The Digital Era 1

Big Data Stakes

**Edited by
Jean-Pierre Chamoux**

The Digital Era 1

The Digital Era 1

Big Data Stakes

Edited by

Jean-Pierre Chamoux

ISTE

WILEY

Contents

Note to Reader

Data telecommunications and digital processing are omnipresent in today's society. Administrations, businesses and even leisure activities produce, disseminate and exploit data related to human activity and trade. These data, the raw material of information industries, are vital for our society. This is why our era can rightly be called the "digital era".

Many experts were consulted prior to producing this series. The main idea took three years to prepare. The overall content, which was finished at the beginning of 2016, includes three successive volumes. This collective work is a trilogy that aims to describe and understand the technical, economic and social phenomena that result from widespread use of the Internet, a digital network that has been present everywhere since the end of the 20th Century.

The first volume summarizes the state of play and issues raised by the enormous amount of data that accompanies human activities: demographic, biological, physical, geographical, political, industrial, economic and environmental data. This data feeds into records, inspires people, guides their businesses and even their countries. This first volume sheds light on the practical, technical and methodological advances that are associated with the Internet and big data.

The second volume explores how and why the digital era is transforming commercial trade and interpersonal relations, as well as our living conditions. How are the media, commerce and trade evolving? How is wealth formed and transmitted? What causes the digital wave in economic and social life described in the first volume? According to Adam Smith, *The Wealth of Nations*[1] has been the

---

1 Smith A., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Campbell R.H and Skinner A.S. ed., Oxford U. Press, 1976.

core of political economy; is it shattered or transformed? This second volume emphasizes that the *new economy* is not stable. Its social dimension is still unclear. But thinking is progressing: didn't it take a good century for the political economy to adapt to industrial society? Why wouldn't it be the same for a knowledge-based society to be delivered, one that even futurists in the 1970s[2] had already announced?

The third volume attempts to summarize the questions that the digital age suggests to our contemporaries: questions about society, interests and politics, which were partly mentioned in the first two volumes, are reformulated and developed in this last volume in order to encourage the reader to reflect and contribute to the debate on the "issues of the century", as Ellul said in 1954[3]. A debate that will hopefully help us understand the digital society that is being built before our very eyes; and perhaps it will help us to get the best and not the worst out of it...

This series is based on competent, attentive and precise specialists. The authors wrote freely, as they should; they inserted their texts within the proposed outline. We owe them a great debt of gratitude. May they be sincerely thanked here for their scholarly assistance. As for the shortcomings, as is customary, only the project coordinator will be accountable[4].

---

2 Kahn H., Wiener A., *The Year 2000*, McMillan, New York, 1967 (introduction by Daniel Bell).
3 Ellul J., *La technique ou l'enjeu du siècle*, Armand Colin, Paris, 1954 (reissued by Economica, Paris, 1970) [English translation: *The Technological Society*, Knopf, New York, 1964].
4 Editor's notes in the book correspond to comments by the coordinator.

Preface

Understanding Digital Society

CONIUGI DILECTISSIMAE

During my childhood, only the local bar, the baker, the grocer, the mayor, the doctor, the notary and the veterinarian of the village had a telephone installed in their homes, mainly to deal with the necessities of their trade. Only a dozen telephone sets existed in our small town in the French countryside. The telephone was rare, expensive, reserved for supposedly important uses, which only a very small minority of users were privileged to have. Less than 20 years later, Fernand Raynaud's memorable sketch, which marks the memory of our generation and even the next one that follows, stigmatized this incredible under-equipment in France: in 1966, while everyone knew *Number 22 at Asnières*[1], our family home still had no automatic telephone, neither did the neighbors, nor did the village cafe. It was not until the government promoted telephone catch-up as an essential and urgent national priority through a huge advertising campaign that the former subscribers of the manual network were finally provided with automatic equipment, which was only established from 1974 onwards in our rural community[2]!

