and consumption performance, or information on the parameters resulting from the flights of equipped planes? And do infant formula producers supply nutrients adapted to babies' needs depending on their age, and only that... with no relation to a free application to monitor baby growth? These questions have an obvious answer: the data here seem to be intimately connected to the product through the customer relationship, and dissociating them proves difficult. Thus, clients are offered improved services from products thanks to the usage information collected. The tractor suggests an optimal harvesting circuit to the farmer, along with the best way to use the inputs dispensed by the tractor, etc. The performance of the use of a tractor is an increasing function of the number of users of the tractors. In fact, the tractor manufacturer benefits from a rich analysis base featuring all of the uses and usage data that it can mobilize and process to improve its models and predictions, and then send them out in the form of targeted services to its customers.

These system integrators are increasing their market power thanks to digital platformization: higher prices than those determined by the competition, increased lock-in capacity, higher switching costs, etc. Based on the standard yardstick used by competition authorities, the following question emerges: To what extent do customers benefit financially? Does the advantage of digital platformization not over-distort the relationship to the seller's advantage?

Secondly, we should look at the digital companies themselves

Focusing on how platformization impacts B2B relations involves anticipating the changes they face. There is no doubt that digital platformization reshapes the interactions between system integrators and the suppliers and sub-contractors thanks to which they manufacture or offer their services. However, complex product services based on components and sub-systems that all come from a system integrator's own production chains no longer exist. Digital platformization reinforces this phenomenon. Thus, when medical equipment manufacturers or aeroplane engine manufacturers integrate sensors and actuators into their products, it is unlikely that they will have designed and produced them themselves. Similarly, some of the processing and analysis of data collected in this way will in the short or longer term take place outside the integrating company. A recent car includes 100 to 150 sensors and features onboard computers, numerous embedded software programmes, and network access. The latter makes it possible to transfer, save and process data on dedicated servers. Only some of these components are designed and produced by the car manufacturer. Reciprocally, most of the new value that can be brought to the vehicle by the components, following data processing, involves a wealth of actors.

For system integrators, digital platformization therefore has paradoxical impacts on digital companies, irrespective of their sector (pharmaceuticals, cars, aeronautics, defence, education, electricity, etc.):

- They are more dependent on external components and skills.
- They also have more power, since digital firms need data on the use of the IoTs they have developed, if only to maintain the service quality of their component or set of components.

As a result of digital technologies, competitive pressures have grown: along with established impacts on consumers come new impacts affecting the structure of markets. The insertion into data value chains changes both companies' capacity to direct their price system downstream and their aptitude to influence their technological constraints upstream.

PLATFORMS FROM AN ECONOMIC ANGLE

Digital platforms that are well known to the public, like Google, Amazon, TripAdvisor, Airbnb and BlablaCar, tend to operate like "two-sided markets". Most often, one "side", which could be one of the markets that the platform sells to, meets the needs of other businesses (B2B); the other side, or the platform's other market, meets the needs of final consumers (B2C). Users of the same type benefit "on their side" from specific services that correspond to belonging to that network and whose usefulness depends on the number of its members. Simultaneously, although indirectly, the users of one side fuel the performance of the other side's network, thus attracting more users to it. These platforms allow intermediation, which is a source of value creation. Some are in fact specifically devised to establish the price of the goods exchanged on them, such as eBay and Datalogix, known as information (or data) brokers.

The economic models of the platforms mentioned above, which are usually classed as B2C platforms, have difficulties gaining a strong foothold among pure-play B2B platforms. In terms of B2B relations, data valorisation is derived from mastering technical principles specific to the problem at hand. They therefore vary according to the socioeconomic sector, and more precisely, according to the defined speciality: e.g. in health, anatomical pathology; in transport, cars; in energy, electricity, or possibly a source of intermittent energy. **On an industrial scale, mastering the "physics" (the underlying physical laws) of the domain is crucial, and embodied in specific business knowhow.** The service value procured results from the capacity of the system solution to correctly anticipate functioning or malfunctioning, i.e. occurrence of a breakdown, pathological risk, demand peak, etc. The datum (e.g. a measurement produced by a sensor) constitutes the source of the value created, and the value itself following the completed processing and analyses. Between the two, numerous processes will have turned the raw data into information that is directly useful to the precision of the processing.

A particular stage in the life cycle of an information block corresponds to the link of an overall value chain. At each stage of the cycle, value is created.

MULTIPLE DATA DIMENSIONS: INDUSTRIAL CONSEQUENCES

Confronted with the domination of GAFAM and BATX[1] , European and French system integrators seeking digital firms tend to look no further to form an alliance. Several of these alliances have recently been concluded by, e.g. Renault, Atos and Sanofi. Nevertheless, it is probably not too late to "play collectively" and develop inter-industry digital platforms in health, construction, and energy. The fierce competition that simultaneously acts on numerous links of the data value chain seems to encourage firms that are not yet platform businesses to buy services from the major established digital platforms.

However, coalitions of complementary businesses on the data value chain would make sense and drive a real economic advantage. No single actor can hold all of the information required for an isolated valorisation. The data

1- The initials of Google, Amazon, Facebook, Apple and Microsoft form the acronym GAFAM, while those of their Chinese equivalents Baidu, Alibaba, Tencent and Xiaomi form BATX.

value contained in a link of the chain is always the result of an exchange-based aggregation. Exchanges, sharing, cross-purchases and bartering are expressed in the two dimensions of data. We tend to immediately think of the volume or quantity of data, translated by the expression "big data". Nevertheless, data are also characterized by their number of attributes (and the impact that these attributes have on the phenomenon analysed). Each individual or data point is characterized by a set of qualities. If the objective is to improve the understanding of a particular disease, then the records of all of the people suffering from that disease at a particular point in time could be usefully enriched. So, what about features like their geographic location, the dynamics of that location over the previous year, their sports activities or lack of physical exercise, the frequency of their alcohol consumption, existence of other diseases and treatment received, etc.? The value resulting from a statistical analysis of the predictors of a common disease for the patients affected depends on the list of the attributes of those patients that can be analysed. **These characteristics feed into each other and thus benefit from as wide and frequent a circulation as possible.** Instead of buying a technical skill and sets of generic data already established by GAFAM, French and European system integrators could build new associations, as close as possible to needs.

ORGANIZED CIRCULATION OF DATA, SOURCE OF VALUE CREATION

The circulation of data needs to be organized at the scale of the field of use. Thus, quality norms and standards must be developed and adopted as close as possible to uses. The security offered by standardized protection of data is an intrinsic part of their value. In the absence of concerted, urgent initiatives, the data underlying value creation in the industrial sectors mentioned run a high risk of ending up in GAFAM business models. Industrial policy could regain its importance here in accompanying the strategies of companies competing with GAFAM: from data collection to final usages data of these industrial companies. Satisfactory use for stakeholders should nevertheless be based on a structuring agreement and a common understanding of the diverse links in the value chain. Establishing a common understanding of the data value chain is the lynchpin of public intervention and of the inter-industry relations that structure ecosystems.

