Toward a geopolitics of data

Thierry Berthier, associate professor in mathematics, researcher, Chair of Cyberdefense and Cybersecurity, Saint-Cyr-Thales-Sogeti, and **Olivier Kempf**, PhD in political science and associate researcher at the Institut de Relations Internationales et Stratégiques (IRIS)

For: In Jean-Pierre Dardayrol (ed.) *The European digital union* [special issue of *Réalités Industrielles*, August 2016]

Summary:
Some pundits present data as a resource, whereas others see it as the prelude to a new currency. The Internet of objects will foreseeably multiply the volume of data in comparison with what we now know. Data will become the key element, economically and politically, in our societies. This trend is mainly arising out of private parties' actions, since public authorities (in France and Europe) seem powerless before this new driving force. Although data are becoming the key instrument of power, we still lack an analytic grid for measuring a nation's power in terms of its digital data.

The data economy in numbers

Sometimes considered the new petroleum, digital data modify our perception of the world by opening onto its full complexity.[1] They fuel the engine of an economy that, within in a few decades, has turned into a "knowledge-based economy". Estimated in 2013 at more than €15 trillion,[2] the share of digital products and services in the world's economy represented nearly a sixth of the total (€100 trillion) for traditional goods and services.[3] In May 2015, the accumulated market value of the fifteen leading world groups in digital technology amounted to $2.4 trillion, the same as France's GDP in 2015.[4] Recall that, in 1995, the market capitalization of these top fifteen amounted to but $16.75 billion.

[1] Article translated from French by Noal Mellott (Omaha Beach, France).
[2] "L'économie des données personnelles: les enjeux d'un business éthique", CIGREF report, 2015. Available at: http://www.cigref.fr/wp/wp-content/uploads/2015/11/CIGREF-Economie-donnees-perso-Enjeux-business-ethique-2015.pdf
[3] Éric Péres, rapporteur, "Les données numériques: un enjeu d'éducation et de citoyenneté", the opinions of the Conseil Économique, Social et Environnemental, January 2015. Available at: http://www.lecese.fr/sites/default/files/pdf/Avis/2015/2015_01_donnees_numeriques.pdf
[4] Kleiner Perkins Caulfield Byers (KPCB), *Internet Trends 2015 – Code Conference, Top15 Digital*. Available at: http://www.kpcb.com/internet-trends

World rank (May 2015)	Company	Home country	Market capitalization (US$MM) in 2015
1	Apple	USA	763,567
2	Google	USA	373,437
3	Alibaba	China	232,755
4	Facebook	USA	226,009
5	Amazon.com	USA	199,139
6	Tencent	China	190,11
7	eBay	USA	72,549
8	Baidu	China	71,581
9	Priceline.com	USA	62,645
10	Salesforce.com	USA	49,173
11	JD.com	China	47,711
12	Yahoo!	USA	40,808
13	Netflix	USA	37,7
14	LinkedIn	USA	24,718
15	Twitter	USA	23,965
Total of the Top 15			**2,415,867**
Source: KPCB, *Internet Trends 2015: – Code Conference*			

Table 1: Ranking by market capitalization of global public Internet companies in 2015:
The GAFA extended family

Regardless of their core business activity, industrial firms are organized around information systems, which condition their strategies, survival and prosperity. Built by using entirely new models, “smart” towns optimize their resources and consumption of energy by integrating, at all levels, data and data-processing. Their architecture combines materials and digital data to offer a new kind of space where information is ubiquitous, *i.e.*, accessible everywhere, for everyone and always.

With more than 1.01 billion sensors and connected devices in 2013 and very likely nearly one hundred billion by 2020, a literal deluge of data is coming that will create wealth, needs and new occupations. In 2015, humanity produced, in a single minute, 200 million e-mails, 15 million SMS messages, 350,000 tweets, 250 gigabytes of data on Facebook and more than 1,740,000 gigabytes of digital information worldwide. Google processes daily more than 24 petabytes (*i.e.*, 24 trillion bytes). The diagram in Figure 1 provides an idea of the magnitude of data used for various purposes.