---

1 The French title of Raynaud's short story: "*Le 22 à Asnières*".
2 Manual network communications were set up by the "telephone ladies", employees of Post Offices who connected the subscriber by hand to the recipient's line; these operators could therefore follow all conversations and especially in rural areas the exchange between correspondents was hardly discreet. However, this was still only the time of the local "small ears"; the process has since then improved considerably with the Internet, taking on a global dimension, the "big ears" that will be mentioned in Volume 3.

Thirteen years on, another telephone revolution took off in France, albeit several years later than in countries such as Sweden: wired telephone equipment having become the norm, a first cellular telephone license, prepared by the team I was coordinating at the French Ministry of Post and Telecommunications, was awarded to a private operator, SFR, at the end of 1987; the first customer of this commercial company was only served in the first quarter of 1989. At the same time, the Post and Telecommunications Administration began the transformation that would split the Post Office and telecoms. After several intermediate stages, France Télécom SA was created thanks to a law in July 1996. In the meantime, the GSM digital cellular service had been put into service by France Télécom in July 1992; it was the first link in a long series of digital communications that gradually equipped the planet and made the telephone what it is today, everywhere in the world: a digital smartphone[3], a multifunctional device that has sold billions worldwide in a just few years!

What a transformation of objects, techniques and modes of interpersonal communication in barely 20 years! But the story does not stop there. At the beginning of the period that is today glorified by history as the one where everything is (supposed to be) possible[4], the students that we were had two traditional instruments for calculations, in addition to our own aptitude for mental arithmetic: a table of logarithms and a slide rule which, just to clarify, is only a personal calculator based on logarithms! These two calculation assistants, which are totally forgotten today, refer not to actual figures, but to analog calculation[5]: add two lengths of a ruler, on a logarithmic scale, for multiplication; or subtract, on the contrary, two lengths of the same ruler to do a division.

Analog calculation has proved its worth for a long time. For the practitioner, it entailed two important consequences: the obligation to remain wary, in any calculation, of the orders of magnitude, and at the same time, to train the calculator, whether he was an engineer, biologist or architect, to remain aware of the error implied by any measurement. In this case, inevitable approximations upon reading

---

3 An equivalent proposed by French *ad hoc* terminology commissions: "ordiphone", a term that use has not endorsed.

4 The years of industrial reconstruction of Western Europe continued after the Second World War; see Fourastié J., *Les Trente Glorieuses*, Hachette, Paris, 1979.

5 Designed and developed in Europe around 1630, this method of calculation was used for over three centuries (mid-17th Century to mid-20th Century) and thus accompanied most of the industrial revolution up to modern times. It remained the only authorized calculation assistant in official examinations and competitions in France until the mid-1980s.

the ruler scale. None of my children knew, during the course of their studies, how to use a slide rule; from high school onwards, they were all allowed to have an electronic pocket calculator; for them, the digital age began with their entry into sixth grade! As for current students in high schools and colleges, my grandchildren's generation, they now take their classroom notes on a tablet, on a smartphone or a laptop. Nobody pays any attention to analog calculation any more: the object, the industries that manufactured them, the logarithm tables printed in the minds of thousands of past high school students have been wiped from existence. In the 1980s, one stroke of a pen drawn by technical progress dissolved the small precise mechanics that manufactured our slide rules. This disappearance heralded, with our lack of taking any serious precautions, the digital disruption that we will find in almost all the chapters of this series, as a consequence of the disruptions that are now largely felt in the economy, trade, education, health and transport.

After my studies at the *Ecole Centrale*, I discovered modern digital calculation when I became a Fullbright scholar at the American university where I was able to program, for the first time, those IBM computers that none of our great schools had yet installed, to make their students discover this new instrument that would, in a few years, flood into companies, transform business management, industrial research, physics, chemistry and biology in less than half a generation. For us, it was a matter of simulating, by digital calculation alone, the physico-chemical phenomena preparing industrial production. When I returned to the Parisian Faculty of Science laboratory, we had to use resources that were occasionally available in Orsay to be able to do calculations that were important at the time, because Parisian universities were generally badly equipped, a bit like the retorts and lab benches used by Pasteur or even Lavoisier!