Thus, without a legally validated agreement between stakeholders, the data gathered from circulating a car could belong to the vehicle owner, the manufacturer, rental company, driver, or the manufacturer of one of the devices connected to embedded or occasionally used digital services (charging stations, passage through tolls, service station cash desks) and used during journeys, etc. In any case, in Europe, provided that these data are of a personal nature, directly or indirectly connected to an individual, the entity that controls them must comply with the general data protection regulation (GDPR), as underlined by the French national data protection commission (CNIL). Which thus guarantees the rights of individuals in terms of protecting and respecting the principles of transparency and information, with their consent when appropriate. If the entity that controls the data and processes them makes them "anonymous" through adequate aggregation and/or other techniques, then the GDPR do not apply.

Looking at transport, this level of complexity is common to the aviation, maritime and rail sectors. Joint efforts should be made to develop common communication standards for the internet of objects, or IoT[2], and even for potentially standardized processing. To what extent do French car manufacturers and their first-tier sub-contractors pool their collection efforts? In what conditions can the information obtained through vocal recognition algorithms be combined with other information from vehicle usage?

Like automobile data, depending on specific usage conditions, health data may be successively or even simultaneously industrial, public or private. Their interoperability is consequently a crucial attribute that potentiates health data. Thus, data from the French health insurance information system (Sniiram) and the information systems medicalization programme (PMSI), which cover over 85% of the French population, could be usefully combined with others. Examples include data from the Parisian hospital network (APHP) or from health systems in other countries.

Pharmaceutical industrials are generally present in several countries. The effectiveness of the treatment proposed for a disease in a given country benefits from the knowledge of this same disease, and the treatment applied, in numerous countries. Opening up these administrative data and the adoption, for the part that can be made anonymous, of internationally recognized exchange standards would have virtuous consequences for patients. However, at the scale of companies and institutions (e.g. hospitals, research centres), this would involve sharing an integrated vision of the data value chain in which each link is clearly distinguished and characterized.

The use of data, at each stage of their lifecycle, involves rules, the most solid solutions for which are technological. These are accompanied by the implementation of methods for developing quality control, operational processes, and appropriate audits. The ethical concerns that are the focus of public debate would be remedied by implementing informed, technical choices that are explainable to the concerned audience. These apprehensions can partly be explained by an insufficient knowledge of computer science in society. Consequently, computer education and training are major socioeconomic issues. Capitalizing financially on digital platforms involves accelerating the dissemination of computer knowledge throughout the French education system[2].

Following this introduction, below we go on to analyse the specific conditions for digital platformization and the relationships between associated companies in the three most advanced sectors in the domain (health, automotive industry and energy). In line with our established criteria (cf. *Pour une politique industrielle du numérique*), which identify a domain's ecosystem as the best place for defining business behaviours, in particular in terms of data prices and value, we highlight the most specific pointers for each sector. The conclusion sums up the three main lessons of this work of collective intelligence.

2- This major subject is worth more than a mention. Here follows a brief reminder of the basics, based on comments by A. Vizinho-Coutry (MathWorks). Educating younger generations is essential to help them understand what they can contribute to society starting from senior high school (because that is when they decide on their next step) and to make them more confident in their career choices. However, it is not enough to simply underline the importance of learning how to code and handle algorithms. Expertise is key. In France, expertise, in particular on systems, is developed mainly in the engineering sciences and the science and technology of industry and sustainable development education branch (STI2D). These syllabuses cover IoT systems, energy management, set-up of personal assistance systems, etc. in practical classes.

1 HEALTH

1.1 "THE QUEST FOR REAL LIFE DATA?"

Expected developments in predictive medicine highlight several of the key challenges of digital platformization. Predictive medicine is based on taking a holistic approach to an individual's health. For preventive medicine to be fully beneficial to patients requires more than collecting information on people once they get ill, i.e. once the patient is affected by a condition, in the health system. Personalized follow-up throughout their lives, including when they are in good health, is the only way to take full advantage of personalized, preventative, predictive and participative medicine, known as 4P[3]. Benevolent vigilance from public authorities as regards continuous monitoring of health data throughout a person's life is likely to foster 4P medicine. The risks of certain diseases could then be better known. Early alerts thanks to predictive analytics would reduce to a maximum the expression of various pathological risks. Personalized monitoring using predictive medicine, at the scale of the vastest possible cohorts, would anticipate costly misdiagnoses. This kind of medicine would totally transform the way the health system currently operates. It would require significant investments in adapted infrastructures, with a high level of interoperability. In Europe, and in France in particular, public operators are generally the guarantors of disinterested, protective operationality. Their mobilization is therefore a prerequisite to establishing a new health economy based on predictive medicine. Specialized digital platforms, each offering preventative medical services, require this infrastructural layer of trust.

The value of data depends on the establishment of standards, which are only developed under restrictive conditions of use. People will only endorse a digital service platform if the contribution they make today guarantees them a personal advantage tomorrow.

Here once again, the technical side is important. Not all businesses have the capacity to carry out anonymized processing of data collected at individual level for personal, confidential results that involve combining immense databases. In the health domain in particular, the anonymization of some data sets continues to raise considerable problems. One solution is pseudonymization, including to conform with the GDPR. Individuals would need to be advised (informed consent, possible right of withdrawal), step by step, throughout the processing of their data. This process may initially seem complex, but it should be seen in terms of a gradual building up of user trust[4].

Moreover, this proven control may not be sufficient to encourage take-up, which requires clear communication with an emphasis on "user benefits". Indeed, the current positive spotlight on machine learning techniques may be put into question, since two layers of argument need developing in relation to trust:

- The benefit for users of sharing their health data throughout their lives.
- The solidity of deep learning algorithms (which even their supporters describe as a "black box").

3- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4204402/
4- Cf. Cahier IP n°5, CNIL, September 2017: https://linc.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip5.pdf

The methodological dimension is worth highlighting by distinguishing four "real-life data stages" : their production and production situations, including the use of instrumentation; their collection and its conditions, including quality control[5]; data aggregation, ranking, indexation, prioritization and organization; statistical processing combined with the use of exact models; restitution of the diagnosis analysis to the patient by the healthcare professional. **At no point in the data value chain does "the machine act on its own".** Although Philips tools for analysing radiological or anatomopathological imaging make preliminary sorting easier, the decision is always made by the radiologist or oncologist. FDA (US Food and Drug Administration) accreditation is based on the anatomopathological system, Digital Pathology – Intellispace Pathology. This device digitizes cytological and histological slides (e.g. biopsies). It is not therefore specifically radiological analysis using artificial intelligence. Similar principles govern the operation of annotation and automatic segmentation tools on these two types of image (x-rays or digitalized slides). With the same target result: identifying suspect cases.

At the scale of the health system, equipped with a clarified and explainable methodological approach, the health authorities could dispose of an improved pharmacovigilance capacity. On a link of the health data value chain, the use of machine learning techniques from social network news feeds would offer a new capacity that would complement the detection of alerts. Computer-assisted medical imaging already benefits from machine learning techniques. Widespread dissemination would have significant consequences on the fluidity of health pathways, appointment-making, diagnosis and delivery, and would most certainly make savings.