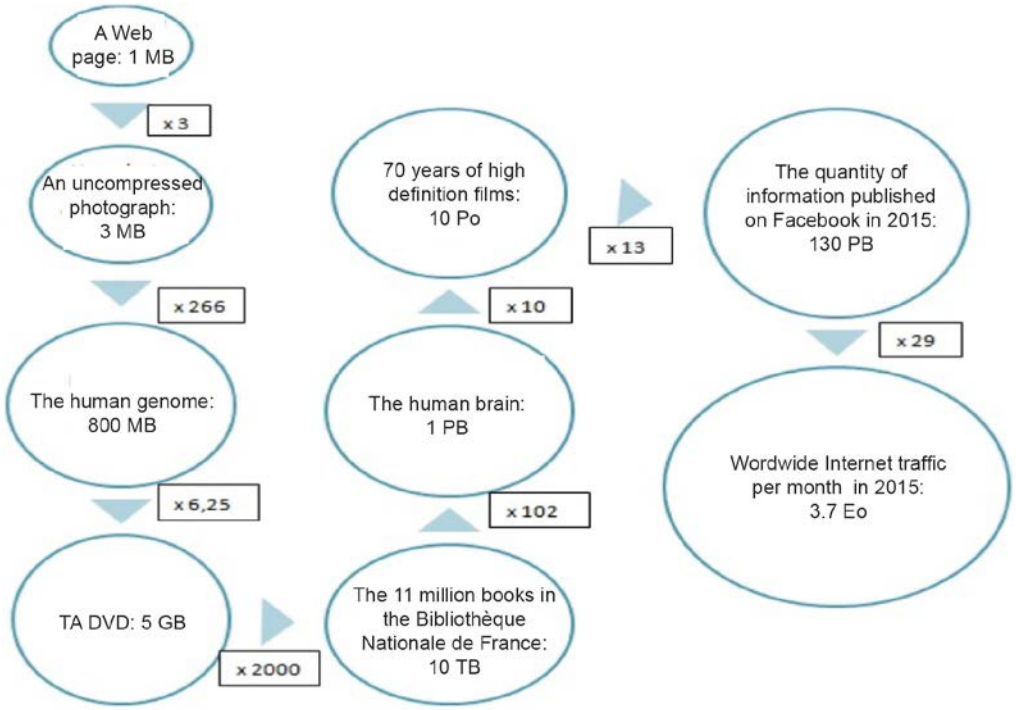

Figure 1: Various natural or digital devices for storing data: Their storage capacity expressed as a power of 10 bytes:
MB= Megabytes (10^6) GB= Gigabytes (10^9) T= Terabytes (10^{12}) P= Petabytes (10^{15}) E= Exabytes (10^{15})

The production of data is now increasing exponentially, but for want of an adapted algorithmic infrastructure, only a part of its potential has been processed. Firms and researchers in the data sciences should devote less attention to storing this mass of data and more to processing it.

These few statistics describe the powerful impact of digital data on the world's economy. However they should not keep us from seeing that this impact has set off tremors reaching beyond the economy into the realm of strategy and geopolitics. Data, as a "virtual" resource, bestow power on those who know how to collect and use them. They become a top-priority issue in the game of power among nations.

Politic power has very little power over data

Data have evidently become a stake in the game of power. This power is, first of all, commercial, since major players make profits from processing big data. By collecting and statistically analyzing huge volumes of data, they intend to draw profiles of consumers and identify their centers of interest. This profiling has an obvious market value since it helps to hone and fine-tune advertising campaigns. Big firms with masses of data can, after processing them, sell "consumer profiles" to other businesses, which can thus optimize their spending on adverting by increasing sales. The companies that have advanced the farthest in this direction are Google, Amazon, Facebook and Apple (the GAFA), to which Microsoft is ever more frequently added. But other players have started doing the same: those directly involved in the Internet (active in developing browsers, in providing access or in telephony) and even those who run big commercial websites (SNCF, Le Bon Coin, and major groups in the media industry).

For public administrations, the stakes are quite different, since the intent is to make available to the public what has been called “open data”.

Big data has become essential to espionage. It should be pointed out that the secret services of major countries (the NSA in the United States, GCHQ in the United Kingdom or DGSE in France) use the same techniques to monitor illicit trafficking activities and to locate potentially dangerous persons.

At present, it is mostly private parties that control data. There are practically no public regulations. We are now observing the first conflicts of interest between private and public interests.

The contention that set the Federal Bureau of Investigation at odds with Apple is telling. One of the terrorists who led, in December 2015, the attack at San Bernardino, California, used an iPhone. The FBI wanted access to the data stored on the phone, but it had been encrypted using Apple’s technology. Apple refused to cooperate with authorities for two reasons. First of all, out of principle: it argued that its services to customers was based on the absolute confidentiality of their messages (thanks to encryption technology). Secondly, out of necessity, for a technical reason: the substantial cost of having so many engineers work on unlocking the phone. Apple essentially refused to set a precedent by invoking fundamental freedoms and, concomitantly, using a clever market-based argument. It thus struck a posture in defense of citizens and their data against a Big Brother government. The company received, in fact, unanimous backing from the major American Internet firms. WhatsApp even announced, a few weeks later, that it was going to launch end-to-end encryption. We know the rest of this story: the FBI used professional hackers to “break” Apple’s code and gain access to the coveted data.