Over the next 10 years, I then witnessed the birth and growth of a generation of management programmers who trained and oversaw the administrative and commercial management of public services (urban planning, land use planning and social services) and businesses: insurance, banks, pension funds, industries, transport and cities. They were my traveling companions and fellow students, and made the strength of the service, research and IT consulting companies (IT Services) who, for more than a quarter of a century, have branded European computing and software engineering. These businesses, which were competitive and rapidly growing, quickly opened up to the international market, maintaining a competitive advantage for a long time. They converged from the mid-1970s with the transmission of remote computing to produce, around 1980, the famous report to the President of the French Republic, christened with the forged name "*Télématique*" that permanently marked the coupling between telecommunications and information technology: telematics (Nora-Minc Report, 1978). Let us make no mistake, this was

the first tangible manifestation of what would then be called "digital convergence", the scope of which we have been constantly expanding since the Internet began to emerge around 1995.

Let us sum up our observations: during a lifetime like mine, a multi-centenary calculation method, the analog calculation, has fallen into oblivion. Only the digital calculator is now used. This changeover, in the time span of just two generations, has two consequences that deserve to be considered:

– a profound change of perspective on measurement and the instruments we use to measure;

– the symbolic overestimation of numbers in our understanding of the world and nature.

By becoming purely digital, our tools now detail each measurement by aligning numbers that are not necessarily significant for our practical needs (temperature, pressure, distance, volume, speed, percentage, etc.): do we really need to know body temperature to two decimal places to diagnose a fever? Is a measurement of land surface per hundredth of a square meter necessary for a real estate agent? Does displaying the result of one-tenth of a percent on a political popularity survey mean anything when you know the level of statistical error associated with the percentages calculated from a survey of 1,600 interviews (approximately +/- 2.5%[6])? The race for display accuracy implicated by the digitization of our tools, to evaluate physical quantities as well as social behaviors is an illusion: accuracy becomes a decoy, a smoke screen that conceals the reality of things, because measurement is only an approximation, an order of magnitude that we seek to evaluate.

We have been going astray for half a century, particularly for those in human sciences. The preliminary thoughts that opened Raymond Aron's 1956–57 lessons titled "*La lutte de classes*"[7] rightly pointed out, with regard to sociological studies: "You can [by survey] determine the proportion of individuals in the community who belong to various classes" (p. 86); and, a little further on: "There will always be a certain difficulty in delimiting (these) groups... the particular of social groups is to be equivocal, to have poorly defined boundaries". So, what is the point of precision, if it is not an illusion? What is important, as I mentioned earlier about the profession of construction engineer, is not to have a precise number, but to identify the order of magnitude of the phenomenon that we observe and describe. I, therefore, insist that

6 See J. L. Boursin, *Les dés et les urnes,* Le Seuil, Paris, 1990, p. 232 ff.
7 Publication titled *Nouvelles leçons sur les sociétés industrielles*, Gallimard, Paris, 1964 (reissued in *Idées*, no. 47, 1972).

with its digital display, this tool is lulling us with illusions, making us believe in a misleading precision to the detriment of what really matters: the appreciation of magnitudes, both in the world of material objects and in the world of living beings with economic and social functions!

Analog calculation required calibration of the measuring tool, in direct relation to empirical knowledge; digital calculation favors a completely different convention that is detached from the sensitive apprehension of our five senses, which are all eminently analogical. By eliminating (or omitting) the doubt and approximation that characterize any measurement, we dissociate the empirical apprehension of an object or phenomenon from its representation by digital measurement; the scientific approach is then more tempted than ever to take off from reality! My point here is not to negate the advent or usefulness of a new time, on the contrary; it is to insist on the fact that the generalization of numbers does indeed induce consequences that have been significant for some time. The more digital calculation is diffused in our society (without us actually looking out for it), the more its effects extend in various forms: changes of reference, of method and reasoning which mark the beginning of the new era. To become aware of this is, without a doubt, a first step towards wisdom!