In the medical data chain, the traceability of origin is a key point. Founding an approach on "real life data" should go hand in hand with a precise description of collection protocols, including accreditation of equipment and staff training. In France, hospitals keep personal data on their patients and are responsible for all data processing. Patients are informed and sign an agreement with the hospital for the outcomes defined in that agreement. Hospitals maintain an absolute respect of standards and are considered to do so by the general public. Implementing indexation and anonymization tasks represents considerable expense. Hospitals pay for the equipment, software and services of the medical apparatus. Gigantic volumes of data need to be managed and stored, the figure of a dozen gigaoctets per patient per year for a hospital has been mentioned[6]. Such a high volume may justify the joint location of data at their place of use[7].

1.2
VALUE AND PRICE IN DIGITAL HEALTH PLATFORMS

These illustrations lead us to think that numerous, varied companies intend to set up digital health platforms. Due to the way that the industry works, the initiative often tends to come from manufacturers of equipment and medical apparatus. They are developing and proposing increasingly complex services thanks to digitalization.

The use of digital platforms calls for pricing mechanisms that involve an intricate interweaving of data. The value created will translate into prices that connect health industry actors with organizations and professionals in the health system – on every link of a chain of relations that ends with the patient.

These prices may be temporary or all-inclusive; the market may also take a bartering form. The prices directly depend on the data that are produced, mobilized or processed by the platform.

5- This is a central health issue, c.f. the Constances cohort: "Quality control is key to the success of Constances because it is crucial that the data to be used for research analyses should be rigorously validated, especially in 'multicentric' projects, where volunteers can carry out examinations in around twenty different centres: the results of these examinations need to be produced so as to be strictly comparable, whichever centre they were done in. Ideally, the same person could be examined on the same day in Lille, Marseille and Bordeaux and have exactly the same results." http://www.constances.fr/coulisses-constances/controle-qualite.php
6-In January 2017, the French national data protection commission (Commission Nationale de l'Informatique et des Libertés – CNIL), authorized the Parisian hospital system (Assistance Publique-Hôpitaux de Paris – AP-HP), which includes 39 hospitals, to establish its own data warehouse (EDS). This warehouse hosts social and administrative information on 8 million patients, 163 million biological examination results and 5 million medical reports. The ambition of EDS is to foster the "acceleration" of scientific research through the studies it authorizes. Cf. https//recherche.aphp.fr/eds/
7- Especially given that the dozen Go mentioned is only an average, at a certain level of digitization, in France. Note that for radiological, genomic and pathologic ("anapath") digital data, the volume can amount to several terabytes per patient.

Within a particular industrial value chain, distinct prices and values correspond to each stage in the lifecycle (raw, indexed, processed, modelled, etc.). However, some of these data can also be exchanged, even valorised, directly on a market. Bartering, exchanges, and commodification usually concern different transformation stages. Identification of the genomic signature of a disease, totally anonymized statistical processing of a group of patients, set-up and maintenance of a knowledge base indicating causalities: each of these transformation stages represents created value.

In health, the end of the data value chain takes the form of a knowledge base indicating causalities. This knowledge base is the result of a rigorous modelling process underpinned by strict quality control. This tool constitutes the "final product" of predictive analysis. Several price models are thus possible, combining purchases and rentals, and based on the intellectual property created. This concerns the algorithms, the underlying model, the processes and the structures. The annotated data can be employed for own use and, under certain conditions, may be sold or exchanged.

It can be useful to make a distinction between source data and their uses for modelling purposes. Based on this distinction, a bartering takes place between hospitals and medical equipment and device manufacturers that organizes the relationship between the parties. Access to the data produced by the equipment is exchanged for the use of models and their results. Similar mechanisms are adopted in other sectors. They are well suited to the research sphere when obtaining results is as uncertain as the associated potential economic gains.

1.2.1 Patient knowledge and consensus on quality

Clinical study phases are long and very expensive. Major pharmaceutical companies are making changes to operationalize 4P medicine (personalized, preventive, predictive and participative). These dimensions essentially rely on intelligent management of health data. The "patient knowledge" required prior to phase-3 clinical trials involves selecting a group of the most suitable people for the tests that need to be carried out. Making a step forward in "patient knowledge" would gain valuable time and lead to significant financial savings for pharmaceutical companies. Bases of digital-twin files of selected patients would be pertinent here. The methods for exploiting these "real life" data are improving through regular use. A market for digital inter-industry health platforms has developed, i.e. population health management. Philips is positioned on this market[8]. The use of this real-life data to reveal early certitudes still faces legal barriers in numerous countries. Undoubtedly, establishing consensus on the quality of data and associated models will foster the emergence of markets. Such consensus will need to involve partnerships between various organizations within health systems. A good example is the HU-PRECIMED project set up by the Association Française des Sociétés de Services et d'Innovation pour les Sciences de la Vie (AFSSI)[9], which aims to organize a precision medicine branch in France.

Partnerships between competing actors, i.e. present on the same link of the health data value chain, are possible when the object of consensus concerns representation standards. Quality norms and standards for data and models consequently constitute a key issue for developing new health markets. The joint technology initiative, Innovative Medicines Initiative (IMI)[10], an EU PPP in its second phase (2014-2020), has a public-private budget of close to 3.3 billion euro and works on developing this kind of model.

1.2.2 Shared medical records and beyond

In France, since November 2018, the implementation of shared medical records (*dossier médical partagé* – DMP, coming under the national health insurance system) could open up economic perspectives to all health system stakeholders.

The DMP is considered to be a digital health record whose data is owned by the patient. The records feature the patients' health history: care over the last 24 months, treatment followed, examination results, medical history, hospital reports, etc.

For the health insurance system, the DMP makes it easier to coordinate and ensure quality care

8-Acquisition of two specialized companies: Wellcentive in 2016 and VitalHealth in 2017.
9- https://www.afssi.fr/blog/hu-precimed/
10- https://www.imi.europa.eu/

between all healthcare professionals, including in hospitals[11]. Patients' healthcare information can be recorded in the DMP by their healthcare professionals or by the patients themselves. They are accessible at any time on the dmp.fr website or via the DMP mobile application. They can be consulted, with the patient's agreement, by her general practitioner or any healthcare professional caring for the patient. At all times, the patient controls access to the data by a secure service. She can add information (e.g. contact person for emergencies) or decide to mask certain documents. She is also informed by email if a new document is added.

In June 2019, over 6 million people had their own DMP. The way it currently works already makes it possible to anticipate savings, and even improvements in patient care. The pressing challenge for healthcare companies is to get access to this unique warehouse to carry out new research, while ensuring all the conditions of anonymity and privacy. The centre for secure data access (*Centre d'Accès Sécurisé aux Données* – CASD)[12] is one of the French specialists that proposes solutions to these safe access issues. The authorization of data access in this kind of warehouse will lead to an expansion in new value-creating services. These additional services will be offered to patients thanks to judicious combinations with other data sources. Beyond regulatory aspects, this kind of service can only emerge if people are convinced of the pertinence of services that they could receive in exchange. In complete confidence and security.