This affair draws our attention to a dual aspect of data: they might be massive but they are also individual. Governments only have indirect means for accessing data, through technical espionage or targeted intrusions (not always easy to orchestrate). Moreover, data bring to light an opposition between two types of power: the one technocommercial and the other governmental. Whereas nation-states do not yet have control over data, technocommercial companies have gained strength and power thanks to their skills in data-processing.

Another dimension of the data economy is visible, but subjacent: how to protect data. Public authorities tirelessly issue warnings about cybercriminality, and exhort cybernauts to keep their computers “clean” so as to lower the exposure to piracy. One method for protecting data is to encrypt data and messages, but this arouses the ire of authorities, since criminals and terrorists are also able to use these techniques. Authorities are, therefore, calling for restrictions on encrypted data. Internet providers easily argue that it is not because cars are used by criminals that they should be outlawed. For them, this analogy holds for encryption.

Public authorities (in particular American) seem to suffer from a sort of schizophrenia — evidence of their dissatisfaction with being unable to exercise control over a new, expanding field that everyone now sees as the grounds for establishing power.

France and Europe are lagging…

The debate on these issues is mostly taking place in the United States, the front runner in the race, both technological and economic, toward cyberspace. Too often adopting an inadequate intellectual approach, France and Europe seem missing.

In France, a country with a Colbertist tradition, making the connection between data and power often leads to evoking the concept of economic sovereignty. One of the keys to success is to understand that power must be, first of all, economic before turning political. Major projects have been launched that leave the public (precisely consumers) unconvinced, even though some of them bore results — such as Minitel, which made France one of the most connected countries in the world during the 1980s. To keep control over the Minitel technology, there was a refusal to let it evolve; to keep the standards from being diffused, a refusal to open the source code — these were some of the

reasons why the Minitel adventure came to an end. Recently, public authorities launched plans for a "sovereign cloud", which has, however, sunk into the technocratic quicksands. The only French operators in the cloud are now private (OVH or Gandi). Likewise, authorities have touted the browser Qwant, which is struggling to carve out a share of the market. The only innovation still surviving is Uhuru, an antivirus software with limited diffusion despite its obvious qualities. In fact, success is a matter of use, which, in turn, is a response to expectations.

Nor has the European Union been very effective, even though this scale of action should be conducive to mass solutions. With a population of 450 million wealthy, educated, connected inhabitants and despite its institutions (which primarily set legal regulations and standards), the EU should provide the stimulus for the emergence of strong digital operators. For the time being, this is not at all the case. We might mention, at most, Junker's program (in favor of economic development) or the recent vote by the European Parliament, which, when adopting the PNR (Passenger Name Record) directive, tacked on a set of protection safeguards for users' data. The Commission has just announced its intention to develop a sovereign scientific cloud, the first step toward launching a European cloud.

France or Europe will acquire power over data only if they develop major operators that will serve as reference marks in the digital economy.

These brief comments paint a contrasting picture, one that is inadequate since we lack a detailed, comparative knowledge of the determinants of digital data power. Let us examine a few aspects…

A nation's digital data power

What are the grounds of a nation's digital power? The determinants, diverse but usually interdependent, correspond to five major capacities that a nation has (or has not) fostered over the past two decades: the capacity of the data infrastructure, the capacity for data-processing, the capacity for attracting the big international players in the digital data business, the capacity for training data scientists and, last of all, the capacity for setting as a priority an ambitious national data policy.