Why speak any further on this part of life that has nothing exceptional, apart from the fact that it is characteristic of what people of my generation have experienced? To bear witness to the fact that, contrary to certain common thoughts, the diffusion of digital objects, practices and digital uses is neither as brutal nor as recent as many of our contemporaries believe, as they are slaves of novelty at any price. On the contrary, it is a long-running technical, economic and social phenomenon that began in the middle of the 20th Century, in the aftermath of the Second World War. The effects of digital technologies are gradually inducing society as a whole, as the steam engine and electricity[8] did in their time. The transformation of customs, habits and communication between humans is progressing in stages. Change is initially only noticeable in a minority, which is directly affected by emerging technical progress; then it spreads more widely, sometimes under the influence of unforeseen circumstances[9], and the majority of humans become receptive to it because they finally feel the effect of this technical

---

8 Reference is made to the comparative study published by Daniel Sichel under the Brookings Institution stamp: *The Computer Revolution*, Washington DC, 1997; Chapter 5 compares the rise in power of the information, electrical, air transport and rail industries over a long period of time in which each of these industries has matured and taken its place in modern times.

9 Such as those that allowed the extraordinary extension of the Internet to large parts of the world population in recent years!

progress in their daily lives or in their jobs. These ruptures and disruptions will be discussed in the three volumes of this series.

No nostalgia dulls my words; I have no reluctance to adapt to change. On the other hand, I believe that concrete testimonies can, in their diversity and multiple perspectives, usefully illustrate contemporary digital practices and place them in a historical perspective outside of which the observer, even if attentive, risks either losing himself or being blinded by novelty. Moreover, we know this perfectly well, but it is worth repeating: our electronic machines do nothing more than what we have taught them to do: calculate, compare, reproduce, record and transmit the data we entrust to them. Exploiting this data to deduce a better understanding of the world, of social facts and of the entire universe, is at the end of the line because, if the lone data, which is not classified, nor treated, nor interpreted, is worthless, the data which is carefully pursued, put into perspective, can contribute to our knowledge, and enlighten our thinking. It is by observing, for example, the displacements of a population, its uses, its customs and demographic vibrancy that we can begin to understand it.

There is evidence that the digital data collected and made available on a human group supports knowledge and helps prepare for action. However, we must want to act and do so consciously. Let's hope that the work gathered in this series will not only help us to understand the digital age, but also to attain a certain mastery of it!

Jean-Pierre CHAMOUX
February 2018

Introduction

Big Data Challenges

"No man steps in the same river twice...
Nothing endures but change."
Heraclitus

Philosopher, born in Ephesus, Asia Minor, early
5th Century BC. He proclaimed that everything is
perpetual: conflict, movement and becoming!

The youngest of us, our students in particular, who are fortunate enough to live in the pivotal part of this new century, know that novelties follow one another, that one item replaces the other and that commercial propaganda often tries to present an old-fashioned costume by disguising it as new: merchants of illusion never tire of these tricks. Let us learn how to arm ourselves, so that we know what is worth believing! Social behavior is changing, and so are technology and the relationships between people. It is a law of evolution: the beginning of this century has already been marked by the installation of digital tools in our cars, our professions, our homes, in most of the infrastructures of modern life: telephones, entertainment, transport, health and banking, all have been invaded.