The massive statistical processing of health data needs to be carried out while preserving patient anonymity. This anonymity should be demonstratable and demonstrated by the actor that carries out the processing. Some say that block chain techniques could be used to this end, as "traceability tools". Regular industrial computing employs tokenization and rapid, precise, reliable processing techniques to prove anonymity through authentication.

To satisfy the needs that they have identified, digital health platforms need to practise a complex dialectic involving access to and use of nominative and anonymized data. The exploitation of aggregated data, e.g. corresponding to standard profiles, would open up numerous possibilities for therapeutic improvements. The Sniiram mega database is designed for health expenditure refunds. It therefore processes nominative data. Given the complexity and stakes of the information handled, upkeep and maintenance of the high-level technology required is expensive. The question then arises of the nature of statistical and algorithmic processing carried out using these data. In particular, in the context of disseminating machine learning techniques. Some machine learning practitioners, focused on discovering the correlations permitted by these data analyses, can overlook basic precautions when using these statistical data. Well-known statistician biases may occur, such as selection and endogeneity biases, presentation bias, and cognitive bias of programmers, along with economic bias, sometimes intentional[13].

11- https://www.ameli.fr/
12- https://www.casd.eu. This public interest group gathers the state represented by INSEE, GENES, the CNRS, Ecole Polytechnique and HEC Paris: it receives "Equipex" funding as part of the government investment for the future plan (Plan Investissement d'Avenir).
13- https://www.telecom-paris.fr/algorithmes-biais-discrimination-et-equite

2 CONNECTED AND AUTONOMOUS VEHICLES

2.1 UNIQUE OPEN LOCATION PLATFORM

The automobile sector is the focus of some of the biggest expectations in terms of predictive analysis, with stiff competition to impose platforms. Manufacturers of cars and equipment are all intent on establishing a platform of their own. Moreover, industry-specific and specialist platforms try to make their mark on the biggest possible number of vehicles on a global scale. However, few inter-industry platforms currently dominate the market. Few use cases exist from which managers, strategists and economists might learn lessons. And the same situation applies to numerous other sectors today. The tension between system integrators and digital firms keen to get a foothold everywhere, independently of a common understanding of the data value chain, hinders the emergence of a dominant design. While not everyone in the sector agrees on the interest of a technological block of "precision mapping services"[14], for the moment, this niche is the best example of a digital inter-industry platform featuring enough advantages to gain a following.

The company HERE Technologies[15] offers mapping and advanced navigation services for cars, and advanced information and alert services for connected vehicles, along with advanced mapping services for industrial applications or the public sector. The company collects and purchases mapping data on road networks, buildings, carparks, road traffic localization data, and data on meteorological conditions. With this material, it draws up very precise dynamic maps for close to 200 countries. It sells and licenses its mapping and navigation services to most automobile manufacturers in the world, with a market share of close to 80% (not including China, Korea and Japan). The company's shareholders, Daimler/Mercedes, BMW, Audi, Bosch, Nokia, Intel, Continental and Pioneer, are its main clients. HERE Technologies also sells services to other companies like Amazon and Garmin, and has developed several applications for the general public to simplify mobility.

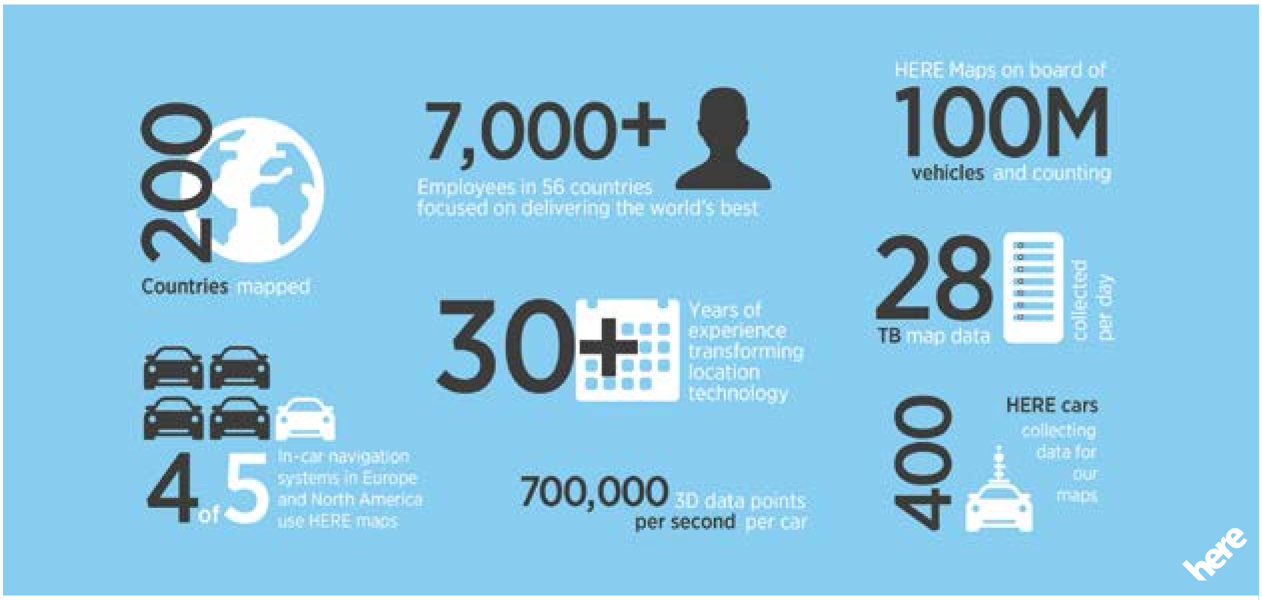
HERE Technologies has shown itself to be opportunistic and, beyond the automotive sector, has developed adjacent markets that can be accessed by advanced mapping services for cars. Its fleet of 400 vehicles circulates constantly all over the world to collect associated contextual information. HERE thus acts as an information broker and sells on this information. It also provides online mapping services. In fact, the HERE platform makes use of multiple, mostly B2B, markets in the following sectors: automobile, transport-logistics, infrastructure and the public sector, the media, insurance, electronics, telco and utilities, mass retail, health and real estate.

HERE Maps and HERE AUTO provide services to car manufacturers, their primary clientele, while HERE Location Services, HERE Mobile SDK and HERE Data Lens target companies in other industrial sectors. In addition, HERE WeGo and HERE SoMo are services aimed at a third category of clients: the consumer-users of multimodal modes of transport. According to D&B Hoovers, the company had a deficit of 255 million euro in 2017/2018.

