Sharing the value added from industrial Big Data fairly

The know-how of EU industrial workers must not be stolen by Big Data monopolists!

The "data economy" under discussion since January 2017 may lead to (1) a concentration of wealth and power in the hands of the few "Big Data Analytics" companies holding a monopolistic grip on industrial data, and (2) a de-skilling of industrial workers because their know-how would be captured by the data and algorithms. IndustriAll Europe proposes policies to prevent these risks: (1) a reinforced Social Dialogue to preserve the professional know-how of industrial workers and to share the value added generated by Big Data Analytics, (2) a licensing regime for the (non-exclusive) rights to access and to process "raw" industrial data, and (3) a protection of "processed" data under a revised regime of "industrial secrets".

Industrial "Big data" and analytics: the revolution ahead

Recent advances in sensor electronics and in digital information processing technologies have made the collection and storage of industrial, machine-generated data much easier and cheaper. Now, any production machine can easily be equipped with a set of sensors, and with a data storage and transmission system, thereby facilitating the logging and storage of output for further analysis. This data is often referred to as "**industrial data**", "machine-generated data" or "sensor-generated data" – to be distinguished from "personal data" (i.e. related to a physical person), which is the purpose of a specific protection in the EU under the General Data Protection Regulation[1].

The **volume of data** being generated by industrial machines equipped with such sensors and data logging devices is **enormous**. As an example, General Electric states that an aircraft engine generates 1 terabyte (= 1,000 Gbytes = 1 million Mbytes) of data for each flight.

The **analysis** of these hoards of data is performed by the recently developed "**Big Data Analytics**" algorithms. These algorithms are an adaptation to (very) large and heterogeneous datasets of methods (statistics, automatic control, Artificial Intelligence based on "neural networks") that were used for decades to extract meaningful information, to be used in the future (e.g. the average life duration of a mechanical piece), from a set of "noisy" data collected in the past (e.g. the life duration of all mechanical pieces of the same model). The (big) difference compared with before is that such statistical extraction of information is now much easier and cheaper to perform, due to the massive availability of low- to zero-cost industrial data. It can thus be applied more systematically, and in more situations.

Workers' know-how in industry risks being captured

Industrial production transforms (raw, intermediate) materials or objects by applying a physical and/ or chemical process on them (e.g. laminating a steel slab, machining mechanical components, steam cracking naphtha to obtain ethylene, pulping wood to obtain paper, tanning hides to obtain leather…).

In all these industrial processes, the main issues driving factory profitability are: (1) the **quality** and yield of the output and (2) the **load factor** of the installation (i.e. the fraction of time when it is operational) – or, reciprocally, (1) the reject rate of the product and (2) the failure rate of the machines.

Industrial installations are already very much driven by **automatic controls** based on processing the data

[1] Full text available in all official EU languages here

immediately generated by the sensors installed on the machine in short feedback loops. However, the rules governing these feedback loops are based on the underlying scientific knowledge of the industrial process of the machine and of the processed material.

This scientific knowledge is imperfect, because:

- the machines are often so large and complex that they are custom-manufactured and individually maintained, so that each of them is different
- the processed material is often of natural origin (mineral ore, oil, cotton, wool, flax, animal hides, wood…), and therefore heterogeneous and not fully controlled
- the environmental conditions in the factory (temperature, humidity, vibrations) are not fully controlled
- the scientific knowledge on the underlying chemical and physical process was acquired on small-scale pilots, and does not translate well to the full-scale industrial installation.

For all these reasons, the automatic controls need to be complemented by the **human know-how of workers**. Workers in industry have learnt empirically about the behaviour of the machine and about the features of the material being processed. Their acquired experience and skills in operating and maintaining each individual machine enables them to fine-tune its operation, and to reach the high level of quality and reliability required[2]. This is particularly true in the continuous processes found in the Chemical, Pharmaceutical, Basic Metals and Materials, Textile and Leather sectors, but it is a very general statement, valid in all industrial sectors.

Similarly, the economically important **incremental** improvements to the **design** of the machine (performed by engineers and technicians in the Mechanical Engineering sector) are based on human feedback from industrial workers, in close, human interaction[3].

This situation will be **dramatically changed** by **Big Data Analytics**.

One way to see "Big Data Analytics" algorithms is that they enable computers to "**learn**" about the behaviour of the machine and of the processed material based on the statistical analysis of the huge amounts of data collected in the past from the same machine (or from other machines of the same type in different factories), just as humans would do.

More specifically, the "Big Data Analytics" algorithms can perform the following tasks:

- relate the **control parameters** of the machine with the **quality** of the produced output, just like a skilled worker in **production** would do
- **anticipate the failures** of the machine, just like a **maintenance** worker
- identify the features of the machine that need **incremental improvement** (e.g. the pieces that deteriorate faster, and which could be reinforced to enhance reliability), just like an engineer or technician in the **Research & Development** department of the machine manufacturer.