What is important to us in the first part of this volume is to appreciate whether or not this omnipresence, depicted in Chapter 1, leads to radical changes or *progressive* evolutions, in every sense of the word: regular transformations and sources of progress. Are its uses and morals affected? What are the underlying domains that facilitate the adoption of digital devices that multiply in our daily lives: television, telephone, electricity meters, games and toys for our children,

---

Introduction written by Jean-Pierre CHAMOUX.

etc.? We emphasized, a few pages earlier, that digital machines are based on calculation and formal logic. For over a century, part of mathematics has focused on the study of numbers, long before computers and huge databases, which required the use of different computing methods than those implemented in the 1970s or 1980s. A foundation of abstract knowledge feeds the powerful software and algorithms that benefit from the accumulated storage of data from space observation, search engines and particle accelerators. Chapter 2, "Mathematical Culture and Massive Data", written by epistemologist and mathematician Jean Dhombres, explains the important contribution of mathematics in this rapidly growing field.

The figures that accompany our daily lives also come from large-scale enumerations, the traditional and ancient model of which is that of population censuses, a practice that dates back to ancient times[1]. Designed and written by statistician, business leader and leading expert in audience research Philippe Tassi, Chapter 3, "From Sample to Big Data: Competing or Complementary Paradigms", illustrates the growing importance of methods used to extend statistics and take advantage of the vast digital content that characterizes the Internet. Analyzing metadata makes it possible to identify and to benefit from sometimes very heterogeneous information lost within a magma of data. Chapter 4, "Researching Forms and Correlations: the Big Data Approach", edited and written by Gilles Santini, an expert on large numbers, surveys and their interpretation, illustrates how to "play with the devil" so that the network's *deus ex machina* can extract useful elements for action from the apparent disorder, to notably inspire commercial uses.

Chapters 5 and 6 were written and edited by Gérard Dréan, a very experienced practitioner in the IT industry and services, whose analysis and inquisitiveness are flawless. Chapter 5, "*Bitcoin*: an Innovative System", describes a virtual universe consisting of 10,000 nodes where operators establish a register and reproduce thousands of copies around the world in order to keep a faithful and lasting record of the commitments already contracted. It is an original invention, intimately inspired by the principles of the Internet: Bitcoin relies on thousands of programmers who share a common goal that of building this unique electronic register that attests to the authenticity, existence, purpose and amount of commitments contracted on the Net by third parties who trust it. These two chapters, linked to one another, form the hinge between the two parts of this book.

---

1 There is evidence of numerous censuses in Imperial China and Pharaonic Egypt, many centuries before Christ. The history of Christianity begins with "the decree of the Emperor Augustus ordering the census of the whole universe... while Quirinius was governor of Syria", a census that forced Joseph to go to Bethlehem where Christ was born, see *Luke, 2, 1-4.*

In the second part of this volume, we have brought together concrete cases that illustrate tactical or strategic challenges that characterize the contemporary digital world. Chapter 6, "*Bitcoin* and other Cyber-currency", focuses on the financial applications of this cooperative, peer-managed system. Fulfilling both the role of an exchange platform and a financial institution, Bitcoin and its emulators emit cyber money. The operators, paid on assignment, receive the counterpart of their contribution in bitcoins for this cross-border job. Multiple projects of a similar nature follow the same itinerary, and only the future will determine whether or not they will contribute to inventing new monetary mechanisms[2].

Chapters 7 and 8 elaborate on the very significant case of healthcare and social benefits which are all preparing themselves, in no apparent order, for the digital era: benefits, hospitalization, medical diagnoses and treatments, pension schemes and social coverages generate enormous flows of data relating to patients, contributions and the living conditions of the population. This heterogeneous whole produces and absorbs a great deal of information about everyone; the corresponding cash flows exceed, in most modern countries, government spending. Chapter 7, "Health and Care: Digital Challenges", was written and edited by Isabelle Hilali, who has been working on integrating digital technologies into healthcare systems for many years. The author provides a general overview of the applications, present and future, within this vast, disconnected and costly ensemble. Chapter 8, "Access to Health Data: Debates and Controversies in France", focuses on the conditions of access to data in a country like France, and describes the exploitation and potential value of existing metadata, to inform as much as generate new research. Based on the literature review she conducted with Gabriella Salzano and Christian Bourret, Joumana Boustany, in this chapter, points out that access to this data is limited, even though exploiting it could lead to the emergence of knowledge and actively contribute to the economy.