14- 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 April 2019 cf.https://www.theverge.com/2019/4/24/18512580/elon-musk-tesla-driverless-cars-lidar-simulation-waymo

15- The background of this company founded in 1985 by Karlin & Collins Inc. with the name Navigation Technologies Corporation merits more space than this report allows. Interested readers should consult https://en.wikipedia.org/wiki/Here_(company)

HERE, key figures



Source : Presentation by J.-E. Grandjean for the FutuRIS working group, 17.01.19

HERE's strategic horizon is the arrival of autonomous vehicles. Real-time navigation of a level 4 or 5 autonomous vehicle (i.e. no driver intervention) requires geolocation with 10cm accuracy. The vehicle's exact position, with respect to all of its environmental constraints, including roads and other vehicles, represents the complex primary matter of all interactive behaviour models. Overall, much of the immense quantity of data captured by car sensors is far from being exploited.

Based on the above observation, HERE Technologies has developed its Open Location Platform (OLP). This platform is used to collect and process in real time data from sensors on connected vehicles in order to provide real-time, high-definition mapping services (HERE HD Live Maps) and information and advanced alert services for connected vehicles (HERE Connected Vehicle Services). Over one million Audi, BMW and Mercedes connected vehicles currently feed this platform in real time thanks to anonymized data. However, the platform is "open" to all, both for inputs and outputs: the added-value services portfolio generated by HERE on OLP is available to all automobile and industrial clients, whether or not they have contributed the underlying data. What is more, any actor can use OLP to develop its own added-value services, for itself or with a view to commercialization.

As a result of integration, the HERE platform directly establishes contracts with manufacturers and thus benefits from extensive, deep-seated access. In its form, the situation is comparable to medical instrument companies that enter into contracts with hospitals. The potential value generated by using this kind of advanced navigation system results from combining and sharing information gathered by different types and brands of sensor. To simplify access would therefore involve a different concentration of information, in a single place, or at least by a single actor who owns the data. In Europe, and internationally, the concept of an "extended vehicle" developed as a norm (ISO 20077-1) is being worked on. Car manufacturers support this solution as the only one capable of ensuring effective, high-level privacy protection.

In parallel, in order to guarantee the interoperability needed for exploiting the data produced, HERE has made public a standard format for ingesting data from sensors: SENSORIS, which stands for Sensor Ingestion Interface Specification. This format describes the technical prerequisites for producing and exchanging data from sensors on circulating vehicles. The data collected by in-vehicle sensors are sent to the cloud and the vehicle maps are updated on the fly. Vehicles that subscribe to the OLP's value-added services are then also alerted in almost real time about any obstacles or accidents on nearby roads.

The HERE Connected Vehicle Services value proposition centres on real-time traffic news, updating of permanent changes in road signs, diverse risk warnings, and detection of free parking spaces. All of these services could be considered as making use of data in the general interest. In conformity with the European directive of 20 June 2019[16], the public sector should thus make them available for no charge. This would open the door to potential economic developments. Value creation follows the processing of data, i.e. cleaning and quality checking, possible indexations, formatting for mass statistical processing, possible drawing up of algorithms, etc. All of these tasks are costly and their results need to attract clearly targeted clienteles.

HERE's "core competency" involves implementing the business model that best serves its profitability research strategy, for each layer of its OLP. At each layer of the platform, i.e. "data market place", "data", "developer environment and platform foundation" and "services and solutions", specific techno-economic conditions apply the potential of which needs to be exploited for a profitable price.

The underlying managerial and governance efforts aim to combine:

- the need for international development in constraining local conditions, with
- a very considerable need for cash to invest in technological infrastructures and the related rare skills.

The specific characteristics of this platform, in other words, its multi-layer character for variable local profitability conditions, constitute a source for driving its business model. The fact that similar types of information can be sold at different margins depending on the target customer could make the main shareholder-clients uncomfortable. What counts is the final use of the data.

2.2 PRIVACY IN PLATFORMS AS A PUBLIC GOOD

This type of platform offers services mainly aimed at other companies. They nevertheless feed off data produced by sensors that implicitly draw up a portrait of final users' behaviour (mobility, health). In actual fact, the data received by HERE Technologies are pseudonymized by the car manufacturers. HERE therefore hosts anonymized but detailed information on vehicles. Although HERE strictly abides by European rules on personal data (GDPR), the practical and legal distinction between industrial data (anonymized) and personal data can sometimes be difficult to make, especially in rural or very low-density areas.

Economics research has attempted to model e-privacy problems[17 18]. The new context of digital platformization is characterized by the intense use of social networks and other internet exchange platforms. In this situation, confidentiality can be considered as a "public good". It has been proven that it only takes a very low percentage of participants in a network to renounce their privacy (e.g. in exchange for using network services) for it to be possible to reconstitute all of the information of all of the participants. Including those who did not consent, and information regarding other people who are not even on that particular network. Since individual consent is insufficient, only a strong regulatory policy is capable of providing some kind of solution. A powerful, generalized regulatory approach is based on technical means and methods that result in the respect of anonymity. The (European) General Data Protection Regulation on protection of private data is perceived as a first positive step in terms of e-freedom.

16- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (revision). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L1024
17- Fairfield, J. A. T., Engel, C., 2015, Privacy as a Public Good, Duke Law Journal, 65, 385-457.
18- 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.

2.3

USING DATA AND ALGORITHMS FOR INDUSTRIAL PURPOSES

Inadequate control of the statistical concepts underlying some applications of inter-industry platforms is likely to have damaging effects. The inclination to model the behaviour of complex systems via machine learning algorithms can also come up against an insufficient consideration a priori of the basic physical principles governing these systems. To obtain forecasts of irreproachable industrial quality, it is vital to qualify relationships, meaning and power, and to be capable of weighing up the different attributes. These data preparation tasks, which require detailed knowledge of the underlying operations, are therefore crucial and costly. Yet the reliability and robustness of the model results depend on them. The failure of IBM Watson Health is mainly due to incorrect annotation of the data. Which requires similar expertise to that of experts who can do without prediction-aid tools, like for example experienced cardiologists or neurologists.

To ensure that data, and thus the results of statistical processing from data, i.e. forecasts, attain industrial quality necessarily requires turning to rare, top-level expertise.

This contrasts with the image of automated, immediate services that we associate with these algorithms when used for Facebook and Amazon. As HERE clearly illustrates, the price of services provided by platforms results from investments in technologies and top-level scientific and technical skills. It is true that economies of scale and network externalities are characteristics of digital platform business models. But the infrastructure costs involved can be high and even sometimes be “sunk costs”.

3 ENERGY

3.1 UBERIZATION SEEN AS A "STRATEGIC RED LINE"

From an energy operator's[19] point of view, there are two types of platform: B2B, which is technological, and B2C, which is about economics. While it might be possible to develop some of the functionalities of a B2C platform towards B2B, the opposite is not true. Adopting this point of view and making it known opens up potential economic relations that rely on trust within inter-industry digital platforms.

The risks of this development – in particular a radical loss of trust – could explain the relative failure of such attempts. Imagine that EDF considers digital platform P to be indispensable to monitor certain parameters of its nuclear power stations. The relationship between EDF and platform P is built on technological foundations. No commercial development needs to be sought based on data or the results of proposed analyses. Companies in the EDF ecosystem are thus protected, along with the quality of interactions, over the long term. Hence, EDF will never try to benefit from valorising the data that originate in third-party systems. Conversely, in a platform relationship with its final customers, added-value services are likely to result in the valorisation of associated data. For example, in the case of an electric vehicle charging station network, EDF could employ the data exchanged between vehicles and the stations to optimize the cost of charging, or even "remunerate" customers for supplying services to the network.