This can be interpreted as a **capture** by these algorithms of the **know-how** of workers in industry:

- of the skilled workers in production and maintenance in the company **using** the machine
- of the R&D engineer or technician in the company **designing** the machine.

This capture of workers' know-how is problematic in itself, since it can be a source of massive **de-skilling** of workers in industrial firms (and in the Mechanical Engineering firms supplying them with machines and equipment).

The situation of workers in industry will also be impacted by the **identity** of the companies having

[2] As an illustration, the Russian steelmaker NLMK reported in 2017 that a skilled workforce enables the reject rate for steel coils to be reduced from 17% to 3% - a determinant factor for factory profitability.

[3] Other, more fundamental improvements to the design of the machine also exist, and are based upon the fundamental scientific and technical knowledge of the underlying physical and chemical process. But these "radical" innovations are rare, whereas the competitive edge of most companies in the Mechanical Engineering sector is based on their capacity to implement many small incremental changes and innovations. This is why the "incremental" innovation process is so important from an industrial competitiveness point of view.

access to industrial data, and operating the Big Data Analytics algorithms, i.e. whether this company is:

- the company **operating** the installation
- the company **designing** the machine, from the Mechanical Engineering sector
- a **fully digital** company exploiting Big Data Analytics algorithms.

We will consider these cases separately hereafter. We will then consider the **policies** regarding the legal and economic **rights** to **access** and **process** industrial data, which influence how the value added will be shared, and between which stakeholders.

Big Data Analytics controlled by the operating company: Social Dialogue and Collective Bargaining are the best adapted institutions

The case where Big Data Analytics is performed by the industrial company operating the installation is the most favourable for workers. Indeed, this is the case when the **value added** from the exploitation of **industrial Big Data** remains **within** the company, so that the existing institutions gathering trade unions and employers, which cope with industrial transformations, i.e. **Social Dialogue** and **Collective Bargaining**, can be used fully.

It is then the purpose for these institutions to discuss:

- how to **preserve** the **professional know-how** of industrial workers, while exploiting Big Data Analytics as a tool to **assist** workers, and not to replace them
- the **sharing** between workers and management of the **value added** being generated by Big Data Analytics algorithms, and by the resulting improvement in the quality of the product and in the reliability of the process.

Big Data Analytics controlled by the designer of machines: a renewal of inter-sectoral bargaining

If Big Data Analytics is performed by the company designing the machine, the balance of power between the user and the designer of the machine is shifted in favour of the latter, and this may influence the sharing of the value added between these two sectors.

However, this would not involve a radical change. Knowledge about the underlying chemical and physical process has always been shared between the machine operator and the machine designer in a fruitful and mutually beneficial inter-sectoral dialogue. The symbiotic inter-dependency between the manufacturers of production machinery, on the one hand, and their industrial customers, on the other hand, is so deep that it forces the relation, and the inter-sectoral bargaining on the sharing of the value-added, to be **balanced**. This symbiotic relationship is loose when the machines are very generic (e.g. machine tools) and much deeper when they are custom-designed, complex industrial installations.

Social Dialogue and Collective Bargaining within the companies of the Mechanical Engineering sector that design and manufacture the machines would be equally as beneficial as in the companies operating the machines, with the same goals.

Big Data Analytics controlled by digital monopolists: a major risk to European industry

The third possible scenario is that in which one purely digital company would exert **monopolistic control** over industrial, sensor-generated data to feed its proprietary Big Data Analytics algorithms. This company monopolising access to industrial data would thus be controlling the **differentiating asset** of the industrial plant in terms of product quality and of process reliability. It would then be in the position to demand **any** price for its Big Data Analytics service as soon as it has locked in its industrial customer, in effect pushing this price up to the level where it **captures** the whole **value added**. The current industrial company would be left with the low-value functions of simply owning the material productive assets (the machines and the buildings) and of running the plant (purchasing the raw material, paying and controlling the workers and selling the final product) in bare survival mode. This evolution would be difficult to revert, since all its **know-how**, and the **skills** embodied in its workers would have been captured by the Big Data Analytics algorithm and by the stock of industrial data feeding it.

This situation is the **worst scenario for European workers in industry**, among the three considered. In this scenario, workers would have no partner with whom to discuss a fair usage of technology, which respects workers' skills and experience, or about the sharing of value added. They would be left with the **low-qualification** tasks of **pure execution** of the orders given by the Big Data Analytics algorithm, with low pay and bad working and employment conditions. Their traditional social partner, the company or even the sector-relevant employers' association, would have lost its relevance and be left with empty pockets and nothing to bargain about. Basically, the value added would have **shifted away** along the supply chain to a new player, the Big Data monopolist. This Big Data monopolist would itself concentrate the value added of the whole industrial sector whose industrial data it controls. Since it hardly employs any workers, it could afford to pay them well, with no incentive to share this value added with the industrial companies of the sector – and even less with their workers.