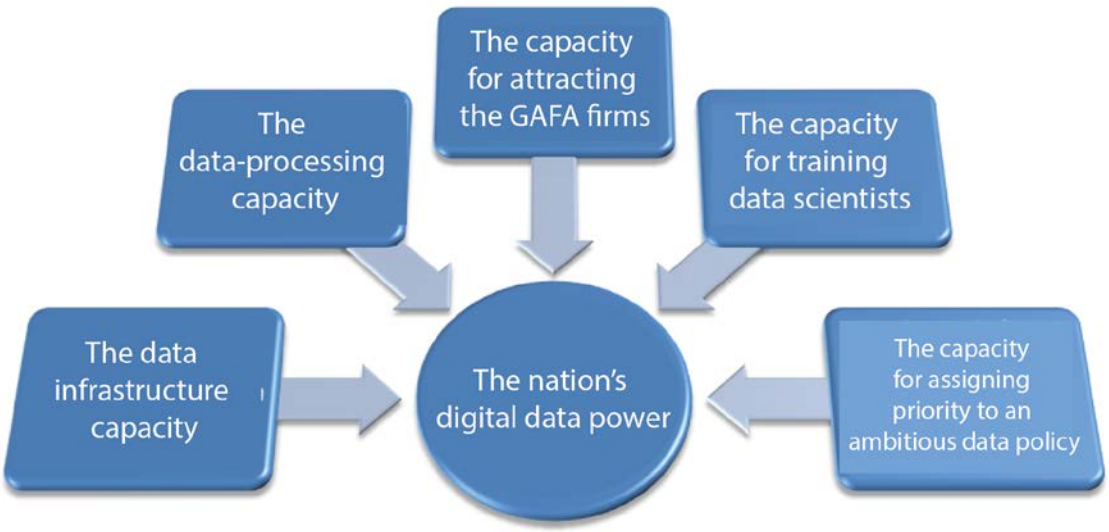

Figure 2: The determinants of a nation's digital data power:

The data infrastructure capacity

This capacity is measured by the density and power of the physical infrastructure inside the nation's border for storing and processing data. It can be evaluated by taking into account: the number of data centers in relation to the national population; these centers' maximal capacities per year for storage and processing; their total annual consumption of energy; the number and ranking (as a function of computing power) of high performance computers installed in the country; and the volume of data processed annually by the country's major operators (telephony, the Internet and networks).

The data-processing capacity

This capacity reflects both how much the tools, developed in the country, for algorithmically processing data are operational and how much the nation depends on foreign (in particular American) installations. It takes into account whether or not a sovereign cloud (durable and secure) exists or is being developed and whether or not a national search engine exists or is being developed that can rival the major commercial ones (Google, Yahoo, Yandex, Baidu, etc.). It also takes under consideration in the country: the number of laboratories, R&D clusters and university research chairs for the data sciences; the number of patents (per year and in relation to the national population) filed in the data sciences; and the number (per year and in relation to the national population) of doctoral dissertations defended in the data sciences. It also brings into the picture: the number of companies (whether startups or big groups) per year receiving citations or awards for their innovations during major international events, shows or competitions; the number of data-centered startups acquired by the big international operators (GAFA); and the number of partnerships between local companies (startups or large groups) and these big international players.

The capacity for attracting the big global players

This capacity has to do with the degree that the country's digital ecosystem attracts the big global players in the data business, in particular the Top 15. To evaluate it, we must look at the operations (research, data and training centers as well as sales outlets) set up inside the country by these global players. We must take into account: the number per year of R&D partnerships formed by the big global groups with research laboratories or high-technology clusters inside the country; the volume per year of acquisitions by these global groups of start-ups inside the country; and the equity investments made by these global operators in innovative national programs. Attention must also be given both to the degree of compatibility of the country's legal environment with the business activities of these big global players and to the level of acceptability of taxes for them. A last point is how (positively or negatively) these big global players perceive the government's digital policies. A question to summarize this point: can national digital policies stimulate investments, durable investments, in the country?

The capacity for training data scientists

The nation's capacity for training data scientists depends on: the national distribution of higher education programs in the data sciences (statistics, machine learning, data-mining); their international reputation; the selectivity of their admission procedures; the number of partnerships between these programs and firms; and the number of firms created by students in these programs before or after graduation.

The capacity for assigning priority to an ambitious data policy

This capacity refers to the priority that the government assigns to digital policy and its measures for fostering a digital ecosystem. The percentage of the GDP reinjected in R&D on digital data is, for sure, the most relevant aggregate criterion for quantifying priority.

A quantitative approach to these five major capacities entails defining formal, numerical indicators. This fifth capacity very much affects the first four but is apparently the least easy to evaluate. In effect, the concept of priority is volatile, fluctuating with the government in office.

To conclude…

Caught up in worldwide competition for the control of big data, France and Europe already have networks of excellence, top-quality technology clusters, scientists, academics and businesses. Coordinating and mutualizing their efforts is the way to create the conditions for competing effectively with the big American and Chinese operators in collecting and processing data. This European cooperation must be set up as soon as possible if EU member states really want to maintain or restore, in part, their national sovereignty by increasing their digital data power. Data give power — let us have the capacity to conquer!