The ownership of the information present and employed can and should be guaranteed by technical solutions – although it requires being able to control this layer underlying the data analysis, and then being capable of communicating on the methodologies involved. Use cases exist that would benefit from wider dissemination. In Europe, the GDPR and PSD2 (revised directive on payment services) stipulate the ownership of user-consumers' data and the obligation to request informed consent. These regulations foster initiatives that provide interesting use cases. From a competition law point of view, competition works thanks to the transparency of information. The clearer the initial distribution of property rights, the less friction will occur that is likely to perturb market operations. **The availability of APIs (application programming interfaces) plays a key role in exchange flows.** APIs provide detailed information on platforms for a particular functionality. Thanks to these "contractual facades", third-party software applications can mobilize one of the platform's functionalities. Their existence makes clearer the sharing of data ownership between the different companies interacting through the platform.

The 170 energy operators in France that act in the building sector are grouped together into the ORE Agency (*opérateurs réseaux énergie* – energy network operators). This agency drives the digitalization of energy networks, and the management and valorisation of data based on an open data approach. ORE promotes and facilitates the energy transition through digitalization. It develops three types of service: visualization of energy data, energy overviews, and tools for market actors (such as requests for certification of switching capacity, or additional remuneration mechanisms). The APIs proposed allow data owners to take control.

19- According to the research by foresight experts at EDF, cf. their participation in our working group on 15 November 2018.

3.2

TOWARDS OPTIMIZED ELECTRICAL MANAGEMENT

EcoStruxure underpins the overall architecture of Schneider Electric's IoT platforms. Each of these platforms operates a set of technical solutions applied to specific energy optimization problems in houses, buildings, data centres, infrastructures and industrial companies. These solutions feature safety, reliability, operational efficiency, sustainability and connectivity aspects. EcoStruxure exists in several versions corresponding to expertise in the sector: *electricity supply, computing, construction, equipment/machinery, factory installation, and network.*

What artificial intelligence departments offer in industry

The Analytics and Artificial Intelligence team at **Schneider Electric** uses all data analysis and optimization techniques to produce applications that are useful to its clients. These are broken down into seven types of functionality:

- Evaluation of performance and benchmarking
- Disaggregation of data and discovery of information
- Data correlation, prediction, revelation of underlying structures and models
- Monitoring and diagnosis of equipment and systems, predictive maintenance
- Planning and scheduling of activities and resources (including energy)
- Context-dependent controls
- Decision support through simulation.

Source: Presentation by C. Le Pape for the FutuRIS working group, 21.02.19

A specific type of services offered by the EcoStruxure platform, i.e. advisors, provides a good illustration of several key features of digital platforms for industry in terms of data. *Ecostruxure Asset Advisor* and *Ecostruxure Microgrid Advisor* demonstrate the benefits of exploiting the complementary nature of multiple data, from different sources, and varied conceptual approaches to predictive analytics. Meteorological and pricing data are thus crossed with data coming directly from electricity systems and their uses. Multi-scale, multi-physical models are combined with data analysis and statistical learning techniques at different points in the resolution of problems, i.e. identification phase of the underlying model, elaboration of a reference model as close as possible to the collected data, validation tests of the model, then tests to identify any divergences with the expected behaviour of the operating system, and lastly to trigger the alert or suitable correcting command.

Smart microgrids[20] are small electricity grids designed to supply reliable, high-quality electricity to a small number of consumers. A study of this type of grid points to the diversity of the data involved and their multiple uses, whose value depends on the context of use. The support provided by a small number of companies, like Schneider Electric, shows how these data can be combined to improve decision-making – both before initial investment and during installation, maintenance and day-to-day control.

A first stage usually involves the technico-economic analysis prior to setting up the microgrid. This can for example involve a network of solar panels and energy storage on a site. EcoStruxure Microgrid Advisor anticipates changes to the site's production and load. These changes depend on weather conditions and the correct use of batteries. Good management of these developments results in proposals to reduce future electricity bills. The lifespan of batteries depends on the number of charging and discharging cycles. Electricity purchase or sales contracts also vary widely. The price of electricity and the revenue made from resale or the contribution to different services rendered to the network (e.g. regulation of frequencies) affect electricity production and consumption. Characteristic data on site operations in terms of production and consumption along with the factors that influence operations are therefore mobilized. Since they provide information on the site's activity, these data are generally considered to be extremely confidential. These same data also have other uses, such as evaluating the energy performance of the site, improving control of heating systems, ventilation and air conditioning, and detecting anomalies.

Electricity pricing, possibly including feed-in tariffs for wind and solar energy, vary structurally from one large region of the world to the next.

20- Cf. The French Energy Regulatory Commission (CRE) website: http://www.smartgrids-cre.fr/index.php?p=microgrids

In simple cases, optimizing the electricity bill often depends on two components in the United States and a single component in France (the first one mentioned below): (i) the total cost of consumption over given time periods (peak, off-peak), which is the unit price of the period multiplied by the energy used during that period (in kWh) and (ii) the price of the highest demand, known as the demand charge, over a given period, e.g. monthly. For operators, a reduction in the demand charge represents a major constraint that is complicated and expensive to guarantee. This demand charge is a key parameter in optimizing the design and then running the microgrid. The value brought by *Microgrid Advisor* to its clients therefore results from the best estimation of the expenses of the demand load, taking into account the cost-advantage economic balance of the solar panels and batteries.

Lastly, the conditions for the economic development of predictive analytics clearly depend on the choice of public policies on norms and regulations. The latter have major technological and industrial implications. A platform can decide to offer a service depending on the electricity pricing conditions in the economic area concerned. In this case, the presence of the demand load component in electricity pricing in the United States makes it more complicated to produce the microgrid but accelerates its profitability. And thus facilitates the sale of services provided by *Microgrid Advisor.*

The connected thermostat *Wiser Air,* only available in Canada and the United States, presents another facet of optimization methods. Wiser Air learns from the actions of housing users who, by indicating whether they are too hot or cold, allow the system to calibrate the desired temperature to match the situation through learning. The right temperature, i.e. that which is the most comfortable for inhabitants, results from their interactions with the system. Once the dynamic setting has been established thanks to human adjustment, *Wiser Air* tunes it more finely so as only to use the quantity of energy necessary to achieve the desired level of comfort. These tests may then possibly be further corrected by humans to progressively converge towards the optimum comfort/price for energy consumption.

3.3

DATA VALUE CHAINS, THREE KEY INDUSTRIAL ISSUES

Three major industrial issues can be highlighted through analysing data from complex systems. It begins when an electricity grid sends out an alert, including breakdown predictions. Questions of interpretation, comprehension, and intervention-repair then arise. And the issues raised can have significant organizational and financial repercussions.