The rights to access and process industrial data: the key to fairly sharing value added

The European Commission has identified the central role of the rights to access and process industrial data in its Communication and supporting Staff Working Document on "**Building a European Data Economy**" – Part 3 "Data Access & transfer" (January 2017).

These documents highlight the current **legal void** regarding industrial data. Public policies protect the economic rights of players that engage in societally useful activities, such as being inventive to improve technology, or investing in equipment, respectively protected by patents or by the Database Directive 96/9/EC. However, in the case of industrial data being collected from pre-installed sensors and data logging equipment on a machine, neither **inventiveness** nor **specific investment** were mobilised, so that there is **no reason** for public authorities to protect it by law. If industrial data collected thus happens to be present where it was not intended to, there is no other recourse than invoking the notion of "trade secrets" protected by the (too broad, and thus rather problematic) Directive 2016/943.

In this policy debate, industriAll European trade union wants the **value added** brought by the usage of Big Data Analytics applied to industrial data to be **fairly distributed**. This value added is the translation, in economic terms, of the improvements in the quality of the product and in the reliability of the process that these techniques can bring.

Reciprocally, industriAll Europe wants to avoid that the know-how and experience of industrial workers in Europe regarding the operation, maintenance and design of industrial processes be transferred from the existing European industrial companies and workers towards Big Data Analytics firms holding monopolistic access to their industrial data.

Considered from the perspective of the public good, industriAll Europe also considers that the **collective benefits** of industrial data are **maximised** when **access** to it is **broadly distributed**. It is when all stakeholders can exploit data according to their needs[4] that they can extract all the information contained in it.

Distributing the value of industrial Big Data fairly: Initial policy reflections

In order to achieve these goals, industriAll Europe recommends that the **legal** and **economic conditions** for accessing and processing **industrial data** be **Fair, Reasonable and Non-Discriminatory (FRAND)**.

In order to concretely operationalise this concept of FRAND economic and legal conditions to access industrial data, industriAll Europe suggests applying the abstract template of **Collective Bargaining** to the discussion between generators and users of Big Data in industry[5].

This would mean that generators of industrial Big Data on the one hand and users of data on the other hand would simultaneously create collective representations of their interests[6]. Collective

[4] E.g. for quality and yield in production companies, for reliability in Mechanical Engineering companies, for scientific knowledge of the process in research laboratories…

[5] A justification for using this template is provided in the Appendix.

[6] A first move in this direction can be identified as the Avanci platform pooling a large number of patent holders in the field of the Internet of Things (incl. Ericsson, Qualcomm, ZTE, KPN, and InterDigital) and providing collective and somewhat transparent terms and conditions for use.

representation of data generators could be built based upon existing sector-specific industry associations. The fairness of procedures would entail transparency of the terms of the discussion and their content, and on the legal enforceability of the collective agreements thus reached. This may imply a revision of competition law regarding cartels.

This approach implies to frame the discussion around the **rights** attached to **accessing and processing data**[7] under a legal regime of **licensing**.

Under this regime, explicit agreements or regulations define who is entitled to mobilise these rights, for how long, where, and for what purpose.

This means specifically that we consider the concept of data "ownership" to be fundamentally **wrong**. The concept of "ownership" is not neutral, it conveys the idea that once the data is "sold", the seller loses any further rights over the way it is further used. In our view, the democratic discussion to take place on the rights attached to industrial data should leave open the option of maintaining a legal connection between the originator of the data (and potentially others), and the data itself, throughout its whole life[8]. The notion of data "ownership" also leads to situations where large players (typically from the immensely wealthy digital economy) would purchase and monopolise data, and then extract a rent from this monopolistic position. As mentioned above, we consider this the worst case scenario for European workers in industry. IndustriAll Europe is therefore **very critical** of proposals to create "*in rem*" property rights[9] on industrial data.

Under our favoured regime of data licensing, the first collector of the data, the entity controlling the physical access to the source of industrial data, could have the specific role of a "**data steward**", in charge of managing the industrial data properly and of concluding the licensing contracts.

The "data steward" would, of course, be entitled to make a reasonable profit from this activity, but would have no monopoly on the usage of the data.

Multiple entities can legitimately claim rights to access and process the same set of industrial data[10].

In order to avoid endless conflicts around equally legitimate claims, we recommend providing a legal framework for **non-exclusive rights over data**, and making this feature of non-exclusivity **mandatory** – in order to cover the largest number of sectors.

Mandating the licensing of data (which does not need to be for free) for the sake of preventing damaging monopolies has a legal precedent, the access to in-vehicle data for the purpose of opening up the market for after-sales services (maintenance and repair of cars)[11]. The obligation to licence should be a sufficiently strong incentive towards fair prices and **fair sharing of the value added**.