Chapter 9, "Artificial Intelligence: Utopia or Progress?", finally attempts to ask in weighted terms the questions raised for decades with the potential convergence between digital machines, software and human intelligence. For centuries, the construction of robots and machines has fascinated mankind; this passion is once again of topical interest, since the joint progress of biology and cybernetics makes it possible to imagine a hybrid creature that would benefit from the common progress of biology and digital machines. Artificial intelligence inspires fantasies, like the sea serpents from fantastic literature used to, among enthusiasts of the new age. It is never useless to recall implicit theories that inspire today's inventors, just as

---

2 Volume 2 of this series illustrates more broadly the effects of digital disruption on economic paradigms and on notions of value, wealth and growth.

yesterday's; they lend their intelligence to man-made objects: "They are not madmen and their goal is not anarchy, but social stability. It is in order to ensure (this) stability that they carry out, by scientific means, the ultimate revolution... in the meantime, we are in the first phase of what is perhaps the penultimate revolution". The best of all possible worlds[3]?

---

[3] Quote taken from the Foreword to Huxley A., *Brave New World*, Vintage Classics, New York, 2007 [French translation reviewed by the author: *Le meilleur des monde* Presse Pocket no. 1 438, Paris, 1972, p. 12]. Huxley was French-speaking; he dedicated his work to Voltaire's *Candide*, whose famous aphorism justifies the French title: "everything is for the best in *the best of all possible worlds"* (our translation).

What's New and Why?

Introduction to Part 1

The information industries have enabled the collection and exploitation of huge data storages that represent the world around us. What used to take months of thankless work to solve a combinatorial problem, sort sources or describe a natural phenomenon in detail is now done in a matter of moments, provided adequate hardware, software and an effective method is used. And, if there is no instrument, if the search engine returns empty handed at our request, nothing prevents us from venturing out, to create what is missing in order to benefit others! The digital wave is transforming the way we act, observe, describe and explain facts, living beings and nature. It favors those who are able to imagine how to gather, process and exploit the resourceful information that hangs around us. Fortunately, digital machines are doing these recurring tasks quickly and well, as we are mentally incapable of managing the abundance and complexity of raw information. The man-machine relationship has been taking advantage of this natural complementarity for the past 50 years, and still retains a great deal of potential for invention.

The Internet broadens and prolongs this trend. Of course, the ubiquity of the network multiplies data, diversifies sources and facilitates its exploration; the web further increases the complexity of the digital world. But, alongside our electronic assistants – phones, tablets and computers that store images, sounds, texts, charts and databases that accumulate on our virtual desktop – the field of exploration and discovery is immense. That's why so many people are rushing to the El Dorado of data that their great predecessors have been ploughing through for the past 15 years: Apple, Amazon, eBay, Google, Leboncoin, Microsoft, Oracle, Symantec and many others, the overwhelming majority of whom were established in North America. In this context, there is still much to be done and much to invent, since change is rapid. The flow of innovation does not dry up: technical discoveries, certainly; but also method or process improvements, transfers of skills and expertise, geographical and

sectoral, etc. Creative imagination is expressed everywhere and often finds the human and financial resources needed to attempt the venture.

The five chapters that follow describe the challenges of such ventures, whose impetus comes more often from the mind than from the hand: from mathematics to medical practice and pharmacopoeia, the field of investigation is very broad and the forms of inventive cooperation are very diverse, as demonstrated on a global scale by the extraordinary success of blockchains, a venture which is located in the antipodes of the multinational giants frequently associated with the digital industry and America. The games are afoot: so, let's get to work!