3.3.1 Understanding the alert of a system in operation

The first step involves adopting the best explanatory model of the problem encountered. This model is most often multi-physical (e.g. combination of thermal, chemical and electromagnetic), which thus requires having an up-to-date supply of models for all of the systems, subsystems and components operated. Next, analyses and tests need to be made from a guaranteed high-quality data source. The data, which come from the functioning of systems in operation, must be captured using a controlled procedure and selected in sufficient quantity to ensure that they are representative. The following step is a systematic exercise of benchmarking algorithms in order to choose the one that provides the most convincing estimation. And all the more so since the information on measuring the reliability of the estimations proposed constitutes one of the characteristics of the tested algorithms. Tests of algorithm robustness are being developed. However, to date, real-time benchmarking of algorithms based on digital twins is rarely implemented on complex systems. The up-to-date theoretical models are fed and completed by standards and guidelines in the field. For our example, in the United States, this indispensable information is available from the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE). This is the case for the choice of expected and legal indicators and performance values, and for interpretation. These data processing and analysis stages require complicated, costly preparatory manipulations.

3.3.2 Moving from a learned model to "real-life" data (beyond the original learning area)

The performance of modelling based on machine learning starts with the data set from which a model was "learned". Using new data, like real-life system data, is not so easy. In fact, using a learned model requires a very strong qualitative and structural resemblance between the learning data and application data sets.

This very close proximity remains a construct. It requires the same quality of cleaning, recalibration, marking and indexation for both sets of data. The learning dataset also need to be representative of situations encountered in the real world. This must be established and verified with an expert. In the case of classification of e.g. causes and anomalies detected on an apparatus or system, the learning data must be labelled with the causes and anomalies observed in the past and the labels must be verified by experts in the field. This use of experts constitutes a bottleneck in the development of specialized models using machine learning. The skills required to validate and label data are valuable. The people that possess them are assets for the organizations that employ them, from designing new projects to providing support for strategic clients and understanding clients' emerging needs. Mobilizing these experts in a significative and repetitive way to prepare machine work, or perhaps to check it, may seem unflattering and less directly profitable than other contributions made by these same experts.

Two techniques are often put forward to adapt or generalize a learned model to a new context. The first involves simulating a digital twin to generate new data, which can be used to learn a new model. In the second, transfer learning is used to transform a model learned earlier in a given context into a model that can be applied to a new context and therefore to data situated outside the initial learning area. The use of this kind of technique does not however do away with the need for validation. In the case of simulation, the need is in fact shifted, since the model used to simulate the real system must be validated itself.

The question of delimiting the learning space and the application space is therefore scientific, technical and economic. The refined data – suitably cleaned, organized, labelled and extended over time to take into account changes in context – have acquired a value that was absent from the initial raw data.

3.3.3 Improving systems in operation, and developing them thanks to data

Companies that supply services through a digital platform operating several products (a transformer, a water heater, an engine, an automatic device) attempt to improve them by accessing data in operation. A connection must be maintained between the conception and the operations. These retroaction loops are channels through which the main factors influencing the performance of the system considered are expressed. The veracity of the alert sent out by the system must be absolutely validated. It is necessary to ensure a realistic understanding of the mechanisms rather than simply observe a certain number of converging correlations.

When a monitored system evolves, the models need to be refined. The methods used can involve systematically and gradually introducing perturbations into the learned model, using simulations, or even digital twins if the system is not too complex. If the tests indicate that the quality of forecasts tends to deteriorate over time, it may be that the correct explanatory factors have not been considered… or that the conditions have effectively changed. Statistical tests can then be carried out to check the degree of anomaly. Most often, in the case where a prediction takes the form of a digital value, the method involves monitoring how the cumulative sum of the values of successive errors evolves, or verifying that the values of the series remain within the interval centred around the average between 2 and 3 standard deviations. The opposite case requires explaining the anomaly: is it due to insufficient information, changes, malfunctioning sensors, incorrect intermediate predictions, an exceptional event? Before improving, or even correcting a model in terms of alerts, a number of precautions need to be taken.

4 RETURN ON DATA VALUE AND PRICES

STRADDLING CORPORATE STRATEGY AND INDUSTRIAL POLICY

Since this document set out to provide an overview, this final section does not aim to close the debate, but instead underlines the three central aspects of this analysis of business-to-business digital platforms, i.e. the establishment of a common understanding of the data value chain; the associated condition of developing and controlling an economy of trust; the exploration of platformonomics aimed at curbing differentiated treatment behaviours.

4.1

FOR A COMMON UNDERSTANDING OF THE DATA VALUE CHAIN, BY CORE BUSINESS

Predictive analytics serves the industrial sphere by combining varied modelling methods that must consider core business skills (i.e. "metier expertises") to be effective. The quality of the predictions depends on the validation, certification, verification and homologation of the data and explanatory models in which digital industrials invest. The value of industrial data, throughout their chain, results from this expert processing, which is usually of a very high level. The investments involved are integrated into the price of using digital platforms. Open standards that foster interoperability are at the origin of financial gains derived from the private appropriations allowed by intellectual property rights associated with another link of the data value chain. The preliminary step must involve establishing a common understanding, business link by business link, mainly at the initiative of the system integrators involved and in cooperation with digital firms.

4.1.1 Cyber-risky environment: a wake-up call

Cybersecurity has evolved considerably over the last five years. Attacks have moved on from being physical and costly to becoming logical and financially profitable. This is because of the flood of connected devices integrated into most everyday objects. For criminals, succeeding in taking control of an object for a given investment brings immense financial returns at the level of its dissemination. The average cost of attacks has increased considerably. Two million euros can now work out to be perfectly profitable, involving, for example, all payment terminals of a certain type, all cars of a particular model or even all brands of one generation, electrical power stations of a particular type, etc. The remarkable wide-scale attacks that took place in 2015 served as a wake-up call to the vulnerability of numerous systems. In the United States, for example, the Federal Bureau of Investigation (FBI) in partnership with the Department of Transportation (DoT) and the National Highway Traffic Safety Administration (NHTSA) published an alert stating, "motor vehicles [are] increasingly vulnerable to remote exploits"[21]. The problem is only getting worse, since the number of connected objects is accelerating, along with the relevant use of such objects in everyday life, and thus the financial value concerned. Vulnerability and incentives for attacks are increasing. The economic interest of cyberattacks is greater, and yet for the time being, they are mostly the work of certain countries.

Faced with this new cyber-risky environment, specialists are taking action. This is illustrated by an increasing number of conferences for

21- Cf. "Motor Vehicules Increasingly Vulnerable to Remote Exploits", FBI & NHTSA, 17 March 2016 cybersecurity professionals, like Black Hat, which started out with one annual conference in the United States and now organizes several each year all round the world. The events are an occasion to share the state of the art of attacks and suggest solutions to tackle them. The US National Institute of Standard and Technology (NIST) published a white paper[22] in October 2018 that lists the security issues facing the internet of objects (IoT). The type of alert proposed originates in the observation that the dissemination of the IoT requires a totally new approach to security. Through a comparison with the way that standard computer peripherals work, the NIST points out that numerous IoT peripherals:

- Interact differently with the physical world.
- Cannot be turned off, managed or monitored in the same way.
- Often have different degrees of efficiency, effectiveness and availability in their cybersecurity and confidentiality functionalities.