---

[7] A list of the rights attached to data, and which could each lead to a specific licensing agreement could be: right to access, to duplicate, to store (and if so, for how long), to modify, to erase, to transfer (and if so, where, to whom, under which conditions, specifically with or without the transfer of the original rights and obligations), to aggregate (i.e. to analyse jointly with the same nature of data collected on other machines / items), to correlate (i.e. to analyse jointly with different nature of data collected on the same machines / items), to exploit (commercially or not, where, when).

[8] This option of retaining a legal connection between the originator of the data and the data itself is made possible by the fact that data, unlike a material item, is a "non-rival" good (the concept of "non-rival" captures the fact that when one entity uses data, it does not deprive any other entity from using it as well, with no loss of benefit).

[9] A right *in rem* is the term used for property rights. Its specific characteristic is enforceable against the world (*erga omnes*) independent of contractual relations.

[10] E.g. from an industrial machine: the manufacturer of the whole machine, in order to improve its design; the manufacturer of each specific module, for the same reason; the maintainer of the whole machine or of each specific module, in order to predict failures and to implement the necessary preventive measures in time; the operator, in order to measure performance and yield; the developer of Big Data Analytics software. This is only an example of the most obvious entities potentially interested in exploiting industrial data. In real life, one could expect the number and variety of legitimate claimants wanting to access a given data set to increase very significantly.

[11] Regulation 715/2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information

Mandatory licencing could be complemented by a **legal framework** regulating the **contractual clauses** of the license (e.g. default clauses and definition of Unfair Contract Terms).

One way to proceed even further in the direction of a broad distribution in the access to and processing of industrial "Big Data" could even be the "**Big data is open data**" principle[12].

These considerations, however, only apply to "raw", unprocessed, sensor- and machine-generated industrial data, resulting directly from the data logging systems present on machines. IndustriAll European trade union considers that the **result** of the **analysis** of such data, based upon human know-how regarding the industrial process, or upon human knowledge incorporated in software, i.e. "processed data", should be protected by a (re-defined, amended and more focused) regulation regarding **industrial secrets**.

[12] Defined in greater detail in our Policy Brief on "the regulation of monopolistic digital platforms", October 2014.

Appendix: Why the abstract template of Collective Bargaining is applicable to achieve FRAND legal & economic conditions to access and process industrial data

In the debate around FRAND legal & economic conditions to access industrial data, we are confronted with **two communities**:

1. the **generators** of industrial data, who want to be remunerated for their investment (even if this investment strongly decreases with the multiplication of low-cost sensors and storage devices); and
2. the **users** of the data, who remunerate the former by exploiting their production.

The debates between these two communities boil down to a discussion on:

- a **price** (the royalty rate – and the royalty base), which entails how the **value added** generated by industrial Big Data will be **shared** between the two communities
- **legal conditions** attached to the use of the data.

The **interests** of the two communities are **structurally opposed** on these topics of discussion:

- generators of data want high prices and strict legal conditions for use
- users of data want low prices and loose legal conditions for use.

However, despite this structural opposition, both communities **need each other**:

- generators of data need the users to create economic value from their production
- users need the data to provide their own customers with the services that they expect.

The usual solution, leaving the problem to individual contracts between parties each belonging to the opposing communities, has very significant economic and social drawbacks:

- it is **economically inefficient**, because it brings a repetition in each individual case of discussions that have already taken place elsewhere, along lines that are absolutely identical, thereby "re-inventing the wheel" and multiplying what is known in economics as "**transaction costs**", and because it introduces a **perverse form of competition** between players, who compete on the price of this essential asset, based on power relationships, instead of competing on the performance of products or processes
- it is **socially unfair**, because it leaves the weaker party in the transaction victim to uneven power relationships that deprive it from legitimate economic benefits – and this is particularly true for the highly fragmented but globally dominant European sector of Mechanical Engineering.

This structure is identical to that encountered around **labour**:

- the two communities are the **workers** and the **employers**: they have structurally opposing interests, but need each other
- direct individual contracts would entail large **power imbalances**
- the terms of the discussion are a price – the **wages**, which translates into how the value added generated by the firm is shared between labour and capital – and the legally-binding regulation of **working conditions**.

The solution developed by the social partners (trade unions and employers' organisations) to solve the question of fairness is **Collective Bargaining**. The concept is that:

- one cannot define *ex ante* what a fair **outcome** will be, but that the effort must go in the definition of fair **procedures**
- the parties negotiate **collectively**, and the outcomes are **legally binding** for all parties represented.

Considering these structural analogies, it can thus be legitimate to propose that the template of Collective Bargaining (originally developed for labour relationships) be used for the licencing conditions of industrial data.