The consequences of this last aspect merit attention. The capacity of establishing reliable measures of the quality of IoT functionalities has become one of the major challenges of these new computer threats. Measures can bring trust. On the other hand, a lack of measures and performance indicators for IoT makes trust a concern. Metrics and measures are the basis of trust. The IoT features relatively young technology. To date, few means exist to measure IoT systems other than by counting or using specific dynamic tests. This lack of objectivized legibility constitutes an obstacle to disseminating the most reliable IoTs. It is complicated to assert that a system is reliable or even to estimate the number of tests that it should be subject to.

4.1.2 Security by design

IoT objects, which might be sensors, aggregators or e-utilities, are often difficult to test and certify, including by their producers, due to interactions with their precise context of use. Some uncertainty persists regarding their permeability. They are therefore attacked, and users do not always even realize it. There is only one way to identify and therefore protect an IoT from being attacked, and that is using formal methods. These are the computer equivalent of a mathematical proof. The commercial solutions available – dedicated operating systems – supply the automatic proof of the theorems underlying the verified system (e.g. see Prove&Run[23]).

The unit cost of adding this kind of proven security to any IoT is very low. For multiple reasons, including insufficient awareness of the risks, such as reputation risks, these solutions by design are not yet widespread. The most suitable sector is defence and national security. Promising specific applications currently include mobile telephones, streamed video websites (the move to secure 4K), and systems that use cameras for secure exchanges, etc.

4.2 ASSOCIATING DATA VALUE AND PROTECTION VALUE. TOWARDS AN INDUSTRY OF TRUST.

The value of data is directly related to the extent to which the data are protected. The connection works both ways. High-value data require a high level of security. The fact of guaranteeing their security in turn increases their value. The choices made by operators and owners – which can be conditioned by technical aspects – in terms of production, storage and/or processing in the cloud, using hybrid edge computing, raise specific security issues. All are identically liable for a by-design security approach. **Whatever the domain, the question of the (geographic) location of data comes after security issues.**

National sovereignty thus initially takes the form of a simple, auditable infrastructural skeleton. These two qualities form the foundation of trust represented by sovereignty. In industrial and commercial terms, the major issue is clearly reputation: it is trust that has been granted and renewed. Any hitch in the granting undermines that reputation. Thus, the clear and simple acknowledgement of the use of the best technology underlying services and products sold can constitute an essential guarantee.

22- https://csrc.nist.gov/CSRC/media/Publications/white-paper/2018/10/17/iot-trust-concerns/draft/documents/iot-trust-concerns-draft.pdf
23- Cf. https://www.provenrun.com/solutions/. The computer security solutions developed by Prove&Run are based on this approach. They attain the highest evaluation assurance level (EAL7 or ITSEC E6). These operating systems secure all types of microprocessor, based on any architecture. Prove&Run's revenues come from licences on its operating systems. However, microprocessors can be designed in several ways, and they are not neutral when it comes to establishing formal proof. The best solution is "sure by design" chips.

From the point of view of companies, what makes an industrial, among other things, is his capacity to correctly gauge the potential risks run by his processes. This aptitude, which is rooted in experience and translated by quality control considerations, leads to measuring the degrees of gravity, which are then associated with the levels of security to put in place. In organizations, the extent of the protection to be implemented, and so the human and financial investments involved, depend on sharing technical data at all decision-making levels. However, sometimes the pertinent decision-making items have trouble reaching the pertinent decision-making levels.

The two key factors in computer trust are auditability and immutability. This is valid for both systems and data. The widest possible dissemination of formal methods capable of effectively guaranteeing auditability and immutability is likely to promote the emergence of a new industry: an industry of trust. This strong promise of a formal guarantee of the execution of a function corresponding exactly to expectations is likely to result in the equitable sharing of the created value. Lastly, a regulatory framework is part of the infrastructural skeleton specific to auditability. The GDPR encourages propagation of the trust that individuals grant to a collection and processing system implemented by those responsible for processing. The latter thus need to provide a practical demonstration that they respect the rules, in total transparency, e.g. proof of implementing security and respect of access and modification rights.

4.3

THE MOVE TO *PLATFORMONOMICS*: THE QUESTION OF DIFFERENTIATED TREATMENTS[24]

As observed in *"Competition policy for the digital era"*[25] by J. CREMER et al. (2019), the contributions made by economics to understanding the consequences of digital platformization largely correspond to the standard doctrine of competition law. While the notion of two-sided markets turns out to be an effective argument for an in-depth, case-by-case examination of the conditions for establishing prices, theoretical censorship is ultimately in the hands of consumers. They should never suffer from any distortion of competition due to the advantageous position that companies may benefit from.

The industrial organization problems tackled by economic theory in the presence of digital platforms include that of differentiated treatments. An e-commerce website is not supposed to treat differently any of its suppliers or products on sale. It plays its marketplace role in a strictly neutral manner. E-commerce websites are theoretically pushed towards adopting this kind of neutral behaviour by the competition between them. However, competition encourages them to stand out individually, more or less radically, and thus avoid easy comparison of the price of products sold. These e-commerce outfits tend to avoid direct confrontation. For example, they might organize themselves so as not to be compatible with each other in the eyes of the suppliers and producers whose products they sell. Or they may use algorithms for customer decision-making that are devised to encourage certain choices and not others, based on criteria not chosen by the customer. Identifying possible differentiated treatments is in fact particularly complex for analysts due to dominant platforms' tendency to develop "own brands". This involves drawing from the data of millions of purchasing acts to offer products similar to best-selling brands at competitive prices and of a similar quality. And without necessarily indicating that the products are "own brand".

From an economic point of view, this kind of operation unfairly increases the profits of platforms and reduces consumers' well-being. The rules that competition authorities aim to guarantee all over the world are therefore jeopardized by platforms' behaviour. To guarantee fair, equitable prices, the legislator is starting to draw on the services of top technicians specializing in algorithms, mass data analysis and platform architecture, etc.

24- The following paragraphs take modest inspiration from the presentation by Doh-Shin JEON (Toulouse School of Economics) on 19 April and do not claim to have a similar depth of reasoning or in-depth technical knowledge of the subject. They are only an extrapolation related to the subject of this report.
25- http://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

These differentiated treatments issues primarily concern e-commerce platforms, and therefore the relationships between consumers and businesses. However, they naturally end up involving technical questions of behaviour between businesses within the value chain. Key challenges include access to and control of mixed data. In the case of digital platforms that connect businesses, practices are not yet stabilized and thus poorly understood. Consequently, economic theory and competition law are lagging behind.

The main recommendation of this report, initiated by the implementation of platforms by system operators themselves, is to establish a common understanding, by sub-sector, of the data value chain. Our strong assumption (cf. *Pour une politique industrielle du numérique*) is that ecosystems are the perfect place to do so. These business ecosystems centred on predictive analysis bring a hope that, with suitable industrial policy measures, could become a reality.

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